FDA/AACR/ASCO
Public Workshop on Brain Tumor Clinical Trial Endpoints

January 20, 2006
Bethesda North Marriott Hotel and Conference Center
North Bethesda, Maryland

Meeting Summary

INTRODUCTION (Dr. Richard Pazdur, FDA)

Dr. Pazdur welcomed everyone in attendance and noted that the purpose of the meeting was to have a wide-ranging discussion about the positive and negative aspects of various endpoints for trials intended to support the approval of new drugs to treat primary brain tumors. This workshop is the fifth in a series evaluating potential endpoints for drug approvals in the most common cancers. Previous workshops have considered endpoints in lung, colon, and prostate cancer and acute leukemia. Issues highlighted at these workshops are subsequently discussed at meetings of the Oncology Drugs Advisory Committee (ODAC), the FDA's statutory advisory body on issues related to oncology drugs.

The primary focus of the discussion should be on endpoints that are ready for incorporation into clinical trials now or in the near future. Workshop participants may identify key issues and areas in which knowledge is limited and may recommend issues or questions for further study. However, it is not the workshop panel's task to make recommendations or arrive at definitive conclusions and no votes will be taken. By law, FDA may take advice only from its statutory advisory committees.

Dr. Pazdur acknowledged that a tremendous need exists to develop new agents for the treatment of brain tumors, that many methodological hurdles need to be overcome in the validation of radiographic endpoints and patient-reported outcomes (PROs) for this type of tumor, and that clinical trial design issues also need to be addressed.

FDA has issued an overarching guidance document on endpoints for registration trials (*Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products*; May 1998; available at http://www.fda.gov/cder/guidance/1397fnl.pdf) and intends to supplement this document with guidances focused on specific tumor types.

The final hour of the workshop will be chaired by representatives of the National Cancer Institute (NCI), who will lead a discussion aimed at identifying areas where further research is needed.

REGULATORY BACKGROUND (Dr. Edwin Rock, FDA)

Dr. Rock briefly reviewed the key pieces of legislation that established the framework for drug regulation in the United States: the Pure Food & Drug Act (1906); the Food, Drug, and Cosmetic Act (FDC, 1938); and the FDC Amendments (1962). The 1962 FDC Amendments for the first

time required sponsors, prior to marketing a new drug, to submit data documenting "substantial evidence of efficacy in adequate and well-controlled studies."

In most cases, efficacy is considered equivalent to clinical benefit. FDA's view of what constitutes clinical benefit has evolved over time. Currently, clinical benefit can be summarized as either longer life or better life; the latter is usually indicated by a direct measure of how the patient feels or functions. Clinical benefit can also be reflected by a surrogate that is not a direct measure of benefit.

In 1992 FDA introduced an alternative pathway to drug approval that is based on surrogates for clinical benefit. The accelerated approval (AA) mechanism was intended to speed medicines to market for serious or life-threatening diseases when an improvement can be shown over available therapy. A drug sponsor may apply for AA based on the demonstration of a favorable effect on a surrogate endpoint that is considered *reasonably likely to predict clinical benefit*. As a condition of approval, the sponsor must agree to provide additional data confirming clinical benefit, which may be generated either by another trial or by a distinct endpoint later in the same trial. Most AAs of cancer drugs have been granted on the basis of a demonstrated tumor response in a refractory setting, often supported by additional information.

For regular drug approval in oncology, survival is the undisputed "gold standard" for evidence of clinical benefit. During the 1990s, survival accounted for about one third of all cancer drug approvals. Demonstration of a favorable effect on how a patient feels or functions, measured by a valid, clinically relevant instrument, can also support regular approval. For example, mitoxantrone was approved for the treatment of hormone-refractory prostate cancer solely on the basis of pain relief, which was defined as a 2-point increase on a 6-point pain scale lasting at least 6 weeks.

Some surrogate endpoints have been accepted for regulatory purposes and may be used as the basis for regular drug approvals in oncology. For example, durable complete response is an accepted surrogate in acute leukemia; partial response is an accepted surrogate for approval of hormonal agents to treat metastatic breast cancer; and disease-free survival is an accepted surrogate for drug approvals in adjuvant breast cancer therapy.

Strengths and Weaknesses of Accepted Oncology Endpoints

Survival. The strength of survival as an endpoint is that is it unequivocal and easily measured. However, trials in which survival is the primary endpoint must be randomized, require a large sample size and lengthy follow-up, and are expensive. Another potential problem is that any beneficial effect of the experimental therapy may be "washed out" by crossover from the control arm to the experimental arm of the trial. This is usually more of a problem when the treatment effect is modest.

Response rate. Radiographic response rate is a surrogate endpoint that is unique to oncology. In the 1990s response rate was the basis for about half of regular approvals and almost all AAs. The strength of response rate as an endpoint is that tumor size reduction can be attributed in its entirety to therapy, whereas both survival and progression-free survival (PFS) are influenced to

some extent by the natural history of the disease. However, the response must be durable and the necessary duration of response is context-specific. It can be difficult to weigh the importance of a partial response vs. a complete response. In addition, response rate does not take into account stable disease, low-level responses that do not meet the criteria for partial response, or baseline disease burden.

Response rate can be effectively assessed for regulatory purposes in a single-arm trial. Acceptable criteria for response, stable disease, and progression must be defined prospectively. Response rate is more credible when supplemented by additional evidence of clinical benefit such as symptom improvement.

Progression-free survival. A strength of PFS as an endpoint is that the sample size and follow-up period are generally shorter than is necessary to show a survival benefit. Additionally, differences in PFS are not obscured by secondary therapy even if a crossover effect exists. Finally, PFS takes into account the potential toxic effects of therapy. However, because of the potential for bias in the interpretation of disease progression, trials in which PFS is a primary endpoint must be meticulously designed and executed and interpretation of progression must be blinded.

Symptom palliation. It is generally accepted that palliation of disease symptoms represents clinical benefit. About one fourth of drug approvals during the 1990s were based in part on symptom palliation. Symptom palliation is not synonymous with global measures of quality of life (QoL) ; the latter has not yet been accepted as the basis of any drug approval in the United States.

Symptom-palliation endpoints can be challenging to use. The development of symptom-palliation measurement instruments must be hypothesis-driven and validated. A measurement instrument's validity is easily compromised by trial design issues or by problems in execution. The credibility of symptom-palliation endpoints can be enhanced by blinding and by association with a biological effect of the drug such as response rate.

Trial Design Considerations

Randomized trials are invaluable for establishing the magnitude of a treatment effect and providing a thorough safety assessment. Blinding is essential whenever bias in measurement or interpretation could be an issue. Measurements must be clinically relevant with explicitly defined prospective analysis. For psychometric instruments and PROs, the concept underlying the instrument must be identified and mapped onto discrete elements of the measurements.

Approvals of Drugs to Treat Primary Brain Tumors

Several challenges have limited the development of effective new therapies to treat primary brain tumors, including the chemoresistance of brain tumors and problems with drug delivery to the central nervous system. Nonetheless, several drugs have been approved to treat this group of diseases:

- Nitrosoureas of DNA alkylating agents capable of crossing the blood-brain barrier after systemic administration
    - Orally administered lomustine (CeeNu), approved in 1976.
    - Intravenous carmustine (BiCNU), approved in 1977.
    - Both approvals based on tumor response rate (as were all drugs approved for cancer treatment prior to the 1980s).

- Carmustine wafer (Gliadel)
    - Synthetic biodegradable polymer impregnated with carmustine.
    - Approved in 1996 for treatment of recurrent glioblastoma multiforme (GBM) as an adjunct to surgery on the basis of a randomized, placebo-controlled trial in 222 glioma patients who progressed following surgery and radiation. Median survival for patients who received carmustine wafers was 7.4 months, vs. 5.5 months for those who received a placebo.
    - Approved in 2003 for initial treatment of high-grade malignant glioma as an adjunct to surgery and radiation. The basis of approval was a randomized, placebo-controlled trial in 240 patients with newly-diagnosed, high-grade glioma undergoing resection craniotomy. Median survival for patients who received carmustine wafers was 13.9 months vs. 11.6 months for those who received a placebo.

- Temozolomide (Temodar)
    - Orally available alkylating agent chemically related to dacarbazine.
    - Granted AA in 1999 on the basis of five durable complete responses among 54 patients with aplastic astrocytomas refractory to both nitrosoureas and procarbazine.
    - Granted regular approval in 2005 after confirmation of clinical benefit was obtained in a trial of 574 patients with newly diagnosed GBM. Patients were randomized following surgery to adjuvant radiation alone or radiation plus temozolomide followed by maintenance temozolomide for 6 months. Median survival was prolonged by 2.5 months in the temozolomide group.

Charge to the Panel

The panel was asked to discuss potential nonsurvival endpoints that may either directly represent clinical benefit or, as potential surrogates, be reasonably likely to predict clinical benefit in primary brain tumors. Questions that should be addressed included the following:

- Are the endpoints analytically valid and/or clinically relevant?
- Are the endpoints now or could they soon be useful, either individually or as composites, for establishing safety and efficacy, therefore supporting the drug approval process?

OVERVIEW: CLASSIFICATION AND TREATMENT OF PRIMARY BRAIN TUMORS; ISSUES AND EFFICACY ENDPOINTS IN GLIOMA CLINICAL TRIALS

(Dr. Howard Fine, NCI)

Primary brain tumors are the leading cause of cancer-related deaths in children and the fourth leading cause of cancer-related deaths in people under the age of 54, Dr. Fine said. A significant increase in the incidence of brain tumors has been observed in people over the age of 60, although the extent to which this observation is an artifact of increased screening remains a matter of debate.

Brain tumors are of several different types, each with a distinct biology. Most of today's discussion will center on gliomas, the most common type of primary brain tumor. Other types of primary brain tumors include embryonal tumors (e.g., medulloblastomas), tumors of the lining of the brain (meningiomas) and tumors of the peripheral nerve cell sheath (e.g., schwannomas, neurofibromas). Brain metastases of systemic tumors present a different set of issues with regard to clinical trial design and will not be discussed today.

Current Treatment Options for Gliomas

Gliomas may be subdivided into benign (World Health Organization [WHO] grade I or "low grade") and malignant (WHO grades II to IV or "high grade") tumors. Radiographically complete surgical resection is generally considered optimal treatment for low-grade gliomas. Radiation therapy can halt disease progression for a time and probably increases survival; issues such as timing, dose, and volume of radiation therapy remain unresolved. The risk of long-term radiation-induced neurocognitive deficits is a significant concern as patients with low-grade gliomas generally live longer than those with malignant tumors. Interest is growing in the use of chemotherapy with agents such as temozolomide to delay radiation therapy. Radiographic responses are possible in patients with low-grade gliomas who receive chemotherapy, but no long-term outcome data are available.

For high-grade gliomas, complete surgical resection is generally considered optimal treatment, although only retrospective data support this. Radiation therapy remains the foundation of treatment. Long-term neurocognitive deficits are less of a concern than in low-grade tumors because patients usually do not live long enough to experience this toxicity.

Three meta-analyses have shown that post-radiation chemotherapy results in a small but statistically significant survival benefit. The definitive European Organization for Cancer Research and Treatment (EORTC) trial showed a benefit for temozolomide given either concurrently with radiation therapy or after radiation therapy to patients with GBM; median survival was increased by about 2.5 months and 2-year survival by about 18%. Two trials of carmustine wafers have shown small but statistically significant increases in survival in both recurrent and newly diagnosed GBM.

With no treatment, median survival from the time of diagnosis for patients with malignant gliomas is 3 months. Surgery may extend median survival to 4 or 5 months; adding radiation to surgery extends it to 10 months. Adding temozolomide chemotherapy to radiation and surgery has now extended median survival to 14 months. Existing therapies are clearly of limited effectiveness and new, more effective therapies are sorely needed.

Obstacles to the Development of More Effective Glioma Therapies

The central nervous system is a unique micro-environment. Because the brain is essential to survival, surgery cannot be performed with wide margins as is done in the resection of systemic tumors. The brain is physiologically different from other tissues and these physiologic differences have profound effects on both tumor biology and on drug delivery. The brain endothelium differs significantly from other endothelial tissue, resulting in the blood-brain barrier. The brain lacks a lymphatic system and is an immunological sanctuary, thus presenting a different set of challenges with regard to the use of immunologic therapies.

Central nervous system tumors differ biologically from systemic tumors. They generally have high drug resistance, both intrinsic and acquired. They are nonmetastatic in that they rarely spread to other organs, but they are highly infiltrative.

Brain tumors present specific pharmacologic challenges. In addition to the problem of the blood-brain barrier, it has become clear within the past decade that hepatic cytochrome P450 isoenzymes are intrinsic to the metabolism of most chemotherapy drugs. Many patients with brain tumors are taking anti-epileptic drugs that induce or inhibit the P450 system. Patients who are on enzyme-inducing anti-epileptic drugs (EIAEDs) have significantly altered drug metabolism. For example, for patients taking phenytoin (Dilantin) or carbamazepine (Tegretol), the maximal tolerated dose (MTD) of paclitaxel or CPT-11 may be 3- to 5-fold higher than for patients with systemic tumors. This has profound implications for clinical trials. It is necessary, for example, to conduct two Phase 1 studies to establish two different MTDs: one for patients who are taking EIAEDs and one for those who are not.

Patients with gliomas are very heterogeneous. Factors such as age, performance status, extent of resection, neurologic deficits, and use of glucocorticoids have significant effects on prognosis. Tumors are also heterogeneous; distinctions between tumor histologies are often unclear, resulting in inter-observer variability rates as high as 30% to 40%. Even tumors with similar histology can have very different genetic characteristics; for example, expression of the HMGT enzyme contributes to resistance to temozolomide. Additionally, the anatomic location of a tumor (e.g., brain stem, thalamus, right frontal lobe) can significantly affect the outcome.

Clinical Trial Design Issues

All of the above issues present challenges to the design of clinical trials of therapies for primary brain tumors. Because gliomas are rare and it is difficult to accumulate sufficient numbers of patients for a clinical trial, clinical researchers have attempted to use historical data to make comparisons. Unfortunately, the literature is severely flawed. Investigator-selected criteria for response are variable and almost always include stable disease. Past trials have often not required

response duration and have not controlled for the effects of glucocorticoids, the type of magnetic resonance imaging (MRI) technology used to measure response, or for important prognostic factors such as tumor type, grade, age, and performance status.

NCI-sponsored brain tumor research consortia are now generating databases that will improve the objective nature of neuro-oncology trials, but these databases are not yet freely available. Moreover, they may be of limited utility because patients enrolled in trials conducted by NCI-sponsored consortia may represent average patients in the community. An additional challenge for the design of clinical trials is that, with the possible recent exception of radiation +/- temozolomide, no agreed-upon standard of care exists upon which to base comparisons.

Clinically Meaningful Endpoints For Patients With Brain Tumors

Survival is both an objective and clinically meaningful endpoint, but it requires large randomized studies in a relatively rare disease. Few adequate historical controls exist to allow non-randomized comparisons. Small patient numbers make it very difficult to study any glioma subtype except GBM. It is difficult to balance hugely important prognostic factors, particularly in the setting of recurrent brain tumors. Finally, survival is not an appropriate endpoint for studies of palliative drugs.

Disease-stabilization endpoints (e.g., PFS, time to progression) offer the advantage of requiring a shorter time to data maturation. Because tumor progression is usually associated with worsening neurological function, tumor stabilization might translate to improved QoL, but few data are available to support this. In a rapidly progressive disease such as GBM, however, progression tends to precede death by a few months at most, so it is unclear how much time is really saved by the use of progression rather than survival as an endpoint. Disease-stabilization endpoints have many of the same disadvantages as survival: the need for large randomized studies in a rare disease, inadequate historical controls for non-randomized comparisons, small patient numbers, and inappropriateness for studies of palliative drugs.

Clinical response is associated with patient symptoms, performance, and QoL. However, patient symptoms are highly subjective. Neurological signs are objective but are affected by significant inter-examiner variability. Symptoms are also affected by concomitant medications (e.g., glucocorticoids, antiepileptics, anticoagulants).

Radiographic response is somewhat objective and is the historical standard, but has many disadvantages. Because gliomas usually do not form “lumps” in the brain, MRI scans are often not looking at the tumor directly but rather at the tumor’s effects on normal brain architecture. Tumors cause several different signal abnormalities on MRI scans.

Gadolinium enhancement is measured as a response criterion in most clinical trials. However, gadolinium enhancement does not measure nonenhancing tumors. This approach tends to measure vascular permeability rather than tumor; factors such as radiation damage and use of glucocorticoids or vascular-stabilizing drugs can affect vascular permeability. No standard way of measuring gadolinium enhancement exists; the Response Evaluation Criteria in Solid Tumors (RECIST) have not been validated in brain tumors.

With regard to PRO and QoL endpoints, although treatments that improve patients' neurological functioning, increase their ability to live independently, and decrease seizures would be valuable, no clear methods currently exist for measuring these parameters.

Conclusions

In summary, few effective treatments exist for primary brain tumors. No systemic therapy is approved for recurrent GBM. The literature from which to derive historical control data is largely undependable. Evaluation of clinical trials is affected by patient and tumor heterogeneity, factors shown to have a greater impact than any given therapy on patient outcome. Survival is currently the only clearly accepted trial endpoint. Treatment that resulted in tumor and symptom stabilization would be considered clinically meaningful and useful, but how best to objectively measure such outcomes remains unclear.

Dr. Fine ended his presentation by posing two questions that he said he hoped the workshop would address:

- What therapeutic outcomes are truly clinically meaningful to patients with gliomas?
- What clinical trial endpoints are representative of those outcomes and how can they be objectively and reproducibly measured?

CLINICAL TRIAL ENDPOINTS FOR APPROVAL: IMAGING-BASED OUTCOMES

Magnetic Resonance Imaging Surrogate Markers of Brain Tumor Therapeutic Response

(Dr. James Provenzale)

Dr. Provenzale began by saying that if he could sum up Dr. Fine's talk in three words, those words would be "validation," "quantification," and "reproducibility." Those words also describe the three issues that imaging scientists face in dealing with brain tumors, he said.

MRI is the imaging technique most commonly used to diagnose and assess therapeutic response in brain tumors. However, conventional MR imaging of brain tumors provides anatomic, but not physiologic, information. In most trials of brain tumor therapies, tumor assessment is based on both tumor size and enhancement characteristics.

The principal advantages of MRI compared with computed tomography (CT) imaging are that MRI makes it possible to image the tumor in multiple planes, offers better image resolution, and offers more advanced imaging techniques. However, MR imaging takes more time than CT scanning and cannot be performed on patients who have incompatible implanted devices such as aneurysm clips and cardiac pacemakers. Thirdly, it can be difficult to perform an adequate MRI scan on a very ill patient who has difficulty lying very still. It can also be difficult to monitor patients who are on respirators or are receiving continuously infused drugs.

CT scanning takes less time to perform than MRI and is useful for answering basic questions. Perfusion imaging can be performed with CT, but the role of this type of imaging in brain tumor

assessment is underexplored. However, the depiction of tumor extent is inferior with CT scanning compared with MRI. Like MRI, CT scanning provides very limited physiologic information.

Currently, brain tumors are assessed in clinical trials primarily by measuring them at the widest point of their diameter in accordance with the RECIST criteria. This provides no information about tumor physiology. At a time when many drugs can alter tumor physiology, new imaging techniques are needed that keep pace with these pharmacologic advances. Additionally, current imaging methods provide only a gross estimation of tumor aggressiveness. Three advanced MR techniques may be able to address these challenges.

- **MR spectroscopy** can be used to obtain metabolic profiles throughout the brain, which can be helpful in trying to determine what is happening in unenhancing areas of a tumor or in tissue adjacent to a tumor.

- **MR diffusion imaging** measures the rate of diffusion of water molecules throughout the brain in both tumor and normal tissue. The presence of tumor cells restricts the diffusion of water molecules in the brain; when the tumor responds to treatment, water molecules can diffuse more readily. Diffusion imaging can be used to measure therapy-induced changes in water mobility within the brain. Preliminary data suggest that this technique may be able to indicate within a few weeks whether or not the tumor is responding to therapy.

- **MR perfusion imaging** is a technique for monitoring the effectiveness of antiangiogenic therapy. Angiogenesis is the development of new blood vessels within tumors, which is essential for tumor growth beyond a few millimeters. Studies in animal models have shown that restricting angiogenesis severely impairs tumor growth. Angiogenic factors in tumors increase both the number and the permeability of blood vessels. High cerebral blood volume (CBV) can be an indicator of tumor aggressiveness. Perfusion imaging techniques can measure CBV and vessel permeability, both of which should decline in the presence of an antiangiogenic agent.

In summary, several advanced MR imaging techniques can provide both physiologic and anatomic information about brain tumors. These techniques, which are currently experimental, need to be used to measure tumor responses to therapy and to determine whether a tumor response correlates with outcome. Secondly, some of these techniques show promise as surrogate biomarkers. The jury is currently out on whether one technique is superior to any other.

Positron Emission Tomography Scanning with FDG in Brain Tumors; Brain Tumor Measurements in Assessing Response to Treatment (Dr. Nicholas Patronas)

Experience over the past 25 years has shown that positron emission tomography (PET) scanning is valuable in assessing tumor growth, providing guidance for surgical biopsy, assessing malignant transformation, addressing the issue of recurrence vs. necrosis after radiation therapy, and evaluating the extent of tumor growth within the cranial cavity, Dr. Patronas said.

As yet, data are sparse on the value of using PET scanning to assess response to treatment. Response assessment may be measured qualitatively (i.e., visually or by means of a ratio of pathologic to normal tissue) or quantitatively (i.e., standardized uptake value [SUV]). Another approach to quantitative measurement, calculation of the rate of glucose utilization, is no longer used.

SUVs are used in a variety of tumor types to measure prognosis and disease progression or regression. However, there are greater challenges in the application of this measurement approach to brain tumors because, unlike many other organs, the brain is highly metabolically active. Factors influencing SUV measurements include the plasma glucose level, the injected dose of the isotope, the time after injection that the scan is performed, use of medications that affect glucose metabolism (e.g., steroids, insulin), partial volume effects, and body weight vs. lean body mass.

Factors influencing image quality and lesion conspicuity on MRI include the signal-to-noise ratio, contrast issues, and image resolution and homogeneity. Factors influencing tumor enhancement include the dose of the administered contrast agent, the compound used, the time delay prior to scanning, medication use, renal function, hemodynamic alterations, and partial volume artifacts (i.e., obliquity of brain sections). It is important to ensure that every time the tumor is measured by linear measurement technique, images are coregistered by date to ensure that the same "slice" is evaluated.

Measuring tumor diameter is probably an outdated methodology, as small percentage changes in diameter can reflect much larger changes in tumor volume. Both manual and automated segmentation techniques provide more accurate measurements of tumor volume than diameter measurement; these techniques have the further advantage of not being operator-dependent. Automated segmentation is more accessible now, is easy to perform, and does not require manual manipulation of the image. In the post-contrast MRI, each tissue type has a unique distribution of pixel intensities. In automated segmentation the intensity of distribution is estimated for cerebrospinal fluid, normal brain tissue, and enhancing tumor. Each pixel's intensity is compared with these distributions and segmented according to its most probable tissue class.

Response and Progression-Free Survival Endpoints for Gliomas (Dr. Karla Ballman)

Dr. Ballman presented analyses of data from the North Central Cancer Treatment Group (NCCTG) database. The first study compared the performance of, and the extent of agreement between, the unidimensional (1D) RECIST criteria, the WHO bidimensional (2D) criteria, and computer-calculated measurements of tumor area and volume. Tumors were classified at various time points as progressive disease, stable disease, or disease regression. All measurements were conducted in newly diagnosed gliomas of different tumor types and different grades. Patients with enhancing tumors generally were older and had higher-grade tumors; patients with non-enhancing tumors generally were younger and had lower-grade tumors.

Agreement among methods was moderate at best. Determination of response by the 1D and 2D criteria did not differ significantly. No evidence of an association between response and survival

was seen for enhancing/nonenhancing tumor measurements. Some evidence of an association between progression and survival was observed for enhancing tumor measurements. Small sample size may explain some of the lack of agreement. Other limitations are that the data are from a single group and the analysis was not done by tumor type and grade.

The second study examined the relationship between PFS at 6 months and overall survival (OS) at 12 months in Phase 2 GBM trials. The study purposes were to determine the relationship between the endpoints, determine whether the relationship was similar in trials of newly diagnosed GBM patients and trials of patients with recurrent GBM, and assess whether it is reasonable to use 6-month PFS in place of 12-month OS as an endpoint for Phase 2 GBM trials. Data were pooled for 1,359 patients in 12 trials, all of which were negative.

Patient-level agreement was moderate for trials of both newly diagnosed and recurrent disease. Trial-level agreement was mixed for both types of trials; correlation was moderate (less than 0.90) and agreement of study results was good (88-90%). PFS at 6 months was strongly associated with OS at 12 months. Once again, all data are from a single cooperative group and all are from negative trials. Importantly, this was not a formal surrogate endpoint analysis.

Is Progression-Free Survival a Clinically Relevant Endpoint for Clinical Trials Testing Treatments For Malignant Glioma at Time of Progression? Report of Data From the North American Brain Tumor Consortium (Dr. Kathleen Lamborn)

Dr. Lamborn presented data from an analysis of 13 single-arm Phase 2 trials involving 611 patients with high-grade gliomas. The trials were performed at multiple institutions participating in the North American Brain Tumor Consortium (NABTC). Entry criteria were similar for all trials: patients were adults with Karnofsky Performance Status (KPS) scores of at least 60, proof of disease progression by imaging, adequate organ function, prior radiation therapy, and a limited number of prior chemotherapies. Evaluable or measurable disease was not required. The primary endpoint for all trials was 6-month PFS. The purpose of the analysis was to determine whether progression status at various time points predicted OS from those time points.

For patients with both Grade 3 and Grade 4 tumors, progression status strongly predicted survival from the time of assessment for each of the planned assessment times (9 weeks, 18 weeks, and 26 weeks) during the first 6 months from the start of the study, indicating that delay in time to progression predicts for improved patient survival. This finding is limited by the fact that the data were not derived from randomized trials and none of the therapies tested was particularly successful. These data nevertheless raise the hope that extending PFS would in turn extend OS, Dr. Lamborn concluded.

This analysis is ongoing. The same results were seen when the data were adjusted for age and performance status and when patients with prior surgery were excluded. Further analysis showed that response was predictive of survival with a hazard ratio of about 0.5. However, this did not alter the strength of progression vs. no progression as a predictor of survival. Analysis of a separate data set, involving patients with both Grade 3 and Grade 4 gliomas at first progression who were treated at the University of California, San Francisco, also concluded that progression status strongly predicted survival.

Dr. Lamborn then discussed the implications for sample size and study duration of using PFS vs. survival as an endpoint for studies aimed at regulatory approval. She estimated that a Phase 3 trial involving patients with Grade 4 tumors could be completed in 1.5 years if 6-month PFS was the primary endpoint vs. 3.5 years if OS was the primary endpoint. For a Phase 3 trial involving patients with Grade 3 tumors, the estimated study duration would be 2.5 years if 6-month PFS was the endpoint vs. 4.2 years if OS was the endpoint.

Dr. Buckner asked whether age, performance status, and extent of resection were associated with differences in PFS outcome. Dr. Lamborn replied that age was associated with PFS outcome much more strongly than was performance status. She did not look at extent of resection because few patients fell into this category. Instead, a second analysis was performed in which patients who had surgery within 30 days of the start of the study were excluded. This made no difference to the results of the analysis. Dr. Lamborn also responded to two other questions concerning the analysis methodology.

Panelist Discussion—Imaging-Based Outcomes

Dr. Barker commented that the discussion about defining response and progression was in the context of no locally delivered therapy. It had not been explicitly stated that none of the reported findings apply to the measurement of response, progression, or PFS following carmustine wafer implantation. He added that trials in recurrent disease must have carefully defined starting points and entry criteria. Particularly for trials involving antiangiogenic agents in recurrent disease, goals for trial endpoints must take into account whether or not patients have measurable disease. All of the endpoints that have been discussed may also be starting points for certain other trials.

Dr. Loeffler noted that when patients are treated with escalating doses of radiation, post-treatment imaging of their tumors almost always appears worse than before, but in most cases these changes are transient. Dr. Fine said that the trials analyzed by Dr. Lamborn all involved chemotherapeutic or targeted agents that would not be expected to cause significant radiographic changes; thus, the findings of her analysis may be relevant only for certain classes of therapies.

Dr. Yung said this highlighted the problem of interpreting MRI data that are acquired close in time to the use of high-dose radiation therapy; in this situation, it is difficult to be sure of the meaning of radiographic changes. Dr. Provenzale commented that studies must take into account the expected effect of a drug or device on the underlying principles of the imaging technique being used; otherwise, conclusions may be misleading. For example, a therapeutic device implanted in the brain might cause changes in water diffusibility or in the permeability of the blood-brain barrier.

Dr. Yung noted that when the first scan is done 2 to 4 weeks after radiation therapy, a high percentage of observed changes are likely to be radiation-induced. One way to resolve this problem might be to discount the findings on this scan. The next scan 2 months later is likely to provide a more accurate picture of disease progression; determinations about discontinuing patients from the study on the basis of progression should be postponed until this point.

Dr. Fine asked whether any imaging modalities can definitively differentiate, for example, treatment-related from tumor-related changes in gadolinium enhancement. Dr. Patronas replied that from a morphological point of view it is not possible to distinguish treatment-related phenomena from tumor progression. Dr. Provenzale agreed that no single imaging tool could meet this need in all circumstances, but expressed hope that in any individual circumstance it might be possible to identify an imaging tool that would answer the question.

Dr. Fine pointed out that the focus of this discussion was whether any imaging modalities were currently validated to the extent that they could be reliably used to assess the efficacy of a drug. He said the data that had been presented suggested that PFS might be a valid predictor of survival (in the context of standard systemically administered agents), but the question was how progression is defined.

Dr. Paoletti said regional distribution of the lesion was an important issue; a 1 mm reduction in tumor volume in a certain part of the brain might have a dramatic clinical effect whereas a larger volume reduction elsewhere in the brain might be clinically meaningless. Companies engaged in drug development would like simple, clear guidance on how to measure disease progression because the RECIST criteria are not appropriate.

Dr. Buckner said the NCCTG data were reasonably convincing that either 1D or 2D measurement of contrast-enhancing tumor was a reasonable endpoint because both were associated with survival. He added that two independent data sets seemed to support the conclusion that, for patients with recurrent glioma who are treated with standard systemic agents, 6-month PFS is predictive of 12-month OS. He pointed out that the analysis of NCCTG data had excluded patients who were treated with stereotactic radiosurgery and implanted carmustine wafers. Additionally, the conclusions of the NCCTG analysis were not affected by inclusion in the database of patients treated with an agent with antiangiogenic properties. Although this agent was inactive according to the study definition, it may still have had biological activity. Dr. Buckner added that patients who have focal therapies are highly selected; for this reason, what could be a confounding variable was likely to be limited to a subset of patients.

Dr. Fine noted that the caveat regarding standard systemic agents was important. Two ongoing, industry-sponsored Phase 3 studies were using convection-enhanced delivery of the investigational agent and that necrosis and breakdown of the blood-brain barrier were anticipated toxicities. Thus, standard measures of radiographic progression (i.e., increased gadolinium enhancement) may not be predictive surrogates for overall antitumor activity or overall clinical benefit. He asked, however, whether it could be known with certainty prospectively that a new targeted agent would not cause MRI changes that would confound the measurement of progression. He added that it is not known whether or not antiangiogenic agents cause necrosis.

Dr. Yung said the data were convincing that 6-month PFS was a useful endpoint not only for recurrent disease but also for newly diagnosed disease. He asked if it was possible in multi-site trials to standardize the parameters for the use of contrast agents (e.g., how much contrast agent to use, the infusion rate, etc.). Dr. Fine asked whether inter-institutional variability in imaging was large enough to affect the results of a trial. Dr. Provenzale responded that in well-designed multi-site trials standardized criteria are used for imaging and compliance at individual sites is

monitored. Dr. Patronas said that the issue of different imaging equipment at different study sites could be addressed by prospectively designing imaging parameters and by requiring all scans of an individual patient to be done on the same instrument.

Dr. Fine asked the panelists whether, in large multi-institutional studies at their institutions, it was routine for a detailed MRI protocol to be followed. The consensus was that it was not. One panelist commented that at his institution a study evaluating contrast agents had failed because of cross-platform discrepancies. He felt that studies would fail unless imaging techniques were standardized across institutions.

Dr. Friedman asked whether studies could be designed in a way that allowed AA to be granted on the basis of interim results, with confirmation of clinical benefit provided (or not) by the final results of the same study. Dr. Pazdur said that FDA advocates this approach to trial design. One drawback, however, is that if the interim results show that the experimental agent appears to offer a benefit, crossover from the control arm may confound the final survival analysis. Dr. Pazdur added that the magnitude of a therapy's effect on an endpoint is an important consideration in regulatory decision-making. For example, a doubling of PFS would be a more compelling result than a 15% improvement.

Dr. Fine reiterated that the question before the panel was whether another endpoint was sufficiently accurate to replace or serve as a surrogate for survival, the current gold standard. Dr. Yung suggested that another way to phrase the question was whether the correlation between 6-month PFS and 12-month OS was significant enough to support the conclusion that the patient is likely to benefit from treatment. Dr. Buckner commented that the evaluation of survival is increasingly being confounded by the use of sequential therapies.

Dr. Provenzale said that the most appropriate technique for imaging and tumor measurement are likely to be different depending on whether scans are being performed at academic medical centers or at community-based centers. No single method will be optimal for all tumors. He recommended the use of two MRI techniques, one based on contrast administration and the second (performed at the same examination) not dependent on contrast administration. To enable the highest degree of confidence that imaging protocols will be followed, trials should be performed at tertiary care centers.

Dr. Fine noted that such a policy would have great implications for the way clinical trials are carried out in the United States. For example, the Radiation Therapy Oncology Group, which conducts most of the large Phase 3 trials in glioblastoma, has a large network of community-based investigators. Dr. Buckner suggested that the problem could be handled by requiring central image review, as occurs in pathology. Dr. Pazdur commented that it was problematic for FDA reviewers when there was a significant difference of opinion between image readers.

Dr. Pazdur asked whether panelists thought that freedom from disease progression constituted a clinical benefit to the patient in and of itself, regardless of whether it was a surrogate for overall survival. Dr. Buckner responded that freedom from progression would be valuable if it were known to result from the treatment, since other variables (e.g., age, performance status) are known to affect PFS.

Dr. Yung said that symptomatic deterioration often precedes radiographic evidence of disease progression. Dr. Fine added that disease progression cannot be defined only radiographically. A patient who is deteriorating clinically, even if shown to be progression-free by MRI scan at 6 months, would likely not feel that he or she was obtaining benefit from the current therapy. Dr. Yung pointed out that the NABTC criteria for lack of disease progression included stable neurologic condition. Dr. Buckner added that in the NCCTG clinical deterioration is considered disease progression even if the results of two consecutive MRI scans showed tumor stability.

It was noted that freedom from progression may result in freedom from therapy, which in brain tumors is often highly toxic, but that it may also be a result of concurrent chemotherapy; in the latter situation, freedom from progression may not be associated with improved QoL.

Several panel members said that no standardized scales are currently used to measure neurologic status. An assessment that a patient has deteriorated neurologically is based on clinical judgment. Dr. Yung noted that instruments used in other neurologic diseases (e.g., multiple sclerosis, dementia, stroke) measure highly specific aspects of neurologic function rather than global neurologic status.

Dr. Pazdur asked the panel if there are circumstances in which response rate would be a useful endpoint for brain tumor studies. He noted that in other tumor types FDA has accepted single-arm studies in which response rate was the primary endpoint. The advantages of single-arm trials are that they are generally less complex to design and require fewer patients. On the other hand, they cannot be used to characterize toxicity or evaluate time-to-event endpoints. Dr. Pazdur added that FDA had felt confident granting AA to temozolomide on the basis of response rate because several patients showed a sustained complete response. However, such sustained complete response rates are rare.

Dr. Buckner replied that if the response rate were of sufficient magnitude (e.g., greater than 30%), it was likely to be associated with clinical benefit; the magnitude of the response rate would outweigh the uncertainties associated with interpreting MRI scans. Dr. Yung said that several meta-analyses of data from negative Phase 2 trials in recurrent GBM had consistently found response rates of around 5% to 7% despite changes in MRI technology over time. Dr. Fine said that if response rate were the primary endpoint it would be important to select patients whose disease was clearly progressing. Dr. Crocker commented that it would also be important to ensure that post-surgical changes were not misinterpreted as a therapeutic response. Dr. Patchell said that the only two issues of ultimate importance for patients were survival and QoL.

Audience Questions and Comments

Dr. Henry Brem, Johns Hopkins University, commented that all local therapies increase tumor enhancement and may very well increase diffusion; for this reason, MRI scans would be a poor way to assess the effectiveness of these therapies. He agreed with Dr. Patchell that improved survival and QoL were the key criteria to be met when assessing therapeutic effectiveness. Dr. Fine noted that the NABTC analysis did not exclude patients who received carmustine wafers as a first-line therapy.

Susan Arbuck of Schering-Plough, Inc., pointed out that the response rate that was the basis for AA of temozolomide was substantive but substantially below the rates that panelists had suggested might be required. The drug subsequently showed a survival benefit and received regular approval on that basis.

In response to a question from a member of the audience, Dr. Rock said that the EORTC study illustrated the value of survival as an endpoint. Six-month PFS was addressed in that study. The panel had heard some provocative, hypothesis-generating discussion about PFS this morning, he added. FDA would be interested to know how PFS maps with other prognostic indicators such as performance status and cognitive function. Dr. Buckner added that overall distribution of PFS was a secondary endpoint in the EORTC study.

Dr. Paoletti noted that many drugs now in development are not cytotoxic. For patients treated with these new-generation agents, stable disease or a minor response rate associated with symptomatic improvement may be very important. A new paradigm is needed for assessing the clinical benefit of these new agents.

Susan Wiener, a patient advocate with the NABTC, said that neurologic exams are indeed a solid measure, although there is no substitute for the physician's clinical judgment. Correlation between the neurologists' assessment and the patient's disease status would generally be high. She said she did not understand why a neurologic exam could not be included as a measure of the patient's response.

Dr. Buckner responded that it would be difficult to mandate a specific tool for use in the assessment of all patients; a global assessment of neurologic status might be more informative. Dr. Yung agreed. Dr. Fine said the development of a standardized neurologic assessment tool would be a worthwhile research effort but no tool currently exists that could be recommended for standard use.

Dr. Patchell said that in a recently completed trial in metastatic disease, clinical criteria and a custom-devised neurologic exam had been used to measure patients' neurologic status. An independent blinded committee reviewed the data and determined that the neurologic exam was as accurate as, and correlated closely with, investigators' clinical judgments of patients' status. Dr. Fine commented that other studies have shown that mental status or neurocognitive deterioration is a better predictor of long-term outcome than radiographic findings. Dr. Patchell said it would be helpful to have an objective scale that could be used across trials. He noted that neurologic function is closely associated with QoL.

CLINICAL TRIAL ENDPOINTS FOR APPROVAL: PATIENT-REPORTED OUTCOMES

Cognitive Testing and Patient-Reported Outcomes in Brain Tumor Clinical Trials (Dr. Christina Meyers)

Most patients with brain tumors suffer from cognitive dysfunction, Dr. Meyers said. The net clinical benefit of cancer therapy includes "beneficial effects on disease-related symptoms and/or quality of life," according to an FDA-NCI working group. Maintaining function is particularly important for patients with brain tumors because long-term remission or cure is unlikely or is often accompanied by significant disability.

Clinical benefit to the patient with a brain tumor includes relief of tumor-specific symptoms, including disruption of brain function. However, anatomic evidence does not correlate well with cognitive function. A patient with a large, slow-growing tumor may have minimal cognitive impairment whereas a patient with a much smaller but rapidly growing tumor may have profound cognitive effects. Treatment-related changes on an MRI scan also do not correlate closely with patient function. For example, focal high-dose radiation therapy causes oxidative stress and inflammatory changes in the brain that may persist long after transient changes on an MRI scan have resolved.

Tumor-specific symptoms are measured by the patient's subjective report of symptoms (headache, nausea, etc.), objective assessment of cognitive function or mood, and objective assessment of function (e.g., independence in activities of daily living). Clinical researchers evaluating cognitive function in patients with brain tumors want to know what if any cognitive problems the patient had prior to treatment and whether treatment regimens improve neurocognitive function as a result of better tumor control, slow expected tumor-related neurocognitive deterioration, and have more or less short- and long-term toxicity.

In a trial of a radiation sensitizer for the treatment of brain metastases, FDA stated that "radiological response alone is not acceptable for approval. However, improvement in neurocognitive function or delay in neurocognitive progression are acceptable endpoints." These alternative endpoints are being used in ongoing and planned trials for both brain metastases and primary brain tumors.

Assessment of cognitive function. Evaluation of cognitive function presents a number of assessment issues. Performance status has little relation to cognitive function and QoL. Brief mental status exams are only sufficiently sensitive to detect serious cognitive impairments such as delirium and significant dementia. Self-reports of cognitive problems correlate poorly with objective test results; patients with brain tumors may have a diminished appreciation of their impairments and may report that their memory is fine when in fact they have significant memory deficits.

Any tool for assessing cognitive function must be brief (i.e., take no more than 30 minutes to administer) and repeatable in alternate forms with minimal practice effect. It must have good psychometric properties—that is, it must measure the intended function reliably over time. It

must be highly sensitive to changes in function, must measure relevant cognitive functions, and must be highly standardized and simple to administer. Most patients must be able to complete the instrument.

To be analytically valid, assessment instruments must reflect population norms—that is, must take into account the expected level of cognitive function in a patient of a specific age and educational level. The degree of change that is considered to reflect either an improvement or a decline in the patient's performance must be prospectively established. Variation in results at different sites or by different examiners must be minimized; formal training, certification, and quality assurance requirements must be built into the trial. In trials of pediatric brain tumors, assessment instruments must be developmentally appropriate and must take into consideration the likelihood of altered long-term cognitive development.

Issues that may confound the assessment of cognitive function (e.g., adjuvant medications such as steroids, medical complications such as seizures) must be identified. Cognitive assessments should be performed at the same time intervals as other staging evaluations such as MRI scans. The frequency of assessment should be relevant to the disease course; for example, assessments may be less frequent in a trial of low-grade glioma than in a GBM trial. The results of cognitive assessments must be correlated with anatomic response and neurologic outcome, although cognitive deterioration may occur before other evidence of progression is apparent.

Patient-reported outcomes. Ideally, a patient-reported outcome (PRO) should be based on disease- or treatment-related symptoms rather than on social function or satisfaction with life. It must have sound psychometric properties, be simple enough to be completed by patients with cognitive deficits, and be sensitive to changes over time.

Several caveats apply to the use of PROs in patients with brain tumors. Patients need to have sufficient cognitive function to complete the instrument. Many symptom assessment instruments have suboptimal psychometric properties (e.g., poor test-retest reliability). Instruments must be able to account for reporting bias so that over- and under-reporters do not simply cancel each other out. Proxy assessments are problematic for subjective symptoms; for example, a caregiver cannot reliably evaluate the severity of a patient's headaches. To reduce missing data, investigators must "buy in" to both cognitive and symptom assessment and encourage patients to complete the instruments. Finally, change in QoL does not parallel cognitive change and cannot be used as a proxy for it.

Standardized approach to assessment. In brain tumor clinical trials it is desirable to be able to compare cognitive and symptom assessment findings in trials of different agents conducted by different investigators. One of the recommendations of the NCI's Brain Tumor Progress Review Group was to develop a "practice guideline protocol," which would include standard content that would enable investigators to select the tools most appropriate for the evaluation of a specific drug or hypothesis.

Which trials? Several issues should be considered in deciding in which trials to use cognitive and symptom assessment as endpoints. For randomized controlled trials, cost effectiveness and whether alternative endpoints should be primary or secondary endpoints are among the issues to

be discussed. It may also be worth considering what value alternative endpoints might add to single-arm Phase 1 or Phase 2 trials. For example, they could be useful in monitoring neurotoxicity. Standard content would permit comparison of findings from different single-arm trials.

Panelist Discussion—Patient-Reported Outcomes

Dr. Pazdur noted that PRO endpoints have been incorporated into many cancer clinical trials in other tumor types, but to date few trials have succeeded in demonstrating a beneficial impact on patients' QoL. There are methodological challenges to the use of PRO endpoints in cancer trials. For example, sponsors often submit to FDA only a single, unblinded, randomized trial with a lot of missing data. Dr. Pazdur emphasized that FDA believes patient QoL is an important outcome in cancer treatment. However, if PROs are to be used as the basis for drug approval, they must be measured with the same rigor as any other endpoint.

Jane Scott, Ph.D., FDA endpoint reviewer, drew a distinction between general QoL (e.g., financial security, quality of personal relationships) and health-related QoL (HR-QoL), which is the aspect that FDA reviewers focus on. HR-QoL is a multidimensional concept that encompasses, but is not necessarily limited to, measurement of symptoms and physical function. However, the ability to accurately measure the impact of a therapy on symptoms would be valuable to FDA even if it could not be directly related to improvement in patient function.

Dr. Scott asked whether panel members have been developing tools that have proved helpful in systematically establishing what a patient's symptoms are and how they change over time. Dr. Meyers replied that a symptom research group at M.D. Anderson Cancer Center, of which she is a member, has developed several psychometrically based symptom assessment tools, which she has used in brain tumor clinical trials. In one recently published trial, patients' symptoms were unchanged except for fatigue, which worsened considerably.

Dr. Scott noted that in other tumor types there is less reason to be concerned that the disease process itself and/or its treatment will erode patients' cognitive function. She asked for information about efforts to develop standardized clinician assessments of patient symptoms and function to complement patient self-reports and about study designs that would enable patients' symptoms and function to be followed as their disease advances. Dr. Meyers acknowledged that self-reported symptom assessments in patients with brain tumors present a high risk of selection bias because only the more highly functioning patients can complete them. Some steps can be taken to compensate for patients' deficits, such as reading questions aloud to patients who have difficulty reading. Patients with the worst cognitive function (who, for example, cannot rate their pain on a numerical scale) most likely will have been withdrawn from the study.

Dr. Scott asked if it would be feasible to design trials so that findings on patient self-reported symptom assessments and objective cognitive tests would trigger radiographic assessment, rather than performing radiographic assessments at fixed time intervals. Dr. Meyers responded that this study design had not been tried at M.D. Anderson for logistical reasons, since patients often have to travel long distances to attend their assessments.

Dr. Pazdur asked for comments on the feasibility of designing a composite endpoint that combined measures of patient function with radiographic findings. Dr. Crocker noted that other factors, such as a patient's dose of seizure medication being too high, could confuse the assessment of patient function. A composite endpoint that combined "soft" endpoints would not be helpful. Dr. Meyers said she knows of trials that have stipulated that a change in function be confirmed at a subsequent assessment to increase confidence in the finding.

Dr. Barker commented that it would be helpful to distinguish whether data were missing because patients were too ill to attend the assessment or because it was inconvenient for them to attend. Dr. Friedman observed that clinical and radiographic findings can be contradictory (i.e., the MRI scan can look good but the patient is clinically worse, or vice versa). Dr. Paoletti urged that an effort be made to develop and validate standardized tools.

Dr. Rock asked Dr. Meyers to describe the metric she developed for trials of motexafin gadolinium (Xcytrin) in patients with brain metastases. Dr. Meyers said that in those trials patients had been assessed monthly with a brief battery of tests that took about 23 minutes to administer. The memory test had six alternate forms. Other tests focused on measuring patients' independence in activities of daily living (e.g., frontal lobe function, motor coordination). Careful certification procedures were employed to ensure the accuracy of test administration. The tests was translated into multiple languages and administered to patients in 7 countries. Multiple comparisons were performed. A test focusing on a single aspect of cognitive function is insufficient because patients will develop different symptoms (e.g., weakness, headaches) depending on factors such as the location of the tumor. Memory function tends to be most sensitive to both tumor and treatment effects.

Dr. Pazdur observed that symptoms can more readily be measured when a particular symptom is a cardinal feature of the disease (e.g., dysphagia in esophageal cancer, bone pain in prostate cancer). Symptoms that are diffuse or ill-defined, or that do not appear until very late in the disease course, are much more difficult to measure. Dr. Yung said that the fact that a patient's symptoms are so dependent on the location of their tumor has so far confounded efforts to develop a standardized approach to symptom assessment in brain tumors. Dr. Lamborn asked if it was possible to prospectively define and follow symptoms on an individual-patient basis. Dr. Meyers said she had no experience with this approach.

Dr. Fine observed that high doses of steroids are a major cause of morbidity in patients with brain tumors. Steroid doses are determined empirically by attempting to find the lowest dose that optimizes the patient's neurologic function. A nontoxic drug that stabilized the vasculature and enabled patients to take lower doses of steroids would provide clinical benefit even if it had no effect on the tumor itself. He asked how a trial of such an agent could be designed to reliably capture this benefit. Dr. Pazdur said such a trial would have to convincingly demonstrate a beneficial effect on steroid doses and on the toxic side effects of steroids. Measuring such changes consistently in an unblinded trial could be challenging. Dr. Scott added that the use of a validated, standardized approach to symptom measurement would be helpful in such a trial design.

Dr. Pazdur noted that PROs have been the basis for approvals of drugs in other therapeutic areas such as neurology and psychiatry. In these cases, however, the approval decision is usually based on review of two blinded, randomized trials. Blinding of trials is problematic in oncology because of factors such as different drug delivery schedules, different toxicities, and patient reluctance to enter blinded trials. Additionally, in most cases, a single pivotal trial is submitted. When the magnitude of change attributable to a new therapy is relatively small, it is difficult to be confident that a beneficial effect on symptoms or QoL is not due to chance or to a placebo effect. Another common problem is that in many trials assessments of PROs and QoL are added on as an afterthought instead of being integrated into the trial.

Dr. Pazdur asked Dr. Scott to describe the factors that reviewers in other therapeutic areas take into account when considering an application for approval based on PROs. Dr. Scott said that in many therapeutic areas reduction or stabilization of symptoms is regarded as an important clinical benefit. It is important at the outset to clearly define the symptom that is to be measured, which sounds simple but in practice can be very nuanced. In particular, patients must understand what is being measured. Symptoms that clinicians consider important may not be the ones that are most bothersome to patients.

The next step is to test the questions to ensure that patients understand them and to try the questionnaire out in studies. A large literature has evolved on the development, calibration, and validation of questionnaires. The literature also addresses what kind of recall patients can reasonably be expected to have of past events. For patients with brain tumors, a disease in which both the condition and its treatment may significantly affect memory, the focus should to the extent possible be on asking patients about their current status. When questionnaires are translated into other languages, care must be taken to ensure that patients' scores are not affected by the language in which they respond to the questions.

Dr. Scott added that in her experience FDA has found symptom and HR-QoL assessments to be most helpful, reliable, and useful for regulatory purposes when the findings are derived from double-blinded randomized trials. It would be problematic, in her opinion, to accept symptomatic improvement as the primary grounds for approval on the basis of a single unblinded study.

Dr. Weiss noted that in rheumatoid arthritis a composite instrument has been developed and validated that combines measurement of symptoms with objective measures of the patient's status. Dr. Scott said that all composite measures must be based on a large amount of data so that reliable judgments can be made as to what the parameters of each element should be. Some questionnaires sacrifice precision to achieve brevity. Some composite instruments combine several different measures into a single global score, making it difficult to pinpoint the precise areas in which the patient obtained benefit. The ability to disaggregate a global score is an important feature of any composite measurement tool.

Audience Questions and Comments

Dr. Elana Farace, Penn State University, stated that she held an NIH grant to study the relationship between global QoL and neurocognitive symptoms over time in patients with malignant glioma; Dr. Meyers is her senior mentor on this grant. She said the discussion at the

morning session suggested that clinicians felt they could assess a patient's overall status on the basis of detailed information about neurological and neurocognitive function, whereas FDA seemed to be talking about global QoL. The latter is more difficult to assess and there is a lack of information about the relationship between neurological/neurocognitive function and QoL. Her data suggest that deterioration in neurocognitive function adversely affects QoL more seriously than does decline in physical function. She added that a large body of data supports the reliability and validity of standardized neuropsychological tests.

In response to a question from a member of the audience, Dr. Scott said that the usefulness of a patient's KPS score often depends on whether the patient was initially high-functioning or low-functioning on the KPS scale. Low-functioning patients may be unable to self-report symptoms and cognitive tests may be a less useful measure of their status. In any case, the KPS score is usually only modestly correlated with patient self-reported symptoms. Dr. Meyers added that the psychometric reliability of the KPS is low; in one published study, there was only 29% agreement between two physicians on what a patient's KPS score was.

Dr. Fine said that both the KPS and the Eastern Cooperative Oncology Group performance status scales were developed for systemic tumors and were essentially surrogates for tumor burden. Because these scales tend to focus on motor function rather than cognitive function, they are unreliable tools for the assessment of patients with primary brain tumors.

Dr. Pazdur said FDA's experience with global QoL measures in other disease areas has been unsatisfactory. He said the agency has in the past suggested to sponsors that they measure time to symptomatic progression rather than time to radiographic progression, using blinded evaluators to minimize bias. He said that patients could continue to be followed for time to symptomatic progression even after a change in therapy, just as patients' survival continues to be monitored after a change in therapy. He asked for comments on this approach. He stressed that FDA is very interested in the use of PROs as endpoints, recognizing that symptoms are a very important issue for patients.

Dr. Fine responded that this approach would require a large randomized trial in which tumor location was controlled for, since tumor location has such a significant effect on the patient's symptoms. Dr. Patronas noted that in his experience symptomatic deterioration may occur before disease progression is evident on the patient's MRI scan. Dr. Crocker said it would important to correlate symptomatic progression with survival. Dr. Buckner said such an endpoint would be very valuable and would probably have a high rate of acceptance but he questioned whether a validated tool currently exists to measure it.

Dr. Kun said the data do not yet exist to document that measurable changes in patient symptoms can be correlated with progression, PFS, or survival in any population with brain tumors. Dr. Yung said the cognitive-function battery described by Dr. Meyers has been validated, but there is a lack of experience in large randomized trials to confirm that it is a valid surrogate. Dr. Meyers noted that the battery has been validated in brain metastases. She added that treatment neurotoxicity (e.g., late radiation effects) is an additional complicating factor. Ms. Wiener (patient advocate) suggested that neurocognitive function might be an appropriate topic for an NIH consensus conference.

Dr. Scott emphasized that in FDA's view the reliable demonstration of a reduction in patient symptoms is a clinical benefit in and of itself, regardless of the long-term survival benefit associated with the therapy. Measurement of symptoms can also be helpful in establishing the appropriate next step (e.g., imaging studies), but the requirement that these be perfectly correlated tends to minimize the value of the outcome itself.

A member of the audience questioned whether the benefit of extending survival may be overestimated when the patient's neurocognitive function is seriously compromised. Tom Nesi, patient representative on the panel, responded by noting that he had cared for his wife, a GBM patient, for 18 months. In his opinion, survival was not a good outcome measure. He said his wife was unconscious for the last 4 weeks of her life. Caregivers and primary care providers would certainly question whether extending survival is always beneficial, he said.

Mr. Nesi added that assessment of the quality of a patient's life must take into account issues such as the effect of polypharmacy (during her illness his wife was taking at least 7 prescription medications); the impact of a sudden, lethal diagnosis on a previously healthy person and on his or her family; and the enormous financial burden of treating the disease.

Dr. Grant Williams, Novartis Corporation, from the audience, suggested that since both symptom progression and imaging progression seemed to have drawbacks as sole endpoints, the solution might be to combine them—that is, to define disease progression at 6 months by means of both symptom and imaging progression.

GENERAL PANEL DISCUSSION

Panel members turned their attention to the general discussion questions posed by FDA.

Individual Endpoints
1.1. What if any non-survival endpoints reflect or predict clinical benefit?

The panel agreed that 6-month progression-free survival (PFS) (with clinical stability as currently defined and without the use of local therapies) is a meaningful endpoint.

1.2. What if any endpoints available now may be reasonably likely to predict clinical benefit?

Dr. Pazdur said the term *reasonably likely* refers to surrogate endpoints that are "reasonably likely to predict clinical benefit," the standard for granting accelerated approval (AA). Symptomatic improvement would be considered direct clinical benefit to the patient, not a surrogate for clinical benefit, and could therefore be used as an endpoint for regular approval.

Dr. Buckner said the neuro-oncology community has accepted radiographic response as a surrogate endpoint in oligodendroglioma because the magnitude of the benefit was dramatic and it was unequivocally treatment-related. Radiographic response is a reasonable endpoint if it convincingly represents a therapeutic effect, he said. Dr. Fine added that the response should be significant (i.e., greater than 15% or 20%) and durable, the patient must be clinically stable or

improving, and the patient's doses of steroids must be stable or decreasing. Dr. Yung said it was generally established that a response must be validated on a second scan.

Dr. Barker asked whether it was necessary to stipulate how one knows that a response is treatment- related. Dr. Provenzale said that the effects of therapy on imaging of the patient's tumor must be understood. Imaging studies, in his opinion, are reflectors of therapy rather than predictors of outcome. Dr. Yung said that because agents are now in use that modify the blood-brain barrier and change edema patterns, outcome measures must correlate with a therapeutic agent's biologic activity. Dr. Fine said the agent's mechanism of action must be considered in determining an appropriate surrogate endpoint; one surrogate is unlikely to be appropriate in all circumstances.

Dr. Pazdur said that the magnitude of response (including the number of complete responses) is important, particularly in a disease characterized by inter-reader variation in response assessment. The presumed effect of a drug is often overestimated; an agent may look promising in a small study, but in a larger trial response rates may be much lower.

Dr. Kun commented that many novel agents such as angiogenesis inhibitors may stabilize disease but not cause tumor shrinkage, which is the conventional means by which response is measured. Dr. Pazdur noted that some recently approved agents had low response rates but large effects on time to progression. He noted that although response can be measured in a single-arm study, time-to-event endpoints must be measured in randomized trials.

1.3 Is it reasonable to allow a period of time for a novel biologic agent to have a biologic effect on a tumor? If so, how much time is reasonable?

Dr. Rock said that this question was specifically relevant to the use of novel biologic therapies that are locally delivered at a tumor site and may initially result in images that appear to show radiologic tumor progression. In response to an earlier comment by Dr. Provenzale regarding the difficulty of making blanket statements about response based on novel MRI techniques, Dr. Rock said FDA did not find this to be a limiting factor. He said the Office of Oncology Drug Products invites drug sponsors to come in at any time to discuss endpoints that they are considering using in registration trials.

Dr. Barker said he believed that initial imaging changes associated not with biologic therapies but with standard external beam radiation can be significant in predicting survival. He said it is increasingly clear that imaging changes that develop during or soon after treatment are an unreliable guide to a patient's prognosis following local therapy and should be interpreted with considerable caution. To improve understanding of the effects of local therapies, including their biological effects, careful consideration should be given during trial design to how much apparent "progression" can be tolerated and for how long before the decision is made to proceed with interventions such as PET scanning or biopsy.

Dr. Lamborn suggested that two separate issues must be differentiated: firstly, the need to ensure that a temporary effect of treatment on imaging is not misinterpreted as disease progression; secondly, the fact that certain agents may require a period of time after delivery before their

effects become apparent. From a statistical perspective, it is acceptable to prospectively plan for allowing some time to elapse before counting apparent radiographic progression as disease progression. In this circumstance, however, it would be necessary to re-evaluate the historical data on PFS that she and Dr. Ballman had presented.

Dr. Yung noted that it may take 8 to 12 weeks for an antiangiogenic agent to exert a modulating effect on the tumor angiogenesis environment. The oncology community has debated the period of time that such agents can be given to patients before it is concluded that they are ineffective. In brain tumor therapy no standard approach to this problem has yet been agreed on.

Dr. Fine said that he knew of very few examples of patients who had been retained on therapy despite apparent evidence of progression who had subsequently responded to therapy. Dr. Pazdur noted that several drugs now used in oncology are continued after progression has been documented; in some but not all cases, this approach was prospectively planned in the studies that led to the drugs' approval. Dr. Meyers pointed out that the patient obtains no benefit from a therapy if his or her condition declines irreversibly during the time spent waiting for a drug to exert its effect. Dr. Buckner suggested that time to treatment failure might be an appropriate component of a composite endpoint.

Dr. Yung said it would be reasonable to allow time for certain classes of drugs to work even if there is apparent radiographic progression, provided that the patient remains clinically stable. Dr. Barker said it would be important to measure the symptomatic deterioration and weigh that against the potential eventual benefit of the therapy.

Composite Endpoints
2.1. What evaluation techniques discussed are complementary?

Dr. Pazdur said that the information FDA sought with this question was whether it would be reasonable to accept a composite endpoint that, for example, combined the findings of two radiologic tests (e.g., MRI and PET), or that combined radiologic and clinical endpoints, or that combined a radiologic endpoint with symptom measurement or patient-reported outcomes (PROs). Dr. Lamborn said that PFS was already a composite endpoint, although its precise components had not been documented.

Dr. Rock asked for comments from the panel on the cognitive function metric described by Dr. Meyers, which she had developed for trials of motexafin gadolinium (Xcytrin) in patients with brain metastases.

Dr. Paoletti observed that the role of PET had not been highlighted in the panel's discussions. Dr. Patronas responded that PET may be useful in some situations to supplement the information obtained from MRI or clinical evaluation but that it has not been validated to assess treatment response in brain tumors. He therefore could not recommend routine use of PET for this purpose in prospective studies. Dr. Provenzale agreed that it would currently be premature to use PET in Phase 3 studies in brain tumors but said it would be helpful to gather exploratory data on the use of PET in well-controlled Phase 2 studies. Dr. Yung said that resolution is currently inadequate in FDG-PET images of brain tumors. Dr. Fine said that PET has an important role to play in

understanding brain tumor biology but cannot be recommended for use in registration studies at this time.

Dr. Buckner said that given uncertainty about whether imaging changes are clinically meaningful in all circumstances, it would be helpful if radiographic evidence of a therapeutic effect could be complemented by evidence of functional or symptomatic improvement.

Endpoint Development
3.1. What if any potential endpoints should be explored apart from those discussed?

Dr. Fine observed that although the panel had not discussed the role of molecular and other biologic markers for segregating patient populations, such markers will play an increasingly important role not only in study design but also in the approval process for drugs to treat brain tumors as well as other cancers.

3.2. What questions should be brought from this workshop to the Oncologic Drugs Advisory Committee (ODAC) for further consideration?

Dr. Pazdur said that ODAC should be asked to consider whether 6-month PFS is an established surrogate for clinical benefit in brain tumor studies or a surrogate that is *reasonably likely* to predict clinical benefit (the standard for granting AA).

Dr. Yung said that ODAC should also be asked to consider the question of whether unidimensional, bidimensional, or volumetric approaches to tumor measurement are optimal. Dr. Provenzale added that some volumetric measurement techniques are highly reproducible and have a low rate of inter-reader variability, a factor that should be considered if such variability is a concern.

Dr. Fine said that ODAC should be asked whether a profound radiographic response rate in a singe-arm trial should be considered a surrogate endpoint that is reasonably likely to predict clinical benefit.

Dr. Lamborn suggested that a significant increase in 6-month PFS in a single-arm trial (e.g., 40% vs. 15%) might also be considered a surrogate endpoint that is reasonably likely to predict clinical benefit. Dr. Pazdur responded that, whereas response rate can be unequivocally considered to be a direct therapeutic effect, disease stabilization is influenced by many factors in addition to the experimental therapy. Randomization is the best way to account for such unknown factors. Because FDA must be satisfied that a drug truly has a therapeutic effect before approving it for marketing, the agency has been reluctant to accept time-to-event endpoints in single-arm trials.

Dr. Weiss said consideration should be given to the importance of obtaining confirmatory data after AA has been granted. Once a drug has been approved, however, it is often difficult to complete the trials necessary to confirm clinical benefit. Dr. Fine noted that in a rare disease such as a primary brain tumor, in which patients have few treatment options, it is difficult to recruit patients to randomized trials because the standard treatments offered in the control arm are

unattractive. Dr. Pazdur said this problem can be addressed by, for example, studying the drug in combination with another therapy (e.g., radiation) in the adjuvant setting or by conducting the confirmatory trial outside of the United States in a country where the drug is not yet approved.

Dr. Weiss asked Dr. Meyers for her suggestions on how to frame questions about PROs and symptom measurement for discussion by ODAC. Dr. Meyers responded that in addition to measuring response to therapy, neurocognitive function should also be measured in Phase 2 trials to provide information about possible injury to normal brain tissue.

Audience Questions and Comments

A member of the audience commented that targeted therapies may be most effective in subsets of patients. He asked what sort of metrics FDA would consider meaningful in a study testing a targeted therapy in a patient subset. Dr. Pazdur responded that this question would require a longer discussion than was possible at this meeting. In general, one would expect to see an above-average therapeutic effect when a targeted therapy is used in a patient subset; for this reason, endpoints other than survival could be considered. However, the agency has not clearly defined what endpoints it would consider specifically for targeted therapies.

Ms. Wiener (patient advocate) said that ODAC should be asked to consider rethinking the endpoints for brain tumor trials so that "longer life" and "better life" are not alternatives but are integrated ("longer life if it is better life").

WORKSHOP SUMMARY (Dr. Henry Friedman)

Dr. Friedman summarized the workshop proceedings, focusing on the following questions:

- *Can a unified set of outcome assessments be applied to primary brain tumors as a group?*

  There was a consensus among panel members that 6-month PFS was an endpoint that should be pursued in trials in the near future.

- *How well do existing and imagined imaging techniques assess or predict clinical benefit?*

  Imaging techniques assess or predict progression reasonably well, although there are concerns about reproducibility. They assess or predict response less well, except in the case of complete responses or a dramatically high response rate.

- *Might a unified PRO metric be validated to assess clinical benefit across both multiple therapeutic approaches and types of primary brain cancers?*

  There was consensus among panel members that PRO metrics are not yet sufficiently developed to be acceptable in registration trials in primary brain tumors.

In response to the comments made by Mr. Nesi (patient representative), Dr. Friedman said that every clinician who treats patients with brain tumors does so with the hope that each patient will

achieve longer survival accompanied by QoL that the clinician would find acceptable for a member of his own family. Extended survival with poor QoL is not a satisfactory outcome. Dr. Friedman suggested that FDA review studies with a view to trying to ensure that improvements in survival are not achieved at the expense of QoL. He added that PFS may be a better endpoint in terms of assuring acceptable QoL because, in his experience, it is uncommon for patients to deteriorate clinically while their tumor is under control.

BIOMARKER AND ENDPOINT RESEARCH PRIORITIES

Questions for discussion:

- *Which endpoints appear most promising and ready or nearly ready for clinical/regulatory application?*
- *What strategies are required to validate the most promising endpoints?*
    - *Are there ongoing or planned clinical trials that could incorporate these endpoints to facilitate validation?*
- *What are the most promising strategies to identify the next generation of promising endpoints/biomarkers for development?*
    - *What are the leading candidates for near-term development?*
- *How should the various promising imaging modalities be developed as biomarkers?*

Dr. Pazdur welcomed attendees to the final workshop session, the purpose of which was to identify endpoint-related issues that should be taken forward into new or existing clinical trials. The discussion was led by Dr. Jeffrey Abrams of NCI's Cancer Therapy Evaluation Program, Dr. Lalitha Shankar of NCI's Cancer Imaging Program, and Dr. Tracy Lugo-Lively of NCI's Cancer Diagnosis Program.

Dr. Abrams said that in brain tumors, the most promising potential endpoints (and those that were the focus of the most discussion during this workshop) seem to be imaging tests and HR-QoL endpoints.

Dr. Abrams noted that NCI's research program in brain tumors is fairly extensive considering the uncommon nature of the disease. NCI supports four Specialized Programs of Research Excellence in brain tumors, two research consortia on brain tumors in adults, and one research consortium on pediatric brain tumors. In addition, several of the NCI-supported cooperative groups, including the American College of Radiology Imaging Network, conduct research on brain tumors. NCI's Cancer Diagnosis Program supports a program for prognostic assessment of clinical cancer tests and the Cancer Imaging Program supports an imaging implementation group. NCI's Division of Cancer Control and Population Sciences supports an HR-QoL initiative.

Dr. Abrams said NCI would welcome opinions on where it should be investing in trying to bring new therapies to patients with brain tumors. For example, should the priority be to incorporate new imaging tests or neuropsychiatric tools early in drug development or to maximize the benefit to patients from drugs such as temozolomide? Should imaging tests focus on measuring tumor shrinkage or on functional imaging? Which imaging techniques should be used?

Dr. Shankar noted that the Cancer Imaging Program is funding several large imaging studies through its grants portfolio and is working to address significant issues such as standardization and validation in the clinical setting that currently impede the use of radiographic studies. Workshops have taken place in an effort to achieve consensus on the use of dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and FDG-PET. Consensus guidelines on the use of FDG-PET were issued in 2005 and are now being applied prospectively in all NCI-sponsored trials in which that technique is used. In November 2004 consensus was achieved on the use of DCE-MRI for body imaging; however, discussions are continuing on the use of this technique for brain imaging. NCI is working with the American College of Radiology to update existing guidelines on a 3-yearly cycle to ensure that they reflect current technology.

To address the logistical difficulties and costs associated with archiving and central reading of images, NCI is working to provide electronic image archiving for prospective studies and to enable images to be accessed and read centrally via the Internet. Experience from multiple trials has shown that central reading of images results in a more reproducible response rate. Data security will be employed to ensure that only investigators involved in a trial can access the data. Archived data that have been anonymized and annotated with clinical information will be available to the research community. NCI is also supporting pilot and early-phase studies to evaluate novel imaging agents.

Dr. Lively said that her branch's research portfolio is focused on the development of tissue- and serum-based prognostic and predictive markers. Most of these markers are not yet sufficiently well developed to be germane to the questions faced by today's panel. Nevertheless, NCI felt it was important for panelists and workshop attendees to be aware of ongoing research in this area. The Diagnostic Evaluation Branch supports both independent research projects and correlative studies associated with clinical trials to discover or confirm the importance of molecular or biochemical markers that could be useful in clinical decision-making. Experience with the approval of targeted agents to treat solid tumors has shown that diagnostic or predictive assays to guide the use of an agent need to be tested and validated before pivotal Phase 3 trials aimed at registration of an agent are begun.

Dr. Abrams asked the panel to suggest what critical trials NCI should be supporting in brain tumors. Dr. Pazdur asked for information about ongoing and proposed Phase 3 trials and the feasibility of embedding endpoints such as 6-month progression-free survival (PFS), neurocognitive testing, or time to symptomatic progression into these trials. Dr. Abrams responded that the only currently ongoing trial is a Radiation Therapy Oncology Group (RTOG) trial comparing intravenous carmustine with temozolomide as adjuvant therapy in high-grade gliomas; this trial has run into difficulty because of a shortage of intravenous carmustine and consideration is being given to converting to oral lomustine.

A large Phase 3 trial (RTOG-0525) comparing standard-dose with dose-dense temozolomide as adjuvant therapy for high-grade gliomas has just been launched in collaboration with the European Organization for Cancer Research and Treatment (EORTC). A Phase 3 trial is planned to compare the effectiveness of temozolomide in patients with and without deletions of

chromosome 1p and/or 19q. Other trials that are being considered would evaluate the role of temozolomide in subsets of patients such as the elderly and those with low-grade gliomas.

Dr. Yung said that RTOG-0525 had been designed with survival as the primary endpoint. It would be feasible to evaluate the correlation between 6-month PFS and overall survival in this trial. RTOG is submitting a separate grant application to evaluate the correlation of biomarkers with response and the effect of treatment on biomarkers. However, this might not be an appropriate trial in which to evaluate DCE-MRI because temozolomide is not a drug that modulates permeability and perfusion.

Dr. Fine said that Phase 3 trials provide a platform for the evaluation and validation of surrogate endpoints; the endpoints need not be related to the study drug. Evaluating endpoints in small groups of patients or in single-arm trials provides no information about the natural history of the disease or about how the endpoints might change in the presence of an effective therapy; evaluating endpoints in Phase 3 studies can address these limitations.

Dr. Abrams suggested that 6-month PFS could be studied as a secondary endpoint at a subset of centers participating in the RTOG trial that have the ability to standardize MRI scans. Dr. Pazdur said that it should be relatively simple to collect data for a "single point in time" endpoint such as 6-month PFS.

Dr. Pazdur added that because neurocognitive dysfunction is a cardinal feature of primary brain tumors, it is important to gain experience with neurocognitive testing in Phase 3 trials. Dr. Yung noted that the neurocognitive battery developed by Dr. Meyers had been validated in brain metastases in large trials supported by industry and by EORTC. Neurocognitive testing could be incorporated into any large trial provided that additional resources were made available to do it.

Dr. Paoletti said that industry would be willing to participate in the development and validation of neurocognitive testing instruments in Phase 2 and 3 randomized trials. He added that it is also extremely important to develop standardized ways of assessing patients' neurologic status and to try to correlate neurologic status with the site of the lesion. Industry would also appreciate guidance from NCI on the optimal approach to take to tumor measurement.

Dr. Barker said he suspected that some drugs now being tested as anti-tumor agents would fail in that capacity but would nevertheless reduce the volume of edema surrounding the enhancing mass and perhaps the apparent size of the tumor itself through the restoration of the disrupted blood-brain barrier, thus relieving symptoms; as such, they could be potential replacements for steroids. This issue could be addressed in small Phase 2 trials if the appropriate methodology existed to standardize across centers the measurement of neurologic changes and the measurement of vascular permeability, volume of peritumoral edema, and enhancing volume.

Dr. Abrams noted that for trials in low-grade gliomas, in which survival is not a useful endpoint, the NCI-supported cooperative groups are trying to develop an HR-QoL instrument that could be used either as a primary endpoint on its own or could be combined with an objective measure into a composite endpoint. At present different groups tend to favor different instruments. There is a need to develop validated instruments that are widely accepted.

Dr. Fine noted that most Phase 3 clinical trials in gliomas are exclusively supported by industry. He asked if it would be feasible for CTEP to fund an investigation of a particular potential endpoint within an industry-supported Phase 3 trial. Dr. Paoletti said he believed industry would be willing to collaborate in this way provided that agreement could be reached on intellectual property issues. Dr. Abrams said this would be a new mechanism for CTEP but he saw no reason why an effort could not be made in this direction. He noted that investigators in the NCI-supported brain tumor consortia work collaboratively with industry on many Phase 2 trials.

In response to a question from Dr. Yung, Dr. Shankar said that NCI is supporting a demonstration project in renal cell carcinoma to test the reliability of DCE-MRI in predicting treatment response and the feasibility of using DCE-MRI in a multi-center study. One hundred out of 300 patients enrolled in the study will receive DCE-MRI. Centers performing DCE-MRI must do so in accordance with trial guidelines and must meet quality assurance standards. Consideration could certainly be given to undertaking a similar study in a subset of patients with brain tumors.

Dr. Yung noted that the brain tumor consortia are currently running several large Phase 2 trials to evaluate agents in the class of tyrosine kinase inhibitors. When the consortia have proposed adding sub-studies to evaluate DCE-MRI, barriers have arisen related to funding or to concerns about the uniformity of imaging. Dr. Abrams responded that funding constraints necessitate limits on the use of MRI in NCI-supported trials. The challenge is to try to identify the trials in which the use of MRI is most likely to move the field forward. Dr. Shankar commented that in a trial in which patients are being routinely evaluated via MRI, the addition of a DCE-MRI evaluation components would add only 15 minutes to a patient visit.

Audience Questions and Comments

A member of the audience commented that a large amount of neurocognitive data already exists from completed trials. She asked whether NCI would be interested in funding a secondary analysis of this data to address some of the questions that had been raised during the panel meeting. Dr. Abrams said this sounded like a good idea and a good way to extract the maximum information from Phase 3 trials. He added that NCI's Division of Cancer Control and Population Sciences might have initiatives in this area of which he was unaware. Dr. Fine cautioned that existing data are relevant to brain metastases of systemic tumors, which have a very different biology and growth characteristics than primary brain tumors. It is therefore unclear whether analysis of neurocognitive and symptom-assessment data from patients with brain metastases will advance the knowledge base concerning primary brain tumors.

Another member of the audience noted that there are reasons to think volumetric measurement of irregularly shaped tumors may be more accurate than measurement of tumor diameter. He asked if it would be possible to evaluate the accuracy of the measurement methods used for images stored in the imaging archive that NCI is developing. Dr. Shankar responded that this issue is still under discussion.

Adjournment

Dr. Pazdur thanked the panelists, NCI representatives, and audience for their participation. The workshop was adjourned.