The Oncologist®

Selection of Response Criteria for Clinical Trials of Sarcoma Treatment

Scott M. Schuetze,[a] Laurence H. Baker,[a] Robert S. Benjamin,[b] Renzo Canetta[c]

[a]Department of Internal Medicine, Division of Hematology/Oncology, University of Michigan, Ann Arbor, Michigan, USA; [b]Department of Sarcoma Medical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA; [c]Bristol-Myers Squibb, Wallingford, Connecticut, USA

Abstract

Soft tissue sarcomas are a heterogeneous group of malignancies arising from mesenchymal tissues. A large number of new therapies are being evaluated in patients with sarcomas, and consensus criteria defining treatment responses are essential for comparison of results from studies completed by different research groups. The 1979 World Health Organization (WHO) handbook set forth operationally defined criteria for response evaluation in solid tumors that were updated in 2000 with the publication of the Response Evaluation Criteria in Solid Tumors (RECIST). There have been significant advances in tumor imaging, however, that are not reflected in the RECIST. For example, computed tomography (CT) slice thickness has been reduced from 10 mm to ≤2.5 mm, allowing for more reproducible and accurate measurement of smaller lesions. Combination of imaging techniques, such as positron emission tomography with fluorine-18-fluorodeoxyglucose ([18]FDG-PET) and CT can provide investigators and clinicians with both anatomical and functional information regarding tumors, and there is now a large body of evidence demonstrating the effectiveness of PET/CT and other newer imaging methods for the detection and staging of tumors as well as early determination of responses to therapy. The application of newer imaging methods has the potential to decrease both the sample sizes required for, and duration of, clinical trials by providing an early indication of therapeutic response that is well correlated with clinical outcomes, such as time to tumor progression or overall survival. The results summarized in this review support the conclusion that the RECIST and the WHO criteria for evaluation of response in solid tumors need to be modernized. In addition, there is a current need for prospective trials to compare new response criteria with established endpoints and to validate imaging-based response rates as surrogate endpoints for clinical trials of new agents for sarcoma and other solid tumors. *The Oncologist* 2008;13(suppl 2):32–40

Introduction

Soft tissue sarcomas are a heterogeneous group of malignancies arising from mesenchymal tissues [1]. Based on the Surveillance, Epidemiology, and End Results (SEER) database, nearly 15,000 new cases of sarcoma, both bone and soft tissue, are diagnosed in the U.S. each year [2]. Anthracyclines and ifosfamide have been established as the most active drugs for the treatment of patients with advanced, soft tissue sarcomas of most histologic subtypes, with the exception of gastrointestinal stromal tumors (GISTs). However, after failure of these drugs, patients with advanced soft tissue sarcomas have few treatment options [3]. Limitations in current therapeutic options for these tumors have prompted the development and evaluation of a very large number of new chemotherapeutic and biologic agents for their treatment [3].

Assessment of new therapies for sarcomas requires agreement on, and consistent use of, endpoints sensitive to the effects of these treatments. Longer survival is the generally accepted gold standard for demonstrating clinical benefit of an oncologic therapy. However, a wide range of surrogate endpoints has been employed as the basis for approving new therapeutic agents [4], and considerable controversy exists regarding which endpoints may be most appropriate for specific tumors [4]. This review article analyzes endpoints in oncology clinical trials, with a focus on sarcomas. This issue is timely, because there has been considerable evolution in approaches for assessment of these tumors and endpoints employed in clinical studies.

CURRENT ISSUES IN CONSIDERING ENDPOINTS

The acceptance of new cancer treatments requires demonstration of their benefit in clinical trials. Results from these studies may be confounded or invalidated by a large number of factors, including invalid comparisons in nonrandomized studies because of patient selection or stage migration, inadequate power of the study (i.e., sample size) to detect clinically meaningful differences in appropriately controlled randomized trials, use of multiple comparisons and retrospective analyses to make inappropriate conclusions from randomized trials, and endpoints that are inappropriate for the cancer under study [5]. Thus, documenting benefit for a new cancer therapy requires attention to study design and data analysis, and particularly to selection of study endpoints. Selection of endpoints for studies of new cancer therapies may be especially challenging because many of them require long-term patient follow-up and thus demand long-term studies. This conflicts with pressure to provide patients with rapid access to new treatments [6]. It has also been noted that there is greater tolerance for agents with high toxicity profiles that demonstrate efficacy, particularly in cancers where treatment options are limited, and that this may result in acceptance of therapies with high risk-to-benefit ratios [7]. The objective of a phase III oncology trial is to determine the risk-to-benefit ratio of an experimental agent or regimen, compared with a standard treatment [8].

Overall Survival

Overall survival (OS) has traditionally been the gold standard as the primary endpoint of phase III trials of cancer therapies [8]. When a randomized trial clearly demonstrates that an experimental drug produces a longer OS time than with standard therapy, approval is likely [9]. A number of advantages are associated with the use of OS as an endpoint for clinical trials of cancer therapies. It is objective and free from ambiguities in interpretation. However, it may be confounded by deaths from causes other than the

patient's cancer [8, 10]. The survival difference should not only be statistically significant but also regarded as clinically significant.

Several additional problems may arise with the use of OS as the primary endpoint in an oncology study. The current availability of multiple effective lines of systemic or local therapy, as well as treatment switching, may obscure the impact of the agent under study upon survival in an intent-to-treat analysis. In addition, the interval from the end of recruitment to primary efficacy analysis for OS is protracted, such that subsequent studies taking the best therapy from the previous trial cannot begin until years after the previous trial has completed recruitment. Trials based on OS, which require a minimum of 5 years to complete, are inevitably lengthy and expensive [10, 11] and inhibit drug development, especially in uncommon cancers such as sarcoma.

Time to Progression and Progression-Free Survival

Time to progression (TTP) and progression-free survival (PFS) time (defined as the time from randomization to death or progression, whichever comes first) are also used extensively as endpoints in clinical trials of cancer therapies. These endpoints are similar except that death is included in PFS [8, 10]. Both PFS and TTP are correlated with OS in patients with rectal cancer and can be considered as surrogates for it in this setting [11]. Perhaps the most important advantage associated with the use of PFS and TTP is that they permit smaller sample sizes and shorter study durations. Progression often occurs months to years before death, and differences in the efficacies of new drugs or therapeutic regimens can be detected with shorter follow-up. Another advantage of these endpoints is that TTP and PFS do not require shrinkage of the tumor mass for detection of differences between treatments, which may make them highly suitable for measuring the benefits of cytostatic agents. In addition, disease progression is often the basis for a change in therapy and thus has high applicability to clinical practice [9–11]. Most importantly, TTP or PFS after a single line of therapy is a direct measure of the benefit from that treatment and is not confounded by subsequent events.

However, both PFS and TTP also have disadvantages as surrogate endpoints for OS in clinical trials. The clinical significance of small differences in TTP or PFS may be unclear (as in OS), especially when one is evaluating toxic treatments, and careful assessment of progression at frequent intervals can be costly and labor-intensive. There are also concerns about ascertainment bias in unblinded trials and questions about the reliability of modest differences in TTP or PFS that are often observed in such studies. Because OS is not the primary endpoint in studies employing either

PFS or TTP as surrogates, a design to cross over to the investigational treatment from the standard or placebo arm when tumor progression occurs can also be considered. A crossover design has the potential to improve patient enrollment and patient benefit. However, it is also a weakness in that it can dilute the contribution of a survival benefit [10, 11].

Response Rates

Response rates are also used to assess efficacy in trials of cancer therapies. Response measures have been variously defined. However, a complete response (CR) generally indicates tumor disappearance, a partial response (PR) indicates a >50% reduction in the tumor cross product (multiplication of the maximum tumor diameter in the axial plane by the largest diameter and its perpendicular dimension on the same image), and stable disease (SD) indicates a <50% reduction to <25% increase in the cross product [12]. Advantages associated with the use of these measures include the lack of a dilution effect with smaller sample sizes, shorter study durations, and tumor shrinkage, which is clearly and solely dependent on a therapeutic intervention, as spontaneous regressions are quite rare. Disadvantages include potential for bias in unblinded studies, variability in results across studies and centers within an individual trial secondary to differences in criteria and methods of assessment, and the fact that a response to treatment may not necessarily equate with clinical benefit. Response to treatment may or may not be a true surrogate for survival [10].

THE NEED FOR REDEFINING RESPONSE

The World Health Organization Criteria

Consensus criteria for defining a response to cancer therapy are essential for comparison of data from studies carried out by different research and clinical trial groups. In 1979, the World Health Organization (WHO) handbook set forth four specific operationally defined criteria for the codification of response evaluation in solid tumors. In this scheme, the lesions are measured bidimensionally and the product of the greatest tumor diameter and the greatest perpendicular distance, summed over all measured tumors, is used in determining responses. The four response categories were: CR, tumor disappearance confirmed at 4 weeks; PR, 50% or greater decrease in tumor size confirmed at 4 weeks; no change (NC), neither PR nor PD criteria met; and progressive disease (PD), 25% or greater increase in tumor size with no CR, PR, or SD documented before increased disease [13, 14].

Three major problems with these definitions gradually became apparent with their use in clinical trials [15, 16]. Methods of integrating the change in tumor size into

response assessments varied among research groups, minimum lesion size and number of lesions documented varied from one study to the next, and what constituted PD was based on the change in size of a single lesion by some researchers and a change in the overall tumor load (including measurements of all lesions) by others. The advent of new technologies, particularly computed tomography (CT) and magnetic resonance imaging (MRI), further confused matters with respect to the relevance of volumetric and three-dimensional measurements versus bidimensional measures in response assessments. The combination of all these factors resulted in a situation in which response criteria were no longer comparable among research organizations. This was the circumstance that the original WHO publication had aimed to avoid.

Response Evaluation Criteria in Solid Tumors

The Response Evaluation Criteria in Solid Tumors (RECIST) were developed in response to problems with the WHO criteria. The RECIST were published in 2000 and are a simplification of four other methods of assessing solid tumor responses. The RECIST are generally similar to the criteria set forth by the 1979 WHO handbook, with the major change being that the RECIST employ unidimensional measurements of the sum of the longest diameters of tumors in the axial plane instead of the conventional bidimensional WHO method of the product of the longest diameter and that perpendicular to it, summed over all measured tumors. The RECIST response categories are: CR, disappearance of tumor sustained for at least 4 weeks; PR, ≥30% decrease in tumor sustained for at least 4 weeks; SD, neither PR nor PD criteria met; and PD, ≥20% increase with no CR, PR, or SD documented before increase of disease [12]. A detailed comparison of the WHO criteria and the RECIST and their associated guidelines is provided in Table 1 [17].

Potential advantages of the RECIST over the WHO criteria [17] include the fact that the RECIST give specific size requirements for measurable lesions at baseline, distinguish target from nontarget lesions, allow the maximum number of target lesions to be followed up to a total of 10, and provide a baseline tumor burden (smallest sum of longest diameters from the start of treatment) for determining PD. The RECIST also state that all target lesions should be measured to determine PD instead of one or more measurable lesions [17].

The RECIST are predicated on unidimensional and bidimensional measurements being comparable and assume metastases are spherical and change proportionally. Application of the WHO criteria and RECIST to the same patients in 14 studies with a wide range of cancers indicated very similar results for all response categories. Results from

Table 1. Comparison of WHO response criteria and RECIST [17]

Characteristic	WHO	RECIST
Measurability of lesions at baseline	1. Measurable, bidimensional (product of LD and greatest perpendicular diameter)[a]	1. Measurable, unidimensional (LD only, size with conventional techniques ≥20 mm; spiral computed tomography ≥10 mm)
	2. Nonmeasurable/evaluable (e.g., lymphangitic pulmonary metastases, abdominal masses)	2. Nonmeasurable: all other lesions, including small lesions. Evaluable is not recommended.
Objective response	1. Measurable disease (change in sum of products of LDs and greatest perpendicular diameters, no maximum number of lesions specified): CR, disappearance of all known disease, confirmed at ≥4 wks; PR, ≥50% decrease from baseline, confirmed at ≥4 wks; PD, ≥25% increase of one or more lesions, or appearance of new lesions; NC, neither PR nor PD criteria met	1. Target lesions (change in sum of LDs, maximum of 5 per organ up to 10 total [more than one organ]): CR, disappearance of all target lesions, confirmed at ≥4 wks; PR, ≥30% decrease from baseline, confirmed at 4 wks; PD, ≥20% increase over smallest sum observed, or appearance of new lesions; SD, neither PR nor PD criteria met
	2. Nonmeasurable disease: CR, disappearance of all known disease, confirmed at ≥4 wks; PR, estimated decrease of ≥50%, confirmed at ≥4 wks; PD, estimated increase of ≥25% in existent lesions or appearance of new lesions; NC, neither PR nor PD criteria met	2. Nontarget lesions: CR, disappearance of all target lesions and normalization of tumor markers, confirmed at ≥4 wks; PD, unequivocal progression of nontarget lesions, or appearance of new lesions; non-PD: persistence of one or more nontarget lesions and/or tumor markers above normal limits
Overall response	1. Best response recorded in measurable disease	1. Best response recorded in measurable disease from treatment start to disease progression or recurrence
	2. NC in nonmeasurable lesions will reduce a CR in measurable lesions to an overall PR	2. Non-PD in nontarget lesion(s) will reduce a CR in target lesion(s) to an overall PR
	3. NC in nonmeasurable lesions will not reduce a PR in measurable lesions	3. Non-PD in nontarget lesion(s) will not reduce a PR in target lesion(s)
Duration of response	1. CR: from date CR criteria first met; to date PD first noted	1. Overall CR: from date CR criteria first met; to date recurrent disease first noted
	2. Overall response: from date of treatment start; to date PD first noted	2. Overall response: from date CR or PR criteria first met (whichever status came first); to date recurrent disease or PD first noted
	3. In patients who only achieve a PR, only the period of overall response should be recorded	3. SD: from date of treatment start; to date PD first noted

[a]Lesions that can only be measured unidimensionally are considered to be measurable (e.g., mediastinal adenopathy, malignant hepatomegaly).
Abbreviations: CR, complete response; LD, longest diameter; NC, no change; PD, progressive disease; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SD, stable disease; WHO, World Health Organization.

this analysis indicated that 91.9% of patients evaluated had the same date of disease progression with the WHO criteria and RECIST; 7.3% had earlier disease progression with the WHO criteria and 0.9% had earlier disease progression with the RECIST (Table 2) [12]. This change is important to PFS since the PFS time by the RECIST will be longer than by the WHO criteria.

Problems with the RECIST in Clinical Trials

The RECIST have been used extensively in clinical trials and are generally viewed as an important advancement over the criteria provided by the WHO [18]. However, these guidelines and criteria do have significant limitations. The RECIST are intentionally terse, and this can lead to confusion in how to apply appropriate measurement techniques across centers [16]. In addition, unidimensional measurements of the type set forth in the RECIST may not be suit-

Table 2. Proportions of patients with disease progression by the WHO criteria and the RECIST [12]

	n of patients	%
Total *n* of progressors	234	100
Progression by appearance of new lesions[a]	118	50
Progression by increase in pre-existing measurable disease	116	50
Same date of disease progression by WHO criteria and RECIST	215	91.9
Different date of disease progression	19	8.1
Earlier PD with WHO criterion	17	7.3
Earlier PD with unidimensional criterion	2	0.9

[a]Also includes a few patients with PD because of a marked increase of nonmeasurable disease.
Abbreviations: PD, progressive disease; RECIST, Response Evaluation Criteria in Solid Tumors; WHO, World Health Organization.

able for all tumor types, most notably those with nonspherical growth patterns (e.g., malignant pleural mesothelioma) [19–22]. Definition of target lesions may also limit the utility of the RECIST. Target lesions defined by the RECIST may not represent burden of disease [23], and changes in tumor characteristics may confound evaluation (e.g., pleural effusions in lung cancer [24]).

There are also limitations of the RECIST with respect to determination of disease progression. Response assessment as measured by the RECIST has been shown to have some discrepancies with WHO-determined responses. These appear to occur most often at the PR–SD and SD–PD "borders." This difference may be problematic when new experimental therapies are compared with conventional agents whose response rates have been established in historical trials. The apparent lower rate of disease progression with the RECIST may mean that more patients remain on therapy, and the percentages of patients with SD thus need to be interpreted with caution [25].

The RECIST also ignore the fact that changes in tumor size may not be directly correlated with disease progression in all therapeutic situations. Qualitative changes in tumors (e.g., myxoid degeneration in GIST) may not be reflected in tumor measurements, and this can result in erroneous classification of the response to treatment. Standard anatomic imaging techniques are often inadequate for evaluating malignancies, particularly when monitoring treatment responses for agents that do not cause tumor shrinkage (i.e., cytostatic agents) or for slow-progressing cancers or those malignancies that metastasize diffusely [26]. Thus, morphologic evaluation based solely on one- or two-dimensional measurements may not directly reflect biological changes in tumors associated with either the disease itself or its treatment [27]. Moreover, anatomical changes in the tumor as described by the RECIST may be detected later than functional changes in some circumstances (e.g., in GISTs treated with imatinib) [18]. The use of a primary tumor for response assessment, if the tumor is localized in a hollow organ (e.g., the esophagus), also makes measurements based on the RECIST difficult [18].

Finally, it is important to remember that the RECIST were developed on the basis of discussions carried out in the 1990s and published in 2000. As a result, they do not reflect many advances in imaging technology that have occurred over the past decade. Newer imaging and image-processing modalities may allow changes not considered in the RECIST to be included in revised response criteria [28]. For example, a comparison of relative values of manual unidimensional measurements and automated volumetry with multidetector-row computed tomography (MDCT) for longitudinal treatment response assessment in patients

with pulmonary metastases indicated that MDCT provided better reproducibility of response evaluation and should be preferred over manual measurements in these patients [28]. The following section further explores the application of newer imaging technologies in assessing the efficacy of therapies for solid tumors.

IMAGING-BASED EVALUATION OF RESPONSE TO CANCER THERAPY

Great advances in image acquisition and processing techniques are improving both staging of solid tumors for treatment planning and evaluation of new therapies. Surrogate markers of tumor response are also being developed with the use of functional imaging techniques that provide greater insight into tumor responses to therapy (e.g., changes in tumor perfusion, permeability, blood volume, and oxygenation). Combination of imaging techniques, such as positron emission tomography (PET) and computed tomography (CT), can provide investigators and clinicians with both anatomical and functional information. There is now substantial evidence that the use of PET with fluorine-18-fluorodeoxyglucose (^{18}FDG-PET), particularly in conjunction with CT, can improve the accuracy of cancer staging with a high sensitivity for detecting small-volume disease [29]. The application of newer imaging methods that permit more rapid and precise evaluation of tumors has the potential to decrease both the sample sizes required for and duration of clinical trials by providing an early indication of therapeutic response that is well correlated with clinical outcomes (e.g., OS) for chemo- and radiotherapy [26]. The use of these new modalities and advances in transmission, storage, quality assurance, and analysis of images could streamline clinical trials of new treatments and accelerate new drug approvals [26].

CT

As noted above, the RECIST for evaluating responses to treatment have been criticized because they do not reflect biological changes in solid tumors induced by new targeted therapies and thus may provide misleading results. Modified objective criteria using a combination of tumor size and density on CT have shown promise in early response evaluation and in predicting long-term outcomes in patients with advanced GISTs treated with imatinib. Results from Choi and colleagues indicated that tumor size determined using the sum of the longest dimensions and the RECIST definitions for a significant change were not reliable and underestimated the tumor response to imatinib during the early post-treatment stage in patients with metastatic GISTs. The mean tumor density, however, decreased significantly 2 months after treatment compared with pretreatment val-

ues. Moreover, evaluation using a combination of tumor size, tumor density, and absence or presence of tumor nodules and tumor vessels was a better indicator of the tumor response to imatinib than tumor density alone [27].

Choi and associates evaluated a series of 40 patients treated with imatinib for recurrent or metastatic GISTs who had undergone both PET and CT evaluation to determine the CT findings that could differentiate those who had a good response by PET and those who did not [30]. They found that a decrease in tumor size of ≥10% or a decrease in tumor density of ≥15% identified 97% of good responders by PET and none of the seven poor responders [30]. They also demonstrated that response defined by these new CT criteria was correlated with longer TTP, whereas response by the RECIST was not [30].

Benjamin and associates confirmed the observations of Choi et al. [30] in a separate group of 58 patients and then evaluated all 98 patients by the RECIST and the Choi criteria. All patients had pretreatment and follow-up CT scans. Disease-specific survival (DSS) and TTP were analyzed by response category. There were 45 (46%) good responders and 53 (54%) poor responders by the RECIST. In contrast, there were 81 (83%) good responders and 17 (17%) poor responders by the Choi criteria [31]. Despite the almost doubling of the response rate when patients were assessed by the Choi criteria versus the RECIST, patients with good responses by the Choi criteria on CT at 8 weeks after the start of treatment had equivalent DSS to that of patients with a CR or PR at any time by the RECIST. In addition, TTP and DSS were significantly correlated with the Choi response group, but not with the response group by the RECIST [31]. These results support the conclusion that the Choi response criteria, which incorporate tumor density and small changes in tumor size on CT, are more sensitive and accurate than the RECIST in assessing the response of GISTs to imatinib treatment. These results have been reproduced at other institutions; however, further validation needs to be completed [32].

Advances in CT technology are likely to further increase its usefulness for the evaluation of cancer therapies. Greater numbers of detectors in CT scanners offer better three-dimensional reconstruction and volumetric measurement [33], but the lack of a sufficient number of centers with appropriate scanners to process data limits the organization of large-scale, multicenter clinical trials. Automated collection and analysis of CT data are vital but not widely available, and manual data collection and analysis are expensive and time-consuming [34]. Another important limitation of CT methods is that heterogeneity of tumors (e.g., hypoxic regions) can confound volumetric measurements [35].

[18]FDG-PET

[18]FDG-PET can assess tumor glucose use with high reproducibility. Following therapy, the decrease in glucose uptake correlates with a reduction in viable tumor cells. In contrast to CT, MRI, or ultrasound, PET imaging allows identification of responding and nonresponding tumors early in the course of therapy. PET imaging can easily demonstrate changes in metabolic activity and indicate, sometimes within hours of the first treatment, whether or not a patient will respond to a particular therapy. [18]FDG-PET has demonstrated efficacy for monitoring therapeutic response in a wide range of cancers, including breast, esophageal, lung, and head and neck cancers, and lymphoma [36, 37].

Effectiveness of [18]FDG-PET

[18]FDG-PET is useful for determining the responses of GISTs to treatment with imatinib and may be superior to standard anatomic criteria for early evaluation of the responses of GISTs to targeted molecular therapies [38]. As noted above, evaluation using the RECIST may be poorly suited to these tumors because they may have a strong positive response to treatment (e.g., decreased FDG uptake, clinical improvement) without major shrinkage [27].

[18]FDG-PET scanning has also been shown to be a useful method for prediction of outcomes in patients with high-grade extremity soft tissue sarcomas treated with chemotherapy. Schuetze and colleagues evaluated 46 patients with high-grade localized sarcomas with [18]FDG-PET. The maximum standardized uptake value (SUV_{max}) of tumors was measured before neoadjuvant chemotherapy and again prior to surgery. Resected specimens were examined for residual viable tumors. Patients with a baseline tumor $SUV_{max} \geq 6$ and a <40% decrease in [18]FDG uptake were at high risk for systemic disease recurrence, estimated to be 90% at 4 years from the time of initial diagnosis. Patients whose tumors had a ≥40% decline in SUV_{max} in response to chemotherapy were at a significantly lower risk for recurrent disease and death after complete resection and adjuvant radiotherapy (Fig. 1) [39]. [18]FDG-PET results have also been shown to correlate closely with histologic responses of tumors to chemotherapy. Results from 36 patients with osteosarcoma or Ewing's sarcoma–family tumors who received neoadjuvant therapy indicated that a good [18]FDG-PET response was concordant with a histologic response in 68%–69% of patients. In addition, a lower SUV_{max} after neoadjuvant chemotherapy (SUV2) was associated with better long-term outcomes in this small study cohort (the 4-year PFS rate was 72% for SUV2 <2.5 versus 27% for SUV2 ≥2.5; $p = .01$ for all patients) [40].

Results from additional studies have demonstrated that [18]FDG-PET is useful for the staging of solid tumors and for
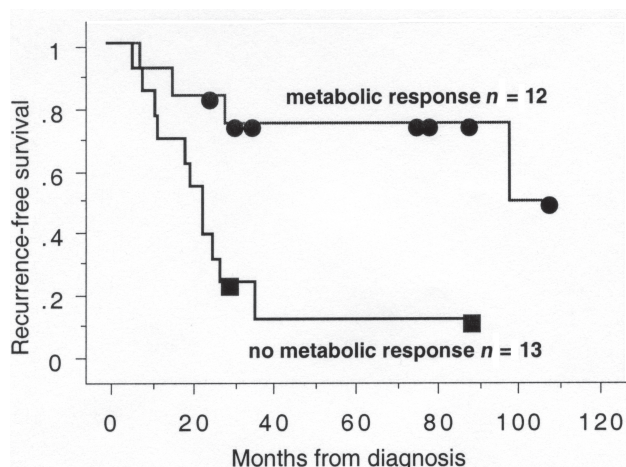
Figure 1. Kaplan–Meier plot demonstrating a significantly higher risk for metastatic disease recurrence in patients with a pretreatment sarcoma maximum standardized uptake value (SUV_{max}) ≥6 and poor metabolic response to neoadjuvant chemotherapy ($p = 0.004$) [43].

assessing responses to neoadjuvant therapy [41, 42]. Review of the results from 16 studies of patients with esophageal carcinoma indicated that the accuracy of ^{18}FDG-PET in assessing response to treatment was similar to that for endo-esophageal ultrasonography and significantly superior to that for CT. The staging value of ^{18}FDG-PET was limited for the detection of locoregional metastases, but the technique was effective for the detection of distant lymphatic and hematogenous metastases [42]. ^{18}FDG-PET has also been shown to be useful for monitoring results of therapy in patients with stage III non-small cell lung cancer [41] and for identification of progressing lesions and for detecting flares in tumor lesions that were previously under control [43, 44].

New positronic substrates will likely expand the utility of PET [44]. The most widely used PET tracer for osteosarcoma is ^{18}FDG. The other clinical PET tracer with reported utility for osteosarcoma imaging is ^{18}F-fluoride ion. ^{18}F-labeled monoclonal antibodies, ^{18}F-fluoromisonidazole, ^{18}F-labeled arginine–glycine–aspartic acid (RGD)-containing glycopeptide, ^{3}H-thymidine, ^{13}N-methionine, and PET of p53 transcriptional activity in osteosarcoma are all being investigated [44].

^{18}FDG-PET Issues
While ^{18}FDG-PET has clearly been demonstrated as an important advance for staging, prognosis, and evaluation of treatment responses in patients with solid tumors, this technique does have significant limitations. These include inability to characterize lesions <1 cm in diameter, difficulty in distinguishing benign post-therapy or unrelated

inflammatory responses from the effects of treatment on tumors, and variability in signal acquisition across instruments and in interpretation and analysis among readers [26]. Multi-institutional studies using ^{18}FDG-PET need to consider the potential impact of this variability on overall results. It has also been noted that ^{18}FDG-PET evaluations can be very subjective, and that even SUV is only a semi-quantitative measure [45]. In addition, static ^{18}FDG uptake indices alone may not enable adequate differentiation between benign and malignant lesions. While quantitative dynamic imaging may provide more helpful information, it is more labor-intensive and costly. Another limitation of ^{18}FDG-PET is variable ^{18}FDG uptake in normal structures and sites of inflammation caused by infection or foreign bodies, leading to false-positive results [46].

MRI
Ongoing efforts are developing more powerful methods for automated classification of MRI spectra, based on the acquisition of large datasets of tumor spectra and use of diffusion- and perfusion-weighted imaging. These methods are useful for distinguishing between tumors and abscesses and for predicting responses to radiotherapy, respectively [47].

Magnetic resonance spectroscopy may predict response in a manner analogous to PET [48]. Both of these methods permit imaging of the entire body and combine functional and anatomical information. ^{18}FDG-PET and MRI spectroscopy are valuable techniques for monitoring tumor response in patients undergoing chemo- and radiotherapy, particularly when evaluating early responses. In contrast, MRI is particularly useful for assessing metastasis and infiltration of bone marrow and the central nervous system.

Dynamic contrast-enhanced MRI (DCE-MRI) is a new imaging method for assessing the physiological state of tumor vascularity in vivo. This method uses available imaging techniques and contrast agents and assays the kinetics of tumor enhancement during bolus i.v. contrast administration [49]. DCE-MRI has been shown to be useful for detecting microvascular changes in tumors in response to isolated limb perfusion within 24 hours of treatment in experimental animals [50] and for correctly predicting tumor responses to therapy in a small cohort of 12 patients with histologically proven high-grade soft tissue sarcoma [51]. DCE-MRI correctly predicted tumor response in 8 of 10 evaluable patients. Early rapidly progressive enhancement was correlated histologically with residual viable tumors, and late and gradual, or absence of, enhancement was correlated with necrosis, predominantly centrally located, or granulation tissue [51]. These preliminary results show that DCE-MRI offers the potential for noninvasive monitoring of responses to isolated limb perfusion in soft tissue sarcomas.

FUTURE DIRECTIONS/CONSENSUS

The results summarized in this review support the conclusion that the RECIST and the WHO criteria for evaluation of response in solid tumors need to be replaced by alternative endpoints. While these criteria have been useful, change is mandated in order to provide more rapid assessment of tumor responses and to reflect the advances in imaging technology that have occurred over the past decade. Prospective trials are needed to compare new response criteria with established endpoints and to validate imaging-based response rates as surrogates for traditional endpoints. New approaches developed for evaluation of sarcomas should subsequently be assessed in other tumor types. However, the underlying biology of tumors may impact the applicability of imaging-based criteria, and response kinetics and landmarks may vary among tumor types. In addition, alternative response criteria may have different prognostic values for cytotoxic versus cytostatic therapies.

REFERENCES

1   Fahn W, Issels RD. Emerging treatments for soft tissue sarcoma of adults. Expert Opin Emerg Drugs 2004;9:313–334.

2   Borden EC, Baker LH, Bell RS et al. Soft tissue sarcomas of adults: State of the translational science. Clin Cancer Res 2003;9:1941–1956.

3   Milano A, Apice G, Ferrari E et al. New emerging drugs in soft tissue sarcoma. Crit Rev Oncol Hematol 2006;59:74–84.

4   Gollob JA, Bonomi P. Historic evidence and future directions in clinical trial therapy of solid tumors. Oncology (Williston Park) 2006;20(suppl 5):10–18.

5   Tannock IF. Some problems related to the design and analysis of clinical trials. Int J Radiat Oncol Biol Phys 1992;22:881–885.

6   National Cancer Institute. Understanding the Approval Process for New Cancer Treatments. Available at http://www.cancer.gov/ clinicaltrials/ learning/approval-process-for-cancer-drugs/page5. Accessed June 1, 2007.

7   Phan A, Patel S. Advances in neoadjuvant chemotherapy in soft tissue sarcomas. Curr Treat Options Oncol 2003;4:433–439.

8   Rubinstein LV. Therapeutic studies. Hematol Oncol Clin North Am 2000;14:849–876, ix.

9   U.S. Food and Drug Administration. ASCO/FDA Lung Cancer Endpoints Workshop. April 15, 2003. Available at http://www.fda.gov/cder/drug/ cancer_endpoints/lungFinalSummary.doc. Accessed June 1, 2007.

10 Scott J, McGettigan G. Regulatory approvals for oncology products based on accelerated clinical development and limited data packages-2. Regulatory Rapporteur July/August 2005:6–15.

11 Glynne-Jones R, Mawdsley S, Pearce T et al. Alternative clinical end points in rectal cancer—are we getting closer? Ann Oncol 2006;17:1239–1248.

12 Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92:205–216.

13 Miller AB, Hoogstraten B, Staquet M et al. Reporting results of cancer treatment. Cancer 1981;47:207–214.

14 WHO Handbook for Reporting Results of Cancer Treatment. Geneva, Switzerland: World Health Organization, 1979:1–45.

15 Tsuchida Y, Therasse P. Response evaluation criteria in solid tumors (RECIST): New guidelines. Med Pediatr Oncol 2001;37:1–3.

16 McHugh K, Kao S. Response evaluation criteria in solid tumours (RECIST): Problems and need for modifications in paediatric oncology? Br J Radiol 2003;76:433–436.

17 Gehan EA, Tefft MC. Will there be resistance to the RECIST (Response Evaluation Criteria in Solid Tumors)? J Natl Cancer Inst 2000;92:179–181.

18 Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: A review of validation studies on tumour assessment. Eur J Cancer 2006;42:1031–1039.

19 van Klaveren RJ, Aerts JG, de Bruin H et al. Inadequacy of the RECIST criteria for response evaluation in patients with malignant pleural mesothelioma. Lung Cancer 2004;43:63–69.

20 Tropine A, Dellani PD, Glaser M et al. Differentiation of fibroblastic meningiomas from other benign subtypes using diffusion tensor imaging. J Magn Reson Imaging 2007;25:703–708.

21 Chong S, Lee KS, Chung MJ et al. Neuroendocrine tumors of the lung: Clinical, pathologic, and imaging findings. Radiographics 2006;26:41–57; discussion 57–58.

22 Miller TR, Pinkus E, Dehdashti F et al. Improved prognostic value of 18F-FDG PET using a simple visual analysis of tumor characteristics in patients with cervical cancer. J Nucl Med 2003;44:192–197.

23 Trillet-Lenoir V, Freyer G, Kaemmerlen P et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: Accuracy of the RECIST criteria. Br J Radiol 2002;75:903–908.

24 Bond M, Bernstein ML, Pappo A et al. A phase II study of imatinib mesylate in children with refractory or relapsed solid tumors: A Children's Oncology Group study. Pediatr Blood Cancer 2007; Jan 29 [Epub ahead of print].

25 Mazumdar M, Smith A, Schwartz LH. A statistical simulation study finds discordance between WHO criteria and RECIST guideline. J Clin Epidemiol 2004;57:358–365.

26 Kelloff GJ, Sullivan DM, Wilson W et al. FDG-PET Lymphoma Demonstration Project Invitational Workshop. Acad Radiol 2007;14:330–339.

27 Choi H, Charnsangavej C, de Castro Faria S et al. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: A quantitative analysis correlated with FDG PET findings. AJR Am J Roentgenol 2004;183:1619–1628.

28 Marten K, Auer F, Schmidt S et al. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. Eur Radiol 2006;16:781–790.

29 Koh DM, Cook GJ, Husband JE. New horizons in oncologic imaging. N Engl J Med 2003;348:2487–2488.

30 Choi H, Charnsangavej C, Faria C et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: Proposal of new computed tomography response criteria. J Clin Oncol 2007;25:1753–1759.

31 Benjamin RS, Choi H, Macapinlac HA et al. We should desist using RECIST, at least in GIST. J Clin Oncol 2007;25:1760–1764.

32 Bulusu VR, Jephcott CR, Fawcett S et al. RECIST and Choi criteria for response assessment (RA) in patients with inoperable and metastatic gastrointestinal stromal tumours (GISTs) on imatinib mesylate. Cambridge GIST study group experience. J Clin Oncol 2007;25(suppl 18S):549s.

33 Gimel P, Hubert J, Iochum S et al.]Contribution of the latest generation CT in preoperative assessment of kidney cancer]. Prog Urol 2002;12:1310–1317. French.

34 Lau D, Seibert A, Gandara D et al. Computer-assisted image analysis of bronchioloalveolar carcinoma. Clin Lung Cancer 2005;6:281–286.

35 Castillo E, Lawler LP. Diagnostic radiology and nuclear medicine. J Surg Oncol 2005;92:191–202.

36 Avril NE, Weber WA. Monitoring response to treatment in patients utilizing PET. Radiol Clin North Am 2005;43:189–204.

37 Hawkins DS, Rajendran JG, Conrad EU 3rd et al. Evaluation of chemotherapy response in pediatric bone sarcomas by [F-18]-fluorodeoxy-D-glucose positron emission tomography. Cancer 2002;94:3277–3284.

38. Van den Abbeele AD. The lessons of GIST—PET and PET/CT: A new paradigm for imaging. *The Oncologist* 2008;13(suppl 2):8–13.

39 Schuetze SM, Rubin BP, Vernon C et al. Use of positron emission tomography in localized extremity soft tissue sarcoma treated with neoadjuvant chemotherapy. Cancer 2005;103:339–348.

40 Hawkins DS, Schuetze SM, Butrynski JE et al. [18F]Fluorodeoxyglucose positron emission tomography predicts outcome for Ewing sarcoma family of tumors. J Clin Oncol 2005;23:8828–8834.

41 Eschmann SM, Friedel G, Paulsen F et al. Repeat 18F-FDG PET for monitoring neoadjuvant chemotherapy in patients with stage III non-small cell lung cancer. Lung Cancer 2007;55:165–171.

42 Westerterp M, Van Westreenen HL, Sloof GW et al. Role of positron emission tomography in the (re-)staging of oesophageal cancer. Scand J Gastroenterol Suppl 2006;(243):116–122.

43 Nanni C, Rubello D, Castellucci P et al. 18F-FDG PET/CT fusion imaging in paediatric solid extracranial tumours. Biomed Pharmacother 2006;60:593–606.

44 Brenner W, Bohuslavizki KH, Eary JF. PET imaging of osteosarcoma. J Nucl Med 2003;44:930–942.

45 Calvo R, Marti-Climent JM, Richter JA et al. Three-dimensional clinical PET in lung cancer: Validation and practical strategies. J Nucl Med 2000;41:439–448.

46 Fukui MB, Blodgett TM, Snyderman CH et al. Combined PET-CT in the head and neck: Part 2. Diagnostic uses and pitfalls of oncologic imaging. Radiographics 2005;25:913–930.

47 Rees J. Advances in magnetic resonance imaging of brain tumours. Curr Opin Neurol 2003;16:643–650.

48 Rahmouni A, Luciani A, Itti E. MRI and PET in monitoring response in lymphoma. Cancer Imaging 2005;5:S106–S112.

49 Rosen MA, Schnall MD. Dynamic contrast-enhanced magnetic resonance imaging for assessing tumor vascularity and vascular effects of targeted therapies in renal cell carcinoma. Clin Cancer Res 2007;13:770s–776s.

50 Preda A, Wielopolski PA, Ten Hagen TL et al. Dynamic contrast-enhanced MRI using macromolecular contrast media for monitoring the response to isolated limb perfusion in experimental soft-tissue sarcomas. MAGMA 2004;17:296–302.

51 van Rijswijk CS, Geirnaerdt MJ, Hogendoorn PC et al. Dynamic contrast-enhanced MR imaging in monitoring response to isolated limb perfusion in high-grade soft tissue sarcoma: Initial results. Eur Radiol 2003;13:1849–1858.