VOLUME 25 · NUMBER 13 · MAY 1 2007

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

We Should Desist Using RECIST, at Least in GIST

Robert S. Benjamin, Haesun Choi, Homer A. Macapinlac, Michael A. Burgess, Shreyaskumar R. Patel, Lei L. Chen, Donald A. Podoloff, and Chuslip Charnsangavej

From the Department of Sarcoma Medical Oncology and the Division of Diagnostic Imaging, The University of Texas M.D. Anderson Cancer Center, Houston, TX.

Submitted May 8, 2006; accepted February 5, 2007.

Supported by National Cancer Institute Contracts No. U01-CA70172-01 and N01-CM-17003.

Presented in part at the 8th Annual Meeting of the Connective Tissue Oncology Society, November 6-8, 2003, Barcelona, Spain; and the 42nd Annual Meeting of the American Society of Clinical Oncology, June 2-6, 2006, Atlanta, GA.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to Robert S. Benjamin, MD, Department of Sarcoma Medical Oncology, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Blvd, Unit 450, Houston, TX 77030; e-mail: rbenjami@mdanderson.org.

0732-183X/07/2513-1760/$20.00

DOI: 10.1200/JCO.2006.07.3411

ABSTRACT

Purpose
Response Evaluation Criteria in Solid Tumors (RECIST) are insensitive in evaluating imatinib-treated gastrointestinal stromal tumors (GISTs). Response by Choi criteria, a 10% decrease in size or a 15% decrease in density on contrast-enhanced CT, correlated well in a small training set of patients who showed response as measured by positron emission tomography, and was more predictive of time to tumor progression (TTP) than response by RECIST. This study was designed to validate these observations in an independent data set.

Patients and Methods
Fifty-eight patients with imatinib-treated GISTs were evaluated by RECIST and Choi criteria. TTP was compared with TTP in the training set. Patients were analyzed initially with follow-up to 28 months, extended to 60 months for survival analysis.

Results
Patients who met Choi response criteria on CT at 2 months had significantly better TTP than those who did not ($P = .0002$), whereas response group by RECIST was not significantly correlated with TTP. Even when the 98 patients from both sets were analyzed together, the response group by RECIST did not correlate significantly with TTP, whereas response group by Choi criteria did correlate significantly with TTP. Disease-specific survival (DSS) was also significantly correlated with response group by Choi criteria ($P = .04$), but not with response group by RECIST.

Conclusion
Choi response criteria are reproducible, more sensitive, and more precise than RECIST in assessing the response of GISTs to imatinib mesylate. Response by Choi criteria, unlike response by RECIST, correlates significantly with TTP and DSS. Response by Choi criteria should be incorporated routinely into future studies of GIST therapy. We should desist using RECIST, at least in GIST.

J Clin Oncol 25:1760-1764. © 2007 by American Society of Clinical Oncology

INTRODUCTION

Recent studies have demonstrated that Response Evaluation Criteria in Solid Tumors (RECIST) are insensitive in evaluating gastrointestinal stromal tumors (GISTs) treated with imatinib.[1-3] We have analyzed a group of 40 patients who underwent baseline and 2-month follow-up evaluation by computed tomography (CT) and positron emission tomography (PET). We demonstrated that response by Choi criteria, a 10% decrease in unidimensional tumor size or a 15% decrease in tumor density on contrast-enhanced CT, correlated well with good response by PET and was more predictive of time to tumor progression (TTP) than response by RECIST.[4,5] Our reviewers appropriately pointed out that our findings should be confirmed in an independent data set before broad conclusions are made. Therefore, the purposes of this study were to validate the correlation with TTP in an independent data set and, if validated, to perform an analysis on the combined group of patients, update the follow-up, and perform a survival analysis.

PATIENTS AND METHODS

Of the 109 patients treated for recurrent or metastatic GIST at our institution from December 2000 to September 2001, 40 patients were included in our training set because they had additional PET imaging. In 11 patients we did not have baseline and 2-month follow-up CTs that could be evaluated for response by Choi criteria, leaving a new group of 58 patients who were evaluated by contrast-enhanced CT, referred to as the test set. All patients had pretreatment and 2-month follow-up CTs. All patients were observed up to 28 months. Follow-up was not updated for this analysis to ensure accurate comparison with our initial patient group. Subsequently, follow-up was updated up to 60 months for survival analysis. CT images

were evaluated for response by a single medical oncologist (R.S.B.) using RECIST and Choi criteria. Whenever response classification was not obvious, images were evaluated by a single radiologist (H.C.), who made the final assessment. Images were reviewed on a clinical patient information system (ClinicStation; University of Texas M.D. Anderson Cancer Center, Houston, TX) that uses Stentor image analysis software (Stentor Inc, Brisbane, CA). TTP was determined by Kaplan-Meier analysis and compared within each response group with the training set by log-rank testing.

RESULTS

Response rates by RECIST and Choi criteria are listed in Table 1 for the initial training set of 40 patients, the test set of 58 patients, and the entire group of 98 patients. When the test set was evaluated by RECIST, there were 28 (48%) responders and 30 (52%) nonresponders. When evaluated by Choi criteria, there were 49 (84%) responders and nine (16%) nonresponders. Similar findings were seen in the training set and the entire group.

TTP was analyzed in all patients in the training set and the test set, and the comparisons were broken down further by Choi response category (Fig 1). There was no difference in the TTP for all patients or either subset, demonstrating the reproducibility of the technique. We had demonstrated previously that patients from our training set who met Choi response criteria on CT at 2 months after treatment had significantly better TTP than those who did not, unlike patients with response by RECIST, whose TTP was not significantly improved.[5] Similarly, patients from the test set who met Choi response criteria on CT at 2 months after treatment had significantly better TTP than those who did not have significantly better TTP ($P = .0002$), whereas response by RECIST was not significantly correlated with TTP (Fig 2). Even when the entire 98 patients were analyzed together, RECIST did not correlate significantly with TTP, whereas Choi criteria did correlate significantly with TTP (Fig 3). Despite an almost doubling of the response rate from 46% to 83% using Choi criteria rather than RECIST, the clinical benefit of response is not diminished. As shown in Figure 4, TTP for responders was identical whether defined by Choi criteria or RECIST. Rather, if patients not qualifying for partial response by RECIST are considered nonresponders, as is traditionally the case, then RECIST was not sensitive in detecting patients truly not benefiting from therapy.

We subsequently updated follow-up on all patients to analyze survival. First, we confirmed that with increased follow-up time, patients with good response by Choi criteria on CT at 2 months after treatment had significantly better TTP than those who did not ($P = .01$), whereas response by RECIST was not significantly correlated with TTP ($P = .74$). We then performed a survival analysis, but censored patients dying without evidence of disease progression at the time of last follow-up, given that a good response to treatment could not be expected to prevent death from drug toxicity or causes unrelated to GIST. Patients with good response by Choi criteria had significantly improved disease-specific survival ($P = .04$) in contrast to those with complete or partial response at any time by RECIST ($P = .45$; Fig 5).

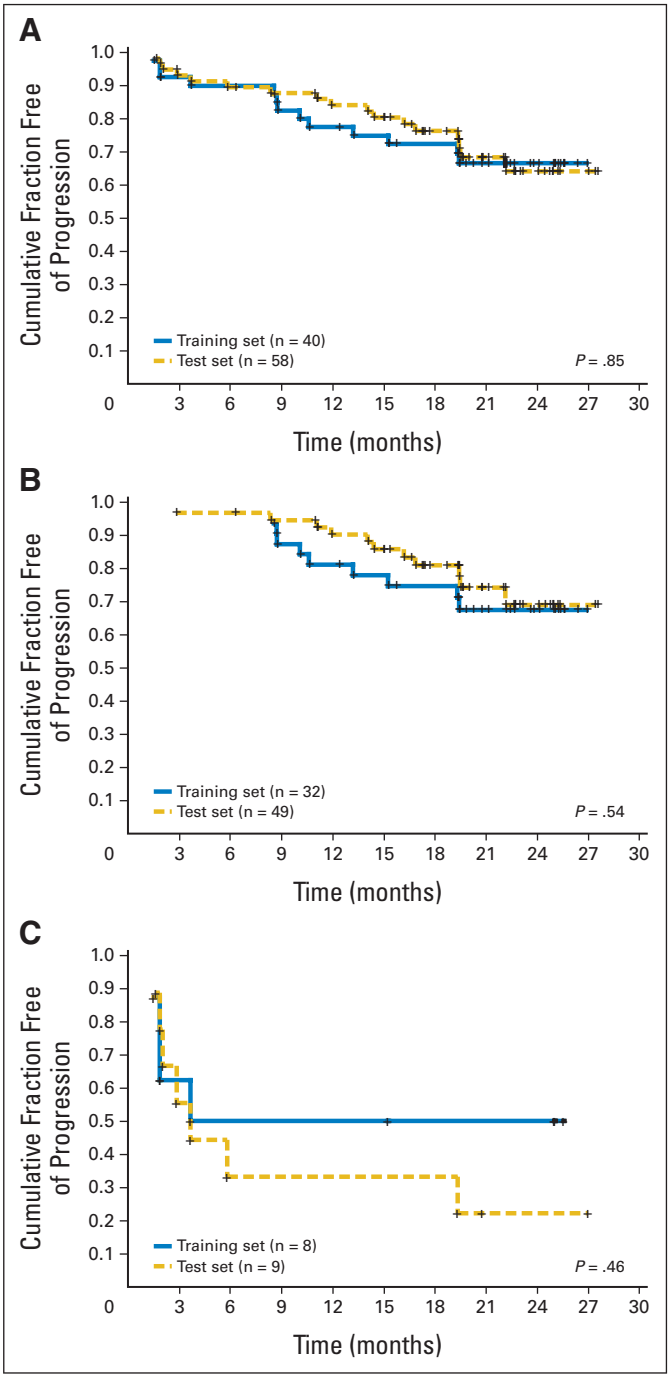

Fig 1. Time to tumor progression in the training set (40 patients) and the test set (58 patients). (A) All patients; (B) patients with good response; (C) patients with poor response.

Table 1. Response Rates of GIST to Imatinib by Choi Criteria and RECIST by Patient Group

	Training Set (n = 40)		Test Set (n = 58)		All Patients (N = 98)	
Response	Choi	RECIST	Choi	RECIST	Choi	RECIST
Responders	32	17	49	28	81	45
Nonresponders	8	23	9	30	17	53
Response rate, %	80	43	84	48	83	46

Abbreviations: GIST, gastrointestinal stromal tumor; RECIST, Response Evaluation Criteria in Solid Tumors.

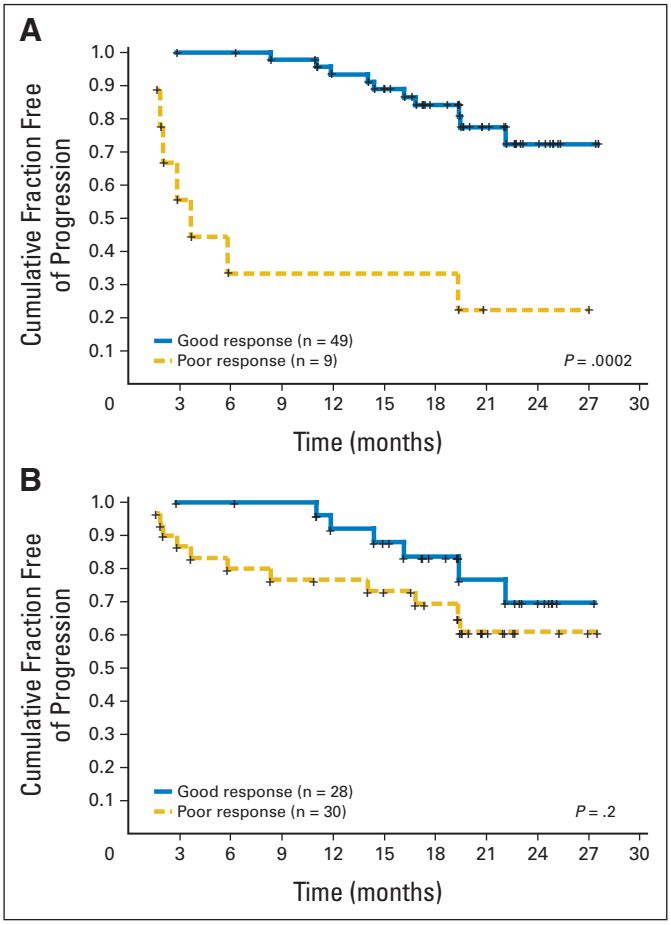

Fig 2. Time to tumor progression in good and poor responders in the test set. (A) Response by Choi criteria; (B) response by Response Evaluation Criteria in Solid Tumors. When the tumor response was evaluated on the basis of Choi response criteria, a significant difference was observed in the long-term prognosis between the good and poor responders ($P = .0002$) for up to 28 months.

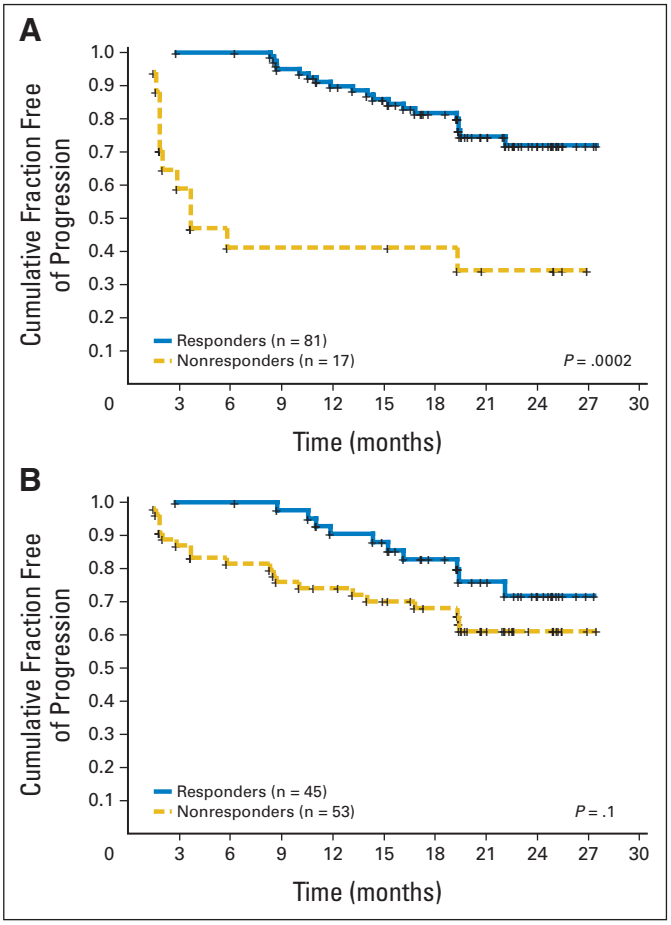

Fig 3. Time to tumor progression in good and poor responders in the entire group of 98 patients by response criteria. (A) Response by Choi criteria; (B) response by Response Evaluation Criteria in Solid Tumors (RECIST). When the tumor response was evaluated on the basis of Choi response criteria, a significant difference was observed in the long-term prognosis between the good and poor responders ($P = .0002$) for up to 28 months, but even with the increased number of patients, no significant difference was observed between good and poor responders by RECIST ($P = .1$).

DISCUSSION

Choi response criteria, incorporating tumor density and using small changes in tumor size on CT, are more sensitive and more precise than RECIST in assessing the response of GISTs to imatinib mesylate. The response rate of GIST to imatinib by Choi criteria in our entire 98-patient group (83%) is the same as the response rate by PET in our training set (83%), is almost double the response rate by RECIST (46%), and correlates more significantly with TTP and survival. Thus these new, now validated, criteria should be used, rather than RECIST, in future studies of patients with GIST.

The tumor response to treatment traditionally has been evaluated solely on the basis of tumor size, whereas Choi criteria employ both size and another quantitative parameter, tumor density. The tumor size criteria of RECIST, although carefully conceived to make a partial response mean the same thing whether determined from unidimensional measurements or bidimensional measurements, are totally arbitrary. The requirement of a 50% shrinkage of a bidimensionally measured mass was based on the limitations of physical examination, the main technique available to assess solid tumors in the 1960s, when these criteria were initially conceived,[6] and the 1970s, when they were verified quantitatively.[7] Now that we can measure the size of lesions with a precision of tenths of millimeters with a computer on cross-sectional images, such as CT or MRI, the restrictions of current response criteria should be re-examined.

RECIST has been validated to correlate with older criteria (WHO) using bidimensional measurements in thousands of patients with solid tumors,[8] but it has not been validated in responding patients with GIST; even if GIST patients were included in the 28 sarcoma patients evaluated in that series, it is unlikely that any would have responded before the advent of imatinib. It is ironic that imatinib received US Food and Drug Administration approval based on its response rate (by modified WHO criteria, on which RECIST was based).

Furthermore, the prognostic value of a response to therapy by RECIST has not been evaluated carefully in any situation. A partial response to therapy should be a surrogate marker for clinical benefit.[6,8] Such benefit is best measured by duration of tumor control (ie, TTP or survival), at least where only the therapeutic intervention under study has any impact. Furthermore, response should be assessable early in the majority of patients to influence decision making about continuation or discontinuation of therapy.

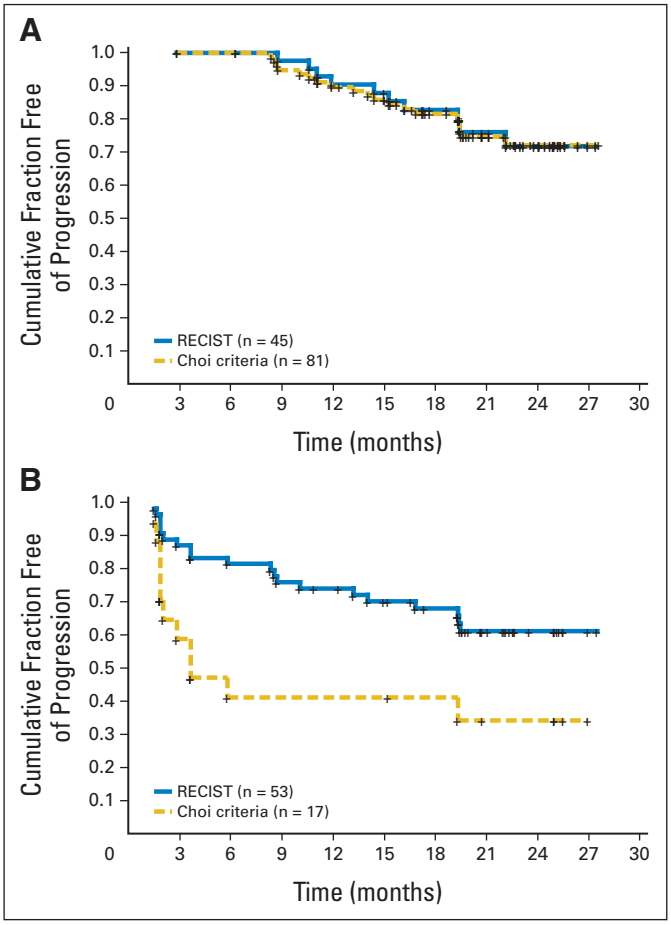

Fig 4. Time to tumor progression in the entire group of 98 patients by response evaluated by Choi criteria or Response Evaluation Criteria in Solid Tumors (RECIST). (A) Good response; (B) poor response. Whether the tumor response was evaluated on the basis of Choi response criteria or RECIST, there was no difference observed in the long-term prognosis between the good responders, despite the increased frequency of good response by Choi criteria. In contrast, poor responders by Choi criteria had significantly shorter time to tumor progression ($P = .0002$) than those with poor response by RECIST.

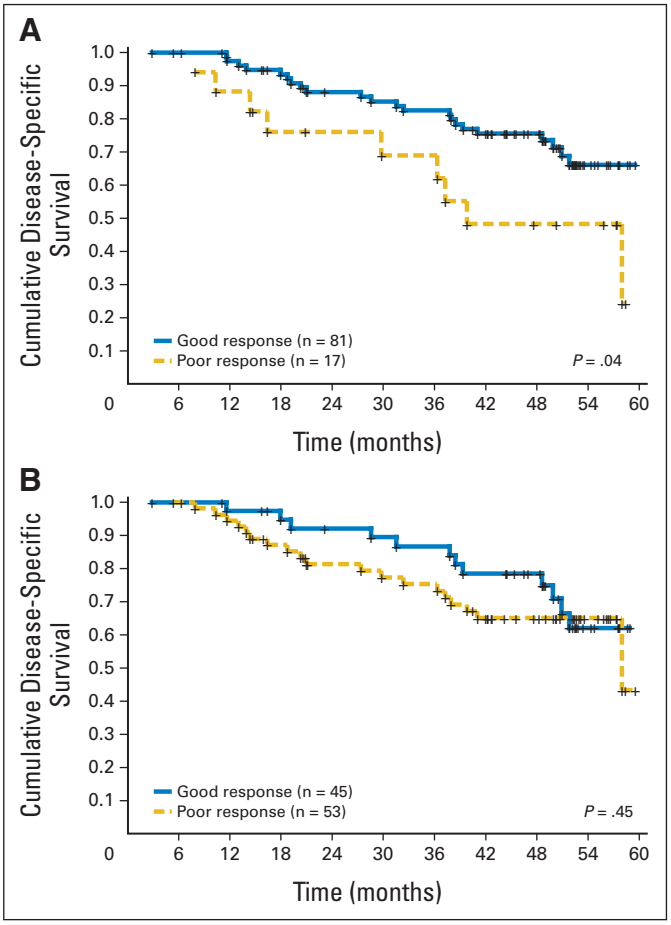

Fig 5. Disease-specific survival in good and poor responders in the entire group of 98 patients by response criteria. (A) Response by Choi criteria; (B) response by Response Evaluation Criteria in Solid Tumors (RECIST). When the tumor response was evaluated on the basis of Choi response criteria, a significant difference was observed in disease-specific survival between the good and poor responders ($P = .04$) with follow-up to 60 months, but no significant difference was observed between good and poor responders by RECIST ($P = .45$).

We are not the first to point out the deficiencies of RECIST. Ratain and Eckhardt,[9] proponents of the randomized discontinuation trial design, have pointed out that "a drug may be active without consistent achievement of high-level tumor regression." Michaelis and Ratain[10] also noted that "end points are needed that can be measured earlier than survival and that can more reliably predict phase III outcome." The Choi criteria address both of these concerns.

The RECIST article represents very careful consideration on the part of its authors.[8] It states that cystic lesions are not measurable. One of the underlying assumptions in all systems of tumor measurement is that tumor size is directly proportional to the number of tumor cells. In responding patients with GIST (and often other sarcomas; osteosarcoma is a notable example), the change in tumor size clearly is not proportional to the number of cells. There can be increasing areas of apparent cyst formation in previously solid tumors. How should one measure the size of these now-unmeasurable lesions? The Choi criteria, which evaluate tumor density quantitatively, address this deficiency of using tumor size as the only parameter of response evaluation. This approach is totally in keeping with the suggestions of the RECIST article, which states, "the guidelines proposed in this document are not meant to discourage the development of new tools that may provide more reliable surrogate end points than objective tumor response for predicting a potential therapeutic benefit for cancer patients."[8] Choi criteria may also apply to other sarcomas, especially at metastatic sites other than the lungs, and to cytotoxic therapy as well as targeted therapy, as observed by one of the authors (R.S.B., unpublished comment). Similar findings in patients with myxoid liposarcoma treated with trabectedin were presented at the 2006 Annual Meeting of the American Society of Clinical Oncology by Grosso et al.[11]

Choi criteria might apply to other tumor types as well. A notable example is renal carcinoma. Treatment with sorafenib produced notable clinical benefit without a meaningful response rate by RECIST, causing the investigators to find alternative means to measure the benefit. In the randomized discontinuation study by Ratain et al, [12] the response rate by traditional criteria was only 4%. The authors, however, considered a 25% decrease in bidimensional measurements (equivalent to a 13.4% decrease in unidimensional measurements) as

sufficient evidence of benefit to continue treatment of all patients who met that criterion, and randomly assigned only patients with less than 25% decrease to receive sorafenib or placebo. Even in the patients randomly assigned to treatment, TTP was significantly longer in patients randomly assigned to sorafenib. These data suggest that a substantial number of patients considered nonresponders by RECIST actually benefited from therapy.

Similarly, the phase III study of sorafenib reported by Escudier had a RECIST response rate of only 2%.[13] We were struck that the images shown of responding patients were similar to those of responding patients with GIST. The tumors showed decreased contrast enhancement and small decreases in size. Similar changes were noted with sunitinib in renal carcinoma[14,15] and in other tumors.[15] Although we cannot assess tumor density quantitatively from published or presented data, we can assess tumor shrinkage from the presented waterfall plots. Analyzing tumor response in the sorafenib phase III study[13] by Choi size criteria, we calculated that the response rate in patients treated with sorafenib was 44% compared with 7% for those treated with placebo. That difference, which might have been even greater had changes in tumor density been considered, is highly statistically significant ($P < .0001$).

In conclusion, CT assessment is a sensitive and specific method to assess the response of metastatic GISTs to imatinib if evaluated by Choi criteria: a decrease in tumor size of $\geq 10\%$ or a decrease in tumor density of $\geq 15\%$. We have demonstrated that these CT criteria have reproducible correlation with TTP, as well as disease-specific survival. RECIST substantially underestimated, especially at the early stage of treatment, the effect of imatinib on metastatic GIST, and was a poor predictor of clinical benefit. We should desist using RECIST, at least in GIST. Choi criteria should be investigated in other solid tumors.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following authors or their immediate family members indicated a financial interest. No conflict exists for drugs or devices used in a study if they are not being evaluated as part of the investigation. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment: N/A **Leadership:** N/A **Consultant:** Robert S. Benjamin, Novartis; Haesun Choi, Novartis; Homer A. Macapinlac, Siemens, GE Healthcare, Radiology Corporation of America; Donald A. Podoloff, Biogen Idec Inc, Siemens, GE Healthcare, Bexxar **Stock:** N/A **Honoraria:** Robert S. Benjamin, Novartis; Haesun Choi, Novartis; Homer A. Macapinlac, Siemens, GE Healthcare, Radiology Corporation of America; Shreyaskumar R. Patel, Novartis, Amgen; Donald A. Podoloff, Bexxar, Biogen Idec Inc, GE Healthcare **Research Funds:** N/A **Testimony:** N/A **Other:** Robert S. Benjamin, Novartis

AUTHOR CONTRIBUTIONS

Conception and design: Robert S. Benjamin, Haesun Choi, Chuslip Charnsangavej
Financial support: Robert S. Benjamin
Administrative support: Robert S. Benjamin, Donald A. Podoloff
Provision of study materials or patients: Robert S. Benjamin, Michael A. Burgess, Shreyaskumar R. Patel, Lei L. Chen
Collection and assembly of data: Robert S. Benjamin, Haesun Choi
Data analysis and interpretation: Robert S. Benjamin, Haesun Choi
Manuscript writing: Robert S. Benjamin, Haesun Choi
Final approval of manuscript: Robert S. Benjamin, Haesun Choi, Homer A. Macapinlac, Michael A. Burgess, Shreyaskumar R. Patel, Lei L. Chen, Donald A. Podoloff, Chuslip Charnsangavej

REFERENCES

1. Stroobants S, Goeminne J, Seegers M, et al: 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). Eur J Cancer 39:2012-2020, 2003

2. Van den Abbeele A, Badawi R, Cliche J, et al: 18F-FDG-PET predicts response to imatinib mesylate (Gleevec) in patients with advanced gastrointestinal stromal tumors (GIST). Proc Am Soc Clin Oncol 21:403, 2002 (abstr 1610)

3. Antoch G, Kanja J, Bauer S, et al: Comparison of PET, CT, and dual-modality PET/CT imaging for monitoring of imatinib (STI571) therapy in patients with gastrointestinal stromal tumors. J Nucl Med 45:357-365, 2004

4. Choi H, Macapinlac H, Burgess M, et al: Correlation of computerized tomography (CT) and positron emission tomography (PET) in patients with metastatic GIST treated at a single institution with imatinib mesylate. Proc Am Soc Clin Oncol 22:819, 2003 (abstr 1359)

5. Choi H, Charnsangavej C, Faria S, 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 25:10.1200/JCO.2006.07.3049

6. Karnofsky DA: Meaningful clinical classification of therapeutic responses to anticancer drugs. Clin Pharmacol Ther 2:709-712, 1961

7. Moertel CG, Hanley JA: The effect of measuring error on the results of therapeutic trials in advanced cancer. Cancer 38:388-394, 1976

8. Therasse P, Arbuck S, Eisenhauer E, 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 92:205-216, 2000

9. Ratain MJ, Eckhardt SG: Phase II studies of modern drugs directed against new targets: If you are fazed, too, then resist RECIST. J Clin Oncol 22:4442-4445, 2004

10. Michaelis LC, Ratain MJ: Measuring response in a post-RECIST world: From black and white to shades of grey. Nat Rev Cancer 6:409-414, 2006

11. Grosso F, Demetri GD, Blay JY, et al: Patterns of tumor response to trabectedin (ET743) in myxoid liposarcomas. J Clin Oncol 24:18s, 2006 (suppl; abstr 9511)

12. Ratain MJ, Eisen T, Stadler WM, et al: Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma. J Clin Oncol 24:2505-2512, 2006

13. Escudier B, Szczylik C, Eisen T, et al: Randomized phase III trial of the Raf kinase and VEGFR inhibitor sorafenib (BAY 43-9006) in patients with advanced renal cell carcinoma (RCC). J Clin Oncol 23:1093s, 2005 (suppl; abstr 4510)

14. Motzer RJ, Michaelson D, Redman BG, et al: Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor, in patients with metastatic renal cell carcinoma. J Clin Oncol 24:16-24, 2006

15. Faivre S, Delbaldo C, Vera K, et al: Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer. J Clin Oncol 24:25-35, 2006