JOURNAL OF CLINICAL ONCOLOGY

Assessment of Imaging Modalities and Response Metrics in Ewing Sarcoma: Correlation With Survival

Vadim S. Koshkin, Vanessa Bolejack, Lawrence H. Schwartz, Richard L. Wahl, Rashmi Chugh, Denise K. Reinke, Binsheng Zhao, Joo H. O, Shreyaskumar R. Patel, Scott M. Schuetze, and Laurence H. Baker

Vadim S. Koshkin, Cleveland Clinic, Cleveland, OH; Vanessa Bolejack, Cancer Research and Biostatistics, Seattle, WA; Lawrence H. Schwartz, Binsheng Zhao, Columbia University Medical Center, New York, NY; Richard L. Wahl, Mallinckrodt Institute of Radiology, Washington University in St Louis, St Louis, MO; Rashmi Chugh, Scott M. Schuetze, Laurence H. Baker, University of Michigan Medical School; Denise K. Reinke, Sarcoma Alliance for Research Through Collaboration, Ann Arbor, MI; Joo H. O, The Catholic University of Korea; Seoul St Mary's Hospital, Seoul, Republic of Korea; and Shreyaskumar R. Patel, The University of Texas MD Anderson Cancer Center, Houston, TX.

ABSTRACT

Purpose

Despite the rapidly increasing use of [^{18}F]fluorodeoxyglucose (FDG) –positron emission tomography (PET), the comparison of anatomic and functional imaging in the assessment of clinical outcomes has been lacking. In addition, there has not been a rigorous evaluation of how common radiologic criteria or the location of the radiology reader (local *v* central) compare in the ability to predict benefit. In this study, we aimed to compare the effectiveness of various radiologic response assessments for the prediction of overall survival (OS) within the same data set of patients with sarcoma.

Methods

We analyzed assessments made during a clinical trial of a novel IGF1R antibody in Ewing sarcoma: PET Response Criteria in Solid Tumors (PERCIST) for functional imaging and WHO criteria (performed locally and centrally), RECIST, and volumetric analysis for anatomic imaging. We compared the effectiveness of the various criteria for the prediction of progression and survival.

Results

For volume analysis, progression—defined as cumulative lesion volume increase of 100% at 6 weeks—was the optimal cutoff for decreased OS ($P < .001$). Assessment of the day-9 FDG-PET scan was associated with reduced OS in progressors compared with nonprogressors ($P = .001$) and with improved OS in responders compared with nonresponders. Significant variations in response (18% to 44%) and progression (9% to 50%) were observed between the different criteria. The comparison of central and local interpretation of anatomic imaging produced similar outcomes. PET was superior to anatomic imaging in identification of a response. Volume analysis identified the most responders among the anatomic imaging criteria.

Conclusion

An early signal with FDG-PET on day 9 and volume analysis were the best predictors of benefit. Validation of the volumetric analysis is required.

J Clin Oncol 34:3680-3685. © 2016 by American Society of Clinical Oncology

INTRODUCTION

The definitions of therapeutic efficacy and progression (which indicates treatment failure) shape critical decisions in the care of oncology patients and for clinical trial end points. Early clinical trials relied on the subjective self-assessment of symptoms by the patient, on the more objective assessment by the investigator made on the basis of radiographic shrinkage, and on physical findings to assess clinical benefit.[1] Bidimensional criteria introduced by the Southwest Oncology Group (SWOG)[2] were followed by unidimensional measurements that led to the Response Evaluation Criteria in Solid Tumors (RECIST)[3] and the updated RECIST 1.1.[4]

Progression varies in terms of equivalent increase in volume among the different criteria: 40% according to WHO, 84% according to SWOG, and 73% according to RECIST. In contrast, the criteria for response have estimated the equivalent volume decrease consistently at approximately 65% to 66%.[5] A new methodology for estimation of change in volume throughout treatment could generate more accurate and objective criteria for the definition of progression and response.

The emergence of functional imaging exemplified by [^{18}F]fluorodeoxyglucose (FDG) – positron emission tomography (PET) has added

more tools to the clinician's armamentarium. Despite the increasing use of FDG-PET in clinical practice, the comparison of anatomic and functional imaging in the assessment of clinical outcomes has been lacking. FDG-PET criteria to assess progression and response use changes in tumor standardized uptake value (SUV), including the European Organization for the Research and Treatment of Cancer (EORTC) criteria and the PET Response Criteria in Solid Tumors (PERCIST 1.0).[6,7] FDG uptake measured after chemotherapy in pediatric patients with Ewing sarcoma[8-10] was predictive of progression-free survival.[11] However, no study has compared anatomic and functional imaging criteria in their abilities to predict progression and response within the same data set of patients with sarcoma. Concerns about imaging performed in the context of multicenter clinical trials are the variability in reader experience and the potential bias of interpretation by knowledge of the clinical course. Many clinical trial designs consequently rely on external, study-specific, expert radiologists to interpret imaging centrally. It is unclear whether this central-reader approach is superior.

We performed an analysis of the SARC (Sarcoma Alliance for Research Through Collaboration) 011 trial to compare various criteria to assess progression and response for both anatomic and functional imaging. We compared anatomic and functional imaging criteria head to head in the assessment of progression and response. We introduce new volume-based criteria for the assessment of anatomic imaging. Finally, we compared the criteria to assess anatomic imaging, such as WHO and RECIST, and included the comparison of local interpretation by sarcoma oncologists (WHO local) with central interpretation of imaging by an expert radiology group (WHO central).

METHODS

SARC011 was a single-arm, multicohort, multicenter, phase II study of patients with recurrent Ewing sarcoma treated with IGF1 receptor (IGF1R) antibody (R1507). A total of 115 patients were enrolled from 31 centers in Europe and North America. Response was evaluated with both FDG-PET and anatomic imaging by computed tomography (CT) or magnetic resonance imaging (MRI). Anatomic imaging was assessed at baseline and at 6 weeks after the start of treatment. It was assessed by the treating
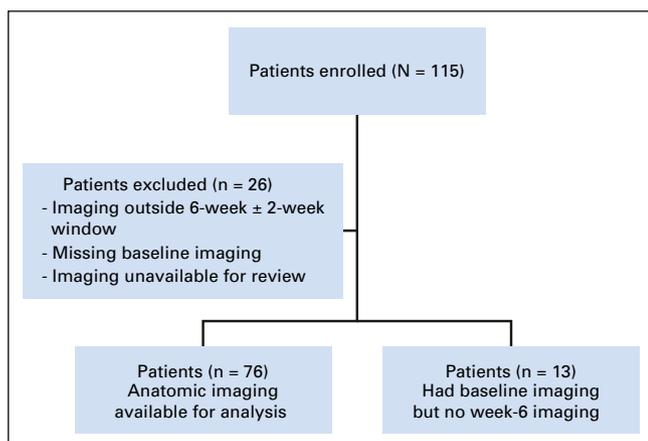
oncologist according to WHO criteria. Imaging was also assessed centrally by an external group of radiologists blinded to the clinical courses of the individual patients. FDG-PET was done at baseline and on day 9 of treatment via central review by experts who used PERCIST 1.0.

The imaging criteria comparison included the following: (1) PERCIST 1.0 criteria for functional imaging (FDG-PET) and for anatomic imaging; (2) WHO criteria on the basis of independent assessment; (3) WHO criteria on the basis of local site measurements; (4) RECIST criteria on the basis of independent assessment; and (5) volumetric criteria (newly defined) on the basis of measurements done by the central radiology group. The central radiology group included radiologists from Columbia University led by Schwartz. These radiologists were blinded to the PET results and clinical outcomes. Follow-up anatomic imaging had to be within 2 weeks of the 6-week mark after the start of treatment to be included in the volumetric analysis.

Anatomic Imaging

Of the 115 patients with Ewing sarcoma enrolled in this trial, 89 had anatomic imaging available for central review. Twenty-six patients were excluded because of inability to complete imaging as a result of sickness or as a result of death (Fig 1). The radiology group measured lesions from all available anatomic imaging scans at baseline and at 6 weeks by using the semiautomated solid tumor segmentation software at Columbia University. The volume of a lesion was calculated after segmentation by multiplying the number of lesion voxels by the voxel volume. Subsequently, the longest line and its longest perpendicular line inside the segmented lesion were automatically determined for all axial images that contained the lesion. The central RECIST measure—that is, the maximal diameter of a lesion—was calculated by multiplying the number of voxels of the longest line by the voxel length; the central WHO measure—that is, the product of the two diameters—was calculated by multiplying the number of voxels of the longest line and its longest perpendicular line by the voxel area. Figure 2 provides an example of how this was done for an irregularly shaped lesion.

Volumetric criteria were selected by establishing the optimal cutoffs for the percent change in volume between baseline and week 6 that were most predictive of overall survival. Volume increase of 100% versus baseline was defined as the optimal cutoff for progression ($P < .001$);

Fig 1. Patients included in anatomic imaging analysis.

Patients enrolled (N = 115)

Patients excluded (n = 26)
- Imaging outside 6-week ± 2-week window
- Missing baseline imaging
- Imaging unavailable for review

Patients (n = 76)
Anatomic imaging available for analysis

Patients (n = 13)
Had baseline imaging but no week-6 imaging

Fig 2. Example of volume calculation and of unidimensional (uni) and bidimensional (bi) measurements for an irregularly shaped lesion.

Case Name	Date	Site	Uni (mm)	Bi (mm²)	Volume (cm³)
5102	9/25/2008	Mediastinum	41.6	1,077.7	20.7
5102	12/18/2008	Mediastinum	39.2	1,078.0	57.0
Percentage change			−5.8%	0.0%	175.4%

Table 1. PERCIST Response Versus RECIST

PERCIST Status	RECIST Status			Total No. of Patients
	Response	Stable Disease	Progressive Disease	
Response	10	11	8	29
Stable disease	4	10	17	31
Progressive disease	0	0	6	6
Total No. of patients	14	21	31	66

Abbreviations: PERCIST, positron emission tomography Response Criteria in Solid Tumors; RECIST, Response Criteria in Solid Tumors.

volume decrease of 45% was most predictive of survival ($P = .4$). On the basis of these criteria, response to treatment was separated into categories of progressive disease, stable disease, and response. A total of 76 patients were included in the volume analysis and had data compared with central evaluation by using WHO and RECIST criteria. Finally, the same 76 patients were separated into categories by progressive disease, stable disease, and response according to WHO criteria as assessed by local sites at baseline and week 6.

A total of 92 patients had interpretable FDG-PET scans that were done at baseline and on day 9 of treatment. All scans were reviewed centrally under the supervision of R.L. Wahl and assessed according to PERCIST 1.0.[6] On the basis of the FDG-PET changes between baseline and day 9, the patients were separated into the same categories (progressive disease, stable disease, response). At the time that this analysis was done, standard criteria for acquisition or quantitative interpretation of PET scans in patients with sarcoma were not yet established.

RESULTS

The comparison of the five different imaging criteria from both anatomic and functional imaging produced interesting results from this rich data set. Functional imaging assessment of progressive disease can be identified as early as day 9 versus at 6 weeks by using any of the anatomic imaging criteria. There was no significance in median survival between patients who responded to

treatment and patients with stable disease for any of the imaging criteria. However, for all of the criteria, there was a trend toward longer survival for patients in the response group compared with the stable disease group. There was variation among the imaging criteria of patients called responders (21% to 35%) and an even greater variation in patients labeled progressors (12% to 50%). PERCIST identified the most patients in the response group: 32 of 92 patients, or 34.8% of the total patients analyzed. Anatomic imaging criteria (volume, WHO local, WHO central, RECIST) identified fewer patients in the response group (average, 21.7% of patients among all four criteria). The contrast between anatomic and functional imaging is even more striking when only the subgroup of 66 patients with interpretable functional imaging, who were also among the 76 patients with interpretable anatomic imaging, was considered. In this subgroup of patients, 43.9% (29 of 66 patients) were responders according to PERCIST, and 90.9% (60 of 66 patients) were nonprogressors. Table 1 lists a comparison of the PERCIST and RECIST response categories among the patients who had both evaluable anatomic and evaluable functional imaging. It shows that PERCIST, in addition to having an advantage of being performed earlier in the course of treatment, also identified more patients with clinical response than did RECIST. Use of PERCIST would lead to fewer patients discontinuing the therapy.

Volumetric criteria were selected by establishing the optimal cutoffs for the percentage of change in volume between baseline and week 6 that were most predictive of overall survival. We did not find a linear correlation between survival and volume. Volume criteria identified more patients in the response group (19 of 76 patients, or 25%) than did WHO central (22.4%), WHO local (18.4%), or RECIST (21.1%; Table 2; Fig 3). Volume criteria also identified more patients with clinical benefit from therapy (response or stable disease): 64.5%. In comparison, 51% were identified by WHO central; 50%, by WHO local, and 55.0%, by RECIST. Functional imaging identified fewer patients in the progressive-disease group than did anatomic imaging criteria. PERCIST classified 12% of patients as having progressive disease. Volumetric analysis identified 36% of patients as having progressive disease; RECIST identified 45%, WHO central identified 49%, and WHO local identified 50%.

Table 2. Median OS for Progressors Versus Nonprogressors in the Study on the Basis of the Imaging Criteria Used

Subgroup	No. of Patients	Nonprogressors				Progressive Disease	
		Response		Stable Disease			
		No. of Patients	Median OS (months)	No. of Patients	Median OS (months)	No. of Patients	Median OS (months)
PERCIST (all patients with PET imaging)	92	32	13.4	49	6.8	11	4.7
Patients with PET and anatomic imaging	66	29	13.0	31	11.4	6	5.5
Volume	76	19	13.9	30	12.9	27	6.6
WHO (central read)	76	14	17.0	25	12.8	37	7.6
WHO	76	17	13.9	21	13.5	38	8.0
RECIST	76	16	17.0	26	12.6	34	7.6

NOTE. OS of patients with progression at 6 weeks (or day 9 for FDG-PET) was significantly reduced compared with nonprogressors on the basis of all assessed criteria ($P < .005$ for all).
Abbreviations: FDG, [18F]fluorodeoxyglucose; OS, overall survival; PET, positron emission tomography; PERCIST, PET Response Criteria in Solid Tumors; RECIST, Response Criteria in Solid Tumors.
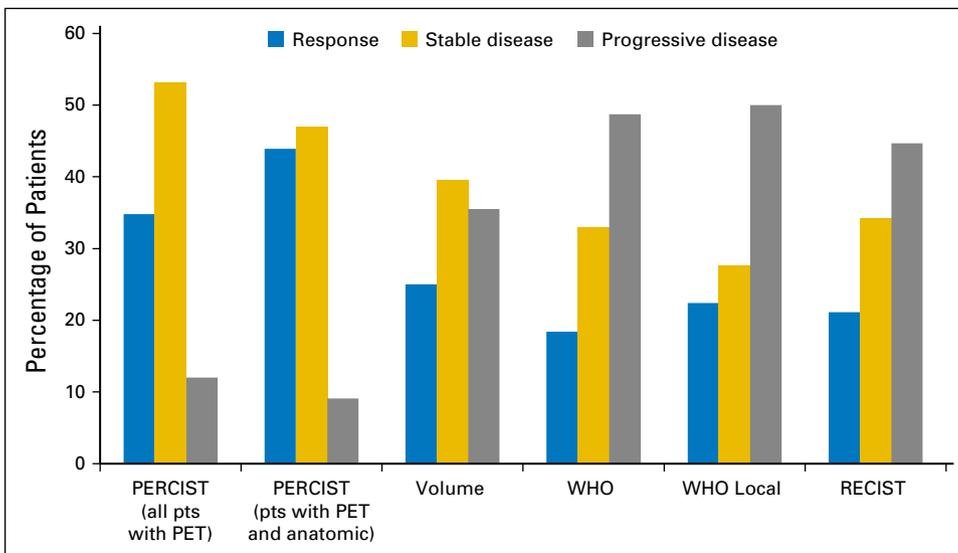
Fig 3. Percentage of patients (pts) in each response category for all criteria. PERCIST, positron emission tomography Response Criteria in Solid Tumors; RECIST, Response Criteria in Solid Tumors.

The comparison of central and local interpretation of anatomic imaging produced similar outcomes among the two criteria (Table 2; Fig 3). Tumor response interpretation by central review was performed exclusively by radiologists, and interpretation by local research teams was performed predominately by treating medical or pediatric oncologists. The differences between central and local interpretation were quite similar (50% v 51%). The comparison of central interpretation of WHO and RECIST in survival is shown in Fig 3.

Patients enrolled in this trial who had baseline anatomic imaging but who did not have week-6 imaging (n = 13) had significantly diminished overall survival compared with the patients who had both baseline imaging and week-6 imaging (P < .001). Patients with baseline imaging but no week-6 imaging had a median overall survival of only 1.1 months (Fig 4). Some trial reports exclude such patients, because the primary objective cannot be measured.

DISCUSSION

We showed that FDG-PET assessed by PERCIST as early as day 9 predicted clinical benefit of IGF1R antibody in Ewing sarcoma. PET compared favorably with anatomic imaging assessed at 6 weeks. FDG-PET was shown to be superior to any of the assessed anatomic imaging criteria in identification of response. PERCIST identified at least 35% of patients who had a response, including several long-term responders. A response rate of 35% in patients with metastatic drug-refractory Ewing sarcoma is impressive and could have been worthy of regulatory approval. PERCIST identified the fewest patients as progressors (day 9), which gave these patients who had experienced progression the opportunity to seek treatment alternatives.

The newly defined volumetrics were shown to be superior to WHO and RECIST in prediction of response, which suggests that volumetric analysis may be a superior method of assessment of clinical response compared with the widely used unidimensional RECIST. All criteria correlated with the outcome. Because of the frequently irregular shape of tumors, assessment of these lesions in only one or two dimensions implies a sacrifice of both accuracy and precision of assessment. Volumetric assessment is now available at many institutions through the availability of tumor segmentation software on Picture Archiving and Communications Systems (ie, PACS) and commercial advanced 3D workstations. The time needed to assess tumors volumetrically is slightly greater than to do so unidimensionally or bidimensionally, but the process is now automated. The analysis presented here suggests that assessment of tumor volume is superior to predict response in clinical trials compared with the currently widely used RECIST and WHO criteria. This requires additional validation with prospective clinical trials.

There were only slight differences between the WHO criteria assessed locally and WHO criteria assessed by a central group of radiologists blinded to patient clinical status or outcome. Concurrence of sarcoma control rates between treating investigative sites and independent central interpretation by radiologists blinded to treatment assignment was previously reported in two randomized trials of trabectedin treatment of advanced liposarcoma and leiomyosarcoma.[8,10] The added benefit of central interpretation is unclear.[12] These data suggest that local experts, at least in sarcoma, accurately interpret anatomic imaging in context of tumor response assessment. Centers that participated in this trial in the United States likely had radiologists more familiar with imaging of patients with sarcoma. The standard practice at European centers is unknown.

Thirteen enrolled patients had no week-6 imaging and had markedly reduced survival. In some clinical trials, such patients have their results excluded. Such exclusion adversely biases the overall trial results. Our results support intention-to-treat analysis as the standard.

Our investigation had a number of shortcomings. It was a retrospective analysis of data in a clinical trial. Only 89 of the 115 patients with Ewing sarcoma in the trial had anatomic imaging available to be reviewed centrally, and only 76 of those patients had imaging available at baseline and at 6 weeks. For the functional imaging analysis, 92 of 115 patients had scans available for review.
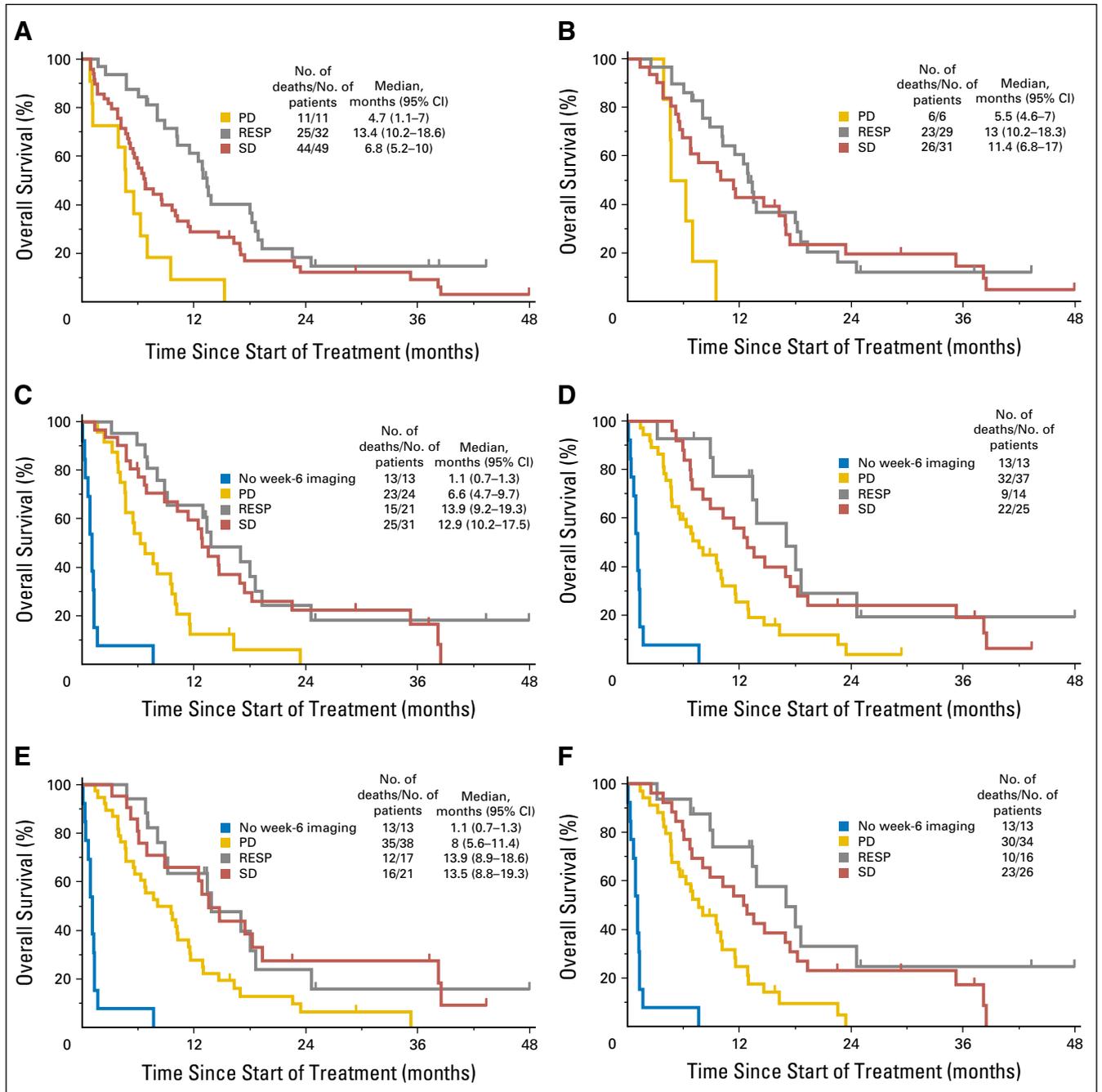
Fig 4. Kaplan-Meier curves of overall survival for each assessment criteria: (A) positron emission tomography Response Criteria in Solid Tumors (PERCIST; all patients), (B) PERCIST (patients with anatomic imaging), (C) total volume, (D) WHO central, (E) WHO local, (F) RECIST. PD, progressive disease; RESP, response; SD, stable disease.

The FDG-PET scans done at different institutions worldwide were not standardized to the same common criteria. During trial accrual (2007-2010), a common standard was not yet established. The investigators involved in this phase II trial were all experts in the management of sarcoma and were practicing at centers of sarcoma excellence; a strong correlation between local and central interpretation of imaging may not be as likely in other cancers or in more broadly conducted clinical trials.

We performed a retrospective analysis of imaging data and survival in a large, phase II clinical trial to evaluate imaging criteria

most suitable for the assessment of progression and response in Ewing sarcoma. Our analysis is the first, to our knowledge, to compare functional imaging (FDG-PET) assessed according to PERCIST with four anatomic imaging criteria, including a newly defined volume criteria. Functional imaging was shown to predict clinical benefit of IGF1R antibody in Ewing sarcoma as early as day 9 of treatment and was superior to anatomic imaging assessed at 6 weeks in identification of response. Newly defined volume criteria were superior to WHO or RECIST in the prediction of response and clinical benefit.

AUTHOR CONTRIBUTIONS

Conception and design: Lawrence H. Schwartz, Shreyaskumar R. Patel, Laurence H. Baker

Administrative support: Laurence H. Baker

Collection and assembly of data: Vadim S. Koshkin, Vanessa Bolejack, Lawrence H. Schwartz, Richard L. Wahl, Rashmi Chugh, Denise K. Reinke, Binsheng Zhao, Joo H. O, Shreyaskumar R. Patel

Provision of study materials or patients: Scott M. Schuetze, Laurence H. Baker

Data analysis and interpretation: Vadim S. Koshkin, Vanessa Bolejack, Lawrence H. Schwartz, Richard L. Wahl, Rashmi Chugh, Joo H. O, Shreyaskumar R. Patel, Scott M. Schuetze

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

REFERENCES

1. Karnofsky DA, Abelman WH, Craver LF, et al: The use of nitrogen mustards in the palliative treatment of carcinoma. Cancer 1:634-656, 1948

2. Green S, Weiss GR: Southwest Oncology Group standard response criteria, end point definitions, and toxicity criteria. Invest New Drugs 10:239-253, 1992

3. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst 92:205-216, 2000

4. Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45:228-247, 2009

5. Oxnard GR, Morris MJ, Hodi FS, et al: When progressive disease does not mean treatment failure: Reconsidering the criteria for progression. J Natl Cancer Inst 104:1534-1541, 2012

6. Wahl RL, Jacene H, Kasamon Y, et al: From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 50: 122S-150S, 2009

7. O JH: Lodge MA, Wahl RL. Practical PERCIST: A simplified guide to PET Response Criteria in Solid Tumors 1.0. Radiology 280:576-584, 2016

8. Demetri GD, von Mehren M, Jones RL, et al: Efficacy and safety of trabectedin or dacarbazine for metastatic liposarcoma or leiomyosarcoma after failure of conventional chemotherapy: Results of a phase III randomized multicenter clinical trial. JCO 34:786-793, 2016

9. Pappo AS, Patel SR, Crowley J, et al: R1507, a monoclonal antibody to the insulin-like growth factor 1 receptor, in patients with recurrent or refractory Ewing sarcoma family of tumors: Results of a phase II Sarcoma Alliance for Research through Collaboration study. J Clin Oncol 29:4541-4547, 2011

10. Demetri GD, Chawla SP, von Mehren M, et al: Efficacy and safety of trabectedin in patients with advanced or metastatic liposarcoma or leiomyosarcoma after failure of prior anthracyclines and ifosfamide: Results of a randomized phase II study of two different schedules. J Clin Oncol 27:4188-4196, 2009

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

12. Amit O, Bushnell W, Dodd L, et al: Blinded independent central review of the progression-free survival end point. Oncologist 15:492-495, 2010

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Assessment of Imaging Modalities and Response Metrics in Ewing Sarcoma: Correlation With Survival

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or jco.ascopubs.org/site/ifc.

Vadim S. Koshkin
No relationship to disclose

Vanessa Bolejack
No relationship to disclose

Lawrence H. Schwartz
Honoraria: Merck KGaA
Consulting or Advisory Role: Novartis, GlaxoSmithKline, Merck, Sharp & Dohme, Celgene, Bioclinica, Icon
Research Funding: Eli Lilly (Inst), Astellas Pharma (Inst), Merck Sharp & Dohme (Inst), Pfizer (Inst), Boehringer Ingelheim (Inst)
Patents, Royalties, Other Intellectual Property: Varian Medical Systems

Richard L. Wahl
Consulting or Advisory Role: Nihon Medi-Physics
Research Funding: Akrivis

Rashmi Chugh
Stock or Other Ownership: Portola Pharmaceuticals
Research Funding: MabVax, Morphotek, Novartis, Lilly, Biomarin, AADi, Sarcoma Alliance for Research Through Collaboration (Inst)
Travel, Accommodations, Expenses: AADi

Denise K. Reinke
Research Funding: Bayer (Inst), Threshold Pharmaceuticals (Inst), Tesaro (Inst)

Binsheng Zhao
No relationship to disclose

Joo H. O
No relationship to disclose

Shreyaskumar R. Patel
Consulting or Advisory Role: Eli Lilly, Johnson & Johnson, CytRx, EMD Serono, Bayer, Eisai, Novartis
Research Funding: Johnson & Jonson (Inst), Morphotek (Inst), Eisai (Inst), PharmaMar (Inst)

Scott M. Schuetze
Honoraria: EMD Serono, Janssen
Consulting or Advisory Role: EMD Serono, Janssen
Research Funding: AB Science (Inst), Janssen (Inst), Threshold Pharmaceuticals (Inst), Amgen (Inst), ZIOPHARM Oncology (Inst), BioMed Valley Discoveries (Inst), CytRx (Inst), Plexxikon (Inst), Lilly (Inst), Sarcoma Alliance for Research Through Collaboration (Inst)

Laurence H. Baker
Consulting or Advisory Role: Teva Pharmaceutical Industries

Acknowledgment