Original Investigation

Soft Tissue Sarcoma Response to Two Cycles of Neoadjuvant Chemotherapy: A Multireader Analysis of MRI Findings and Agreement with RECIST Criteria and Change in SUVmax

Jennifer L. Favinger, MD, Daniel S. Hippe, MS, Darin J. Davidson, MD, Saeed Elojeimy, MD, Eira S. Roth, MD, Antoinette W. Lindberg, MD, Alice S. Ha, MD

Rationale and Objectives: When soft tissue sarcomas are treated with neoadjuvant chemotherapy, the number of cycles of chemotherapy is usually dependent on the tumor's initial response. Popular methods to assess tumor response include Response Evaluation Criteria in Solid Tumors (RECIST) criteria, which rely solely on tumor size, and maximum standardized uptake value (SUVmax) reduction in positron emission tomography (PET), which requires an expensive and high radiation test. We hypothesized that contrast-enhanced magnetic resonance imaging (MRI) may offer a good alternative by providing additional information beyond tumor size.

Materials and Methods: Following IRB approval, a retrospective review identified patients with soft tissue sarcomas who underwent both PET and MRI before and after two cycles of neoadjuvant chemotherapy. Five readers independently examined the MRI exams for: changes in size, T2 or T1 signal, necrosis and degree of enhancement. Readers then made a subjective binary assessment of tumor response to therapy. Each reader repeated the anonymized randomized reading at least 2 weeks apart. 18 F-FDG PET exams were interpreted by a nuclear medicine specialist. The maximum standardized uptake values (SUVmax) for pre and post-chemotherapy exams were compared. Intra- and inter-reader agreement was assessed using Cohen's kappa and Light's kappa, respectively. .

Results: Twenty cases were selected for this multireader study, of which 9 (45%) were responders and 11 were nonresponders by SUVmax. Using all MRI criteria, 43% were classified as responders based on MRI and 1.5% were classified as responders by RECIST criteria. Using PET as the reference, the sensitivity and the specificity of the MRI diagnosis for response using all findings were 50% and 63%, respectively. There was fair to moderate intrareader (kappa = 0.37) and inter-reader (kappa = 0.48) agreement for the MRI diagnosis of response. None of the individual MRI signal characteristics were significantly different between the PET responders and nonresponders. Additionally, no MRI findings were significantly different between those with and without good clinical responses.

Conclusion: By our assessment, there is a poor correlation between tumor response by RECIST criteria and PET SUVmax. In addition, varying MR features did not help in diagnosing tumor response. Imaging of tumor response remains a challenging area that requires further research.

Key Words: Sarcoma; Neoadjuvant chemotherapy; PET-CT; MRI.

Acad Radiol 2017; ■:■■–■■

From the Department of Radiology, University of Washington, Box 354755, 4245 Roosevelt Way NE, Seattle, WA 98105 (J.L.F., D.S.H., A.S.H.); Department of Orthopedics, University of Washington, Seattle, Washington (D.J.D.); Department of Radiology, University of New Mexico, Albuquerque, New Mexico (S.E.); Department of Radiology, Irving Radiological Associates, Irving, Texas (E.S.R.); Department of Orthopedics, Seattle Children's Hospital, Seattle, Washington (A.W.L.). Received June 9, 2017; revised October 4, 2017; accepted October 9, 2017. All work was performed at the University of Washington. **Address correspondence to:** A.S.H. e-mail: aha1@uw.edu

https://doi.org/10.1016/j.acra.2017.10.013

INTRODUCTION

Soft tissue sarcomas are rare connective tissue tumors with high mortality that are challenging to treat (1,2). Surgery and radiation therapy are the hallmarks of local disease control, but neoadjuvant and adjuvant chemotherapies have been shown to improve outcomes in select circumstances, such as large, high-grade, and recurrent tumors (3–5). Different chemotherapy regimens are used (6); at our institution, the typical protocol is to start with two cycles of AIM (doxorubicin, ifosfamide, and mesna). After the initial

two cycles, the decision to use more chemotherapy is made based on the tumor's initial response, both clinically and on imaging, in addition to the patient's tolerance of the treatment (7).

The traditional use of Response Evaluation Criteria in Solid Tumors (RECIST), which uses change in tumor size to assess tumor response, has been shown to be unreliable in soft tissue sarcomas (8). Studies have shown that a significant reduction in maximum standardized uptake value (SUVmax) on [F-18]Fluorodeoxyglucose/positron emission tomography-computed tomography (^{18}F-FDG/PET-CT) correlates better with histopathologic response (8) and with improved outcomes (9,10). However, positron emission tomography-computed tomography (PET-CT) is an expensive, time-consuming, and high-radiation test. PET-CT, in conjunction with contrast-enhanced magnetic resonance imaging (MRI), has been shown to be useful when distinguishing viable tumor from post-treatment changes (11). Our providers typically obtain contrast-enhanced MRI at various time points during treatment to assess tumor response, but the criteria to assess response based on MRI have not been well established. Currently, patients with sarcoma at our institution undergo a mixture of PET-CT and contrast-enhanced MRIs after chemotherapy to assess for response, sometimes one or the other, and only occasionally both imaging modalities at the same time.

The objective of the present study was to evaluate the utility of contrast-enhanced MRI in assessing post-treatment response using a retrospective multireader study evaluating the agreement between MRI and PET-CT for determining a positive or a negative response of the soft tissue sarcoma to two cycles of neoadjuvant chemotherapy. We hypothesized that contrast-enhanced MRI may offer a good alternative to PET-CT by providing additional information about tumor response beyond just tumor size.

MATERIALS AND METHODS

Following IRB approval, a retrospective review identified patients with soft tissue sarcomas who underwent both positron emission tomography (PET) and MRI before and after two cycles of neoadjuvant chemotherapy at our institution between April 2008 and November 2013. All initial MRI and PET examinations must have been performed before the initiation of treatment, and all follow-up examinations were obtained within 1–3 weeks following two cycles of neoadjuvant chemotherapy. The MRI protocols varied slightly, but all included both T1- and T2-weighted fat-suppressed sequences, and all but one examination included T1-weighted fat-suppressed postcontrast imaging in multiple planes.

Five MRI readers independently examined the pair of before and after chemotherapy magnetic resonance examinations, blinded to PET and other clinical information. We specifically chose readers with variable radiology experience levels: two of the readers were fellowship-trained musculoskeletal radiologists, two were orthopedic oncologists, and one was a senior radiology resident. Readers were not blinded to the temporal ordering of the examinations. Each reader measured the tumor in three dimensions on both examinations and then chose the longest tumor dimension on both examinations. Each reader then specifically rated each pair of examinations for changes in size, T2 signal, T1 signal, central necrosis, and degree of enhancement. A five-point Likert-like scale was used, with levels definitely decreased (−2), probably decreased (−1), no change (0), probably increased (+1), and definitely increased (+2). Readers then made a subjective binary assessment of tumor response or nonresponse to therapy, as if they were making a real-time clinical decision about the patient. The readers were then asked to comment on which findings most influenced their binary selection. Each reader repeated the anonymized randomized reading at least 2 weeks later to allow assessment of intrareader agreement.

[F-18]Fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET) examinations were interpreted by a single nuclear medicine specialist not involved in the MRI reader study and blinded to the clinical data. The percentage change between the SUVmax values for pre- and postchemotherapy examinations were compared, with greater than 35% reduction defined as tumor response by PET (10). Following RECIST criteria, a 30% or greater reduction in the longest tumor dimension was considered a responder. Moreover, in regard to good clinical response, if the patient received additional cycles of neoadjuvant chemotherapy after the two initial cycles, this was considered a clinical responder.

Continuous and ordinal variables were summarized as mean ± standard deviation and binary variables as number (%). Intra- and inter-reader agreement for continuous measurements was assessed using the intraclass correlation coefficient. Intra- and inter-reader agreement for binary or ordinal variable was assessed using the Cohen kappa and the Light kappa, respectively. Kappa was unweighted for binary variables and linearly weighted for ordinal variables. To interpret kappa, ≤0.20 was considered as poor agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and >0.80 as excellent agreement (12).

The Mann-Whitney test and area under the receiver operating characteristic curve were used to evaluate how well individual MRI findings, averaged across readers and reading sessions, could differentiate between responders and nonresponders by ^{18}F-FDG PET. The nonparametric bootstrap and percentile method was used to calculate 95% confidence intervals (CIs) for estimates of area under the receiver operating characteristic curve, intraclass correlation coefficient, and kappa by resampling patients to preserve the dependence among multiple reads of the same patient.

All statistical calculations were conducted with the statistical computing language R (version 3.1.1; R Foundation for Statistical Computing, Vienna, Austria) (13). Throughout, two-sided tests were used, with statistical significance defined as $P < 0.05$. No adjustments were made for multiple comparisons.