Advances in Radiation Oncology (2019) **4**, 413-421

advances
in radiation oncology

www.advancesradonc.org

Scientific Article

MRI Radiomic Features Are Independently Associated With Overall Survival in Soft Tissue Sarcoma

Matthew B. Spraker MD, PhD [a,*], Landon S. Wootton PhD [b], Daniel S. Hippe MS [c], Kevin C. Ball MD [d], Jan C. Peeken MD [e,f,g], Meghan W. Macomber MD [b], Tobias R. Chapman MD [h], Michael N. Hoff PhD [c], Edward Y. Kim MD [b], Seth M. Pollack MD [i,j], Stephanie E. Combs MD [e], Matthew J. Nyflot PhD [b,c]

[a]Department of Radiation Oncology, Washington University in St. Louis, St. Louis, Missouri; Departments of [b]Radiation Oncology and [c]Radiology, University of Washington, Seattle, Washington; [d]Aurora St. Luke's Medical Center, Department of Diagnostic Radiology, Milwaukee, Wisconsin; [e]Department of Radiation Oncology, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany; [f]Institute of Innovative Radiation therapy, Department of Radiation Sciences, Helmholtz Zentrum München, Neuherberg, Germany; [g]Deutsches Konsortium für Translationale Krebsforschung, Munich, Germany; [h]Beth Israel Deaconess Medical Center, Department of Radiation Oncology, Harvard Medical School, Boston, Massachusetts; [i]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; and [j]Division of Medical Oncology, University of Washington, Seattle, Washington

Received 5 November 2018; accepted 12 February 2019

Abstract

Purpose: Soft tissue sarcomas (STS) represent a heterogeneous group of diseases, and selection of individualized treatments remains a challenge. The goal of this study was to determine whether radiomic features extracted from magnetic resonance (MR) images are independently associated with overall survival (OS) in STS.

Methods and Materials: This study analyzed 2 independent cohorts of adult patients with stage II-III STS treated at center 1 (N = 165) and center 2 (N = 61). Thirty radiomic features were extracted from pretreatment T1-weighted contrast-enhanced MR images. Prognostic models for OS were derived on the center 1 cohort and validated on the center 2 cohort. Clinical-only (C), radiomics-only (R), and clinical and radiomics (C+R) penalized Cox models were constructed. Model performance was assessed using Harrell's concordance index.

Results: In the R model, tumor volume (hazard ratio [HR], 1.5) and 4 texture features (HR, 1.1-1.5) were selected. In the C+R model, both age (HR, 1.4) and grade (HR, 1.7) were selected along with 5 radiomic features. The adjusted c-indices of the 3 models ranged from 0.68 (C) to 0.74 (C+R) in the derivation cohort and 0.68 (R) to 0.78 (C+R) in the validation cohort. The radiomic

Sources of support: Dr. Nyflot reports funding from the R&E Foundation of the Radiological Society of North America during the study.

Conflicts of interest: Daniel Hippe reports research grants from GE Healthcare, Philips Healthcare, Toshiba American Medical Systems, and Siemens Medical Solutions USA outside the submitted work. The other authors declare no potential conflicts of interest.

* Corresponding author. 4921 Parkview Place, CAM LL, CB 8224, St. Louis, MO 63110.
E-mail address: mspraker@wustl.edu (M.B. Spraker).

https://doi.org/10.1016/j.adro.2019.02.003

features were independently associated with OS in the validation cohort after accounting for age and grade (HR, 2.4; $P = .009$).

Conclusions: This study found that radiomic features extracted from MR images are independently associated with OS when accounting for age and tumor grade. The overall predictive performance of 3-year OS using a model based on clinical and radiomic features was replicated in an independent cohort. Optimal models using clinical and radiomic features could improve personalized selection of therapy in patients with STS.

Introduction

Soft tissue sarcomas (STS) exhibit a wide range of behavior, from indolent tumors with long patient survival to highly aggressive disease that can be rapidly fatal. Surgical resection is the cornerstone of treatment for localized disease, and the addition of radiation therapy (RT) and chemotherapy have allowed for limb-sparing surgery and improved local and distant control.[1,2] Clinical and pathologic factors, such as tumor size and grade, help guide selection of optimal adjuvant therapies, but this remains a critical challenge for individual patients.

Prior studies of STS reported promise for quantitative imaging biomarkers. These have mainly analyzed positron emission tomography/computed tomography (PET/CT) and found that quantitative features[3-5] and structural features (eg, tumor boundary heterogeneity)[4,6] were associated with patient outcomes. More recently, radiomics has been used to study quantitative imaging biomarkers in STS. *Radiomics* refers to the process of extracting multiple quantitative imaging biomarkers (ie, features) to convert medical images into minable data, which are then used for outcome modeling and clinical decision support.[7] The quantitative imaging biomarkers can capture complex image characteristics, such as image texture, which refers to the structural relationship of voxels within the image. Initial radiomics studies of STS have found that radiomics features may be related to the risk of developing distant metastases[8] and overall survival (OS)[6] in STS.

Although most prior work has studied PET/CT, magnetic resonance imaging (MRI) is the standard initial imaging option for soft tissue tumors, and thus a radiomic model based on MRI would be broadly applicable for sarcoma patients. A previous study evaluated the association among radiomic features extracted from MRI, PET/CT, and fused MRI–PET/CT images to predict lung metastases.[8] They found that MRI textures alone are generally not useful compared with fused PET/CT-MRI images, which offered the best prognostic ability. However, that study is limited by a small cohort and lack of validation data set, thus the utility of MRI-based radiomic features is not thoroughly investigated. The purpose of this study is to evaluate the hypothesis that quantitative imaging features extracted from pretherapy T1-weighted MR images would be predictive of overall survival (OS) in patients with STS and to validate the model on an external international data set.

Methods and Materials

Data collection

The institutional review board of each institution approved this study, which retrospectively analyzed pretreatment contrast-enhanced T1-weighted MRIs from 2 independent cohorts of patients with biopsy-proven STS. Patients were treated with curative intent with surgery, radiation therapy, or chemotherapy based on multidisciplinary review. Patients younger than 18 years were excluded. Patients with Kaposi or primary bone sarcomas were also excluded because of the unique clinical contexts or treatment pathways associated with these diseases. A total of 199 patients were treated at center 1 from 2000 to 2017 and 78 patients were treated at center 2 from 2007 to 2015. This study focused on American Joint Committee on Cancer (AJCC) version 7 stage II-III patients only, which encompasses nonmetastatic patients with large (ie, >5 cm) or higher grade (ie, >1) tumors. AJCC stage II-III patients were specifically studied because multimodality therapy is often used in these patients and selection of ideal therapy in individual patients in this group is challenging. Therefore patients with stage I ($n = 27$ in center 1, $n = 2$ in center 2) and IV ($n = 7$ in center 1, $n = 5$ in center 2) STS were excluded (Fig. 1). Patients with image artifacts as a result of multiple MRI acquisitions within the tumor area were also excluded ($n = 9$ in center 2). The final cohorts were 165 and 61 patients, respectively.

Patient and tumor characteristics were collected from the electronic medical record. Maximum tumor extent was extracted from radiology reports. Pathologic characteristics, including French Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) grade and margin status, were extracted from reports issued by an STS pathologist at each institution. Histologic type was reclassified into one of 11 groups based on histologic

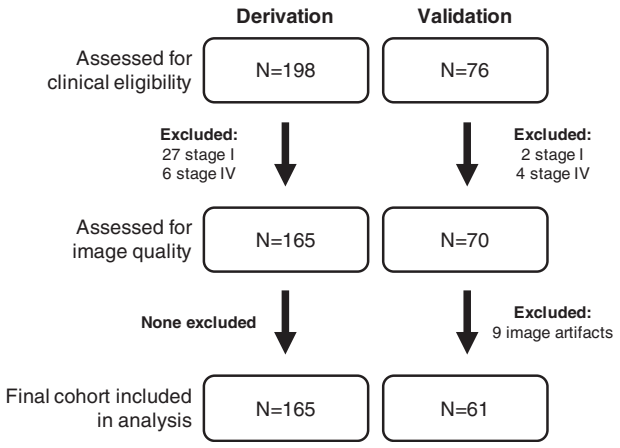

Fig. 1. CONSORT diagram for the present study.

differentiation and patterns of failure, as described in prior work by Taylor and colleagues.[9]

MRI data for the center 1 cohort were acquired from the institutional picture archiving and communication system. MRIs consisted of contrast-enhanced, T1-weighted, 2-dimensional (2D) or 3-dimensional (3D) acquisitions completed using 0.7T to 3T General Electric Healthcare, Siemens Healthineers, Philips Healthcare, and Hitachi Medical Systems MRI scanners. Image matrix sizes ranged from 128 to 704, 123 to 343, and 18 to 160 and resolutions ranged from 0.31 to 1.88, 0.62 to 2.40, and 2 to 11 mm along frequency, phase, and slice directions, respectively. Five image acquisitions were completed using gradient echo techniques with flip angles ranging from 10° to 30° and repetition time (TR) and echo time (TE) ranging from 3.98 to 7.70 and 1.96 to 3.36 ms, respectively. The remaining scans employed spin echo or inversion recovery–based sequences with flip angles ranging from 80° to 180° and TR and TE ranging from 400 to 5500 and 7.5 to 51 ms, respectively. Full DICOM (Digital Imaging and Communications in Medicine) information was not available for the center 2 cohort because some DICOM tags were removed during anonymization. However, parameters were generally similar to the derivation cohort. MR images were acquired using 1.5T to 3T General Electric Healthcare, Siemens Healthineers, and Philips Healthcare MRI scanners with TR and TE ranging from 102 to 1311 and 4.8 to 39 ms, respectively. Reconstructed image matrix sizes (acquisition sizes may be smaller) ranged from 176 to 704, 176 to 704, and 7 to 82 and resolutions ranged from 0.28 to 1.41, 0.28 to 1.41, and 3 to 7.27 mm along frequency, phase, and slice directions, respectively.

Tumor delineation and image processing

A radiation oncologist (M.S., M.M., T.C., or J.P.) or radiologist (K.B.) evaluated each image for quality and manually segmented the gross tumor, which was defined as all enhancing tumor on contrast-enhanced T1 MRI. This was completed using MIM software (version 6.6, MIM Software Inc, Cleveland, OH) for the center 1 cohort and iPlan RT (version 4.1.2, Brainlab, Munich, Germany) for the center 2 cohort. Images were discretized into 64 gray levels by taking the minimum and maximum value for each image and using equal bin widths before feature extraction. The images were processed in their native resolution. The PORTS software package (https://nciphub.org/groups/ports/overview, https://www.mathworks.com/matlabcentral/fileexchange/55587-ports-3d-image-texture-metric-calculation-package[10]) was used to extract 30 features derived from 5 categories: tumor volume, intensity histograms, gray tone spatial dependence matrices,[11] neighborhood gray tone difference matrices (NGTDM),[12] and gray level zone size matrices (GLZSM)[13] (Table E1; available online at https://doi.org/10.1016/j.adro.2019.02.003). All features were calculated in 3 dimensions.

Exploratory and univariate analyses

Initial model building was performed using the center 1 (derivation) cohort. OS was defined as the time from initial pathologic diagnosis to death. Cox proportional hazard models were used for exploring univariate associations between OS and each clinical and imaging feature. Before inclusion in the Cox model, right skewed variables were transformed (log or cube-root) to reduce skewness. Hazard ratios (HRs) for continuous variables were scaled by the standard deviation so they could be interpreted as changes per 1 standard deviation increase in the corresponding variable. Follow-up was truncated at 1096 days (3 years) for all data analyses to reduce violations of the proportion hazard assumption identified during the univariate analysis.[14] The univariate analysis results were not used for feature selection or derivation of the multivariate model.

Feature selection and model derivation

Collinearity among the radiomic texture features was assessed using the R^2 statistic. To reduce collinearity among the texture features before model training, a backward elimination procedure was applied. First, the feature with the largest R^2 value was eliminated and the R^2 values for the remaining features were recalculated. These steps were then repeated until all remaining features had a corresponding R^2 <50%.

After feature selection, 3 primary models for predicting OS were trained: a clinical-only model (model C), a radiomics-only model (model R), and a combined clinical and radiomics model (model C+R). C contained only age and grade as prespecified predictors. R contained tumor volume and the collinearity-filtered texture features. C + R contained all predictors included in C and R.

Penalized Cox regression fit using the LASSO (least absolute shrinkage and selection operator) algorithm was used for all 3 models.[15] The LASSO-penalty parameter was selected to minimize the partial likelihood deviance computed using leave-one-out cross-validation. The application of the LASSO penalty resulted in a second round of feature selection based on the association with OS rather than collinearity. Several alternative feature selection and multivariate modeling approaches were also considered during model development (Table E1; available online at https://doi.org/10.1016/j.adro.2019.02.003), but none provided a clear performance benefit compared with the primary approach.

Internal and external validation of model performance

For each model, the selected features were combined into a prognostic index (PI). Each PI was computed as a linear combination of the selected features, weighted by their corresponding regression coefficients, and centered so that median (PI) = 0 in the derivation data set.[16] PI values <0 indicate the individual has a lower risk than the median patient, PI values >0 indicate the individual has higher risk than the median patient, and exp(PI) is the overall hazard ratio for the individual versus the median patient.

Overall model performance for discriminating survival was assessed using Harrell's concordance index (c-index).[17] For internal testing and validation within the derivation cohort, the 0.632 bootstrap was used.[17] All of the feature selection and model training steps described in the previous section were included in the 0.632 bootstrap routine. External validation of the models was performed using the center 2 (validation) cohort with fixed weights for the PIs based on the derivation cohort.

Model performance was further summarized in both cohorts using Kaplan-Meier estimates of OS in different risk groups and receiver operating characteristic curves for death within 3 years, where the latter were calculated using an inverse probability weighted estimator to account for censoring.[18] Throughout, the nonparametric bootstrap and percentile methods were used to calculate confidence intervals (CIs) and to compare the c-indices between models.[19] Independent associations between components of the PI and OS were tested in the validation cohort using Cox models.[20] All statistical calculations were conducted with the statistical computing language R (version 3.1.1; R Foundation for Statistical Computing, Vienna, Austria). Two-tailed tests were used with statistical significance defined as $P < .05$.

Results

Patient characteristics for the derivation (center 1) and validation (center 2) cohorts are shown in Table 1. The 2 cohorts were similar except that the derivation cohort had more patients who received chemotherapy (54.5% vs 14.8%) and had a wider range of histologic types than the validation cohort. Almost all patients in the deviation and validation cohorts received external beam radiation therapy (96.3% vs 100%) to similar doses (Table 1). No patients were treated with brachytherapy. The median follow-up of both cohorts was similar (median, 3.1 years

Table 1 Patient characteristics

Variable	Site*		$P^{\dagger}$
	Center 1 (N = 165)	Center 2 (N = 61)	
Age	53 (19-88)	57 (21-88)	.20
Sex			.65
Male	96 (58.2)	33 (54.1)	
Female	69 (41.8)	28 (45.9)	
Location			.62
Extremity	122 (73.9)	43 (70.5)	
Other	43 (26.1)	18 (29.5)	
Pathology size			.55
<5 cm	28 (17.0)	8 (13.1)	
>5 cm	137 (83.0)	53 (86.9)	
Grade[‡]			.73
1	4 (2.4)	0 (0.0)	
2	65 (39.4)	25 (41.7)	
3	96 (58.2)	35 (58.3)	
Margins[‡]			.56
Negative	110 (70.5)	32 (76.2)	
Positive	46 (29.5)	10 (23.8)	
Histologic type			**<.001**
Pleomorphic sarcoma	80 (48.5)	25 (41.0)	
Adipocytic	21 (12.7)	17 (27.9)	
Uncertain differentiation	17 (10.3)	7 (11.5)	
Smooth muscle	17 (10.3)	5 (8.2)	
Fibro-/myofibroblastic	11 (6.7)	0 (0.0)	
Nerve sheath	9 (5.5)	0 (0.0)	
Skeletal muscle	5 (3.0)	0 (0.0)	
Chondro-osseous	3 (1.8)	0 (0.0)	
Vascular	2 (1.2)	0 (0.0)	
Fibrohistiocytic	0 (0.0)	7 (11.5)	
Chemotherapy			**<.001**
Yes	90 (54.5)	9 (14.8)	
No	75 (45.5)	52 (85.2)	
Radiation therapy dose (total, Gy)[§]	50 (18-74)	50 (28-70)	.32
Tumor volume (cm^3)	153 (1-3694)	147 (4-918)	.34

* Values are n (%) or median (range).

† Fisher's exact test (categorical variables) or the Wilcoxon rank-sum test (continuous variables).

‡ Grade was missing in 1 center 2 case; margins were missing in 9 center 1 cases and 20 center 2 cases.

§ In the derivation cohort, 9 patients in center 1 were not included in the dose calculations. Six of these patients did not receive radiation therapy, and the 3 other patients received radiation therapy, but detailed dose information was not available.

Table 2 Primary prognostic models for overall survival based on the derivation data set

	Hazard ratio*		
Clinical variables	Model C: Age + grade	Model R: Radiomics only	Model C+R: Age + grade + radiomics
Age	1.6		1.4
Grade 3 (vs grades 1 and 2)	2.1		1.7
Radiomics variables			
Tumor volume†		1.5	1.5
Histogram: Skewness		1.2	-
Histogram: Kurtosis‡		1.1	1.2
NGTDM: Complexity		-	-
GLZSM: Small zone size emphasis		-	-
GLZSM: Small zone/low gray emphasis†		1.2	1.2
GLZSM: Zone size nonuniformity†		1.5	1.3

Abbreviations: C = clinical only; C+R = clinical + radiomics; GLZSM = gray level zone size matrices; NGTDM = neighborhood gray tone difference matrices; R = radiomics only.

* Hazard ratio (HR) is per 1 SD increase for continuous variables; HR > 1 indicates higher risk of death; - indicates the variable was "deselected" by the LASSO.

† Variable was log transformed before entry into the model to reduce right skewness.

‡ Variable was cube rooted before entry into the model to reduce right skewness.

vs 3.0 years; $P = .97$), as were rates of OS (log-rank test $P = .80$), with a total of 56 deaths (33.9%) and 21 deaths (34.4%) reported in the derivation and validation cohorts, respectively.

The median follow-up of both cohorts was similar (median, 3.1 years vs 3.0 years; $P = .97$), as were rates of OS (log-rank test $P = .80$), with a total of 56 deaths (33.9%) and 21 deaths (34.4%) reported in the derivation and validation cohorts, respectively. When follow-up was truncated at 3 years, there were 37 deaths (22.4%) and 16 deaths (26.2%) reported in the derivation and validation cohorts, respectively. Overall there were 83 (50.3%) and 28 (45.9%) instances of tumor progression in the 2 cohorts (median progression-free survival, 3.2 vs 2.5 years; $P = .99$). Progression was locoregional in 28 of these patients (33.7%) and 5 patients (17.9%), respectively, and distant in 55 (66.3%) and 23 (82.1%) ($P = .15$).

Univariate analysis of OS in the derivation cohort is summarized in Table E2 (available online at https://doi.org/10.1016/j.adro.2019.02.003). Among clinical variables, only age (HR, 1.7; $P = .003$) and grade (HR, 2.4 for grade 3 vs grades 1-2; $P = .022$) were significantly associated with OS. Among radiomic variables, tumor volume (HR, 2.3; $P < .001$) and 25 texture features were significantly associated with OS. After applying the iterative collinearity filter, 6 texture features met the collinearity criterion of $R^2 < 50\%$ for model inclusion: skewness ($R^2 = 40.0\%$; univariate HR, 1.5; $P = .015$), kurtosis ($R^2 = 35.7\%$; HR, 1.3; $P = .083$), complexity ($R^2 = 29.4\%$; HR, 0.9; $P = .53$), small zone size emphasis ($R^2 = 45.4\%$; HR, 0.7; $P = .034$), small zone or low gray emphasis (SZLGE, $R^2 = 43.9\%$; HR, 1.1; $P = .58$), and zone size nonuniformity (ZSNU, $R^2 = 44.9\%$; HR, 1.9; $P < .001$) (Tables E2 and E3; available online at https://doi.org/10.1016/j.adro.2019.02.003). Texture features generally had low to moderate associations with tumor volume, with absolute Spearman's correlation coefficients ranging from 0.09-0.65.

The C, R, and C+R models are shown in Table 2. In the R model, the LASSO selected tumor volume (HR, 1.5) and 4 of 6 texture features (skewness, kurtosis, SZLGE, and ZSNU; HR, 1.1-1.5). In the C+R model, both age (HR, 1.4) and grade (HR, 1.7) were selected along with tumor volume (HR, 1.5) and 3 texture features (kurtosis, SZLGE, and ZSNU; HR, 1.2-1.5). Heat maps of the texture features selected for the C+R model for a patient who is high risk (patient 1) and low risk (patient 2) for death at 3 years are shown in Figure 2. During internal validation, bootstrap 0.632 c-indices were statistically significant for all models ($P < .001$) and ranged from 0.68 (C model) to 0.74 (C+R model) in the derivation cohort.

In the external validation cohort, the radiomics component of the C+R PI was independently associated with OS in the validation cohort after accounting for the clinical component (HR, 2.4; $P = .009$). In addition, the C model (c-index, 0.70; 95% CI, 0.57-0.82; $P < .001$) and R model (c-index, 0.6; 95% CI, 0.55-0.80; $P < .001$) had overall performances as measured by the c-index that were similar to the initial estimates from the derivation cohort, whereas the validation C+R model performed slightly better than the derivation model with a c-index of 0.78 (95% CI, 0.66-0.88; $P < .001$). C-indices in the validation cohort were similar whether follow-up was truncated at 3 years or full follow-up was used (Table E4; available online at https://doi.org/10.1016/j.adro.2019.02.003). When stratified using the median risk threshold (PI = 0), sensitivity was 82% and specificity was 62% for OS at 3 years in the derivation cohort. Sensitivity and

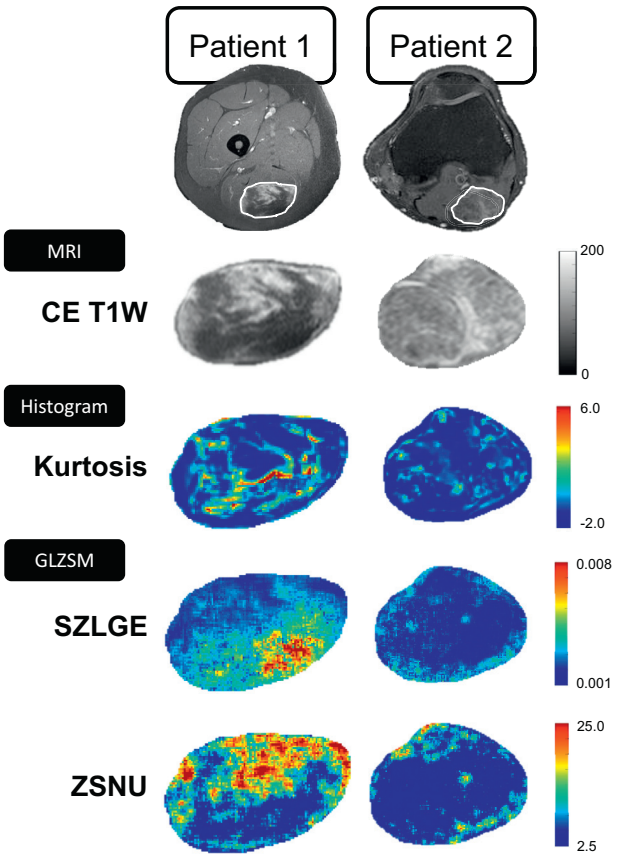

Fig. 2. Single-slice contrast-enhanced T1-weighted MRI and texture feature heat maps are shown for patients identified as high risk (patient 1) and low risk (patient 2) for death at 3 years by the clinical + radiomics model. Patient 1 (high risk) is a 45 year old with AJCC stage III pleomorphic sarcoma (FNCLCC grade 3) of the lower extremity treated with sequential preoperative chemotherapy and radiation therapy to 50 Gy followed by surgical resection (negative margins). This patient survived 18.5 months. Patient 2 (low-risk) is a 44 year old with AJCC stage III myxoid/round cell liposarcoma (FNCLCC grade 3) of the lower extremity treated with neoadjuvant chemotherapy followed by surgical resection (positive margins) followed by adjuvant radiation therapy to 66 Gy. This patient was still alive after 88.5 months of follow-up. *Abbreviations:* AJCC = American Joint Committee on Cancer; CE T1W = contrast-enhanced T1-weighted; FNCLCC = Fédération Nationale des Centres de Lutte Contre le Cancer; GLZSM = gray level zone size matrix; SZLGE = small zone/large gray emphasis; ZSNU = zone size nonuniformity.

specificity in the validation cohort were similar (79% [95% CI, 58%-100%] and 68% [95% CI, 50%-84%]), as were the receiver operating characteristic curves (Fig. 3). Kaplan-Meier curves summarizing OS in the higher and lower risk subgroups from all 3 models for both cohorts are shown in Figure 4.

To investigate the contribution of tumor volume versus texture features as a driver of performance in the radiomic model, models similar to the primary R and C+R models but using tumor volume only (R_A and C+R_A) or texture

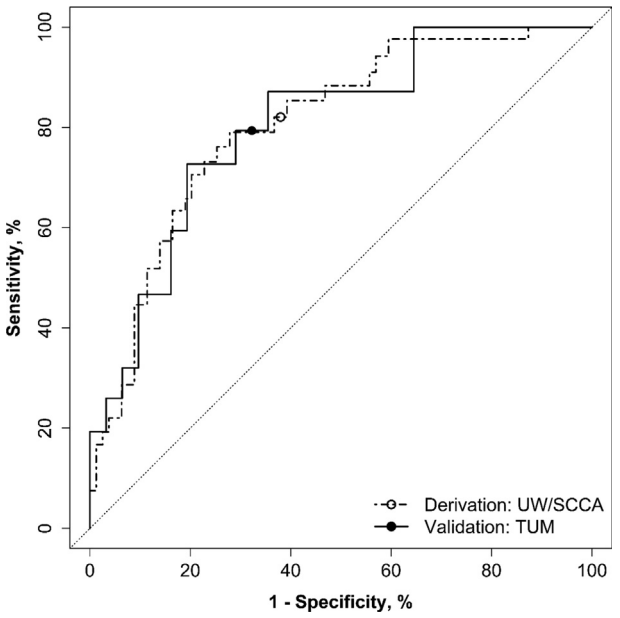

Fig. 3. Receiver operating characteristic curves of the final clinical + radiomics models in the derivation cohort (dashed curve) and the validation cohort (solid curve). The points indicate the performance of the median risk threshold based on the derivation cohort. *Abbreviations:* SCCA = Seattle Cancer Care Alliance; TUM = Technical University of Munich; UW = University of Washington.

features only (R_B and C+R_B) were constructed using the derivation cohort (Table E4; available online at https://doi.org/10.1016/j.adro.2019.02.003). In the validation cohort, tumor volume and texture features appeared to have similar overall impact on discrimination of OS, with R_A and R_B (c-index, 0.68 vs 0.65) and C+R_A and C+R_B (c-index, 0.75 vs 0.75) having similar c-indices in the validation cohort. From Cox analysis, both the tumor volume component of the C+R_A PI (HR, 2.2; $P = .033$) and the texture features component of the C+R_B PI (HR, 2.2; $P = .017$) significantly improved prediction of OS over the clinical component alone. However, the tumor volume component of the C+R PI (*HR*, 1.4; $P = .38$) did not significantly improve prediction of OS over the clinical and texture feature components alone, nor did the texture feature component of the C+R PI (HR, 1.5; $P = .11$) significantly improve prediction of OS over the clinical and volume components alone.

Discussion

This is the first externally validated study to evaluate whether radiomic features extracted from pretherapy T1 MRIs are predictive of OS in a cohort of more than 200 patients with STS. There were 2 important findings. First, radiomic features alone were together significantly predictive of OS, had similar prognostic performance as a model based on age and grade alone, and remained

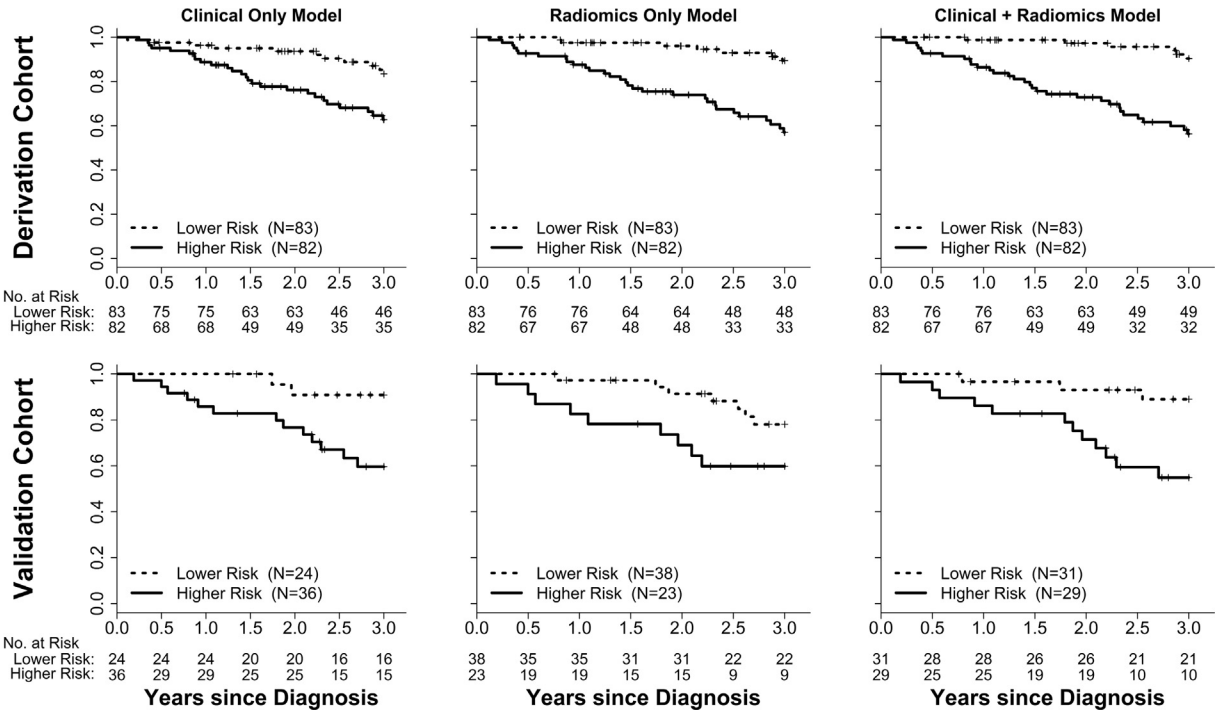

Fig. 4. Kaplan-Meier estimates of overall survival under each model in the derivation (top row) and validation cohorts (bottom row). For each model the cohort was risk stratified into lower and higher risk groups using the median risk value in the derivation cohort.

independently associated with OS after accounting for age and grade. Second, a combined model based on both clinical and radiomic features appeared to perform the best and was similarly predictive of OS in independent derivation and validation cohorts. Prognostic performance was similar to other radiomic studies in non-small cell lung cancer,[21-23] glioblastoma,[24] and hepatocellular carcinoma.[25]

Optimal selection of therapy for STS remains challenging, and attempts to individualize treatment based on clinical attributes alone have remained unsuccessful. For instance, a recent randomized trial by Gronchi et al[26] attempted to individualize chemotherapy regimens based on a clinical feature, tumor histotype, with the goal of sparing toxicity associated with standard, anthracycline-based chemotherapy. Unfortunately, counter to the goal of the trial, histotype-tailored chemotherapy was associated with worse progression-free survival compared with standard chemotherapy.[26] Our study found that radiomic features were significantly predictive of OS in STS independent of known clinical features (eg, age and grade) and appeared to improve prediction of OS when combined with clinical features compared with clinical features alone. Therefore MRI-based radiomic features represent a promising and readily available biomarker that may be useful in selecting therapy for patients with STS, such as identifying those most likely to benefit from chemotherapy or radiation therapy dose intensification.

MRI-based radiomic features are attractive in STS because MRI is the standard imaging modality to evaluate soft tissue tumors and they are acquired on almost all patients. Yet most prior work evaluating the relationship between imaging variables, tumor features, and patient outcomes in STS is in PET/CT, which may not be clinically available. For instance, increasing maximum standard uptake value within a tumor PET image is associated with higher tumor grade,[27] worse OS,[28,29] and worse progression-free survival[3] in patients with STS. More complex PET features related to spatial heterogeneity and texture[6] are associated with OS. One prior study completed a radiomic analysis of the association between radiomic features extracted from PET/CT and MRI and the development of lung metastases and achieved the greatest prediction by extracting texture features from fused FDG (fluorodeoxyglucose)–PET/MRI scans.[8] Although our study is not directly comparable because of the use of different imaging modalities, modeling endpoints, and the use of external versus internal validation, some similarities emerge. For instance, 3 of the 4 variables in the model using fused MRI and PET/CT scans were GLZSM (zone-size) features, which were similarly selected in our models. Similarly, CT-based GLZSM variables have been found to be significantly associated with OS in head and neck[30] and non-small cell lung cancers (NSCLCs).[21]

This study also found that tumor volume and texture features were each separately associated with OS after accounting for age and grade in the validation cohort. However, the combination of volume and texture features did not produce a significant improvement in prediction over models with only one or the other. It has been suggested that the high levels of correlation between tumor volume and texture features can limit interpretation

of radiomic studies.[31,32] However, this primarily has been a concern in PET, which is characterized by fewer voxels analyzed because of lower image resolution. The radiomic features modeled in this study had only low to moderate correlation with each other, suggesting that correlation between tumor volume and texture features is not a confounding factor for T1 MRI in sarcoma. It should be noted that although the texture features investigated here are commonly applied to medical imaging, including other empirical feature classes or representation learning methods (eg, convolutional neural networks) might improve model performance.

This study had several limitations related to its retrospective nature. First, this study retrospectively examined images from STS patients treated at institutions using different techniques. Specifically, there was a significantly higher proportion of patients treated with chemotherapy at center 1 than at center 2. However, multiple randomized trials have found that chemotherapy use has marginal efficacy with respect to OS in unselected patients in STS,[1] and therefore these differences are not expected to substantially affect the results of the present study. This is consistent with the finding of similar death rates in the 2 cohorts despite different chemotherapy use rates. Second, the 2 cohorts had different proportions of STS histotypes, and all histotypes were analyzed together because relatively small numbers of cases in most individual histotypes precluded a reliable subgroup analysis. An analysis of model performance in pleomorphic sarcomas, our largest subgroup, is currently underway. Third, the tumors were manually delineated by a team of practicing radiation oncologists or a radiologist. Although prior work has found substantial agreement between radiation oncologists delineating STS gross tumor volumes,[33] interobserver variability could affect radiomic features. Investigations evaluating the impact of interobserver variability in tumor volume delineation will be an important direction for future study. Finally, regarding imaging protocols, MRIs were collected on different scanners using different protocols and in different years. Prior work in PET and CT has found that radiomics feature values can vary significantly as a result of image noise and acquisition and reconstruction parameters such as voxel size.[34,35] Although the impact of these factors on MRI is unknown, harmonization of MRI imaging parameters may further increase the predictive power of radiomic models. The model results were consistent despite these cohort differences, which suggests that MRI-based radiomics models may translate well into sarcoma practice. Nevertheless, designing future radiomics studies to include more uniform data sets with respect to STS histotype and imaging and treatment protocols, such as those from cohorts treated in clinical trials, would eliminate these possible sources of variability.

Recent work in other disease sites suggests important future directions for MRI-based radiomics studies in STS. For instance, radiogenomics studies have been successful in correlating radomics features with gene expression patterns in NSCLC,[21] low-grade glioma,[36] and renal cell carcinoma.[37] Another study has identified a radiomic biomarker for CD8 cell tumor infiltration in NSCLC and response to immune therapy.[38] Prior work suggests that individual histotypes of STS have differential immunophenotypes,[39] which may affect response to therapy. Such future radiogenomics studies in STS may allow for noninvasive profiling of this heterogeneous group of tumors and serial response assessment as patients undergo therapy.

Conclusions

In conclusion, this study found, in independent derivation and validation cohorts, that radiomic features extracted from MRI are promising biomarkers for predicting OS in patients with STS. Optimal models using clinical and radiomic features could improve personalized selection of therapy in patients with STS.

Acknowledgments

The authors thank Dr. Larry Pierce for development of the PORTS software for radiomic feature extraction.

Supplementary data

Supplementary material for this article can be found at available online at https://doi.org/10.1016/j.adro.2019.02.003.

References

1. Sarcoma Meta-Analysis Collaboration. Adjuvant chemotherapy for localised resectable soft tissue sarcoma in adults. *Cochrane Database Syst Rev*. 2000:CD001419.
2. Pisters PW, Pollock RE, Lewis VO, et al. Long-term results of prospective trial of surgery alone with selective use of radiation for patients with T1 extremity and trunk soft tissue sarcomas. *Ann Surg*. 2007;246:675-681. discussion 681-672.
3. Eary JF, O'Sullivan F, Powitan Y, et al. Sarcoma tumor FDG uptake measured by PET and patient outcome: A retrospective analysis. *Eur J Nucl Med Mol Imaging*. 2002;29:1149-1154.
4. O'Sullivan F, Roy S, O'Sullivan J, Vernon C, Eary J. Incorporation of tumor shape into an assessment of spatial heterogeneity for human sarcomas imaged with FDG-PET. *Biostatistics*. 2005;6:293-301.
5. Skamene SR, Rakheja R, Dahlstrom KR, et al. Metabolic activity measured on PET/CT correlates with clinical outcomes in patients with limb and girdle sarcomas. *J Surg Oncol*. 2014;109:410-414.
6. Wolsztynski E, O'Sullivan F, Keyes E, O'Sullivan J, Eary JF. Positron emission tomography-based assessment of metabolic gradient and other prognostic features in sarcoma. *J Med Imaging (Bellingham)*. 2018;5:024502.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016;278:563-577.
8. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the

prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol*. 2015;60:5471-5496.
9. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer*. 2011;11:541-557.
10. Kinahan P, Pierce L, Nyflot M, et al. *An Open-Science Toolkit for Image Texture Analysis of PET Oncology Images*. Chicago, IL: Radiologcal Society of North America; 2015.
11. Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybernet*. 1973;SMC-3:610-621.
12. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybernet*. 1989;19:1264-1274.
13. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: Application to cell classification. *IEEE Trans Biomed Eng*. 2014;61:630-637.
14. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81:515-526.
15. Tibshirani R. The LASSO method for variable selection in the Cox model. *Stat Med*. 1997;16:385-395.
16. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: A review. *BMC Med*. 2010;8:21.
17. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
18. Uno H, Cai T, Tian L, Wei J. Evaluating prediction rules for t-year survivors with censored regression models. *J Am Stat Assoc*. 2012;102:527-537.
19. Davison A, Hinkley D. *Bootstrap Methods and their Application*. Cambridge, UK: Cambridge University Press; 1997.
20. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Stat Med*. 2013;32:1467-1482.
21. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
22. Huang Y, Liu Z, He L, et al. Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. 2016;281:947-957.
23. van Timmeren JE, Leijenaar RTH, van Elmpt W, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiother Oncol*. 2017;123:363-369.
24. Chaddad A, Sabri S, Niazi T, Abdulkarim B. Prediction of survival with multi-scale radiomic analysis in glioblastoma patients. *Med Biol Eng Comput*. 2018;56:2287-2300.
25. Cozzi L, Dinapoli N, Fogliata A, et al. Radiomics based analysis to predict local control and survival in hepatocellular carcinoma patients treated with volumetric modulated arc therapy. *BMC Cancer*. 2017;17:829.
26. Gronchi A, Ferrari S, Quagliuolo V, et al. Histotype-tailored neoadjuvant chemotherapy versus standard chemotherapy in patients with high-risk soft-tissue sarcomas (ISG-STS 1001): An international, open-label, randomised, controlled, phase 3, multicentre trial. *Lancet Oncol*. 2017;18:812-822.
27. Schwarzbach MH, Dimitrakopoulou-Strauss A, Willeke F, et al. Clinical value of [18-F]] fluorodeoxyglucose positron emission tomography imaging in soft tissue sarcomas. *Ann Surg*. 2000;231:380-386.
28. O'Sullivan F, Roy S, Eary J. A statistical measure of tissue heterogeneity with application to 3D PET sarcoma data. *Biostatistics*. 2003;4:433-448.
29. Kubo T, Furuta T, Johan MP, Ochi M. Prognostic significance of (18)F-FDG PET at diagnosis in patients with soft tissue sarcoma and bone sarcoma; systematic review and meta-analysis. *Eur J Cancer*. 2016;58:104-111.
30. Cheng NM, Fang YH, Lee LY, et al. Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. *Eur J Nucl Med Mol Imaging*. 2015;42:419-428.
31. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: Relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med*. 2014;55:414-422.
32. Hatt M, Majdoub M, Vallières M, et al. 18F-FDG PET uptake characterization through texture analysis: Investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med*. 2015;56:38-44.
33. Baldini EH, Abrams RA, Bosch W, et al. Retroperitoneal sarcoma target volume and organ at risk contour delineation agreement among NRG sarcoma radiation oncologists. *Int J Radiat Oncol Biol Phys*. 2015;92:1053-1059.
34. Mackin D, Fave X, Zhang L, et al. Measuring computed tomography scanner variability of radiomics features. *Invest Radiol*. 2015;50:757-765.
35. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE. Quantitative radiomics: Impact of stochastic effects on textural feature analysis implies the need for standards. *J Med Imaging (Bellingham)*. 2015;2:041002.
36. Han Y, Xie Z, Zang Y, et al. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an MRI-based radiomics signature in lower-grade gliomas. *J Neurooncol*. 2018;140:297-306.
37. Karlo CA, Di Paolo PL, Chaim J, et al. Radiogenomics of clear cell renal cell carcinoma: Associations between CT imaging features and mutations. *Radiology*. 2014;270:464-471.
38. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: ＼An imaging biomarker, retrospective multicohort study. *Lancet Oncol*. 2018;19:1180-1191.
39. Pollack SM, He Q, Yearley JH, et al. T-cell infiltration and clonality correlate with programmed cell death protein 1 and programmed death-ligand 1 expression in patients with soft tissue sarcomas. *Cancer*. 2017;123:3291-3304.