CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Assessment of the prognostic value of radiomic features in 18F-FMISO PET imaging of hypoxia in postsurgery brain cancer patients: secondary analysis of imaging data from a single-center study and the multicenter ACRIN 6684 trial
Authors	Muzi, Mark;Wolsztynski, Eric;Fink, James R.;O'Sullivan, Janet N.;O'Sullivan, Finbarr;Krohn, Kenneth A.;Mankoff, David A.
Publication date	2020-03-01
Original Citation	Muzi, M., Wolsztynski, E., Fink, J. R., O'Sullivan, J. N., O'Sullivan, F., Krohn, K. A. and Mankoff, D. A. (2020) 'Assessment of the Prognostic Value of Radiomic Features in 18F-FMISO PET Imaging of Hypoxia in Postsurgery Brain Cancer Patients: Secondary Analysis of Imaging Data from a Single-Center Study and the Multicenter ACRIN 6684 Trial', Tomography, 6(1), pp. 14-22. doi: 10.18383/j.tom.2019.00023
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.mdpi.com/2379-1381/6/1/14 - 10.18383/j.tom.2019.00023
Rights	© 2020 The Authors. Published by Grapho Publications, LLC This is an open access article under the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/) - http://creativecommons.org/licenses/by/4.0/
Download date	2026-11-16 10:25:00
Item downloaded from	https://hdl.handle.net/10468/11085

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

TOMOGRAPHY® RESEARCH ARTICLE

Assessment of the Prognostic Value of Radiomic Features in ^{18}F-FMISO PET Imaging of Hypoxia in Postsurgery Brain Cancer Patients: Secondary Analysis of Imaging Data from a Single-Center Study and the Multicenter ACRIN 6684 Trial

Mark Muzi[1], Eric Wolsztynski[2,4], James R. Fink[1], Janet N. O'Sullivan[2], Finbarr O'Sullivan[2], Kenneth A. Krohn[1], and David A. Mankoff[3]

[1]Department of Radiology, University of Washington, Seattle, WA; [2]Department of Statistics, University College, Cork, Ireland; [3]Department of Radiology, University of Pennsylvania, Philadelphia, PA and; [4]Insight Centre for Data Analytics, Cork, Ireland

Corresponding Author:
Mark Muzi, PhD
1959 NE Pacific Street, RR650 Neurology UW356465, Seattle, WA 98195 USA;
E-mail: muzi@uw.edu

Key Words: Fluoromisonidazole, PET imaging, brain cancer, ACRIN 6684, radiomics

Abbreviations: ^{18}F-fluoromisonidazole (^{18}F-FMISO), University of Washington (UW), American College of Radiology Imaging Network (ACRIN), World Health Organization (WHO), Response Assessment in Neuro-Oncology Working Group (RANO), Karnofsky Performance Status (KPS), time-to-progression (TTP), overall survival (OS), overall survival at 1 year from diagnosis (OS-1), 3-dimensional iterative reconstruction method (3D IR), ordered subset expectation maximization reconstruction method (OSEM), point spread function reconstruction method (PSF), standard uptake value (SUV), maximum voxel in a tissue region in SUV units (SUVmax), average SUV from a 1-cm circular ROI centered over the hottest voxel (SUVpeak), maximal activity voxel in a tissue region normalized by the blood activity (T/Bmax), hypoxic volume of pixels in a region that are above T/B = 1.2 hypoxic threshold (HV), gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM)

ABSTRACT

Hypoxia is associated with resistance to radiotherapy and chemotherapy in malignant gliomas, and it can be imaged by positron emission tomography with ^{18}F-fluoromisonidazole (^{18}F-FMISO). Previous results for patients with brain cancer imaged with ^{18}F-FMISO at a single center before conventional chemoradiotherapy showed that tumor uptake via T/Bmax (tissue SUVmax/blood SUV) and hypoxic volume (HV) was associated with poor survival. However, in a multicenter clinical trial (ACRIN 6684), traditional uptake parameters were not found to be prognostically significant, but tumor SUVpeak did predict survival at 1 year. The present analysis considered both study cohorts to reconcile key differences and examine the potential utility of adding radiomic features as prognostic variables for outcome prediction on the combined cohort of 72 patients with brain cancer (30 University of Washington and 42 ACRIN 6684). We used both ^{18}F-FMISO intensity metrics (T/Bmax, HV, SUV, SUVmax, SUVpeak) and assessed radiomic measures that determined first-order (histogram), second-order, and higher-order radiomic features of ^{18}F-FMISO uptake distributions. A multivariate model was developed that included age, HV, and the intensity of ^{18}F-FMISO uptake. HV and SUVpeak were both independent predictors of outcome for the combined data set ($P < .001$) and were also found significant in multivariate prognostic models ($P < .002$ and $P < .001$, respectively). Further model selection that included radiomic features showed the additional prognostic value for overall survival of specific higher order texture features, leading to an increase in relative risk prediction performance by a further 5%, when added to the multivariate clinical model..

INTRODUCTION

Evolving and immature tumor vasculature can lead to chronic tissue hypoxia, which is associated with resistance to radiotherapy and chemotherapy in malignant gliomas, leading to treatment failure and recurrence (1). Chronic hypoxia in patients with glioma has been quantified by using ^{18}F-fluoromisonidazole (^{18}F-FMISO) positron emission tomography (PET) imaging. This tracer is irreversibly reduced and trapped when the level of O_2 is below

© 2020 The Authors. Published by Grapho Publications, LLC This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
ISSN 2379-1381 https://doi.org/10.18383/j.tom.2019.00023

3 mmHg—the threshold for the oxygen enhancement effect of ionizing radiation (2). Several studies have shown that metrics of brain tumor hypoxia measured by ^{18}F-FMISO PET were associated with time-to-progression (TTP) and/or overall survival (OS) (3–6). Previous studies at the University of Washington (UW) of 22 patients imaged with ^{18}F-FMISO PET following surgical resection and before radiotherapy with concomitant chemotherapy showed that the maximum ^{18}F-FMISO tissue–blood ratio (T/Bmax) and the hypoxic volume (HV) were associated with shorter OS and TTP (4). However, in a multicenter clinical trial of 42 patients imaged at a similar time point [ACRIN 6684 (5)], T/Bmax and HV were not found to be prognostic, although SUVpeak (defined as the average SUV from a 1-cm-diameter circular region centered on the tumor voxel with greatest activity) predicted OS at 1 year (OS-1) (5). The current analysis considered both the UW and ACRIN 6684 study cohorts with a view to examining differences in the analytical results, and to also explore the utility of agnostic measures of ^{18}F-FMISO uptake using radiomic features to further assess patient risk on the combined cohort of patients.

METHODOLOGY

Patient Characteristics

The study consisted of 2 cohorts of newly diagnosed patients with high-grade glioma imaged with ^{18}F-FMISO PET after histopathological diagnosis and before initial standard treatment of radiation with concomitant chemotherapy. For the UW cohort, 30 patients (27 WHO grade IV glioblastoma multiforme and 3 WHO grade III anaplastic astrocytoma; median age, 58 years [range, 32 to 77], females, 9; and males, 21) underwent ^{18}F-FMISO imaging studies after surgical resection and before standard treatment (see online supplemental Table S1). Patients were selected and recruited by an experienced neuro-oncologist at the UW-affiliated hospital system on the basis of histopathologic diagnosis, clinical performance status, and presence of residual tumor volume on postsurgical contrast-enhanced T1 magnetic resonance imaging (MRI). Signed informed consent, as approved by the respective Investigational Review Boards and Radiation Safety Committees, was obtained for all patients before ^{18}F-FMISO imaging. All UW patients were followed through periodic clinical evaluations that included Karnofsky Performance Status (KPS) and routine contrast-enhanced MRI surveillance for tumor progression and OS. Tumor progression was determined by RANO criteria for high-grade glioma (7). The majority of these patients (28 out of 30) were included in previous reports examining the relationship of ^{18}F-FMISO imaging for the prediction of TTP and OS (4, 8).

The second cohort (n = 42) was recruited through the American College of Radiology Imaging Network (ACRIN) 6684 trial from 11 participating academic sites including UW, and the patients included in the cohort had histopathologically proven, newly diagnosed glioblastoma multiforme. The trial was approved by the institutional review boards at all participating sites. Before enrollment, all patients signed an informed consent document. Eligible patients had undergone surgical resection before planned standard-of-care treatment of radiation with chemotherapy. In addition, patients could receive investigational agent(s) at any time during the clinical trial. Study candidates were required to have evidence of residual tumor following surgery based on MRI imaging, although the minimal amount of residual tumor was not specified and, unlike the UW cohort, the presence of contrast-enhancement was not required. The ACRIN protocol thus had much less restrictive inclusion criteria. Other key patient characteristics included a median age of 60 years (range, 30–77 years), with 15 females and 27 males. Patients were followed every 3 months for tumor progression and survival for a minimum of at least 1 year or until death (median follow-up period until death, 227 days; range, 74–1034 days), while the patients in the UW group was followed continuously until death and therefore had a longer follow-up. Eleven of the 42 ACRIN patients were alive at the end of the study the follow-up period, which terminated 12 months after the last study patient was imaged. Routine clinical parameters considered for statistical analysis included age, gender, and KPS.

Scanner Qualification

The UW tomography (GE Advance, GE Medical Systems, Waukesha, WI) was calibrated on a quarterly basis for conversion of counts to Becquerel per milliliter using the ACR flood phantom containing known activities of ^{18}F imaged separately from a patient and reconstructed using identical parameters as patient emission images. Sample aliquots of the flood phantom were used to cross-calibrate the gamma well counter (Packard COBRA II, Meriden, CT) where patient blood samples were assayed to determine ^{18}F concentration. For the ACRIN study participants (11 institutions), PET/computed tomography (CT) scanners (GE Discovery STE, GE Discovery RX, GE Discovery LS, Siemens ECAT Exact HR+, Siemens Biograph 40, and Siemens Biograph 64) were prequalified by ACRIN using ACR flood phantom scans of known activity and sample patient images and every 2 years following initial accreditation. The phantom scans required reconstruction with identical parameters as patient emission images. Scanners were required to undergo cross-calibration to their local gamma well counter (Packard, Perkin-Elmer, Capintec, Beckman, Baird Atomic) at each participating center for determining blood activity within a week of patient scanning. Cross-calibrations were done at the time of scanner calibration by counting 1-mL aliquots from the calibration phantom following a standardized protocol developed by UW for ACRIN. All qualification data were centrally reviewed. One scanner, in which 3 patients were enrolled, was a prototype Siemens brain MR/PET scanner (no model number) that could not be calibrated with a head coil in place.

Radiosynthesis

^{18}F-FMISO was initially prepared at UW using the glycidyl tosylate method (9), and then changed to the method developed by Lim (10) as modified by Adamsen (11). In all cases, the same purification by high-performance liquid chromatography was used. The product-specific activity ranged from 37 to 74 GBq/μmol at the time of injection with >98% radiochemical purity. ^{18}F-FMISO was administered by venous injection of a 10-mL solution of isotonic saline containing <10% (v/v) ethanol USP. The average injected dose for the UW studies was 259 MBq (range, 218–350 MBq). Early studies were done under Radioactive Drug Research Committee, but the majority was performed under

Investigational New Drug approval. The ACRIN patients' radiotracer doses were supplied by local production of ^{18}F-FMISO or through Cardinal Health with product characteristics specified by the NCI-held IND (#76042). High-performance liquid chromatography purification was not required. The mean injected dose was 237 MBq (range, 185–285 MBq).

PET Imaging Acquisition

For both cohorts, imaging with ^{18}F-FMISO PET was performed at a clinically relevant time after surgical tissue confirmation of GBM diagnosis and within 2 weeks before the start of conventional therapy. After patient immobilization, venous lines were established in each arm, one for tracer injection and the other for blood sampling. At UW, this was followed by either a 25-minute transmission scan for the GE Advance or for the ACRIN cohort, a low-dose CT scan for PET/CT scanners. For both cohorts, static 20-minute ^{18}F-FMISO PET emission scans were acquired in 3-dimensional, high-resolution mode starting ∼110 min after injection (mean, 118; range 90–155 minutes; n = 72) of ^{18}F-FMISO. Three venous blood samples were acquired at 5, 10, and 15 minutes after scan initiation for quantitative assessment of hypoxia status by normalizing tissue uptake to the blood concentration as previously described (8). Averaged blood activity was decay-corrected to the injection time and converted to the same units as the scanner (Bq/mL).

Image Reconstruction

UW ^{18}F-FMISO PET emission images were reconstructed by 3D-filtered backprojection (12) with corrections for attenuation, scatter, random events, dead time, and sensitivity, and these resulted in an in-plane spatial resolution of 2.34 mm with a postreconstruction 6-mm Hanning filter. Most ACRIN studies (39 out of 42) were scanned on PET/CT scanners, which used CT-based attenuation correction. A prototype Siemens brain MR/PET scanner used an estimated attenuation map based on the magnetic resonance (MR) signal (13). Reconstruction of emission images for the ACRIN studies used various iterative 3D techniques (3-dimensional iterative reconstruction method [3D IR], ordered subset expectation maximization reconstruction method [OSEM], point spread function reconstruction method [PSF]), wherein postreconstruction filter sizes ranged from 2 to 6 mm and reconstructed in-plane image resolution varied between 1.25 and 6 mm with corrections for attenuation, scatter, random events, decay, deadtime, and sensitivity (see online supplemental Table S2). Tomography sensitivity was calibrated as described in aforementioned sections.

Image Analysis

For the ACRIN cohort, ^{18}F-FMISO images were analyzed at UW by an experienced team that performed analysis of the original UW cohort and served as the central analysis laboratory for the ACRIN 6684 ^{18}F-FMISO PET studies. Images acquired with ^{18}F-FMISO PET were decay-corrected to the time of radiotracer injection and converted to SUV units by normalization of the emission image to the recorded dose and patient weight. Traditionally, volumes of interest (VOIs) were either segmented from the contrast-enhancing volumes on the MR T1+ contrast images and dilated by several voxels or manually constructed on the ^{18}F-FMISO image based on areas of ^{18}F-FMISO tracer uptake (4). Both methods proved adequate for routine extraction of ^{18}F-FMISO uptake information but proved inadequate for radiomic texture feature extraction. Therefore, VOIs over tumor were segmented from either coregistered Gd-enhanced MRI fluid-attenuated inversion recovery images or T2-weighted images obtained within 2 weeks of PET image acquisition, and before conventional therapy. Because ^{18}F-FMISO is a freely diffusible tracer (14) and tumor hypoxia may exist beyond the margins of Gd-enhancement on MRI, each tumor VOI was segmented to include the entire area of abnormal fluid-attenuated inversion recovery signal hyperintensity taking care not to include postsurgical changes such as the resection cavity and extra-axial fluid collections. An experienced neuroradiologist reviewed each MRI-based tumor volume for all patients. The segmented tumor VOI was then applied to the ^{18}F-FMISO PET SUV image using PMOD software (PMOD Tech., Version 3.8, Zurich CH) for data extraction. Owing to the relatively large size of the MRI-segmented tumor VOIs (4 cc–750 cc), partial volume correction was not required. In addition, a circular 1-cm-diameter SUVpeak region of interest was centered over the hottest voxel in the tumor. Extracted image data from ^{18}F-FMISO scans were normalized using the average measured blood activity (Bq/mL) during the scan to produce tissue–blood ratio (T/B) values for all voxels in each image slice. The pixel with maximum T/B value (T/Bmax) and the hypoxic volume (HV) of voxels above the T/B threshold of 1.2, indicating significant hypoxia, were determined to quantify the volume of tissue hypoxia in each tumor region (4). To include the 3 patients scanned on the MRI/PET scanner, their image data were scaled by the average ratio of blood to cerebellar cortex activity, as these values are highly correlated (8) for standard PET imaging. Conventional PET and hypoxia parameters extracted from tumor regions included VOI volume, SUVmax, SUVmean, SUVpeak, T/Bmax, HV, and TLH (total lesion hypoxia = SUVmean × VOI).

Radiomic Analysis

We evaluated a set of radiomic features assessing texture characteristics of the hypoxia distribution of the 72 patients with brain cancer from the 2 cohorts who had ^{18}F-FMISO PET studies with concurrent blood sampling. An adaptive thresholding method (30% of SUVmax) was then applied to the segmented tumor VOIs, yielding volumes that were restricted from void regions, but including voxels with ^{18}F-FMISO uptake (Figure 1). Preprocessing of ^{18}F-FMISO emission images for radiomic analysis included spatial normalization of the image data and tumor VOI segmentation mask, based on the segmentations used for standard analyses noted above, and interpolated in 3 dimensions at the lowest resolution of the 2 cohorts (1 mm) via trilinear interpolation. The exploratory investigation on the association to outcome considered the following biomarkers detailed in the Image Biomarker Standardization Initiative (IBSI) reference manual (15) for analysis: 12 first-order statistical variables reflecting intensity (IBSI Section 3.4), 20 histogram features of the discretized uptake values (IBSI Section 3.5), 22 texture features of the gray-level co-occurrence matrix (GLCM) (IBSI Section 3.6), 16 GLRLM-based features emphasizing discretized images with long-run lengths, or sequences of voxels with the

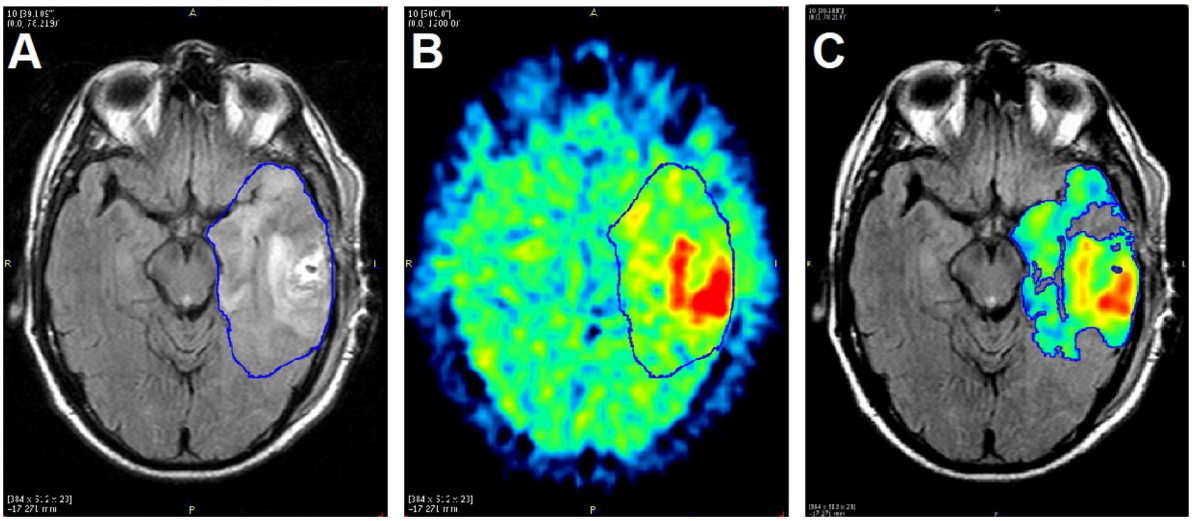


Figure 1. Image segmentation for data extraction. The tumor region is segmented from the magnetic resonance imaging (MRI) fluid-attenuated inversion recovery (FLAIR) (blue line) as a region of hyperintensity (A). The region is applied to the ^{18}F-fluoromisonidazole (^{18}F-FMISO) standard uptake value (SUV) image (B). Adaptive thresholding of the ^{18}F-FMISO SUV image yields the tumor volume for radiomic analysis that avoids regions of low uptake (C).

same discretized intensity (IBSI Section 3.7) and 16-GLSZM-based features evaluating entropy (or uniformity) in the distribution of groups of connected voxels with the same discretized intensity (IBSI Section 3.8). Following the guidelines of Vallières at al. (16) on responsible radiomics, the online supplemental online data provides a description of the radiomic extraction process and ontology of the radiomic features used in this report. An open-source software implementation of radiomic features in R (mia 1.0.4, https://github.com/ericwol/mia) was used for radiomic quantitation of the ^{18}F-FMISO uptake distribution. Certain features based on the GLCM (eg, contrast, entropy, dissimilarity, homogeneity) have been shown to have high predictive value for PET imaging (17–19). In total, there were 97 variables for each patient that were considered for model selection—11 conventional clinical and PET variables and 86 radiomic PET features.

Statistical Analysis

Comparisons between cohorts for parameters such as uptake time and the 97 other variables were made using a 2-sample Wilcoxon (Mann–Whitney) test for equal distributions (20). To examine the relationship of clinical, PET, and textural biomarkers to OS, the following methods were used.

Univariate Model. Univariate Cox proportional model analyses were applied to all variables on each cohort and the combined cohort to assess the statistical significance of each variable separately for OS prognosis. In these analyses, significance levels were set at the 5% level with Bonferroni correction to limit the risk of a type 1 error (false positives).

Multivariate Model. Multivariate prognostic modeling was performed in 2 steps: first to explore the prognostic value of a model that comprises conventional clinical variables, thereby selecting the most parsimonious base reference clinical model, before evaluating the potential for prognostic improvement by adding radiomic features.

In the first step, we performed optimal subset selection by evaluating all possible Cox proportional hazards models comprising combinations of the clinical variables to identify the AIC (Akaike information criterion)-minimizing (21) model from only conventional clinical and PET variables (see online supplemental Figure S1). The analysis resulted in a base clinical model with a parameter combination of age, HV of the tracer, and the intensity of the tracer (T/Bmax and SUVpeak).

In the second step, preliminary feature elimination was performed to remove radiomic variables that had a strong correlation with other clinical or radiomic variables. We eliminated features with a Pearson correlation >80% in absolute value. Forward stepwise selection (22) was then carried out to yield a second multivariate Cox model. This search allowed selection in both forward and backward directions of any remaining conventional or radiomic variables to the base clinical model from the first step. This step tested the additional prognostic benefit from radiomic features.

The hazard ratio along with its 95% confidence interval and associated *P*-value based on Wald's statistic were reported for the Cox models. Prognostic significance was assessed by bootstrapped median *P*-value (using 1000 bootstrap resamples), and considered both at the 5% significance level and with Bonferroni correction. The likelihood ratio test (LRT) was used to confirm statistical significance of the addition of textural variables to the base clinical model (23, 24). The concordance index of the multivariate prognostic model assessed potential gain in predictive performance over the base clinical model. Kaplan–Meier survival curve estimates from diagnosis to time-to-death were generated. All analyses were done using R (CRAN R-Project; https://cran.r-project.org).

Table 1. ^{18}F-FMISO Tumor Uptake Values

		SUVmax (g/mL)	SUVpeak (g/mL)	T/Bmax (unitless)	HV (mL)
UW	Mean	3.15	2.53	2.13	28.40
	SD	1.11	1.04	0.71	35.51
	Median	2.94	2.15	2.00	12.28
	Max	6.53	5.64	4.26	131.90
	Min	1.85	1.34	1.27	1.59
ACRIN	Mean	3.20	2.49	2.10	13.39
	SD	1.14	0.84	0.75	11.49
	Median	2.96	2.26	2.02	10.42
	Max	6.26	5.00	4.14	41.18
	Min	1.80	1.45	0.95	0.00
ALL	Mean	3.18	2.51	2.11	19.64
	SD	1.12	0.92	0.73	25.43
	Median	2.94	2.24	2.00	11.39
	Max	6.53	5.64	4.26	131.90
	Min	1.80	1.34	0.95	0.00

^{18}F-FMISO PET uptake parameters for the single-center UW, multicenter ACRIN 6684 and the combined cohorts of all patients. T/Bmax is the maximum tumor voxel normalized by the blood activity at the time of scanning, and HV refers to the hypoxic volume of pixels above the T/B threshold of 1.2 indicating significant hypoxia.

RESULTS

Considering patient characteristics of the 2 cohorts, the UW one had lower KPS (mean, 80; range, 60–100 KPS) than the ACRIN cohort (mean, 87; range, 70–100 KPS; Wilcoxon test, $P = .028$). The uptake time between injection and scanning for the UW cohort (mean, 128 minutes; range, 100–155 minutes) was longer than the ACRIN cohort on the average (mean, 111 minutes; range, 90–145 minutes; Wilcoxon test, $P < .001$). While the UW group had a combination of grades III and IV patients, removing grade III patients did not yield significantly different results. Methodological differences in attenuation, reconstruction algorithm, and filter size varied significantly between cohorts (see online supplemental Table S1). Conventional PET uptake parameters for ^{18}F-FMISO (SUVmax, SUVpeak, T/Bmax, and HV) appear in Table 1. A notable difference in distribution between UW and ACRIN cohorts was observed for only KPS, on the basis of Bonferroni-corrected Wilcoxon tests. At the 5% significance level, without Bonferroni correction, only one other variable (run-length nonuniformity, a GLRLM-based texture feature) yielded a difference in cohort distributions ($P = .002$). Reconstruction methods are known to have a substantial effect on radiomic features (25, 26), and there were clearly reconstruction differences between the 2 cohorts.

Univariate Analysis

PET parameters were assessed on the basis of overall patient survival status at last follow-up, including censoring information. Univariate Cox proportional modeling combining the 2 cohorts showed that 24 variables were significant without Bonferroni correction, but only 9 were significant for prognosis at the 5% significance level with Bonferroni correction (Table 2). In particular, age, SUVpeak, and HV were all independent predictors of outcome in the combined data set after Bonferroni correction ($P < .001$ for all). Comparing the results between the 2 data sets, the UW cohort had all 9 variables from Table 2 remain significant at the 5% significance level, with HV and SUVpeak both remaining independent predictors after Bonferroni correction. Six of the 9 variables were significant independent predictors at the 5% significance level in the ACRIN group (Table 2), where HV, SUVpeak, and GLCM energy were not significant.

Multivariate Analysis

Following selection of the best parameter subset for model composition, the most parsimonious base conventional parameter Cox model with optimal AIC value included age, HV, SUVpeak, and T/Bmax, and is summarized in Table 3. All 4 variables in this model were statistically significant prognostic markers at the 5% significance level, with HV and age remaining significant after Bonferroni correction, using threshold significance level $0.05/P_{clinical} < 0.005$, where $P_{clinical} = 11$ clinical variables.

The second multivariate prognostic model obtained by extending the base conventional model is presented in Table 4. It illustrates the enhanced prognostic potential obtained by adding texture features to the base model. Although clinical variables were permitted in the model selection procedure, texture features were selected over conventional variables by the more objective stepwise selection algorithm. This supports a potential benefit in introducing texture analyses for postresection risk characterization in glioma. The likelihood ratio test confirmed the statistical significance ($P < .010$) of the additional contribution of this set of texture variables. We observed a reasonably high concordance index of 77.4% for the model of Table 2. This index measures the pairwise probability of assessing a lower patient risk, based on

Table 2. Univariate Analysis Hazard Ratios (*P*-value)

Parameter	Combined		ACRIN Cohort		UW Cohort	
Age	1.94	**(<.001)**	2.45	(.002)	1.58	(.017)
SUVpeak	1.71	**(<.001)**	1.46	(.070)	2.24	**(.001)**
HV	1.80	**(<.001)**	0.98	(.957)	1.85	**(.001)**
SUVmean	1.87	**(<.001)**	2.03	(.001)	1.70	(.004)
Median	1.82	**(<.001)**	2.36	(.001)	1.53	(.012)
p10	1.71	**(<.001)**	2.08	(.006)	1.48	(.026)
p90	1.71	**(<.001)**	1.64	(.002)	2.02	(.001)
GLCM Energy	1.73	**(<.001)**	1.27	(.451)	1.73	(.001)
RMS	1.88	**(<.001)**	1.99	(.001)	1.72	(.003)

Variables that were found statistically significant in univariate survival analyses after Bonferroni correction, with associated *P*-values for the combined, ACRIN, and UW cohorts respectively. *P*-values in bold indicate significance with respect to the Bonferroni correction threshold of 0.05/97 = 0.0005. Median, p10, p90, and RMS, respectively, correspond to the sample median, 10th and 90th percentiles, and root mean square (quadratic mean) of the tumor SUV values.

the model, given the longer survival time, and is equivalent to the AUC value in an ROC analysis. The addition of the texture features to the base clinical model led to an increase in relative risk prediction performance, from 72.2% to 77.4%.

Figure 2 shows the Kaplan–Meier OS curve estimates obtained from the base (dashed red lines) and extended (solid blue lines) multivariate Cox models. Dichotomous separation of the combined cohort into higher- and lower-risk groups (lower and higher curves respectively) was obtained by maximizing the statistical separation in terms of the log-rank test statistic. The separation of high-risk from low-risk patients was statistically significant for both models (log-rank test, $P < .001$).

DISCUSSION

We investigated the association of ^{18}F-FMISO PET hypoxia parameters with survival between a single-center (UW) cohort and a multicenter (ACRIN 6684) cohort. Although both studies showed prognostic value for ^{18}F-FMISO, the 2 studies found the association with different measures of tracer uptake. Previous results on a subset of 22 patients of the UW cohort showed that age, T/Bmax, and HV were significant in predicting OS (4) in univariate analyses. A similar result for OS was shown here using a larger UW cohort of 30 patients (Table 2), while in the multicenter ACRIN cohort, only SUVpeak was significant in univariate analyses for OS-1. Although all patients in both studies had pathologically proven brain tumors imaged before conventional therapy, the 2 cohorts appear different with respect to traditional ^{18}F-FMISO uptake parameters, particularly HV. A potential explanation of the differences between the 2 individual cohorts most likely lies in patient selection. The UW cohort had larger hypoxic volumes of ^{18}F-FMISO and incidentally had lower KPS, indicating greater disease involvement, although there were no differences in postsurgery contrast-enhancement volume on MRI T1 between the UW cohort and the ACRIN cohort. Unlike the UW group, a few of the ACRIN patients had no ^{18}F-FMISO uptake and therefore no measurable hypoxic volume. For example, hypoxic volumes were <4 cc in nearly 30% (12 out of 42) of patients in the ACRIN cohort compared with in 7% (2 out of 30) of patients in the UW cohort. Low residual tumor volume at the time of ^{18}F-FMISO imaging may not permit adequate assessment of hypoxia by any measure, such as tracer distribution, intensity, or structural characteristics of spatial uptake distribution. A smaller HV limits image count statistics in the ACRIN data set and along with the added noise of variability in cross-calibration where multiple sites differ in scanner, image reconstruction, counters, and blood sampling/counting skill levels may well explain why T/Bmax and HV did not work better for univariate ACRIN data set analysis. In addition, the T/B threshold of 1.2 that defines the volume of hypoxic voxels within the tumor was developed using filtered backprojection image data, and could well be different for images using an estimation reconstruction analysis.

Heterogeneity of treatment in the ACRIN cohort versus relatively uniform treatment in the single-center UW study may also affect study outcome. The ACRIN cohort had many patients with alternative therapies (Lomustine, Paclitaxel, Poliglumex, Selinexor, Vorinostat, Dasatinib, Carmustine, Bevacizumab) during and following conventional therapy; the UW cohort had

Table 3. Base Multivariate Cox Model Analysis

Variable	Effect	HR (95% CI)		Median *P*
Age	0.74	2.10	(1.52, 2.91)	**<.001**
HV	0.54	1.72	(1.23, 2.41)	**.002**
SUVpeak	0.66	1.93	(1.18, 3.17)	.006
TBmax	−0.50	0.61	(0.39, 0.96)	.023

The base multivariate Cox model was comprised of 4 conventional variables, showing predictor effects, associated hazard ratio (HR) with associated 95% confidence interval (CI), and bootstrap median *P*-values obtained from 1000 bootstrap resamples. All 4 variables are significant at the 5% level; *P*-values in bold indicate significance after bonferroni correction ($P < .005$). This model has a concordance index of 0.722.

Table 4. Extended Multivariate Cox Model Analysis

Variable	Effect	HR (95% CI)		Median *P*
Age	0.84	2.31	(1.65,3.25)	**<.001**
HV	0.67	1.95	(1.28, 2.97)	.006
SUVpeak	0.97	2.63	(1.52, 4.56)	**.001**
TBmax	−0.75	0.47	(0.28, 0.80)	.010
HIST CoV[a]	1.29	3.65	(1.72, 7.73)	**.001**
HIST QCOD	−0.59	0.55	(0.36, 0.86)	.015
HIST Kurtosis	−0.55	0.58	(0.34, 0.98)	.048
GLRLM Long runs emphasis	−1.26	0.28	(0.14, 0.58)	**.001**
GLSZM Zone size entropy	0.74	2.10	(1.18, 3.71)	0.020
GLSZM Low gray-level zone emphasis	0.41	1.51	(0.94, 2.45)	0.110

Extended multivariate Cox model obtained by selecting additional variables (conventional or radiomic) to the base model, showing predictor effects, associated hazard ratio (HR) with associated 95% confidence interval (CI), and bootstrap median *P*-values obtained from 1,000 bootstrap resamples. Nine of the 10 variables are significant at the 5% level; *P*-values in bold indicate significance after Bonferroni correction. This model has a concordance index of 0.774. The likelihood ratio test confirmed the significant contribution ($P = .010$) of the additional set of radiomic variables.

[a] HIST CoV, HIST QCOD, and HIST Kurtosis, respectively refer to the coefficient of variation (CoV), quartile coefficient of dispersion (QCOD), and kurtosis of the histogram of discretized uptake values; these 3 features evaluate various aspects of discretized uptake distribution variability. GLRLM long runs emphasize discretized images with long-run lengths (ie, sequences of voxels with the same discretized intensity). GLSZM zone size entropy evaluates entropy (or uniformity) in the distribution of groups of connected voxels with the same discretized intensity. GLSZM low gray-level zone emphasis highlights discretized images containing more low-intensity regions.

fewer agents (Carboplatin, Irinotecan, Bevacizumab) and they were used only after conventional therapy had been completed. Treatments during and after ^{18}F-FMISO imaging may modulate the time to progression and potentially delay death, altering the survival outcome variables. Notable differences between the early survival characteristics of the 2 cohorts, the degree of tumor burden as observed with ^{18}F-FMISO HV, and the inclusion of alternative therapies may partially explain these findings. Both

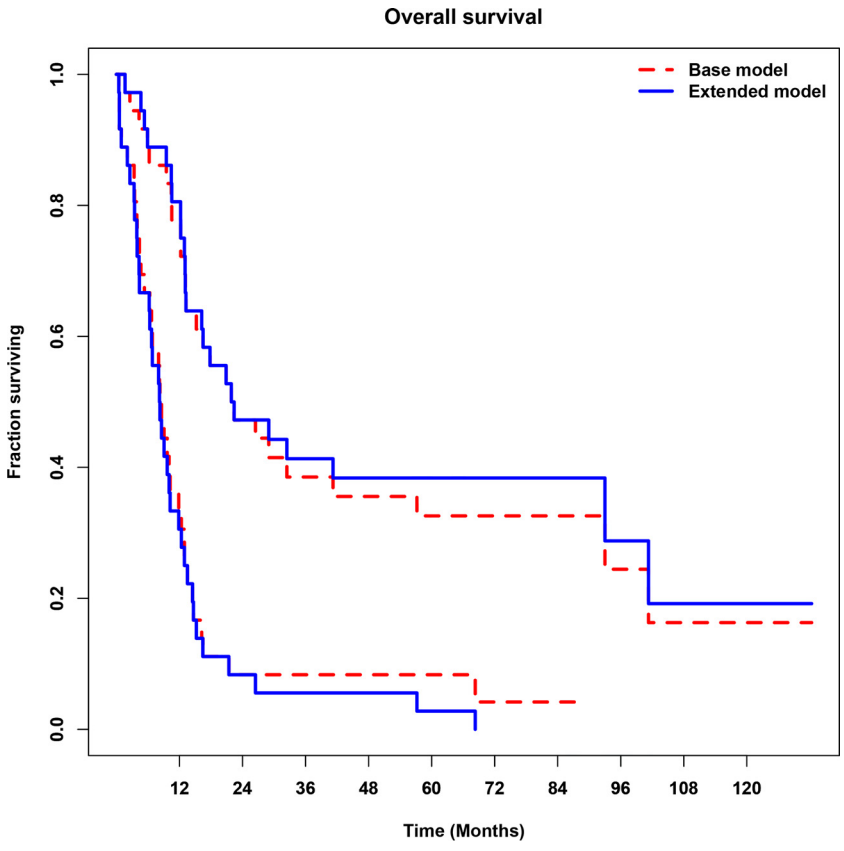


Figure 2. Survival analysis for the base and extended multivariate models on the combined cohorts. Kaplan–Meier analysis of overall survival (OS) with stratification into lower- and higher-risk groups (the higher and lower curves, respectively) based on maximized statistical separation in terms of the log-rank test statistic ($P < .001$), for the base 4-variable clinical model comprised age, hypoxic volume (HV), T/Bmax (maximal activity voxel in a tissue region normalized by the blood activity) and SUVpeak (average SUV from a 1-cm circular ROI centered over the hottest voxel; dashed red lines), and its extension using additional radiomic features (solid blue lines). The figure shows how the latter model led to different patient classification.

independent studies were carried out on limited sample sizes, and as such, differences in (or lack of) significance for some of the variables may also be a direct consequence of low statistical power.

The present analysis on the aggregated cohort highlighted the preserved significance of both sets of markers from each cohort. Each of the 4 variables in the multivariate base model of Table 3 was a statistically significant jointly prognostic marker at the 5% significance level. The multivariate model applied to the combined cohort thus assesses complementary aspects of ^{18}F-FMISO uptake by including the tracers' overall distribution (HV) and intensity of uptake (SUVpeak and T/Bmax).

Texture features, such as GLCM entropy and other radiomic measures could improve the prediction of patient outcome. The addition of 6 radiomic variables, identified via a formal statistical stepwise selection process, to the initial 4-variable clinical model led to an added 5% in concordance index, and their inclusion contributed to a multivariate Cox proportional hazards model that was deemed statistically significant ($P < .001$).

LIMITATIONS

The modest sample size and constraints on follow-up in this study are important to highlight. These issues arose owing to the limited access to FMISO radiotracer imaging for patients with glioma during the period following pathological confirmation and before conventional therapy. ACRIN funding limited patient follow-up to be a minimum of 1 year. At the end of the study, 9 patients were alive and thereafter lost to follow-up, and thus were censored.

Selection of the prognostic multivariate models presented in Tables 3 and 4 could be performed in a number of different ways. The base multivariate clinical model of Table 3 was selected using conventional best-subset selection, an exhaustive comparison of all possible models on the basis of some goodness-of-fit criterion adjusted for model dimension. We use AIC for this test, as it is one of the most commonly used metrics and a reasonable choice considering the relatively small number of candidate variables. The extended model of Table 4 is subject to selection bias, as it was constrained to include the 4 conventional variables of the base model of Table 3. As such it should be validated on an independent ^{18}F-FMISO PET-imaged glioma cohort. The modest sample size prevented the ability to carry out convincing training and test validation. A much larger cohort might have facilitated such an approach and thereby reduced the lingering risk of optimistic bias associated with retrospective analysis of small data sets. However, our models can be considered as demonstrations of prognostic potential, as an illustration of the opportunity for potential prognostic enhancement provided by radiomic analysis of ^{18}F-FMISO imaging data when combined with standard clinical models.

CONCLUSIONS

Differing significance of quantitative metrics for ^{18}F-FMISO uptake resulted between a single-center cohort and a multicenter cohort of glioma patients for the assessment of hypoxia using ^{18}F-FMISO PET before combined radiotherapy and chemotherapy. These may be explained by single-center patient selection bias for patients in the UW cohort with larger tumors that allow adequate assessment of hypoxia and by a substantial subset of multicenter patients with low or no HV in the other cohort. The present analysis on the aggregated cohort highlighted the preserved significance of ^{18}F-FMISO PET markers from both sets. Multiparametric modeling on the combined cohort showed that all of SUVpeak, HV and T/Bmax together with age have a significant predictive value when combined in a multivariate prognostic model. Our results also showed that adding selected radiomic features to this prognostic model further increased predictive performance for the combined group of patients and might be considered in future studies and potential future clinical applications.

Supplemental Materials

Supplemental Data: https://doi.org/10.18383/j.tom.2019.00023.sup.01

ACKNOWLEDGMENTS

We would like to recognize the substantial effort of the ACRIN and ECOG-ACRIN clinical trial teams that helped direct the study and provide the ACRIN data for this report. The imaging and clinical data for ACRIN 6684 are publicly available on TCIA (The Cancer Imaging Archive) at https://wiki.cancerimagingarchive.net/display/Public/ACRIN-FMISO-Brain.

This study was supported by NIH/NCI R50-CA211270, P01-CA042045, U01-CA079778, U01-CA080098, and Science Foundation Ireland Grants PI 11/1027 and 12/RC/2289-P2.

Disclosure: The authors have nothing to disclose.

REFERENCES

1. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff RO. European Organisation for Research and Treatment of Cancer Brain Tumour and Radiation Oncology Groups; National Cancer Institute of Canada Clinical Trials Group. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol. 2009;10:459–466.
2. Moeller BJ, Richardson Ra Fau - Dewhirst MW, Dewhirst MW. Hypoxia and radiotherapy: opportunities for improved outcomes in cancer treatment. Cancer Metastasis Rev. 2007;26:241–248.
3. Cher LM, Murone C, Lawrentschuk N, Ramdave S, Papenfuss A, Hannah A, O'Keefe GJ, Sachinidis JI, Berlangieri SU, Fabinyi G, Scott AM. Correlation of hypoxic cell fraction and angiogenesis with glucose metabolic rate in gliomas using 18F-fluoromisonidazole, 18F-FDG PET, and immunohistochemical studies. J Nucl Med 2006;47:410-418.
4. Spence AM, Muzi M, Swanson KR, O'Sullivan F, Rockhill JK, Rajendran JG, Adamsen TCH, Link JM, Swanson PE, Yagle KJ, Rostomily RC, Silbergeld DL, Krohn KA. Regional hypoxia in glioblastoma multiforme quantified with [18F]fluoromisonidazole positron emission tomography before radiotherapy: correlation with time to progression and survival. Clin Cancer Res. 2008;14:2623–2630.
5. Gerstner ER, Zhang Z, Fink JR, Muzi M, Hanna L, Greco E, Prah M, Schmainda KM, Mintz A, Kostakoglu L, Eikman EA, Ellingson BM, Ratai EM, Sorensen AG, Barboriak DP, Mankoff DA, Group AT. ACRIN 6684: assessment of tumor hypoxia in newly

diagnosed glioblastoma using 18F-FMISO PET and MRI. Clin Cancer Res. 2016;22:5079–5086.
6. Yamaguchi S, Hirata K, Toyonaga T, Kobayashi K, Ishi Y, Motegi H, Kobayashi H, Shiga T, Tamaki N, Terasaka S, Houkin K. Change in 18F-fluoromisonidazole PET is an early predictor of the prognosis in the patients with recurrent high-grade glioma receiving Bevacizumab treatment. PLoS One. 2016;11:e0167917.
7. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, Degroot J, Wick W, Gilbert MR, Lassman AB, Tsien C, Mikkelsen T, Wong ET, Chamberlain MC, Stupp R, Lamborn KR, Vogelbaum MA, van den Bent MJ, Chang SM. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. J Clin Oncol. 2010;28:1963–1972.
8. Muzi M, Peterson LM, O'Sullivan JN, Fink JR, Rajendran JG, McLaughlin LJ, Muzi JP, Mankoff DA, Krohn KA. 18F-fluoromisonidazole quantification of hypoxia in human cancer patients using image-derived blood surrogate tissue reference regions. J Nucl Med. 2015;56:1223–1228.
9. Grierson JR, Link JM, Mathis CA, Rasey JS, Krohn KA. A radiosynthesis of fluorine-18 fluoromisonidazole. J Nucl Med. 1989;30:343–350.
10. Lim JL, Berridge MS. An efficient radiosynthesis of [18F]fluoromisonidazole. Appl Radiat Isot. 1993;44:1085–1091.
11. Adamsen TCH, Grierson JR, Krohn KA. A new synthesis of the labeling precursor for F-18-fluoromisonidazole. J Label Compd Radiopharm. 2005;48:923–927.
12. Kinahan PE, Rogers JG. Analytic 3D image reconstruction using all detected events. IEEE Trans Nucl Sci. 1989;36:964–968.
13. Catana C, van der Kouwe A, Benner T, Michel CJ, Hamm M, Fenchel M, Fischl B, Rosen B, Schmand M, Sorensen AG. Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype. J Nucl Med. 2010;51:1431–1438.
14. Krohn KA, Link JM, Mason RP. Molecular imaging of hypoxia. J Nucl Med. 2008;49:129S–148S.
15. Zwanenburg A, Leger S, Lock S. Image Biomarker Standardisation Initiative (reference manual). arXiv preprint arXiv:1612.07003; 2019. https://arxiv.org/pdf/1612.07003.pdf.
16. Vallieres M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible radiomics research for faster clinical translation. J Nucl Med. 2018;59:189–193.
17. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, Corcos L, Visvikis D. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med. 2011;52:369–378.
18. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J Nucl Med. 2014;55:37–42.
19. Hatt M, Majdoub M, Vallieres M, Tixier F, Le Rest CC, Groheux D, Hindie E, Martineau A, Pradier O, Hustinx R, Perdrisot R, Guillevin R, El Naqa I, Visvikis D. 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.
20. Hollander M, Wolfe DA. Nonparametric Statistical Methods. 2nd ed. 1999: NY: John Wiley & Sons.
21. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19:716–723.
22. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. 2013, New York: Springer-Verlag.
23. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological). 1972;34:187–220.
24. Shakir H, Deng Y, Rasheed H, Khan T. Radiomics based likelihood functions for cancer diagnosis. Sci Rep. 2019;9:9501.
25. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, Tham IW, Townsend D. Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med. 2015;56:1667–1673.
26. Altazi BA, Zhang GG, Fernandez DC, Montejo ME, Hunt D, Werner J, Biagioli MC, Moros EG. Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. J Appl Clin Med Phys. 2017;18:32–48.