

---

---

# Repeatability of $^{18}\text{F}$ -FDG PET Radiomic Features in Cervical Cancer

John P. Crandall<sup>1</sup>, Tyler J. Fraum<sup>1</sup>, MinYoung Lee<sup>1</sup>, Linda Jiang<sup>1</sup>, Perry Grigsby<sup>2</sup>, and Richard L. Wahl<sup>1,2</sup>

<sup>1</sup>Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, Missouri; and <sup>2</sup>Department of Radiation Oncology, Washington University in Saint Louis, St. Louis, Missouri

Knowledge of the intrinsic variability of radiomic features is essential to the proper interpretation of changes in these features over time. The primary aim of this study was to assess the test–retest repeatability of radiomic features extracted from  $^{18}\text{F}$ -FDG PET images of cervical tumors. The impact of different image preprocessing methods was also explored. **Methods:** Patients with cervical cancer underwent baseline and repeat  $^{18}\text{F}$ -FDG PET/CT imaging within 7 d. PET images were reconstructed using 2 methods: ordered-subset expectation maximization (PET<sub>OSEM</sub>) or ordered-subset expectation maximization with point-spread function (PET<sub>PSF</sub>). Tumors were segmented to produce whole-tumor volumes of interest (VOI<sub>WT</sub>) and 40% isocontours (VOI<sub>40</sub>). Voxels were either left at the default size or resampled to 3-mm isotropic voxels. SUV was discretized to a fixed number of bins (32, 64, or 128). Radiomic features were extracted from both VOIs, and repeatability was then assessed using the Lin concordance correlation coefficient (CCC). **Results:** Eleven patients were enrolled and completed the test–retest PET/CT imaging protocol. Shape, neighborhood gray-level difference matrix, and gray-level cooccurrence matrix features were repeatable, with a mean CCC value of 0.81. Radiomic features extracted from PET<sub>OSEM</sub> images showed significantly better repeatability than features extracted from PET<sub>PSF</sub> images ( $P < 0.001$ ). Radiomic features extracted from VOI<sub>40</sub> were more repeatable than features extracted from VOI<sub>WT</sub> ( $P < 0.001$ ). For most features (78.4%), a change in bin number or voxel size resulted in less than a 10% change in feature value. All gray-level emphasis and gray-level run emphasis features showed poor repeatability (CCC values  $< 0.52$ ) when extracted from VOI<sub>WT</sub> but were highly repeatable (mean CCC values  $> 0.96$ ) when extracted from VOI<sub>40</sub>. **Conclusion:** Shape, gray-level cooccurrence matrix, and neighborhood gray-level difference matrix radiomic features were consistently repeatable, whereas gray-level run length matrix and gray-level zone length matrix features were highly variable. Radiomic features extracted from VOI<sub>40</sub> were more repeatable than features extracted from VOI<sub>WT</sub>. Changes in voxel size or SUV discretization parameters typically resulted in relatively small differences in feature value, though several features were highly sensitive to these changes.

**Key Words:** FDG; PET; radiomics; repeatability

**J Nucl Med** 2021; 62:707–715

DOI: 10.2967/jnumed.120.247999

**T**he role of PET/CT with  $^{18}\text{F}$ -FDG in oncologic diagnosis, staging, and treatment monitoring is well established (1,2). With the increasing use of functional imaging modalities such as PET/CT, interest in the quantification of image data has grown. Despite the existence of established quantitative frameworks for response assessment such as RECIST and PERCIST, most interpretations in clinical practice are still based primarily on subjective visual interpretations (3,4).

Tumor metabolism is commonly quantified with SUV metrics, including SUV<sub>max</sub>, SUV<sub>mean</sub>, and SUV<sub>peak</sub> (5). Radiomic features are complex quantitative imaging biomarkers purported to provide additional information beyond intensity-based SUV metrics (6). Extraction of radiomic features typically incorporates image preprocessing, segmentation, and feature calculation. Each of these steps has been shown to affect the outcome of radiomic analyses (7).

$^{18}\text{F}$ -FDG PET is now part of the standard of care for the evaluation of cervical cancer in the United States (8). In cervical cancer patients, radiomic features extracted from  $^{18}\text{F}$ -FDG PET images have been shown to predict both survival and disease recurrence (9,10). As PET is also useful for monitoring response to cervical cancer treatment, there is interest in using radiomic methods to further this purpose (11–13). However, to use these novel features effectively, their intrinsic variability must be formally quantified. Unfortunately, studies focusing on the repeatability of radiomic features are limited and have shown conflicting results (14,15).

The primary aim of this study was to assess the repeatability of radiomic features extracted from PET images of cervical cancer patients. The effects of reconstruction, segmentation, voxel resampling, and SUV discretization methods on repeatability were also explored. These images were collected as part of a repeatability study of several commonly used PET/MRI and PET/CT quantitative imaging metrics (16).

## MATERIALS AND METHODS

### Subjects

This prospective study was approved by the Washington University Institutional Review Board, and all volunteers provided written informed consent (ClinicalTrials.gov identifier NCT02717572). From June 2016 through May 2017, 17 patients with histologically proven malignancy were enrolled. Two patients who failed to complete the second imaging session were excluded, and another was excluded because of a lack of tumor  $^{18}\text{F}$ -FDG uptake. For the current study focusing on cervical cancer, 3 patients without cervical cancer were also excluded.

Each patient underwent double baseline  $^{18}\text{F}$ -FDG PET/CT imaging separated by at least 24 h and at most 7 d. Unless otherwise noted, the imaging procedures used in this study conformed with the European

---

Received Apr. 23, 2020; revision accepted Sep. 2, 2020.

For correspondence or reprints contact: Richard L. Wahl, Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 S. Kingshighway Blvd., Campus Box 8131, St. Louis, MO 63110.

E-mail: rwahl@wustl.edu

Published online Oct. 2, 2020.

COPYRIGHT © 2021 by the Society of Nuclear Medicine and Molecular Imaging.

**TABLE 1**  
Radiomic Feature Groups

Feature group	Features (n)	Radiomic feature names
Standard intensity	5	SUV <sub>max</sub> , SUV <sub>mean</sub> , SUV <sub>peak</sub> , SUV <sub>SD</sub> , total-lesion glycolysis
Nonstandard intensity	4	Skewness, kurtosis, entropy, uniformity
Shape	3	Metabolic tumor volume, sphericity, compacity
GLCM	6	Homogeneity, energy, contrast, correlation, entropy, dissimilarity
GLRLM	11	Short- and long-run emphasis, low and high gray-level run emphasis, short-run low and high gray-level emphasis, long-run low and high gray-level emphasis, gray-level nonuniformity, run length nonuniformity, run percentage
NGLDM	3	Coarseness, contrast, busyness
GLZLM	11	Short- and long-zone emphasis, low and high gray-level zone emphasis, short-zone low and high gray-level emphasis, long-zone low and high gray-level emphasis, gray-level nonuniformity, zone length nonuniformity, zone percentage

Association of Nuclear Medicine tumor imaging guidelines (2). The same PET/CT scanner and approximate <sup>18</sup>F-FDG dose were used for both imaging sessions. Patients were instructed to fast and to avoid liquids other than water for at least 6 h before the planned <sup>18</sup>F-FDG administration time. Blood glucose levels were measured immediately before the <sup>18</sup>F-FDG injection.

### Imaging

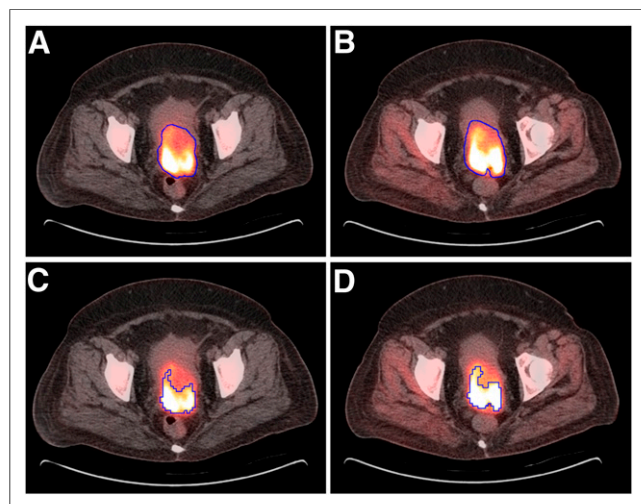
All images were acquired using a Biograph 40 scanner (Siemens AG). Static PET data were collected for 15 min over a single station (tumor centered within a 21.6-cm field of view), starting 60–70 min after intravenous injection of 370 MBq of <sup>18</sup>F-FDG. A CT was performed immediately before PET imaging, using a tube potential of 120 kV, maximum tube current of 80 mAs (CARE Dose 4D [Siemens] tube current modulation), pitch of 0.8, and rotation time of 0.5 s. PET images were reconstructed using ordered-subset expectation maximization (PET<sub>OSEM</sub>) or ordered-subset expectation maximization with point-spread function (PET<sub>PSF</sub>). Image reconstruction parameters, as well as additional acquisition and postacquisition details, can be found in Supplemental Table 1 (supplemental materials are available at <http://jnm.snmjournals.org>).

### Image Analysis

MIM, version 6.9.3 (MIM Software), was used for image segmentation (Fig. 1). For each PET/CT session, the lesion volume of interest (VOI) was manually delineated by 1 expert nuclear medicine reader to generate a whole-tumor contour (VOI<sub>WT</sub>). On the basis of previous work involving segmentation of cervical tumors, a 40% isocontour (VOI<sub>40</sub>) was also generated, containing all voxels with an SUV of at least 40% of the SUV<sub>max</sub> of the VOI<sub>WT</sub> (17). Before radiomic feature calculation, PET image intensities were normalized to decay-corrected injected activity per kilogram of body weight (SUV [g/mL]). The effects of SUV discretization and spatial resampling were explored by discretizing to a fixed number of bins (32, 64, or 128), over the full SUV range in each image, and either no spatial resampling (4.07 × 4.07 × 5.00 mm) or resampling to 3-mm isotropic voxels. Radiomic features were then extracted using LIFEX, version 5.28, which complies with the Imaging Biomarker Standardization Initiative recommendations (18). Intensity, shape, and textural features (Table 1) were calculated from each VOI. Detailed descriptions, formulas, and computation parameters for each feature can be found on the LIFEX website (<https://www.lifexsoft.org>).

### Statistical Methods

Repeatability was assessed using the Lin concordance correlation coefficient (CCC), which provides a quantification of agreement between 2 repeated measures (19). In this study, a CCC value of less than 0.80 was considered unrepeatability, a value of at least 0.80 was considered repeatable, and a value of at least 0.95 was considered highly repeatable. Including the repeatability calculations for all reconstruction, segmentation, voxel resampling, and SUV discretization methods, we generated 24 CCC values for each radiomic feature.

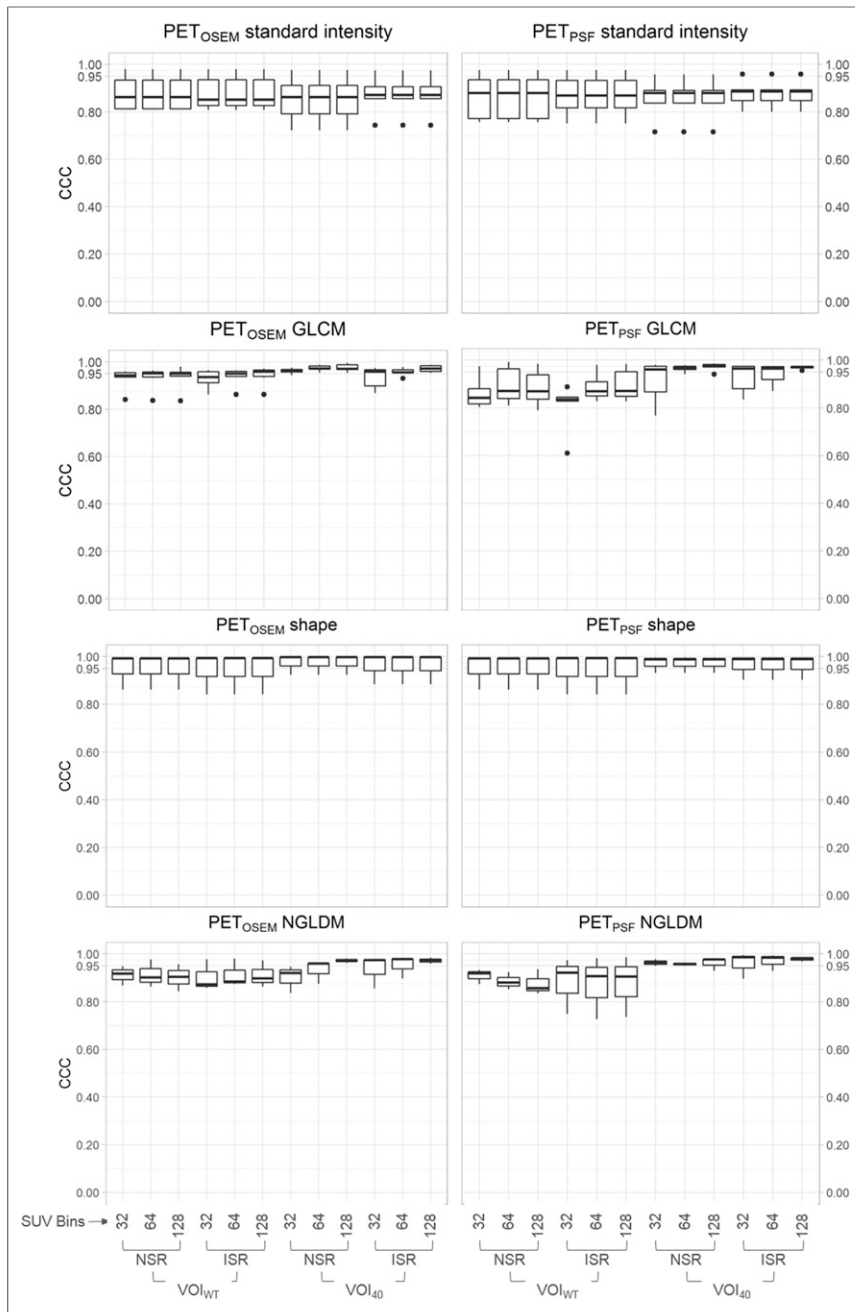


**FIGURE 1.** Representative images. Tumors were manually delineated to generate VOI<sub>WT</sub> (baseline image [A] and repeat image [B]). VOI<sub>40</sub> was created by removing all voxels with SUVs ≤ 40% of SUV<sub>max</sub> of VOI<sub>WT</sub> (baseline image [C] and repeat image [D]).

**TABLE 2**  
Patient Characteristics

Characteristic	Value
Age (y)	46.8 ± 11.0
Height (m)	1.66 ± 0.06
Weight (kg)	77.8 ± 17.3

Qualitative data are numbers followed by percentages in parentheses; continuous data are mean ± SD.



**FIGURE 2.** Repeatability radiomic feature groups. Box plots show repeatability ranges for all features within specified feature group. ISR = isotropic voxels; NSR = no spatial resampling.

Paired *t* tests or Wilcoxon signed-rank tests were used to assess groupwise repeatability differences. Group normality was evaluated using D'Agostino–Pearson tests. Data were analyzed using R, version 3.6.2 (<http://cran.r-project.org/>), and Excel, version 2016 (Microsoft Corp.). A *P* value of less than 0.05 was considered significant, unless otherwise indicated. A Bonferroni adjustment for multiple comparisons was applied when necessary to control for type I errors.

## RESULTS

Seventeen patients were enrolled, and test–retest images from 11 female patients (Table 2) were eligible for radiomic analysis. The

median time between imaging sessions was 2 d, with a range of 1–7 d. Blood glucose levels (mean ± SD) before <sup>18</sup>F-FDG administration were 92.2 ± 8.1 and 93.2 ± 20.9 mg/dL during visits 1 and 2, respectively. The mean administered <sup>18</sup>F-FDG dose was 367.7 ± 20.2 MBq during visit 1 and 371.0 ± 16.8 MBq during visit 2. The mean visit 1 <sup>18</sup>F-FDG uptake time was 60.4 ± 1.4 min, and the mean visit 2 <sup>18</sup>F-FDG uptake time was 61.4 ± 3.1 min.

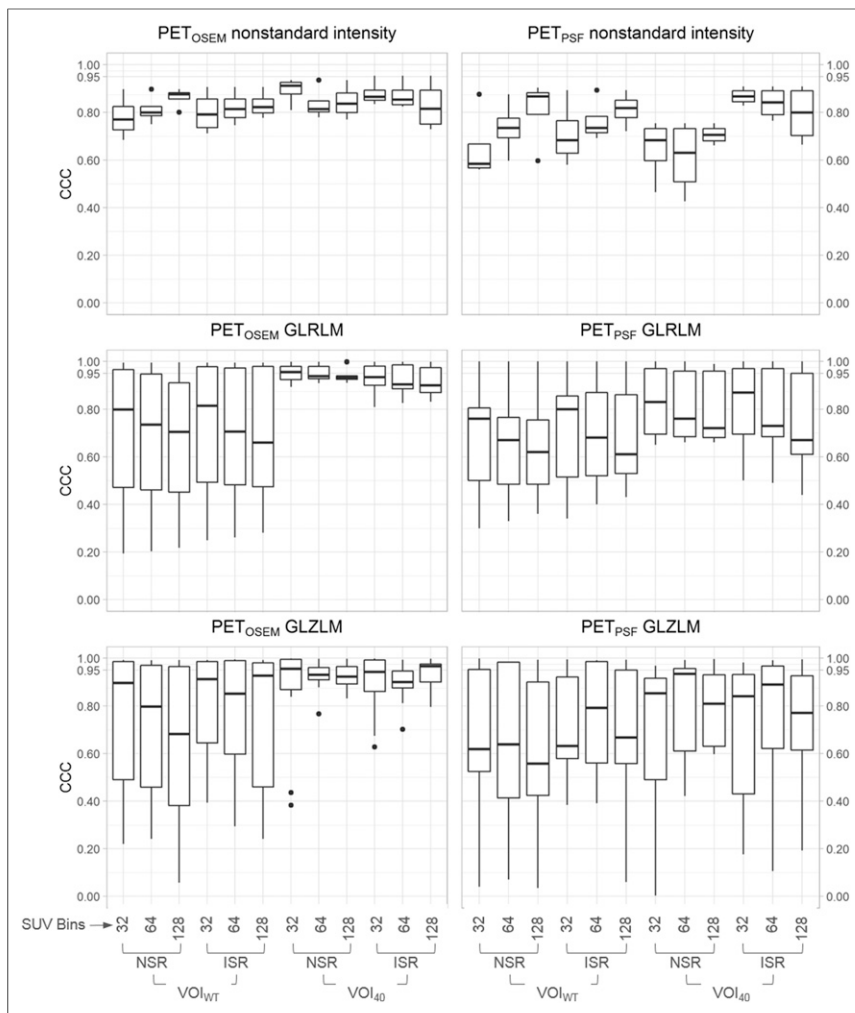
## Radiomic Feature Repeatability

When assessed as groups, standard-intensity, shape, neighborhood gray-level difference matrix (NGLDM), and gray-level cooccurrence matrix (GLCM) features showed consistent repeatability, with mean CCC values greater than 0.80 (Fig. 2). Nonstandard-intensity, gray-level zone length matrix (GLZLM), and gray-level run length matrix (GLRLM) features were mostly unrepeatability (56.7% of CCC values were less than 0.80) when extracted from PET<sub>PSF</sub> images (Fig. 3). Nonstandard-intensity, GLZLM, and GLRLM features were also mostly unrepeatability (55.8% of CCC values were less than 0.80) when extracted from VOI<sub>WT</sub> segmentations.

Highly repeatable radiomic features (mean CCC values above 0.95) included metabolic tumor volume, compacity, entropy (GLCM-calculated), gray-level nonuniformity (GLRLM-calculated), run length nonuniformity, coarseness, long-zone high gray-level emphasis, gray-level nonuniformity (GLZLM-calculated), and zone length nonuniformity (Supplemental Tables 2 and 3). Nearly all (98.6%) GLCM CCC values were greater than 0.80, and many (61.8%) were greater than 0.95. Overall, the combination of PET<sub>OSEM</sub> with isotropic resampling, SUV discretized to 64 bins, and tumors segmented using VOI<sub>40</sub> resulted in the most stable features (76.7% were repeatable and 39.5% highly repeatable). The fewest repeatable radiomic features resulted from PET<sub>PSF</sub> images with no spatial resampling, SUV discretized to 128 bins, and tumors segmented manually (39.5% were repeatable and 20.9% highly repeatable).

## PET Reconstruction Method

The mean CCC values for radiomic features extracted from PET<sub>OSEM</sub> and PET<sub>PSF</sub> images are provided in Table 3. Shape and textural (i.e., GLCM, GLRLM, NGLDM, and GLZLM) features had significantly higher CCC values when extracted from PET<sub>OSEM</sub> images than when extracted from PET<sub>PSF</sub> images (0.86 and 0.79, respectively; *P* < 0.001; Fig. 4A). GLRLM and GLZLM features were impacted by reconstruction method more than were other features (Fig. 5A). Standard-intensity–based features extracted from



**FIGURE 3.** Unrepeatable radiomic feature groups. ISR = isotropic voxels; NSR = no spatial resampling.

both PET reconstructions had approximately the same mean CCC value (0.87;  $P = 0.494$ ; Fig. 2).

### Segmentation Method

Radiomic features were less repeatable when extracted from  $VOI_{WT}$  segmentations than when extracted from  $VOI_{40}$  segmentations (Tables 4 and 5). Using a CCC cutoff of 0.80, 56.2% of  $PET_{PSF}$   $VOI_{WT}$  features and 65.5% of  $VOI_{40}$  features were repeatable. When extracted from  $PET_{OSEM}$  images, 68.6% of  $VOI_{WT}$  features and 89.9% of  $VOI_{40}$  features were repeatable. Shape, GLCM, GLRLM, and NGLDM feature groups extracted from  $PET_{PSF}$  images were found to have significantly (all  $P < 0.015$ ) lower CCC values when extracted from  $VOI_{WT}$  than when extracted from  $VOI_{40}$ . Nonstandard-intensity, GLCM, GLRLM, NGLDM, and GLZLM features extracted from  $PET_{OSEM}$  images were found to have significantly lower (all  $P < 0.023$ ) CCC values when extracted from  $VOI_{WT}$  images than when extracted from  $VOI_{40}$  images.

### Spatial Resampling

The repeatability of most radiomic features was robust against spatial resampling changes, with 78.5% of features showing less than a 5% relative difference in CCC value between resampling

methods (Figs. 4C and 5C). Features extracted from  $PET_{PSF}$  images were more sensitive to spatial resampling changes than were features extracted from  $PET_{OSEM}$  images (Tables 6 and 7). When extracted from  $VOI_{WT}$ , GLRLM features showed greater repeatability after isotropic resampling ( $P < 0.045$ ). The opposite was true when extracted from  $VOI_{40}$ , with GLRLM features showing greater repeatability before spatial resampling ( $P < 0.005$ ). Few radiomic feature groups (21%) were found to have mean CCC values that differed by more than 0.03 before and after spatial resampling.

### SUV Discretization

The repeatability of most radiomic features showed little sensitivity to changes in SUV discretization (Figs. 4D and 5D). Most (60.3%) textural feature CCC values varied by less than 5% among bin number groups. GLCM feature repeatability was largely insensitive to SUV discretization changes, with 80.2% of CCC values varying by less than 5% within SUV bin groups. GLZLM features were considerably impacted by changes in the number of SUV bins, with 69.3% of CCC values varying by 5% or more (and by as much as 228.6%) across SUV bin groups.

### Lesion Volume Analysis

The absolute relative difference between test–retest radiomic values was correlated with mean test–retest metabolic tumor volume to assess the influence of lesion volume on feature repeatability. After controlling for multiple comparisons, we found no significant correlations between metabolic tumor volume and absolute relative difference

in feature value (Supplemental Tables 4 and 5).

## DISCUSSION

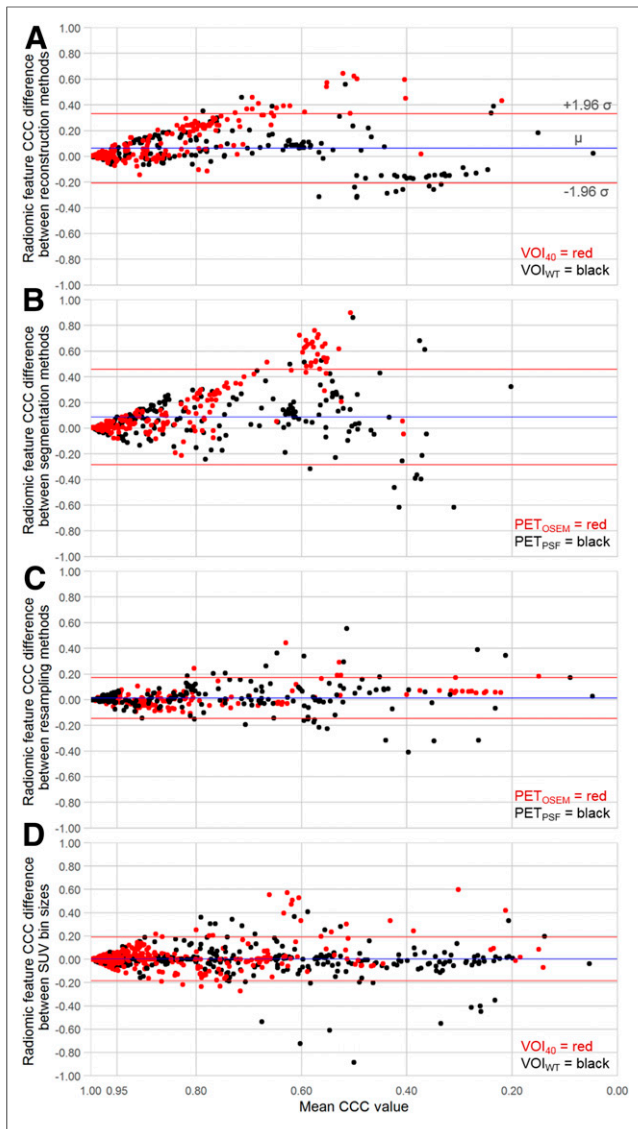
The increasing examination of alternative imaging biomarkers such as radiomic features requires an understanding of their test–retest repeatability. The calculation of these features involves several steps, including image reconstruction, segmentation, and preprocessing, all of which have been shown to impact feature stability (15). In the current study, patients with cervical cancer underwent double baseline  $^{18}F$ -FDG PET/CT studies. PET images were reconstructed using 2 different methods, and tumors were delineated manually and with a semiautomated technique. Radiomic features were then extracted following various image preprocessing methods.

Features calculated using run length and zone length matrices were found to have low repeatability, whereas shape, GLCM, and NGLDM features were consistently repeatable. The repeatability of most GLRLM and GLZLM features was quite sensitive to changes in any step of the radiomic extraction process. Tixier et al., using images of esophageal tumors, similarly concluded that GLCM features were repeatable and that GLZLM features were unrepeatable (20). In another study, GLZLM features also showed poor repeatability

**TABLE 3**  
Comparison of Radiomic Feature Repeatability Between PET Image Reconstruction Methods

Feature group	Feature	Mean PET <sub>PSF</sub> CCC	Mean PET <sub>OSEM</sub> CCC	<i>P</i>
Standard intensity	SUV <sub>mean</sub>	0.82	0.86	<0.001*
	SUV <sub>SD</sub>	0.89	0.83	0.066
	SUV <sub>max</sub>	0.88	0.86	<0.001*
	SUV <sub>peak</sub>	0.78	0.82	<0.001*
	Total-lesion glycolysis	0.97	0.98	<0.001*
Nonstandard intensity	Skewness	0.85	0.92	0.012
	Kurtosis	0.74	0.83	0.003
	Entropy	0.69	0.80	0.040
	Energy	0.70	0.80	0.003
Shape	Metabolic tumor volume	0.99	1.00	0.313
	Sphericity	0.88	0.88	0.016
	Compacity	0.99	1.00	0.313
GLCM	Homogeneity	0.89	0.97	<0.001*
	Energy	0.89	0.94	0.110
	Contrast	0.90	0.95	0.176
	Correlation	0.93	0.91	<0.001*
	Entropy	0.93	0.96	0.266
GLRLM	Dissimilarity	0.91	0.96	0.176
	Short-run emphasis	0.89	0.96	0.007
	Long-run emphasis	0.89	0.97	0.002
	Low gray-level run emphasis	0.52	0.58	0.176
	High gray-level run emphasis	0.62	0.78	<0.001*
	Short-run low gray-level emphasis	0.52	0.59	0.176
	Short-run high gray-level emphasis	0.62	0.78	<0.001*
	Long-run low gray-level emphasis	0.55	0.58	0.569
	Long-run high gray-level emphasis	0.71	0.82	0.002
	Gray-level nonuniformity	0.99	1.00	<0.001*
NGLDM	Run length nonuniformity	0.99	0.99	0.036
	Run percentage	0.89	0.96	0.009
	Coarseness	0.96	0.97	0.493
	Contrast	0.88	0.89	0.522
GLZLM	Busyness	0.93	0.92	0.092
	Short-zone emphasis	0.54	0.74	0.005
	Long-zone emphasis	0.92	0.97	0.036
	Low gray-level zone emphasis	0.43	0.56	0.176
	High gray-level zone emphasis	0.54	0.73	<0.001*
	Short-zone low gray-level emphasis	0.44	0.53	0.447
	Short-zone high gray-level emphasis	0.59	0.77	<0.001*
	Long-zone low gray-level emphasis	0.55	0.77	<0.001*
	Long-zone high gray-level emphasis	0.97	0.98	0.970
	Gray-level nonuniformity	0.98	0.99	0.009
GLZLM	Zone length nonuniformity	0.94	0.98	0.003
	Zone percentage	0.83	0.94	0.007

\*Significant *P* value, based on paired Wilcoxon signed-rank test. After correcting for multiple comparisons, *P* value of less than 0.001 was considered significant.



**FIGURE 4.** Bland–Altman plots showing effect on radiomic feature repeatability of changes in reconstruction (PET<sub>PSF</sub> CCC subtracted from PET<sub>OSEM</sub> CCC [A]), segmentation (VOI<sub>WT</sub> CCC subtracted from VOI<sub>40</sub> CCC [B]), spatial resampling (no-spatial-resampling CCC subtracted from resampling to 3-mm isotropic voxel CCC [C]), and SUV discretization (32-bin CCC subtracted from 64-bin CCC, 32-bin CCC subtracted from 128-bin CCC, and 64-bin CCC subtracted from 128-bin CCC [D]). Plots were generated by calculating mean of, and difference between, corresponding radiomic feature CCC values of each method.

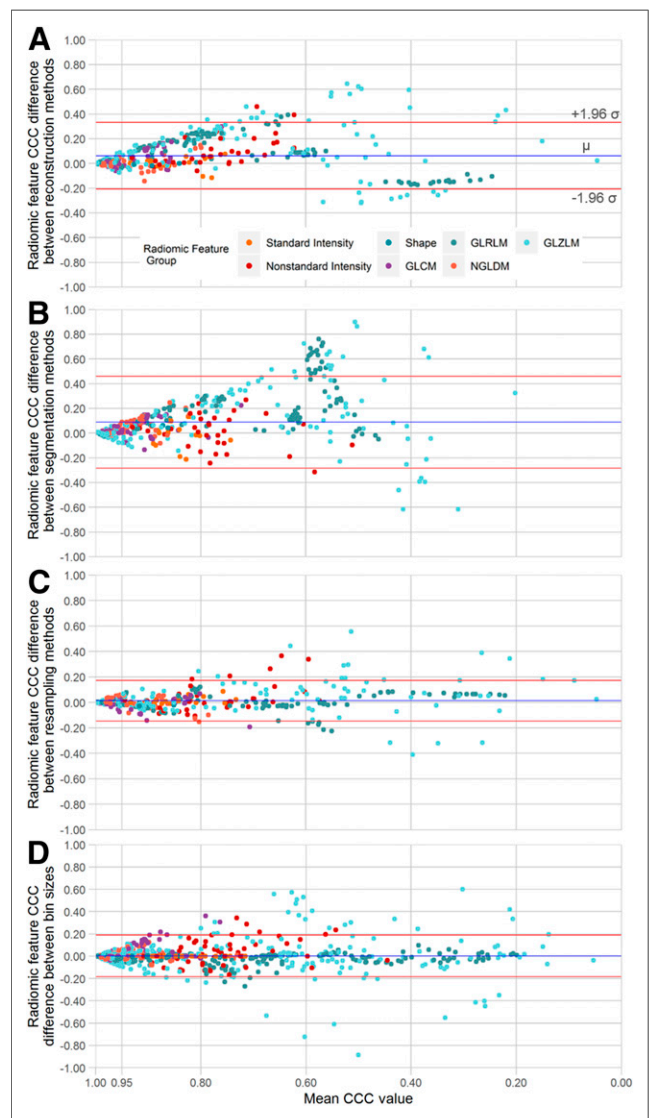
when extracted from PET images of lung cancer using a PET<sub>OSEM</sub>, 50% isocontour segmentation, and isotropic spatial resampling (21). In our study, GLZLM features were mostly unrepeatably, but using a PET<sub>OSEM</sub> and VOI<sub>40</sub>, 89.4% of GLZLM CCC values were designated as repeatable.

Two feature groups, GLRLM and GLZLM, were not homogeneous in terms of repeatability within each group. Gray-level emphasis and gray-level run emphasis features within the GLRLM group showed poor repeatability. However, the remaining GLRLM features (non-uniformity, run percentage, and run-emphasis calculations) were highly repeatable. Likewise, most GLZLM features were unrepeatably and highly sensitive to changes in image reconstruction, segmentation, and

preprocessing methods. However, GLZLM nonuniformity and run percentage features were consistently repeatable.

GLCM features were consistently repeatable in this study—a finding that has also been described in other work (20–22). Entropy, in particular, has consistently been found to be reproducible and repeatable, as well as a significant predictor of patient response (7,23,24). In our study, entropy was consistently repeatable, though somewhat sensitive to SUV discretization. When SUV was discretized to 32 bins, mean entropy CCC was 0.90, but when SUV was discretized to 64 or 128 bins, mean entropy CCC increased to 0.97.

The PET reconstruction methods used here had a substantial impact on repeatability. This impact was intensified by segmentation method, as shown in Figure 4. Features that were otherwise mostly unrepeatably showed mostly high repeatability when PET<sub>OSEM</sub> and VOI<sub>40</sub> were used. Yan et al. found that 5%–56% of textural features showed a large variation between values (920%) when reconstruction settings were varied (25). Reconstruction method was also found by



**FIGURE 5.** Bland–Altman plots showing how changes in reconstruction (A), segmentation (B), spatial resampling (C), and SUV discretization (D) affected repeatability of radiomic feature groups.

**TABLE 4**  
Groupwise Comparison of Repeatability of Radiomic Features Extracted from PET<sub>OSEM</sub> Images Using 2 Different Segmentation Methods

Feature group	Mean VOI <sub>WT</sub> CCC	Mean VOI <sub>40</sub> CCC	<i>P</i>
Standard intensity	0.88	0.86	0.057
Nonstandard intensity	0.82	0.86	0.010
Shape	0.95	0.97	0.140
GLCM	0.93	0.96	<0.001*
GLRLM	0.70	0.94	<0.001*
NGLDM	0.91	0.94	0.023
GLZLM	0.73	0.90	<0.001*

\*Significant *P* value, based on paired *t* test. After correcting for multiple comparisons, *P* value of less than 0.004 was considered significant.

Mean value calculation includes CCC values from both resampling methods and each SUV bin level (i.e., 6 CCC values per feature per segmentation).

Gallivanone et al. to have a strong impact on the stability of radiomic features (26). Using an anthropomorphic phantom, test–retest PET images were acquired and reconstructed using PET<sub>OSEM</sub>, PET<sub>PSF</sub> and PET<sub>PSF</sub> with time of flight. They concluded that fewer than 20% of radiomic features were robust against changes in reconstruction method.

Numerous segmentation methods have been combined with radiomic analyses. One study segmented phantom ROIs using an adaptive Bayesian method or a 60% isocontour and found that less than 20% of radiomic features were stable between these segmentations (26). Using VOI<sub>40</sub> significantly improved the repeatability of most radiomic features in our study, especially when used on images reconstructed without point-spread function. This improvement was particularly dramatic in some features otherwise found to be unrepeatably. The median CCC values of GLRLM and GLZLM features rose above the repeatability cutoff only when using VOI<sub>40</sub>, and all features within both groups were repeatable when also extracted from PET<sub>OSEM</sub> images. Automated or semiautomated segmentation approaches help to prevent the selection of normal tissue near tumor edges but often underestimate volume and may not capture regions of necrosis. Although this study showed a significant improvement

in repeatability with the implementation of an SUV threshold, the utility of textural analyses may suffer when using such methods. Though the threshold applied in this study was based on its previous use in this patient population, other segmentation methods (including fixed or adaptive approaches) may further improve radiomic repeatability and should be evaluated.

Interpolation to isotropic voxel sizes has routinely been used in radiomic analyses (27,28). Since many features are sensitive to changes in voxel size, isotropic resampling affects the radiomic feature value and, therefore, reproducibility (29,30). However, the effect of voxel size resampling on radiomic feature repeatability has not been previously studied, and a standard approach to spatial resampling in the context of radiomic analysis has not been described. In the current study, voxels either were not resampled or were downsampled to 3-mm isotropic voxels. This method requires information inference, whereas upsampling involves information loss. Here, most features showed similar repeatability between resampling methods, though GLRLM and GLZLM features showed low mean CCC values and high differences in CCC value between resampling methods.

**TABLE 5**  
Groupwise Comparison of Repeatability of Radiomic Features Extracted from PET<sub>PSF</sub> Images Using 2 Different Segmentation Methods

Feature group	Mean VOI <sub>WT</sub> CCC	Mean VOI <sub>40</sub> CCC	<i>P</i>
Standard intensity	0.87	0.87	0.984
Nonstandard intensity	0.75	0.74	0.784
Shape	0.95	0.97	0.015
GLCM	0.87	0.95	<0.001*
GLRLM	0.68	0.81	<0.001*
NGLDM	0.88	0.96	<0.001*
GLZLM	0.68	0.73	0.139

\*Significant *P* value, based on paired *t* test. After correcting for multiple comparisons, *P* value of less than 0.004 was considered significant.

Mean value calculation includes CCC values from both resampling methods and each SUV bin level (i.e., 6 CCC values per feature per segmentation).

**TABLE 6**

Groupwise Comparison of Repeatability of Radiomic Features Extracted from PET<sub>OSEM</sub> Images Using 2 Different Spatial Resampling Methods

Feature group	Mean VOI <sub>WT</sub> CCC			Mean VOI <sub>40</sub> CCC		
	NSR	ISR	<i>P</i>	NSR	ISR	<i>P</i>
Standard intensity	0.88	0.88	0.857	0.85	0.87	0.021
Nonstandard intensity	0.82	0.82	0.970	0.86	0.86	0.890
Shape	0.95	0.94	0.090	0.97	0.96	0.250
GLCM	0.93	0.94	0.333	0.97	0.95	0.068
GLRLM	0.69	0.71	0.005	0.95	0.92	0.001*
NGLDM	0.91	0.91	0.928	0.93	0.95	0.031
GLZLM	0.69	0.76	<0.001*	0.90	0.90	0.249

\*Significant *P* value, based on paired Wilcoxon signed-rank test. After correcting for multiple comparisons, *P* value of less than 0.002 was considered significant.

NSR = no spatial resampling; ISR = resampling to 3-mm isotropic voxels.

Mean value calculation includes CCC values from each SUV bin level (i.e., 3 CCC values per feature per resampling method).

Resampling of image intensity values is a crucial step in radiomic analysis. Downsampling from a nearly unlimited set of intensity values to a discrete number of bins effectively reduces image noise and allows for comparison of radiomic values between image datasets. SUV discretization has consistently been shown to have a significant impact on radiomic feature values (7,23). Though the feature value may fluctuate with varying bin numbers or sizes, consistent repeatability of a discretization method may still allow for effective comparison between repeated evaluations. We found most feature groups to be consistently repeatable among SUV bin sizes, with GLCM features performing particularly well and GLZLM features performing quite poorly. On the basis of the current study and other recent radiomic repeatability studies, GLZLM features may not be useful for test–retest assessments (7,31). In this study, we explored varying a fixed number of SUV bins. Another approach is to vary the size of the bins, and future studies should explore the repeatability of this alternative method.

This study was limited by a small sample size, though the number of patients included was similar to other published radiomic repeatability studies (20,21). Future studies are needed to validate if these results represent radiomic features extracted from cervical tumors, generally. When performing this study, we attempted to adhere to the Imaging Biomarker Standardization Initiative guidelines, which only recently became available. Performing additional repeatability studies of other cancer types using the methods reported here, and guided by the Imaging Biomarker Standardization Initiative recommendations, may be informative. Repeatability thresholds used in this study, as in test–retest studies generally, are somewhat arbitrary and intended to illustrate only which features may be sensitive to changes in radiomic and image analysis methods. Additional sources of radiomic feature variability should also be explored, since any of the various factors that impact SUV repeatability could, in principle, affect radiomic feature stability.

**TABLE 7**

Groupwise Comparison of Repeatability of Radiomic Features Extracted from PET<sub>PSF</sub> Images Using 2 Different Spatial Resampling Methods

Feature group	Mean VOI <sub>WT</sub> CCC			Mean VOI <sub>40</sub> CCC		
	NSR	ISR	<i>P</i>	NSR	ISR	<i>P</i>
Standard intensity	0.86	0.87	0.670	0.86	0.88	0.376
Nonstandard intensity	0.73	0.76	0.156	0.65	0.83	<0.001*
Shape	0.95	0.94	0.683	0.97	0.96	0.729
GLCM	0.88	0.86	0.259	0.95	0.95	0.490
GLRLM	0.66	0.70	<0.001*	0.82	0.79	0.045
NGLDM	0.89	0.88	0.655	0.96	0.97	0.345
GLZLM	0.63	0.72	0.001*	0.75	0.71	0.077

\*Significant *P* value, based on paired Wilcoxon signed-rank test. After correcting for multiple comparisons, *P* value of less than 0.002 was considered significant.

NSR = no spatial resampling; ISR = resampling to 3-mm isotropic voxels.

Mean value calculation includes CCC values from each SUV bin level (i.e., 3 CCC values per feature per resampling method).



## CONCLUSION

Certain radiomic feature groups (shape, GLCM, and NGLDM) were repeatable, whereas others (GLRLM and GLZLM) were highly variable. Using a fixed-threshold segmentation method increased the repeatability of most features. Most often, changes in voxel size or SUV discretization parameters resulted in relatively small differences in feature value, though several features were highly sensitive to these changes.

## DISCLOSURE

This project was supported by NIH grants U01CA140204, 5P30CA006973, and 1R01CA190299. No other potential conflict of interest relevant to this article was reported.

### KEY POINTS

**QUESTION:** Are radiomic features extracted from PET images of cervical cancer repeatable?

**PERTINENT FINDINGS:** Using certain image preprocessing techniques, many radiomic features (especially shape, GLCM, and NGLDM) are stable and repeatable. The impact of PET reconstruction and segmentation methods on radiomic feature repeatability can be substantial.

**IMPLICATIONS FOR PATIENT CARE:** Certain radiomic features are repeatable and may be useful in the management of cervical cancer patients.

## REFERENCES

- Hess S, Blomberg BA, Zhu HJ, Hoiland-Carlson PF, Alavi A. The pivotal role of FDG-PET/CT in modern medicine. *Acad Radiol*. 2014;21:232–249.
- Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging—version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328–354.
- Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50(suppl 1):122S–150S.
- Schwartz LH, Litiere S, de Vries E, et al. RECIST 1.1: update and clarification—from the RECIST committee. *Eur J Cancer*. 2016;62:132–137.
- Lodge MA. Repeatability of SUV in oncologic <sup>18</sup>F-FDG PET. *J Nucl Med*. 2017;58:523–532.
- Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–446.
- Altazi BA, Zhang GG, Fernandez DC, et al. 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.
- Koh WJ, Abu-Rustum NR, Bean S, et al. Cervical cancer, version 3.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2019;17:64–84.
- Chen SW, Shen WC, Hsieh TC, et al. Textural features of cervical cancers on FDG-PET/CT associate with survival and local relapse in patients treated with definitive chemoradiotherapy. *Sci Rep*. 2018;8:11859.
- Reuzé S, Orhac F, Chargari C, et al. Prediction of cervical cancer recurrence using textural features extracted from <sup>18</sup>F-FDG PET images acquired with different scanners. *Oncotarget*. 2017;8:43169–43179.
- Lin AJ, Dehdashti F, Grigsby PW. Molecular imaging for radiotherapy planning and response assessment for cervical cancer. *Semin Nucl Med*. 2019;49:493–500.
- Schwarz JK, Grigsby PW, Dehdashti F, Delbeke D. The role of <sup>18</sup>F-FDG PET in assessing therapy response in cancer of the cervix and ovaries. *J Nucl Med*. 2009;50(suppl 1):64S–73S.
- Sarabhai T, Tschischka A, Stebner V, et al. Simultaneous multiparametric PET/MRI for the assessment of therapeutic response to chemotherapy or concurrent chemoradiotherapy of cervical cancer patients: preliminary results. *Clin Imaging*. 2018;49:163–168.
- Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys*. 2018;102:1143–1158.
- Zwanenburg A. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging*. 2019;46:2638–2655.
- Fraum TJ, Fowler KJ, Crandall JP, et al. Measurement repeatability of <sup>18</sup>F-FDG-PET/CT versus <sup>18</sup>F-FDG-PET/MRI in solid tumors of the pelvis. *J Nucl Med*. 2019;60:1080–1086.
- Olsen JR, Esthappan J, DeWees T, et al. Tumor volume and subvolume concordance between FDG-PET/CT and diffusion-weighted MRI for squamous cell carcinoma of the cervix. *J Magn Reson Imaging*. 2013;37:431–434.
- Nioche C, Orhac F, Boughdad S, et al. LIFEX: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res*. 2018;78:4786–4789.
- Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255–268.
- Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in <sup>18</sup>F-FDG PET. *J Nucl Med*. 2012;53:693–700.
- Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. *Acta Oncol*. 2013;52:1391–1397.
- Desseroit MC, Tixier F, Weber WA, et al. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non-small cell lung cancer tumors: a repeatability analysis in a prospective multicenter cohort. *J Nucl Med*. 2017;58:406–411.
- Leijenaar RT, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep*. 2015;5:11075.
- Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline <sup>18</sup>F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med*. 2011;52:369–378.
- Yan J, Chu-Shern JL, Loi HY, et al. Impact of image reconstruction settings on texture features in <sup>18</sup>F-FDG PET. *J Nucl Med*. 2015;56:1667–1673.
- Gallivanone F, Interlenghi M, D'Ambrosio D, Trifiro G, Castiglioni I. Parameters influencing PET imaging features: a phantom study with irregular and heterogeneous synthetic lesions. *Contrast Media Mol Imaging*. 2018;2018:5324517.
- Zwanenburg A, Leger S, Vallières M, Löck S, for the Image Biomarker Standardisation Initiative. Image biomarker standardisation initiative. arXiv.org website. <https://arxiv.org/abs/1612.07003>. Published December 21, 2016. Revised December 17, 2019. Accessed February 12, 2021.
- Larue RT, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br J Radiol*. 2017;90:20160665.
- Bailly C, Bodet-Milin C, Couespel S, et al. Revisiting the robustness of PET-based textural features in the context of multi-centric trials. *PLoS One*. 2016;11:e0159984.
- Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on <sup>18</sup>F-FDG PET radiomic features: multi-scanner phantom and patient studies. *Eur Radiol*. 2017;27:4498–4509.
- Papp L, Rausch I, Grahovac M, Hacker M, Beyer T. Optimized feature extraction for radiomics analysis of <sup>18</sup>F-FDG PET imaging. *J Nucl Med*. 2019;60:864–872.