^{18}F-FDG PET/CT Qualitative and Quantitative Evaluation in Neurofibromatosis Type 1 Patients for Detection of Malignant Transformation: Comparison of Early to Delayed Imaging With and Without Liver Activity Normalization

Alin Chirindel[1,2], Muhammad Chaudhry[1,3], Jaishri O. Blakeley[4], and Richard Wahl[1,5]

[1]Division of Nuclear Medicine, Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, Maryland; [2]Department of Radiation Oncology, University Medical Center, Frieburg, Germany, and German Cancer Research Center (DKFZ), Partner Site Freiburg, Freiburg, Germany; [3]Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia; [4]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland; and [5]Mallinckrodt Institute of Radiology, Washington University in St. Louis School of Medicine, St. Louis, Missouri

^{18}F-FDG PET/CT has shown increased accuracy, compared with morphologic imaging, in differentiating malignant peripheral nerve sheath tumors (MPNSTs) from benign neurofibromas (BNFs) in patients with neurofibromatosis type 1 (NF1). Delayed ^{18}F-FDG PET imaging typically enhances malignant tumor to background. Our goal was to compare the effectiveness of early (1-h) and delayed (4-h) ^{18}F-FDG PET/CT imaging in differentiating MPNSTs from BNFs in patients with NF1, with and without liver activity normalization. **Methods:** NF1 patients presenting new symptoms or enlarging lesions were clinically evaluated with early and delayed ^{18}F-FDG PET/CT imaging. SUL_{max} (maximum standardized uptake value derived for lean body) and $SUL_{max/liver}$ (lesion uptake adjusted to mean liver activity) were obtained for all sites identified with abnormal metabolic activity. Qualitative and quantitative evaluations, including receiver-operating-characteristic (ROC) comparison of early and delayed imaging sessions, were performed. Histopathology and clinical follow-up (1–9 y) were considered as a gold standard. **Results:** Forty-one NF1 patients with early and delayed ^{18}F-FDG PET/CT scans were identified, and 93 lesions were retrospectively analyzed, representing 24 MPNSTs (all histologically confirmed) and 69 BNFs (26 histologically confirmed). Qualitative evaluation on early imaging showed sensitivity, specificity, positive predictive value, and negative predictive value for separating MPNSTs from BNFs of 91%, 84%, 67%, and 96% versus 91%, 81%, 63%, and 96%, respectively, on 4-h delayed imaging. The mean SUL_{max} was significantly higher for MPNSTs than BNFs on both early scans (6.5 vs. 2.0, $P < 0.01$) and delayed imaging (8.3 vs. 2.3, $P < 0.02$). However, SUL_{max} overlap between benign and malignant lesions persisted even after normalization to mean liver activity. ROC-derived best SUL_{max} cutoffs were 3.2 on early (area under the curve, 0.973) and 4.1 on delayed scans (area under the curve, 0.978). ROC analysis for $SUL_{max/liver}$ improved test specificity (94% vs. 87%, $P < 0.05$) on early and (93% vs. 88%, $P < 0.05$) on delayed imaging. **Conclusion:** Qualitative interpretation of ^{18}F-FDG PET/CT discriminates MPNSTs from BNFs in NF1 patients with similar accuracy on both early and delayed imaging. Quantitative data showed better sensitivity on delayed acquisition and best test specificity with lesion SUL_{max} normalization to liver activity, more so than with delayed imaging at 4 h.

Key Words: ^{18}F-FDG PET/CT; early and delayed; NF1; quantitative and qualitative

J Nucl Med 2015; 56:379–385
DOI: 10.2967/jnumed.114.142372

Neurofibromatosis type 1 (NF1) is one of the most common autosomal-dominant Mendelian diseases, with a worldwide estimated prevalence of 1 in 3,000 (*1*). The NF1 gene (cr17q11.2) codes a protein called neurofibromin, which is part of the p21-ras oncogene family (*2*). Clinical features in patients with NF1 typically include café-au-lait spots, cutaneous and plexiform neurofibromas (benign neurofibromas [BNFs]), optic glioma, Lisch nodules, and bone dysplasia (*3*). Individuals with plexiform neurofibromas harbor an increased risk of transformation into malignant peripheral nerve sheath tumors (MPNSTs), with a relative lifetime risk of 8%–13% (*4,5*).

Although of profound prognostic and therapeutic consequences, distinguishing between benign and malignant lesions proves to be difficult. Both benign and malignant lesions (especially in a synchronous context) have similar clinical manifestations such as changes in consistency or size, unremitting pain, or new neurologic findings (*6*). Morphologic imaging cannot reliably differentiate benign from malignant transformed lesions, especially in tumors with significant heterogeneity (*7*). Accurate histologic evaluation is often challenging due to tumor sampling error, leading to extensive and possibly repeated surgical interventions (*8,9*).

^{18}F-FDG PET/CT metabolic imaging has been shown to be able to detect soft-tissue sarcomas, with positive correlation between tumor intensity uptake and histologic grade (*10*). However, when standard qualitative and quantitative ^{18}F-FDG PET was applied in NF1, mixed success was noticed because of both false-negative (FN) and false-positive (FP) identification of MPNST (*11–14*). Modifications to PET acquisition and postimaging analysis have been attempted to improve test performance (Table 1), particularly by adding delayed PET imaging to the acquisition protocol and by

Received Sep. 2, 2014; revision accepted Nov. 21, 2014.
For correspondence or reprints contact: Richard L. Wahl, Mallinckrodt Institute of Radiology, Washington University in St. Louis School of Medicine, 510 Kingshighway Blvd., St. Louis, MO 63110.
E-mail: wahlr@mir.wustl.edu
Published online Feb. 5, 2015.

TABLE 1
Qualitative and Quantitative Performance of ^{18}F-FDG PET in NF1

Study	NF1 patient	Lesion (MPNST by pathology)	Qualitative performance		Mean SUV_{max}		Quantitative performance		
			Sensitivity	Specificity	BNF	MPNST	SUV cutoff	Sensitivity	Specificity
One-time-point PET									
Ferner et al. (*21*)	18	23 (7)	100	87	1.54*	5.4*	2.5	N/A	N/A
Cardona et al. (*22*)	13	25 (10)	100	75	1.1*	2.6*	1.8	100	83
Brenner et al. (*24*)	16	N/A (16)	N/A	N/A	N/A	5.7	N/A	N/A	N/A
Bensaid et al. (*25*)	38	49 (6)	N/A	N/A	N/A	N/A	Lesion/liver > 1.5	100	86
Bredella et al. (*12*)	45	50 (18)	95	72	1.5	8.5	N/A	N/A	N/A
Ferner et al. (*11*)	105	114 (29)	89	95	1.5	5.7	2.5–3.5	N/A	N/A
Benz et al. (*26*)	34†	40 (17)	N/A	N/A	3.4	12	6.1‡	94	91
Tsai et al. (*23*)	20	26 (10)	N/A	N/A	7.6	2.5	4	100	94
Salamon et al. (*15*)	49	152 (18)	N/A	N/A	8.6	2.6	3.5‡	100	80
							Lesion/liver > 2.6	100	90
Combemale et al. (*13*)	113	145 (40)	N/A	N/A	8.6	2.6	1.5‡	97	76
							Lesion/liver > 1.5	97	76
Dual-time-point PET									
Warbey et al. (*14*)	62	85 (21)	97	87	Early, 2§	Early, 7§	Early, 2.35‡,§	Early, 100§	Early, 60§
					Delayed, 1.9‖	Delayed, 8.1‖	Delayed, 3.1‡,‖	Delayed, 100‖	Delayed, 77‖

*Median.
†21 sporadic disease.
‡From ROC.
§Average, 108 min.
‖Average, 252 min.
N/A = not applicable.

normalizing lesion standardized uptake value (SUV) to normal tissue activity (*11,14,15*). The rationale for dual-time PET imaging is that the activity within benign lesions reportedly typically plateaus after 30 min whereas malignant tumors have rising SUVs over approximately 4 h, allowing for a better separation (*16*). However, this approach has residual FP and FN rates, it is resource-intensive, and it exposes the patients to additional radiation from CT.

Qualitative PET evaluation in patients with NF1 has demonstrated overall good sensitivity (89%–100%) and specificity (72%–95%) in differentiating BNF from MPNSTs (*17–19*), although an explicit set of criteria for accurate visual interpretation has never been validated. Quantitatively, several SUV cutoffs have been proposed to best detect and separate malignant from benign lesions in NF1 patients (Table 1). The wide range for these quantitative uptake thresholds (1.5–6.1) may be due to differences in acquisition protocols (different imaging time points, partial-volume effects), scanner performance, and analysis methods (*16,20*). Such variability limits the use of ^{18}F-FDG PET in clinical practice.

At Johns Hopkins University, our practice has been to acquire whole-body early (1-h after injection) and delayed (4-h from injection) imaging when possible for all NF1 patients. In this study, we assess the utility of visual and quantitative criteria derived from early and delayed whole-body ^{18}F-FDG PET/CT scans to discriminate BNFs from MPNSTs.

MATERIALS AND METHODS

Patient Population

This retrospective study of clinically acquired PET scans was approved by the institutional review board committee. The radiology database was queried with the key words "NF1, neurofibromatosis or neurofibroma" from January 2003 until August 2013. Forty-one NF1 patients were identified with early (1-h) and delayed (4-h) PET scans and appropriate clinical data (Fig. 1). Seventy-four early and delayed PET sessions were evaluated, representing 41 baseline and 33 follow-up studies (18 patients with 1 follow-up, 6 with 2, and 1 with 3) with 93 lesions (24 MPNSTs and 69 BNFs) included (Table 2).

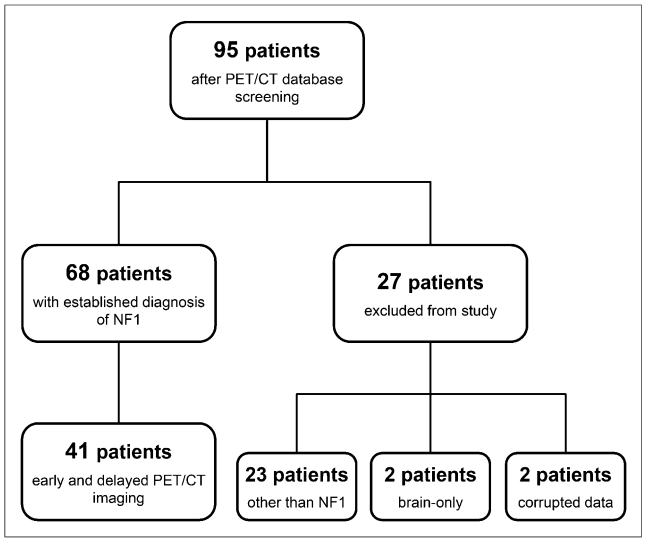

FIGURE 1. Flow diagram showing process of patients' identification.

^{18}F-FDG PET/CT Acquisition

After at least a 4-h fast, the patients received an intravenous injection of ^{18}F-FDG according to a weight-based formula (for adults, 1.3 × 7.4 MBq/kg, and for children, 7.4 MBq/kg), with mean injected activity of 566.1 ± 181.3 MBq (range, 111–925 MBq). All ^{18}F-FDG PET/CT scans were acquired with a vertex-to-toes protocol and were obtained on a Discovery Rx-VCT (GE Healthcare) lutetium oxyorthosilicate crystal, 64-slice scanner in 3-dimensional acquisition mode and 4.15 min per bed position. Reconstruction was performed using an ordered-subset expectation maximization algorithm, with a 128 × 128 matrix, 21 subsets, 2 iterations, 3-mm postreconstruction gaussian filter, standard Z filter, 4.7-mm pixels, and 3.27-mm slice thickness. PET data were reconstructed with and without CT-based attenuation correction and decay-corrected.

The mean baseline serum glucose was 93.1 mg/dL (±13.1) with mean uptake time of 65.3 min (±10.8) for the early and 248.3 min (±22.3) for the delayed scans.

Qualitative Lesion Identification and Evaluation

Early ^{18}F-FDG PET/CT scans were evaluated by 1 board-certified nuclear medicine physician with 18 mo of additional clinical PET/CT fellowship training. Sites of abnormally increased metabolic activity were qualitatively dichotomized as either suspected malignant or benign. Interpretation criteria for malignant lesions were as follows:

- Intensity rule = sites of abnormal metabolic activity associated with morphologic lesions and demonstrating significantly more intense ^{18}F-FDG uptake relative to liver activity.
- Anatomic rule = sites of metabolic activity, satisfying the intensity rule, without obvious morphologic correlation, which were identified at concerning locations (musculature, nerve root/plexus) or appeared asymmetric, compared with the contralateral side.

All other sites of abnormal metabolic activity (with or without morphologic correlation) that did not satisfy the previously explained qualitative rules were interpreted as benign lesions.

The qualitative evaluation for the delayed scans was performed independently from the early qualitative analysis so that all the lesions could again be dichotomized as benign or suspected malignant.

Lesion Inclusion Criteria

On baseline early and delayed ^{18}F-FDG PET/CT scans, lesions accepted for analysis were as follows: all sites fulfilling the qualitative criteria for a suspected malignant lesion; all additional sites with clinical or prior imaging suspicion for malignancy; and up to 5 lesions per scan, which were qualitatively evaluated as benign lesions (if >5 benign lesions were identified, the most ^{18}F-FDG–avid or qualitatively concerning lesions were included).

Additionally included on the follow-up (7–29 mo) early and delayed PET/CT scans were new or increasingly suspicious lesions on follow-up clinical/morphologic examination and new sites of metabolic activity satisfying the qualitative criteria for suspected malignant lesions.

TABLE 2
Patient, PET/CT Scan, and Lesion Characteristics

Patient characteristic	*n*
No. of patients	41 (14 men and 27 women)
Median age (y)	36 (range, 8–77)
Median clinical follow-up (y)	3 (range, 1–9)
PET/CT (early and delayed scans)	74
Baseline sessions	41
Follow-up sessions	33 (7–29 mo)
Lesions evaluated	93
BNF	69 (26 on pathology)
MPNST	24 (24 on pathology)

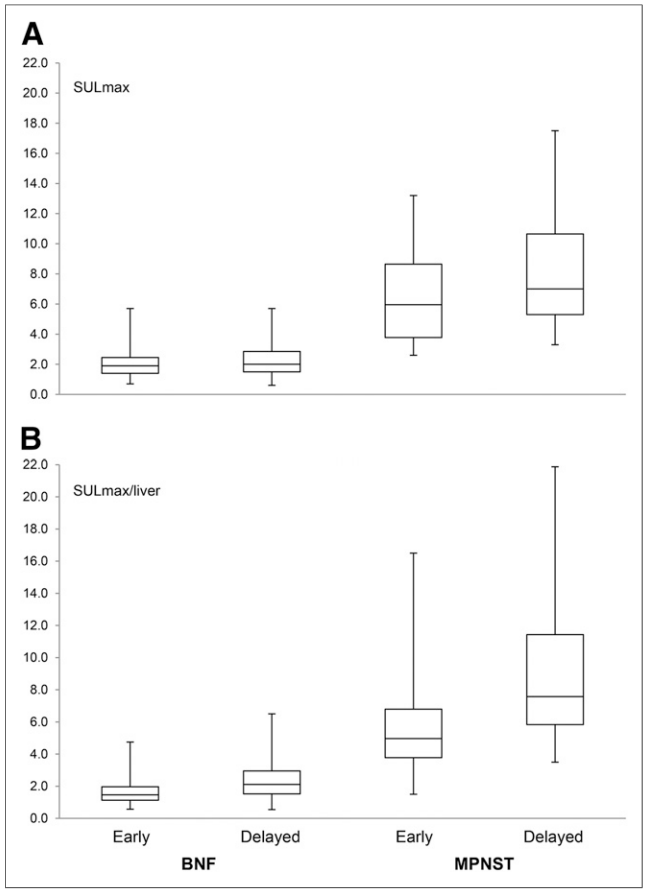

FIGURE 2. Whisker plots (median value with first and third quartiles) for BNF and MPNST SUL_{max} on early and delayed PET imaging. (A) Unadjusted. (B) Liver activity–adjusted SUL_{max}.

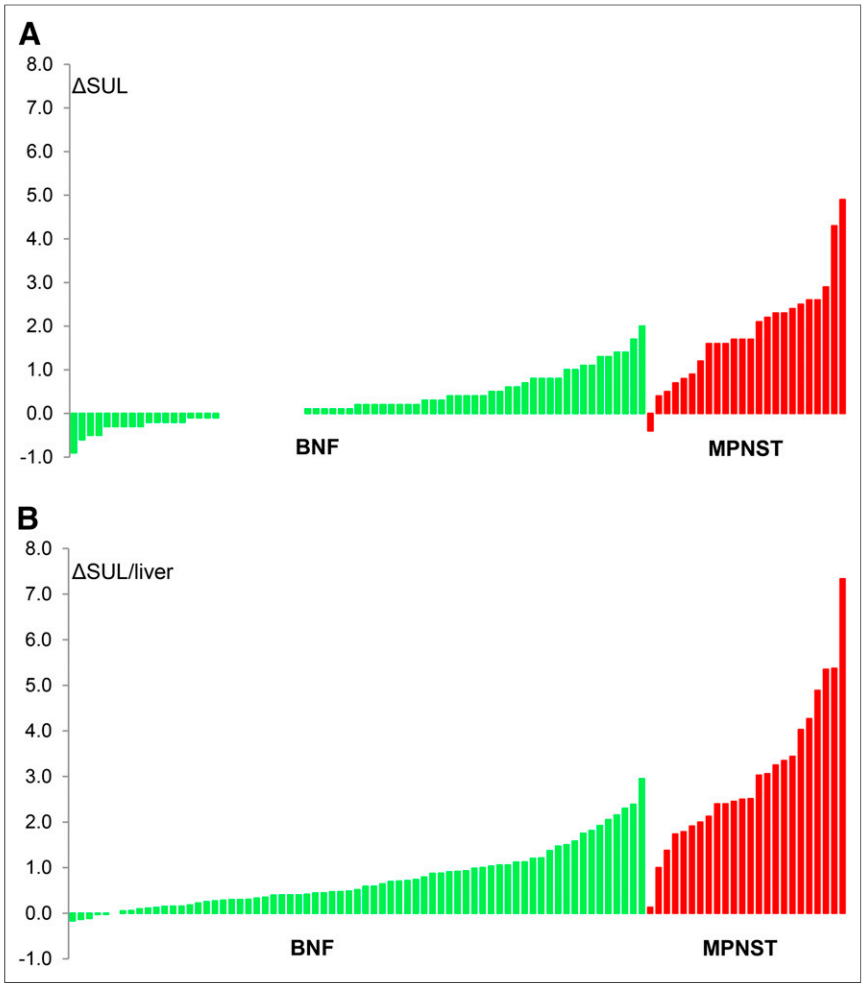

FIGURE 3. Waterfall plots showing absolute differences between early and delayed SUL_{max} of BNFs (green bars) and MPNSTs (red bars) representing raw data (A) and liver-adjusted data (B). Upward shift in all ΔSUL can be noticed with liver normalization, although significant overlap between ΔSUL for BNFs and MPNSTs persists.

The duplicate lesions from the initial scan that were not of clinical/imaging concern and that appeared stable on follow-up PET/CT scans were not included in the analysis.

Quantitative Analysis

SUVs for lean body mass (SUL_{max}) were measured on an Advanced Workstation (software 4.6; GE Healthcare) by placing a volume of interest on the axial PET images with CT cross-reference to ensure correct lesion localization. We tried to minimize differences in measuring lesion activity by starting with a predefined volume of interest. When necessary, manual adjustments were applied to accommodate lesions' extensions and to avoid potentially contaminating intense activity within adjacent normal tissue. The SUL_{max} represented lesion highest SUL and was measured on both the early and the delayed scans at the equivalent image level (*20*).

Mean liver uptake (calculated for each PET session) represented the average activity within a 30-mm spheric volume of interest placed in the right middle lobe of the liver. Normalization of lesion SUL_{max} was performed to the corresponding PET session liver activity (lesion SUL_{max}/liver SUL_{mean} [lesion uptake adjusted to mean liver activity]).

Statistical Analysis

Histopathology from biopsy or surgery and clinical follow-up (median, 3 y; range, 1–9 y) obtained from the pathology department and chart review was regarded as the gold standard. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated using standard formulae. Descriptive statistics for lesion SUL_{max} were performed using the SPSS software package (version 20.0; IBM).

Nonparametric Mann–Whitney–Wilcoxon tests as well as receiver-operating-characteristic (ROC) curves were applied for comparative analyses, with a significance level of 0.05.

RESULTS

Early and Delayed PET/CT Scans

Qualitative Analysis. There were 93 sites identified with abnormal metabolic activity, of which 89 were recorded from the 41 baseline and 4 from follow-up scans (Table 2).

Qualitative evaluation of the early imaging reported 60 sites as benign lesions and 33 as suspected malignant. Subsequent analysis of delayed ^{18}F-FDG PET/CTs showed 58 benign and 35 suspected malignant; hence, 2 of 60 benign interpretations from early evaluation were interpreted as suspected malignant on late evaluation. Final histopathology and clinical follow-up revealed 69 BNFs and 24 MPNSTs, with surgical excision or biopsy available for 26 of BNF and 24 MPNST. We acknowledge that histopathology data are not available for all benign lesions because patients are not routinely referred for surgery if indices suggest a benign lesion. However, all patients were carefully evaluated, and none of the lesions showed changes compatible with malignant transformation during clinical or imaging follow-up (median, 3 y; range, 1–9 y).

The correlation with pathology and clinical follow-up revealed 11 FP and 2 FN interpretations on early ^{18}F-FDG PET/CTs and 13 FPs and 2 FNs on delayed scans. The 2 discordant sites between early and delayed imaging proved to be FP interpretations on the delayed evaluation.

The sensitivity, specificity, positive predictive value, and negative predictive value for detection of MPNST versus BNF on early images were 91% (95% confidence interval [CI], 73–99), 84% (95% CI, 73–92), 67% (95% CI, 48–82), and 96% (95% CI, 88–99), respectively. On delayed images, the sensitivity and negative predictive value were similar, with a slightly decreased specificity of 81% (95% CI, 70–90) and positive predictive value of 63% (95% CI, 45–79).

Quantitative Analysis. The average early SUL_{max} for the lesions identified on histology or follow-up as BNF was 2.0 (±0.9), and for those confirmed as malignant it was 6.5 (±2.9). On the delayed imaging, the average BNF SUL_{max} was 2.3 (±1.2), and average MPNST SUL_{max} was 8.3 (±3.8). Nonparametric Mann–Whitney *U* tests comparing the distribution and ranking of the SUL_{max} showed

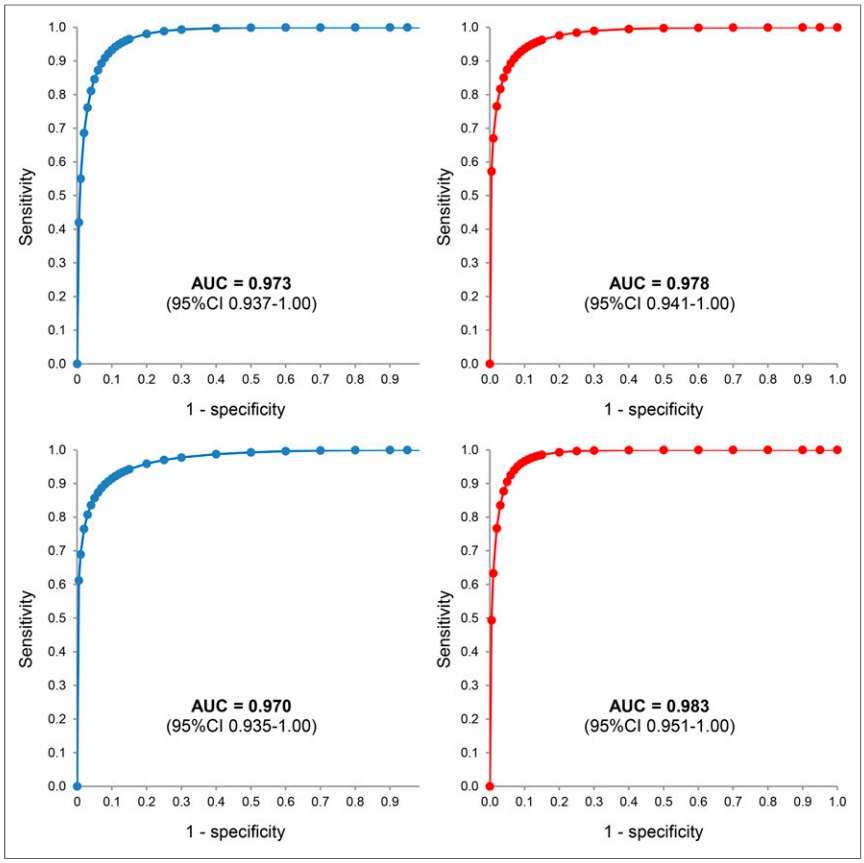

FIGURE 4. ROC diagrams for early (blue curves) and delayed (red curves) scans for unadjusted (top) and liver activity–adjusted (bottom) lesion SUL_{max}.

a statistically significant difference between BNF versus MPNST on both early ($P < 0.01$) and delayed ($P < 0.02$) PET scans.

However, no statistical difference was noticed between the early and delayed SUL_{max} for either BNFs (Z of −1.13, $P = 0.26$) or MPNSTs (Z of −1.8, $P = 0.07$). Whisker plots of SUL_{max} for benign and malignant lesions from both the early and the delayed imaging are displayed in Figure 2.

The mean value of BNF SUL_{max} change (delayed scans minus early scans) was 0.3 (±0.7), with a mean increase of 14% (range, −41% to +110%). For MPNST lesions, the average SUL_{max} change was 1.9 (±1.2) corresponding to a mean 30% increase (range, −6% to +70%). There was statistical significance for both absolute and percentage change between the SUL_{max} of BNF and MPNST ($P < 0.05$).

When the BNF SUL_{max} were adjusted to liver activity (lesion SUL_{max}/liver SUL_{mean}), the mean absolute change became 0.7 (±0.7), and the mean percentage change was 46% (±38). The mean liver-adjusted MPNST SUL_{max} absolute change was 3.0 (±1.6), and the mean percentage change was 61% (±37). Statistical significance was reached only for absolute change ($P < 0.05$) but not for the percentage increase ($P > 0.09$) between the liver-adjusted SUL_{max} for BNF and MPNST.

Histograms of absolute SUL_{max} change in BNF and MPNST lesions between delayed and early images are presented in Figure 3.

ROC analysis of lesion SUL_{max} for the early PET/CT scans revealed an area under the curve (AUC) of 0.973 (95% CI, 0.937–1.00, $P < 0.05$), best SUL_{max} cutoff of 3.2 (for 92% sensitivity and 87% specificity), and maximal sensitivity threshold of 2.5 (77% specificity). For the delayed imaging, the AUC was 0.978 (95% CI, 0.947–1.00, $P < 0.05$), with a best cutoff of 4.1 (96% sensitivity and 88% specificity) and a 100% sensitivity threshold of 3.3 (81% specificity) (Fig. 4).

When lesion SUL_{max} were normalized to liver activity, the AUC for early ROC evaluation was 0.970, with a best SUL_{max} of 2.7 (92% sensitivity and 94% specificity), and delayed AUC was 0.983, with a best SUL_{max} cutoff of 4.3 (96% sensitivity and 93% specificity). When lesion/liver uptake was examined for 100% sensitivity, the cutoffs were 1.5 (51% specificity) on early scans and 3.5 (87% specificity) on late scans.

ROC analysis for the absolute and percentage change of unadjusted lesion SUL_{max} showed an AUC of 0.899 and 0.742, respectively. For liver-normalized absolute and percentage change, the AUC was 0.921 and 0.638, respectively.

DISCUSSION

^{18}F-FDG PET has been reported to be a good, but imperfect, test for distinguishing benign from malignant tumors. Efforts to improve the diagnostic performance of the test have included serial acquisitions and quantitative analyses beyond simple qualitative assessments. Our retrospective study evaluated the added benefit of a late acquisition protocol and qualitative and quantitative interpretation of ^{18}F-FDG PET imaging to differentiate MPNST from BNF in patients with NF1 (*14*).

Visual evaluation of metabolic activity within lesions was performed in the context of regional background/liver uptake, and it was adjusted toward the final radiologic interpretation according to the location and morphologic features from corresponding CT because this is consistent with clinical application of this technique. We achieved reasonable sensitivity and specificity (91% and 84%, respectively) for the early visual assessment (1 h), with similar sensitivity and slightly decreased specificity (80%) for the delayed visual evaluation (4 h). These results are consistent with prior reported performance values (*11,12,14*).

One intrinsic advantage of the qualitative approach is that no specific imaging process or patient information is required before attempting a successful interpretation. This method does not rely on lesion SUV measurement or fixed thresholding, which may vary substantially among scanners, reconstruction protocols, display/analysis software, and radiotracer uptake times (*16,20*). However, we observed that qualitative evaluation on delayed images did not improve PET diagnostic accuracy versus early images.

In fact, the slight drop in the performance of qualitative assessment on late imaging was driven by 2 FP interpretations, which were confirmed with histopathology (Fig. 5). The FP interpretation highlights a caveat of visual interpretation on delayed images: the increasing contrast between lesion ^{18}F-FDG uptake and decreasing physiologic liver and soft-tissue background activity. Therefore, stable BNF metabolic activity can result in a more pronounced subjective interpretation, with wrong classification of benign lesions and hence decreased

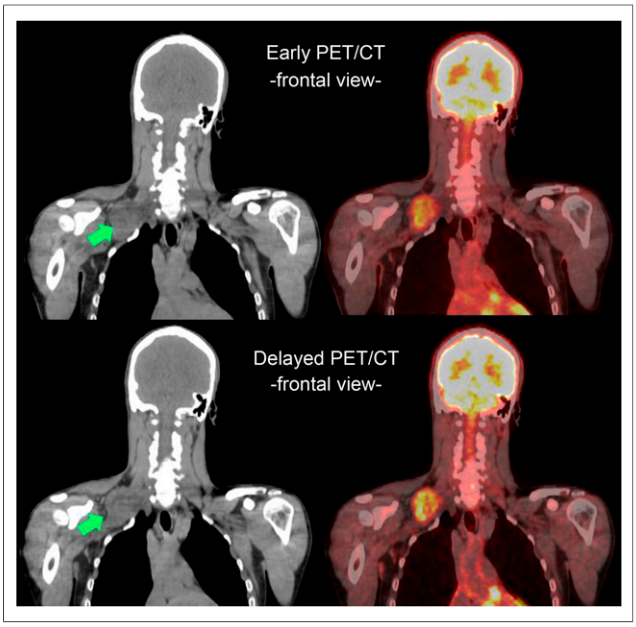

FIGURE 5. Early PET/CT imaging (upper) shows heterogeneously intense ^{18}F-FDG uptake (early SUL_{max} = 3.2) in brachial plexus lesion (green arrows), which demonstrates increased intensity (delayed SUL_{max} = 5.0) on delayed PET/CT imaging (lower). Histology from surgical excision revealed benign plexiform neurofibroma and patient remained clinically asymptomatic on follow-up (>12 mo).

specificity. We acknowledge that a multiinterpreter qualitative analysis may yield results different from current single-interpreter evaluation, although strictly predefined qualitative interpretation criteria should limit the extent of divergent interpretation.

In the quantitative evaluation, we used the SUL_{max} (derived from lean body mass) and not the more common maximum SUV (SUV_{max}; derived from patient weight) as a more reliable measure of tissue activity to account for variation of individual body habitus (*20*). There was a significant difference between mean SUL_{max} for benign and malignant lesions on both early ($P < 0.01$) and delayed ($P < 0.02$) imaging, further supporting the hypothesis that PET can be used as an imaging discriminator for BNF and MPNST in NF1. The mean SUL_{max} for BNFs and MPNSTs were 2.0 (±1.0) and 6.5 (±2.9) on the early PET and 2.3 (±1.2) and 8.3 (±3.8) on the delayed PET scans, respectively. Further work should evaluate lesion activity in the context of lesions' size and deep versus superficial location.

Time-dependent analysis of benign lesion activity showed an unexpected pattern with increasing SUL_{max} from early to delayed imaging for 59% (41/69) of all benign sites. In fact, the absolute measure was more than 1 unit in 8 cases (7 with pathologic correlate), and percentage ΔSUL_{max} was above 30% in 16 BNF cases (12 with histologic proof). After normalization to liver activity, even more benign sites had interval-increased SUL_{max} on delayed scans: 19 sites with absolute $\Delta SUL_{max/liver}$ greater than 1 unit and 44 sites with percentage $\Delta SUL_{max/liver}$ greater than 30%. It is important to note that prior studies have described a rather universal pattern of decreasing ^{18}F-FDG uptake in benign NF1 lesions (*14,21*). This observation of increased SUL_{max} on late imaging even in pathologically confirmed benign lesions challenges the hypothesis that late acquisition successfully overcomes the limitations in specificity of early ^{18}F-FDG PET.

For the malignant sites, our analysis showed increasing ^{18}F-FDG uptake from early to delayed scans in all but 1 lesion (percentage ΔSUL_{max}, −6%), which was, in fact, identified on pathology as high-grade malignancy. Therefore, universal interpretation of decreasing/stable ^{18}F-FDG uptake as BNF and increasing uptake as malignant could lead to FN and FP in quantitative PET evaluation (Fig. 6). Absolute and percentage change ROC analysis showed best performance in differentiating malignant versus benign for the absolute change liver-adjusted SUL_{max} (AUC, 0.921). However, absolute and percentage change were inferior to test performance directly from raw or liver-adjusted BNF and MPNST SUL_{max} (AUC, 0.970–0.983).

Computed best joint-operating cutoff points were 3.2 and 4.1 for early and delayed ROC curves, respectively. These are slightly higher than previously reported values (3.1 and 3.5, respectively) by Warbey et al. (*22*), even after accounting for different weight-based formulas used to compute SUL_{max} and SUV_{max} (*23*). Nevertheless, review of the literature (Table 1) demonstrates a wide range of best SUV_{max} cut points (1.5–6.1), suggesting that an interinstitutional standardization is advisable before any multicenter cooperation.

Two recent studies have proposed that normalization of lesion ^{18}F-FDG uptake to liver activity could improve PET performance in NF1 patients. Salamon et al. (*15*) reported increased test accuracy (100% sensitivity and 90% specificity) when a threshold of lesion/liver greater than 2.6 was used, whereas Combemale et al. (*13*) reported best results (97% sensitivity and 76% specificity) for a threshold of lesion/liver greater than 1.5.

When applied to our data, neither of these 2 normalized cutoff values appeared to improve ^{18}F-FDG PET performance in discriminating BNF from malignant lesions: the first suggested threshold missed too many MPNSTs (90% sensitivity) whereas the second generated unnecessary surgical interventions

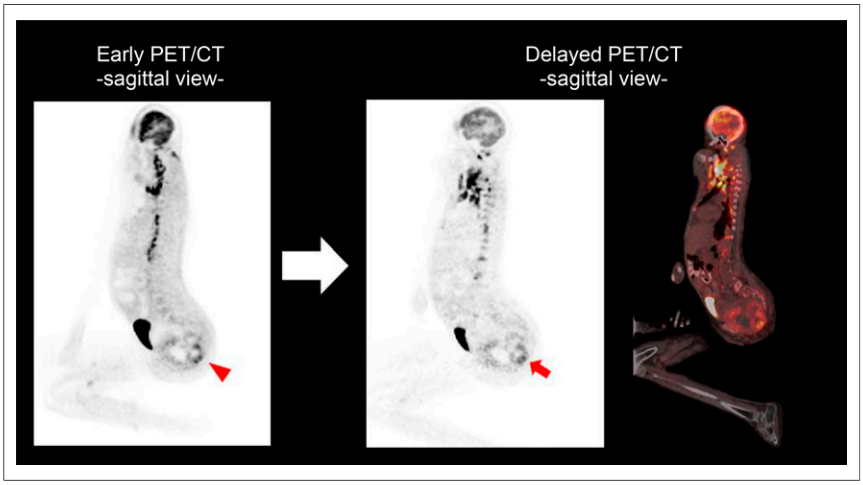

FIGURE 6. FN interpretation with quantitative evaluation. Early PET/CT imaging shows highly heterogeneous metabolic activity (early SUL_{max} = 4.3) within pelvis (red arrowhead), which persists albeit with slightly decreased intensity (delayed SUL_{max} = 3.7) on delayed imaging (red arrow). Initial fine-needle aspiration was inconclusive, and complete surgical excision showed MPNST. There are extensive bone deformities in this NF1 patient.

(51% specificity). In our study, the best liver-normalized cutoff was 2.7 on early imaging (92% sensitivity and 94% specificity) and 4.3 on delayed imaging (96% sensitivity and 93% specificity). Similar to SUV_{max} cut points, lesion/liver thresholds will need to be selected and standardized at each institution and for each trial based on the emphasis desired on sensitivity versus specificity. The liver normalization has considerable potential advantages over the absolute determination of lesion activity, because determining relative values is easier than determining absolute radiotracer uptake.

In summary, we found in 41 NF1 patients with early and delayed acquisition of ^{18}F-FDG PET and pathologic confirmation of diagnosis that for qualitative evaluation the addition of a delayed acquisition protocol did not substantially improve test accuracy in differentiating BNF versus MPNST. Further, direct comparison of lesion SUL_{max} from early to delayed sessions was not helpful because of unpredictable and confounding increasing activity in more than 50% of BNF. ROC analysis did improve sensitivity over qualitative assessment, and liver normalization had an incremental benefit for test specificity. These effects were similar for both early and delayed acquisitions.

Our data suggest that there is not significant additive information from late acquisition to off-set the patient burden of extended periods of fasting, additional radiation exposure, or the institutional resource demands. On the basis of these results, we would suggest qualitative and quantitative assessment of ^{18}F-FDG PET at 1 h, with the understanding that applying this single imaging strategy alone will yield rare FP and FN results and therefore require multidisciplinary collaboration for interpretation of the results in the setting of each clinical scenario. The calibration for liver activity thresholds and SUV_{max} cut points should be chosen on the basis of center-specific validated data and according to clinical priorities (i.e., sensitivity vs. specificity). Similarly, this means that if ^{18}F-FDG PET is to be applied to multicenter clinical studies, intercenter calibrations and intracenter reliability are required before ^{18}F-FDG PET could be used as an endpoint across sites and time points.

CONCLUSION

The qualitative interpretation of standard ^{18}F-FDG PET images (at 60 min) provides good clinical utility for distinguishing BNF from MPNST (91% sensitivity and 84% specificity) in NF1 patients.

Quantitative data provided better sensitivity on delayed imaging, yet the highest specificity was achieved with lesion SUL_{max} normalization to liver activity, more so than delayed acquisition.

Quantitation must be interpreted in the context of center-specific ROC analysis. Multiinstitutional standardization is advised for setting a meaningful best SUV/SUL cutoff for future therapeutic clinical trials in which ^{18}F-FDG PET is used as an endpoint.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENT

We thank Julia Buchanan at Johns Hopkins for her helpful review and advice in the preparation of this manuscript.

REFERENCES

1. Friedman JM. Epidemiology of neurofibromatosis type 1. *Am J Med Genet.* 1999;89:1–6.
2. Rasmussen SA, Friedman JM. NF1 gene and neurofibromatosis 1. *Am J Epidemiol.* 2000;151:33–40.
3. Ferner RE, Huson SM, Thomas N, et al. Guidelines for the diagnosis and management of individuals with neurofibromatosis 1. *J Med Genet.* 2007;44:81–88.
4. Nguyen R, Jett K, Harris GJ, Cai W, Friedman JM, Mautner VF. Benign whole body tumor volume is a risk factor for malignant peripheral nerve sheath tumors in neurofibromatosis type 1. *J Neurooncol.* 2014;116:307–313.
5. Evans DG, Baser ME, McGaughran J, Sharif S, Howard E, Moran A. Malignant peripheral nerve sheath tumours in neurofibromatosis 1. *J Med Genet.* 2002;39:311–314.
6. Ferner RE, Gutmann DH. International consensus statement on malignant peripheral nerve sheath tumours in neurofibromatosis 1. *Cancer Res.* 2002;62:1573–1577.
7. Wasa J, Nishida Y, Tsukushi S, et al. MRI features in the differentiation of malignant peripheral nerve sheath tumors and neurofibromas. *AJR.* 2010;194:1568–1574.
8. Woodruff JM. Pathology of tumors of the peripheral nerve sheath in type 1 neurofibromatosis. *Am J Med Genet.* 1999;89:23–30.
9. Spurlock G, Knight SJ, Thomas N, Kiehl TR, Guha A, Upadhyaya M. Molecular evolution of a neurofibroma to malignant peripheral nerve sheath tumor (MPNST) in an NF1 patient: correlation between histopathological, clinical and molecular findings. *J Cancer Res Clin Oncol.* 2010;136:1869–1880.
10. Benz MR, Tchekmedyian N, Eilber FC, Federman N, Czernin J, Tap WD. Utilization of positron emission tomography in the management of patients with sarcoma. *Curr Opin Oncol.* 2009;21:345–351.
11. Ferner RE, Golding JF, Smith M, et al. [^{18}F]2-fluoro-2-deoxy-D-glucose positron emission tomography (FDG PET) as a diagnostic tool for neurofibromatosis 1 (NF1) associated malignant peripheral nerve sheath tumours (MPNSTs): a long-term clinical study. *Ann Oncol.* 2008;19:390–394.
12. Bredella MA, Torriani M, Hornicek F, et al. Value of PET in the assessment of patients with neurofibromatosis type I. *AJR.* 2007;189:928–935.
13. Combemale P, Valeyrie-Allanore L, Giammarile F, et al. Utility of ^{18}F-FDG PET with a semi-quantitative index in the detection of sarcomatous transformation in patients with neurofibromatosis type 1. *PLoS ONE.* 2014;9:e85954.
14. Warbey VS, Ferner RE, Dunn JT, Calonje E, O'Doherty MJ. FDG PET/CT in the diagnosis of malignant peripheral nerve sheath tumours in neurofibromatosis type-1. *Eur J Nucl Med Mol Imaging.* 2009;36:751–757.
15. Salamon J, Veldhoen S, Apostolova I, et al. ^{18}F-FDG PET/CT for detection of malignant peripheral nerve sheath tumours in neurofibromatosis type 1: tumour-to-liver ratio is superior to an SUVmax cut-off. *Eur Radiol.* 2014;24:405–412.
16. Lodge MA, Lucas JD, Marsden PK, Cronin BF, O'Doherty MJ, Smith MA. A PET study of 18FDG uptake in soft tissue masses. *Eur J Nucl Med.* 1999;26:22–30.
17. Fisher MJ, Basu S, Dombi E, et al. The role of [^{18}F]-fluorodeoxyglucose positron emission tomography in predicting plexiform neurofibroma progression. *J Neurooncol.* 2008;87:165–171.
18. Karabatsou K, Kiehl TR, Wilson DM, Hendler A, Guha A. Potential role of 18fluorodeoxyglucose-positron emission tomography/computed tomography in differentiating benign neurofibroma from malignant peripheral nerve sheath tumor associated with neurofibromatosis 1. *Neurosurgery.* 2009;65:A160–A170.
19. Moharir M, London K, Howman-Giles R, North K. Utility of positron emission tomography for tumour surveillance in children with neurofibromatosis type 1. *Eur J Nucl Med Mol Imaging.* 2010;37:1309–1317.
20. 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;50:122S–150S.
21. Ferner RE, Lucas JD, O'Doherty MJ, et al. Evaluation of 18fluorodeoxyglucose positron emission tomography (18FDG PET) in the detection of malignant peripheral nerve sheath tumours arising from within plexiform neurofibromas in neurofibromatosis 1. *J Neurol Neurosurg Psychiatry.* 2000;68:353–357.
22. Cardona S, Schwarzbach M, Hinz U, et al. Evaluation of F18-deoxyglucose positron emission tomography (FDG-PET) to assess the nature of neurogenic tumours. *Eur J Surg Oncol.* 2003;29:536–541.
23. Tsai LL, Drubach L, Fahey F, Irons M, Voss S, Ullrich NJ. [^{18}F]-fluorodeoxyglucose positron emission tomography in children with neurofibromatosis type1 and plexiform neurofibromas: correlation with malignant transformation. *J Neurooncol.* 2012;108:469–475.
24. Brenner W, Friedrich RE, Gawad KA, et al. Prognostic relevance of FDG PET in patients with neurofibromatosis type-1 and malignant peripheral nerve sheath tumours. *Eur J Nucl Med Mol Imaging.* 2006;33:428–432.
25. Bensaid B, Giammarile F, Mognetti T, et al. Utility of 18 FDG positon emission tomography in detection of sarcomatous transformation in neurofibromatosis type 1. *Ann Dermatol Venereol.* 2007;134:735–741.
26. Benz MR, Czernin J, Dry SM, et al. Quantitative F18-fluorodeoxyglucose positron emission tomography accurately characterizes peripheral nerve sheath tumors as malignant or benign. *Cancer.* 2010;116:451–458.