CrossMark

ORIGINAL ARTICLE

Assessing the role of ^{18}F-FDG PET and ^{18}F-FDG PET/CT in the diagnosis of soft tissue musculoskeletal malignancies: a systematic review and meta-analysis

Elba C. Etchebehere [1,2] · Brian P. Hobbs [3] · Denái R. Milton [3] · Osama Malawi [4] ·
Shreyaskumar Patel [5] · Robert S. Benjamin [5] · Homer A. Macapinlac [1]

Abstract

Purpose Twelve years ago a meta-analysis evaluated the diagnostic performance of ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in assessing musculoskeletal soft tissue lesions (MsSTL). Currently, PET/CT has substituted PET imaging; however, there has not been any published meta-analysis on the use of PET/CT or a comparison of PET/CT with PET in the diagnosis of MsSTL. Therefore, we conducted a meta-analysis to identify the current diagnostic performance of ^{18}F-FDG PET/CT and determine if there is added value when compared to PET.

Methods A systematic review of English articles was conducted, and MEDLINE PubMed, the Cochrane Library, and Embase were searched from 1996 to March 2015. Studies exploring the diagnostic accuracy of ^{18}F-FDG PET/CT (or dedicated PET) compared to histopathology in patients with MsSTL undergoing investigation for malignancy were included.

Results Our meta-analysis included 14 articles composed of 755 patients with 757 soft tissue lesions. There were 451 (60 %) malignant tumors and 306 benign lesions. The ^{18}F-FDG PET/CT (and dedicated PET) mean sensitivity, specificity, accuracy, positive predictive value, and negative predictive value for diagnosing MsSTL were 0.96 (0.90, 1.00), 0.77 (0.67, 0.86), 0.88 (0.85, 0.91), 0.86 (0.78, 0.94), and 0.91 (0.83, 0.99), respectively. The posterior mean (95 % highest posterior density interval) for the AUC was 0.92 (0.88, 0.96). PET/CT had higher specificity, accuracy, and positive predictive value when compared to a dedicated PET (0.85, 0.89, and 0.91 vs 0.71, 0.85, and 0.82, respectively).

Conclusion ^{18}F-FDG PET/CT and dedicated PET are both highly accurate in the diagnosis of MsSTL. PET/CT is more accurate and specific and has a higher positive predictive value than PET.

Keywords Meta-analysis · ^{18}F-FDG · PET/CT · Sarcoma · Musculoskeletal · Soft tissue tumors

✉ Elba C. Etchebehere
Ecetchebehere@mdanderson.org; elba.etchebehere@gmail.com

1   Department of Nuclear Medicine, The University of Texas MD Anderson Cancer Center, 1400 Pressler, FCT 16.6005, Unit 1483, Houston, TX 77030, USA

2   Department of Nuclear Medicine, Sirio Libanes Hospital, São Paulo, Brazil

3   Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

4   Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

5   Department of Sarcoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Introduction

The incidence of soft tissue sarcomas has been steadily rising; however, they still represent less than 1 % of newly diagnosed cancers in adults and 6 % of childhood malignancies [1]. Consequently, imaging studies evaluating soft tissue tumors have a limited number of soft tissue sarcomas reported and thus the differentiation from a benign soft tissue lesion may be challenging.

Many studies have demonstrated that ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET)/CT (or dedicated PET) is able to adequately diagnose soft tissue tumors

separating benign from malignant lesions as well as staging patients with soft tissue sarcomas. Furthermore, ^{18}F-FDG PET/CT has an added value in staging, grading, and evaluating response to therapy, thus improving patient management.

An extensive meta-analysis was conducted 12 years ago evaluating the diagnostic performance of ^{18}F-FDG PET scanners in the assessment of soft tissue lesions. The authors found a good sensitivity and specificity in discriminating benign from malignant soft tissue lesions [87 and 79 %, respectively, with standardized uptake value (SUV) cutoff of 2.0 and 70 and 87 %, respectively, with SUV cutoff of 3.0] [2]. Currently, however, ^{18}F-FDG PET/CT is widely performed and has substituted dedicated PET imaging due to the perception of improved sensitivity and specificity in sarcomas [3–5]. It has been purported that PET/CT would enhance diagnostic performance for identifying sarcomas, but heretofore the relative effectiveness of discriminating benign from malignant lesions with these modalities has yet to be established.

Therefore, we undertook a systematic review of the available studies to identify the current diagnostic performance of the combined use of ^{18}F-FDG PET/CT and dedicated PET in terms of sensitivity, specificity, and accuracy. We also undertook a subanalysis to determine if the use of PET/CT has an added value when compared with a dedicated PET in diagnosing musculoskeletal soft tissue tumors through meta-analysis.

Materials and methods

Information sources

A systematic review of English articles was conducted, and MEDLINE PubMed, Cochrane Library, and Embase were searched from 1996 to March 2015. Search terms included positron emission tomography (MeSH), soft tissue sarcoma (MeSH), 18f fluorodeoxyglucose (MeSH), humans (MeSH), and musculoskeletal tumors (all fields).

Study selection

Studies investigating the diagnostic accuracy of ^{18}F-FDG PET or ^{18}F-FDG PET/CT in patients with musculoskeletal soft tissue lesions undergoing investigation for malignancy were eligible for inclusion. Studies with at least ten patients with musculoskeletal soft tissue lesions were selected. Histopathology had to be the reference standard in ≥50 % of the patients studied in order to be included. Among those without histopathology, the follow-up period had to be ≥6 months. All studies had to present true-positive (TP), false-negative (FN), true-negative (TN), and false-positive (FP) results at the patient level. Lastly, the types of soft tissue lesions had to be explicitly reported in order to qualify for analysis.

Studies reporting SUV values were included as well as those with quantification methods that are not used in the daily clinical practice such as total lesion glycolysis, metabolic tumor volume, and ^{18}F-FDG kinetic analyses.

Exclusion criteria

Review articles, editorials or letters, comments, conference proceedings, case reports, errata, preclinical studies, duplicate publications, and animal studies were excluded from this review. Also excluded were the terms PET/MRI, Ewing's sarcoma, and osteosarcoma. Subsequently, only studies with sufficient data to reassess sensitivity and specificity of ^{18}F-FDG PET/CT (or dedicated PET) were included in the meta-analysis.

We also excluded studies evaluating only tumor grading and prognosis or tumors submitted to prior chemotherapy or surgery [6, 7]. Additionally, some studies were excluded because there was no clear separation of patients with musculoskeletal and abdominal lesions, the interpretation of PET images to distinguish malignant from benign lesions was unclear, and some studies addressed only the role of PET after therapy and recurrence.

Data collection process

The titles and abstracts were reviewed according to the established selection criteria and rejected when ineligible. The remaining publications were submitted to full-text review to determine eligibility. When necessary a consensus was obtained for equivocal articles.

All data were recorded at the lesion level. The relevant articles were obtained based on the following criteria: image used to assess lesions was either ^{18}F-FDG PET/CT (or dedicated PET), the analysis for lesion differentiation was quantitative (as opposed to visual), the total number of soft tissue lesions was presented, reference standard was histopathology in ≥50 % of the patients, and the TP, FN, TN, and FP lesions could be determined. Additionally, we documented the first author's name, the year of publication, age, sex, total number of lesions (soft tissue and bone), types of lesions, the reader level summary, and when reported, the SUV classification threshold that yielded the best performance for discriminating malignant from benign lesions (based on body weight in the majority of the studies). When more than one cutoff value was determined, only the value that led to highest sensitivity was used in the analysis.

The quality of the studies (including risk of bias), was assessed using the revised tool for the quality assessment of diagnostic accuracy studies included in systematic reviews 2 (QUADAS-2) [8]. We classified the risk of bias and concerns regarding patient selection, index test (^{18}F-FDG PET/CT and dedicated PET studies), reference standard (mainly

histopathology), and flow and timing (time lapsed from index test to reference standard test). Studies with a low risk of bias were classified as "L," unclear or intermediate risk as "U," and high risk as "H."

Statistical analysis

A binormal Bayesian receiver-operating characteristic (ROC) model [9] was used to integrate the available data to jointly estimate sensitivity, specificity, and infer the resulting inter-study mean ROC curve in meta-analysis. The model assumes that classification arises from unobserved Gaussian random variables. The distribution of latent traits underlying conspicuity for malignancy was assumed to be $N(a,b)$, while traits for benign lesions were assumed to generate from a standard normal distribution $N(0,1)$. Location parameters presumed zero-centered Gaussian priors with precision of 0.10. Gamma priors were used for inverse variance hyperparameters with shape = 0.05 and rate = 20. A Bayesian logistic random effects model was used to evaluate positive and negative predictive values in meta-analysis. The model assumes that the log odds of a positive (negative) test being a TP (TN) result is exchangeable across studies with Gaussian distribution and random hierarchical mean and standard deviation. For both analyses, the hierarchical mean assumed a Gaussian prior with mean fixed at the sample estimated log odds and variance = 2. The inter-study standard deviation was assumed uniform(0, 10). Analyses were undertaken separately for each modality as well as in combination. The numbers of observed positive (malignant) and negative (benign) lesions (for each study as well as overall) were combined with posterior distributions for sensitivity and specificity to infer accuracy. Diagnostic performance was inferred from the area under the ROC curve (AUC) as well as partial AUC restricted to the domain yielding at least 80 % specificity. The resultant posterior mean and 95 % highest posterior density (HPD) intervals are reported for each summary. For one study, which failed to report TP classifications, the missing value for sensitivity was imputed via sampling from its posterior predictive distribution assuming the presence of two malignant lesions. Bayesian computation used Markov chain Monte Carlo (v3.2.3, OpenBUGS Foundation).

An optimal classification threshold was inferred using the posterior distributions of specificity and sensitivity for each of the 12 studies for which the "best" classification threshold for SUV was ascertained. The approach considered 1-specificity and sensitivity as x and y Cartesian coordinates of a binary classification system where perfect discrimination is represented by x=0, y=1. For each study, the reported threshold was paired with the length of the line segment connecting the point (1-posterior mean specificity, posterior mean sensitivity) to (0, 1) to characterize the Euclidean distance from perfect discrimination. An optimal threshold was selected using

weighted piecewise cubic spline interpolation as the numerical value that minimized this distance. The influence of each threshold-length pair was weighted in proportion to the number of lesions evaluated in the study. The approach makes the implicit assumption of equal loss among FN and FP errors.

Results

Eligible studies

We initially identified 316 studies and after removing duplicates, 191 studies were screened (Fig. 1). Case reports, reviews, and systematic reviews were then excluded. Thus, 83 full-text articles were assessed for eligibility. Another 39 articles were removed because of one of the following reasons: less than ten patients reported with musculoskeletal soft tissue lesions, non-^{18}F-FDG tracers, and the use of non-PET or PET/CT equipment, redundant data, and unclear separation of soft tissue from skeletal lesions. Consequently, 44 studies were eligible for the qualitative analysis. After excluding studies that lacked information of TP, TN, FP, and FN results and studies not pertaining to diagnosis of soft tissue lesions, the meta-analysis was conducted on the remaining 14 studies.

Among the 14 articles included in the meta-analysis, ^{18}F-FDG PET/CT was used in 4 (29 %) of them to assess the tumors, while ^{18}F-FDG PET was employed in the remaining 10 (71 %).

A cutoff from the SUV was used to categorize lesions as malignant or benign in 90 % of the studies. In all but two studies [10, 11] only one lesion per patient was assessed. Thus, lesions were considered to be independent in the meta-analysis.

Diagnoses of soft tissue lesions

A total of 947 patients with suspicion of soft tissue and bone lesions were studied. After removing the patients that presented bone tumors, 755 patients with soft tissue lesions remained. In three studies the patients had more than one soft tissue lesion evaluated by histopathology [10–12]. In the other studies the patients had either one soft tissue lesion [13–20] or the information regarding TP, TN, FP, or FN was not available for all patients [21–23]. Therefore, a total of 757 lesions with histology information were evaluated in 755 patients (Table 1). Although in some studies the patients had more than one lesion [10], in other studies [12] the patients were imaged with multiple radiotracers and the number of ^{18}F-FDG images was less than the total number of patients in the study. The mean age was 49.2 years and 48 % were women. The mean number of PET/CT (or dedicated PET) readers (reader level summary) was 2.0.

Fig. 1 Flowchart of study selection
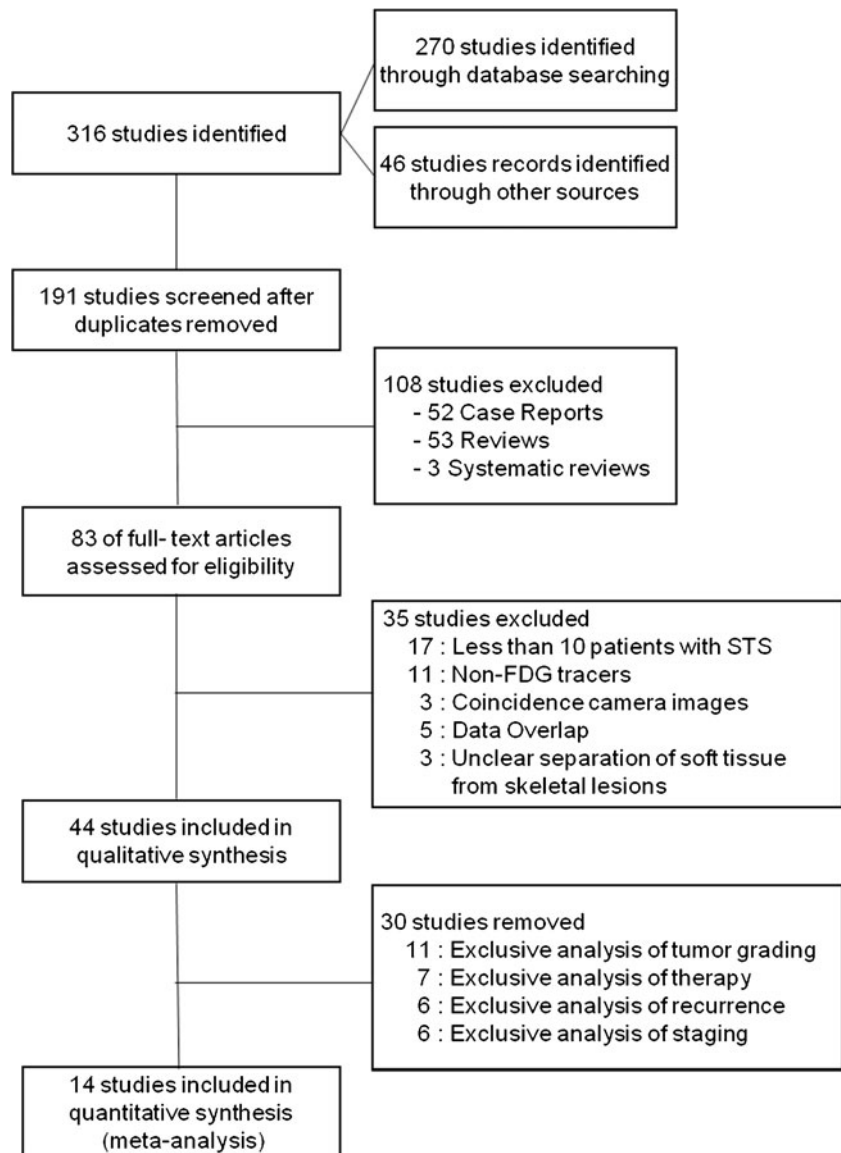
Among all 755 patients studied with 757 lesions, there were 451 (60 %) histologically confirmed malignant tumors and 306 benign lesions (40 %). All tumor types and frequencies are listed in Table 2. The main types of malignant soft tissue sarcomas were classified according to WHO criteria and relate to the tissue of origin [24]. Other unexpected malignant lesions detected were lymphoma ($n=5$), melanoma ($n=5$), squamous cell carcinoma ($n=1$), and metastases ($n=2$). Other benign lesions detected were sarcoidosis ($n=1$), bursa ($n=1$), and cysts ($n=1$).

Methodological quality assessment

The risk of bias was unclear for patient selection in 4 studies [10, 14, 16, 18], for the index test in 2 studies [11, 18], for reference test in all 14 studies, and for flow and timing in 4

studies [11, 14, 16, 18]. The risk of bias was high for flow and timing in one study [10]. The applicability of the included studies was adequate and all classified as "low."

SUV values were calculated with body weight in 86 % of the studies and were unclear in two studies [13, 19]. The mean time of image acquisition post-radiotracer injection was 60 min (40–120 min). The dose delivered ranged from 120 to 740 MBq. In five studies the average dose delivered instead of the dose range was described [10, 11, 14, 15, 23]. All acquisition parameters as well as the optimal cutoff SUV values (related to the TP, TN, FP, and FN results obtained at the lesion level) for each study are listed in TABLE 1S.

Histopathology was used as the standard of reference in 86 % of the studies. In three studies histopathology was performed in only 56 % [10], 97 % [16], and in 91 % of the patients [11]; the remaining patients were clinically followed

Table 1 Characteristics of patients included in the meta-analysis

First author (year)	Age		Sex		RS	D	NT	NSTL	% HP
	Min.	Max.	F	M					
Nieweg [21]	18	82	9	13	2	P	22	22	100
Schulte [13]	1	89	47	55	2	P	102	102	100
Lodge [14]	18	76	15	14	1	P	29	29	100
Lucas [15]	6	85	13	17	1	P	30	31	100
Ferner [20]	12	62	10	8	1	R	18	18	56
Watanabe [16]	12	77	29	26	1	U	55	35	97
Dimitrakopoulou-Strauss [21]	NA	NA	NA	NA	1	P	56	56	100
Ahmed [12]	24	73	10	12	2	P	22	25	100
Suzuki [18]	27	87	29	28	4	R	57	44	100
Shin [11]	6	79	44	39	2	R	83	43	91
Charest [19]	2	90	102	110	2	R	212	160	100
Okazumi [20]	NA	NA	NA	NA	1	P	117	46	100
Amini [22]	8	89	42	58	4	R	100	100	100
Leal [23]	17	88	21	23	3	P	44	44	100
Mean	13	81	NA	NA	1.9	NA	NA	NA	95
Total	NA	NA	371	403	NA	NA	947	755	NA

RS reader level summary, *D* study design, *NT* total number of patients in the study, *NSTL* total number of patients in the study with soft tissue lesions; *% HP* percentage of patients with histopathology as reference test, *Min.* minimum, *Max.* maximum, *F* female, *M* male, *P* prospective, *R* retrospective, *NA* data not available, *U* unclear

up for at least 6 months, since sarcoma growth may vary from the very aggressive that may grow quickly to the low-grade tumors that are slow growing.

Performance of PET/CT and PET for diagnosis of soft tissue lesions

Table 3 presents an overview of the 14 studies incorporated in the meta-analysis, including the number of TP, FN, TN, and FP, where appropriate. Specificity was computed for all 14 studies, while sensitivity was computed for 13 of the 14 studies (the number of TP and FN was not provided in one study [12]). In addition, Figs. 2 and 3 display the posterior mean ROC curve as well as the observed sensitivity and 1-specificity results for each study.

The posterior distributions of AUC, sensitivity, specificity, accuracy, positive predictive value, and negative predictive value for ^{18}F-FDG PET and ^{18}F-FDG PET/CT are summarized (combined and separately) in Table 4. For the combined modalities, the posterior mean (95 % HPD interval) for AUC was 0.92 (0.88, 0.96) and for sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were 0.96 (0.90, 1.00), 0.77 (0.68, 0.86), 0.88 (0.85, 0.91), 0.86 (0.77, 0.94), and 0.91 (0.83, 0.99), respectively. A trend toward enhanced performance for PET/CT when compared to a dedicated PET was present, but failed to attain statistical significance. When assessing the two modalities separately,

AUC, specificity, accuracy, and positive predictive values were higher for ^{18}F-FDG PET/CT compared with ^{18}F-FDG PET (Table 4 and Fig. 4). In addition, the partial AUC, which is restricted to the domain of the ROC curve yielding at least 80 % specificity, was also higher for ^{18}F-FDG PET/CT [0.17 (0.15, 0.19)] compared with ^{18}F-FDG PET [0.10 (0.04, 0.16)].

Lastly, the SUV thresholds used in the 12 studies that had evaluable SUV values were assessed to determine the optimal value that yielded enhanced sensitivity and specificity. Since two studies were outliers because of extremely low [18] and extremely high SUV values [22], they were excluded. Thus, the optimal SUV threshold value found with the remaining ten studies was 2.4 (Fig. 5).

Discussion

To our knowledge this is the first systematic review and meta-analysis conducted to evaluate the performance of ^{18}F-FDG PET/CT (separate or combined with dedicated ^{18}F-FDG PET) in the diagnosis of musculoskeletal soft tissue tumors. Our meta-analysis included 755 patients and we only selected studies where we could clearly separate bone from soft tissue tumors through anatomical imaging modalities and with TP, TN, FP, and FN results at the patient level. To that effect, our AUC index was 0.92, and therefore the diagnostic test applied is highly accurate in distinguishing benign from malignant

Table 2 Number and frequency of soft tissue lesions

Soft tissue tumors	n	%
Malignant soft tissue tumors	451	60
Leiomyosarcoma (and other smooth muscle tumors)	54	7.2
Angiosarcoma (and other vascular tumors)	14	1.9
Liposarcoma (and other adipocytic tumors)	109	14.4
Malignant fibrohistiocytic (and other MFH tumors)	78	10.3
Malignant schwannoma	8	1.1
Peripheral nerve sheath tumor	30	4.0
Chondro-osseous tumors	8	1.1
Rhabdomyosarcoma (and other skeletal muscle tumors)	13	1.7
Fibroblastic/myofibroblastic tumors	23	3.0
Tumors of uncertain differentiation[a]	50	6.6
Synovial sarcoma	39	5.2
PNET & Ewing's	11	1.5
Other cancers	14	2.0
Benign soft tissue tumors	306	40
Hemangioma (and other benign vascular tumors)	25	3.3
Myxoma and other tumors of uncertain differentiation	10	1.3
Lipoma (and other adipocytic tumors)	71	9.4
Pigmented villonodular synovitis	11	1.5
Postsurgical or posttraumatic lesions[b]	21	2.8
Infection and inflammation	39	5.2
Neurofibroma/pericytic tumors	34	4.5
Schwannoma	52	6.9
Fibroblastic/myofibroblastic tumors	32	4.2
Chondro-osseous tumors	7	0.9
Other benign lesions	4	0.5

MFH malignant fibrohistiocytic, *PNET* primitive neuroectodermal tumor

[a] Excludes synovial sarcoma, PNET, and Ewing's

[b] Includes scars, hematomas, and thrombophlebitis

Table 3 Studies included in meta-analysis

First author (year)	No. of lesions	Image type	TP	FN	TN	FP
Nieweg [21]	22	PET	18	0	3	1
Schulte [13]	102	PET	65	2	23	12
Lodge [14]	29	PET	12	0	13	4
Lucas [15]	31	PET	18	1	9	3
Ferner [10]	22	PET	7	0	13	2
Watanabe [16]	35	PET	7	0	21	7
Dimitrakopoulou-Strauss [17]	56	PET	43	0	3	10
Ahmed [12]	22	PET	NA	NA	14	8
Suzuki [18]	44	PET	18	2	20	4
Shin [11]	44	PET/CT	20	5	13	6
Charest [19]	160	PET/CT	118	8	34	0
Okazumi [20]	46	PET	14	18	12	2
Amini [22]	100	PET/CT	47	9	39	5
Leal (2014) [23]	44	PET/CT	19	0	20	5
Total	757	–	406	45	235	69

TP true-positive, *FN* false-negative, *FP* false-positive, *TN* true-negative, *NA* data not available

lesions. Overall we found the sensitivity, specificity, accuracy, positive predictive value and negative predictive value for differential diagnosis of musculoskeletal soft tissue tumors to be 96, 77, 88, 86, and 90 %, respectively.

Although we initially attempted to evaluate the performance of only ^{18}F-FDG PET/CT (not dedicated PET), we found only a few studies that met our selection criteria [11, 19, 22, 23]. Although there were only four studies that met our criteria, we believe that our work represents the best concurrent comparison between these modalities for lesion discrimination based on the current literature, which is limited by a lack of high-quality studies and/or the lack of reproducible/coherent reporting practices in the radiological literature.

When comparing the performance of PET/CT vs dedicated PET scanners, there was a trend towards higher AUC (0.95 vs 0.90) and specificity (0.85 vs 0.71) for PET/CT despite the small number of studies using PET/CT. We also found a trend towards better sensitivity for dedicated PET (0.91 vs 0.96), although the partial AU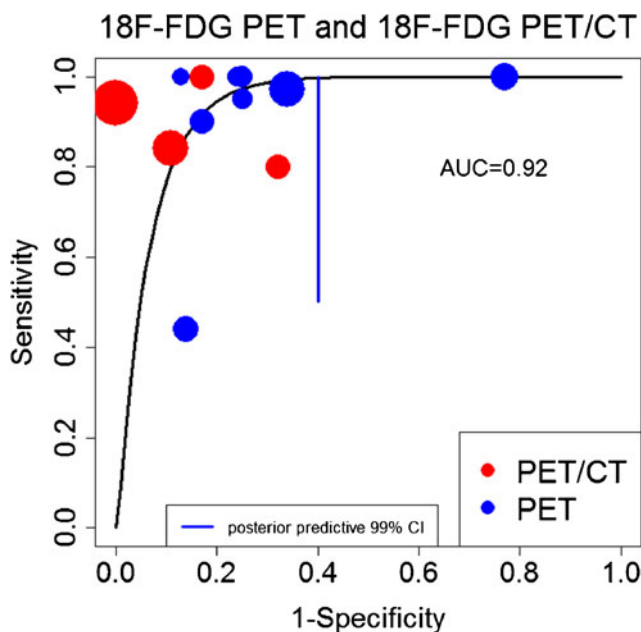C suggests that PET/CT yields better performance for sensitivity over the domain of high specificity (0.17 vs 0.10). However, none of the results attained statistical significance for comparing PET/CT to dedicated PET. The increased specificity and accuracy of PET/CT are most likely due to the added benefit of the anatomical (CT) portion of the PET/CT images. CT accurately measures the tumor diameter and evaluates tumor depth and contact of the lesion with the adjacent structures such as joints, bones, and vessels. To that effect, CT plays an essential role in diagnosis of these extremity masses by helping to evaluate possible aggressiveness, which consequently increases specificity in the diagnosis of soft tissue malignancies, complementing the metabolic information that PET provides. While the previous meta-analysis [2] found a good sensitivity and specificity in discriminating benign from malignant soft tissue lesions, by evaluating TP, TN, FP, and FN findings at the lesion level and excluding studies in which initial diagnosis was not the main purpose of the study, we could determine the added benefit of PET/CT in comparison to dedicated PET.

The SUV cutoff value that best separated benign from malignant musculoskeletal soft tissue lesions was 2.4. Among the 12 studies analyzed, 10 reported a threshold between 2 and 4. Two studies were outliers. These discrepancies may be due to the type of lesion studied and/or the acquisition time. ^{18}F-FDG uptake may increase with time in soft tissue malignancies and decrease with time in inflammatory lesions. In one study [22] the suggested optimal SUV cutoff value was 5.2; however, the authors included abscesses (which exhibit high uptake). In the other study [18] the suggested optimal cutoff was 0.8; however, the authors only evaluated lipomatous lesions (which
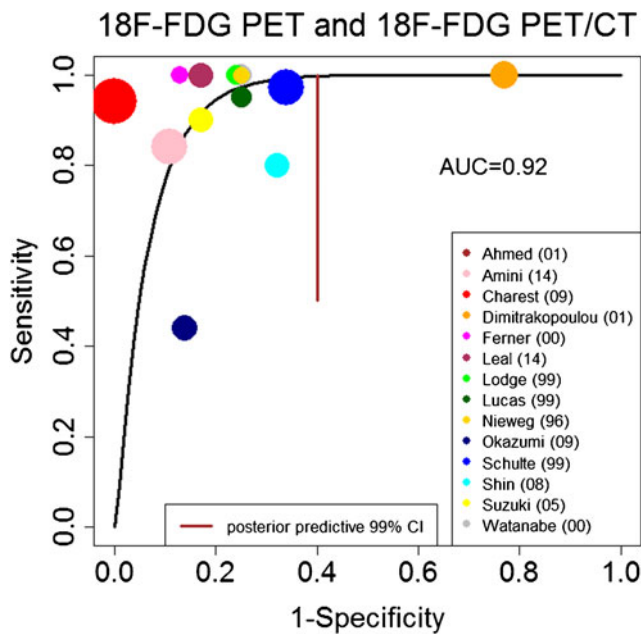
Fig. 2 Observed sensitivity, specificity, and estimated posterior mean ROC curve—^{18}F-FDG PET and ^{18}F-FDG PET/CT. Each point in the scatterplot depicts the observed sensitivity and 1-specificity for one study. *Color* is used to indicate the study. The *line* depicts the estimated mean ROC curve obtained from meta-analysis. The *vertical line* represents the 99 % posterior predictive credible interval that is obtained for the Ahmed et al. [12] study when predicting sensitivity under the assumption of two malignant lesions

Table 4 Results from meta-analysis of diagnostic performance for differentiating between malignant and benign soft tissue musculoskeletal lesions
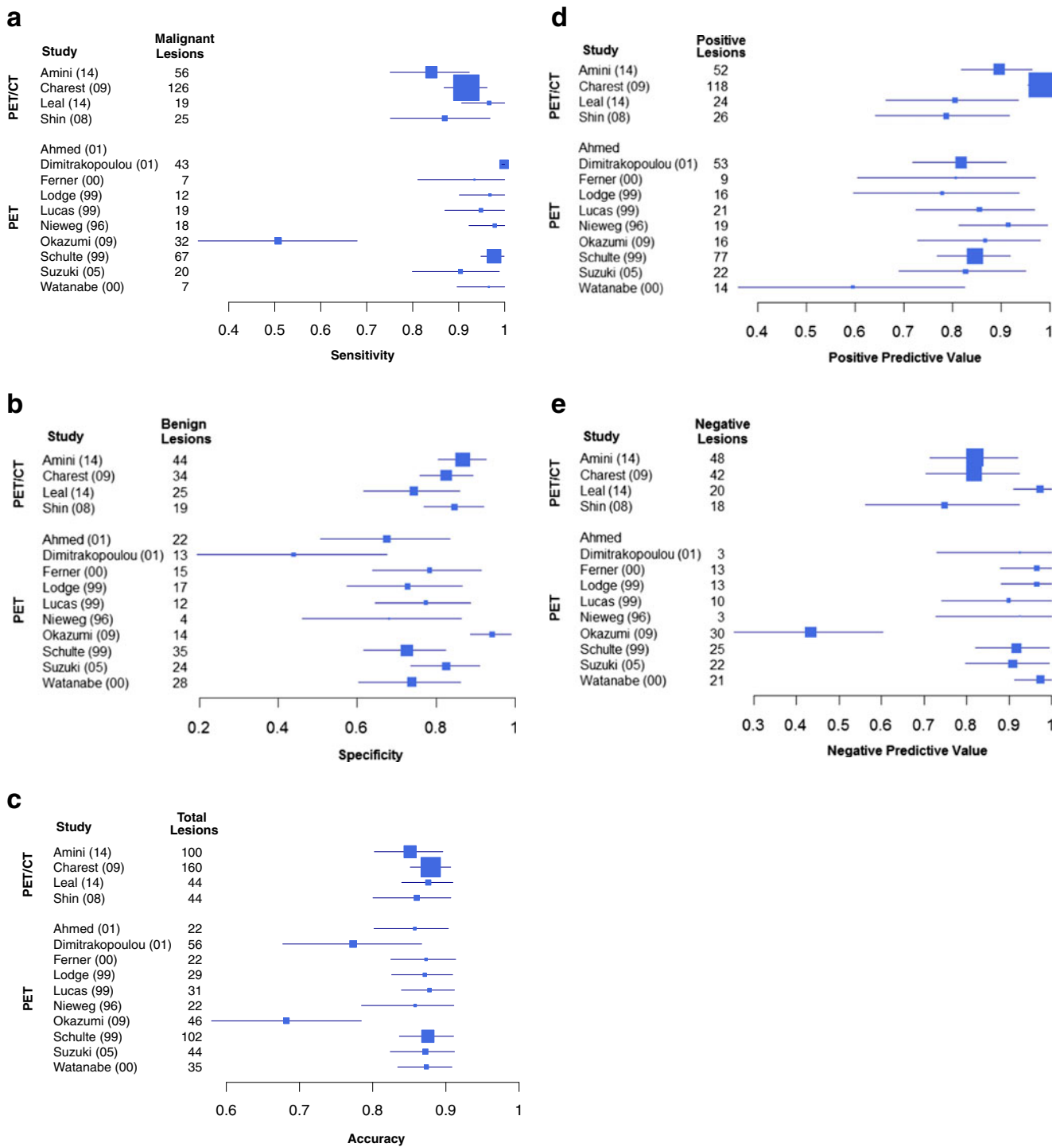
Measure	Posterior mean (95 % HPD interval)		
	PET/CT combined with dedicated PET (n=14)	Only dedicated PET (n=10)	Only PET/CT (n=4)
AUC	0.92 (0.88, 0.96)	0.90 (0.83, 0.96)	0.95 (0.91, 0.98)
Partial AUC	0.13 (0.08, 0.17)	0.10 (0.05, 0.16)	0.17 (0.15, 0.19)
Sensitivity	0.90 (0.88, 1.00)	0.96 (0.89, 1.00)	0.91 (0.85, 0.97)
Specificity	0.77 (0.68, 0.86)	0.71 (0.57, 0.85)	0.85 (0.71, 0.98)
Accuracy	0.88 (0.85, 0.91)	0.85 (0.80, 0.90)	0.89 (0.84, 0.93)
PPV	0.86 (0.77, 0.94)	0.82 (0.74, 0.89)	0.91 (0.73, 1.00)
NPV	0.91 (0.83, 0.99)	0.92 (0.81, 1.00)	0.83 (0.63, 0.99)

PPV positive predictive value, *NPV* negative predictive value

have low uptake) and the images were acquired earlier (40 min after injection). The mean scan time among the 14

studies analyzed was 60 min. In one study [14] the images were acquired 120 min after ^{18}F-FDG injection, and even though higher SUV values were expected to occur, still the optimal SUV cutoff for discriminating benign from malignant lesions in that study was in the range of 2.0 to 4.0 (SUV=3.0).

Unfortunately, FP results may occur due to increased ^{18}F-FDG uptake in benign lesions such as pigmented villonodular synovitis/tenosynovial giant cell tumor, hibernoma, sarcoidosis, myositis ossificans, abscesses, and inflammation [11]. There were 69 FP cases (9.0 %) and 45 (6.0 %) FN cases. FN ^{18}F-FDG PET/CT studies are generally due to myxoid liposarcomas, low-grade fibromyxoid sarcomas, well-differentiated liposarcomas in general, and spindle cell tumors. However, the low ^{18}F-FDG uptake is associated with significantly prolonged survival [6].

Inherent to any meta-analysis, some limitations exist. For instance, all possible studies with our selection criteria may not have been retrieved during the data collection process. Additionally, as studies are conducted in a variety of institutions and in more than one scanner in the same institution, there is variability in SUV measurements, ranging between 10 and 40 % [25]. SUV variability also arises from new features available such as time-of-flight mode. Despite this SUV variability and the SUV cutoff values being reported post hoc, the consistency of SUV discriminative values (range 2.0–3.0) in differentiating malignant from benign musculoskeletal soft tissue lesions makes this limitation of minor impact in daily clinical practice [26]. Risks of bias at the patient level are unclear in four studies, as it was not stated if the studies were prospective or retrospective, if case–control designs were avoided, or if the study had inappropriate exclusions. Additionally, since 36 % of the studies were retrospective, a selection bias could have occurred because the outcome measures had already occurred at the time of the patient selection. The index tests were interpreted in a blind manner in 12 studies. In


Fig. 3 Observed sensitivity, specificity, and estimated posterior mean ROC curve of the observed results by modality—^{18}F-FDG PET and ^{18}F-FDG PET/CT. Each point in the scatterplot depicts the observed sensitivity and 1-specificity for one study. *Color* is used to indicate the modality used in the study. The *line* depicts the estimated mean ROC curve obtained from meta-analysis based on both modalities

Fig. 4 Forest plots for **a** sensitivity, **b** specificity, **c** accuracy, **d** positive predictive value, and **e** negative predictive value depicting the posterior means and 95 % HPD intervals obtained from meta-analysis of diagnostic performance for differentiating between malignant and benign soft tissue musculoskeletal lesions. Results for malignant lesions were not reported for the Ahmed et al. [12] study

two studies [11, 18] it is unclear if the index test results were interpreted without knowledge of the results of the reference standard. Likewise, possible risks of bias could have been due to the reference standard. Although some of the dedicated PET studies might have difficulties in separating pure soft tissue tumors from bone tumors, patients were referred for PET imaging after undergoing other imaging modalities (ultrasound, conventional radiography, CT, or MRI); those exams are able to separate soft tissue from bone masses. The final differentiation of the pure soft tissue tumors from bone lesions was performed by histopathology. Although the majority of the studies used histopathology as the reference test,
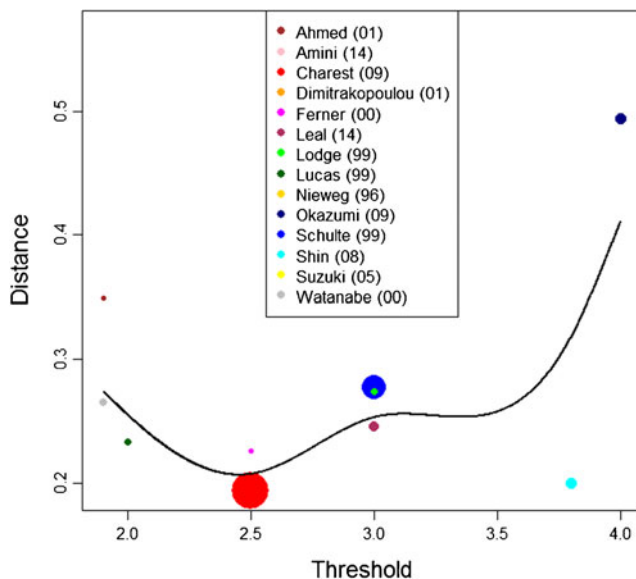
Fig. 5 Threshold analysis was used to determine the optimal SUV cutoff for differentiating between malignant and benign soft tissue musculoskeletal lesions. The optimal SUV threshold value found was 2.4

in all of them it was unclear if histopathology results were blinded to the results of the index test. However, biopsies should be image-guided for optimal tissue sampling so this bias will probably exist in all high-quality institutions/studies. In regards to flow and timing, there was an adequate interval between the index test and reference standard, and in nine studies the patients received the same reference standard (histopathology). Although the remaining five studies used both follow-up and histopathology as reference standards (which characterizes considerable verification bias), in one of those studies [10] only 56 % of the patients had histopathology results, potentially resulting in a high risk of bias. Finally, we were only able to analyze four PET/CT studies with 346 data points (lesions). Moreover, because our meta-analysis facilitates joint estimation of sensitivity and specificity, each of the four studies contributed eight diagnostic summaries to the meta-analysis of PET/CT. However, the results might be considered limited by the quality of the available literature, for which it is often the case that only one of sensitivity or specificity is reported, which precludes evaluation of accuracy, and gold standards are difficult to ascertain. To that effect, we could not attain statistical significance when comparing the performance of PET to concurrent PET/CT, and can only state that the evidence "trends" towards enhanced diagnostic performance for PET/CT.

MRI is important for soft tissue lesion characterization and surgical planning as well as T staging of musculoskeletal soft tissue malignancies. However, the specific characteristics on MRI that describe a soft tissue lesion as malignant (size> 5 cm; ill-defined borders; presence of vessel and neural invasion; compartmental extent; invasion of the surrounding tissues; deep location; heterogeneous T1/T2 signal intensity; and

surrounding edema [27, 28]) are not present in all cases. A prospective multi-institutional study of 548 untreated soft tissue tumor lesions found a sensitivity of 93 %, specificity of 82 %, and accuracy of 85 % of MRI to distinguish benign from malignant lesions [29]. However, other reports have shown that not only are the prevalence of these features variable, many benign lesions display the same characteristics with accuracies ranging from 25 to 90 % [30–32]. Particularly in lesions lower than 5 cm or in those that are not growing, there may not be a clear indication for a biopsy. Consequently, diagnosis of small soft tissue malignancies are a challenge and the American College of Radiology in 2012 classified the majority of the studies relating to diagnosis of soft tissue masses according to relevance and design as exhibiting important study design limitations or not useful as primary evidence because of either: not a clinical study, invalid study design, or conclusions solely based on expert consensus [33].

Significant clinical implications for the management of soft tissue lesions arise from the outcomes of our meta-analysis. Although MRI is fundamental for T staging and surgical planning, until this date and to our knowledge no meta-analyses have been conducted evaluating the performance of MRI in the diagnosis of musculoskeletal soft tissue lesions. ^{18}F-FDG PET/CT offers additional information. For example, from our meta-analysis there were 2 % of lesions unrelated to musculoskeletal soft tissue tumors (lymphomas, melanomas, metastases, squamous cell carcinoma, etc.). ^{18}F-FDG PET/CT does not depend on tumor size to differentiate benign from malignant lesions, as tumors <5 cm may have high ^{18}F-FDG uptake [23]. Unfortunately, we were not able to determine, from our data, the incidence of malignant lesions <5 cm and its possible impact on patient management. The management of malignant soft tissue lesions requires a thorough history, physical examination, and a cautiously planned biopsy that will provide precise histological diagnosis and grading. While biopsies of soft tissue lesions are easily performed, the risk of sampling errors should not be overlooked because of the heterogeneous composition of most soft tissue sarcomas (as they may arise in fat, muscles, blood vessels, nerves, and other connective tissues). ^{18}F-FDG avidity occurs in the majority of these tumors because the uptake is related to tumor cellularity, biological behavior, and the extent of activation of inflammatory cells [34]. For example, on MRI, a liposarcoma may present itself as a heterogeneous mass with extensive "dirty fat" and some areas of non-adipose solid muscle density mass. The differential diagnosis would be a fibrous component of a liposarcoma or a true dedifferentiated component. The latter would be ^{18}F-FDG avid and should guide biopsy for the correct diagnosis. Therefore, ^{18}F-FDG PET/CT is not only a sensitive method to detect these lesions, it is a powerful tool to guide biopsies. Furthermore, since ^{18}F-FDG PET/CT is a whole-body imaging modality, staging is a vital part of the analysis. Malignant soft tissue sarcomas have a significant

metastatic risk that ranges from 20 to 100 % depending on the histological type and grade [24]. Although metastases often arise in the lung, they may be located in the lymph nodes and ^{18}F-FDG PET/CT is the most sensitive method to detect these lesions. For example, ^{18}F-FDG PET/CT is imperative in staging epithelioid hemangioendotheliomas and angiosarcomas as they may be multifocal and/or metastasize to lungs, liver, bone, and bone marrow [19, 35]. Additionally, occult lesions in rare tumors (e.g., synovial, epithelioid, alveolar soft part and clear cell sarcomas, extraskeletal Ewing, desmoplastic round cell tumors and PEComas) may not be identified by conventional imaging, nonetheless are easily detected by ^{18}F-FDG PET/CT [36, 37].

Conclusion

In conclusion, both ^{18}F-FDG PET/CT and dedicated PET are highly sensitive and accurate in the diagnosis of malignant soft tissue lesions. ^{18}F-FDG PET/CT is more specific and accurate than dedicated PET (85 % vs 71 % and 89 % vs 85 %, respectively) for diagnosing malignant soft tissue lesions. Our systematic review provided some evidence to support a SUV threshold of 2.4 for discriminating between malignant and benign lesions. ^{18}F-FDG PET/CT whole-body imaging should be used for initial diagnosis of suspicious musculoskeletal soft tissue tumors.

References

1. Landis MD, Lehmann BD, Pietenpol JA, Chang JC. Patient-derived breast tumor xenografts facilitating personalized cancer therapy. Breast Cancer Res 2013;15:201. doi:10.1186/bcr3355.
2. Ioannidis JP, Lau J. 18F-FDG PET for the diagnosis and grading of soft-tissue sarcoma: a meta-analysis. J Nucl Med 2003;44:717–24.
3. Treglia G, Salsano M, Stefanelli A, Mattoli M, Giordano A, Bonomo L. Diagnostic accuracy of 18F-FDG-PET and PET/CT in patients with Ewing sarcoma family tumours: a systematic review and a meta-analysis. Skeletal Radiol 2012;41:249–56. doi:10.1007/s00256-011-1298-9.
4. Tateishi U, Hosono A, Makimoto A, Sakurada A, Terauchi T, Arai Y, et al. Accuracy of 18F fluorodeoxyglucose positron emission tomography/computed tomography in staging of pediatric sarcomas. J Pediatr Hematol Oncol 2007;29:608–12. doi:10.1097/MPH.0b013e318142b5ab.
5. Gerth HU, Juergens KU, Dirksen U, Gerss J, Schober O, Franzius C. Significant benefit of multimodal imaging: PET/CT compared with PET alone in staging and follow-up of patients with Ewing tumors. J Nucl Med 2007;48:1932–9. doi:10.2967/jnumed.107.045286.
6. Hong SP, Lee SE, Choi YL, Seo SW, Sung KS, Koo HH, et al. Prognostic value of 18F-FDG PET/CT in patients with soft tissue sarcoma: comparisons between metabolic parameters. Skeletal Radiol 2014;43:641–8. doi:10.1007/s00256-014-1832-7.
7. Benz MR, Allen-Auerbach MS, Eilber FC, Chen HJ, Dry S, Phelps ME, et al. Combined assessment of metabolic and volumetric changes for assessment of tumor response in patients with soft-tissue sarcomas. J Nucl Med 2008;49:1579–84. doi:10.2967/jnumed.108.053694.
8. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36. doi:10.7326/0003-4819-155-8-201110180-00009.
9. Hellmich M, Abrams KR, Sutton AJ. Bayesian approaches to meta-analysis of ROC curves. Med Decis Making 1999;19:252–64.
10. Ferner RE, Lucas JD, O'Doherty MJ, Hughes RA, Smith MA, Cronin BF, et al. Evaluation of (18)fluorodeoxyglucose positron emission tomography ((18)FDG PET) in the detection of malignant peripheral nerve sheath tumours arising from within plexiform neurofibromas in neurofibromatosis 1. J Neurol Neurosurg Psychiatry 2000;68:353–7.
11. Shin DS, Shon OJ, Han DS, Choi JH, Chun KA, Cho IH. The clinical efficacy of (18)F-FDG-PET/CT in benign and malignant musculoskeletal tumors. Ann Nucl Med 2008;22:603–9. doi:10.1007/s12149-008-0151-2.
12. Ahmed AR, Watanabe H, Aoki J, Shinozaki T, Takagishi K. Schwannoma of the extremities: the role of PET in preoperative planning. Eur J Nucl Med 2001;28:1541–51.
13. Schulte M, Brecht-Krauss D, Heymer B, Guhlmann A, Hartwig E, Sarkar MR, et al. Fluorodeoxyglucose positron emission tomography of soft tissue tumours: is a non-invasive determination of biological activity possible? Eur J Nucl Med 1999;26:599–605.
14. 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.
15. Lucas JD, O'Doherty MJ, Cronin BF, Marsden PK, Lodge MA, McKee PH, et al. Prospective evaluation of soft tissue masses and sarcomas using fluorodeoxyglucose positron emission tomography. Br J Surg 1999;86:550–6. doi:10.1046/j.1365-2168.1999.01090.x.
16. Watanabe H, Shinozaki T, Yanagawa T, Aoki J, Tokunaga M, Inoue T, et al. Glucose metabolic analysis of musculoskeletal tumours using 18fluorine-FDG PET as an aid to preoperative planning. J Bone Joint Surg Br 2000;82:760–7.
17. Dimitrakopoulou-Strauss A, Strauss LG, Schwarzbach M, Burger C, Heichel T, Willeke F, et al. Dynamic PET 18F-FDG studies in patients with primary and recurrent soft-tissue sarcomas: impact on diagnosis and correlation with grading. J Nucl Med 2001;42:713–20.
18. Suzuki R, Watanabe H, Yanagawa T, Sato J, Shinozaki T, Suzuki H, et al. PET evaluation of fatty tumors in the extremity: possibility of using the standardized uptake value (SUV) to differentiate benign tumors from liposarcoma. Ann Nucl Med 2005;19:661–70.
19. Charest M, Hickeson M, Lisbona R, Novales-Diaz JA, Derbekyan V, Turcotte RE. FDG PET/CT imaging in primary osseous and soft tissue sarcomas: a retrospective review of 212 cases. Eur J Nucl Med Mol Imaging 2009;36:1944–51. doi:10.1007/s00259-009-1203-0.

20. Okazumi S, Dimitrakopoulou-Strauss A, Schwarzbach MH, Strauss LG. Quantitative, dynamic 18F-FDG-PET for the evaluation of soft tissue sarcomas: relation to differential diagnosis, tumor grading and prediction of prognosis. Hell J Nucl Med 2009;12: 223–8.

21. Nieweg OE, Pruim J, van Ginkel RJ, Hoekstra HJ, Paans AM, Molenaar WM, et al. Fluorine-18-fluorodeoxyglucose PET imaging of soft-tissue sarcoma. J Nucl Med 1996;37:257–61.

22. Amini B, Madewell JE, Chuang HH, Haygood TM, Hobbs BP, Fox PS, et al. Differentiation of benign fluid collections from soft-tissue sarcomas on FDG-PET/CT. J Cancer 2014;5:328–35. doi:10.7150/jca.8310.

23. Leal AL, Etchebehere M, Santos AO, Kalaf G, Pacheco EB, Amstalden EM, et al. Evaluation of soft-tissue lesions with (18)F-FDG PET/CT: initial results of a prospective trial. Nucl Med Commun 2014;35:252–9. doi:10.1097/MNM.0000000000000041.

24. Fletcher CDM. The evolving classification of soft tissue tumours: an update based on the new WHO classification. Histopathology 2006;48:3–12. doi:10.1111/j.1365-2559.2005.02284.x.

25. Boktor RR, Walker G, Stacey R, Gledhill S, Pitman AG. Reference range for intrapatient variability in blood-pool and liver SUV for 18F-FDG PET. J Nucl Med 2013;54:677–82. doi:10.2967/jnumed.112.108530.

26. Ewald B. Post hoc choice of cut points introduced bias to diagnostic research. J Clin Epidemiol 2006;59:798–801. doi:10.1016/j.jclinepi.2005.11.025.

27. Robinson E, Bleakney RR, Ferguson PC, O'Sullivan B. Oncodiagnosis panel: 2007: multidisciplinary management of soft-tissue sarcoma. Radiographics 2008;28:2069–86. doi:10.1148/rg.287085167.

28. Chhabra A, Soldatos T. Soft-tissue lesions: when can we exclude sarcoma? AJR Am J Roentgenol 2012;199:1345–57. doi:10.2214/AJR.12.8719.

29. Gielen JL, De Schepper AM, Vanhoenacker F, Parizel PM, Wang XL, Sciot R, et al. Accuracy of MRI in characterization of soft tissue tumors and tumor-like lesions. A prospective study in 548 patients. Eur Radiol 2004;14:2320–30. doi:10.1007/s00330-004-2431-0.

30. Berquist TH, Dalinka MK, Alazraki N, Daffner RH, DeSmet AA, el-Khoury GY, et al. Soft tissue masses. American College of Radiology. ACR appropriateness criteria. Radiology 2000;215(Suppl):255–9.

31. White LM, Wunder JS, Bell RS, O'Sullivan B, Catton C, Ferguson P, et al. Histologic assessment of peritumoral edema in soft tissue sarcoma. Int J Radiat Oncol Biol Phys 2005;61:1439–45. doi:10.1016/j.ijrobp.2004.08.036.

32. Panzarella MJ, Naqvi AH, Cohen HE, Damron TA. Predictive value of gadolinium enhancement in differentiating ALT/WD liposarcomas from benign fatty tumors. Skeletal Radiol 2005;34: 272–8. doi:10.1007/s00256-004-0884-5.

33. National Guideline C. ACR Appropriateness Criteria® soft-tissue masses.

34. Dong A, Wang Y, Dong H, Gong J, Cheng C, Zuo C, et al. Inflammatory myofibroblastic tumor: FDG PET/CT findings with pathologic correlation. Clin Nucl Med 2014;39:113–21. doi:10.1097/RLU.0b013e3182952caa.

35. Rest CC, Botton E, Robinet G, Conan-Charlet V, Bizais Y, Visvikis D. FDG PET in epithelioid hemangioendothelioma. Clin Nucl Med 2004;29:789–92.

36. Zhang WD, Li CX, Liu QY, Hu YY, Cao Y, Huang JH. CT, MRI, and FDG-PET/CT imaging findings of abdominopelvic desmoplastic small round cell tumors: correlation with histopathologic findings. Eur J Radiol 2011;80:269–73. doi:10.1016/j.ejrad.2010.06.046.

37. Arora VC, Price AP, Fleming S, Sohn MJ, Magnan H, LaQuaglia MP, et al. Characteristic imaging features of desmoplastic small round cell tumour. Pediatr Radiol 2013;43:93–102. doi:10.1007/s00247-012-2485-0.