^{18}F-FDG PET for the Diagnosis and Grading of Soft-Tissue Sarcoma: A Meta-Analysis

John P.A. Ioannidis, MD[1,2]; and Joseph Lau, MD[1]

[1]Tufts–New England Medical Center Evidence-Based Practice Center, Division of Clinical Care Research, Department of Medicine, Tufts–New England Medical Center, Boston, Massachusetts; and [2]Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

PET using ^{18}F-FDG is increasingly used for the diagnosis and grading of tumors. Several studies have been performed that evaluate the diagnostic and grading performance of ^{18}F-FDG PET for soft-tissue sarcoma, but each study has had a limited sample size. Therefore, we undertook a comprehensive meta-analysis of the evidence. **Methods:** Relevant studies were identified from MEDLINE and EMBASE. Diagnostic and grading performance were evaluated for qualitative visualization; standard uptake value (SUV, cutoffs of 2.0 and 3.0); and metabolic rate of glucose (MRG, cutoff of 6.0 μmol/100 g/min). Quantitative data synthesis included independent weighting of sensitivity and specificity, construction of summary receiver operating characteristic curves, and pooled analyses. **Results:** The meta-analysis included 15 studies with 441 soft-tissue lesions (227 malignant, 214 benign). For diagnosis of malignant versus benign lesions, typical pairs of sensitivity and specificity estimates from the summary receiver operating characteristic curves were 92% and 73% for qualitative visualization; 87% and 79% for SUV 2.0; 70% and 87% for SUV 3.0; and 74% and 73% for MRG 6.0. Diagnostic performance was similar for primary and recurrent lesions. By qualitative interpretation, ^{18}F-FDG was positive in all intermediate/high-grade tumors (95% confidence interval [CI], 97.3%–100%), 74.4% (95% CI, 58.6%–85.9%) of low-grade tumors, and 39.3% (95% CI, 29.1%–50.3%) of benign lesions (including 11 of 12 inflammatory lesions). Using an SUV cutoff of 2.0, respective rates were 89.4% (95% CI, 79.4%–95.6%), 33.1% (95% CI, 15.6%–55.3%), and 19.1% (95% CI, 10.6%–30.5%). Limited data on comparisons with MRI and CT showed no differences against ^{18}F-FDG PET in diagnosing recurrent and metastatic disease. **Conclusion:** ^{18}F-FDG PET has very good discriminating ability in the evaluation of both primary and recurrent soft-tissue lesions. ^{18}F-FDG PET may be helpful in tumor grading but offers inadequate discrimination between low-grade tumors and benign lesions.

Key Words: ^{18}F-FDG PET; meta-analysis; soft-tissue sarcoma

J Nucl Med 2003; 44:717–724

PET using ^{18}F-FDG is increasingly being established as a key imaging technique in the diagnosis, grading, staging, and assessment of response to therapy of diverse types of tumors (1,2). ^{18}F-FDG PET is considered as a potentially major advance in clinical practice, because it may offer information about not only the anatomic extent but also the behavior of tumors, thus helping to guide therapeutic choices. Soft-tissue sarcoma (STS) is a malignancy for which evidence on the diagnostic performance of ^{18}F-FDG PET has been accumulating rapidly in recent years. While STS is not a common tumor, accounting for only 1% of cancer diagnoses in the United States (3), it carries a high mortality rate, and tumor grade is a strong predictor of outcome (4). Accurate imaging studies that can provide additional information on grading and staging may be useful in the management of the disease. ^{18}F-FDG PET has been evaluated in several series of patients with suspected or proven STS with the aim to assess whether this imaging modality may be useful in differentiating malignant from benign soft-tissue lesions and in providing adequate information to grade STS. However, most of the studies have been small and used diverse qualitative or quantitative diagnostic parameters. Therefore, we undertook a meta-analysis to combine results across studies and to obtain a more reliable estimate of the diagnostic performance of ^{18}F-FDG PET in the assessment of soft-tissue lesions.

MATERIALS AND METHODS

Scope and Eligibility Criteria

We addressed the performance of ^{18}F-FDG PET as a diagnostic test for differentiating STS from benign soft-tissue lesions and for grading STS. We considered all studies with ^{18}F-FDG PET evaluations on at least 3 subjects with soft-tissue lesions, at least 1 of which was STS. Reports including only STS cases were eligible, provided that they had collected information on tumor grade. Reports including both soft-tissue lesions and lesions at other sites (e.g., bone) were included, provided that ^{18}F-FDG PET information on soft-tissue lesions could be separated.

We conducted MEDLINE and EMBASE searches (last update, February 2002) using various search terms for PET (PET, positron emission tomography, ^{18}F-FDG, and fluorodeoxyglucose) and sarcoma and limited the search to "human subjects." Searches were

also conducted using names of specific histologic types of STS (liposarcoma, malignant fibrous histiocytoma, leiomyosarcoma, fibrosarcoma, malignant schwannoma, synovial sarcoma, and peripheral nerve sheath tumor). Furthermore, we also perused the references of retrieved articles to find additional studies and communicated with expert investigators for additional data and clarifications. We set no language restrictions.

Data Extraction

We extracted information on authors, year of publication, age, number of subjects, benign and malignant lesions evaluated, inclusion and exclusion criteria, number of primary lesions and of lesions assessed for recurrence, study design (prospective, retrospective, or unclear), histologic types, technical characteristics of PET, diagnostic parameters considered, definition and interpretation of the reference test (biopsy and type thereof, operative histologic diagnosis, other imaging, clinical, other, and unspecified), and potential for verification bias. Verification bias refers to incomplete confirmation of the results of the test under investigation with the reference test (e.g., no biopsy performed when PET suggests benign lesion).

For each report, we recorded the number of true positives, false positives, true negatives, and false negatives for ^{18}F-FDG PET in diagnosing malignant versus benign lesions using the following prespecified parameters: (a) qualitative visualization (simple visualization, qualitative interpretation by experts, or assessment based on a tumor-to-background ratio [TBR] $\geq$ 3.0 without correction for dose of ^{18}F-FDG, weight, and plasma glucose); (b) standard uptake value (SUV) $\geq$ 2.0; (c) SUV $\geq$ 3.0; and (d) metabolic rate of glucose (MRG) $\geq$ 6 μmol/100 g/min. We also separated evaluations for primary lesions from evaluations for potential recurrences. Furthermore, whenever information was provided on tumor grade, we recorded the number of lesions that were positive by PET (based on each of the above definitions) for intermediate/high-grade (G II/III) and low-grade (G I) tumors. Benign lesions were separated into noninflammatory and inflammatory ones.

For reports that had also used CT scans or MRI, we evaluated the performance of each test in diagnosing primary disease, local recurrence, and metastases. Data were compared on the same patients when 2 imaging procedures had been performed in parallel. For reports describing longitudinal evaluation of therapeutic response with serial ^{18}F-FDG PET, the baseline (pretherapy) data were used in the quantitative synthesis. In addition, we extracted descriptive information regarding the impact of ^{18}F-FDG PET on patient management.

Statistical Analysis

Data were combined quantitatively to provide summary information across studies for each of the 4 prespecified diagnostic definitions. We estimated the overall number of true positives, false negatives, true negatives, and false positives, and we estimated the overall sensitivities and specificities using a random-effects model that incorporated between-study variability. We also performed data synthesis using the summary receiver operating characteristic (SROC) approach that takes into consideration the interdependence between sensitivity and specificity (5,6). Combining sensitivity and specificity data independently across studies may underestimate both parameters and provides no information about the effect of diagnostic threshold. However, the overall random-effects estimates fall close to the SROC curve and can provide a useful indicator of where most investigations operated. The SROC curve shows how the true-positive rate (sensitivity) changes as a function of false-positive rate (1 − specificity) across all studies, when the same diagnostic criterion is used to classify cases as benign or malignant. It is estimated by the equation D = a + bS, where D is the difference of the logit of the true-positive rate and the logit of the false-positive rate and S is the respective sum. Both unweighted and weighted regressions were evaluated. When sufficient data are available, the area under the curve (AUC) can also be estimated. For all analyses, data on primary lesions were treated separately from data on evaluation of recurrences whenever feasible. The main analyses combined all data. Subgroup analyses were also performed for each subgroup (primary and recurrent). We also report typical pairs of sensitivity and specificity from these curves, taking as reference value for specificity the one estimated by random effects weighting of specificities across the included studies.

For the analysis of grading, we calculated the percentage of subjects with each type/grade of lesion when ^{18}F-FDG PET was positive by each of the prespecified diagnostic criteria. Exact 95% confidence intervals (CIs) are also provided. Differences in pooled proportions were tested by Fisher exact test. Analyses were conducted in SPSS (SPSS Inc.), Meta-Test (Joseph Lau), and StatXact 3.0 (Cytel Inc.).

RESULTS

Eligible Studies

Thirty-three potentially relevant articles were identified and examined in full text. Of those, 18 articles were rejected: 10 articles were rejected because they had <3 soft-tissue lesions; 3 other reports (with a total of 37 patients) (7–9) were rejected because they included only STS cases without grading or only benign lesions; and another 2 overlapping reports were rejected (10,11) because they included data on both bone and soft-tissue lesions that could not be separated. We attempted to contact authors for more data but this was not successful. The remaining articles (12–29) were scrutinized for potential overlap. Among 3 overlapping reports by the Heidelberg team (12–14), only the most comprehensive was retained (14). There were 4 publications by the Groningen team (15–18). The most recent (16) was excluded because it overlapped, to a considerable extent, with previous reports and provided only MRG data, whereas the earlier reports also had additional parameters. Two other reports (17,18) had partial overlap: one examined only primary lesions, whereas the other examined some of the same primary lesions and recurrent lesions. To avoid duplication, we retained the first report (17) and the recurrent-lesions group from the other report (18). Finally, the St. Thomas team had published 5 reports (19–23), but they all provided largely nonoverlapping, complementary information. Thus, 15 reports were analyzed (Table 1) (14,15,17–29).

In all, a total of 421 subjects with soft-tissue lesions had been assessed with ^{18}F-FDG PET, and 416 could be evaluated (5 patients had either a PET technical failure or no diagnostic documentation). There were 441 evaluable lesions (227 malignant and 214 benign). Three small studies included only malignant lesions documented by biopsy.

TABLE 1
Characteristics of Studies Evaluating ^{18}F-FDG PET for Diagnosis or Grading of STS

Study	Mean age (range)	Patients with STL	Evaluable patients* (lesions)	Malignant (%)	Inclusion/exclusion criteria	Primary/ recurrent	Design
Schwarzbach 2000	54 (16–78)	50	47 (56)	35 (63)	Consecutive patients suspected of STS from clinical symptoms and MRI or CT	19/37	Prospective
Watanabe 2000	47 (12–77)	37	37 (37)	7 (19)	Patients referred for STM evaluation after CT, MRI, or angiography	33/4	Unclear
Griffeth 1992	50 (16–84)	21	19 (20)	10 (50)	Patients referred for STM evaluation after CT or MRI	10/10?	Unclear
Schulte 1999	49 (1–89)	102	102 (102)	66 (65)	Patients with STL suggestive of benign or malignant tumor on US or MRI	88/14	Prospective
Lucas 1999	51 (6–85)	30	30 (31)	19 (61)	Consecutive patients with suspected STS from clinical symptoms and MRI	31/0	Unclear
Nieweg 1996	50 (18–82)	22	22 (22)	18 (82)	Patients suspected of STS from clinical findings	22/0	Prospective
Hain 1999	39 (11–81)	16	16 (16)	2 (13)	Patients after amputation for STS, evaluated for recurrence	0/16	Retrospective
Kern 1988	30 (12–63)	4	4 (4)	3 (75)	Patients referred with STM	4/0	Unclear
Lucas 1998	51 (3–84)	62	62 (72)	19 (26)	STS patients evaluated for local recurrence	0/72	Retrospective
Lodge 1999	50 (18–76)	29	29 (29)	12 (41)	STM suspected to be malignant based on clinical examination and MRI	29/0	Prospective
van Ginkel 1996	49 (18–80)	20	7 (7)	7 (100)	Patients with biopsy-proven recurrent STS, assessed before HILP therapy	0/7	Prospective
Kole 1997	54 (32–83)	17	17 (17)	15 (88)	Patients with proven or suspected local recurrence of STS	0/17	Prospective
Ferner 2000	26 (12–62)	15	15 (19)	5 (26)	Patients with neurofibromatosis with symptomatic visible neurofibromas	19/0	Unclear
Adler 1990	ND (41–85)	5	5 (5)	5 (100)	Patients with liposarcoma of the thigh	5/0	Prospective?
Jones 1996	54 (46–65)	4	4 (4)	4 (100)	Histologically confirmed STS with planned radiotherapy/hyperthermia	4/0	Unclear

*In Schwarzbach et al. (2000), 3 patients were excluded because of no biopsy or technical failure; in Griffeth et al. (1992), 2 patients were excluded because of no diagnostic documentation or follow-up; in van Ginkel et al. (1996), 20 patients with STS were assessed, but those with primary lesions overlapped with patients included in Nieweg et al. (1996). Thus, only the 7 recurrent lesions were considered in all analyses and are listed in table.

STL = soft-tissue lesion; STM = soft-tissue mass; US = ultrasound; HILP = hyperthermic isolated limb perfusion; ND = no data.

Whenever both soft-tissue lesions and other lesions were included (Watanabe, Kern, Ferner, Jones), only data on soft-tissue lesions are presented in table. Median age is provided in Schwarzbach et al. (2000) and in Schulte et al. (1999).

Otherwise, both malignant and benign lesions were included in 12 reports. There was a total of 264 primary lesions and 177 lesions evaluated for recurrence. Seven studies addressed only primary lesions, 4 addressed only lesions suspected for recurrence, and 4 addressed both.

Diagnoses

We retrieved specific histologic diagnoses on malignant tumors from all studies with one exception (*22*). Among 208 malignant tumors with recorded histology, the most common were liposarcoma and variants (*n* = 56), malignant fibrous histiocytoma (*n* = 43), leiomyosarcoma (*n* = 21), fibrosarcoma (including myxofibrosarcoma) (*n* = 13), malignant schwannoma (*n* = 13), synovial sarcoma (*n* = 12), and peripheral nerve sheath tumor (*n* = 11). These types accounted for over 80% of STS. Of note, soft-tissue malignancies also included 1 case of non-Hodgkin lymphoma and 1 case of metastatic nasopharyngeal carcinoma.

Specific diagnoses were available for all 214 benign lesions. The most common diagnoses were postsurgical or posttraumatic lesions (typically scars, *n* = 113), lipoma (*n* = 23), neurofibroma (*n* = 21), hemangioma (*n* = 13), and schwannoma (*n* = 12). Overall, there were 199 benign noninflammatory lesions and 15 benign lesions with acute or chronic inflammation (infectious or noninfectious).

^{18}F-FDG PET and Reference Test Characteristics

A variable amount of radiopharmaceutical was used across studies (148–407 MBq). Imaging was typically performed after fasting from a few hours to overnight. With one exception (*21*), all studies attempted some qualitative interpretation of the PET images, based mostly on various nonquantitative criteria with consensus between radiologists, and in 3 studies, a crude TBR was also used. SUV was estimated in 9 studies, all of which had primary lesions, and 3 also included an evaluation of recurrences (total of 12 case

series). MRG was estimated in 5 studies (3 of primary lesions, 2 evaluations of recurrences). Histology was the typical reference standard for all lesions, but there were some notable exceptions. In 3 studies (*19,20,22*), most or all of the patients with benign-appearing lesions did not have histologic confirmation, and diagnosis depended only on clinical and radiologic criteria. This represents considerable verification bias. A small number of patients with benign lesions did not have histologic confirmation in a fourth study (*14*) (Table 2). Typically there was no clear mention about whether ^{18}F-FDG PET evaluations were performed without knowledge of histologic results.

Data Synthesis: Diagnosis of Malignant Versus Benign Lesions

Summary estimates are presented in Table 3 for all the examined diagnostic parameters. Qualitative visualization interpretations were available on 398 subjects from 13 case series comparing malignant and benign lesions. There were 184 true-positives (patients with malignancies detected as such by ^{18}F-FDG PET), 16 false-negatives (patients with malignancies considered benign by ^{18}F-FDG PET), 145 true-negatives (patients with benign lesions correctly identified by ^{18}F-FDG PET), and 53 false-positives (patients with benign lesions incorrectly classified as malignant by

^{18}F-FDG PET). Respective numbers from 10 case series with SUV data with a cutoff of 2.0 (or 3.0) were 79 (or 60) true-positives, 21 (or 40) false-negatives, 89 (or 97) true-negatives, and 19 (or 11) false-positives. Data on MRG were limited ($n = 66$ from 4 case series) and included 30 true-positives, 12 false-negatives, 18 true-negatives, and 6 false-positives. The overall accuracy rate was 82.7%, 80.8% (or 75.5%), and 72.7%, respectively, with these diagnostic parameters. The studies seemed to operate at different parts of the SROC curve, and independent estimates of sensitivity and specificity (Table 3) or simple pooled accuracy rates might be underestimating the true performance of the test. The diagnostic performance was better when formal SROC analyses were undertaken (Table 3 and Fig. 1). Both SUV and simple qualitative visualization had somewhat better diagnostic performance than MRG. An AUC would be meaningful only for qualitative visualization, when there were both sufficient data and a wide range (0%–100%) of specificity estimates in individual studies. AUC was estimated at 0.91 (0.92 with weighted regression). There was no evidence of a different overall diagnostic performance (based on AUC estimates) when primary lesions were examined separately from evaluations for recurrence. However, diagnostic performance trended toward larger sensi-

TABLE 2
Characteristics of ^{18}F-FDG PET Imaging and of Reference Standards

Study	^{18}F-FDG dose (MBq)	Measures	Reference standard	Potential verification bias
Schwarzbach 2000	370–440	Visualization, SUV	Biopsy, except for 7 patients followed clinically and radiologically for 4–27 mo	Limited
Watanabe 2000	185–350	Visualization, SUV	Biopsy, surgical excision or postmortem, except for 1 patient (clinical and radiologic criteria)	Very limited
Griffeth 1992	370	Visualization (TBR), SUV	Biopsy or surgical excision, except for 1 patient (clinical and radiologic criteria)	Very limited
Schulte 1999	120–300	Visualization (TBR)	Excisional, incisional, or needle biopsy followed by surgical resection when indicated	No
Lucas 1999	350	Visualization, SUV	Open biopsy, followed by surgical excision when appropriate	No
Nieweg 1996	187–407	Visualization, SUV, MRG	Biopsy (no details)	No
Hain 1999	320	Visualization	Unclear, presumably biopsy performed only when indicated	Considerable
Kern 1988	185	Visualization, MRG	Biopsy and tumor resection	No
Lucas 1998	350	Visualization	Biopsy (and operative diagnosis), when PET or MRI suggested recurrence, otherwise clinical and imaging follow-up	Considerable
Lodge 1999	350	SUV (at 60, 120, 255 min),* MRG	Biopsy followed by appropriate surgical excision	No
Van Ginkel 1996	370	Visualization, MRG	Biopsy of irresectable tumors	No
Kole 1997	370	Visualization, MRG	Biopsy (no details)	No
Ferner 2000	350	Visualization, SUV	Biopsy, except for 9 patients assessed clinically and radiologically	Considerable
Adler 1990	148–278	Visualization (TBR), SUV	Biopsy presumably (no details)	No?
Jones 1996	370	Visualization, SUV	Needle or incisional biopsy followed by surgical resection	No

*SUV data at 60 min are used in quantitative synthesis.

TABLE 3
Summary Diagnostic Performance of ¹⁸F-FDG PET for Diagnosing Malignant vs. Benign Soft-Tissue Lesions

Diagnostic parameter	Case series (subjects)	Independent estimates (95% CI)*		Typical SROC estimates (sensitivity, specificity)†	
		Sensitivity	Specificity	Unweighted	Weighted
Qualitative visualization	13 (398)	88% (81–93)	73% (56–86)	(90%, 73%)	(92%, 73%)
Primary lesions	7 (231)	95% (89–98)	59% (38–78)	(95%, 59%)	(96%, 59%)
Recurrent lesions	5 (146)	81% (69–89)	91% (81–96)	(74%, 91%)	(81%, 91%)
SUV ≥ 2.0	10 (208)	76% (64–85)	79% (71–86)	(92%, 79%)	(87%, 92%)
Primary lesions	7 (157)	79% (67–88)	77% (67–85)	(93%, 77%)	(91%, 77%)
Recurrent lesions	3 (51)	63% (44–78)	91% (71–98)	(70%, 91%)	(72%, 91%)
SUV ≥ 3.0	10 (208)	58% (48–67)	87% (79–92)	(90%, 87%)	(70%, 87%)
Primary lesions	7 (157)	60% (48–72)	86% (77–92)	(82%, 86%)	(66%, 86%)
Recurrent lesions	3 (51)	52% (35–69)	91% (71–98)	(62%, 91%)	(62%, 91%)
MRG ≥ 6.0 μmol/100 g/min	4 (66)	70% (54–82)	73% (53–87)	(72%, 73%)	(74%, 73%)
Primary lesions	3 (52)	75% (57–87)	72% (50–87)	(76%, 72%)	(76%, 72%)
Recurrent lesions	1 (14)	58% (29–83)	100% (16–100)	NE	NE

*Independent random effects pooling of sensitivity and specificity across studies.
†Pairs of sensitivity and specificity from SROC curves; specificity values are selected from random-effects estimates.
NE = not estimable.

tivity and smaller specificity in primary lesions than in evaluations of recurrent lesions (Table 3). ¹⁸F-FDG PET was highly specific in ruling out recurrence but had limited sensitivity for diagnosing recurrence.

Data Synthesis: Grading

In the 11 case series in which tumor grading was assessed and qualitative visualization was used, all 133 (100% [95% CI, 97.3%–100%]) intermediate/high-grade malignant lesions and 32/43 (74.4% [95% CI, 58.6%–85.9%]) low-grade malignant lesions were detected. The respective rate for benign lesions was 35/89 (39.3% [95% CI, 29.1%–50.3%]) in the same studies or 74/196 (37.8% [95% CI, 31.0%–44.9%]) when studies without tumor grading were included. Rates differed significantly between the 3 groups. Of 12 inflammatory benign lesions, 11 were positive on ¹⁸F-FDG PET (91.7% [95% CI, 61.5%–99.8%]).

In the 8 case series in which tumor grading was assessed and SUV was estimated, values above 2.0 were seen in 59/66 (89.4% [95% CI, 79.4%–95.6%]) of intermediate/high-grade malignant lesions, 8/24 (33.3% [95% CI, 15.6%–55.3%]) of low-grade malignant lesions, and 13/68 (19.1% [95% CI, 10.6%–30.5%]) of benign lesions. SUV values above 3.0 were seen in 45/66 (68.2% [95% CI, 55.6%–79.1%]), 3/24 (12.5% [95% CI, 26.6%–32.4%]), and 8/68 (11.8% [95% CI, 5.2%–21.9%]), respectively. With either cutoff, there was no significant difference between the low-grade malignant lesions and the benign lesions, whereas intermediate/high-grade malignant lesions differed significantly from both other groups. Inferences were similar when data on benign lesions were included from studies without tumor grading (data not shown).

MRG data were limited, but the inferences were similar. Values ≥ 6.0 μmol/100 g/min were seen in 32/35 (91.4%

[95% CI, 76.9%–98.2%]) G II/III tumors, 1/13 (7.7% [95% CI, 0.2%–36.0%]) G I tumors, and 6/24 (25.0% [95% CI, 9.8%–46.7%]) benign lesions (all noninflammatory).

Comparison with MRI and CT

Several studies on primary lesions clarify that part of the inclusion criteria was the prior performance of CT, MRI, or ultrasound, suggestive of malignancy. Because MRI/CT and ¹⁸F-FDG PET are performed in series, rather than in parallel, one cannot address their comparative accuracy for primary soft-tissue lesions. The same applies to studies on evaluation of recurrence, with 2 exceptions: In one study (22), ¹⁸F-FDG PET sensitivity was 13/17 (76%) and specificity was 47/50 (94%) versus 15/17 (88%) and 48/50 (96%), respectively, for paired MRI. ¹⁸F-FDG PET was wrong in 3 cases in which MRI was correct. The study may have substantial verification bias. In the other study (15), PET had 92% sensitivity (12/13) and 100% specificity (2/2). Paired MRI failed to diagnose 3 malignant lesions (including the one also misdiagnosed as benign by PET) and misdiagnosed both benign lesions (scar and *Ascaris* mass). PET was correct in 4 cases in which MRI was wrong.

Data on the comparative performance of ¹⁸F-FDG PET for the diagnosis of distant metastasis were also sparse. In 1 study (22), PET was positive by qualitative interpretation in 13/15 lung metastases (sensitivity, 87%) and was negative in all 55 cases without metastasis (specificity, 100%). Paired CT scans had 100% sensitivity and 96% specificity. PET failed to detect 2 lung metastases seen on CT but was correct in 2 cases in which CT was falsely positive. The study suffers from substantial verification bias for patients with negative imaging. Inconclusive retrospective paired CT and PET data were also given by 1 of the excluded studies (7) on 8 patients and were also subject to verification bias.
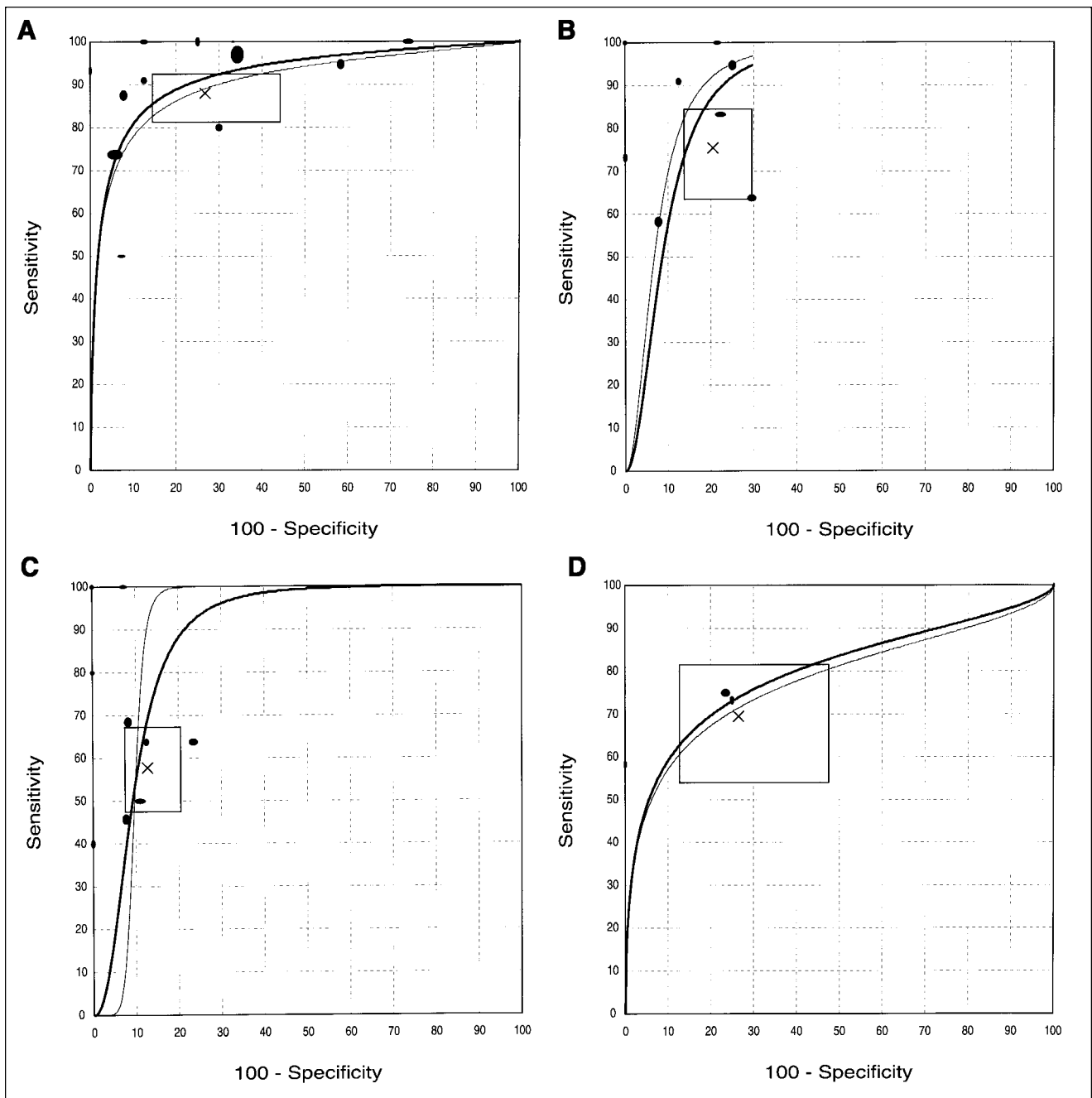
FIGURE 1. SROC curves for (A) qualitative visualization, (B) standard uptake value (cutoff, 2.0), (C) standard uptake value (cutoff, 3.0), and (D) metabolic rate of glucose (cutoff, 6.0 μmol/100 g/min). Curves show trade-off between true-positive rate (sensitivity) and false-positive rate (100 − specificity) across all pertinent studies. Each study is shown by eclipse with diameters approximately proportional to number of subjects evaluated for sensitivity (vertical dimension) and specificity (horizontal dimension) of study. Two SROC curves are shown based on weighted (bold line) and unweighted (thin line) calculations. SROC curves should be used for inferences of diagnostic accuracy preferably within range of sensitivity and specificity values of studies that are included in their calculations. X = random effects estimates of sensitivity and specificity; horizontal and vertical dimensions of rectangle = corresponding 95% confidence intervals.

Longitudinal Evaluation of Treated STS

^{18}F-FDG PET was used for the evaluation of response to therapy in 3 case series with at least 4 subjects involving radiotherapy hyperthermia ($n = 4$); hyperthermic isolated limb perfusion with tumor necrosis α, interferon γ, and melphalan ($n = 20$); and various modalities ($n = 8$), for a total of only 32 patients (*24,18,7*). PET showed clear changes in the treated tumors but did not discriminate partial from complete response in the largest study (*18*).

DISCUSSION

^{18}F-FDG PET has very good, but not perfect, performance in differentiating benign from malignant soft-tissue lesions. Practically all intermediate/high-grade STSs are accurately diagnosed on the basis of qualitative interpretation; almost all of them have SUV $\geq$ 2.0, and the large majority has MRG $\geq$ 6 μmol/100 g/min. Most low-grade tumors are also diagnosed on qualitative interpretation, but they usually have SUV < 2.0 and rarely have MRG $\geq$ 6 μmol/100 g/min. Neither SUV nor MRG can differentiate low-grade tumors from benign lesions, and inflammatory lesions have quantitative characteristics similar to those of high-grade malignancies.

There were no major differences in the diagnostic performance when variable diagnostic parameters were considered, including qualitative visualization, single SUV measurements, or dynamic estimation of MRG. Paradoxically, the estimated diagnostic performance was at its worst with the more sophisticated MRG. The differences could easily be caused by chance. However, one might have expected probably an opposite trend with worse diagnostic performance for subjective parameters. It is possible that the diagnostic performance of the qualitative evaluations might have been spuriously inflated if the evaluators had not been unaware of the clinical and histologic information of each subject. In fact, although we standardized the cutoffs for SUV and MRG in the meta-analysis, the qualitative evaluations used a variety of rules across studies. With the exception of studies using TBR, these rules were largely subjective and depended on consensus between radiologists. Post hoc modulation of diagnostic thresholds may also spuriously inflate the estimated performance of a diagnostic test. Furthermore, it is unknown whether some publication bias may be operating against the publication of studies that may have found less promising results (30). Finally, there was a large scatter of sensitivity and specificity values with the qualitative rules used. Simple qualitative interpretation is unlikely to offer a generalized standard for routinely interpreting ^{18}F-FDG PET scans across different centers.

SUV and MRG are more objective measures in this regard. For SUV, a cutoff of 2.0 was probably preferable to a cutoff of 3.0 in adequately discriminating between benign and malignant lesions. Nevertheless, even the 2.0 cutoff will give false-positive results for almost one fifth of the benign lesions, whereas it will miss two thirds of low-grade tumors. MRG had an even worse performance and, overall, was not a good discriminating parameter, although data were very limited. Although we had to exclude 1 MRG study with nonseparable data on bone and soft-tissue lesions (10,11), the composite data seem similar to what we observed in the remaining studies. Similarly, neither SUV nor MRG is perfect in assigning tumor grade to a malignant lesion. We should acknowledge, however, that tumor grade in STS may sometimes be difficult to establish even with histologic examination. Thus, even a perfect imaging test may spuriously seem to have less than perfect concordance with histologic readings. By measuring the metabolic activity of a lesion, it is possible that ^{18}F-FDG PET values may offer additional long-term prognostic information besides histology. However, no long-term prognostic analyses have been performed with ^{18}F-FDG PET on STS to date.

There are no good quality data on the comparative diagnostic performance of ^{18}F-FDG PET against CT scans or MRI for the diagnosis of primary soft-tissue lesions, because all studies to date have used these imaging tests in series, not in parallel. There is limited evidence suggesting an approximately equivalent diagnostic performance of ^{18}F-FDG PET and MRI for assessing local recurrence and even more limited evidence suggesting approximately equivalent performance of PET and CT scan in searching for distant metastatic disease. The one study of adequate sample size pertains to lung metastatic disease. ^{18}F-FDG PET has been considered to be useful technique for the evaluation of lung nodules in general (31), and it is conceivable that this may apply also in the evaluation of lung lesions in patients with STS. There are very limited data on the usefulness of PET in assessing the response of STS to therapy. One study (16) suggests that tyrosine PET may be superior to ^{18}F-FDG PET for assessing response to therapy, but the data are equally limited. Thus, based on the current evidence, it is unclear whether PET can offer any advantage over traditional imaging modalities. There is considerable enthusiasm about the ability of ^{18}F-FDG PET to assess the response of various tumors to therapy (1,32), but more data are needed for STS.

By assembling a large number of subjects, the meta-analysis has managed to estimate the diagnostic accuracy of ^{18}F-FDG PET in STS diagnosis and grading, decreasing the uncertainty inherent in isolated case series. ^{18}F-FDG PET seems to be a promising technology, and its diagnostic performance is very good. The ability to gain insight into the metabolic parameters of a lesion with ^{18}F-FDG PET has generated extensive enthusiasm (1,2,32). However, despite the experimental enthusiasm and the wide field of potential applications (1,2), routine use needs to be scrutinized carefully (33). The incremental value of ^{18}F-FDG PET against other imaging modalities and its proper role in the management and prognostic assessment of STS remain largely unknown.

CONCLUSION

^{18}F-FDG PET is a useful imaging modality in the evaluation of both primary and recurrent soft-tissue lesions. ^{18}F-FDG PET may be clinically helpful in tumor grading, whereas it may not always offer adequate discrimination between low-grade tumors and benign lesions. Future research should better clarify the incremental diagnostic yield and long-term clinical benefit derived from the use of ^{18}F-FDG PET when compared with or added to other imaging modalities. The cost-effectiveness of different imaging strategies would also need to be delineated, but it is hoped

that by better diagnosis, grading, and staging, PET may actually reduce the cost and complications of invasive diagnostic and therapeutic interventions in patients with various forms of cancer (*2,34*), including STS. Finally, ^{18}F-FDG may need to be compared against other newer radiopharmaceuticals (*35,36*) in STS diagnosis.

REFERENCES

1. Delbeke D, Martin WH. Positron emission tomography imaging in oncology. *Radiol Clin North Am.* 2001;39:883–917.
2. Eary JF. Nuclear medicine in cancer diagnosis. *Lancet.* 1999;354:853–857.
3. National Cancer Institute. SEER Cancer Statistics Review. Available at: http://seer-cancer.gov. Accessed on February 26, 2003.
4. Pisters PW, Leung DH, Woodruff J, Shi W, Brennan MF. Analysis of prognostic factors in 1041 patients with localized soft tissue sarcomas of the extremities. *J Clin Oncol.* 1996;14:1679–1689.
5. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* 1993;12:1293–1316.
6. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making.* 1993;13:313–321.
7. El-Zeftway H, Heiba SI, Jana S, et al. Role of repeated F-18 fluorodeoxyglucose imaging in management of patients with bone and soft tissue sarcoma. *Cancer Biother Radiopharmaceut.* 2001;16:37–46.
8. Shulkin BL, Mitchell DS, Ungar DR, et al. Neoplasms in a pediatric population: 2-[F-18]-fluoro-2-deoxy-D-glucose PET studies. *Radiology.* 1995;194:495–500.
9. 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;28:1541–1551.
10. Eary JF, Conrad EU, Bruckner JD, et al. Quantitative [F-18]fluorodeoxyglucose positron emission tomography in pretreatment and grading of sarcoma. *Clin Cancer Res.* 1998;4:1215–1220.
11. Folpe AL, Lyles RH, Sprouse JT, Conrad EU III, Eary JF. (F-18) fluorodeoxyglucose positron emission tomography as a predictor of pathologic grade and other prognostic variables in bone and soft tissue sarcoma. *Clin Cancer Res.* 2000;6:1279–1287.
12. Schwarzbach M, Willeke F, Dimitrakopoulou-Strauss A, et al. Functional imaging and detection of local recurrence in soft tissue sarcomas by positron emission tomography. *Anticancer Res.* 1999;19:1343–1349.
13. Dimitrakopoulou-Strauss A, Strauss LG, Schwarzbach M, et al. Dynamic PET ^{18}F-FDG studies in patients with primary and recurrent soft-tissue sarcomas: impact on diagnosis and correlation with grading. *J Nucl Med.* 2001;42:713–720.
14. Schwarzbach MH, Dimitrakopoulou-Strauss A, Willeke F, et al. Clinical value of [18-F] fluorodeoxyglucose positron emission tomography imaging in soft tissue sarcomas. *Ann Surg.* 2000;231:380–386.
15. Kole AC, Nieweg OE, van Ginkel RJ, et al. Detection of local recurrence of soft-tissue sarcoma with positron emission tomography using [^{18}F]fluorodeoxyglucose. *Ann Surg Oncol.* 1997;4:57–63.
16. Kole AC, Plaat BEC, Hoekstra HJ, Vaalburg W, Molenaar WM. FDG and L-[1–11C]-tyrosine imaging of soft-tissue tumors before and after therapy. *J Nucl Med.* 1999;40:381–386.
17. Nieweg OE, Pruim J, van Ginkel RJ, et al. Fluorine-18-fluorodeoxyglucose PET imaging of soft-tissue sarcoma. *J Nucl Med.* 1996;37:257–261.
18. van Ginkel RJ, Hoekstra HJ, Pruim J, et al. FDG-PET to evaluate response to hyperthermic isolated limb perfusion for locally advanced soft-tissue sarcoma. *J Nucl Med.* 1996;37:984–990.
19. Hain SF, O'DohertyJ, Lucas JD, et al. Fluorodeoxyglucose PET in the evaluation of amputation for soft tissue sarcoma. *Nuclear Med Commun.* 1999;20:845–848.
20. Ferner RE, Lucas JD, O'Doherty MJ, et al. Evaluation of 18fluorodeoxyglucose 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;68:353–357.
21. 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.
22. Lucas JD, O'Doherty MJ, Wong JC, et al. Evaluation of fluorodeoxyglucose positron emission tomography in the management of soft-tissue sarcomas. *J Bone Joint Surg Br.* 1998;80:441–447.
23. Lucas JD, O'Doherty MJ, Cronin BF, et al. Prospective evaluation of soft tissue masses and sarcomas using fluorodeoxyglucose positron emission tomography. *Br J Surg.* 1999;86:550–556.
24. Jones DN, McCowage GB, Sostman HD, et al. Monitoring of neoadjuvant therapy response of soft-tissue and musculoskeletal sarcoma using fluorine-18-FDG PET. *J Nucl Med.* 1996;37:1438–1444.
25. Griffeth LK, Dehdashti F, McGuire AH, et al. PET evaluation of soft-tissue masses with fluorine-18 fluoro-2-deoxy-D-glucose. *Radiology.* 1992;182:185–194.
26. Kern KA, Brunetti A, Norton JA, et al. Metabolic imaging of human extremity musculoskeletal tumors by PET. *J Nucl Med.* 1988;29:181–186.
27. Schulte M, Brecht-Krauss D, Heymer B, 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.
28. Adler LP, Blair HF, Williams RP, et al. Grading liposarcomas with PET using [^{18}F]FDG. *J Comput Assist Tomogr.* 1990;14:960–962.
29. Watanabe H, Shinozaki T, Yanagawa 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;82:760–767.
30. Easterbrook P, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* 1991;337:867–872.
31. Lowe VJ, Fletcher JW, Gobar L, et al. Prospective investigation of positron emission tomography in lung nodules. *J Clin Oncol.* 1998;16:1075–1084.
32. Price P, Jones T. Can positron emission tomography (PET) be used to detect subclinical response to cancer therapy? The EC PET Oncology Concerted Action and the EORTC PET Study Group. *Eur J Cancer.* 1995;31:1924–1927.
33. Balk EM, Lau J. PET scans and technology assessment: déjà vu? *JAMA.* 2001;285:936–937.
34. Gambhir SS, Hoh CK, Phelps ME, Madar I, Maddani J. Decision tree sensitivity analysis for cost effectiveness of FDG PET in the staging and management of non-small-cell lung carcinoma. *J Nucl Med.* 1996;37:1428–1436.
35. Shields AF, Mankoff DA, Link JM, et al. [C-11]thymidine and FDG to measure therapy response. *J Nucl Med.* 1998;39:1757–1762.
36. Shields AF, Grierson JR, Dohmen BM, et al. Imaging proliferation in vivo with [F-18]FLT and positron emission tomography. *Nat Med.* 1998;4:1334–1336.