

Comparison of ¹⁸F-FDG PET/CT and ultrasound in staging of patients with malignant melanoma

Philipp Weber, MD, MSc^{a,*} , Andreas Arnold, MD^{b,c}, Joachim Hohmann, MD, MSc^{c,d}

Abstract

Objective: To evaluate the sensitivity and specificity of 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography (¹⁸F-FDG PET/CT) and ultrasound (US) for staging patients with malignant melanoma.

Methods: In total, 258 patients (112 men and 146 women; mean age, 61 ± 16years) met the primary inclusion criteria for malignant melanoma without further malignancy proven by histopathology. This was a retrospective study of the diagnostic accuracy. All data were obtained from the hospital's patient and radiology information system. Patients formed a consecutive series and were examined by ¹⁸F-FDG PET/CT and 176 additionally by US (US as a whole [wUS], peripheral lymph nodes [pUS], abdomen [aUS]), with a total of 584 ¹⁸F-FDG PET/CT and 697 US. ¹⁸F-FDG PET/CT and US revealed 824 and 726 lesions, respectively. Per-patient, per-examination, and per-lesion analyses were also performed. The reference standards used were histopathology or resection of lesions, and follow-up controls using other imaging methods.

Results: Significant differences ($P < .05$) were found in the per-examination for the sensitivity of ¹⁸F-FDG PET/CT (0.80) compared to wUS (0.63) and pUS (0.61), and the specificity of ¹⁸F-FDG PET/CT (0.96) compared to wUS (0.98) and aUS (0.99). In the PLA, there were significant differences in sensitivity and specificity for ¹⁸F-FDG PET/CT (0.83, 0.91) compared to wUS (0.61, 0.98), pUS (0.60, 0.98), and aUS (0.61, 0.99).

Conclusion: ¹⁸F-FDG PET/CT is preferable to US for detecting both lymph node and abdominal metastases.

Abbreviations: ¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, AJCC = American Joint Committee on Cancer, pUS = peripheral lymph nodes, ASR(W) = age-standardized rate per 100,000 inhabitants per year using the world standard population, aUS = abdomen, M = organ metastases, MM = malignant melanoma, N = lymph nodes, PEA = per-examination, PLA = per-lesion, PPA = per-patient, RIS = radiological information system, T = tumor, US = ultrasound, wUS = whole.

Keywords: ¹⁸F-FDG PET/CT, malignant melanoma, staging, ultrasound

1. Introduction

1.1. Epidemiology

The incidence of malignant melanoma (MM) has steadily increased. In Switzerland, the incidence of MM among men increased by 83.3% in the 4-year periods from (1988, 1992) to (2013–2017) and among women by 63.5%.^[1] In 2020, the incidence of melanoma in Switzerland was 21.6 ASR(W) (age-standardized rate per 100,000 inhabitants per year using the world standard population). This was the sixth highest incidence of cancer in Switzerland in 2020.^[2] In the same year, it reached 11.4 ASR(W) for both sexes in Europe and 3.4 ASR(W) worldwide. In 2020, 324,635 new cases of MM were reported worldwide.

This accounts for 1.7% of all cancer cases worldwide. The sex distribution was f:m = 1:1.15.

1.2. Classification and staging of MM

MM is classified according to the TNM system,^[3] in which the extent of tumor (T), infestation of lymph nodes (N), and presence of organ metastases (M) are described (Table 1). Tumor stage according to the TNM system at initial diagnosis primarily determines the prognosis and therapy of MM.^[4]

The T stage, and thus, the depth of tumor invasion, is identified by histopathological examination after biopsy or tumor resection. Radiological/nuclear medical imaging detects lymph

This study was funded by the Krebsliga beider Basel, Basel, Switzerland (grant number 01-2013).

The authors declare no conflict of interest.

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

All relevant data are within the paper and its supporting information files.

Trial registration: ClinicalTrials.gov identifier: NCT04253574 5, February 2020 (retrospectively registered).

^a Department of Radiology and Nuclear Medicine, University Hospital of Basel, Basel, Switzerland, ^b Group Practice "Dermatology on the Rhine", Basel, Switzerland, ^c Faculty of Medicine, University of Basel, Basel, Switzerland, ^d Institute of Radiology and Nuclear Medicine, Cantonal Hospital of Winterthur, Winterthur, Switzerland.

* Correspondence: Philipp Weber, MD, MSc, Department of Radiology and Nuclear Medicine, University Hospital of Basel, Petersgraben 4, CH-4031 Basel, Switzerland (e-mail: ph.weber70@gmail.com).

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Weber P, Arnold A, Hohmann J. Comparison of ¹⁸F-FDG PET/CT and ultrasound in staging of patients with malignant melanoma. *Medicine* 2022;101:42(e31092).

Received: 24 July 2022 / Received in final form: 9 September 2022 / Accepted: 13 September 2022

<http://dx.doi.org/10.1097/MD.0000000000031092>

Table 1**TNM system for the classification of MM.^[3]**

Expansion of the tumor: T		Lymph nodes: N		Organ metastases: M	
Tis	Melanoma in situ	N0	No affected lymph nodes	M0	No organ metastases
T0	No tumor present	N1	1 affected lymph node	M1	Metastases in i.e. lung, liver
T1	Thickness < 1 mm	N2	2 to 3 affected lymph nodes	Mx	Existence of organ metastases not assessable
T2	1 mm < thickness < 2 mm	N3	4 or more affected lymph nodes		
T3	2 mm < thickness < 4 mm	Nx	Affection of lymph nodes not assessable		
T4	Thickness > 4 mm				
Tx	Tumour extent not assessable				

MM = Malignant melanoma.

node involvement (N stage) and the presence of organ metastases (distant metastases, M Stage).

Multiple TNM combinations with similar therapy and prognosis are classified into 5 stages by the American Joint Committee on Cancer (AJCC).^[3] Here, stages 0 (Tis, N0, M0), I (T1a-T2a, N0, M0), and II (T2b-T4b, N0, M0) are defined by the thickness of the primary tumor and its possible ulceration (a/b). Stage III (any T, \geq N1, M0) is determined by evidence of lymph node metastases and stage IV (any T, any N, M1) by presence of organ metastases.

1.3. Imaging modalities for the staging of MM

Determining which imaging modality offers the highest sensitivity and specificity for primary staging and restaging of MM and to what extent a combination of different imaging modalities is required, is a controversial issue in current discussions.^[4,5] The question of a necessary combination of several radiological imaging modalities is also of health economic relevance.

Patients with MM undergo either 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography (¹⁸F-FDG PET/CT) and/or ultrasound (US) of the peripheral lymph nodes (pUS) (cervical, axillary, and inguinal) and the abdomen (aUS) for (re)staging. Previous studies have investigated both modalities in different patient populations.^[4,5] At our clinic, most patients with MM have undergone a combined examination using both modalities for several years. This makes it possible to evaluate both modalities in the same patient population, and thus, to close this data gap.

1.4. Objective

This study directly compared the 2 imaging modalities, ¹⁸F-FDG PET/CT and US (as a whole (wUS), pUS, and aUS) for staging and restaging of MM in the same patient population. Per-patient (PPA), per-examination (PEA), and per-lesion (PLA) analyses were retrospectively performed using the existing data from the hospital's patient and radiology information system (RIS).

The primary question is the principle equivalence of the 2 imaging methods, which is established in all 3 analyses as a null hypothesis H_0 . As a further null hypothesis H_0 , the equivalence of the combined application compared with the individual applications of the 2 imaging modalities is asserted. The goal is to reject the null hypothesis, H_0 , using significant differences in sensitivity and specificity.

2. Materials and Methods

The present study was approved by the Northwestern and Central Swiss Ethics Committees (Ref. No. EK: 243/12). Informed consent for publication was obtained from all patients involved.

2.1. Patients

A total of 308 patients underwent 1 or more ¹⁸F-FDG PET/CT and/or 1 or more US scans between September 1998 and August 2014 for primary staging or restaging of suspected or confirmed MM, which defined the inclusion criteria for the study. Exclusion criteria were the absence of a malignancy or existing malignancy other than MM, alone or in combination with MM. 50 patients were excluded from the study, as shown in the flow diagram (Fig. 1): 2 patients without malignancy, 10 patients with a different type of malignancy, 38 patients with MM and at least 1 other type of malignancy. This resulted in 258 patients (112 [43%] women and 146 [57%] men; mean age, 61 \pm 16 years) for evaluation within the study.

All 258 patients underwent at least 1 ¹⁸F-FDG PET/CT, 27 (10%) as primary staging and restaging, 22 (9%) as primary staging only, and 209 (81%) as restaging only.

Additionally, 176 patients were examined at least once by US (pUS and/or aUS), with 24 patients (14% of patients with at least 1 US) receiving both primary staging and restaging, 12 (7%) patients only primary staging, and 140 (79%) patients only restaging. The corresponding values were (167, 21 (13% of patients with at least 1 pUS), 9 (5%), 137 (82%)), and (107, 11 (10% of patients with at least 1 aUS), 15 (14%), 81 (76%)) for pUS and aUS, respectively.

2.2. Tumor classification

The melanoma subtypes were nodular (n = 68, 35% of the classified melanomas), superficial spreading (n = 61, 32%), acral lentiginous (n = 12, 6%), and lentigo maligna (n = 7, 4%) (Table 2).^[6] Forty-four (23%) patients had rare subtypes, unclassifiable melanomas, or mixed forms. 66 patients had MM, without any characterization.

At the initial diagnosis, 224 patients (87% of all included patients) had only a primary tumor and 34 (13%) patients showed metastases only. Only 9 (3%) of the 34 patients with metastasis had original primary tumors. In 18 (7%) patients, the primary tumor was unknown, and in another 7 (3%) patients no information was available on the primary tumor.

In 48 (19%) patients, a primary tumor or metastasis to the head or neck was found at the initial diagnosis, in 102 (39%) patients on the trunk, in 42 (16%) patients on the upper limbs, and in 66 (26%) patients on the lower limbs (Table 3).

Forty-one (16%) patients had stage I MM at initial diagnosis, according to the AJCC staging system.^[3] For stage II, III, and IV the number and percentage of patients included in the study were 87 (34%), 54 (21%), and 8 (3%), respectively. The MM of a single patient was not classifiable by stage, and in 67 (26%) patients, data from the RIS did not include the disease stage at initial diagnosis.

2.3. Nuclear medical and radiological modalities

PET/CT was performed using 2-18fluoro-2-deoxy-D-glucose (¹⁸F-FDG) as a radiopharmaceutical. pUS was conducted using a linear transducer, whereas aUS was performed using a convex transducer.

2.4. Findings

The ¹⁸F-FDG PET/CT examinations were either reported by a nuclear medicine resident and visualized by a nuclear medicine

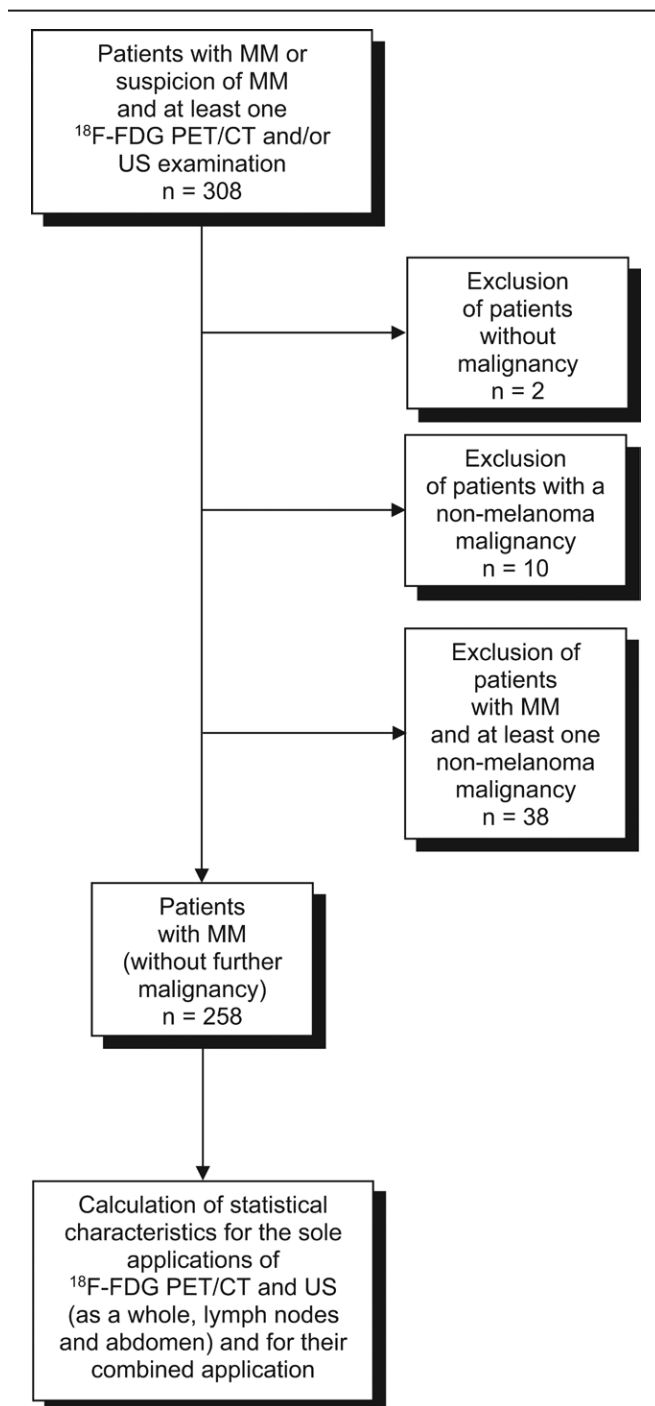


Figure 1. Flow diagram to include patients in the study. MM = Malignant melanoma.

specialist or, alternatively, the reporting was initially performed by a nuclear medicine specialist. Radiological reporting on ^{18}F -FDG PET/CT was either countersigned by a specialist in diagnostic radiology or executed by a dual specialist in diagnostic radiology and nuclear medicine.

US examinations were performed and reported by either a radiology resident or diagnostic radiology specialist. When an initial US examination was performed by a resident, it was performed by a specialist in diagnostic radiology.

All investigators had full access to the results of previous examinations of the patients.

Table 2

Characterisation of the MMs of the patients included: Melanoma subtype and melanoma lesion at first diagnosis.

Melanoma subtype	
Nodular	68 (35%)
Superficial-spreading	61 (32%)
Acral lentiginous	12 (6%)
Lentigo maligna	7 (4%)
Other (e.g. mixed forms)	44 (23%)
Without characterization	66
Lesion at first diagnosis	
Primary tumor	224 (87%)
Metastasis	34 (13%)
•Primary tumor known	9 (3%)
•Primary tumor unknown	18 (7%)
•No information on the primary tumor	7 (3%)

MM = Malignant melanoma.

Table 3

Characterisation of the MMs of the patients included: Localisation of melanoma lesion and stage of melanoma according to AJCC staging^[3] at first diagnosis.

Localisation at first diagnosis, either as primary tumor or metastasis	
Head and neck	48 (19%)
Trunk (incl. axillary and inguinal)	102 (39%)
Upper limbs	42 (16%)
Lower limbs	66 (26%)
Initial stage according to AJCC Staging ^[3]	
Stage I	41 (16%)
Stage II	87 (34%)
Stage III	54 (21%)
Stage IV	8 (3%)
Not classifiable	1 (0%)
Not specified	67 (26%)

AJCC = American Joint Committee on Cancer, MM = Malignant melanoma.

2.5. Evaluation criteria

In ^{18}F -FDG PET/CT examinations, increased ^{18}F FDG uptake by the lymph nodes and organ lesions was regarded as a positive criterion.

In the US, a short axis diameter of lymph nodes greater than 1 cm was considered a positive criterion, as was hypoechogenicity of the lymph nodes, loss of fatty hilum, and a blurred or irregular border. In organ lesions, hypoechogenic presentation, hypoechogenic rim (halo), and space-occupying and/or infiltrative components were also criteria for positive assessment.

New lesions in both modalities were per se evaluated as positive.

2.6. Reference standard

The reference standard for the results of ^{18}F -FDG PET/CT and US was histopathological clarification or resection of an individual lesion (primary tumor, recurrences, lymph node metastases, organ metastases), as well as follow-up controls using other imaging methods such as computed tomography (CT), magnetic resonance imaging, or scintigraphy within a period of 3 months before to 3 months after an examination.

By comparing the findings of ^{18}F -FDG PET/CT or US with the reference standard, each positive finding was classified as a true positive or false positive. In the absence of a reference standard, a positive finding of ^{18}F -FDG PET/CT or US was assessed as “unknown”. If both ^{18}F -FDG PET/CT and US were positive with regard to a lesion within a period of 3 months, these 2

findings were evaluated as true positive in the sense of mutual referencing.

If ^{18}F -FDG PET/CT and/or US showed no positive results, these examinations were classified as false negative in the presence of a reference standard; otherwise, they were classified as true negative.

2.7. Data analysis

2.7.1. PPA. Within the PPA, each patient was assigned for the evaluation of his or her first examination of the 3 radiological modalities: ^{18}F -FDG PET/CT, pUS, and aUS.

The first examination, which included all lesions identified in the same way, also received this evaluation.

If a reference standard was available for all lesions in the first examination, the examination was considered true positive if at least 1 of the lesions was found to be true positive. For the examination to be considered a false negative, all of its lesions must have been detected as false negative.

If there was no reference standard for a negative examination, it was considered as a true negative. However, if there were lesions in an examination for which no reference standard existed, the examination was considered false positive.

If an initial examination included both lesions for which a reference standard existed and lesions for which no reference standard existed, the examination was considered to be true positive if at least 1 lesion was found to be true positive. If there were no lesions determined to be true positive and the examination included at least 1 lesion found to be false positive, the examination was designated as false positive. For a false negative evaluation of an examination, lesions were only allowed to be identified as false negative.

If an imaging modality at the initial examination did not present any findings substantiated by a reference standard, this was noted as “unknown only” for the corresponding patient. If neither ^{18}F -FDG PET/CT nor pUS/aUS was performed on a patient, this was also noted. Consequently, if only non-referenced findings or non-performed radiological modalities were available for a patient, no data were included in the preparation of the PPA for this patient.

2.7.2. PEA. Within the PEA, each examination of the 3 radiological modalities ^{18}F -FDG PET/CT, pUS, and aUS was assigned a single evaluation. The same evaluation scheme was applied as in the PPA.

2.7.3. PLA. Within the PLA, ^{18}F -FDG PET/CT and US findings were evaluated at the level of individual lesions. If no reference standard was available for a lesion, it was evaluated as “unknown”. These data were not included in the PLA.

2.8. Statistics

The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy for ^{18}F -FDG PET/CT and US (wUS, pUS, and aUS) were calculated using the evaluations obtained by data management for all patients,

examinations and lesions. The 95% confidence intervals (CIs) for the binomial distributions were determined using the Wilson method.^[7]

Within the PPA, the differences between the sensitivities and specificities of ^{18}F -FDG PET/CT and US (wUS, pUS, and aUS) were checked for significance (significance level $\alpha = .05$) using the 2-sided McNemar test for associated samples.^[8] For the 2×2 cross tables, a degree of freedom of 1 was found.^[9] Due to the partially small number of cases, a Yates correction of 0.5 was applied.^[10]

Fisher's exact test was used within the PEA and PLA to evaluate differences between the sensitivities and specificities of ^{18}F -FDG PET/CT and US (wUS, pUS, and aUS) for significance (significance level $\alpha = .05$).^[11]

2.9. Combined application of ^{18}F -FDG PET/CT and ultrasound

In addition, the results of the individual lesions were analyzed in terms of the combined use of ^{18}F -FDG PET/CT and US. An attempt was made to assign each lesion analyzed by ^{18}F -FDG PET/CT to pUS or aUS within a period of 3 months before to 3 months after the ^{18}F -FDG PET/CT examination. The pUS or aUS examination with the smallest time interval to the ^{18}F -FDG PET/CT scan was considered.

Combined lesion analysis provided a new evaluation scheme (Table 4). Combinations of identical evaluations of the individual applications resulted in the same evaluation for the combined application. Combinations of different single evaluations were evaluated in such a way, that a single true positive or true negative evaluation was sufficient to obtain a true positive or true negative evaluation for the combination of ^{18}F -FDG PET/CT and US.

In the analysis of the combined use of ^{18}F -FDG PET/CT and US (wUS, pUS, and aUS), the sensitivity, specificity, PPV, NPV, and accuracy were determined. These statistical values were also calculated for the same lesions for the separate applications of ^{18}F -FDG PET/CT and US (wUS, pUS, and aUS).

To test for significant differences between the sensitivities or specificities of the combined applications of ^{18}F -FDG PET/CT and US (wUS, pUS, and aUS) and the sole applications the 2-sided McNemar test for paired samples was performed for a significance level $\alpha = .05$.^[8] The cross tables used for this purpose had 2 columns and 2 rows (2×2 cross tables). This resulted in degrees of freedom of 1.^[9] Due to the partially small number of cases, a Yates correction of 0.5 was used again.^[10]

3. Results

3.1. Per-patient analysis (PPA)

Of 258 patients included in the study, 245 underwent ^{18}F -FDG PET/CT as part of the PPA. The remaining 13 patients underwent a first ^{18}F -FDG PET/CT scan with lesions that could not be confirmed by the reference standard (“unknown only”). Of the 245 patients evaluated, 47 underwent primary staging with ^{18}F -FDG PET/CT and 198 underwent restaging with ^{18}F -FDG PET/CT.

Table 4

Evaluation scheme of the combined application of ^{18}F -FDG PET/CT and ultrasound (as a whole, lymph nodes and abdomen) with the same (concordance) or different (discordance) evaluations of the sole examinations.

Examination	Concordance				Discordance			
^{18}F-FDG PET/CT	True positive	False positive	True negative	False negative	True positive	False positive	True negative	False negative
Ultrasound	True positive	False positive	True negative	False negative	False negative	True negative	False positive	True positive
Combined	True positive	False positive	True negative	False negative	True positive	True negative	True negative	True positive

^{18}F -FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography.

Primary US was evaluated 274 times in 176 patients, with 98 patients receiving primary pUS and primary aUS, 69 patients receiving primary pUS only, and 9 patients receiving primary aUS only. One primary pUS demonstrated lesions that could not be proven using the reference standard. Consequently, 273 US examinations were included in the PPA, 54 as primary staging, and 219 as restaging. The corresponding figures for pUS and aUS were (166, 29, 137) and (107, 25, 82), respectively.

¹⁸F-FDG PET/CT showed the highest sensitivity and a similarly high specificity as US (wUS, pUS, and aUS), the highest PPV, but the lowest NPV, and a similarly high accuracy (Table 5).

There were no significant differences between the sensitivities and specificities in any of the comparisons, whereas the difference between the sensitivities of ¹⁸F-FDG PET/CT and wUS was only slightly insignificant

($P = .052$) (Table 6). Consequently, the null hypothesis H_0 for the equivalence of the methods cannot be rejected in the PPA.

Table 5
Statistical characteristic values of the per-patient analysis.

Statistical characteristic value	¹⁸ F-FDG PET/CT	Ultrasound		
		as a whole	lymph nodes	abdomen
n	245	273	166	107
Sensitivity	0.71 [0.61; 0.79]	0.48 [0.34; 0.62]	0.46 [0.32; 0.61]	0.56 [0.27; 0.81]
Specificity	0.96 [0.91; 0.98]	0.97 [0.94; 0.99]	0.97 [0.92; 0.99]	0.98 [0.93; 0.99]
PPV	0.92 [0.84; 0.96]	0.79 [0.62; 0.90]	0.82 [0.61; 0.93]	0.71 [0.36; 0.92]
NPV	0.82 [0.75; 0.87]	0.90 [0.85; 0.93]	0.85 [0.79; 0.90]	0.96 [0.90; 0.98]
Accuracy	0.85 [0.80; 0.89]	0.89 [0.84; 0.92]	0.85 [0.79; 0.90]	0.94 [0.88; 0.97]

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, n = number of referenced first examinations, NPV = negative predictive value, PPV = positive predictive value.

Table 6
Test for significance between sensitivities and specificities of the per-patient, per-examination and per-lesion analysis (significance level $\alpha = .05$).

Pro-patient analysis		
	18F-FDG PET/CT	
US as a whole	Sensitivity $P = .052$	Specificity $P = .85$
US lymph nodes	$P = .080$	$P = .84$
US abdomen	$P = .45$	$P = .45$
Pro-examination analysis		
	18F-FDG PET/CT	
US as a whole	Sensitivity $P = .0018$	Specificity $P = .014$
US lymph nodes	$P = .0016$	$P = .067$
US abdomen	$P = .30$	$P = .026$
Pro-lesion analysis		
	18F-FDG PET/CT	
US as a whole	Sensitivity $P < .001$	Specificity $P < .001$
US lymph nodes	$P < .001$	$P < .001$
US abdomen	$P = .03$	$P < .001$

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, US = ultrasound.

3.2. PEA

As part of the PEA, ¹⁸F-FDG PET/CT was performed 606 times on all 258 patients included in the study, 50 of which were primary staging and 556 were restaging examinations. On average, each patient received $n = 2.3$ times a ¹⁸F-FDG PET/CT. In 22 cases, ¹⁸F-FDG PET/CT only contained lesions for which no reference standard existed. Consequently, 584 ¹⁸F-FDG PET/CT examinations were included in the PEA.

US was executed a total of 707 times, 61 times for primary staging, and 646 times for restaging. In total, 176 patients were included in this study. On average, each patient received $n = 4.0$ times a US. This resulted in corresponding figures for pUS (478, 33, 445, 167, 2.7) and aUS (229, 28, 201, 107, 2.0).

In 10 cases, US produced only lesions without a reference standard, which is why 697 of the 707 US examinations were finally considered for the PEA. The appropriate values for pUS and aUS were (478, 8, 470) (229, 2, and 227), respectively.

¹⁸F-FDG PET/CT showed the highest sensitivity, lowest specificity, highest PPV, lowest NPV, and lowest accuracy compared to the corresponding US values (wUS, pUS, and aUS) (Table 7).

In the PEA, the sensitivity of ¹⁸F-FDG PET/CT was significantly higher than that of wUS and pUS for a significance level of $\alpha = .05$ (Table 6). No significant difference was observed between the sensitivities of ¹⁸F-FDG PET/CT and aUS. The specificity of ¹⁸F-FDG PET/CT in the PEA was significantly lower than that of wUS and aUS. The specificities of ¹⁸F-FDG PET/CT and pUS did not differ significantly. The null hypothesis H_0 of the equivalence of the methods can be rejected in the context of the PEA for ¹⁸F-FDG PET/CT and wUS because of the significant differences between the sensitivities and specificities. For the comparison between ¹⁸F-FDG PET/CT and pUS, this applies only to sensitivity, and for the comparison between ¹⁸F-FDG PET/CT and aUS, only to specificity.

3.3. PLA

In the PLA, 1108 lesions were examined by ¹⁸F-FDG PET/CT, of which 98 lesions were studied during primary staging and 1010 lesions during restaging. ¹⁸F-FDG PET/CT was performed 606 times for all 258 patients. On average, this resulted in $n = 2.3$ ¹⁸F-FDG PET/CT examinations per patient as well as in $n = 1.8$ lesions per ¹⁸F-FDG PET/CT and $n = 4.3$ lesions per patient. In total, 284 lesions could not be verified using the reference standard. Consequently, 824 lesions were included in the PLA for ¹⁸F-FDG PET/CT.

Table 7
Statistical characteristic values of the per-examination analysis.

Statistical characteristic value	¹⁸ F-FDG PET/CT	Ultrasound		
		As a whole	Lymph nodes	Abdomen
n	584	697	470	227
Sensitivity	0.80 [0.75; 0.85]	0.63 [0.52; 0.72]	0.61 [0.50; 0.72]	0.69 [0.42; 0.87]
Specificity	0.96 [0.93; 0.97]	0.98 [0.97; 0.99]	0.98 [0.96; 0.99]	0.99 [0.97; 1.00]
PPV	0.91 [0.86; 0.94]	0.85 [0.74; 0.91]	0.85 [0.73; 0.92]	0.82 [0.52; 0.95]
NPV	0.90 [0.87; 0.93]	0.95 [0.93; 0.96]	0.93 [0.90; 0.95]	0.98 [0.95; 0.99]
Accuracy	0.90 [0.88; 0.92]	0.94 [0.92; 0.95]	0.92 [0.89; 0.94]	0.97 [0.94; 0.99]

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, n = number of referenced examinations, NPV = negative predictive value, PPV = positive predictive value.

Using US, 748 lesions were assessed, of which 63 in the context of primary staging and 685 in the context of restaging. In total, 707 US examinations were performed in 176 patients. This resulted in corresponding figures for pUS (511, 35, 476, 478, 167) and aUS (237, 28, 209, 229, 107).

On average, $n = 4.0$ US examinations per patient, $n = 1.1$ lesions per US as well as $n = 4.3$ lesions per patient were identified. The corresponding values for pUS and aUS were (2.7, 1.1, and 3.1) and (2.0, 1.0, and 2.2), respectively.

For wUS, 22 lesions could not be verified using a reference standard. Thus, for wUS 726 of totally 748 lesions were considered for the PLA. For pUS and aUS, the respective values were (511, 19, and 492) and (237, 3, and 234) respectively.

Mutual referencing was only possible for 2 lesions without a reference standard, which were identified as positive by ^{18}F -FDG PET/CT and pUS.

^{18}F -FDG PET/CT showed the highest sensitivity, lowest specificity, highest PPV, lowest NPV, and lowest accuracy (Figs. 2, 3 and Table 8) compared to the appropriate values of US (wUS, pUS, and aUS).

The sensitivity of ^{18}F -FDG PET/CT in the PLA was significantly higher than that of US (wUS, pUS, and aUS) ($\alpha = .05$) (Table 6). The specificity of ^{18}F -FDG PET/CT was significantly lower than that of US (wUS, pUS, and aUS). This can reject the null hypothesis H_0 of the equivalence of the methods used in the PLA.

3.4. Combined application of ^{18}F -FDG PET/ct and US within the PLA

A total of 362 ^{18}F -FDG PET/CT and wUS combinations were evaluated in this study. Of these, 264 combinations concerned

pUS and 98 were combinations with aUS. The mean time interval between ^{18}F -FDG PET/CT and US was 12.4 days. For the combination of ^{18}F -FDG PET/CT and pUS, this interval was 11.1 days and 16.1 days for the combination with aUS.

Of the 362 combinations, US examination was performed 155 times on a day before ^{18}F -FDG PET/CT, 101 times on the same day, and 106 times on a day after ^{18}F -FDG PET/CT. The corresponding values for the combination of ^{18}F -FDG PET/CT and pUS were (264, 101, 85, 78) and (98, 54, 16, 28) for the combination of ^{18}F -FDG PET/CT and aUS.

The combination of ^{18}F -FDG PET/CT and aUS achieved the highest sensitivity, the highest NPV, and the highest accuracy in comparison to the appropriate values for the combination of ^{18}F -FDG PET/CT and wUS, as well as to those for the combination of ^{18}F -FDG PET/CT and pUS (Fig. 4, Tables 9, 10 and 11). The specificity was the same for all 3 combinations. The combination of ^{18}F -FDG PET/CT and pUS showed the highest PPV.

The sensitivity of the combination of ^{18}F -FDG PET/CT and wUS was significantly higher ($\alpha = .05$) than that of the corresponding ^{18}F -FDG PET/CT examinations alone ($P = .044$) and that of the corresponding US examinations alone ($P < .001$) (Table 12). The specificity of the combination of ^{18}F -FDG PET/CT and wUS was significantly higher than that of the appropriate ^{18}F -FDG PET/CT examinations alone ($P = .044$) and that of the appropriate US examinations alone ($P = .014$). Based on these results, for the combination of ^{18}F -FDG PET/CT and wUS, the null hypothesis H_0 of the equivalence of the combined application and individual applications can be rejected in comparison to both ^{18}F -FDG PET/CT and wUS.

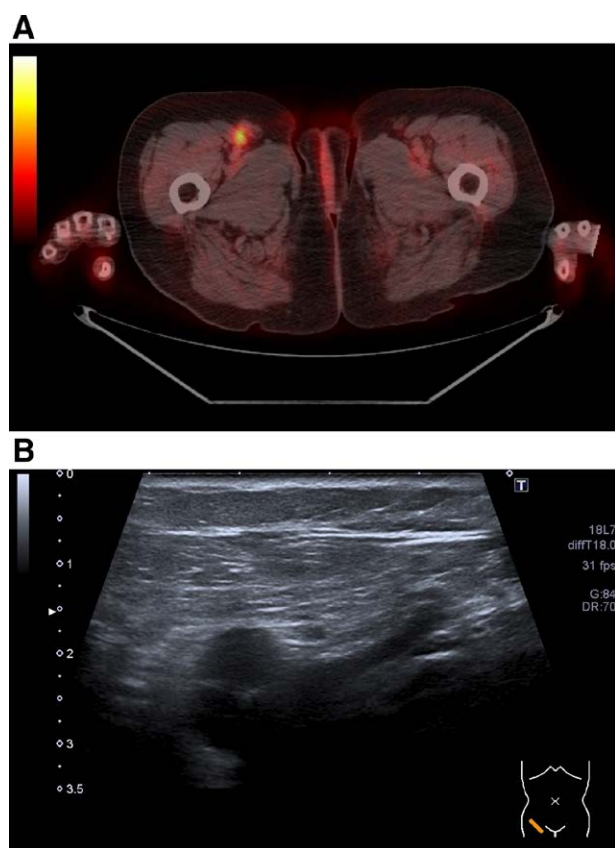


Figure 2. Lymph node metastasis inguinal right. (A) Detected as true positive by ^{18}F -FDG PET/CT. (B) Not evaluated as lymph node by US (false negative). ^{18}F -FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography.

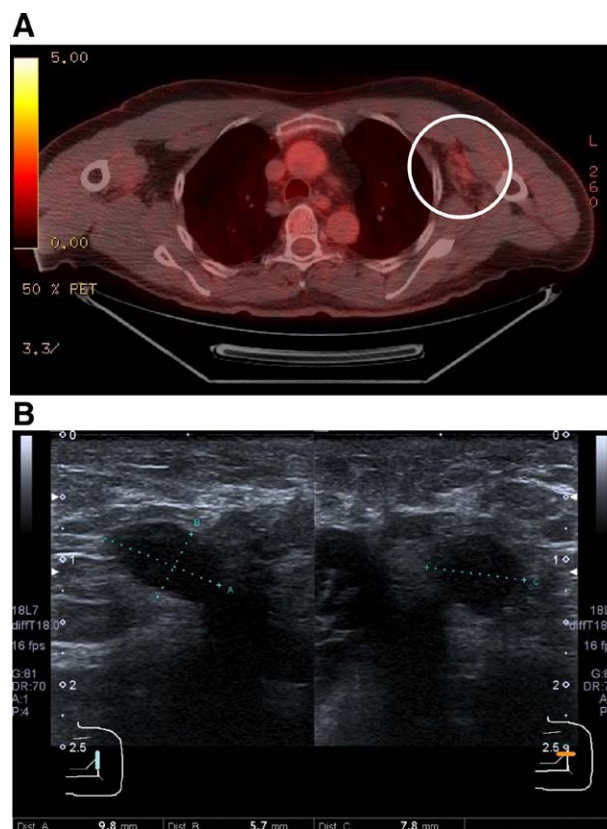


Figure 3. Lymph node metastasis axillary left (white circle). (A) With non-pathologically increased activity in ^{18}F -FDG PET/CT (false negative). (B) Clearly morphologically conspicuous in US (true positive). ^{18}F -FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography.

Table 8
Statistical characteristic values of the per-lesion analysis.

Statistical characteristic value	¹⁸ F-FDG PET/CT	Ultrasound		
		As a whole	Lymph nodes	Abdomen
n	824	726	492	234
Sensitivity	0.83 [0.79; 0.86]	0.61 [0.51; 0.69]	0.60 [0.50; 0.70]	0.61 [0.39; 0.80]
Specificity	0.91 [0.88; 0.93]	0.98 [0.97; 0.99]	0.98 [0.95; 0.99]	0.99 [0.97; 1.00]
PPV	0.91 [0.88; 0.93]	0.85 [0.75; 0.91]	0.85 [0.74; 0.91]	0.85 [0.58; 0.96]
NPV	0.83 [0.79; 0.86]	0.93 [0.91; 0.95]	0.92 [0.89; 0.94]	0.97 [0.94; 0.98]
Accuracy	0.87 [0.84; 0.89]	0.92 [0.90; 0.94]	0.91 [0.88; 0.93]	0.96 [0.93; 0.98]

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, n = number of referenced lesions, NPV = negative predictive value, PPV = positive predictive value.

The combination of ¹⁸F-FDG PET/CT and pUS had a significantly higher sensitivity than ¹⁸F-FDG PET/CT alone ($P = .044$) and pUS alone ($P < .001$). The same combination showed a significantly higher specificity than pUS alone ($P = .014$) but not than ¹⁸F-FDG PET/CT alone ($P = .080$). Consequently, the null hypothesis H_0 of the equivalence of the methods can be rejected for the combination of ¹⁸F-FDG PET/CT and pUS compared to pUS alone. Compared to ¹⁸F-FDG PET/CT alone, this is only possible for sensitivity.

The sensitivity of the combination of ¹⁸F-FDG PET/CT and aUS achieved the same value as that of the corresponding ¹⁸F-FDG PET/CT examinations alone (P not calculable) and was not significantly higher ($P = .15$) than that of the corresponding aUS examinations alone, although the latter had a broad 95% CI. The specificity of the combination of ¹⁸F-FDG PET/CT and aUS was exactly the same as that of the appropriate aUS examinations alone (P not calculable) and was not significantly higher than that of the appropriate ¹⁸F-FDG PET/CT examinations alone ($P = .62$). This means that for the combination of ¹⁸F-FDG PET/CT and aUS in comparison with the appropriate sole applications, the null hypothesis H_0 of the equivalence of the methods cannot be rejected.

4. Discussion

This study compared the detection of MM lesions using ¹⁸F-FDG PET/CT and US in the same patient population. This also enabled a combined evaluation of both radiological modalities. The previous gap in the related data could be closed.

4.1. General results

The sensitivity of ¹⁸F-FDG PET/CT was 0.71 (0.61; 0.79) within the PPA, 0.80 [0.75; 0.85] within the PEA, and 0.83 [0.79; 0.86] within the PLA.

In the “Final Report on Positron Emission Tomography (PET) and PET/CT in MM” of the German Institute for Quality and Efficiency in Health Care (IQWiG),^[4] 2 subgroup analyses are listed, 1 for patients with MM at a low stage (AJCC stages I and II,^[3] 6 primary studies) and 1 for patients with MM at an advanced stage (AJCC stages III and IV, 4 primary studies). Sensitivities from 0.00 to 0.17 are reported for the group of patients with MM at a low stage for PET and PET/CT in 4 primary studies, and of 0.67 and 1.00 for a further 2 primary studies. Sensitivities of 0.68 to 0.87 have been reported for patients with MM at an advanced stage.

In a meta-analysis by Xing et al,^[5] values of 0.11 [0.01; 0.50] for primary staging and 0.65 [0.20; 0.93] for restaging were reported for the sensitivity of PET/CT with regard to the detection of lymph node metastases. The corresponding values for organ metastases are listed as 0.80 [0.53, 0.93] and 0.86 [0.76; 0.93], respectively.

The sensitivities of ¹⁸F-FDG PET/CT determined in the present study are similar to those of PET or PET/CT in the IQWiG report and to those of PET/CT in the meta-analysis by Xing et al, with the exception of the 4 primary studies listed in the IQWiG report for the subgroup of patients with a MM at a low stage with sensitivities ranging between 0.00 to 0.17 and the detection of lymph node metastases using PET/CT in the context of primary staging with a sensitivity of 0.11 [0.01; 0.50] reported in the meta-analysis of Xing et al

The specificities of ¹⁸F-FDG PET/CT determined in the present study within the PPA (0.96 [0.91; 0.98]), PEA (0.96 [0.93; 0.97]) and PLA (0.91 [0.88; 0.93]) are comparable to those of PET and PET/CT in IQWiG’s final report (0.77–1.00 for patients at a low stage of MM and 0.92–0.98 for patients at an advanced stage of MM) as well as to those of PET/CT in the meta-analysis by Xing et al (0.97 [0.78; 1.00] for primary staging and 0.99 [0.92; 1.00] for restaging of lymph node metastases, 0.87 [0.54; 0.97] and 0.91 [0.79; 0.97] for the appropriate values regarding the detection of organ metastases).

For pUS, sensitivities were identified in the present study in the context of the PPA (0.46 [0.32; 0.61]), PEA (0.61 [0.50; 0.72]), and PLA (0.60 [0.50; 0.70]). In the meta-analysis by Xing et al the sensitivity for primary staging of lymph node metastases (0.60 [0.33; 0.83]) was - with a comparatively broad 95% CI - higher than the corresponding value for pUS in the PPA and practically the same as the corresponding values in the PEA and PLA. The sensitivity of restaging of lymph node metastases in the meta-analysis of Xing et al (0.96 [0.85; 0.99]) was significantly higher for a significance level $\alpha = .05$ than all values determined in the present study for the sensitivity of pUS.

In the present study, pUS was used to examine 511 lesions; only 35 (7%) were examined in the context of primary staging. Therefore, this significant difference cannot be explained solely by the facts that no distinction was made between primary staging and restaging by pUS in the present study and that possible small metastases (micrometastases) are known to be difficult to diagnose in primary staging.

The present study revealed very high values for the specificity of pUS in the PPA (0.97 [0.92; 0.99]), PEA (0.98 [0.96; 0.99]), and PLA (0.98 [0.95; 0.99]), which can also be found in the meta-analysis by Xing et al for both primary staging (0.97 [0.88; 0.99]) and restaging (0.99 [0.95; 1.00]) of lymph node metastases.

The sensitivity and specificity values for ¹⁸F-FDG PET/CT and pUS determined for the same patient population in the present study are, therefore, largely similar to those determined in studies with patient populations for either PET/CT or US. An exception is the sensitivity of pUS for restaging in existing studies with separate patient populations, which is, on average, significantly higher (significance level $\alpha = .05$).

4.2. Combined application of 18f-FDG PET/ct and US

For all 3 combinations studied, the sensitivity of the combined application of ¹⁸F-FDG PET/CT and US showed an equally large or narrower 95% CI than the single applications. Consequently, the probability that the true value will be close to the calculated value is higher. This indicates the combined use of ¹⁸F-FDG PET/CT and US. The insignificantly higher sensitivities of the combinations to the sole use of ¹⁸F-FDG PET/CT speak against it. Regarding specificity, in all 3 combinations studied, the values of the sole applications were equal to those of the combinations of ¹⁸F-FDG PET/CT and US.

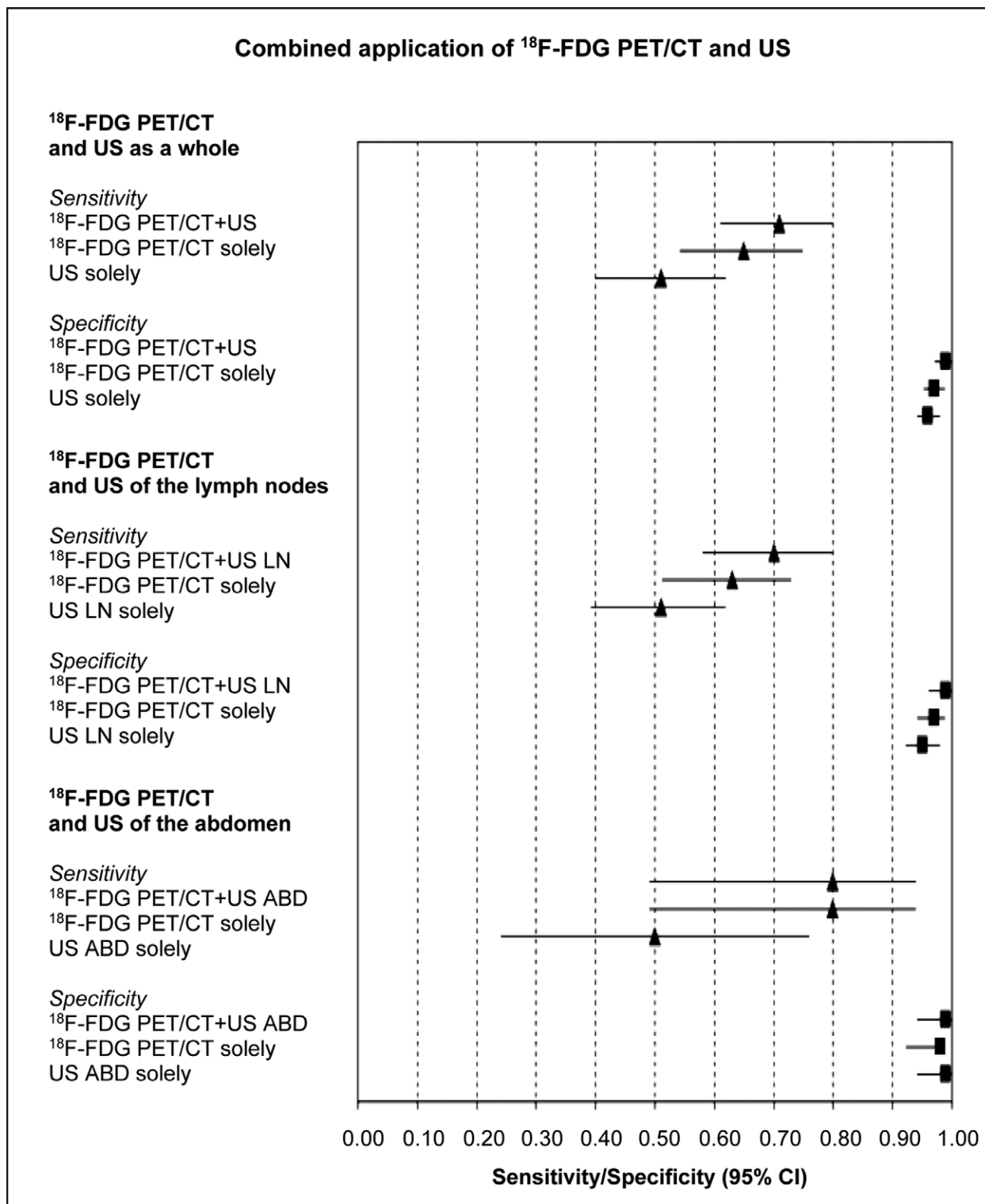


Figure 4. Forest plot for the sensitivity (triangular marks) and specificity (rectangular marks) of the combined application of ¹⁸F-FDG PET/CT and ultrasound (US) (as a whole, lymph nodes (LN) and abdomen (ABD)) as well as of the corresponding sole applications (significance level $\alpha = .05$, 95% confidence interval (CI)). ¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography.

5. Conclusion

Compared with US, ¹⁸F-FDG PET/CT has a higher sensitivity and minimally lower specificity. The combined application of ¹⁸F-FDG PET/CT and US has a slight advantage compared to the sensitivity of ¹⁸F-FDG PET/CT alone. In terms of specificity, the sole applications of ¹⁸F-FDG PET/CT and US, as well as

the combined application of these 2 radiological modalities, are equivalent. Overall, it is preferable to use ¹⁸F-FDG PET/CT as the sole application when considering costs. Combined examination has only a few advantages.

Due to the higher sensitivity of ¹⁸F-FDG PET/CT for lymph node metastases in the present study and because ¹⁸F-FDG PET/CT can be used to detect additional MM lesions throughout the

Table 9

Statistical characteristic values of the combined application of ¹⁸F-FDG PET/CT and ultrasound as a whole and the associated sole applications.

Statistical characteristic value	¹⁸ F-FDG PET/CT and US as a whole	Sole application of ¹⁸ F-FDG PET/CT	Sole application of US as a whole
n	362	362	362
Sensitivity	0.71 [0.61; 0.80]	0.65 [0.54; 0.75]	0.51 [0.40; 0.62]
Specificity	0.99 [0.97; 1.00]	0.97 [0.95; 0.99]	0.96 [0.94; 0.98]
PPV	0.95 [0.86; 0.98]	0.86 [0.75; 0.93]	0.80 [0.66; 0.89]
NPV	0.93 [0.89; 0.95]	0.91 [0.87; 0.94]	0.88 [0.84; 0.91]
Accuracy	0.93 [0.90; 0.95]	0.90 [0.87; 0.93]	0.87 [0.83; 0.90]

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, n = number of referenced lesions, NPV = negative predictive value, PPV = positive predictive value, US = ultrasound.

Table 10

Statistical characteristic values of the combined application of ¹⁸F-FDG PET/CT and ultrasound of the lymph nodes and the associated sole applications.

Statistical characteristic value	¹⁸ F-FDG PET/CT and US lymph nodes	Sole application of ¹⁸ F-FDG PET/CT	Sole application of US lymph nodes
n	264	264	264
Sensitivity	0.70 [0.58; 0.80]	0.63 [0.51; 0.73]	0.51 [0.39; 0.62]
Specificity	0.99 [0.96; 1.00]	0.97 [0.94; 0.99]	0.95 [0.92; 0.98]
PPV	0.96 [0.86; 0.99]	0.88 [0.75; 0.94]	0.79 [0.65; 0.89]
NPV	0.91 [0.86; 0.94]	0.88 [0.83; 0.92]	0.85 [0.80; 0.89]
Accuracy	0.92 [0.88; 0.94]	0.88 [0.84; 0.92]	0.84 [0.79; 0.88]

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, n = number of referenced lesions, NPV = negative predictive value, PPV = positive predictive value, US = ultrasound.

Table 11

Statistical characteristic values of the combined application of ¹⁸F-FDG PET/CT and ultrasound of the abdomen and the associated sole applications.

Statistical characteristic value	¹⁸ F-FDG PET/CT and US abdomen	Sole application of ¹⁸ F-FDG PET/CT	Sole application of US abdomen
n	98	98	98
Sensitivity	0.80 [0.49; 0.94]	0.80 [0.49; 0.94]	0.50 [0.24; 0.76]
Specificity	0.99 [0.94; 1.00]	0.98 [0.92; 0.99]	0.99 [0.94; 1.00]
PPV	0.89 [0.56; 0.98]	0.80 [0.49; 0.94]	0.83 [0.44; 0.97]
NPV	0.98 [0.92; 0.99]	0.98 [0.92; 0.99]	0.95 [0.88; 0.98]
Accuracy	0.97 [0.91; 0.99]	0.96 [0.90; 0.98]	0.94 [0.87; 0.97]

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, n = number of referenced lesions, NPV = negative predictive value, PPV = positive predictive value, US = ultrasound.

Table 12

McNemar test for significance between the sensitivities and specificities of the combined use of ¹⁸F-FDG PET/CT and ultrasound (as a whole, lymph nodes and abdomen) compared to the associated sole applications (significance level $\alpha = 0.05$).

Combined application Sole application	¹⁸ F-FDG PET/CT and US as a whole		¹⁸ F-FDG PET/CT and US lymph nodes		¹⁸ F-FDG PET/CT and US abdomen	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
¹⁸ F-FDG PET/CT	$P = .0442$	$P = .0442$	$P = .0442$	$P = .0801$	$p = N/A$	$P = .6171$
US as a whole	$P = .0001$	$P = .0140$	–	–	–	–
US lymph nodes	–	–	$P = .0005$	$P = .0140$	–	–
US abdomen	–	–	–	–	$P = .1489$	$P = N/A$

¹⁸F-FDG PET/CT = 2-18fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, N/A = not applicable, US = ultrasound.

body, ¹⁸F-FDG PET/CT should be preferred in the search for lymph node metastases.

Based on the available data from this study, the use of aUS cannot be justified, neither as a sole application, nor in combination with ¹⁸F-FDG PET/CT.

5.1. Limitations

This study has a risk of information bias owing to the retrospective data evaluation of the entire patient population. This occurs especially for true-negative and false-positive assessments of lesions by the examined radiological imaging modalities in the absence of an explicit negation or reference standard in the RIS data.

Author contributions

Conceptualization: Philipp Weber, Andreas Arnold, Joachim Hohmann.

Data curation: Philipp Weber, Joachim Hohmann.

Formal analysis: Philipp Weber, Andreas Arnold, Joachim Hohmann.

Funding acquisition: Joachim Hohmann.

Methodology: Philipp Weber, Andreas Arnold, Joachim Hohmann.

Project administration: Joachim Hohmann.

Software: Philipp Weber.

Supervision: Andreas Arnold, Joachim Hohmann.

Validation: Philipp Weber, Andreas Arnold, Joachim Hohmann.

Writing – original draft: Philipp Weber.

Writing – review & editing: Philipp Weber, Andreas Arnold, Joachim Hohmann.

References

- [1] National Agency for Cancer Registration (NACR). Cancer Incidence 2013-2017: new cases, rates and trend by cancer site and period. Neuchâtel (CH): Swiss Federal Statistical Office. 2020. Available at: <https://www.bfs.admin.ch/bfsstatic/dam/assets/14816237/master> [Access date August 17, 2021].
- [2] The Global Cancer Observatory. Cancer today. Lyon (F): International Agency for Research on Cancer. 2020. Available at: <http://gco.iarc.fr/today> [Access date August 31, 2021].
- [3] Gershenwald JE, Scolyer RA, Hess KR, et al. Melanoma staging: evidence-based changes in the American joint committee on cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67:472–92.
- [4] Institute for Quality and Efficiency in Health Care. Positron Emission Tomography (PET) and PET/CT in Malignant Melanoma: Executive Summary of Final Report D06-01E, Version 1.0. Cologne, Germany: Institute for Quality and Efficiency in Health Care; 2011.
- [5] Xing Y, Bronstein Y, Ross MI, et al. Contemporary diagnostic imaging modalities for the staging and surveillance of melanoma patients: a meta-analysis. J Natl Cancer Inst. 2011;103:129–42.
- [6] El Sharouni MA, van DPJ, Witkamp AJ, et al. Subtyping cutaneous melanoma matters. JNCI Cancer Spectrum. 2020;4:pkaa097.
- [7] Heckert N, Filliben J, Croarkin C. et al. NIST/SEMATECH e-Handbook of statistical methods. 2002. Updated 2012. Available at: <http://www.itl.nist.gov/div898/handbook/> [Access date April 24, 2022].

- [8] Sundjaja JH, Shrestha R, Krishan K. McNemar and Mann-Whitney U tests. Updated July 2021. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2022. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK560699/> [access date April 23, 2022].
- [9] Taylor C. Degrees of freedom for independence of variables in two-way table. ThoughtCo. 2020. Available at: [thoughtco.com/degrees-of-freedom-in-two-way-table-3126402](https://www.thoughtco.com/degrees-of-freedom-in-two-way-table-3126402) [Access date April 23, 2022].
- [10] [Data analysis with SPSS: Pearson chi-square test (Contingency Analysis)]. University of Zurich. 2021. Available at: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/pearsonzush.html [Website in German]. [Access date November 30, 2021].
- [11] Renesh B. Fisher's exact test of independence in R [with example]. Data science blog. 2022. Available at: <https://www.reneshbedre.com/blog/fisher-exact-test.html> [Access date April 23, 2022].