Position emission tomography with or without computed tomography in the primary staging of Hodgkin's lymphoma

Martin Hutchings
Annika Loft
Mads Hansen
Lars M. Pedersen
Anne Kiil Berthelsen
Susanne Keiding
Francesco D'Amore
Anne-Marie Boesen
Lone Roemer
Lena Specht

Background and Objectives. In order to receive the most appropriate therapy, patients with Hodgkin's lymphoma (HL) must be accurately stratified into different prognostic staging groups. Computed tomography (CT) plays a pivotal role in the conventional staging. The aim of the present study was to investigate the value of positron emission tomography using 2-[18F]fluoro-2-deoxy-D-glucose (FDG-PET) and combined FDG-PET/CT for the staging of HL patients, and the impact on the choice of treatment.

Design and Methods. Ninety-nine consecutive, prospectively included patients had FDG-PET and CT in their staging work-up. Sixty-one of the 99 patients had combined FDG-PET/CT. A standard of reference for each nodal region and organ was determined using all available information including scan results, histology and a minimum of one year's clinical follow-up data. The lack of a satisfactory diagnostic gold standard limits the reliability of accuracy calculations.

Results. FDG-PET would have upstaged 19% of patients and downstaged 5% of patients, leading to a different treatment in 9% of patients. For FDG-PET/CT, the corresponding figures are 17%, 5%, and 7%. In nodal regions, the sensitivity of FDG-PET and FDG-PET/CT seemed higher than that of CT (92% and 92% vs. 83%). FDG-PET identified more false positive nodal sites than did CT and FDG-PET/CT (1.6% vs 0.7% and 0.5%). FDG-PET and FDG-PET/CT were highly sensitive for evaluating organs (86% and 73%) while CT detected 37% of involved organs.

Interpretation and Conclusions. FDG-PET and FDG-PET/CT have a substantial potential impact on staging and choice of treatment and the methods tend to upstage rather than downstage patients. FDG-PET and FDG-PET/CT seem to have a higher diagnostic accuracy than CT in the staging of HL. However, care should be taken so patients with an excellent prognosis and at risk of over-treatment do not receive more intensive treatment because of these staging methods.

Key words: FDG-PET, FDG-PET/CT, Hodgkin, lymphoma, staging.

Haematologica 2006; 91:482-489

From the Department of Clinical Physiology and Nuclear Medicine, PET and Cyclotron Unit (MH, AL, AKB), and the Departments of Haematology (MH), Radiotherapy (MH, AKB, LS), Radiology (AKB), and Oncology (LS), Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; the Department of Haematology (LMP), University Hospital Herlev, Copenhagen, Denmark; the PET Center (SK), the Department of Haematology (FD), and the Department of Radiology (LR), Aarhus University Hospital, Aarhus, Denmark.

Correspondence:
Martin Hutchings, MD, PET and Cyclotron Unit, Department of Clinical Physiology and Nuclear Medicine, The Diagnostic Centre - Copenhagen University Hospital
9 Blegdamsvej, DK-2100
Copenhagen Ø, Denmark,
E-mail:hutchings@dadlnet.dk

The long-term cure rate of Hodgkin's lymphoma (HL) is over 80% due to modern combination chemotherapy and radiotherapy. The improved survival has revealed serious long-term adverse effects of the treatment, including cardiopulmonary disease and secondary malignancies. HL patients have an excess mortality directly related to these late treatment effects.[1-4] In order to reduce the long-term adverse effects of treatment, therapeutic strategies are becoming more tailored to the individual patient.[5] Individualized HL therapy requires an early and reliable estimation of each patient's prognosis. Pre-treatment prognostic factors, such as clinical stage, number of involved regions, B-symptoms, extranodal disease, bulky disease, age, blood counts and biochemical parameters, have been shown to predict survival in large cohort studies.[6-8] The initial treatment strategy is largely determined by measures of disease dissemination, the single most important factor at present being the clinical stage.[9]

Computed tomography (CT) plays a pivotal role in the conventional staging of lymphoma patients. CT has replaced more complicated procedures such as laparotomy (with splenectomy), lymphangiography and mediastinoscopy and is now the method of choice for identifying sites of disease not detectable by clinical examination. However, CT fails to identify a considerable number of sites, especially abdominal ones.[10] During the last decades, tomographic nuclear medicine imaging modalities have been introduced into the management of HL. Gallium scintigraphy was introduced in the early 1970s as a valuable addition to the anatomical imaging modalities.[11] Positron emission tomography using 2-[18F]fluoro-2-deoxy-D-glucose (FDG-PET) is now considered superior to gallium scintigraphy.[12,13] A number of investigations have examined the properties of FDG-PET in the staging of HL.[14-21] These studies have included from 20 to 44 patients and most have been performed in a retrospective fashion. Only two studies have directly assessed the region-by-region accuracy of FDG-PET in the staging of HL.[16,19] More recently, combined FDG-PET/CT has emerged as an important imaging modality, but the value of FDG-PET/CT in the management of HL has not been thoroughly assessed. The aim of the present study was to investigate, in a large number of patients and in a prospective setting, the diagnostic accuracy of FDG-PET and FDG-PET/CT and their impact on the choice of treatment strategy.

Design and Methods

Patients

This study was a collaboration between the lymphoma treatment centers at Copenhagen University Hospital, Rigshospitalet (RH), Herlev Hospital (HER), and Aarhus University Hospital (AUH) and the PET centers at RH and AUH. Ninety-nine consecutive patients with newly diagnosed HL were prospectively included in the protocol from November 2001 until June 2004. Exclusion criteria were diabetes mellitus, pregnancy and age under 18 years. Sixty-six patients were treated at RH, 16 patients at HER and 17 patients at AUH. Lymph node biopsies were obtained and histologically subtyped according to the WHO classification.[22] The clinical data listed in Table 1 were obtained, and all patients underwent initial staging PET along with standard staging procedures, including CT. Sixty-one of the 66 patients from RH had their staging scans performed as PET/CT investigations. Clinical follow-up data were recorded at regular visits to the lymphoma clinic. The study was approved by the local human investigations ethical committee and performed in accordance with the revised Helsinki declaration.

Treatment

Early stage disease was treated according to the Nordic Lymphoma Group protocols.[23] Patients with advanced stage disease were treated with anthracycline-containing chemotherapy. Depending on the stage and site of presentation, patients were given either chemotherapy alone or a combination of chemotherapy and radiotherapy. Radiotherapy was given with megavoltage energies using an involved field technique to deliver 30-36 Gy to the tumor in 1.8 Gy daily fractions and five fractions per week.

PET and CT scans

^{18}F-FDG was produced in on-site cyclotron and chemistry facilities. All FDG-PET scans were performed as whole-body scans (mid-brain to upper thigh) after a 6-hour fast. Patients were scanned 45-90 minutes after intravenous injection of approximately 400 MBq ^{18}F-FDG. Sixty-one patients from RH were scanned in a GE LS Discovery PET/CT scanner (General Electric Medical Systems, Milwaukee, Wisconsin, USA) with emission scans of 3 minutes per bed position, 16 patients from HER and six patients from RH were scanned (at the RH PET center) in a GE Advance PET scanner, and 17 patients from AUH were scanned using a Siemens/CTI ECAT Exact HR47-PET scanner (Siemens/CTI, Knoxville, TN, USA) with emission scans of 5 minutes per bed position following transmission scans. High resolution images were produced with ordered subset expectation maximation (OSEM) iterative reconstruction, using transmission scans for correction, or CT data when available. The OSEM algorithm was applied to ratio sinograms using attenuation-weighted iterative reconstruction (two iterations, 28 subsets) and subsequent smoothing with a Hanning filter.[24] Diazepam was given orally to some patients before FDG-administration to avoid muscular uptake of the tracer. CT scans covered the cervical, thoracic and abdominal regions with a section thickness of 5 mm. All patients were given oral and intravenous contrast agents.

Table 1. Patients' characteristics.

	Patients with staging PET	*Patients with staging PET/CT*
No.	99	61
Age (years)		
Mean	40.5	41.4
Median	36.2	37.3
Range	18.6–79.2	18.6–79.2
Follow-up (months)		
Mean	22,7	24.4
Median	20.8	23.8
Range	2.0-40.8	2.0-40.8
2-year progression-free survival	80.2%	80.5%
Gender		
Male	61 (62%)	34 (56%)
Female	38 (38%)	27 (44%)
Clinical stage (conventional staging)		
I	22 (22%)	11 (18%)
II	42 (42%)	24 (39%)
III	27 (27%)	18 (30%)
IV	8 (8%)	8 (13%)
No. of regions		
Mean	3.10	3.20
Median	3	3
Range	1-10	1-8
Extranodal disease		
Yes	17 (17%)	16 (26%)
No	82 (83%)	45 (74%)
B-symptoms		
Yes	52 (53%)	36 (59%)
No	47 (48%)	25 (41%)
Bulky disease		
Yes	31 (31%)	18 (30%)
No	68 (69%)	43 (71%)
Histological type		
Nodular sclerosing	61 (62%)	51 (66%)
Mixed cellularity	20 (20%)	17 (22%)
CHL, NOS	8 (8%)	3 (4%)
NLP	10 (10%)	6 (8%)
IPS (ref. #6, values 1-7)		
Mean	2.77	2.85
Median	3	3
Range	1-6	1-6
First-line treatment		
ABVD	85 (86%)	52 (85%)
ABV/MOPP	3 (3%)	3 (5%)
ABVD/COPP	2 (2%)	0 (0%)
BEACOPP esc.	2 (2%)	1 (2%)
PVAG	2 (2%)	2 (3%)
Radiotherapy only	5 (5%)	3 (5%)
Clinical outcome		
Progression	18 (18%)	12 (20%)
Death	5 (5%)	3 (5%)

B-symptoms: unexplained pyrexia, night sweats or weight loss; CHL-NOS: classical HL, not otherwise specified; NLP: nodular lymphocyte predominance HL; IPS: International Prognostic Score; ABVD : adriamycin, bleomycin, vinblastine, dacarbazine; ABV/MOPP: adriamycin, bleomycin, vinblastine, mechlorethamine, vincristine, procarbazine, prednisolone; ABVD/COPP: cyclophosphamide, vincristine, procarbazine, prednisolone; BEACOPP: bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisolone; PVAG: prednisolone, vinblastine, doxorubicin, gemcitabine.

Data analysis

PET images were displayed as projections and as transaxial, coronal and sagittal tomographic sections. Two experienced nuclear medicine physicians read all scans, and differences were decided by consensus. The nuclear medicine physicians were blind to the CT results and all other clinical information except the diagnosis, and the radiologists were blind to the results of PET. The clinicians were also unaware of the PET results, which thus had no impact on the treatment given. The PET and CT images from the 61 PET/CT scans were initially read separately, with no communication between the nuclear medicine physicians and the radiologists, and with no fusion of the images. At a minimum of one year after diagnosis, the fused PET/CT scans were opened and read by an experienced nuclear medicine physician and an experienced radiologist together. They were blind to the identity and all clinical information about the patients. In this way, PET/CT was regarded as a modality of its own, and not merely as the function of the separate findings on PET and CT. The hilar regions were analyzed as included in the mediastinum, since these regions are very difficult to distinguish on PET scans. The standardized uptake value (SUV) was calculated for 60 of the 61 patients examined in the RH PET/CT-scanner.[25] One staging PET/CT scan could not be analyzed for SUV since the body weight was not recorded and the patient died after just a single course of ABVD (adriamycin, bleomycin, vinblastine, dacarbazine) treatment. Regions of interest (ROI) were drawn representing lymph node regions and organs on all transaxial and coronal slices. Counts were normalized for injection dose and body weight using the following formula:

$$\frac{\text{Activity concentration (Bq/mL) x body weight (g)}}{\text{injected activity (Bq)}}$$

SUV_{max} was recorded as the maximum value in each region or organ. Maximum values were used under the assumption that this procedure enhances the reproducibility of the measurements.

Reference standard

In order determine the diagnostic accuracy of a new method, the results of the method must be compared to those of a gold standard method. The optimal gold standard would require sampling of biopsies from all nodal regions and all internal organs. For obvious practical and ethical reasons, this is impossible. Instead, a reference standard for each region or organ was established at a minimum of one year after the diagnosis. A region or organ with involvement seen on both PET and CT was regarded as a true positive focus and a site with no suspicious signs on PET and CT was regarded as a true negative. Discrepant findings were assessed at a consensus conference after a minimum follow-up of one year. This consensus conference was carried out after the analysis of the combined staging PET/CT images. At the consensus conference all available clinical information was taken into consideration. In eight cases, there was histological evidence to prove or disprove the presence of disease (three lymph node, three bone marrow, and two liver biopsies). For all other discrepant findings the status of the region or organ was determined using information from treatment monitoring and follow-up examinations (CT and PET, or PET/CT, was performed after two, four and six to eight cycles of chemotherapy). For example, a region with a marginally enlarged, PET-negative lymph node would be labeled *not involved*, provided that the node did not shrink during treatment while other enlarged nodes regressed. On the other hand, if a small (<1 cm and radiologically normal), PET-positive lymph node disappeared and changed to PET-negative during treatment, along with the regression of other masses, it would be labeled *involved*.[26] When regarding calculations of diagnostic accuracies based on such a reference standard, there are serious limitations which must be acknowledged. These reservations are discussed in detail below.

Table 2. Sensitivity and specificity of CT, PET and PET/CT region-by -region.

	Percentage involved°	*Sensitivity*			*Specificity*		
		CT	*PET*	*PET/CT*	*CT*	*PET*	*PET/CT*
Left cervical region	72%	85%	90%	95%	96%	96%	95%
Right cervical region	62%	82%	93%	89%	100%	90%	96%
Left axilla	31%	80%	94%	85%	99%	99%	98%
Right axilla	22%	67%	86%	75%	97%	96%	100%
Mediastinum*	66%	95%	99%	100%	97%	91%	100%
Retroperitoneum	34%	91%	94%	100%	99%	97%	100%
Left iliac region	9%	50%	78%	100%	100%	99%	100%
Right iliac region	14%	77%	93%	91%	100%	100%	100%
Left inguinal region	8%	43%	75%	83%	99%	99%	100%
Right inguinal region	10%	67%	90%	71%	100%	100%	100%
Spleen	20%	37%	80%	83%	100%	99%	92%
Liver	4%	100%	75%	50%	100%	100%	100%
Lungs	10%	56%	100%	71%	99%	91%	96%
Bones	16%	13%	88%	70%	100%	96%	100%

**Including hilar regions; °percentage of patients with involvement of the region/organ.*

Statistical analysis

The diagnostic accuracies are given as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Receiver operating characteristics (ROC) curves were used to optimize the cut-off points for SUV_{max}. All tests were two-sided and 5% was taken as the level of statistical significance. All data analyses were performed using the statistical software package SPSS 13.0 (SPSS inc., Chicago, IL, USA).[27,28]

Results

Staging accuracy in nodal regions

The frequency of involvement and the sensitivity/specificity of CT, PET and PET/CT are listed in Table 2. The cervical regions, the axillae and the inguinal regions were regarded as peripheral regions and the mediastinum, retroperitoneum and the pelvic regions were regarded as deep regions. Involvement was seen in 268/980 regions on CT, in 316/990 regions on PET and in 192/610 regions on PET/CT. A reference standard was established for all patients, determining that 325 of the

Table 3. **Overall accuracy rates of CT, PET and PET/CT for nodal staging, using qualitative PET assessment.**

	CT	*PET*	*PET/CT*	*SUV_{max}**
No. of regions	980	990	610	600
True positive	261 (27%)	300 (30%)	189 (31%)	172 (29%)
False positive	7 (0.7%)	16 (1.6%)	3 (0.5%)	20 (3.3%)
True negative	657 (58%)	649 (66%)	402 (66%)	393 (66%)
False negative	55 (5.6%)	25 (2.5%)	16 (2.6%)	15 (2.5%)
Sensitivity	82.6% (78.0-86.4)	92.3% (89.5-94.4)	92.2% (87.7-95.1)	92.0% (88.1-94.7)
Specificity	98.9% (97.8-99.5)	97.6% (96.4-98.4)	99.3% (97.8-99.7)	95.2% (93.1-96.6)
PPV	97.4% (94.7-98.7)	94.9% (92.5-96.6)	98.4% (95.5-99.5)	89.6% (85.4-92.9)
NPV	92.3% (90.1-94.0)	96.3% (94.9-97.3)	96.2% (93.9-97.6)	96.3% (94.5-97.6)

*PPV: positive predicitive value; NPV: negative predictive value. *SUV analyses of 60 patients who underwent staging PET/CT.*

Table 4. **Sensitivity, specificity and predictive values for nodal staging.**

	CT	*PET*	*PET/CT*	*SUV_{max}*
All nodal regions				
No.	980	990	610	600
Sensitivity	82.6%	92.3%	92.2%	92.0%
Specificity	98.9%	97.6%	99.3%	95.2%
PPV	97.4%	94.9%	98.4%	89.6%
NPV	92.3%	96.3%	96.2%	96.3%
Peripheral regions				
No.	588	594	366	360
Sensitivity	78.8%	90.6%	87.5%	91.0%
Specificity	98.7%	97.4%	98.8%	94.0%
PPV	96.9%	94.8%	97.2%	87.1%
NPV	90.2%	95.3%	94.2%	95.1%
Mediastinum				
No.	98	99	61	60
Sensitivity	95.3%	98.5%	100%	100%
Specificity	97.1%	91.2%	100%	95.2%
PPV	98.4%	95.5%	100%	97.5%
NPV	91.7%	96.9%	100%	100%
Abdominal and pelvic regions				
No.	294	297	183	180
Sensitivity	81.5%	91.2%	97.7%	86.5%
Specificity	99.6%	98.8%	100%	97.2%
PPV	97.8%	94.5%	100%	88.9%
NPV	96.0%	97.9%	99.3%	96.5%

No.: number of regions in the analysis; PPV: positive predictive value; NPV: negative predictive value.

Table 5. **Overall accuracy rates of CT, PET and PET/CT for organ staging.**

	CT	*PET*	*PET/CT*	*SUV_{max}*
No. of organs	392	396	244	180*
True positive	17 (4%)	43 (11%)	24 (10%)	16 (9%)
False positive	1 (0%)	12 (3%)	6 (2%)	7 (4%)
True negative	345 (88%)	334 (84%)	205 (84%)	152 (84%)
False negative	29 (7%)	7 (2%)	9 (4%)	5 (3%)
Sensitivity	37.0% (26.3-49.1)	86.0% (76.0-92.2)	72.7% (58.6-83.4)	76.2% (58.4-87.9)
Specificity	99.7% (98.7-99.9)	96.5% (94.5-97.8)	97.2% (94.6-98.5)	95.6% (92.1-97.6)
PPV	94.4% (78.5-98.8)	78.2% (67.8-85.9)	80.0% (65.7-89.3)	69.6% (52.4-82.6)
NPV	92.2% (89.6-94.2)	97.9% (96.2-98.9)	95.8% (92.9-97.5)	96.8% (93.6-98.4)

SUV_{max} values were calculated for the spleen, liver, and lungs, but not for the bones. PPV: positive predictive value; NPV: negative predictive value.

990 nodal regions had initial involvement. The overall predictive values and sensitivity/specificity for nodal staging of CT, PET and PET/CT are given in Table 3. Table 4 shows the predictive values and sensitivity/specificity for peripheral regions and deep regions above and below the diaphragm separately.

Staging accuracy in organs

Table 5 shows the predictive values and sensitivity/specificity of CT, PET and PET/CT for detection of organ involvement. Organs considered were spleen, liver, lungs and bones (bone marrow). In Table 6 the sensitivity and specificity are shown for the spleen, lungs and bones separately. Since only four patients were found to have liver involvement, the results for liver involvement are not shown.

Quantitative analysis of FDG-PET data

SUV analyses was performed on 60 PET/CT scans. Logistic regression analyses showed highly significant correlations between SUV_{max} and the reference standard in all sites except the liver, which was involved in only three patients (*data not shown*). For each of the three anatomical locations, the SUV_{max} distribution with and without disease involvement is shown in Figure 1. ROC curves were drawn for each site and they were analyzed independently. The optimal SUV_{max} cut-off point was 4 g/mL in the peripheral nodal regions and 5 g/mL in the deep nodal regions and the organs (*data not shown*). These cut-off values were used for all calculations of SUV_{max} accuracy. The predictive values and sensitivity/specificity of SUV_{max} are displayed in Tables 3 and 4 (nodal regions) and Tables 5 and 6 (organs) along with the accuracies of qualitatively assessed PET, CT, and PET/CT.

Potential impact on staging and treatment strategy

Compared with conventional staging, FDG-PET would have upstaged 19 patients (19%) and downstaged five patients (5%) (Table 7A). This would have led to a change in treatment strategy in nine patients (9%), had the staging relied on FDG-PET alone. Seven patients would have moved from early to advanced stage disease (IA→IIIA:1, IIA→IIIA: 1, IIA→IVA: two, IB→IIB: three). Two patients would have moved from advanced to early stage disease (IIIA→IIA: 1, IIB→IB: one). Among the patients in whom FDG-PET/CT, was performed, this method would have upstaged ten patients (16%) and downstaged three patients (5%) compared with CT (Table 7B), leading to a change of therapy in four patients (7%). All four patients would have moved from early to advanced stage disease (IIA→IIIA:

one, IB→IIB: three). Table 7C shows that FDG-PET/CT upstaged six patients and downstaged five patients compared with FDG-PET. FDG-PET/CT would have moved five patients to a different treatment group than FDG-PET (8%). Three patients would have moved from early to advanced stage disease (IIA→IIIA: one, IB→IIB: two) and two patients would have moved from advanced to early stage disease (IIIA→IA: one, IIB→IB: one).

Figures 2 and 3 show images of a patient upstaged by PET/CT from stage III to stage IV. The PET/CT images in Figure 2 clearly show FDG-PET-positive foci in the liver not detected by CT alone, while no pathological FDG uptake was seen in mesenteric lymph nodes that were abnormal according to conventional morphological criteria. Neither the hepatic nor the mesenterial foci were biopsy-proven sites of disease involvement. The liver foci were FDG-PET-negative after two cycles of ABVD while the mesenteric lymph nodes remained marginally enlarged at the latest follow-up 18 months after diagnosis. The reference standard was based on these findings. In contrast the focally FDG-avid bone marrow displayed in Figure 3 was not seen on CT but bone marrow involvement was proven by biopsy. Of the seven patients who would have been upstaged to a more advanced treatment group by FDG-PET, only one had experienced progressive disease after a median follow-up of 24 months. All three patients who would have been upstaged by FDG-PET/CT are in continued complete remission. For comparison, 18 of the 99 patients had experienced progression during the follow-up period.

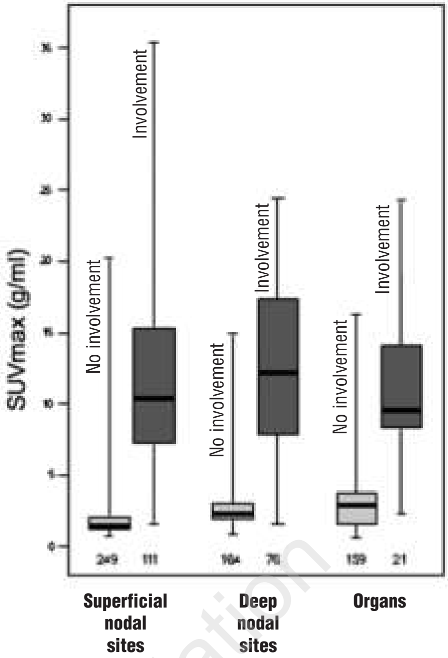

Figure 1. Box plots showing the distributions of SUVmax. SUV analyses were performed on 60 staging PET/CT scans. Region/organs with no involvement are represented by light gray boxes on the left in each of the three sections, while regions/organs with involvement are represented by darker gray boxes on the right. The black horizontal bars represent the median value, gray boxes represent the interquartile range (IQR, the values between the 25 and 75 percentiles), and whiskers represent the range. The numbers of regions/organs in the groups are given below the box plots.

Table 6. Sensitivity, specificity and predictive values for organ staging.

	CT	PET	PET/CT	SUV_{max}
All organs				
Sensitivity	37.0%	86.0%	72.7%	76.2%
Specificity	99.7%	96.5%	97.2%	95.6%
PPV	94.4%	78.2%	80.0%	69.6%
NPV	92.2%	97.9%	95.8%	96.8%
Spleen				
Sensitivity	36.8%	80.0%	83.3%	66.7%
Specificity	100%	98.7%	91.8%	95.8%
PPV	100%	94.1%	71.4%	80.0%
NPV	86.8%	95.1%	95.7%	92.0%
Lungs				
Sensitivity	55.6%	100%	71.4%	100%
Specificity	89.8%	91.0%	96.3%	96.3%
PPV	83.3%	55.6%	71.4%	75.0%
NPV	95.7%	100%	96.3%	100%
Bones				
Sensitivity	13.3%	87.5%	70.0%	
Specificity	100%	96.4%	100%	
PPV	100%	82.4%	100%	
NPV	86.5%	97.6%	94.4%	

PPV: positive predictive value; NPV: negative predictive value.

Discussion

The present study shows that FDG-PET and FDG-PET/CT have a strong potential impact on the staging of HL. The results indicate a higher staging accuracy of FDG-PET and FDG-PET/CT than of CT, although this finding is subject to serious reservations, as discussed in detail below. In 2001, Jerusalem *et al.* undertook the first thorough study of region-by-region accuracy of FDG-PET in HL. They scanned 33 patients before initial treatment or before treatment of relapse and evaluated the impact on nodal staging. In order to determine the method's sensitivity, a reference standard was based on the results of both conventional staging procedures including CT and FDG-PET. Biopsy results, response to treatment and follow-up data were used in cases of discrepant results. The sensitivity of FDG-PET for detecting involved lymph node regions was 95% in peripheral regions, 96% in thoracic regions, and 78% in abdominal/pelvic regions. The corresponding sensitivities for the conventional staging procedures (including CT) were 80%, 81%, and 86%.[16] In 2002, Weihrauch *et al.* applied a similar approach. They examined 22 patients and found involvement of 72 lymph node regions. No false positive lesions were recognized (probably in part due to the limitations of the reference standard), so both methods were regarded as having 100% specificity. The sensitivity of FDG-PET and CT was 88% and 74%, respectively.[19] The results of the present study indicate that FDG-PET is more sensitive than CT for overall nodal staging (92.3% vs. 82.6%, Table 3). The sensitivity was 91% in peripheral regions, 99% in the mediastinum, and 91% in abdominal/pelvic regions. The corresponding sensitivities for CT were 79%, 95%, and 82% (Table 4). FDG-PET produced a higher number of

Table 7A. FDG-PET vs. conventional methods impact on staging.

κ=0.66 (weighted)	*FDG-PET staging*				*Total*
	I	*II*	*III*	*IV*	
Conventional staging					
I	15	6	1	0	22
II	4	30	3	5	42
III	0	1	22	4	27
IV	0	0	0	8	8
Total	**19**	**37**	**26**	**17**	**99**

Table 7B. FDG-PET/CT vs. conventional methods impact on staging.

κ=0.71 (weighted)	*FDG-PET/CT staging*				*Total*
	I	*II*	*III*	*IV*	
Conventional staging					
I	9	4	0	0	13
II	1	18	2	1	22
III	0	0	15	3	18
IV	0	0	2	6	8
Total	**10**	**22**	**19**	**10**	**61**

Table 7C. FDG-PET/CT vs. FDG-PET: impact on staging.

κ=0.75 (weighted)	*FDG-PET/CT staging*				*Total*
	I	*II*	*III*	*IV*	
FDG-PET staging					
I	8	2	0	0	10
II	1	20	2	0	23
III	1	0	14	2	17
IV	0	0	3	8	11
Total	**10**	**22**	**19**	**10**	**61**

false positive results than CT did, resulting in a slightly lower specificity, although this was not statistically significant. For organ staging, our results point towards FDG-PET having a higher sensitivity than CT (86.0% vs. 37.0%). A number of false positive findings on FDG-PET (3% of all organs) resulted in FDG-PET having a lower specificity than CT (96.5% vs. 99.7%, Table 5).

A recent study by Allen-Auerbach *et al.*[29] showed a higher overall staging accuracy in lymphoma using FDG-PET/CT than FDG-PET alone. Their analysis of 53 patients with non-Hodgkin's lymphoma and 20 with HL did not include an analysis of accuracy, but evaluated the different methods' ability to refer patients to the correct Ann Arbor stage. Schaefer *et al.* compared the diagnostic properties of dual modality FDG-PET/low-doseCT with high-resolution contrast-enhanced CT. Their retrospective study included 19 patients referred for primary staging (11 with HL and eight with high-grade non-Hodgkin's lymphoma. Results were only presented on a per-patient basis. Lymph node involvement was seen in all 19 patients with both methods. Organ involvement was present in four patients, and this was found in three patients with FDG-PET/low-dose CT and in only one patient with contrast-enhanced CT.[30]

The present study is the first to attempt an analysis of the region-by-region accuracy of FDG-PET/CT in HL. Our results indicate that FDG-PET/CT is equivalent to FDG-PET alone for nodal staging except in the mediastinum and the abdominal and pelvic regions where FDG-PET/CT seems to have a higher sensitivity than both FDG-PET and CT (Table 4). These regions are often difficult to analyze with FDG-PET due to physiological FDG uptake in normal structures (bowel and urinary tract), which are easier to distinguish from tumor tissue with FDG-PET/CT. For organ staging, FDG-PET/CT seems to have no obvious advantage over FDG-PET, but seems to represent a compromise between the high sensitivity and relatively low specificity of FDG-PET and the high specificity and low sensitivity of CT (Tables 5 and 6). When comparing FDG-PET and FDG-PET/CT, it must be kept in mind that the study populations are not identical, since 38 patients were studied by FDG-PET, but not FDG-PET/CT. Direct comparison of accuracy in different study populations is methodologically questionable, and the conclusions must be regarded as such. Since optimal cut-off points for SUV$_{max}$ were determined using the same material that was later analyzed, the determination of accuracy for SUV$_{max}$ should be regarded as hypothesis-generating only. With this reservation in mind, the SUV$_{max}$ data show the general tendency for an SUV$_{max}$ cut-off be less accurate than qualitative evaluation of FDG-PET images, whether FDG-PET or FDG-PET/CT is used. For both nodal regions and organs, the sensitivity and negative predictive value seem roughly as good as those with qualitative evaluation, whereas the specificity and positive predictive value are somewhat lower. It is surprising that SUV analysis gives a higher false positive rate than visual analysis of FDG-PET. This might be due in part to the problem that the groups compared are not identidal. Nevertheless, a number of patients had regions of relatively intense FDG-uptake, which were not regarded as positive with visual analysis. The reason for this is not clear. In the practical clinical setting, SUV analysis is less likely to be used when qualitative PET reading is straightforward. It would have been interesting to investigate the accuracy of SUV$_{max}$ in the sites and organs for which the qualitative assessment of FDG-PET was particularly difficult. In the present study this was not possible, since the status of a nodal site or organ was reported as either positive or negative.

Histological evidence is the gold standard for the diagnosis of lymphoma, but for obvious ethical reasons it is not possible to obtain biopsies from all lymph node regions and organs of interest. This is the background for the reference standard used in this study, as well as in the previous studies described. We believe that this represents the best possible compromise between feasibility and reliability. One can argue that such a compromise should not be made in the first place. Given that HL is one of the most common indications for FDG-PET and FDG-PET/CT, we preferred this compromise to no study at all. However, there are important problems with the reference standard, which seriously limit our ability to draw reliable conclusions regarding the accuracy of the methods. There is a strong risk that our conclusions are biased in favor of FDG-PET and FDG-PET/CT. A number of PET-positive foci, which were not seen on CT, disappeared during treatment. These foci were all labeled *involved*, although there was in fact no proof of malignancy. There

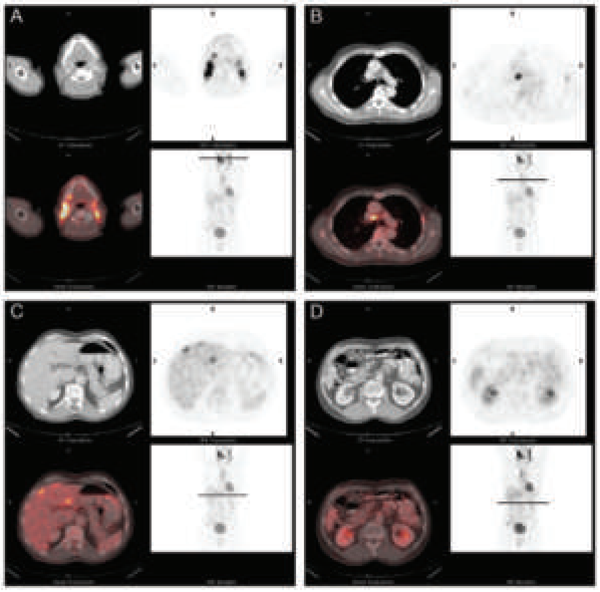

Figure 2. A 74-year old male who, after physical examination and CT was regarded as having stage III disease. Each section contains a transaxial CT, PET, and PET/CT image as well as an anterior/posterior PET projection image that indicates the position of the transaxial images. PET and PET/CT revealed liver involvement not seen on CT (C), while marginally enlarged mesenteric lymph nodes showed no FDG uptake (D).

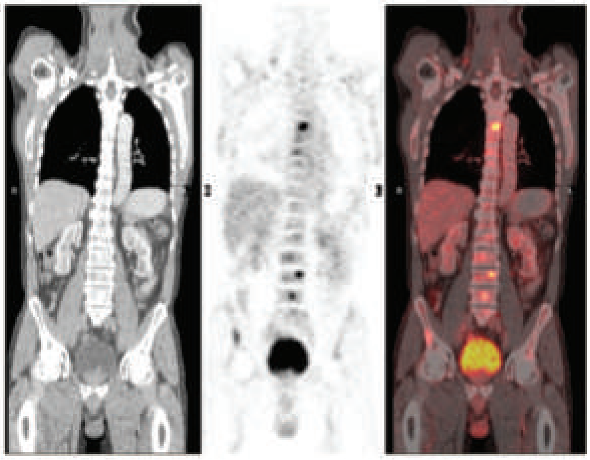

Figure 3. The same patient as in Figure 2. Conventional imaging showed nothing abnormal in the bones but PET/CT revealed pathological FDG uptake. Bone marrow involvement was proven by biopsy.

are a number of reasons why a benign FDG-PET positive focus can disappear. For example, infectious or inflammatory processes are likely to metabolise less FDG after a few months' treatment, either due to the effect of cytostatic therapy or just due to spontaneous resolution.

Numerous of investigations have focused on the potential impact of FDG-PET on staging and choice of therapy in HL.[14-21] These studies show that 11-41% of patients are upstaged by FDG-PET compared with conventional staging procedures and 0-28% are downstaged by FDG-PET. The fraction of patients in whom FDG-PET findings would potentially change the treatment strategy ranges from 3% to 25%. Our results displayed in Table 7a show that in this study 19% of patients were upstaged by FDG-PET and 5% of patients were downstaged by FDG-PET. FDG-PET would have changed the treatment strategy in 9% of all patients, the majority of whom (7/9) would have received a more intensive chemotherapy regimen. FDG-PET/CT has an impact on staging and choice of therapy which is comparable to that of FDG-PET alone (Table 7B). Table 7C underlines this by showing little difference between the staging with FDG-PET/CT compared with the staging with FDG-PET alone. FDG-PET and FDG-PET/CT have a strong impact on the staging and if used, would result in more advanced stage patients and more patients receiving prolonged courses of chemotherapy. It is not known whether the group of patients who are upstaged by FDG-PET and FDG-PET/CT would benefit from more intensive, and potentially more harmful, therapy. However, only one out of seven patients who would have been upstaged to an advanced treatment group by FDG-PET experienced progression during the 2-year follow-up period, compared with 18 out of all 99 patients. Likewise, none of the three patients who would have been upstaged to the advanced treatment group by FDG-PET/CT experienced progression, compared with 12 out of all 61 patients.

In conclusion, the present study indicates that FDG-PET and FDG-PET/CT are highly accurate in the staging of HL. Both FDG-PET and FDG-PET/CT seem superior to CT in all aspects of staging. FDG-PET/CT shows the same high sensitivity as FDG-PET in nodal regions and organs, but due to fewer false positive results, FDG-PET/CT has a higher specificity in nodal regions. The most obvious advantage of FDG-PET/CT is shown in the thorax and abdomen/pelvis, where both the sensitivity and specificity of this combined investigation seem higher than those of FDG-PET. Most importantly FDG-PET and FDG-PET/CT have a substantial potential impact on the staging and choice of treatment. The benefit for the patients is less clear. The patients in our study who were upstaged by FDG-PET and FDG-PET/CT have not so far shown an increased risk of relapse. This could change with longer follow-up. Will patients have a different outcome if their treatment plans are changed according to the FDG-PET results? This can only be answered in a controlled clinical trial. Given that FDG-PET/CT is already part of the staging work-up in a large number of lymphoma treatment centers, such trials are unlikely to be performed in the future. However, if the methods are adopted into the staging work-up under existing treatment guidelines, they are likely to result in a (possibly unnecessary) shift to more intensive therapy for a number of patients. In a disease in which treatment-related late effects are a stronger cause of morbidity and mortality than the disease itself, this is problematic. Modern, individualized HL therapy aims to reduce toxicity without impairing efficacy. For example, The HD13 study of the German Hodgkin Study Group (GHSG) investigates modifications of the ABVD regimen to achieve a less toxic therapy.[31] Leading centers advocate the use of FDG-PET/CT-guided intensity-modulated radiotherapy for HL and the most recent guidelines from the EORTC-GELA

Lymphoma Study Group for early-stage HL introduce involved-node radiotherapy in order to reduce the irradiated volumes.[32,33] Such regimens require as accurate a staging as possible. We believe that FDG-PET and FDG-PET/CT improve the quality of HL staging. However, the methods should only be implemented with great care and introduced along with steps to generally reduce treatment intensity, so they do not merely result in more intensive therapy to patients with an excellent prognosis who are already at risk of over-treatment.

MHu: designed the research, analyzed PET images, analyzed data and wrote the manuscript; ALJ: took part in the design of the study, produced and analyzed PET images, critically reviewed and approved the final version of the manuscript; MHa and SK took part in the design of the study, enrolled patients, critically reviewed the manuscript and approved its final version. LMP: was responsible for the design of the study, enrolled patients, recorded clinical data, critically reviewed the manuscript and approved its final version. AKBand LR: analyzed CT images, critically reviewed the manuscript and approved its final version. FDA: designed and performed the research, enrolled patients and recorded clinical data; AMB: designed the research, enrolled patients and recorded clinical data; LS: enrolled patients, supervised data analysis and manuscript writing, critically reviewed the manuscript and approved its final version. The authors declare that they have no potential conflict of interest.

Manuscript received November 1, 2005. Accepted February 2, 2006.

References

1. Hancock SL, Hoppe RT. Long-term complications of treatment and causes of mortality after Hodgkin's disease. Semin Radiat Oncol 1996;6:225-242.
2. Henry-Amar M, Somers R. Survival outcome after Hodgkin's disease: a report from the international data base on Hodgkin's disease. Semin Oncol 1990; 17:758-68.
3. Mauch P, Ng A, Aleman B, Carde P, Constine L, Diehl V, et al. Report from the Rockefellar Foundation Sponsored International Workshop on reducing mortality and improving quality of life in long-term survivors of Hodgkin's disease: July 9-16, 2003, Bellagio, Italy. Eur J Haematol Suppl 2005;10:68-76.
4. Hoppe RT. Hodgkin's disease: complications of therapy and excess mortality. Ann Oncol 1997; 8 Suppl 1:115-8.
5. Kogel KE, Sweetenham JW. Current therapies in Hodgkin's disease. Eur J Nucl Med Mol Imaging 2003; 30 Suppl 1:S19-S27.
6. Hasenclever D, Diehl V. A prognostic score for advanced Hodgkin's disease. International Prognostic Factors Project on Advanced Hodgkin's Disease. N Engl J Med 1998;339:1506-14.
7. Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, Young RC, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. J Clin Oncol 1989; 7: 1630-6.
8. Specht L. Prognostic factors in Hodgkin's disease. Semin Radiat Oncol 1996; 6: 146-61.
9. Gupta RK, Gospodarowicz MK, Lister TA. Clinical Evaluation and Staging. In: Mauch P, Armitage JO, Diehl V, Hoppe R, Weiss LM, eds. Hodgkin's disease. Philadelphia: Lippincott Williams and Wilkins, 19990. p. 223-40.
10. Munker R, Stengel A, Stabler A, Hiller E, Brehm G. Diagnostic accuracy of ultrasound and computed tomography in the staging of Hodgkin's disease. Verification by laparotomy in 100 cases. Cancer 1995; 76:1460-6.
11. Front D, Israel O. Present state and future role of gallium-67 scintigraphy in lymphoma. J Nucl Med 1996; 37:530-2.
12. Kostakoglu L, Leonard JP, Kuji I, Coleman M, Vallabhajosula S, Goldsmith SJ. Comparison of fluorine-18 fluorodeoxyglucose positron emission tomography and Ga-67 scintigraphy in evaluation of lymphoma. Cancer 2002; 94:879-88.
13. Wirth A, Seymour JF, Hicks RJ, Ware R, Fisher R, Prince M, et al. Fluorine-18 fluorodeoxyglucose positron emission tomography, gallium-67 scintigraphy, and conventional staging for Hodgkin's disease and non-Hodgkin's lymphoma. Am J Med 2002; 112:262-8.
14. Bangerter M, Moog F, Buchmann I, Kotzerke J, Griesshammer M, Hafner M, et al. Whole-body 2-[^{18}F]-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for accurate staging of Hodgkin's disease. Ann Oncol 1998;9: 1117-22.
15. Hueltenschmidt B, Sautter-Bihl ML, Lang O, Maul FD, Fischer J, Mergenthaler HG, et al. Whole body positron emission tomography in the treatment of Hodgkin disease. Cancer 2001;91:302-10.
16. Jerusalem G, Beguin Y, Fassotte MF, Najjar F, Paulus P, Rigo P, et al. Whole-body positron emission tomography using 18F-fluorodeoxyglucose compared to standard procedures for staging patients with Hodgkin's disease. Haematologica 2001;86:266-73.
17. Menzel C, Dobert N, Mitrou P, Mose S, Diehl M, Berner U, et al. Positron emission tomography for the staging of Hodgkin's lymphoma-increasing the body of evidence in favor of the method. Acta Oncol 2002;41:430-6.
18. Partridge S, Timothy A, O'Doherty MJ, Hain SF, Rankin S, Mikhaeel G. 2-Fluorine-18-fluoro-2-deoxy-D glucose positron emission tomography in the pretreatment staging of Hodgkin's disease: influence on patient management in a single institution. Ann Oncol 2000; 11:1273-9.
19. Weihrauch MR, Re D, Bischoff S, Dietlein M, Scheidhauer K, Krug B, et al. Whole-body positron emission tomography using 18F-fluorodeoxyglucose for initial staging of patients with Hodgkin's disease. Ann Hematol 2002;81:20-5.
20. Wiedmann E, Baican B, Hertel A, Baum RP, Chow KU, Knupp B, et al. Positron emission tomography (PET) for staging and evaluation of response to treatment in patients with Hodgkin's disease. Leuk Lymphoma 1999;34:545-51.
21. Munker R, Glass J, Griffeth LK, Sattar T, Zamani R, Heldmann M, et al. Contribution of PET imaging to the initial staging and prognosis of patients with Hodgkin's disease. Ann Oncol 2004; 15: 1699-704.
22. World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: IARC Press. 2001.
23. Glimelius B, Specht L, Nome O, Tuurpeeniemi-Hujanen T. Treatment of adult patients with early stage Hodgkin's disease. Accessed 18-12-2005. http://www.roc.se/PDFfiles/Vardprogram/Protokoll_hd.pdf.
24. Erdogan H, Fessler JA. Ordered subsets algorithms for transmission tomography. Phys Med Biol 1999;44:2835-51.
25. Weber WA, Schwaiger M, Avril N. Quantitative assessment of tumor metabolism using FDG-PET imaging. Nucl Med Biol 2000;27:683-7.
26. Hutchings M, Eigtved AI, Specht L. FDG-PET in the clinical management of Hodgkin lymphoma. Crit Rev Oncol Hematol 2004;52:19-32.
27. Landau S, Everitt BS. A Handbook of Statistical Analyses using SPSS. Boca Raton: Chapman & Hall/CRC. 2004.
28. Collett D. Modelling Survival Data in Medical Research. 2nd ed. Boca Raton: Chapman & Hall/CRC. 2003.
29. Allen-Auerbach M, Quon A, Weber WA, Obrzut S, Crawford T, Silverman DH, et al. Comparison between 2-deoxy-2-[18F]fluoro-D-glucose positron emission tomography and positron emission tomography/computed tomography hardware fusion for staging of patients with lymphoma. Mol Imaging Biol 2004; 6:411-6.
30. Schaefer NG, Hany TF, Taverna C, Seifert B, Stumpe KD, von Schulthess GK, et al. Non-Hodgkin lymphoma and Hodgkin disease: coregistered FDG PET and CT at staging and restaging--do we need contrast-enhanced CT? Radiology 2004; 232:823-9.
31. Klimm B, Diehl V, Pfistner B, Engert A. Current treatment strategies of the German Hodgkin Study Group (GHSG). Eur J Haematol Suppl 2005;66:125-34.
32. Yahalom J. Transformation in the use of radiation therapy of Hodgkin lymphoma: new concepts and indications lead to modern field design and are assisted by PET imaging and intensity modulated radiation therapy (IMRT). Eur J Haematol Suppl 2005;66:90-7.
33. Girinsky T, Pichenot C, Beaudre A, Ghalibafian M, Lefkopoulos D. Is intensity-modulated radiotherapy better than conventional radiation treatment and three-dimensional conformal radiotherapy for mediastinal masses in patients with Hodgkin's disease, and is there a role for beam orientation optimization and dose constraints assigned to virtual volumes? Int J Radiat Oncol Biol Phys 2006;64:218-26.