A functional dynamic scoring model to elucidate the significance of post-induction interim fluorine-18-fluorodeoxyglucose positron emission tomography findings in patients with Hodgkin's lymphoma

Eldad J. Dann,[1,5] Rachel Bar-Shalom,[3,5] Ada Tamir,[5] Ron Epelbaum,[2,5] Irit Avivi,[1,5] Menachem Ben-Shachar,[2] Diana Gaitini,[4] and Jacob M. Rowe[1,5]

[1]Department of Hematology & Bone Marrow Transplantation; [2]Department of Oncology and [3]Department of Nuclear Medicine, [4]Department of Radiology, Rambam Medical Center, Haifa, and [5]Bruce Rappaport Faculty of Medicine, Technion, Israel Institute of Technology, Haifa, Israel

ABSTRACT

Background
The findings of interim fluorine-18-fluorodeoxyglucose positron emission tomography (FDG-PET/CT) predict progression-free survival of patients with Hodgkin's lymphoma. Historically, the assessment was based on a static all-or-none scoring system. However, the clinical significance of any positivity in interim FDG-PET/CT has not been defined.

Design and Methods
Ninety-six patients with Hodgkin's lymphoma who underwent interim FDG-PET/CT were evaluated using dynamic and visual scores, employing mediastinal or liver blood pool uptake as a comparator. FDG-PET/CT was prospectively defined as positive if any abnormal F^{18}FDG uptake was present. In a retrospective analysis dynamic score 0 indicated resolution of all disease sites; score 1 defined a single residual focus; score 2 denoted a reduction in the number of foci; score 3 defined a reduction in intensity with no reduction in number; and score 4 indicated no change in the number and intensity of foci or appearance of new foci.

Results
The dynamic visual score review reduced the number of positive interim studies from 24 to 6 if a score of 2 or less was considered negative, with significantly better specificity (96%) as compared to static visual scores (78%-86%). The 5-year progression-free survival and overall survival rates in patients who had a negative dynamic score were 92% and 97%, respectively; the corresponding figures for patients with positive results were 50% and 67%.

Conclusions
A dynamic visual score may be a better indicator for tailoring therapy than static visual scoring.

Key words: F^{18}FDG-PET, FDG-PET/CT, Hodgkin's lymphoma, prognostic factors, dynamic score, specificity.

Citation: Dann EJ, Bar-Shalom R, Tamir A, Epelbaum R, Avivi I, Ben-Shachar M, Gaitini D, and Rowe JM. A functional dynamic scoring model to elucidate the significance of post-induction interim fluorine-18-fluorodeoxyglucose positron emission tomography findings in patients with Hodgkin's lymphoma. Haematologica 2010;95:1198-1206. doi:10.3324/haematol.2009.016105

©2010 Ferrata Storti Foundation. This is an open-access paper.

EJD and RBS contributed equally to this manuscript.

Acknowledgments: the authors would like to thank Sonia Kamenetsky for her devoted secretarial support, Dorit Dotan for data management and Nili Stein for statistical analyses. The authors also thank Professor Ora Israel for her helpful comments on the manuscript.

Manuscript received on August 24, 2009. Revised version arrived on December 26, 2009. Manuscript accepted on December 28, 2009.

Correspondence: Eldad J. Dann, MD, Department of Hematology and Bone Marrow Transplantation Rambam Medical Center and Technion, P.O. Box 9602, Haifa 31096, Israel. E-mail: e_dann@rambam.health.gov.il

Introduction

The current approach to the therapy of Hodgkin's lymphoma (HL) is aimed at attaining prolonged progression-free survival with the minimum possible long-term, treatment-related toxicity. While tailoring risk-adapted therapy for HL has been based on predefined and validated criteria, an individual interim assessment of response to treatment using 18-fluorine-18-fluorodeoxyglucose positron emission tomography (FDG-PET/CT) has been suggested to be a more accurate predictor of prognosis.[1] Previous studies of patients with HL showed that positive interim scintigraphy correlated with a low progression-free survival, while negative interim studies correlated with high progression-free survival rates[2-4] These studies were carried out in patients treated with ABVD (adriamycin, bleomycin, vinblastin, dacarbazine) or similar regimens.[5] The criteria for determining response from FDG imaging using either PET or PET/CT are a matter of ongoing debate.[6] It has not yet been determined whether a single residual site of minimal uptake truly represents resistant disease and it is still unclear whether response assessment should be based on the static one-point-in-time metabolic data provided by FDG-PET/CT during treatment or on the dynamics of FDG-PET/CT findings over a short period between pre-therapy and an interim study. Such a dynamic visual review could more accurately reflect the sequence of metabolic response. Furthermore, it is still unclear whether early modification of therapy affects the progression-free survival of patients with only a single site of residual uptake of FDG in the interim FDG-PET/CT. As escalating therapy is potentially toxic, it is important to increase therapy only in those patients who are clearly at high risk of treatment failure. The present trial assessed the prognostic performance of interim FDG-PET/CT in HL patients treated with various chemotherapy protocols, using four different scores to define response on FDG-PET/CT (Table 1). A static visual binary score was initially used prospectively for response evaluation on interim FDG-PET/CT. FDG-PET/CT studies were then retrospectively reviewed and compared using a dynamic visual score and two additional static visual scores with either mediastinal or liver blood pool uptake as a comparator.[7-9]

Design and Methods

This analysis was based on 96 patients with classic HL, treated between November 2001 and March 2006 at Rambam Health Care Campus (RHCC), Haifa, Israel, who underwent an interim FDG-PET/CT study after their first (n=15) or second (n=81) cycle of chemotherapy, and who had been followed until disease progression or for at least 1 year of complete remission. FDG-PET/CT was performed after the first cycle of chemotherapy in 15 patients treated before September 2003, from which time onward there was a consensus to perform the scintigraphic study after two cycles of chemotherapy.

Sixty-three individuals were treated with risk-adapted BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, oncovin, procarbazine, prednisone) protocols.[10] Individuals with an International Prognostic Score[11] score of 2 or less were considered to have a standard risk and were treated with two initial cycles of standard BEACOPP. Patients with a score of 3 or more were defined as being at high risk and received two initial cycles of escalated BEACOPP. Following interim FDG-PET/CT, the therapy was tailored as previously reported.[10] Patients who were initially treated with standard BEACOPP and had a positive interim FDG-PET/CT received an additional four cycles of escalated BEACOPP. Patients whose initial therapy was escalated BEACOPP also received four additional cycles following a positive interim FDG-PET/CT. If interim PET/CT was negative following 2 cycles of therapy, additional 4 cycles of standard BEACOPP were administered. If FDG uptake was present in only a single site at the end of therapy, major attempts were made to obtain tissue specimens for histological examination prior to taking a decision about therapy failure.

All patients underwent baseline and interim FDG-PET/CT studies at the Department of Nuclear Medicine of the RHCC and were prospectively included in the study. Interim FDG-PET studies were scheduled for day 10-14 following the last day of chemotherapy and were conducted at a mean of 13.6±1.4 days. All participants had additional FDG-PET/CT scans following completion of chemotherapy and then every 6 months for 2 years and once during the third year.

FDG-PET/CT was performed using a dedicated PET scanner with a full ring bismuth germinate detector and multi-slice CT (Discovery LS, General Electric Healthcare, Waukesha, WI, USA). Patients were instructed to fast for 6 h and blood glucose was measured to ensure that the level was lower than 11 mMol/L prior to injection of 370-444 MBq ^{18}F-FDG. The FDG-PET/CT acquisition protocol included an initial helical CT (140 Kv, 80 mA, 4

Table 1. Patients' characteristics.

Parameter	Data	Percent (%)
Gender (M/ F), n.	49/47	51/49
Median age, years (range)	30 (17-57)	
B symptoms, yes/no, n.	43/53	45/55
Bulky mediastinal mass, n.	10	10
Early disease (Ia, IIa)	2/31	34
Advanced disease Ann Arbor stage Ib,IIb,III, IV	63	66
Patients in whom the International Prognostic Score was applied in the different Ann Arbor groups, n.	1/18/19/25	-
International Prognostic score: no score/ 0/1/2	33/5/10/22	56
International Prognostic score 3/4/5-7	16/6/4	28
Initial chemotherapy regimen: ABVD	33	34
Ann Arbor Stage (I,II)/(III, IV),	25/8	76 / 24
B Symptoms (no/yes)	26/7	79 / 21
International Prognostic Score: no score /0/1/2	20/4/5/3	97
International Prognostic Score 3	1	3
Radiation therapy	22	67
Initial chemotherapy regimens: standard BEACOPP	41	43
Ann Arbor stage (I,II)/(III, IV),	24/17	59/ 41
B symptoms (no/yes)	25/16	61/ 39
International Prognostic Score no score /0/1/2	13/1/5/18	90
International Prognostic Score 3/4	2/2	10
Radiation therapy	15	36
Initial chemotherapy regimen: escalated BEACOPP	22	23
Ann Arbor stage (I,II)/(III, IV),	3/19	14/86
B symptoms (no/yes)	2/20	10/90
International Prognostic Score /2	1	5
International Prognostic Score 3/4/5-7	13/4/4	95
Radiation therapy	2	9

slices, 0.5 s per rotation, pitch 6:1, slice thickness 4.25 mm), followed by FDG-PET acquisition in two-dimensional mode for 4 min per field of view. FDG-PET data were reconstructed using order subsets expectation maximization. Data obtained from CT acquisition were used for low noise attenuation correction of FDG-PET emission data and for fusion of attenuation-corrected PET images with corresponding CT images. PET, CT, and fused PET/CT images were reviewed in axial, coronal, and sagittal planes, and in maximal-intensity projection three-dimensional cine mode, using the manufacturer's review station (Xeleris; General Electric Healthcare, Waukesha, WI, USA).

The interim FDG-PET/CT was initially assessed visually and considered negative for the presence of active HL when it showed no foci of increased ^{18}F-FDG uptake, other than those related to physiological biodistribution of the tracer or to a known benign process (static visual score) as defined in Table 2. In order to optimize the definition of positive studies and thus allow for better prediction of and correlation with risk of disease progression FDG-PET/CT results were further analyzed retrospectively using a 5-point dynamic visual score developed to evaluate the response to therapy. This score is based on a visual comparison of the findings of the interim and pre-treatment FDG-PET/CT studies as detailed in Table 2. All interim FDG-PET/CT results were re-analyzed using this dynamic visual score by a dedicated nuclear medicine specialist (R.B.S.) who was unaware of the patients' outcome. The performance of the 5-point dynamic visual score was then compared with that of two previously recommended static scores[7-9] based on visual interpretation criteria of the single interim FDG-PET/CT study, comparing the presence and intensity of FDG uptake in a residual lesion with mediastinal or liver uptake and with the size of residual masses on the CT component. These scoring systems were suggested by Juweid *et al.* for the Consensus of Imaging Subcommittee (CIS) of the International Harmonization Project in Lymphoma[7] and at the International Congress for Malignant Lymphoma held in Lugano in 2008[9] (Table 2).

All patients signed informed consent to participation in this study, which was performed in accordance with the Declaration of Helsinki guidelines, approved by the Institutional Review Board (approval n. 1376) and registered in the NIH clinical studies website (*clinicaltrials.gov identifier: NCT 00396916*).

Statistical methods

The primary end-point of the study was progression-free survival, calculated from diagnosis to the occurrence of disease progression, relapse or death from any cause.[8] The secondary end-point was overall survival, calculated from diagnosis to death from any cause. Complete remission was defined as the disappearance of all disease manifestations, at the end of therapy. Primary progressive disease was defined as persistence of disease, or the appearance of new sites of FDG uptake on the FDG-PET/CT scans during the first 3 months after completion of therapy. Relapse was defined as the appearance of new clinical or imaging findings, which were proven by biopsy to be foci of disease, in patients who had remained in complete remission for more than 3 months after the completion of therapy.[12] A positive interim FDG-PET/CT

Table 2. Summary of four scoring systems used to define metabolic response on interim F^{18}FDG PET/CT.

Static visual score	Score	Dynamic visual score (current study)	F^{18}FDG uptake in mediastinal blood pool as comparator[7,8]	F^{18}FDG uptake in liver blood pool as comparator[9]
No abnormal F^{18}FDG uptake	0	No abnormal F^{18}FDG uptake	No abnormal F^{18}FDG uptake	No abnormal F^{18}FDG uptake
	1	A single residual focus of abnormal F^{18}FDG uptake. If only a single site on baseline: a markedly decreased intensity compared to baseline	**Residual mass ≥2 cm:** Lesion uptake < mediastinum	**Residual mass ≥2 cm:** Lesion uptake < liver uptake
	2	More than one site of residual uptake but with a marked decrease in number of disease sites compared to baseline.	**Residual mass ≥2 cm:** Lesion uptake=mediastinum	**Residual mass ≥2 cm:** Lesion uptake=liver uptake
Any focus of abnormal F^{18}FDG uptake (not related to physiological or benign tracer uptake).	3	Reduced intensity of uptake with no change in their number compared to baseline	**Residual mass ≥2 cm:** Moderately increased uptake compared with mediastinum OR Residual mass <2 cm: any focus of abnormal F^{18}FDG uptake (not related to physiological or benign uptake)	**Residual mass ≥2 cm:** Lesion uptake moderately increased compared with liver uptake OR Residual mass <2 cm: any focus of abnormal F^{18}FDG uptake (not related to physiological or benign uptake)
	4	No change in either number or intensity of sites or the appearance of new sites of disease	**Residual mass ≥2 cm:** Markedly increased uptake compared with mediastinum **OR** **Residual mass <2 cm:** any focus of abnormal F^{18}FDG uptake (not related to physiological or benign uptake)	**Residual mass ≥2 cm:** Lesion uptake markedly increased compared with liver uptake **OR** **Residual mass <2 cm:** Any focus of abnormal F^{18}FDG uptake (not related to physiological or benign uptake)
	Negative Positive FDG-PET/CT	Score 0-2 Score ≥3	Score 0-2 Score ≥3	Score 0-2 Score ≥3

study was defined as truly predictive when the patient had disease progression. A falsely positive predictive interim study was defined as a positive interim study in a patient who achieved complete remission and did not have progressive disease during the follow-up. Truly negative studies were defined as negative interim FDG-PET scans in patients who had continued remission during follow-up. A falsely negative predictive interim FDG-PET was defined as a negative interim study in a patient who had disease progression during follow-up.

Progression-free survival and overall survival were evaluated using Kaplan-Meier curves and the log-rank test.[13] The positive predictive value (PPV) was defined as the number of truly positive patients divided by the number of positive interim FDG-PET/CT results. The negative predictive value (NPV) was calculated as the number of truly negative patients divided by the number of patients with negative interim FDG-PET findings. All *P* values were two-sided with the level of statistical significance being 0.05. Accuracy was defined as the number of true positive and true negative outcomes divided by the number of all tested participants. Specificity, sensitivity and accuracy values were compared using McNemar's test. Statistical analyses were performed using SPSS15.0 (SPSS, Chicago, IL, USA).

Results

The characteristics of the 96 patients reported in this study are presented in Table 1. Ninety of the 96 patients (93%) included in the study achieved complete remission. Six patients had primary progressive disease (6%). Three relapses were recorded at a median follow-up of 59 months (range, 11-71 months), 11, 13 and 41 months following diagnosis (4, 5 and 36 months after the end of therapy). Interim FDG-PET/CT results using the static visual score and dynamic visual score, and the corresponding patients' outcome are presented in Figure 1A and 1B.

Patients were treated according to predefined risk criteria. Eight patients with early, favorable HL according to the EORTC criteria[12] (stage I or II, non-bulky, no B symptoms, less than four disease sites, age < 50 years, erythrocyte sedimentation rate <50 mm/h) received four cycles of ABVD and radiation therapy to the involved field. One of these patients who had a positive interim FDG-PET study was given dose-escalated chemotherapy. Twenty-five patients with intermediate risk HL[12] were planned to receive six cycles of ABVD. Two of these patients had primary progressive disease, one of them irrespective of treatment escalation after the second cycle, because of which was performed a positive interim FDG-PET.

Two patients who had two cycles of ABVD followed by positive interim FDG-PET/CT had escalation of therapy and received escalated BEACOPP. When initial therapy prior to positive FDG-PET/CT had been two cycles of standard BEACOPP, treatment was increased to escalated BEACOPP for four additional cycles of therapy (n=9). Patients in whom initial therapy was escalated BEACOPP had four additional cycles of this same therapy (n=3).

Thirty-nine patients were treated with radiation therapy following chemotherapy . Radiation was administered at 180 cGy fractions for a total dose of 2520 to 4000 cGy. Indications for radiotherapy were bulky mediastinal mass (n = 10), positive interim FDG-PET/CT (n = 11), or both (n = 2), or early disease (n =16) (Table 1).

Prospective static visual review of FDG-PET/CT results

The interim FDG-PET/CT was positive according to static visual criteria in 24 patients. Eleven of these patients had chemotherapy dose escalation followed by radiation therapy in 6 patients. Five of the 24 patients (21%) had primary progressive disease, three irrespective of dose escalation, and 19 were in remission at a follow-up of 41 to 79 months (median 66 months) (Figure 1A). In only 3 of the 19 patients was FDG-PET/CT positive at a single site at the end of therapy and eventually became negative without added therapy.

Five-year progression-free survival and overall survival rates for patients with positive interim FDG-PET/CT studies were 79% (95% CI: 71-87) and 87% (95% CI: 80-94), respectively. Seventy-three patients had a negative interim FDG-PET/CT according to the visual static score and 69 of them (95%) achieved complete remission. The 5-year progression-free survival rate for patients with negative interim FDG-PET/CT was 94% (95% CI: 89-99) and the overall survival rate was 99% (95% CI: 97-100); these rates were both significantly better than those of patients with positive interim studies (*P*=0.02 and *P*=0.017, respectively (Figure 2).

Out of the 41 individuals treated initially with standard BEACOPP (Figure 1A and Table 3), 11 had positive interim FDG-PET results. Nine of these latter patients had therapy intensified to escalated BEACOPP, with 5-year progression-free and overall survival rates of 73% and 82%, respectively. Two of these nine patients had primary progressive disease in spite of the interim escalation of therapy and seven patients had prolonged remission. Thirty patients with negative interim FDG-PET/CT had prolonged remissions with a median of 61 months (range, 28-93 months).

Twenty-two patients were initially treated with escalated BEACOPP. Six patients had a positive interim study (Figure 1A). Three patients who had a positive interim FDG-PET showing a single residual mediastinal mass and, therefore, received four further cycles of escalated BEACOPP followed by radiation therapy remained in complete remission. Two of these three patients had a positive scan 3 months following radiation therapy. However, biopsies appeared negative and they had no evidence of disease progression. Sixteen out of the 22 patients initially treated with escalated BEACOPP had interim negative FDG-PET/CT studies which resulted in reduction of their therapy to four cycles of standard BEACOPP and 15 out of these 16 patients remained in complete remission at a median follow-up of 70 months (range, 41-89 months); one patient had primary progressive disease (Figure 1A).

Thirty-three patients received initial therapy with ABVD. Seven of these patients had a positive interim FDG-PET/CT, two of whom had primary progressive disease and five of whom remained in complete remission at a median of 54 months (range, 43-79 months). Twenty-six of the 33 patients initially treated with ABVD had interim negative FDG-PET/CT; one of them had primary progressive disease, two relapsed and 23 remained in complete remission at a median of 55 months (range, 25-84 months) (Figure 1A).

Negative and positive predictive values, specificity, sensitivity and accuracy of interim PET/CT using the static visual score were 94%, 21%, 78%, 55%, 76%, respectively, for the whole study population and were similar to those obtained for patients treated with the different

chemotherapy protocols, as presented in Table 3.

Although the data obtained indicated that, overall, at least 21% of patients with a positive visual interim FDG-PET/CT would not be disease-free in 4 years, the PPV was difficult to interpret because of the relatively small number of positive studies (n=24) as well as the dilemma concerning outcome interpretation following therapy modification based on the results of interim scintigraphy. An additional calculation was, therefore, performed, excluding the eight patients who had dose intensification and no evidence of disease progression. This analysis eliminated the possibility that intensification of therapy was the cause of the low PPV. The PPV increased from 21% to 31% and the specificity from 78% to 86% (Table 3).

Retrospective dynamic visual review of FDG-PET/CT results

Using the 5-point dynamic visual score detailed above for interpretation of interim FDG-PET/CT results and defining score 2 as the cut-off point between positive and negative studies, the number of positive interim studies markedly decreased from 24 to 6, as presented in Figure 3 and Table 3. There were six patients with positive FDG-PET/CT studies, of whom four had a score of 3 and two had a score of 4, and 90 patients with negative interim PET/CT studies, including 72 with a score of 0, 13 with a score of 1 and 5 with a score of 2. Three of six patients with a positive interim FDG-PET/CT (two with a score of 3 and one with a score of 4) had disease progression, in spite of therapy escalation. Six patients with a negative interim FDG-PET/CT had treatment failure (7%), including 4 of 72 with score 0 (5.5%), 2 of 13 with score 1 (15%) and none of 5 patients with score 2. Eighty-four patients were in prolonged complete remission. The study results were then re-analyzed to retrospectively define the cut-off score which provided the highest accuracy and predictive values for detecting treatment failure by interim FDG-PET/CT (Table 4). According to this analysis FDG-PET/CT studies with scores 0, 1 or 2 were considered as negative and those with scores 3 or 4 as positive.

The 5-year progression-free and overall survival rates of

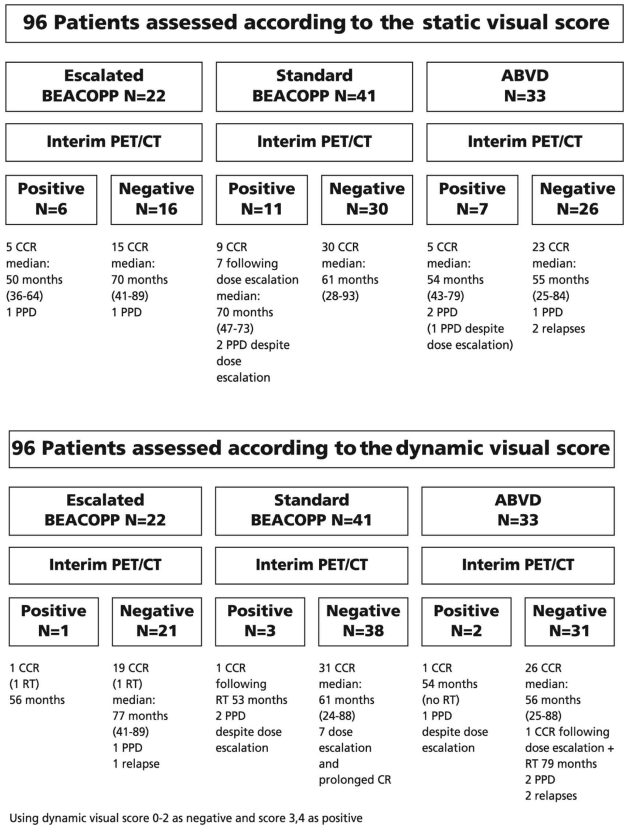

Figure 1. (A) Clinical outcome according to initial therapy protocol and interim FDG-PET/CT results using the binary visual score (B). Clinical outcome according to initial therapy protocol and interim FDG-PET/CT results using the dynamic visual score.

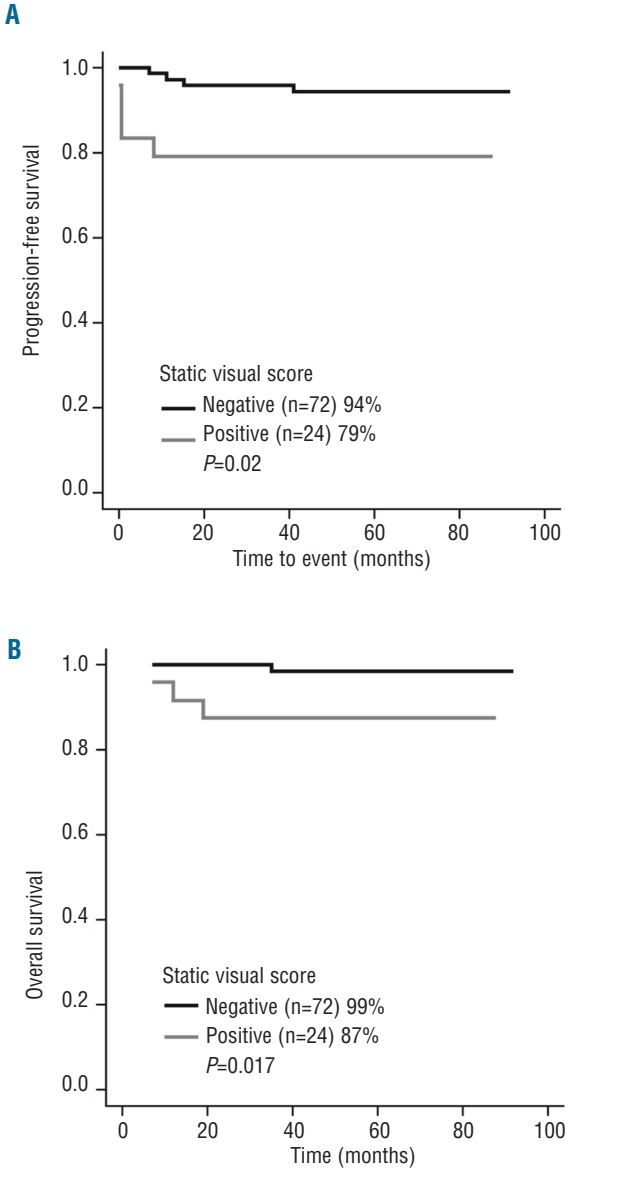

Figure 2. Five-year progression-free survival (A) and overall survival (B) of patients with positive (n=24) and negative (n=72) interim FDG PET/CT using the binary visual score.

patients with negative interim PET/CT using the dynamic visual score were 93% (95% CI: 88-98) and 98% (95% CI: 95-100), respectively, these rates being significantly better than the 50% (95% CI: 40-60) and 67% (95% CI: 58-76), respectively, for patients with a positive interim PET/CT (P=0.0001 and P<0.0001, respectively) (Figure 3).

The NPV, PPV, specificity, sensitivity and accuracy of dynamic visual review of interim PET/CT for predicting treatment failure were 93%, 50%, 96%, 33% and 91%, respectively. Specificity and accuracy were significantly better than those provided by interim PET/CT using the static visual score (P<0.0001 and P=0.0001, respectively) (Table 3).

Interim FDG-PET/CT results were also retrospectively assessed according to the guidelines of the Consensus of the Imaging Subcommittee.[7] Using these criteria, with either mediastinal or liver uptake as a comparator, there were 21 or 16 positive interim PET/CT studies respectively, and 75 or 80 negative studies. The NPV was 93% or 92%, PPV was 19% by both scores, specificity 80% or 85%, sensitivity 44% or 33% and accuracy 77% or 80%, respectively (Table 3).

When the four scoring systems were compared, the dynamic visual score provided the highest specificity and accuracy (96% and 91%) compared to the static visual score (P<0.0001 and P=0.001) or the two other static scores using either mediastinal blood pool as the comparator (P=0.0001 and P=0.001)[7] or liver blood pool uptake as the comparator (P<0.002 and P=0.006, respectively) (Table 3). The progression-free survival rate for patients with moderately or markedly increased uptake compared to that of the mediastinum was 81% (95% CI: 73-89) *versus* 93% (95% CI: 88-98) for patients with negative studies (P=0.06) (Figure 4A). For patients with an interim FDG-PET/CT uptake higher than liver uptake the 5-year progression-free survival rate was 81% (95% CI: 73-88)

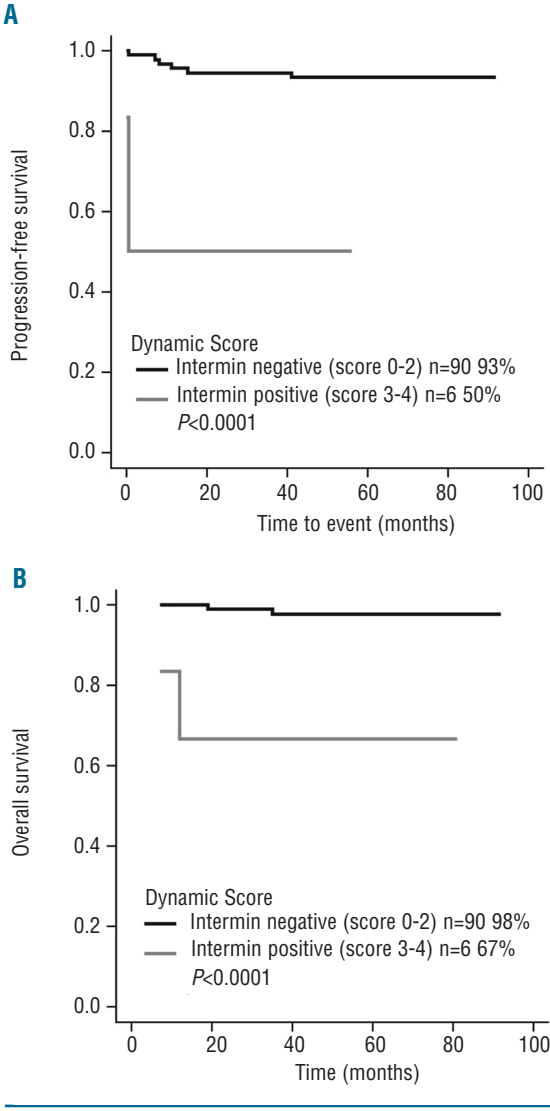

Figure 3. Five-year progression-free survival (A) and overall survival (B) of patients with positive (n=6) and negative (n=90) interim FDG PET/CT using the dynamic visual score.

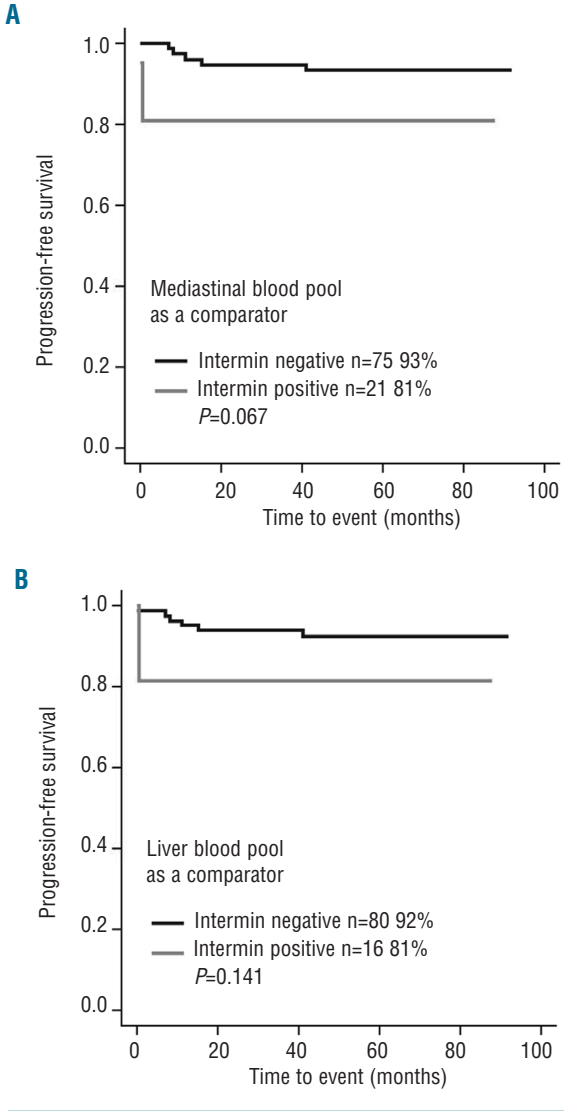

Figure 4. Five-year progression-free survival of patients with positive and negative interim FDG PET/CT using the mediastinal blood pool uptake (A) and liver blood pool uptake (B) as comparators.

whereas that for patients with negative studies was 92% (95% CI: 87-97) (*P*=0.14) (Figure 4B).

Therapy changes were a major limitation of the current study given the possibility that there could be shifts between true positive and false positive results if the adapted therapy was more effective than the unmodified therapy. In the additional analysis performed to partially resolve this bias, excluding patients who had escalation of therapy and no evidence of disease progression, the difference in specificity between the dynamic visual score (96%) and all three static scores remained statistically significant whether the comparator used was the visual score (86%; *P*=0.008), mediastinal blood pool uptake (87%; *P*=0.016) or liver blood pool uptake (89%; *P*=0.03). The difference in accuracy between interim FDG-PET/CT using the dynamic visual score and all three static scores (92% *versus* 83%) was not statistically significant.

Table 3. Performance of interim PET/CT for predicting treatment failure, using four scoring systems, according to therapeutic protocol. (*P* values are indicated in asterisks).

	NPV	PPV	Specificity	Sensitivity	Accuracy
Static visual score					
All patients (n=96)	(68/72) 94%	(5/24) 21%	(68/87) 78%*	(5/9) 55%	(73/96) 76%
(95% CI)	(88-99)	(5-37)	(69 -87)	(23-88)	(67-84)[b]
ABVD (n=33) (95% CI)	(23/26) 88% (75-100)	(2/7) 29% (0-63)	(23/28) 82% (68-96)	2/5 (0-83)	–
BEACOPP (n=41)	(30/30) 100%	(2/11) 18%	(30/39) 77%	2/2	–
Escalated BEACOPP (N=22)	(15/16) 94%	(1/6) 17%	(15/20) 75%	1/2	–
[®]Modified cohort (N=88) (95% CI)	(68/72) 94% (88-99)	(5/16) 31% (8-54)	(68/79) 86%# (78-94)	(5/9) 55% (22 -87)	(73/88) 83% (75-91)
Static score by CIS using mediastinal blood pool uptake as comparator[7,8]	(70/75) 93%	(4/21)19%	(70/87) 80%**	(4/9) 44%	(74/96) 77%[b]
(95% CI)	(87-99)	(2-36)	(72-88)	(12-76)	(68-85)
[®]Modified cohort (N=88) (95% CI)	(69/74) 93% (87-99)	(4/14)28% (4-51)	(69/79) 82%## (73-90)	(4/9) 44% (12-76)	(73/88) 83% (75-91)
Static score using liver blood pool uptake as comparator[9]	(74/80) 92%	(3/16) 19%	(74/87) 85%[a]	(3/9) 33%	(77/96) 80%[c]
(95% CI)	(86 -98)	(0-38)	(77-92)	(2 -64)	(72-88)
[®]Modified cohort (N=88) (95% CI)	(70/76) 92% (86-98)	(3/12) 25% (0-49)	(70/79) 88%@ (81-95)	(3/9) 33% (2 -64)	(73/88) 83% (75-91)
Dynamic visual score					
All patients (n=96) (95% CI)	(84/90) 93% (88-98)	(3/6) 50% (10-90)	(84/87) 96% (93-100)	(3/9) 33% (2 -64)	(87/96) 91% (85-96)
ABVD (n=33)	(27/31) 87%	(1/2) 50%	(27/28) 96%	1/5	
BEACOPP (n=41)	(38/38) 100%	(2/3) 66%	(38/39) 97%	2/3	
Escalated BEACOPP (n=22)	(19/21) 90%	0/1	(19/20) 95%	0/1	
[®]Modified cohort (N=88) (95% CI)	(76/82) 93% (87-98)	(3/6) 50% (10-90)	(76/79) 96% (91.6-100)	(3/9) 33% (2-64)	(79/86) 92% (83.5-96)

*Interim FDG-PET/CT analysis: dynamic visual scores of ≤2 were considered as negative studies; scores of ≥3 were considered positive. When compared to the dynamic visual score: *P <0.0001, ** P=0.0001, @ P=0.03, #P=0.008, ##P=0.016, [a]P=0.002, [b]P=0.001, [c]P= 0.006. [®]Excluding the eight patients who had intensification of therapy and no disease progression; CI: confidence interval.*

Discussion

During the last decade, F^{18}FDG imaging using either PET or PET/CT has become a major strategy both for staging patients with HL and for interim evaluation of their response to therapy. Hutchings *et al.*[4] and Gallamini *et al.*[14] have demonstrated that the findings of interim FDG-PET can predict response to therapy. A negative study result is associated with favorable disease-free survival, while a positive study predicts for a worse prognosis. A recently published prospective trial demonstrated a high prognostic predictive value of FDG-PET studies performed following two cycles of ABVD in a cohort of 260 patients with early unfavorable or advanced disease, with a PPV of 86% and a NPV of 95% at 2.19 years of follow-up.[1] Following publication of these results, several ongoing randomized trials are currently assessing the role of interim F^{18}FDG-PET imaging in patients with early favorable, unfavorable and advanced disease.[15,16] It has not yet been determined whether the high predictive value of FDG-PET is maintained while using aggressive therapeutic protocols such as escalated BEACOPP or standard BEACOPP. It is also still unclear what the precise definition of response on interim FDG-PET/CT studies should be. In contrast to FDG PET/CT studies at the end of therapy, it may be that early prediction of good response during therapy does not necessarily require complete disappearance of FDG uptake at all sites of disease.

In the current study, we used a new simple and reproducible dynamic visual score which expresses the dynamic metabolic trend from diagnosis to early during initial treatment and not only the metabolic status of disease at a single time point during therapy. Using this score, FDG PET/CT was found to be highly predictive of both progression-free and overall survival, similar to previously reported data using other interpretation criteria. The current study, however, included patients treated with three different protocols, ABVD, BEACOPP or escalated BEACOPP, who have been followed for longer periods (median 59 months; range, 7-91 months) than previously reported. The NPV of FDG-PET/CT in patients treated with any of these three regimens was similar to that reported by Gallamini *et al.*[1] for patients treated with ABVD. Regarding PPV, our data indicated that at least 21%-31% of patients with a positive interim FDG-PET/CT would not be disease-free at 4 years. Nevertheless, the PPV in the present study, obtained with either method of scoring (21-50%), was markedly lower than previously reported values (69%-86%).[1,3,4,14] Assessment of the PPV could have been negatively biased in our cohort of patients because of the use of a binary visual score in which every abnormal uptake was considered positive, while in other studies minimal residual uptake or even uptake equal to or slightly higher than mediastinal uptake was considered negative.[1,3] Since treatment intensification for patients with positive FDG-PET who had no disease progression is another potential bias, due to a possible shift from true positive to false positive if the intensified therapy was more efficient, the analyses were also recalculated omitting those patients who received treatment intensification, revealing a similar PPV of 25-50%. Another possible explanation for the low PPV may be the low rate of treatment failure in the present cohort (9 patients, 9%). The low PPV may also reflect a greater potency of BEACOPP regimens to elimi-

Table 4. Evaluation of different dynamic score cut-off points for definition of negative and positive interim FDG-PET/CT.

96 patients	NPV (95% CI)	PPV (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Accuracy % (95% CI)
Score 0 versus 1-4	94% (89-99)	21% (4.6-37)	78% (69-70)	56% (23 -88)	76% (67.4-84.)
Score 0-1 versus 2-4	93% (87-98)	27% (1-53)	91% (85-97)	33% (2-64)	85% (78-92)
Score 0-2 versus 3,4	93% (88-98)	50% (10-90)	96% (93-100)	33% (2-64)	91% (85-96)

nate a single residual site uptake. This possibility is supported by the fact that while 43/50 patients treated with ABVD who had a positive interim FDG-PET were reported to have a positive FDG-PET also at the end of therapy,[1] in the current study only 5 of 24 patients with positive interim FDG-PET/CT studies (21%) ultimately remained positive. All of these five patients had primary progressive disease. Three additional studies reported a low PPV following BEACOPP therapy. Avigdor *et al.*[17] reported a similar low PPV of 46% in a cohort of 45 patients with advanced HL and International Prognostic Score of 3 or more, treated with two cycles of escalated BEACOPP followed by interim FDG-PET/CT. Markova *et al.* described a cohort of 50 patients treated with standard or escalated BEACOPP who had an interim FDG-PET following four cycles of therapy. While the NPV was found to be high, the PPV was low and only two of 14 patients with a positive interim study had disease progression.[18] Gallamini *et al.* reported a PPV of 60%, a NPV of 88% and specificity of 92% in a cohort of 30 patients with advanced HL treated with four cycles of escalated BEACOPP followed by four cycles of standard BEACOPP with a median follow up of 2 years.[19]

The major issue we assessed in the present study was the lack of an accurate, accepted definition of response on interim FDG-PET/CT. In order for FDG-PET/CT to provide information of practical clinical significance early during treatment, FDG-PET/CT response criteria need to be clearly defined and standardized. Juweid *et al.*[7] published criteria for the definition of metabolic response of lymphoma on FDG-PET at conclusion of therapy, based on both the intensity of F^{18}-FDG uptake and the size of the residual mass at initial sites of disease. These criteria were adopted in the Revised Response Criteria for Malignant Lymphoma.[8] At the end of therapy, any increased F^{18}FDG uptake in a mass smaller than 2 cm or residual masses of 2 cm or more with an uptake greater than that in the mediastinum was suggested to be positive, indicating the presence of active lymphoma.[7] To date, there are insufficient data to determine whether these end-of-therapy criteria apply for interim FDG-PET/CT therapy assessment as well. It does, however, seem that early during treatment, a higher threshold should be used to define the presence of active lymphoma, especially if intensification of treatment is anticipated based on a positive study. Hutching and Mikhaeel evaluated 85 patients who underwent interim FDG-PET.[3] A positive interim FDG-PET was defined as "increased uptake suspicious of malignant disease, which did not have a benign explanation." The presence of "minimal residual uptake" was considered negative. Using these criteria the PPV of interim FDG-PET increased from 41% to 62%.[3] Juweid further raised the threshold for defining increased F^{18}FDG uptake as positive for active lymphoma, suggesting that above normal liver F^{18}FDG uptake should be used to signify a positive interim FDG-PET/CT. Meignan *et al.* are using this same threshold in the H10 intergroup study.[20] This group, however, has also shown that a comparison between interim FDG-PET and baseline FDG-PET results provides a significantly higher inter-observer agreement, emphasizing the importance of pre-treatment scanning for subsequent assessment of response, which is in agreement with the guidelines of the National Cancer Institute.[21] In the current study, the initial prospective methodology for the interim FDG-PET/CT interpretation defined any abnormal F^{18}FDG uptake as positive. A low PPV of 21% to 31% and a specificity of 75% to 88% indicated the potential for unwarranted therapeutic escalation. In our retrospective analysis, the performance of interim FDG-PET/CT studies was further refined by considering the dynamic sequence of response on FDG-PET/CT, as reflected by the comparison of changes in the number and intensity of F^{18}FDG-avid disease sites between the pre-therapy and interim studies (Table 3). The specificity of the dynamic visual score remained stable whether the whole group of patients was included or whether the ones who had escalation of therapy were excluded. This is due to the fact that the number of false positives was smaller when the dynamic visual score assessment method was used (3 patients only). The dynamic scoring system in this cohort of patients provided significantly superior specificity and accuracy compared to the three static scoring systems.

In conclusion, the current study confirms the value of interim FDG PET/CT for predicting treatment failure in HL in a cohort of patients followed for a longer time than previously reported. FDG-PET/CT demonstrated a consistently similar performance, with a high NPV in patients treated with three different chemotherapy protocols. The data also suggest that for practical purposes patients with an interim FDG-PET score of 0 or 1 should be managed as having a negative study and reduction of therapy from escalated BEACOPP to the standard regimen should be considered for such individuals. On the other hand, patients with an interim FDG-PET score of 3 or 4 are at high risk of disease progression.

Importantly, while the limitations of studying a relatively small number of subjects is recognized, the findings of the present investigation nevertheless suggest that a dynamic visual scoring system improves the PPV and specificity of interim FDG-PET/CT as compared to a static FDG-PET/CT scoring at a single time point during therapy. Larger studies and further incorporation of a dynamic scoring system into prospective studies relying on interim FDG-PET/CT monitoring are needed to validate and confirm these findings.

Authorship and Disclosures

EJD and RBS contributed equally to this manuscript. EJD and RBS were principal investigators and take primary responsibility for the paper. AT participated in the statistical analysis. RE, IA, MBS recruited the patients. DG interpreted vital data. JMR wrote the paper.

The authors reported no potential conflicts of interest.

References

1. Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. J Clin Oncol. 2007;25(24):3746-52.
2. Kirby AM, Mikhaeel NG. The role of FDG PET in the management of lymphoma: what is the evidence base? Nucl Med Commun. 2007;28(5):335-54.
3. Hutchings M, Mikhaeel NG, Fields PA, Nunan T, Timothy AR. Prognostic value of interim FDG-PET after two or three cycles of chemotherapy in Hodgkin lymphoma. Ann Oncol. 2005;16(7):1160-8.
4. Hutchings M, Loft A, Hansen M, Pedersen LM, Buhl T, Jurlander J, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. Blood. 2006;107(1):52-9.
5. Terasawa T, Lau J, Bardet S, Couturier O, Hotta T, Hutchings M, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. J Clin Oncol. 2009; 27(11):1906-14.
6. Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET scan in lymphoma. Leuk Lymphoma. 2009:50(8):1257-60.
7. Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guermazi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. J Clin Oncol. 2007;25(5):571-8.
8. Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. J Clin Oncol. 2007;25(5):579-86.
9. Gallamini A. The prognostic role of positron emission tomography scan in Hodgkin's lymphoma. Hematology Education: The Education Program for the Annual Congress of the European Hematology Association. 2009;3:144-50.
10. Dann EJ, Bar-Shalom R, Tamir A, Haim N, Ben-Shachar M, Avivi I, et al. Risk-adapted BEACOPP regimen can reduce the cumulative dose of chemotherapy for standard and high-risk Hodgkin lymphoma with no impairment of outcome. Blood. 2007; 109(3):905-9.
11. 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(21):1506-14.
12. Noordijk EM, Carde P, Dupouy N, Hagenbeek A, Krol AD, Kluin-Nelemans JC, et al. Combined-modality therapy for clinical stage I or II Hodgkin's lymphoma: long-term results of the European Organisation for Research and Treatment of Cancer H7 randomized controlled trials. J Clin Oncol. 2006;24(19):3128-35.
13. Kaplan EL, Meier P. Non parametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457-81.
14. Gallamini A, Rigacci L, Merli F, Nassi L, Bosi A, Capodanno I, et al. The predictive value of positron emission tomography scanning performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. Haematologica. 2006;91(4):475-81.
15. Andre M RJ, van der Maazen RW, Reman o, van't Veer M, Lutgenburg E, Girinski T, et al. The H10 EORTC/GELA randomized intergroup trial on early FDG-PET scan guided treatment adaptation versus standard combined modality treatment in patients with supradiaphragmatic stage I/II Hodgkin's lymphoma. Haematologica. 2007;92:P7.
16. Johnson P FM, Enblad G, Burton C, Smith P, Qian W, O'Doherty M, Radford J. Response-adapted therapy using FDG-PET scanning after initial ABVD: the UK NCRI, Italian and Nordic trial in advanced Hodgkin lymphoma. Haematologica. 2007;92:P7.
17. Avigdor A BS, Dann E, Levi I, Perez-Avraham G, Shemtov N, Shimoni A, et al. Combined escalated BEACOPP-ABVD therapy for advanced Hodgkin's lymphoma patients with high IPS score: an effective regimen and low positive predictive value of early FDG-PET/CT. Haematologica. 2007;92:P66.
18. Markova J, Kobe C, Skopalova M, Klaskova K, Dedeckova K, Plutschow A, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. Ann Oncol. 2009;20(7):1270-4.
19. Gallamini A, Bonfante V, Levis A, Di Raimondo F, Merli F, Vitolo U, et al. Early interim FDG-PET during intensified BEACOPP therapy shows a lower predictive value than during conventional ABVD chemotherapy. Haematologica. 2007;92(Suppl 2):144.
20. Meignan M, Itti E, Bardet S, Lumbroso J, Edeline V, Olivier P, et al. Development and application of a real-time on-line blinded independent central review of interim PET scans to determine treatment allocation in lymphoma trials. J Clin Oncol. 2009;27(16):2739-41.
21. Shankar LK, Hoffman JM, Bacharach S, Graham MM, Karp J, Lammertsma AA, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. J Nucl Med. 2006;47(6):1059-66.