Observational Study

Medicine®

OPEN

Pediatric Hodgkin Lymphoma

Predictive value of interim ^{18}F-FDG PET/CT in therapy response assessment

Cristina Ferrari, MD[a], Artor Niccoli Asabella, MD, PhD[a,*], Nunzio Merenda, MD[a], Corinna Altini, MD[a], Margherita Fanelli, MStat[a], Paola Muggeo, MD[b], Francesco De Leonardis, MD[b], Teresa Perillo, MD[b], Nicola Santoro, MD[b], Giuseppe Rubini, MD[a]

Abstract

We investigated the prognostic value of interim ^{18}F-FDG PET/CT (PET-2) in pediatric Hodgkin lymphoma (pHL), evaluating both visual and semiquantitative analysis.

Thirty pHL patients (age ≤16) underwent serial ^{18}F-FDG PET/CT: at baseline (PET-0), after 2 cycles of chemotherapy (PET-2) and at the end of first-line chemotherapy (PET-T). PET response assessment was carried out visually according to the Deauville Score (DS), as well as semiquantitatively by using the semiquantitative parameters reduction from PET-0 to PET-2 (ΔΣSUVmax0–2, ΔΣSUVmean0–2). Final clinical response assessment (outcome) at the end of first-line chemotherapy was the criterion standard, considering patients as responders (R) or nonresponders (NR). Disease status was followed identifying patients with absence or relapsed/progression disease (mean follow-up: 24 months, range 3–78).

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of visual and semiquantitative assessment were calculated; furthermore, Fisher exact test was performed to evaluate the association between both visual and semiquantitative assessment and outcome at the end of the first-line chemotherapy. The prognostic capability of PET-2 semiquantitative parameters was calculated by ROC analysis and expressed as area under curve (AUC). Finally, progression-free survival (PFS) was analyzed according to PET-2 results based on the 5-point scale and semiquantitative criteria, using the Kaplan–Meier method.

Based on the outcome at the end of first-line chemotherapy, 5 of 30 patients were NR, the remnant 25 of 30 were R. Sensitivity, specificity, PPV, NPV, and accuracy of visual analysis were 60%,72%,30%,90%,70%; conversely, sensitivity, specificity, PPV, NPV, and accuracy of semiquantitative assessment were 80%, 92%, 66.7%, 95.8%, 90%. The highest AUC resulted for ΔΣSUVmax0–2 (0.836; cut-off <12.5; sensitivity 80%; specificity 91%). The association between ΔΣSUVmax0–2 and outcome at the end of first-line chemotherapy resulted to have a strong statistical significance ($P=0.0026$). Both methods demonstrated to influence PFS, even if the semiquantitative assessment allowed a more accurate identification of patients with a high risk of treatment failure ($P=0.005$).

Our preliminary results showed that PET-2 visual assessment, by using Deauville criteria, can be improved by using the semiquantitative analysis. The SUV max reduction (ΔΣSUVmax0–2) evaluation might provide a support for the interpretation of intermediate scores, predicting with good confidence those patients who will have a poor outcome and require alternative therapies.

Abbreviations: Δ = decrease, Σ = sum, ^{18}F-FDG PET/CT = fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, ABVD = doxorubicin, bleomycin, vinblastine, dacarbazine, AUC = area under curve, CECT = contrast-enhanced computed tomography, COPP/ABV = cyclophosphamide, vincristine, prednisone, procarbazine/ doxorubicin, bleomycin, vinblastine, CR = complete response, DS = Deauville Score, FN = false negative, FP = false positive, MIP = maximum-intensity-projection, NPV = negative predictive value, NR = non-responders, PD = progressive disease, PET-0 = PET baseline, PET-2 = PET ad interim, PET-T = PET at the end of first-line chemotherapy, PFS = progression-free survival, pHL = Pediatric Hodgkin lymphoma, PPV = positive predictive value, PR = partial response, R = responders, ROC = receiver-operating-characteristics, SD = stable disease, SUVmax = maximum standardized uptake value, SUVmean = mean standardized uptake value, TG = therapeutic groups, TN = true negative, TP = true positive, VOI = volume of interest.

Keywords: Deauville criteria, interim ^{18}F-FDG PET/CT, pediatric Hodgkin lymphoma, pHL, semiquantitative analysis

1. Introduction

Pediatric Hodgkin lymphoma (pHL) constitutes approximately 40% of all childhood lymphomas and represents the most common malignancy in adolescents and young adults.[1] Up to 80% of pHL treated with the current radio/chemotherapy protocols will be cured, with 5-year event-free survival rates >90% achieved by the current therapy protocols.[2,3] Unfortunately, a significant number of patients will suffer from treatment-related morbidity and mortality caused by anthracyclines, alkylating agents, and radiotherapy. In several series, mortality resulting from secondary cancers and heart diseases has exceeded lymphoma-related deaths after 15 to 20 years of follow-up.[4–6] For these reasons, reducing treatment-associated toxicity

Editor: Gilbert Fruhwirth.

The authors report no conflicts of interest.

[a] Nuclear Medicine Unit, D.I.M., [b] Department of Pediatric Hematology and Oncology, University of Bari "Aldo Moro", Bari, Italy.

[*] Correspondence: Artor Niccoli Asabella, Piazza G. Cesare, 11, 70124 Bari, Italy (e-mail: artor.niccoliasabella@uniba.it).

Medicine (2017) 96:5(e5973)

Received: 28 April 2016 / Received in final form: 19 December 2016 / Accepted: 1 January 2017

http://dx.doi.org/10.1097/MD.0000000000005973

Table 1

Therapeutic groups and related treatment approach, according to the treatment optimization protocol (AIEOP-LH2004). (Available at: http://www.aieop.org/web//?q=node/376).

TG	Treatment approach
TG1 stage IA–IIA; no bulky mediastinal disease (M/T <0.33); no hilar lung lymph nodes involvement; <4 sites of disease involved	3 cycles of ABVD (doxorubicin, bleomycin, vinblastine,dacarbazine) ± radiotherapy of involved sites
TG2 patients not included in TG1 and in TG3	4 cycles of COPP/ABV (cyclophosphamide, vincristine, prednisone, procarbazine/ doxorubicin, bleomycin, vinblastine) + radiotherapy of involved sites
TG3 stage IIIB-IVA - IVB bulky mediastinal disease (M/T ≥0.33) any stage	6 cycles of COPP/ABV (cyclophosphamide, vincristine, prednisone, procarbazine/ doxorubicin, bleomycin, vinblastine) + radiotherapy of involved sites.

TG = therapeutic group.

while maintaining high cure rates is the main goal of current therapeutic strategies.

Owing to appearance of metabolic changes earlier than anatomical ones, fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) has proven to be the modality of choice for monitoring and for tailoring response-adapted treatment strategies both for early assessment during therapy (interim PET) and for remission assessment at the end of treatment.[7–11] Interim PET has demonstrated to be an accurate predictor of prognosis, stronger than the currently available prognosis scores, and able to perform tailored treatment regimens in adult affected by HL.[12–14] A negative interim PET is associated with 90% to 95% progression-free survival (PFS) regardless of consolidative radiotherapy.[11]

These results have been recently reproduced in a pediatric population but, although ongoing pediatric protocols recommend interim PET evaluation with subsequent PET-guided intensification or reduction of the amount of treatment, experience is still limited and there is a lack of standardization in this subset of patients.

The interim PET response criteria, based on the 5-point scale introduced by the first international workshop on interim PET in lymphoma and widely used to assess early response to therapy in adults, did not preclude interobserver reproducibility issues and mostly have not been sufficiently validated in pediatric patients.[15,16] Alternative approaches to visual analysis are developing to improve the accuracy and reproducibility of interim PET, mainly based on ^{18}F-FDG PET/CT semiquantitative parameters.[17]

Aim of this study was to evaluate the prognostic value of interim PET in pediatric patients affected by HL analyzing both methods: visual assessment according to Deauville score (DS) and semiquantitative assessment through ^{18}F-FDG PET/CT semiquantitative parameters.

2. Methods

2.1. Patients

This retrospective study included 30 pediatric patients with histological diagnosis of HL performed on excisional biopsy of lymphoid tissue according to the 2008 World Health Organization classification[18] and measurable nodal and/or extranodal disease. Exclusion criteria were age younger than 1 year and older than 16 years at date of initial diagnosis, previous chemotherapy and/or history of malignancy, life-threatening impairment of organ function and diabetes mellitus.

In addition to conventional evaluation, that included patient's physical examination, standard laboratory tests, bone marrow biopsy, and contrast-enhanced computed tomography (CECT) of the chest, abdomen, and pelvis, all patients underwent serial ^{18}F-FDG PET/CT, according to the systematic PET evaluation for ^{18}F-FDG-avid lymphoma: at baseline (PET-0), ad interim that is after 2 cycles of chemotherapy (PET-2), and at the end of first-line chemotherapy (PET-T).

Staging was performed according to the Ann Arbor classification;[19] then, all patients were subsequently stratified to the appropriate therapeutic groups (TG), according to treatment optimization protocol (AIEOP-LH2004) and subjected to the related treatment regimen, as reported in Table 1 (Available at: http://www.aieop.org/web//?q=node/376).

After the end of first-line chemotherapy, all patients were restaged according to the Cheson's Revised Response Criteria,[20] classifying patients in complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Response to therapy was assessed on the basis of both PET-T results and size of residual lesions measurable on CECT images, taking into account also physical examinations, blood tests, and a possible biopsy and histopathological examination of lesions suspected for relapse. All these information were verified by an interdisciplinary tumor board, which gave a final clinical response assessment for the outcome at the end of the first-line chemotherapy, considered the standard of reference. Patients in CR and PR with a tumor volume reduction >75% were considered responders (R), whereas patients in PR with a tumor volume reduction <75%, SD, and PD were considered non responders (NR). NR patients at the end of first-line chemotherapy were subjected to further treatment (second-line chemotherapy, radiotherapy, and/or autologous stem cell transplant).

Disease status was followed in reference to the available clinical and imaging information, identifying patients with absence of disease and patients with relapse and/or PD at follow-up. Mean follow-up from the end of first-line chemotherapy was 24 months (range 3–78). All patients gave their consent for the use of their data for clinical research.

Patients' clinical and pathological characteristics are described in Table 2.

2.2. ^{18}F-FDG PET/CT Technique

Images were acquired with a Discovery LSA PET/CT device (GE Healthcare, Waukesha, WI) that integrates a PET (advance n x I) with 16-slice CT scanner (light speed plus). All patients were instructed to fast for at least 8 hours before ^{18}F-FDG administrations and had capillary blood glucose of <160 mg/mL. The image acquisition was obtained 50 minutes after the intravenous injection of 4.6 MBq/kg of ^{18}F-FDG. Patients were hydrated by drinking 500 mL of water and urinated as needed. The CT scan

Table 2

Patients' clinical and pathological characteristics.

	n = 30	
Sex		
Male	16	53.3%
Female	14	46.7%
Median age, y	12.8	Range 2–16
Histological subtype		
Nodular sclerosis HL	27	90%
Lymphocytic-predominant HL	1	3.3%
Mixed cellularity HL	2	6.7%
Site		
nodal	30	100%
extranodal	13	43.3%
Bulky disease		
yes	18	60%
no	12	40%
Stage		
II A/B	9/7	50%
III A/B	3/5	26.7%
IV A/B	2/4	23.3%
TG		
TG1	0	0%
TG2	5	16.7%
TG3	25	83.3%

HL = Hodgkin lymphoma, TG = therapeutic group.

was carried out from the external acoustic meatus to the root of the thigh with patients lying on their back with hands above their head. The CT acquisition parameters were 340 mA (auto), 120 kV, slice thickness 3.75 mm, tube rotation time 0.8 ms, and collimation field of view of 50 cm. The CT images were reconstructed with a filtered back projection. The CT data were used for the attenuation correction of PET scanning, which was performed immediately after the acquisition of CT images. The CT scans were performed without administration of contrast enhancer. The PET acquisition was obtained in caudal-cranial direction; PET was reconstructed with a matrix of 128×128, ordered subset expectation maximum iterative reconstruction algorithm (2 iterations, 28 subsets), 8-mm Gaussian filter, and 50-cm field of view. The CT, PET, and coregistered PET/CT images were reviewed in transaxial, coronal and sagittal planes along with maximum-intensity-projection (MIP) whole-body images.

2.3. ^{18}F-FDG PET/CT Interpretations and Analysis

All ^{18}F-FDG PET/CT images (PET-0, PET-2, and PET-T) were evaluated in consensus by 2 experienced nuclear physicians blinded to the corresponding patients' medical history and clinical results. Disagreements were resolved with a third observer as referee.

^{18}F-FDG PET/CT exams were considered positive in case of any increased ^{18}F-FDG uptake detected in nodal basins or extranodal sites, unrelated to physiologic or benign uptake (infection or inflammation) and considered as involved by disease. In particular, PET-2 images were graded as negative or positive by a side-by-side comparison with PET-0, as follows: PET-0 and PET-2 images were directly visualized on Xeleris workstation and compared through the Volumetrix MI software to assess the presence or absence in the interim PET of abnormal ^{18}F-FDG uptake in sites of known disease detected on the baseline study, as showed in Figure 1.

The analysis was performed both visually and by means of semiquantitative parameters.

2.3.1. PET-2 Visual analysis. The visual response to therapy assessment was carried out according to the Deauville 5-point scale Criteria.[21] Briefly, a DS of 1 indicated no residual uptake above the background level, a DS of 2 indicated residual uptake less than or equal to the mediastinum, a DS of 3 indicated residual uptake greater than the mediastinum but not greater than the liver, a DS of 4 indicated residual uptake moderately increased compared with the liver, and a DS of 5 indicated residual uptake markedly increased compared with the liver or new sites of disease. PET-2 was considered visually positive when the residual ^{18}F-FDG uptake was superior to the liver uptake (DS 4 or 5), whereas it was considered visually negative in the remnant cases (DS 1, 2, or 3).

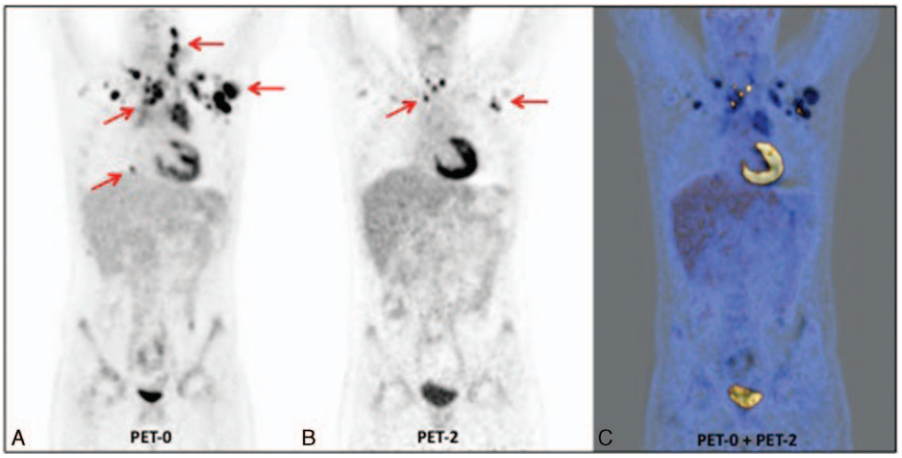

Figure 1. ^{18}F-FDG PET/CT performed in a 16-year-old female affected by Nodular Sclerosis Hodgkin lymphoma at stage IIIA. (A) PET-0 (coronal view): intense ^{18}F-FDG uptake in the upper diaphragmatic lymph nodal basins (red arrows), considered involved by disease. (B) PET-2 (coronal view): residual involvement in the mediastinal and axillary lymph nodes (red arrows) that showed moderately increased ^{18}F-FDG uptake compared to the liver. (C) Overlapped PET-0 (black areas)/PET-2 (yellow areas): the direct comparison of the 2 images provides an intuitive response evaluation. ^{18}F-FDG PET/CT = fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; PET-0 = PET baseline; PET-2 = PET ad interim.

2.3.2. PET-2 Semi-quantitative analysis. Volume of interest (VOI) was drawn semiautomatically on each nodal basins or extranodal sites with pathologically increased ^{18}F-FDG uptake visually detected on axial, coronal, or sagittal PET/CT slices, both at PET-0 and PET-2, if still positive, as showed in Figure 2. Semiquantitative analysis was performed calculating max and mean standardized uptake values (SUVmax and SUVmean), using the maximum and mean activity values within each VOI, corrected for the injected activity dose and patient's body weight, employing the following formula:

$$\mathrm{SUV} = \frac{\text{tissue activity}_{\mathrm{max/mean}}\,(\mathrm{kBq/mL})}{\text{injected activity (MBq)/weight (kg)}}$$

The sum (Σ) of all SUVmax and SUVmean detected on each site involving both at PET-0 (ΣSUVmax-0, ΣSUVmean-0) and PET-2, if still positive (ΣSUVmax-2, ΣSUVmean-2), was calculated in each patient. Then, the PET-2 semiquantitative response assessment was carried out by considering the decrease (Δ) of each parameter from PET-0 to PET-2 (ΔΣSUVmax 0–2, ΔΣSUVmean 0–2).

2.4. Statistical analysis

As previously described about the PET-2 visual response assessment, a DS ≥4 was considered discriminating in distinguishing patients with a poor response from those with a good response to therapy. Conversely, regarding the semiquantitative analysis, receiver-operating characteristics (ROC) curves were used for determining the best cutoff value of PET-2 semiquantitative parameters, expressed as area under curve (AUC), able to distinguish responders from nonresponders patients with the best sensitivity and specificity.

By using the final clinical response assessment as standard of reference, regarding the qualitative visual rating of the interim PET, patients with PET-2 negative (DS 1–3) or PET-2 positive (DS ≥4) and considered CR at the end of first-line chemotherapy, were considered as true-negative (TN) and as false-positive (FP) respectively. Conversely, patients still affected by lymphoma at the end of first-line chemotherapy and classified to be PET-2 positive (DS ≥4) or PET-2 negative (DS 1–3), were referred to be true-positive (TP) and false-negative (FN) respectively.

Similarly, patients were considered TN, TP, FN, or FP with the semiquantitative assessment, by using the cut-off we found with the ROC analysis.

Then, the PET-2 sensitivity, specificity, positive and negative predictive values (PPV and NPV) of both visual and semiquantitative assessment were calculated for the outcome prediction. Moreover, diagnostic accuracy of PET-2 visual and semiquantitative analysis was calculated, expressed as a portion of correctly classified patients (TP and TN) among all patients.

Fisher exact test was performed to evaluate whether any association between both visual and semi-quantitative assessment and outcome at the end of first-line chemotherapy existed.

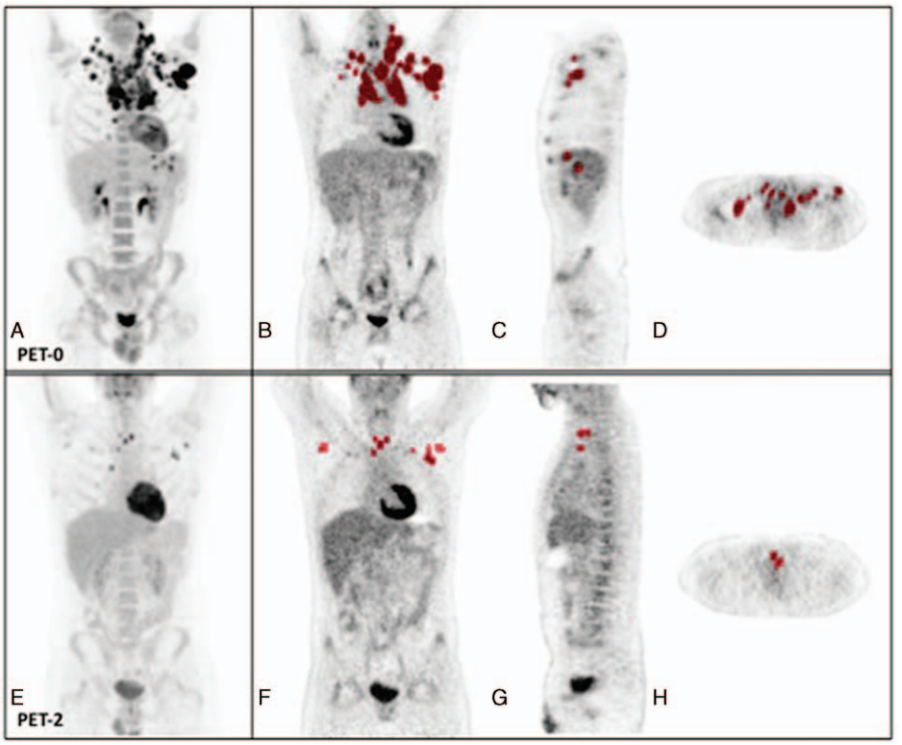

Figure 2. ^{18}F-FDG PET/CT performed in a 16-year-old female affected by Nodular Sclerosis Hodgkin Lymphoma at stage IIIA, at baseline (A–D) and after 2 cycles of chemotherapy (E–H). (A, E) MIP whole-body image of PET-0 and PET-2, respectively. VOI was drawn semiautomatically on each nodal basins with pathologically increased ^{18}F-FDG uptake visually detected on PET-0, and still present on PET-2, in the coronal (B, F), sagittal (C, G), and transaxial plans (D, H), respectively (red areas). The Σ of SUVmax and SUVmean of each VOI at baseline were 20.60 and 11.40, respectively, whereas the Σ of SUVmax and SUVmean of each VOI at PET-2 were: 9.90 and 6.50, respectively. ΔΣSUVmax 0–2 and ΔΣSUVmean 0–2 resulted to be: 10.70 and 4.90, respectively. Abbreviations: ^{18}F-FDG PET/CT = fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography; PET-0 = PET baseline; PET-2 = PET ad interim; MIP = maximum-intensity-projection; VOI = volume of interest; SUVmax = maximum standardized uptake value; SUVmean = mean standardized uptake value; Σ = sum; Δ = decrease; ΔΣSUVmax 0–2 = decrease of the sum of SUVmax from PET-0 to PET-2; ΔΣ SUVmean 0–2 = decrease of the sum of SUVmean from PET-0 to PET-2.

Finally, progression-free survival (PFS) was analyzed according to PET-2 results based on the 5-point scale and semiquantitative criteria, using the Kaplan–Meier method and compared using the log-rank test. The PFS was defined as the time from the beginning of treatment until progression, relapse, or death from any cause or the date of last follow-up. P values <0.05 were regarded as statistically significant.

Statistical analysis was performed using MedCalc software version 14.12.0 (MedCalc Software bvba, Ostend, Belgium).

3. Results

3.1. Treatment outcome

Based on the final clinical response assessment at the end of the first-line chemotherapy, 5 of 30 (16.7%) patients were considered NR. All of them were in TG3. Among them, 2 became R after second-line treatment (nodular sclerosis HL staged IVA; nodular sclerosis HL staged IIB), another one remained NR (nodular sclerosis HL staged IVB), whereas the other 2 patients died (mixed cellularity HL staged IIB; nodular sclerosis HL staged IIA). All the other patients (25/30; 83.3%) were considered R.

3.2. PET-2 visual assessment

By using the visual analysis, 20 of 30 patients (66.7%) resulted PET-2-negative, with DS = 1 in 15 of 20 (75%), DS = 2 in 1 of 20 (5%), and DS = 3 in 4 of 20 (20%), respectively, whereas 10 of 30 patients (33.3%) were PET-2 positive and all of them had DS = 4. No patients with DS = 5 occurred in our sample.

Among the PET-2-negative patients, 2 of 20 resulted FN compared to the final clinical response assessment at the end of first-line chemotherapy (mixed cellularity HL staged IIB, TG3; nodular sclerosis hl staged IVA, TG3), whereas 7 of 10 PET-2-positive patients resulted FP. Among FP patients, 3 were staged II, 2 were staged III, and 2 staged IV; just 1 was in TG2, the remnant 6 were in TG3. All patients were affected by Nodular Sclerosis HL.

Visual PET-2 response assessment, according to Deauville criteria, showed a sensitivity of 60%, specificity of 72%, PPV of 30%, NPV of 90%, and accuracy of 70%. No association between visual assessment and outcome at the end of first-line chemotherapy was found (Fisher exact test $P = 0.300$). (Table 3)

3.3. PET-2 semiquantitative assessment

The mean values (±s.d.) of ΔΣSUVmax 0–2 were 44.72 (±6.91) and 16.16 (± 8.15) in R and NR groups, respectively, whereas the mean values (±s.d.) of ΔΣ SUVmean 0–2 were 5.30 (± 0.64) and 2.88 (± 0.74) in R and NR groups, respectively.

The performance of the ROC curves in determining what is the best semiquantitative parameter and its optimal cutoff for identifying good and poor responders is presented in Figure 3. The highest AUC resulted for ΔΣSUVmax 0–2 (0.836) and the best cut-off was <12.5 with a sensitivity of 80%, and a specificity of 91%

The semiquantitative PET-2 response assessment showed a sensitivity of 80%, specificity of 92%, PPV of 66.7%, NPV of 95.8%. and accuracy of 90%. The association between the semiquantitative response assessment (ΔΣSUVmax 0–2) and outcome at the end of first-line chemotherapy resulted to be statistically significant (Fisher exact test, $P = 0.0026$). (Table 3)

In Figures 4 and 5 are represented 2 exemplar cases of discordant results between visual and semiquantitative assessment.

Table 3

Diagnostic values of interim ^{18}F-FDG PET/CT (PET-2) considering visual and semiquantitative assessments respectively.

	Assessement approach of PET-2	
	Visual (DS)	Semiquantitative (ΔΣ SUVmax 0–2)
TP (n)	3	4
TN (n)	18	23
FP (n)	7	2
FN (n)	2	1
Sensitivity (%)	60	80
Specificity (%)	72	92
PPV (%)	30	66.7
NPV (%)	90	95.8
Accuracy (%)	70	90
Fisher exact test	$P = 0.300$	$P = 0.0026$

The semiquantitative analysis has higher sensitivity, specificity, PPV, NPV, and accuracy than visual assessment. Moreover, Fisher exact test results demonstrate that there is a statistically significant association between ΔΣSUVmax0–2 evaluation and outcome at the end of first-line chemotherapy. ^{18}F-FDG PET/CT = fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, ΔΣSUVmax 0–2 = decrease of the sum of SUV max from PET-0 to PET-2, DS = Deauville score, FN = false-negative, FP = false-positive, NPV = negative predictive value, PET-2 = PET ad interim, SUVmax = maximum standardized uptake value, TN = true-negative, TP = true-positive.

3.4. Influence of PET-2 results on patient outcomes according to visual and semiquantitative analysis

Patients with a visually positive PET-2 showed a lower PFS than PET-2-negative patients (Log Rank = 6.483, $P = 0.011$) (Fig. 6). The semiquantitative assessment allowed a more accurate identification of patients with a high risk of treatment failure: patients who did not reach a ΔΣSUVmax 0–2 >12.5 had a significantly lower PFS than those who did (log rank = 7.948, $P = 0.005$) (Fig. 7).

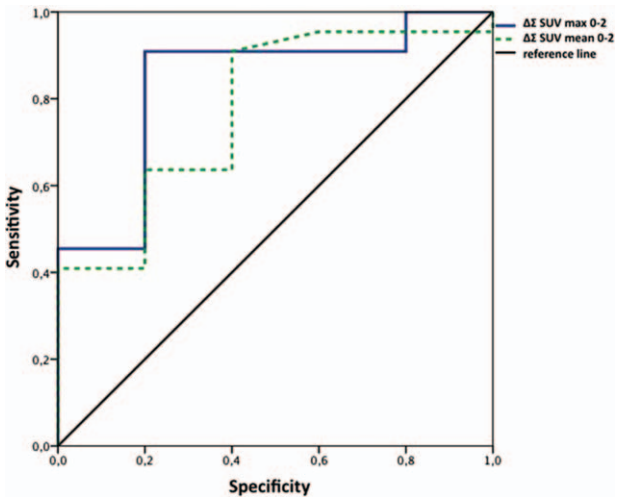

Figure 3. ROC curves of the semiquantitative PET-2 parameters. The highest AUC resulted for ΔΣSUVmax 0–2 (0.836; cut-off <12,5, sensitivity 80%, specificity 91%) that proved to be suitable to separate good from poor responder patients at the end of first-line of treatment. ΔΣSUVmax 0–2 = decrease of the sum of SUVmax from PET-0 to PET-2, ΔΣSUVmean 0–2 = decrease of the sum of SUVmean from PET-0 to PET-2, AUC = area under curve, ROC = receiver-operating-characteristics.

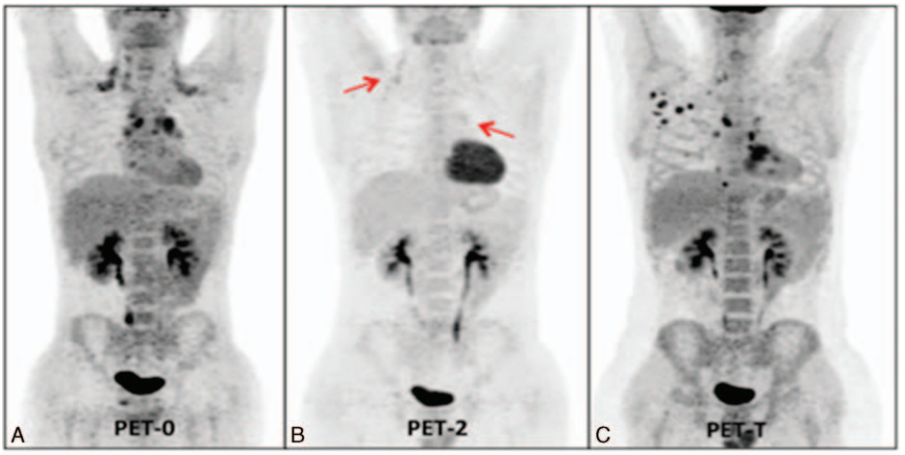

Figure 4. (A–C) ^{18}F-FDG PET/CT performed in a 16-year-old male affected by Mixed Cellularity Hodgkin lymphoma at stage IIB. (A) PET-0 (MIP whole-body image): intense ^{18}F-FDG uptake in the upper diaphragmatic lymph nodal basins, considered involved by disease; (B) PET-2 (MIP whole-body image): residual involvement in the neck and mediastinal lymph nodes (red arrows) that showed a mild ^{18}F-FDG uptake, greater than mediastinum but not greater than liver. Visual analysis considered the patient negative (DS = 3), whereas the semiquantitative analysis considered the patient as a poor responder (ΔΣSUVmax 0–2 = 9,70); (C) PET-T (MIP whole-body image): persistent and new sites of upper diaphragmatic lymph node involvement. Patient was defined nonresponder at the end of the first-line chemotherapy; he died at 40 months' follow-up despite of additional treatment approach. ΔΣSUVmax 0–2 = decrease of the sum of SUVmax from PET-0 to PET-2, ^{18}F-FDG PET/CT = fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, DS = Deauville score, MIP = maximum-intensity-projection, PET-0 = PET baseline, PET-2 = PET ad interim, PET-T = at the end of first-line chemotherapy.

4. Discussion

The aim of ongoing therapy optimization protocols in pHL is to focus on reduction of therapy-related toxicity while maintaining high survival rates. The use of cytotoxic therapy causes toxic effects including myelosuppression, neuropathy, pulmonary fibrosis, and cardiac damage; later effects also include risks of myelodysplasia and leukemia, particularly in patients treated with alkylating agents. Additionally, radiation therapy can cause mucositis and xerostomia and significantly increases the secondary cancer risk. Mortality, resulting from secondary cancers and heart disease treatment-related, that arises after 15 to 20 years of follow-up is the main problem to solve, especially in the pediatric population that has a longer life expectancy when compared to adults.[6]

In this respect, the main goal of interim PET in pHL would be to separate accurately nonresponders, who fail first-line therapy and needs a second-line, more aggressive one as soon as possible, from responders who are at a lower risk for relapse and can be successfully treated with reduced amounts of chemotherapeutic agents and radiation therapy.[2,3]

It has been previously demonstrated that ^{18}F-FDG PET/CT performed early during treatment has a high prognostic value in adults with HL.[13] Gallamini et al[12] demonstrated the prognostic value of interim PET after 2 cycles of ABVD in a

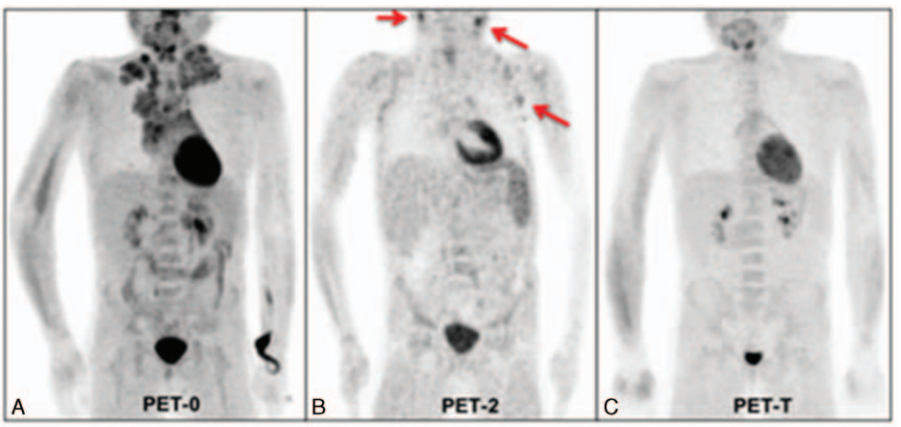

Figure 5. (A–C) ^{18}F-FDG PET/CT performed in a 9-year-old male affected by nodular sclerosis Hodgkin lymphoma at stage IIIB. (A) PET-0 (MIP whole-body image): intense ^{18}F-FDG uptake in the upper diaphragmatic lymph nodal basins, considered involved by disease; (B) PET-2 (MIP whole-body image): residual involvement in the neck and axillary lymph nodes (red arrows) that showed moderately increased of ^{18}F-FDG uptake compared to the liver. Visual analysis considered the patient as positive (DS = 4), whereas the semiquantitative analysis considered the patient as a good responder (ΔΣSUVmax 0–2 = 31,00); (C) PET-T (MIP whole-body image): negative. Patient was considered responder at the end of first-line chemotherapy and still responder at 10 months' follow-up. ^{18}F-FDG PET/CT = fluorine-18 fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography, ΔΣSUVmax 0–2 = decrease of the sum of SUVmax from PET-0 to PET-2, DS = Deauville score, MIP = maximum-intensity-projection, PET-0 = PET baseline; PET-2 = PET ad interim, PET-T = at the end of first-line chemotherapy .

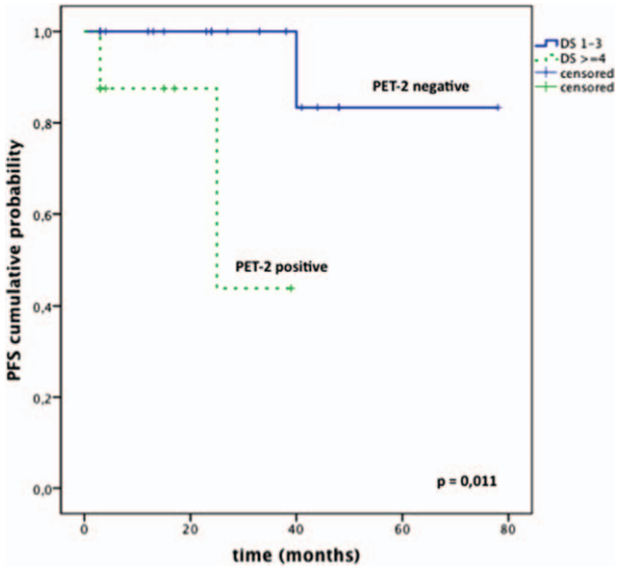

Figure 6. Kaplan–Meier plot showing PFS for PET-2 results according to the visual analysis (Deauville 5-point-scale). Patients with a visually positive PET-2 showed a lower PFS than PET-2-negative patients (Log rank = 6.483, P = 0.011). DS = Deauville score, PET-2 = PET ad interim, PFS = progression-free survival.

cohort of 260 patients with 2-year PFS of 13% for patients with positive interim PET and 95% for patients with negative scans. More recently, Biggi et al[8] in a retrospective analysis of an international cohort of patients with advanced HL using scores 1, 2, and 3 to define CR after 2 ABVD cycles reported a NPV of 94% and PPV of 73% for 3-year PFS.

The few studies performed in children with HL demonstrated that using a merely visual criterion to read out interim PET almost provides a high NPV with 94.4% associated with a poor and variable PPV ranging between 11% and 75%.[3,14,22,23] The

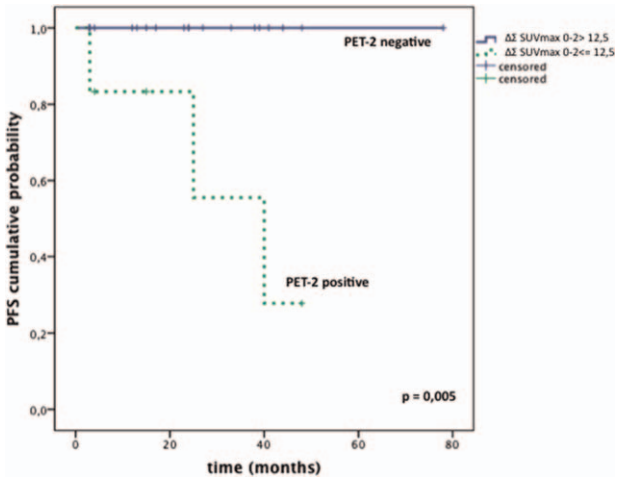

Figure 7. Kaplan–Meier plot showing PFS for PET-2 results according to the semiquantitative assessment (ΔΣSUVmax 0–2). Patients who did not reach a ΔΣSUVmax 0–2 >12.5 had a significantly lower PFS than those who did (Log rank = 7.948, P = 0.005). The semiquantitative assessment allowed a more accurate identification of patients with a high risk of treatment failure than visual analysis. ΔΣSUVmax 0–2 = decrease of the sum of SUVmax from PET-0 to PET-2, PET-2 = PET ad interim, PFS = progression-free survival.

prospective study by Furth et al[3] reported that the visual assessment of interim ^{18}F-FDG PET/CT had a very low PPV of 14% but excellent NPV of 100% in a pediatric patient population with HD, in which 38 of 40 children (95%) reached complete remission.

According to the literature data, our results concerning the PET-2 visual assessment showed a strong NPV (90%) but an impaired PPV (30%). The poor PPV was determinated by an excess of false-positive results produced by the visual analysis; more than half of patients considered poor responders on visual analysis resulted FP compared to the outcome at the end of first-line chemotherapy.

As assessed also by other authors, FP results related to the visual PET-2 evaluation could proceed from numerous causes.[14,17] Unlike other tumor types, HL is characterized by a heterogeneous cellular infiltrate and only a small fraction of the tumor mass is composed of malignant cells with the rest being reactive infiltrative cells.[24] Since antineoplastic therapy affects mainly the tumor cells, the residual ^{18}F-FDG uptake that can be observed at interim PET may be probably related more to the microenvironment surrounding the tumor cells (inflammatory cells) than to the tumor cells themselves.

Moreover, the ^{18}F-FDG uptake used as the background reference, that is the liver uptake, can be influenced by different factors (therapy, benign condition, blood glucose level) and it can be different also in the same patients between the baseline study from interim PET. Visual interpretation can be also influenced by an interobserver variability especially for those lesions small in size or slightly enhanced and/or by the brightness of the background that can vary depending on the physiological uptake pattern. Recently, Kluge et al[25] analyzed 100 interim PET/CT scans from pediatric HL with the aim to assess the inter-reader reliability of the visual analysis by using Deauville scale. They showed that inter-reader reliability of the complete 5-point DS is poor, in particular in cases assigned for DS 2 or 3. The authors confirmed that the distinction in DS 1, 2, 3 versus 4, 5 is the most reliable criterion for clinical decision-making.

In the present study, we used a third expertise nuclear physician opinion for those few cases of discrepancies; only 3 of 30 cases were discordant and actually they were DS 2 or 3.

To overcome these limitations, there has been a growing interest in literature in considering the semiquantitative approach that supports visual analysis, for the correct interpretation of interim PET.[1,2,17,24–26] Kluge et al[25] concluded that the use of a semiautomatic algorithm for comparison of the residual uptake intensity with a reference levels might help to avoid inter-reader discordances, making the interim PET results more reliable.

By using the cut-off we found in our sample, ΔΣSUVmax 0–2 analysis performed significantly better than visual assessment with a significantly better PPV for ΔΣSUVmax 0–2 analysis (66.7%) than for visual analysis (30%) (Table 3). This result leads to better specificity and accuracy for the semiquantitative method. Our results showed that among the 10 of 30 patients (33.3%) with a PET-2 visually positive, just 3 patients had a ΔΣSUVmax 0–2 <12.5. Thus, 7 patients could be reclassified as good responders according to the semiquantitative analysis. All these patients confirmed to have a favorable outcome at the end of first-line chemotherapy and follow-up. It is important to note that all these patients presented a residual mass on PET-2 with relatively low ^{18}F-FDG uptake (median SUVmax of 4.67; range, 3.40–8.20) that remained superior to the liver uptake. In these

patients, ΔΣSUVmax 0–2 analysis can help in distinguishing which positive results may be related to a significant residual lymphoma, improving the PET-2 positive predictive value.

Inversely, the NPV was quite similar with the 2 interpretation criteria: 2 of the 20 patients (10%) with a visually negative PET-2 experienced treatment failure (either progressive disease or a relapse), leading to a NPV of 90% for visual analysis, whereas 1 of the 24 patients (4.2%) who achieved a ΔΣSUVmax 0–2 >12.5 failed treatment, leading to a NPV of 95.8%.

In Figures 4 and 5 are represented 2 exemplar cases of discordant results between visual and semi-quantitative assessment.

Although to a lesser extent, semiquantitative assessment can also generate false-positive results. This occurred in 2 patients in which ΣSUVmax at baseline was low, leading to a small ΔΣSUVmax 0–2, lower than the defined cutoff value. In both 2 cases, the PET-2 visual assessment was negative (DS = 1 and 3). The same occurred in the study of Rossi et al,[17] even if the study was focused on adult patients affected by HL. The authors suggested that, in patients whose tumor exhibits a baseline SUVmax <10 and for whom SUVmax reduction after 2 cycles of chemotherapy does not reach the defined cut-off, the use of visual analysis can be recommended. In our study group, PET-2 semiquantitative analysis using ΔΣSUVmax 0–2 with a cutoff value of 12.5 resulted in a distinct improvement in discriminating good from poor responder patients and predicting the outcome at the end of the first-line chemotherapy, compared to mere visual response assessment. However, given the large range of SUVmax reductions and the relatively small number of patients in our study, we know that this cutoff value is not recommended to be used dogmatically.

Based on the final clinical response assessment, 5 of 30 (16.7%) patients were considered NR at the end of the first-line chemotherapy. Four of five patients would be correctly identified by integrating information from both visual and semiquantitative methods. The last one was considered FN by visual assessment but TP by semiquantitative analysis.

As reported also by Rossi et al,[17] our results confirm that both methods demonstrated to influence PFS, even if the semiquantitative assessment allowed a more accurate identification of patients with a high risk of treatment failure: a low ΔΣSUVmax 0–2 value is related significantly to a lower PFS ($P = 0.005$)

Finally, even if we did not found no association between visual assessment according to DS and outcome at the end of the first-line chemotherapy, probably influenced by the small size of our sample, according to Hussien et al,[26] we consider the qualitative reading of the images still a valuable cornerstone for easily judging PET-2 scans as negative and/or for defining those anatomical regions where to place VOIs for semiquantitative assessment in case of PET-2 positivity. Our results aim to emphasize the integrated role of visual and semiquantitative analysis to achieve the best performance of interim PET.

5. Conclusion

This single-center retrospective study, performed in a homogeneous series of 30 pediatric patients with HL, confirms the high NPV and low PPV of the visual assessment of interim ^{18}F-FDG PET/CT, by using Deauville criteria. Our results showed that semiquantitative analysis of PET-2, by using response using SUVmax reduction (ΔΣSUVmax 0–2) is more accurate than visual analysis and suggest that its use in addition to visual analysis helps to reduce the relatively high number of false-positive and the interpretation of the intermediate scores. Moreover, it can improve the prognostic value of interim ^{18}F-FDG PET/CT, specifically for predicting with good confidence those patients who will have a poor outcome requiring alternative therapies.

Larger series are warranted to confirm these preliminary results.

References

[1] Qiu L, Chen Y, Wu J. The role of 18F-FDG PET and 18F-FDG PET/CT in the evaluation of pediatric Hodgkin's lymphoma and non-Hodgkin's lymphoma. Hell J Nucl Med 2013;16:230–6.

[2] Sharma P, Gupta A, Patel C, et al. Pediatric lymphoma: metabolic tumor burden as a quantitative index for treatment response evaluation. Ann Nucl Med 2012;26:58–66.

[3] Furth C, Steffen IG, Amthauer H, et al. Early and late therapy response assessment with [18F]fluorodeoxyglucose positron emission tomography in pediatric Hodgkin's lymphoma: analysis of a prospective multicenter trial. J Clin Oncol 2009;27:4385–91.

[4] Aleman BM, van den Belt-Dusebout AW, Klokman WJ, et al. Long-term cause specific mortality of patients treated for Hodgkin's disease. J Clin Oncol 2003;21:3431–9.

[5] Ng AK, Bernardo MP, Weller E, et al. Long-term survival and competing causes of death in patients with early-stage Hodgkin's disease treated at age 50 or younger. J Clin Oncol 2002;20:2101–8.

[6] Cerci JJ, Pracchia LF, Linardi CC, et al. 18F-FDG PET after 2 cycles of ABVD predicts event-free survival in early and advanced Hodgkin lymphoma. J Nucl Med 2010;51:1337–43.

[7] Zinzani PL, Rigacci L, Stefoni V, et al. Early interim 18F-FDG PET in Hodgkin's lymphoma: Evaluation on 304 patients. Eur J Nucl Med Mol Imaging 2012;39:4–12.

[8] Biggi A, Gallamini A, Chauvie S, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: Interpretation criteria and concordance rate among reviewers. J Nucl Med 2013;54:683–90.

[9] Maggialetti N, Ferrari C, Minoia C, et al. Role of WB-MR/DWIBS compared to 18F-FDG PET/CT in the therapy response assessment of lymphoma. Radiol Med 2016;121:132–43.

[10] Barnes JA, LaCasce AS, Zukotynski K, et al. End-of-treatment but not interim PET scan predicts outcome in nonbulky limited-stage Hodgkin's lymphoma. Ann Oncol 2011;22:910–5.

[11] Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. J Clin Oncol 2014;32:3048–58.

[12] Gallamini A, Hutchings M, Rigacci L, et al. Early interim 2-[18F]fluoro-2-deoxy-Dglucose 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;25:3746–52.

[13] Hutchings M, Loft A, Hansen M, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression free survival in Hodgkin lymphoma. Blood 2006;107:52–9.

[14] Ilivitzki A, Radan L, Ben-Arush M, et al. Early interim FDG PET/CT prediction of treatment response and prognosis in pediatric Hodgkin disease-added value of low-dose CT. Pediatr Radiol 2013;43:86–92.

[15] Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET scan in lymphoma. Leuk Lymphoma 2009;17:1–4.

[16] Furth C, Amthauer H, Hautzel H, et al. Evaluation of interim PET response criteria in paediatric Hodgkin's lymphoma–results for dedicated assessment criteria in a blinded dual-centre read. Ann Oncol 2011;22:1198–203.

[17] Rossi C, Kanoun S, Berriolo-Riedinger A, et al. Interim 18F-FDG PET SUVmax reduction is superior to visual analysis in predicting outcome early in Hodgkin lymphoma patients. J Nucl Med 2014; 55:569–73.

[18] Campo E, Swerdlow SH, Harris NL, et al. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. Blood 2011;117:5019–32.

[19] Lister TA, Crowther D, Sutcliffe SB, 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.
[20] Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. J ClinOncol 2007;25:579–86.
[21] Meignan M, Gallamini A, Meignan M, et al. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. Leuk Lymphoma 2009;50:1257–60.
[22] Meany HJ, Gidvani VK, Minniti CP. Utility of PET scans to predict disease relapse in pediatric patients with Hodgkin lymphoma. Pediatr Blood Cancer 2007;48:399–402.
[23] Levine JM, Weiner M, Kelly KM. Routine use of PET scans after completion of therapy in pediatric Hodgkin disease results in a high false positive rate. J Pediatr Hematol Oncol 2006;28: 711–4.
[24] Canellos GP. Residual mass in lymphoma may not be residual disease. J Clin Oncol 1988;6:931–3.
[25] Kluge R, Chavdarova L, Hoffmann M, et al. Inter-reader reliability of early FDG-PET/CT response assessment using the Deauville scale after 2 cycles of intensive chemotherapy (OEPA) in Hodgkin's lymphoma. PLoS One 2016;11:e0149072.
[26] Hussien AE, Furth C, Schönberger S, et al. FDG-PET response prediction in pediatric Hodgkin's lymphoma: impact of metabolically defined tumor volumes and individualized SUV measurements on the positive predictive value. Cancers (Basel) 2015;7: 287–304.