

# Diagnostic accuracy of FDG PET/CT for clinical evaluation at the end of treatment of HL and NHL: a comparison of the Deauville Criteria (DC) and the International Harmonization Project Criteria (IHPC)

Federico Fallanca<sup>1</sup> · Pierpaolo Alongi<sup>1,2</sup> · Elena Incerti<sup>1</sup> · Luigi Gianolli<sup>1</sup> · Maria Picchio<sup>1</sup> · Irfan Kayani<sup>3</sup> · Jamshed Bomanji<sup>3,4</sup>

Received: 17 November 2015 / Accepted: 31 March 2016 / Published online: 7 May 2016  
© Springer-Verlag Berlin Heidelberg 2016

## Abstract

**Purpose** To evaluate the accuracy and prognostic value of FDG PET/CT for response assessment after treatment in patients with Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL) when using the Deauville Criteria (DC) and the International Harmonization Project Criteria (IHPC).

**Methods** This retrospective study included 101 patients (35 HL, 66 NHL) who underwent early restaging FDG PET/CT after treatment. Scans were evaluated using the IHPC and DC. Two thresholds of positivity for the DC were used: a score of at least 3 (DC3, i.e. scores 3–5) and a score of at least 4 (DC4, i.e. a score of 4 or 5). Accuracy was assessed using conventional diagnostic procedures, multidisciplinary team case notes, further PET/CT scans and/or follow-up. Progression-free survival and overall survival were computed using the Kaplan-Meier method. The Cox proportional hazards model was used to identify predictors of outcome.

**Results** Sensitivity, specificity, positive predictive value, negative predictive value and accuracy of FDG PET/CT for early restaging were, respectively, 92 %, 87 %, 74 %, 92 % and 86 % using DC4, 97 %, 76 %, 64 %, 98 % and 84 % using DC3, and 97 %, 67 %, 57 %, 98 % and 76 % using the IHPC. FDG PET/CT positivity was associated with a worse

cumulative survival rate over a 2-year period when using DC4 in comparison with the IHPC (20 % vs. 49 %;  $p < 0.05$ ) and DC3 (47 %;  $p < 0.05$ ). Cox regression analysis showed different risks of progression in patients positive on FDG PET/CT using the IHPC, DC3 and DC4 (hazard ratios 1.57, 0.7 and 3.2, respectively).

**Conclusion** FDG PET/CT using DC4 showed higher diagnostic accuracy for HL and NHL than FDG PET/CT using either the IHPC or DC3, indicating its value in predicting clinical outcome after treatment.

**Keywords** Deauville Criteria · International Harmonization Project Criteria · Hodgkin lymphoma · Non-Hodgkin lymphoma · <sup>18</sup>F-FDG PET/CT · End of therapy · Restaging

## Introduction

Lymphoma consists of over 50 histologically and biologically distinct lymphoid malignancies, classified into Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL). Each subtype has a different prognosis and requires a specific treatment approach, depending mainly on the stage of the disease. Remarkable advances have been made in diagnosis and treatment, with an improvement in 5-year survival rate documented in Europa, particularly for NHL [1, 2]. The introduction of FDG PET/CT as a staging and response assessment tool in FDG-avid lymphoma has contributed to better clinical management in lymphoma patients and it has been included among the standard criteria of evaluation [3–5]. When used for remission assessment in HL and aggressive NHL, PET/CT shows a high predictive value; in indolent lymphomas, too, PET/CT appears to be a good predictor of outcome, especially in high tumour burden follicular lymphomas (FL) [6–8].

✉ Jamshed Bomanji  
jamshed.bomanji@nhs.net

<sup>1</sup> Unit of Nuclear Medicine, IRCCS San Raffaele Scientific Institute, Milan, Italy

<sup>2</sup> Nuclear Medicine Unit, Department of Radiological Sciences, San Raffaele G. Giglio Institute, Cefalù, Italy

<sup>3</sup> University College London Hospital, London, UK

<sup>4</sup> Department of Nuclear Medicine, University College Hospitals NHS Foundation Trust, 235 Euston Rd, London NW1 2BU, UK

At the end of treatment, residual masses are frequently detected by CT scan, but PET/CT is able to differentiate between viable pathological cells and fibrotic scar tissue with greater accuracy [9]. It is known that FDG uptake revealed by PET may reflect several phenomena and does not unequivocally indicate the presence or absence of malignancy. In particular, inflammatory changes related to treatment or to residual disease are not infrequent at the end of treatment and may be associated with higher FDG uptake. Nevertheless, the likelihood of malignancy increases with increasing FDG uptake. The main challenge is to identify an accurate threshold for discriminating between these two conditions.

In 2007, the Imaging Subcommittee of the International Harmonization Project (IHP) in Lymphoma proposed a standardization of performance and interpretation of PET examinations [10]. According to the IHP criteria (IHPC), response to therapy is assessed on the basis of both PET tracer uptake and size of residual lesions measurable on CT images. Although the IHPC have been validated only in small cohorts of HL and NHL patients and the interobserver variability has not been assessed [11, 12], these criteria have been strongly recommended for use in clinical practice and research trials [13]. Alternative criteria, known as the Deauville Criteria (DC), have also been proposed for evaluation of PET scans based on the application of a five-point scale using the mediastinum and liver activity as the reference standard. The prognostic value of PET using the DC after one to three cycles of chemotherapy (interim PET) in patients with HL and aggressive NHL and the potential benefits for clinical management have been investigated [14–16]. Furthermore, assessment of the DC in multicentre studies has revealed good diagnostic accuracy and intercentre concordance [17, 18]. More recently, the five-point scale has been recommended for reporting both interim PET and post-therapy PET (ptPET) [4, 5]. Nevertheless, the true effectiveness of the IHPC and DC remains unclear and there is also uncertainty over whether PET/CT performed at the end of therapy for HL and NHL is of prognostic value when using these criteria.

Therefore, the aims of this study were:

- To evaluate the accuracy of PET/CT performed at the end of therapy for HL and NHL when adopting the standardized criteria (IHPC and DC) in defining status of disease
- To assess which of the two models (IHPC or DC) for the interpretation of ptPET provides better prediction of the risk of progression and survival.

## Materials and methods

This retrospective analysis included consecutive patients with lymphoma obtained from the central PET databases of the

departments of nuclear medicine of University College London Hospital (UCLH) and San Raffaele Scientific Institute of Milan (OSR) for the period April 2008 to October 2013.

The inclusion criteria was as follows:

1. Diagnosis of lymphoma confirmed by histology (fine-needle or open biopsy) available at the respective centres
2. Availability of conventional imaging and non-imaging data for staging disease
3. FDG PET/CT scan performed at staging, with evidence of FDG-avid disease
4. FDG PET/CT scan performed at the end of conventional treatment, at least 3 weeks after the end of chemotherapy or 6 weeks following radiotherapy, in accordance with the IHPC
5. Available clinical information in the form of multidisciplinary team case notes and/or imaging information for validation of patient PET findings
6. Posttherapy surveillance, until death or for at least 12 months after the PET/CT scan to evaluate the risk of progression and the survival rate
7. Written informed consent for the FDG PET/CT scan and the anonymous publication of disease-related information

Patients with CNS lymphoma, patients with concomitant malignancy and retroviral positivity, and patients on other research protocols were excluded from the analysis. Finally, 101 patients (60 from UCLH; 41 from OSR) were evaluated: 55 men and 46 women with a median age of 46 years (mean 47 years, range 6–87 years). Of the 101 patients, 35 had HL and 66 NHL. All were treated according to the each institution's standard protocol.

Disease status was followed by reference to the available clinical and imaging information, identifying patients negative for the presence of disease and patients with relapse and/or progression of disease (PD) at follow-up (median duration of follow-up in the overall cohort 18 months; 95 % confidence interval, CI, 15–20 months). From these data we also calculated time of PD, progression-free survival (PFS) and overall survival (OS). The study was performed in compliance with the principles of the Declaration of Helsinki. Patient characteristics are summarized in Table 1.

## FDG PET/CT imaging

Both centres involved in the study use PET/CT tomographs to perform clinical studies and both comply with international guidelines concerning patient preparation, radiopharmaceutical administration and dose, radiation protection, imaging acquisition and quality control of scanners.

Studies were acquired using a Discovery™ PET/CT tomograph (GE Healthcare, Milwaukee, WI) or a Gemini

**Table 1** Characteristics of the patient population

Histotype	No. of patients	Age (years)			Sex (M/F)	Stage			
		Median	Mean	Range		I	II	III	IV
<b>Hodgkin lymphoma</b>									
Classical	30	29	35.5	16–80	18/12	–	7	13	10
Nodular lymphocyte-predominant	5	16	14.0	6–20	4/1	–	3	2	–
Total	35	28	33.6	6–8	22/13	–	10	15	10
<b>Non-Hodgkin lymphoma</b>									
Diffuse large B-cell lymphoma	37	60	59.8	30–87	19/18	1	5	11	20
Follicular lymphoma	16	55.5	56.0	30–76	6/10	1	2	10	3
Mucosa-associated lymphoid tissue lymphoma	2	–	40.0	34–46	1/1	–	–	1	1
Marginal zone lymphoma	2	–	54.4	43–66	1/1	–	–	–	2
Mantle cell lymphoma	2	–	53.5	40–67	2/0	–	1	1	–
Lymphoplasmacytic lymphoma	2	–	44.5	13–72	1/1	–	–	2	–
T-lymphoblastic lymphoma	2	–	40.0	32–48	1/1	–	1	1	–
Anaplastic large T-cell lymphoma	1	–	–	21	1/0	–	–	–	1
Anaplastic ALK+ lymphoma	1	–	–	43	1/0	–	–	–	1
Burkitt lymphoma	1	–	–	27	0/1	–	–	–	1
Total	66	55.5	51.8	13–87	33/33	2	9	26	29
All	101	46	47	6–87	55/46	2	19	41	39

GXL16 PET/CT tomograph (Philips Medical Systems, Veenpluis Best, The Netherlands). PET images were acquired for each bed position from the base of the skull to the mid-thigh and CT (90–120 mA, 140 kV, 0.8 s per tube rotation) was used for nonuniform attenuation correction. FDG PET/CT images were interpreted by two experienced nuclear medicine physicians who had access to the results of previous imaging and clinical information, but were unaware of follow-up results at the time of the ptPET evaluation in order to avoid any significant influence on image evaluation. The whole-body FDG PET/CT scans in each patient were evaluated qualitatively using either the IHPC as described by Juweid et al. [10] or the DC (five-point scale, see below) [16].

### FDG PET/CT analysis

Any previously involved nodal or extranodal site identified on pretherapy staging was evaluated. According to the IHPC, ptPET was considered positive when a lymph node lesion greater than 2 cm on CT images showed residual uptake more intense than the mediastinal blood pool activity; if the lesion was smaller than 2 cm, the surrounding background activity was considered as the reference [19]. Additional rules were considered for lung nodules, residual uptake in the bone marrow and hepatic and splenic parenchyma, in accordance with the established IHPC [10]. Using the DC, the following five-point scale was applied: 1 no residual uptake, 2 residual uptake less than or equal to that in the mediastinum, 3 residual uptake greater than that in the mediastinum but less than that in the

liver, 4 residual uptake moderately higher than that in the liver, and 5 residual uptake markedly higher than that in the liver or the presence of new sites of uptake. If a new site of increased uptake was observed, all potential causes other than PD, such as inflammation/infection, were considered; for example, diffusely increased bone marrow or spleen uptake, even if more intense than the liver uptake, was often due to a postchemotherapy reaction [18]. Two different thresholds of positivity for the DC were used: a score of at least 3 (DC3, i.e. scores 3–5) and a score of at least 4 (DC4, i.e. a score of 4 or 5).

The ptPET results were considered true-positive (TP) or false-positive (FP) and true-negative (TN) or false-negative (FN) according to the corresponding disease state at the time of the PET examination, as recorded in the clinician's or multidisciplinary team's case notes.

In order to estimate the risk of progression and the survival rate, calculating PFS and OS, available posttherapy surveillance data were considered. Assessment of disease state comprised routine conventional diagnostic procedures such as ultrasonography, contrast-enhanced CT, MRI, and biopsy of suspicious residual disease (when performed). Also cases judged doubtful were included in the analysis if the suspected disease state was confirmed by a short-term controlled follow-up. In the accuracy analysis, PET/CT findings were classified as positive or negative using each of the three models of evaluation mentioned above (DC3, DC4 and IHPC). For each model, negative FDG PET/CT results were classified as TN when assessment of disease state confirmed negative results or findings due to inflammatory changes. Positive FDG PET/CT

findings were considered TP when assessment of disease state indicated the presence of PD. FDG PET/CT findings were considered FP or FN when the reference yielded discordant evidence.

### Statistical analysis

The D'Agostino normality test was used to assess the normality of the distributions of the variables under study. The Wilcoxon-Mann-Whitney test (nonparametric) and Student's *t* test (parametric) were used as appropriate. To evaluate the usefulness of FDG PET/CT in restaging, we calculated sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), accuracy and likelihood ratios (*LH+* likelihood ratio for positive results, *LH-* likelihood ratio for negative results) using standard methods [20–22]. ROC curve analysis was performed to confirm sensitivity and specificity and to calculate the area under the curve (AUC). Simple linear regression was used to evaluate the reproducibility of the IHPC and DC. OS was defined as the time from the PET assessment performed at the end of therapy to the date of death or last clinical/imaging information. PFS was obtained by calculating the time from the PET assessment performed at the end of therapy to the appearance of clinical or radiological progression. PFS and OS were calculated using the Kaplan-Meier method. Multivariate Cox proportional hazards models were fitted in all samples to determine whether the IHPC, DC3 and DC4 were significantly associated with survival. A Cox regression model was built using a step-wise selection procedure, with the *p* value set to 0.05 for a feature to be entered into or left out of the model. The relationships between outcome and the included variables are summarized in terms of hazard ratios (HR) and 95 % CI. All statistical analyses were performed using IBM SPSS Statistics software, version 20.0 (IBM Corp., Armonk, NY).

### Results

Of the 101 patients studied, 31 showed therapy failure at the end of treatment (group 1). Of these 31 patients, 21 had experienced further PD at the time of study observation. Despite salvage therapy, 3 of the 21 died of their disease, and 13 were still ill (6 with PD, 6 with stable disease, and 1 with partial remission) at the end of the time of observation; the remaining 5 achieved complete remission and were free of disease until the end of the study. The mean PFS in group 1 was 9.7 months (95 % CI 6.6–13 months), with a median of 11 months (95 % CI 3–12 months).

The remaining 70 of the 101 patients were judged free of disease after the end of treatment (group 2). Of these 70 patients, 9 experienced relapse in the first year after the end of treatment (mean 5 months, median 5 months, range 3–

8 months), and 5 of these 9 experienced further subsequent PD, including 1 with lymphoblastic T lymphoma, who died. The other 4 patients achieved complete remission after salvage therapy. One of the 70 patients, who was cured of lymphoma, died of a haemorrhagic stroke (Fig. 1). The mean PFS in group 2 was 20 months (95 % CI 17.5–23.3 months) with a median of 15 months (95 % CI 14–18 months). The mean OS in group 1 was 23.1 months (95 % CI 19.6–27 months) with a median of 18 months (95 % CI 16–24 months), and mean OS in group 2 was 23 months (95 % CI 20–26 months) with a median of 17 months (95 % CI 15–21 months).

The performance of ptPET was evaluated according to the corresponding disease state at the time of the PET examination.

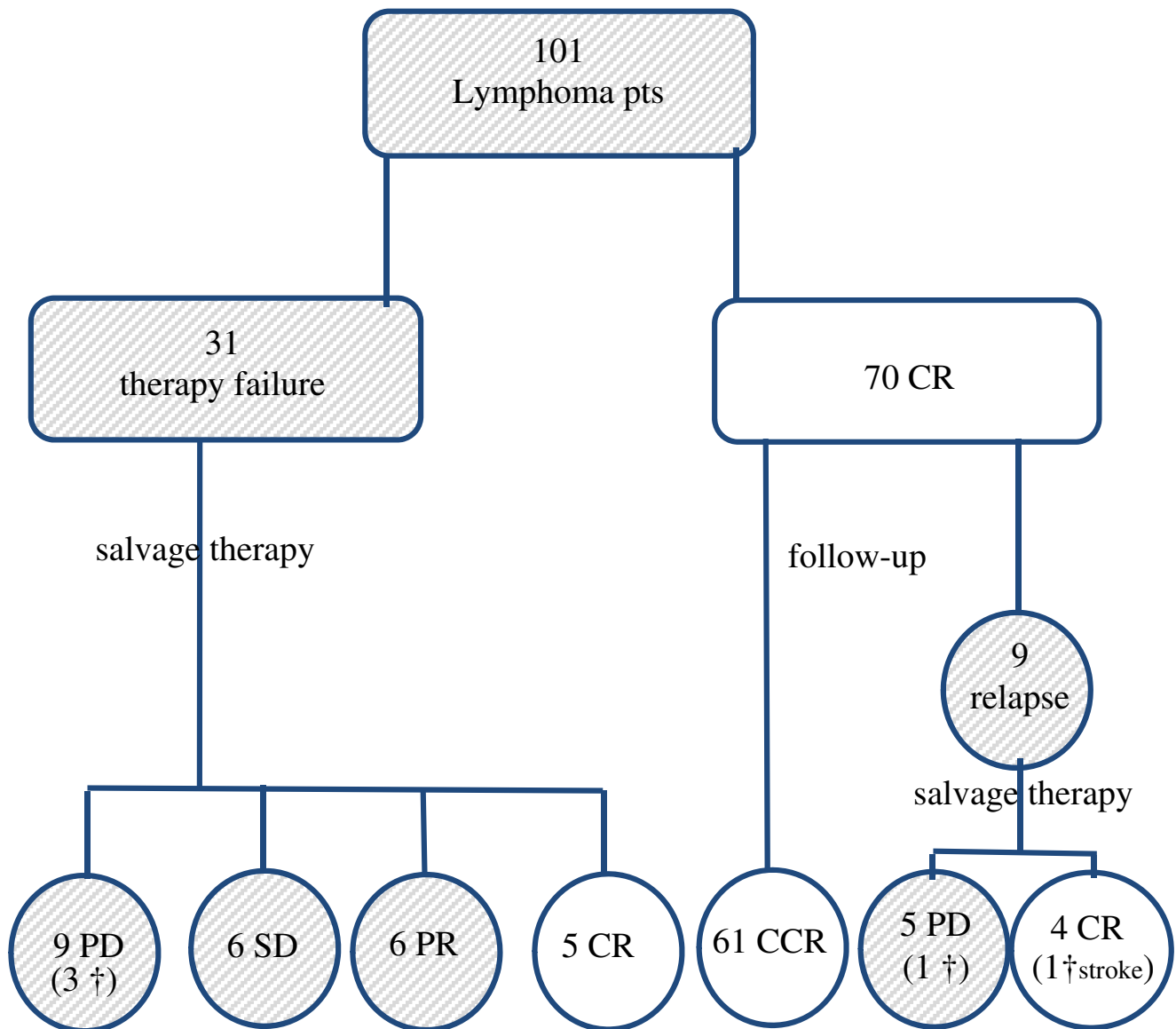
### IHPC evaluation

In group 1, 30 of the 31 patients were positive on PET (TP) using the IHPC. In one patient who received chemotherapy for stage III scleronodular HL (cHL), residual disease was suspected on the basis of the clinical examination 1 month after the end of chemotherapy, although ptPET was negative using both the IHPC and DC. Two months later, a further PET/CT scan and a contrast-enhanced CT scan confirmed lymph node disease in the right laterocervical chain, as assessed by previous clinical evaluations. In this patient the ptPET result was judged FN. In group 2, 47 of the 70 patients were negative on ptPET (TN). The sensitivity, specificity, PPV, NPV and accuracy of ptPET were 97 %, 67 %, 57 %, 98 % and 76 %, respectively (Table 2). In the likelihood ratio analysis, *LH+* was 2.9 and *LH-* was 0.04. The AUC was 0.82 (95 % CI 0.73–0.90).

In the Kaplan-Meier analysis, the 2-year cumulative PFS and OS rates in patients positive on FDG PET/CT were 49 % and 53 %, respectively. In contrast, the 2-year cumulative PFS and OS rates in patients negative on FDG PET/CT were 75 % and 74 %, respectively (Fig. 2). The differences in PFS and OS rates between patients positive and negative on FDG PET/CT were significant (*p* < 0.05). The multivariate Cox regression analysis for the risk of PD in patients positive on FDG PET/CT using the IHPC showed HR of 1.57 (CI 95 % 0.37–6.4, *p* = 0.03; Table 3).

### DC evaluation

In group 1, 30 of the 31 patients were positive on PET using DC3, and 26 of the 31 using DC4 (TP). In group 2, 53 of the 70 patients were negative on ptPET using DC3 and 61 of the 70 using DC4. The sensitivity, specificity, PPV, NPV and accuracy of ptPET were 97 %, 76 %, 64 %, 98 % and 84 %, respectively, using DC3, and 92 %, 87 %, 74 %, 92 % and 86 %, respectively, using DC4 (Table 2). In the likelihood ratio analysis, *LH+* was 4.04 for DC3 and 7.07 for DC4,



**Fig. 1** Outcome in patients (CR complete remission, PD progression of disease, SD stable disease, PR partial remission, CCR continued complete remission, dagger symbol died)

and LH- was 0.03 for DC3 and 0.09 for DC4. The AUC was 0.86 for DC3 (95 % CI 0.78–0.93) and 0.90 for DC4 (95 % CI 0.83–0.97).

In the Kaplan-Meier analysis, the 2-year cumulative PFS rates in patients positive on FDG PET/CT using

DC3 and DC4 were 47 % and 20 %, respectively. In contrast, the 2-year cumulative PFS rates in patients negative on FDG PET/CT using DC3 and DC4 were 74 % and 73 %, respectively (Figs. 3 and 4). The differences in PFS rates observed between patients positive and negative on

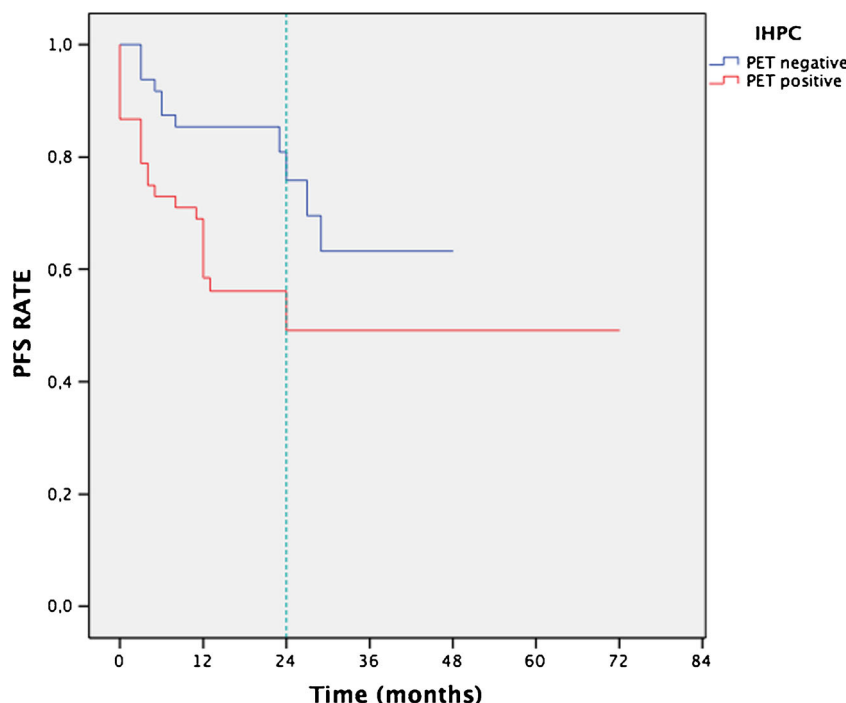
**Table 2** Diagnostic performance of FDG PET/CT performed at the end of treatment using different evaluation criteria

Criteria	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
IHPC	97	67	57	98	76
DC3 <sup>a</sup>	97	76	64	98	84
DC4 <sup>b</sup>	92	87	74	92	86

<sup>a</sup> Deauville Criteria considering scores 3–5 as positive

<sup>b</sup> Deauville Criteria considering scores 4 and 5 as positive

**Fig. 2** Kaplan-Meier analysis of progression-free survival in relation to the PET/CT results using the IHPC



FDG PET/CT were significant ( $p < 0.05$ ) using both DC3 and DC4.

The 2-year cumulative OS rates in patients positive on FDG PET/CT were 48 % and 36 % using DC3 and DC4, respectively, and in patients negative on FDG PET/CT were 74 % and 75 % using DC3 and DC4, respectively ( $p < 0.05$  for both DC3 and DC4). The multivariate Cox regression analysis for the risk of PD in patients positive on FDG PET/CT using DC3 showed HR 0.7 (CI 95 % 0.1–3.7,  $p = 0.07$ ), and using DC4 showed HR 3.2 (CI 95 % 0.9–9.1,  $p = 0.006$ ); Table 3). Simple linear regression between DC4 and the IHPC scores (highest values) showed a partial but not complete correlation ( $R^2 = 0.5$ ,  $f$  ratio 74;  $p < 0.0001$ ).

**Discussion**

The diagnostic role of PET assessment at the end of treatment in patients with lymphoma has been widely addressed in the literature, with PET being considered of particular value in patients with radiological evidence of residual disease [4]. Determination of residual disease is important in both HL and aggressive NHL, where the objective of therapy is the complete eradication of disease. In the case of FL and other NHL subtypes characterized by lower clinical aggressiveness, guidelines do not recommend the routine use of FDG PET/CT for response assessment at the end of therapy. However, there is a consensus among clinicians that wider clinical use of FDG PET/CT is

**Table 3** Risk of disease progression and 2-year cumulative PFS and OS rates in patients undergoing FDG PET/CT at the end of treatment, using the IHPC, DC3 and DC4

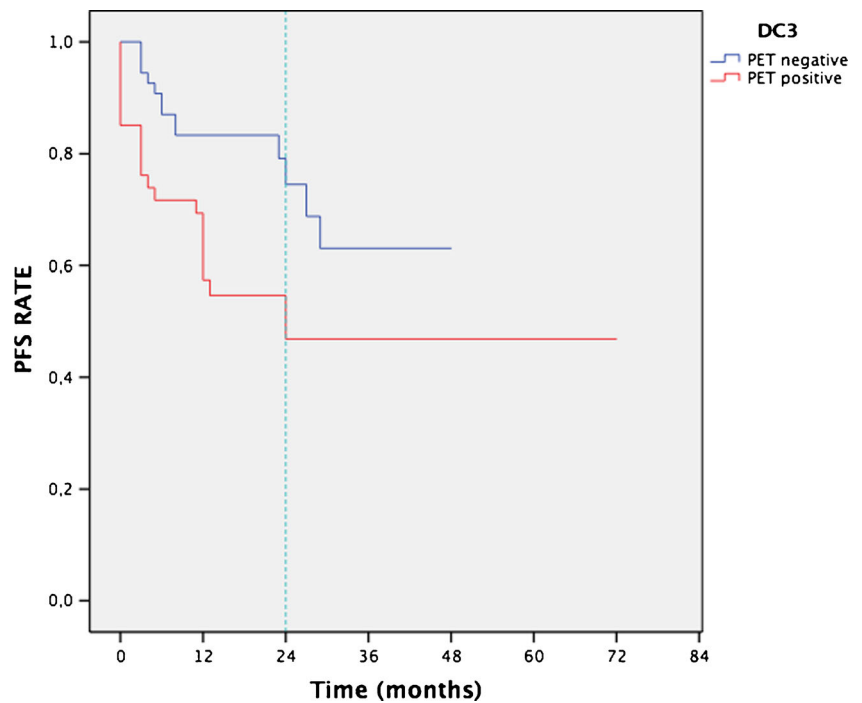
Variable		Cox regression analysis		Kaplan-Meier analysis	
		HR (95 % CI)	<i>p</i> value	2-year PFS rate (%) <sup>*</sup>	2-year OS rate (%) <sup>*</sup>
IHPC	Positive	1.57 (0.37–6.4)	0.03	49	53
	Negative	–	–	75	74
DC3 <sup>a</sup>	Positive	0.7 (0.1–3.7)	0.07	47	48
	Negative	–	–	74	74
DC4 <sup>b</sup>	Positive	3.2(0.9–9.1)	0.006	20	36
	Negative	–	–	73	75

<sup>\*</sup> $p < 0.05$  related to positive vs. negative values for each criteria

<sup>a</sup> Deauville Criteria considering scores 3–5 as positive

<sup>b</sup> Deauville Criteria considering scores 4 and 5 as positive

**Fig. 3** Kaplan-Meier analysis of progression-free survival in relation to the PET/CT results using DC3

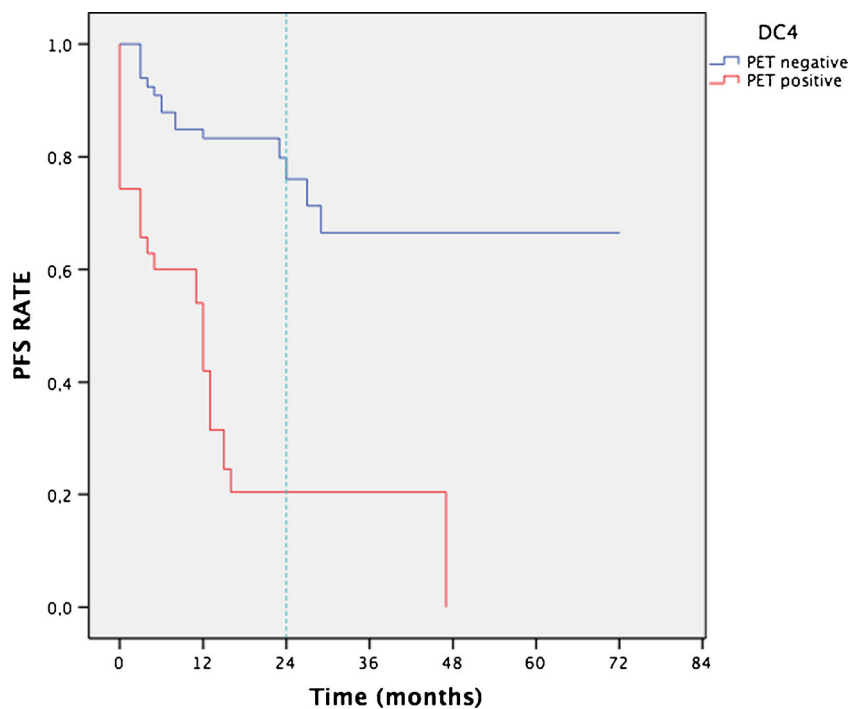


appropriate at least in FL, because emerging evidence indicates that metabolic evaluation at the end of treatment permits better prognostic stratification of patients [7]. In order to obtain a scenario very similar to everyday reality, consecutive patients treated for HL and NHL were included in our series, which included patients with FL and other NHL histotypes with the proviso that evidence of FDG-

avid disease was present at staging and either chemotherapy or radiotherapy was implemented.

Among our patient cohort, significant variability in the diagnostic performance of FDG PET/CT was observed depending on the interpretation criteria adopted (i.e. IHPC, DC3 or DC4). The IHPC and DC3 gave a very high sensitivity of 97 %, and their specificities were 67 % and

**Fig. 4** Kaplan-Meier analysis of progression-free survival in relation to the PET/CT results using DC4



76 %, respectively. Due to high FP rates, limited PPVs (57 % for the IHPC and 64 % for DC3) were observed; on the other hand, NPV was 98 % for both criteria. Using DC4, sensitivity was lower (92 %), specificity was higher (87 %), PPV was higher (74 %), and NPV was adequate (92 %). Previous FDG PET/CT studies in HL patients have shown NPVs of 95–100 % and PPVs of more than 95 % [23, 24]. In recent studies, mostly involving interim PET in aggressive NHL, NPV for ptPET was found to range between 80 % and 100 %, while PPV was lower and more variable from 50 % to 100 % [25–28].

As indicated above, using the different FDG PET/CT evaluation criteria, we obtained a high NPV (92–98 %) and a more variable PPV (57–74 %). Taking into account the fact that about two-thirds of patients enrolled in our study were affected by NHL, this variability in PPV and NPV appears similar to the range of results from evaluation of the cumulative data from previous studies. This observation could support the hypothesis that the principal cause of the variation in results obtained in different studies was the use of different criteria for image interpretation, even if heterogeneity with respect to instrumentation and acquisition protocols may also explain some of the variability [29].

We found that DC, and in particular DC4, provided better diagnostic accuracy than the IHPC. Although the IHPC were based mainly on the authors' expertise, given the limited available literature, they represent a milestone on the road toward standardization in the use of PET for response assessment in lymphoma. A qualitative evaluation of residual uptake has been judged adequate for determining whether a PET scan is positive or negative. The feasibility of using semiquantitative data is in any case severely limited by variability with regard to scanner and image reconstruction, and such data do not appear essential for ptPET interpretation, even if some authors have hypothesized that evaluation of relative reductions in standardized uptake values (SUV) may be beneficial for response assessment [10, 30].

The diagnostic accuracy of the DC for interim PET reporting has previously been tested in other retrospective studies (e.g. Gallamini et al. [15]) and the use of the DC is currently under evaluation in prospective studies [31]. Based on the positive experiences with the use of the DC in reporting interim PET, it has been proposed that these criteria might also be employed for the evaluation of ptPET [4, 32], particularly in clinical trials. It is to be noted in this context that discordant interpretations among reviewers can occur in more challenging cases, e.g. in discriminating physiological from pathological uptake in scans with prominent brown fat uptake, in distinguishing the healing process from residual disease in patients with pathological fractures, in differentiating misregistered physiological gut uptake from liver activity,

and in interpreting uptake when different arm positioning has been used in baseline and interim studies [18]. In our analysis, instances of discordant scoring among the observers using different criteria were resolved by consensus. Analysis of concordance among the observers was not performed because good agreement in the application of the DC has already been demonstrated in previous studies [17, 18].

While the use of DC4 yielded the best accuracy in our study, the rate of FN results was markedly higher than when using DC3 and the IHPC (18 % vs. 3 %). One patient who experienced recurrence after therapy for stage III cHL was negative with the use of all three criteria. The available literature data indicate that at the end of treatment Deauville scores of 1 and 2 reliably represent a complete metabolic response, while scores of 4 and 5 indicate residual disease. Experts suggest a cautious approach in evaluating a score of 3, and our data support this recommendation, even if current evidence shows that the majority of patients with this score have a good prognosis. Generally, such a score warrants prompt further clinical investigations to ensure that there is no delay in initiating salvage therapy, should it be required. Especially in response-adapted trials exploring treatment de-escalation, a score of 3 should be considered to represent an inadequate response to therapy, to avoid under-treatment [4]. These considerations also suggest that the threshold selected for PET positivity may vary depending on the clinical context, the histological subtype and the therapies used.

As shown by Chiaravalloti et al., in contrast to mediastinal blood pool activity, the intensity of liver uptake can vary during chemotherapy. However, while statistically significant differences in liver SUV were found between PET at staging and interim PET, no relevant changes in liver SUV were found between PET at staging and ptPET. Therefore, changes in hepatic metabolism should not affect ptPET scoring in any way [33].

To date, few reported studies have focused on the prognostic value of PET when performed at the end of therapy in patients with HL or NHL using current standard of care therapy outside clinical trials, and the results of these studies have been heterogeneous. One study, by Barnes et al. [34], showed a lack of value of interim PET in predicting progression of HL or NHL, whereas scans carried out at completion of therapy were highly predictive of PFS and OS. Similarly, Trotman et al. demonstrated that FDG PET/CT status at the end of immunochemotherapy induction in patients with FL is strongly predictive of outcome [35]. Nevertheless, there have been no previous comparisons of the use of the DC and IHPC for PET assessment at the end of therapy in a patient population mimicking a standard clinical scenario.

Our survival results obtained using Cox regression and Kaplan-Meier analysis revealed significant differences

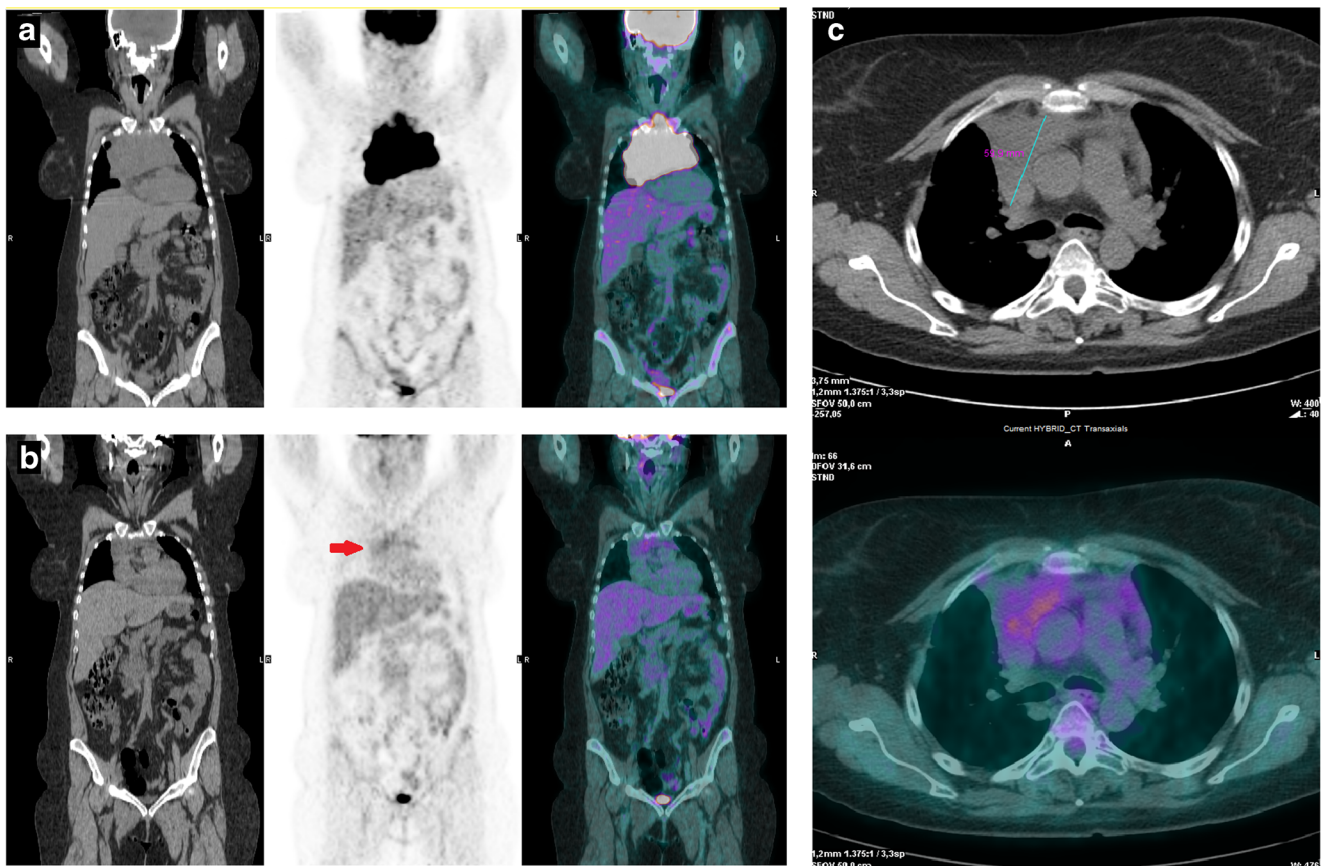


when using the IHPC, DC3 and DC4. A positive PET/CT using DC4 was strongly associated with a higher risk of PD and increased mortality, and this model of evaluation proved to be a better predictor than DC3 or the IHPC. Thus, evaluation of PFS and OS at 2 years of follow-up showed a clear association between a positive examination and worse outcome, which was most pronounced when using DC4: PFS, OS and HR were 20 %, 36 % and 3.2 using DC4; 47 %, 48 % and 0.7 using DC3; and 49 %, 53 % and 1.57 using the IHPC, respectively; Figs. 2, 3 and 4). Conversely, a negative scans predicted favourable outcomes, but with similar results among all three models (Table 3).

It is expected that ongoing prospective studies on PET-tailored therapy will yield further data that will assist in defining the best strategy for the use of the DC in particular situations. In future, it would also be useful to evaluate whether the introduction of a semiquantitative approach, such as SUV analysis comparing a baseline and ptPET/CT studies, could improve accuracy. Another issue warranting study is

whether the integration of ptPET/CT results with other prognostic indices, such as the Follicular Lymphoma International Prognostic Index (FLIPI) or lactate dehydrogenase levels, might improve prediction of outcome in lymphoma patients.

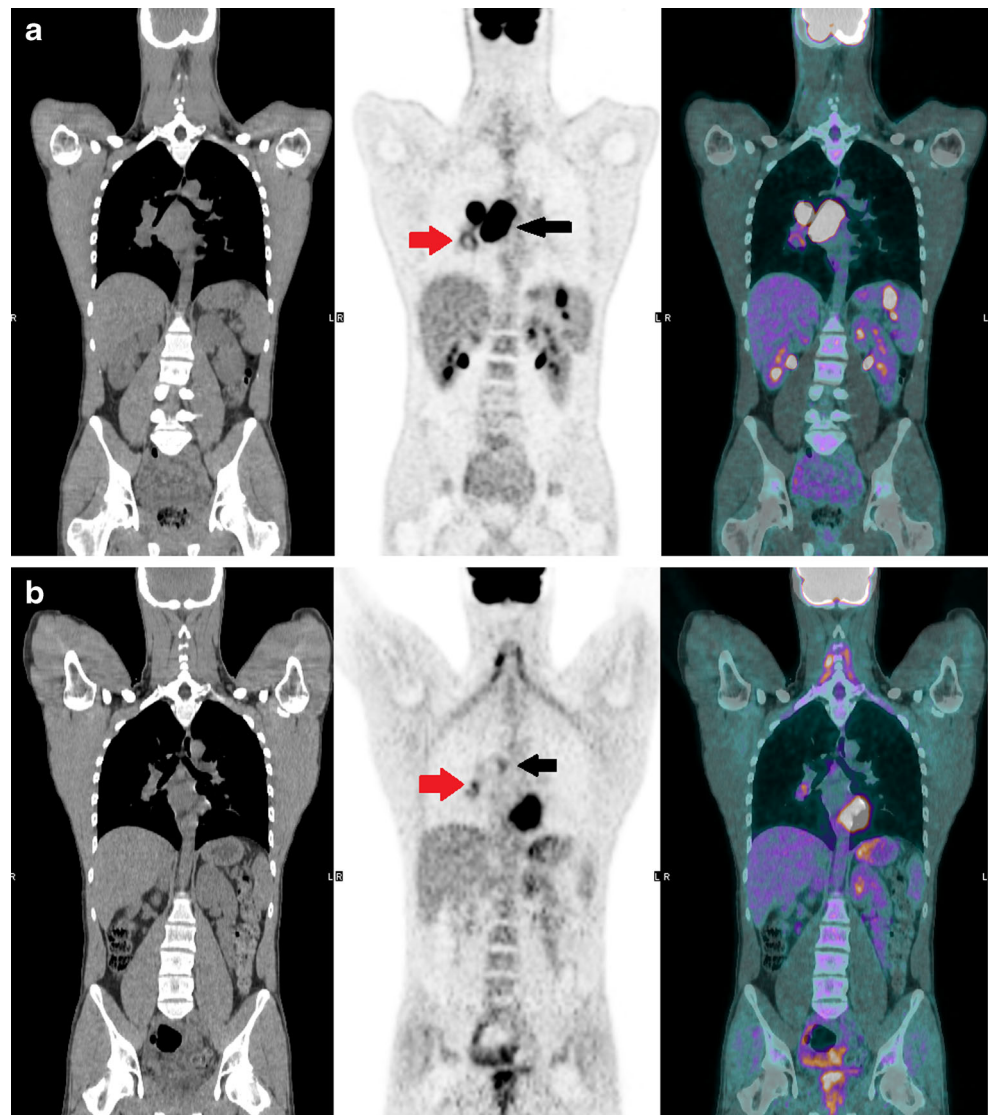
The main limitation of this study was its retrospective nature. Consecutive patients treated for different histological subtypes of lymphoma were included in our series. Different therapeutic approaches were used depending on the patients' characteristics, with a consequent impact on overall outcome. In addition, the instrumentation and acquisition protocols used varied between the centres, and in some cases also over the observation period, preventing reliable semi-quantitative analysis. On the other hand, these very heterogeneous characteristics simulated a scenario very similar to everyday reality and permitted a comparison among the different PET interpretation criteria. The use of the IHPC and DC systems in clinical reality is illustrated in Figs. 5 and 6 that show PET/CT images and their interpretation in two example patients.



**Fig. 5** A 39-year-old female patient with stage II diffuse large B-cell lymphoma. **a** Coronal PET/CT images at staging: there is intense tracer uptake in the upper mediastinum corresponding to a bulky mass. **b** Corresponding coronal ptPET images: at the end of chemotherapy, residual uptake equal to liver uptake is apparent (red arrow). This result

corresponds to a Deauville score of 3 (DC3). **c** Axial ptPET images: residual mediastinal tissue greater than 2 cm on CT images shows residual uptake more intense than the mediastinal blood pool activity. This result corresponds to a positive IHPC score. This patient showed a complete response and was still free of disease after a follow-up of 4 years

**Fig. 6** A 28-year-old female patient with stage IV Burkitt lymphoma. **a** Coronal PET/CT images at staging: there is evidence of disease at the right pulmonary hilar (*red arrow*) and mediastinal lymph nodes, with involvement of the subcarinal region (*black arrow*). **b** Corresponding coronal ptPET images: at the end of chemotherapy, residual uptake slightly greater than liver uptake is apparent in the subcarinal (*black arrow*) and right pulmonary hilar (*red arrow*) lymph nodes. This result corresponds to a Deauville score of 4 (DC4) and a positive IHPC score. This patient showed a partial response and reached a complete response after salvage therapy 2 months later



Another important limitation of this study was that the disease state was histologically confirmed only when this was required by the clinical management. Such an investigation to confirm disease recurrence was therefore available in only four patients; in most patients the presence of disease was confirmed by instrumental investigations (41 of the 101 patients) or on the basis of follow-up (101 patients). The results of our study in a heterogeneous population simulating a routine clinical situation indicate that the DC are simple to apply and certainly more effective than the IHPC. Nevertheless, the experience so far derived using the DC in research protocols cannot be immediately translated into clinical ptPET/CT evaluation, although these criteria may be a valuable tool capable of assisting in image interpretation. Future large prospective studies in homogeneous populations

are required to define the best strategy for DC use in particular clinical situations.

## Conclusion

This retrospective study indicates that, in lymphoma patients, the value of FDG PET/CT for response assessment after completion of treatment is influenced by the interpretation criteria adopted. Differences in the risk of PD and survival rate were shown in patients with positive PET using the IHPC, DC3 and DC4. The best performance and survival prediction results were obtained using DC4, but further studies are needed to definitively confirm the reliability of this approach in the clinical setting.

**Compliance with ethical standards****Funding** None.**Conflicts of interest** None.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

**Informed consent** Informed consent was obtained from all individual participants included in the study

**References**

- De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999–2007 by country and age: results of EURO CARE-5 – a population-based study. *Lancet Oncol.* 2014;15:23–34.
- Sant M, Minicozzi P, Mounier M, Anderson LA, Brenner H, Holleczeck B, et al. Survival for haematological malignancies in Europe between 1997 and 2008 by region and age: results of EURO CARE-5, a population-based study. *Lancet Oncol.* 2014;15:931–42.
- Johnson SA, Kumar A, Matasar MJ, Schöder H, Rademaker J. Imaging for staging and response assessment in lymphoma. *Radiology.* 2015;276:323–38.
- Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol.* 2014;32:3059–68.
- Cheson BD. Staging and response assessment in lymphomas: the new Lugano classification. *Chin Clin Oncol.* 2015;4:5.
- Pyo J, Won Kim K, Jacene HA, Sakellis CG, Brown JR, Van den Abbeele AD. End-therapy positron emission tomography for treatment response assessment in follicular lymphoma: a systematic review and meta-analysis. *Clin Cancer Res.* 2013;19:6566–77.
- Bodet-Milin C, Eugène T, Gastinne T, Frampas E, Le Gouill S, Kraeber-Bodéré F. FDG-PET in follicular lymphoma management. *J Oncol.* 2012;2012:370272.
- Luminari S, Biasoli I, Versari A, Rattotti S, Bottelli C, Rusconi C, et al. The prognostic role of post-induction FDG-PET in patients with follicular lymphoma: a subset analysis from the FOLL05 trial of the Fondazione Italiana Linfomi (FIL). *Ann Oncol.* 2014;25:442–7.
- Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. *J Nucl Med.* 2009;50 Suppl 1:21S–30.
- 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;25:571–8.
- Brepeols L, Stroobants S, De Wever W, Spaepen K, Vandenberghe P, Thomas J, et al. Aggressive and indolent non-Hodgkin's lymphoma: response assessment by integrated international workshop criteria. *Leuk Lymphoma.* 2007;48:1522–30.
- Brepeols L, Stroobants S, De Wever W, Spaepen K, Vandenberghe P, Thomas J, et al. Hodgkin lymphoma: response assessment by revised international workshop criteria. *Leuk Lymphoma.* 2007;48:1539–47.
- Cheson BD. New response criteria for lymphomas in clinical trials. *Ann Oncol.* 2008;19:35–8.
- Gallamini A, Kostakoglu L. Interim FDG-PET in Hodgkin lymphoma: a compass for a safe navigation in clinical trials? *Blood.* 2012;120:4913–20.
- Gallamini A, Patti C, Viviani S, Rossi A, Fiore F, Di Raimondo F, et al. Early chemotherapy intensification with BEACOPP in advanced-stage Hodgkin lymphoma patients with a interim-PET positive after two ABVD courses. *Br J Haematol.* 2011;152:551–60.
- Meignan M, Gallamini A, Haioun C. Report on the First International Workshop on Interim-PET-Scan in Lymphoma. *Leuk Lymphoma.* 2009;50:1257–60.
- Barrington SF, Qian W, Somer EJ, Franceschetto A, Bagni B, Brun E, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging.* 2010;37:1824–33.
- Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, 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.
- Juweid ME, Wiseman GA, Vose JM, Ritchie JM, Menda Y, Wooldridge JE, et al. Response assessment of aggressive non-Hodgkin's lymphoma by integrated International Workshop Criteria and fluorine-18-fluorodeoxyglucose positron emission tomography. *J Clin Oncol.* 2005;23:4652–61.
- Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. *BMJ.* 1994;308:1552.
- Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ.* 1994;309:102.
- Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ.* 2004;329:168–9.
- Cerci JJ, Pracchia LF, Linardi CCG, Pitella FA, Delbeke D, Izaki M, 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.
- Engert A, Haverkamp H, Kobe C, Markova J, Renner C, Ho A, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet.* 2012;379:1791–9.
- Micallef IN, Maurer MJ, Wiseman GA, Nikcevic DA, Kurtin PJ, Cannon MW, et al. Epratuzumab with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone chemotherapy in patients with previously untreated diffuse large B-cell lymphoma. *Blood.* 2011;118:4053–61.
- Mikhaeel NG, Timothy AR, Hain SF, O'Doherty MJ. 18-FDG-PET for the assessment of residual masses on CT following treatment of lymphomas. *Ann Oncol.* 2000;11 Suppl 1:147–50.
- Cashen AF, Dehdashti F, Luo J, Homb A, Siegel BA, Bartlett NL. 18F-FDG PET/CT for early response assessment in diffuse large B-cell lymphoma: poor predictive value of International Harmonization Project interpretation. *J Nucl Med.* 2011;52:386–92.
- Pregno P, Chiappella A, Bellò M, Botto B, Ferrero S, Franceschetti S, et al. Interim 18-FDG-PET/CT failed to predict the outcome in diffuse large B-cell lymphoma patients treated at the diagnosis with rituximab-CHOP. *Blood.* 2012;119:2066–73.

29. Zijlstra JM, Lindauer-van der Werf G, Hoekstra OS, Riphagen HL, Huijgens PC. 18F-fluoro-deoxyglucose positron emission tomography for post-treatment evaluation of malignant lymphoma: a systematic review. *Haematologica*. 2006;91:522–9.
30. Strigari L, Attili A, Duggento A, Chiaravalloti A, Schillaci O, Guerrisi MG. Quantitative analysis of basal and interim PET/CT images for predicting tumor recurrence in patients with Hodgkin's lymphoma. *Nucl Med Commun*. 2016;37:16–22
31. Horning SJ, Juweid ME, Schöder H, Wiseman G, McMillan A, Swinnen LJ, et al. Interim positron emission tomography scans in diffuse large B-cell lymphoma: an independent expert nuclear medicine evaluation of the Eastern Cooperative Oncology Group E3404 study. *Blood*. 2010;115:775–7. quiz 918.
32. Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Müller SP, 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.
33. Chiaravalloti A, Danieli R, Abbatiello P, Di Pietro B, Travascio L, Cantonetti M, et al. Factors affecting inpatient liver and mediastinal blood pool 18F-FDG standardized uptake value changes during ABVD chemotherapy in Hodgkin's lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41:1123–32.
34. Barnes JA, LaCasce AS, Zukotynski K, Israel D, Feng Y, Neuberger D, 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.
35. Trotman J, Fournier M, Lamy T, Seymour JF, Sonet A, Janikova A, et al. Positron emission tomography-computed tomography (PET-CT) after induction therapy is highly predictive of patient outcome in follicular lymphoma: analysis of PET-CT in a subset of PRIMA trial participants. *J Clin Oncol*. 2011;29:3194–200.