

# FDG PET/CT Response Assessment Criteria for Patients with Hodgkin's and Non-Hodgkin's Lymphoma at End of Therapy: A Multiparametric Approach

Ur Metser<sup>1</sup> · Ravi Mohan<sup>1</sup> · Vaughan Beckley<sup>1</sup> · Hadas Moshonov<sup>2</sup> · David Hodgson<sup>3</sup> · Grainne Murphy<sup>1</sup>

Received: 15 July 2015 / Revised: 20 August 2015 / Accepted: 4 September 2015 / Published online: 29 September 2015  
© Korean Society of Nuclear Medicine 2015

## Abstract

**Purpose** Based on the International Harmonization Project (IHP) criteria, positron emission tomography (PET) response assessment of residual nodal masses in patients with lymphoma after completion of therapy is performed visually using mediastinal blood pool as the reference. The primary objective of this study was to define the optimal reference for PET response assessment. Secondary aim was to assess if morphological criteria on computed tomography (CT) may improve performance of PET.

**Methods** This institutional review board approved retrospective study included 137 patients, with Hodgkin's ( $n=43$ ) or non-Hodgkin's lymphoma ( $n=94$ ) assessed for residual masses ( $n=180$ ) after completion of therapy with pathology and clinical and imaging surveillance data (mean, 19 months) as the standard of reference. Two readers independently assessed response by IHP and Deauville criteria. The addition of morphological parameters on CT was assessed in relation to therapy response.

**Results** Based on the standard of reference, 36 patients (26.3 %) had residual lymphoma. For IHP and Deauville criteria, sensitivity, specificity and accuracy were 97.2 %, 97.2 % ( $p=1$ ); 79.2 %, 92.1 % ( $p<0.001$ ); and 83.9 %, 93.4 % ( $p=0.001$ ), respectively. Of the morphological parameters assessed, only change in size over course of therapy was significant ( $p<0.003$ ) and improved specificity for IHP-based interpretation to 90.4 % ( $p=0.008$ ).

**Conclusions** Using liver as the visual reference to determine PET positivity for lymphoma patients being assessed for residual masses at the end of therapy improves specificity, yet maintains the high sensitivity of PET in identifying residual disease. The addition of change in size after therapy improves specificity of PET when using IHP-based but not Deauville-based interpretation.

**Keywords** Fluorodeoxyglucose F18 · Positron emission tomography · Computed tomography · Lymphoma · Therapy · Response

✉ Ur Metser  
ur.metsers@uhn.ca

<sup>1</sup> Joint Department of Medical Imaging, University Health Network, Mount Sinai Hospital and Women's College Hospital, University of Toronto, Toronto, ON, Canada

<sup>2</sup> Office of Research and Development, Joint Department of Medical Imaging, University Health Network, Mount Sinai Hospital and Women's College Hospital, University of Toronto, Toronto, ON, Canada

<sup>3</sup> Department of Radiation Oncology, Princess Margaret Hospital, University Health Network, University of Toronto, Toronto, Ontario, Canada

## Introduction

Traditionally, assessment of response to therapy for lymphoma was based on the International Workshop Criteria (1999), which are based on computed tomography (CT) size measures of nodes, bone marrow biopsy and clinical and biochemical information [1]. <sup>18</sup>F-Fluorodeoxyglucose (FDG) positron emission tomography (PET) has emerged as a useful functional imaging tool for staging, restaging, and response assessment of lymphomas [2–19]. Its main advantage over morphological imaging modalities in assessing response to therapy is its ability to distinguish between viable tumor and

necrosis or fibrosis in residual masses often present after treatment.

In 2007, revised response criteria incorporating PET were developed by the Response Assessment Subcommittee of the International Harmonization Project (IHP) to ensure consistency across clinical trials. The definition proposed for a positive PET is focal or diffuse FDG uptake above the mediastinal blood pool (MBP) for residual nodal masses  $\geq 2$  cm, and any visible uptake above background for masses  $< 2$  cm, using visual assessment. Other specific criteria were provided for extranodal disease [3]. The IHP recommendations were based on a study assessing therapy response in 50 patients with Hodgkin's lymphoma (HL) or aggressive non-Hodgkin's lymphoma (NHL) [20], but little further data validating these criteria exist [21, 22].

Other response assessment measures, initially proposed for the interpretation of interim PET, include the Gallamini criteria (Gallamini-c), in which a semiquantitative uptake value (SUV) cutoff of 3.5 is used to discriminate adequate from inadequate response [23], and Deauville criteria (Deauville-c; Table 1) [24, 25]. However, few studies have validated these criteria as a surrogate for response at end of therapy. In one prospective trial on patients with high tumor burden follicular lymphoma being treated with immunochemotherapy, the Deauville-c were predictive of outcome [26]. A retrospective study on 69 patients with aggressive NHL reported higher accuracy for Deauville-c and Gallamini-c compared with IHP-c due to the low positive predictive value (PPV) of IHP-c [27]. Results from a further study including 66 patients with Hodgkin's lymphoma also suggest that using liver as reference may be a good approach for interpretation of post-therapy PET/CT [28]. However, the results of the latter studies need to be interpreted with caution due to the small number of patients who had residual disease. The primary objective of the current study is to examine the optimal IHP-c, Deauville-c and Gallamini-c in assessing residual masses at end of therapy for patients with HL and NHL. Secondary aims include: (1) assessment of interobserver agreement for IHP-c and Deauville-c; (2) comparison of quantitative and qualitative PET interpretation; (3) evaluation of combined morphological

and functional parameters for end of therapy response assessment.

## Patients and Methods

### Patient Population

This retrospective study was approved by our institution's Research Ethics Board and written informed consent was waived. Patient selection schema and demographic data are presented in Fig. 1. There were 43 patients with HL and 94 patients with NHL (diffuse large B-cell lymphoma,  $n=70$ ). There were 180 residual masses at end of therapy on CT, including 147 (81.7 %) nodal and 33 extranodal sites including bone ( $n=12$ , 6.7 %), lung ( $n=5$ , 2.8 %), spleen ( $n=3$ , 1.7 %), liver ( $n=1$ , 0.5 %), peritoneum ( $n=1$ , 0.5 %), or other ( $n=11$ , 6.1 %). Demographic data, including pathology, clinical status and time to disease recurrence were recorded from the electronic patient records.

### Standard of Reference

The standard of reference was based on histopathology ( $n=15$ ), or clinical follow-up (range, 2–67 months; mean,  $20.7 \pm 12.3$ ) and imaging surveillance (range, 1–48 months; mean,  $14.6 \pm 10.9$ ). A positive biopsy or interval increase in size of the mass on surveillance was considered indicative of residual viable disease at time of PET. Morphological stability over a 6-month period and lack of clinical evidence for disease recurrence was considered as evidence of lack of residual viable disease (mean disease free interval,  $14.8 \pm 13.2$  months; range, 0–51). Follow-up consisted of routine hematology clinic visits, at 1- to 6-month intervals, as clinically indicated. Imaging surveillance was routinely performed with CT. MR imaging (spine, bone or brain) was obtained in five patients for unexplained symptoms or indeterminate CT findings. According to the standard of reference, 36 patients (26.3 %) had residual disease, while 101 (73.7 %) were in remission at time of PET. Of those in remission at time of PET, 87 patients (86.1 %) remained disease-free at time of census (mean, 20.7 months after PET) and 15 (14.9 %) developed recurrence, either at site of residual mass ( $n=6$ ), or elsewhere ( $n=9$ ).

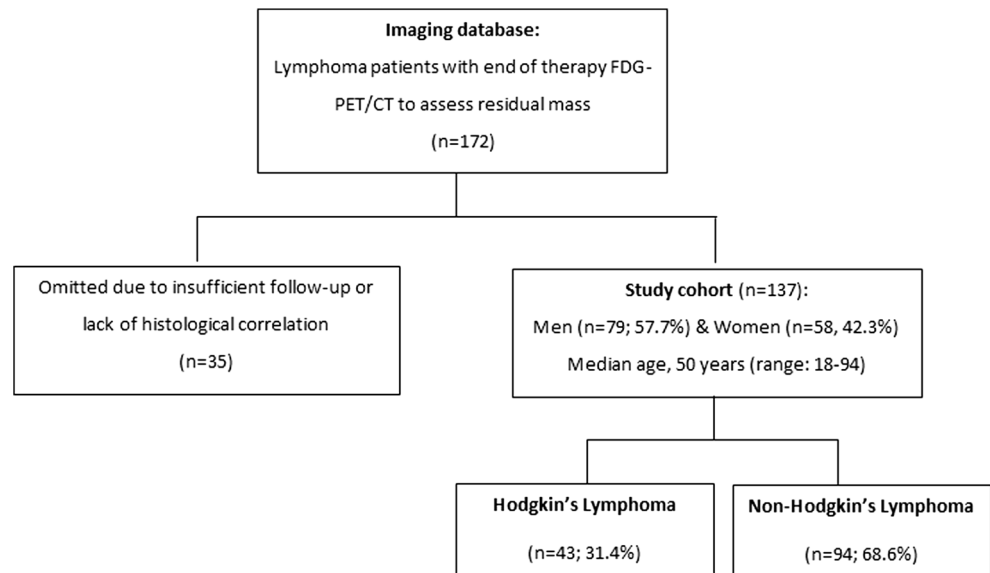
**Table 1** The Deauville interpretation criteria. The most intense uptake in a residual mass is scored on a five-point scale [24]

| Deauville criteria |  |
|--------------------|--|
| Score 1            | No uptake  |
| Score 2            | Uptake $\leq$ mediastinum                              |
| Score 3            | Uptake $>$ mediastinum but $\leq$ liver                |
| Score 4            | Uptake moderately $>$ liver                            |
| Score 5            | Uptake markedly $>$ liver and/or new lesions           |
| Score X            | New areas of uptake unlikely to be related to lymphoma |

### PET/CT Acquisition & Interpretation

Whole-body PET scans were performed in three-dimensional (3D) mode with a dedicated in-line PET/CT scanner (Biograph; Siemens, Oakville, ON, Canada). Patients were asked to fast for at least 6 h before undergoing the examination. Data were acquired 60–70 min after an intravenous injection of approximately 5 MBq/kg body weight of FDG (up

**Fig. 1** Patient selection schema and demographic data



to 550 MBq). First, a spiral CT scan from the skull base to the pelvis was obtained using the following parameters: 130-kV peak; 105 mAs; scan width, 5 mm; feed/rotation, 8.4 mm. No intravenous iodinated contrast was administered. On completion of CT, PET scans of the same area were acquired for 3 min/bed position, with 5–7 bed positions per patient. PET/CT was interpreted on a dedicated fused imaging workstation (Thinking Systems, Petersburg, FL). Interpretation was performed independently by two readers (U.M., R.M.; with 12 and 4 years of experience in interpreting PET), blinded to the standard of reference or outcome data. This was performed in two separate sessions using IHP-c and Deauville-c for each study at least 2 weeks apart to minimize recall bias.

For each patient,  $SUV_{max}$  and  $SUV_{mean}$  of residual or any positive lesion, of MBP and background liver were recorded, using a region of interest covering less than two-thirds of the lesion's diameter, to minimize partial volume effects. Of the 180 residual lesions, four lesions showed focal FDG uptake which could be explained by a benign process (such as thymic hyperplasia). These were scored as negative for residual disease and were excluded from quantitative analysis.

The performance of IHP-c, Deauville-c and Gallamini-c was evaluated on a lesion and patient level, with overall patient score corresponding to the highest score of any lesion. For this study, lesions with a score of  $\geq 4$  on by Deauville-c were considered positive for residual disease. Interobserver agreement was assessed for the IHP-c and Deauville-c. In order to determine performance of visual inspection to SUV measurements, we calculated quantitative IHP and Deauville scores for all lesions by comparing lesion SUV to SUV of the reference being evaluated. For quantitative IHP-c, lesions were classified with a score of 0 if lesion SUV  $\leq$  MBP SUV, and score of 1 if lesion SUV  $>$  MBP SUV. For quantitative

Deauville-c, a Deauville score of 1 was if no lesion was present; a score of 2, if lesion SUV  $\leq$  MBP SUV; 3, if  $>$  MBP SUV but  $\leq$  liver SUV; 4, if moderately (up to twice)  $>$  liver SUV; 5, if lesion was new or if lesion SUV was significantly (more than twice)  $>$  liver SUV.

#### Assessment of Morphological Parameters on Contrast-enhanced CT

Morphological parameters were assessed by two readers (G.M., V.B.) in consensus for all residual nodal masses that had contrast-enhanced CT available at baseline and end of therapy ( $n=138$  masses). These included: (1) size (long and short axis); (2) density obtained for masses assessed at baseline and post-therapy with portal venous phase acquisition ( $n=120$ ); (3) borders (well-defined vs infiltrative); (4) homogeneity (homogeneous, if attenuation mostly uniform vs heterogeneous). The percentage change in size and density were calculated. The association between morphological parameters and response to therapy was assessed. Finally, the predictive value of combining PET and morphological parameters was determined.

#### Statistical Analysis

Statistical analyses were performed using SPSS software version 20 (IBM SPSS, Chicago IL). Descriptive statistics are provided for demographic and clinical characteristics. Continuous variables were described using mean, standard deviation and range, and categorical variables using frequency and percentage. Diagnostics performance measures of visual PET assessment (IHP, Deauville and Gallamini) for the diagnostic of residual masses were calculated using the standard of reference described above. Raw performance estimates were

calculated at both the lesion and patient levels and are reported with exact binomial 95 % confidence intervals. Performance analyses were also evaluated for quantitative PET measurements using SUV values as well as for the combination of morphological and functional parameters. Logistic regression analysis was used to compare performance measures between these various methods. Generalized estimating equations were used to adjust for correlations of repeated measures. Overall percent of agreement and the Kappa coefficient were used to assess the interobserver agreement between readers. Classification based on visual inspection was compared to that based on quantitative measurements using McNemar and McNemar-Bowker tests. Association between quantitative morphological parameters and treatment response was examined via the two-sample *t*-test or the Mann–Whitney *U* test, whereas the Chi-square test and Fisher's exact test were used to assess association between qualitative parameters and treatment response. A *p* value <0.05 was considered statistically significant.

## Results

### Lesion-level Analysis

Of the 176 lesions which could be quantitatively evaluated,  $SUV_{max}$  was significantly higher for positive than negative lesions;  $9.73 \pm 6.72$  (median, 8.5) and  $1.94 \pm 0.99$  (median, 2), respectively;  $p < 0.001$ . Performance of IHP-c, Deauville-c and Gallamini-c in assessing residual masses at end of therapy is presented in Table 2. The distribution of Deauville scores was as follows: 102 (56.7 %), score=1 or 2; 19 (10.6 %), score=3; 59 (32.8 %), score=4 or 5. When classifying score 3 as positive, specificity decreased to 82.3 %, similar to that achieved using IHP-c. Comparing IHP-c and Deauville-c on a lesion level, sensitivity and NPV were similar ( $p=1$  for both); however, specificity, PPV and overall accuracy were significantly improved using Deauville-c ( $p$ -values <0.001, 0.006 and 0.001, respectively). Similarly, IHP-c and Gallamini-c achieved similar values of sensitivity and NPV ( $p$  values=0.289 and 0.345 respectively); however,

using Gallamini-c resulted in significantly higher specificity, PPV and overall accuracy ( $p$  value <0.001, = 0.01 and =0.023, respectively). No significant differences in sensitivity, specificity, PPV, NPV and overall accuracy were found when comparing Deauville-c and Gallamini-c ( $p=0.25, 1, 0.917, 0.375, 0.453$ , respectively).

### Patient-level Analysis

Performance of IHP-c, Deauville-c and in assessing patient disease status at end of therapy is presented in Table 3. When classifying score 3 lesions as positive, specificity decreased to 78.2 %, similar to that achieved using IHP-c. There was similar sensitivity and NPV for IHP-c and Deauville-c on patient level analysis; however, specificity and overall accuracy were significantly higher using the Deauville-c ( $p < 0.001$  for both).

### Interobserver Agreement

There was excellent agreement between the readers in assigning IHP scores for lesions (93.4 %, Kappa=0.854; SE=0.047), but only moderate agreement using the five-point Deauville scale (56.9 %, Kappa=0.433; SE=0.053). However, when combining Deauville scores 1 and 2 (definitely negative) and scores 4 and 5 (definitely positive) agreement was excellent (92.7 %, Kappa=0.858; SE=0.042). There were only two lesions, both of which were residual mediastinal masses that were interpreted as negative by one reader and positive by the other.

### Comparison of Visual Inspection and Quantitative Measurements

Based on  $SUV_{max}$ , there were a higher proportion of positive lesions than with visual assessment using IHP-c (108 [61.4 %] vs 76 [43.2 %] lesions;  $p < 0.001$ ). Similar findings were found when using  $SUV_{mean}$  (103 [58.5 %] vs 76 [43.2 %] positive lesions;  $p < 0.001$ ). Agreement between visual inspection and  $SUV_{max}$  based Deauville was achieved in 114/176 (64.8 %) lesions ( $p < 0.001$ ); with a higher rate of score 3 assigned using

**Table 2** Lesion-level analysis: comparing IHP-c, Deauville-c and Gallamini-c

| Method                   | Sensitivity (95 % CI)            | Specificity (95 % CI)              | PPV (95 % CI)                    | NPV (95 % CI)                      | Accuracy (95 % CI)                 |
|--------------------------|----------------------------------|------------------------------------|----------------------------------|------------------------------------|------------------------------------|
| IHP-c                    | 94.6 % (85.13, 98.88)<br>(53/56) | 81.5 % (73.50, 87.86)<br>(101/124) | 69.7 % (58.13, 79.75)<br>(53/76) | 97.1 % (91.80, 99.40)<br>(101/104) | 85.6 % (79.56, 90.34)<br>(154/180) |
| Deauville-c <sup>a</sup> | 92.9 % (82.71, 98.02)<br>(52/56) | 94.4 % (88.71, 97.70)<br>(117/124) | 88.1 % (77.07, 95.09)<br>(52/59) | 96.7 % (91.75, 99.09)<br>(117/121) | 93.9 % (89.33, 96.91)<br>(169/180) |
| Gallamini-c              | 87.5 % (75.93, 94.82)<br>(49/56) | 94.2 % (88.35, 97.62)<br>(113/120) | 87.5 % (75.93, 94.82)<br>(49/56) | 94.2 % (88.35, 97.62)<br>(113/120) | 92.1 % (87.01, 95.58)<br>(162/176) |

<sup>a</sup> On the Deauville five-point scale, scores of  $\geq 4$  were considered positive for lymphoma

**Table 3** Patient-level analysis: comparing IHP-c and Deauville-c

| Method                   | Sensitivity (95 % CI)             | Specificity (95 % CI)              | PPV (95 % CI)                     | NPV (95 % CI)                     | Accuracy (95 % CI)                  |
|--------------------------|-----------------------------------|------------------------------------|-----------------------------------|-----------------------------------|-------------------------------------|
| IHP-c                    | 97.22 % (85.47, 99.93)<br>(35/36) | 79.21 % (69.99, 86.64)<br>(80/101) | 62.50 % (48.55, 75.08)<br>(35/56) | 98.77 % (93.31, 99.97)<br>(80/81) | 83.94 % (76.70, 89.65)<br>(115/137) |
| Deauville-c <sup>a</sup> | 97.22 % (85.47, 99.93)<br>(35/36) | 92.08 % (84.99, 96.52)<br>(93/101) | 81.40 % (66.60, 91.61)<br>(35/46) | 98.94 % (94.21, 99.97)<br>(93/94) | 93.43 % (87.90, 96.95)<br>(128/137) |

<sup>a</sup> On the Deauville five-point scale, scores of  $\geq 4$  were considered positive for lymphoma

SUV<sub>max</sub> than visually (46 [26.1 %], and 18 [10.2 %], respectively). Agreement between visual inspection and SUV<sub>mean</sub> based Deauville was achieved in 117/176 (66.5 %) lesions ( $p < 0.001$ ); with a higher rate of score 3 assigned using SUV<sub>mean</sub> than visually (44 [25 %], and 18 [10.2 %], respectively). Comparison of the performance of visual and quantitative measures using SUV<sub>mean</sub> is summarized in Table 4. Visual IHP assessment resulted in higher specificity, PPV, and overall accuracy compared with quantitative assessment. Visual Deauville also had better performance measures than quantitative assessment, but these did not reach statistical significance.

### Morphological Parameters

1. *Size*: Pre and post treatment short- and long-axis diameter of residual nodal masses is presented in Table 5. There was a difference in percentage decrease in short- and long-axis diameter of negative and positive lesions ( $p < 0.003$  and  $p < 0.002$ , respectively), with a significantly greater decrease in diameter of negative lesions. Logistic regression model used to determine the predicted probability of having a positive response based on percentage change in short axis showed that for a 20 % and 45 % decrease in short axis after therapy yields a sensitivity of 44.4 % and 73.5 % and specificity of 86.5 % and 63.5 %, respectively. Decrease in 11 % and 35 % in long-axis result in similar sensitivity and specificity profiles.

2. *Density*: Before therapy, the mean density of positive and negative lesions was similar ( $60.7 \pm 18.5$  HU and  $59.8 \pm 17.6$  HU, respectively;  $p = 0.829$ ). After therapy, the mean density of positive and negative lesions was  $54.1 \pm 29.4$  HU and  $47.9 \pm 25$  HU, respectively ( $p = 0.305$ ). The percent change in density before and after therapy was smaller for positive lesions compared with negative lesions: 7 % ( $\pm 45.8$ ) and 16 % ( $\pm 45.2$ ), respectively, but this did not reach significance ( $p = 0.351$ ).

3. *Borders*: Borders of the residual masses post treatment were infiltrative in 30/104 negative lesions (28.8 %), and 7/34 positive lesions (20.6 %);  $p = 0.345$ .

4. *Homogeneity*: No significant association was found between residual lesion homogeneity and response, with 84/104 negative lesions (80.7 %), and 28/34 positive lesions (82.4 %) having homogeneous density, respectively ( $p = 0.838$ ).

### Combination of Morphological and Functional Parameters

#### Change in Size and IHP Criteria

The optimal performance was obtained when adding a cutoff of 50 % change in short-axis diameter to all positive lesions by IHP-c. This resulted in a sensitivity, specificity, PPV, NPV and accuracy of 82.4 % (28/34), 90.4 % (94/104), 73.7 % (28/38), 94 % (94/100) and 88.4 % (122/138), respectively. These

**Table 4** Comparison of Visual Inspection and Quantitative Measurements

| Method                        | Sensitivity        | Specificity          | PPV                 | NPV                  | Accuracy             |
|-------------------------------|--------------------|----------------------|---------------------|----------------------|----------------------|
| Visual IHP                    | 94.64 %<br>(53/56) | 81.45 %<br>(101/124) | 69.74 %<br>(53/76)  | 97.12 %<br>(101/104) | 85.56 %<br>(154/180) |
| SUV <sub>mean</sub> IHP       | 94.64 %<br>(53/56) | 58.33 %<br>(70/120)  | 51.46 %<br>(53/103) | 95.89 %<br>(70/73)   | 69.89 %<br>(123/176) |
| <i>p</i> value                | 1                  | <0.001               | 0.01                | 1                    | <0.001               |
| Visual Deauville              | 92.86 %<br>(52/56) | 94.35 %<br>(117/124) | 88.14 %<br>(52/59)  | 96.69 %<br>(117/121) | 93.89 %<br>(169/180) |
| SUV <sub>mean</sub> Deauville | 87.50 %<br>(49/56) | 91.67 %<br>(110/120) | 83.05 %<br>(49/59)  | 94.02 %<br>(110/117) | 90.34 %<br>(159/176) |
| <i>p</i> value                | 0.375              | 0.250                | 0.430               | 0.370                | 0.07                 |



**Table 5** Short- and long-axis diameter of nodal masses (in centimeters, with standard deviation) before and after therapy and percent change in diameter

|                    | <i>n</i> | Short-axis diameter    |                         |                    | Long-axis diameter     |                         |                    |
|--------------------|----------|------------------------|-------------------------|--------------------|------------------------|-------------------------|--------------------|
|                    |          | PreTx Mean ( $\pm$ SD) | PostTx Mean ( $\pm$ SD) | % Change           | PreTx Mean ( $\pm$ SD) | PostTx Mean ( $\pm$ SD) | % Change           |
| <b>Negative</b>    | 104      | 4.7 ( $\pm$ 2.7)       | 2.2 ( $\pm$ 1.1)        | 46.0 ( $\pm$ 26.0) | 8.2 ( $\pm$ 5.1)       | 5.1 ( $\pm$ 3.4)        | 34.4 ( $\pm$ 25.6) |
| <b>Positive</b>    | 34       | 3 ( $\pm$ 2.2)         | 1.9 ( $\pm$ 1.3)        | 9.7 ( $\pm$ 64.4)  | 5.2 ( $\pm$ 4.1)       | 4.2 ( $\pm$ 2.8)        | 6.6 ( $\pm$ 48.1)  |
| <b>All lesions</b> | 138      | 4.3 ( $\pm$ 2.7)       | 2.2 ( $\pm$ 1.2)        | 37.1 ( $\pm$ 41.9) | 7.5 ( $\pm$ 5.0)       | 4.9 ( $\pm$ 3.3)        | 27.5 ( $\pm$ 34.5) |

parameters were compared to performance measures of IHP-c alone ( $p=0.125, 0.008, 0.326, 0.287, 0.388$ ; respectively).

#### *Change in Size and Deauville Criteria*

Lesions with Deauville score 1 or 2 were considered negative and those with score 4 or 5 were considered positive, regardless of change in size. Lesions with score of 3 were further categorized by change in short axis before and after therapy. No specific cutoff value of change in size resulted in improved performance of Deauville-c. Although sensitivity was higher (97.1 %) for lesions with at least 50 % decrease in short-axis diameter, using this cutoff would reduce specificity to 89.4 %.

## Discussion

This study, as a previous smaller scale study, confirms that the optimal standard of reference to define positivity of a post therapy PET scan in lymphoma is background liver rather than MBP. Using liver uptake as reference to define positivity maintains high sensitivity in identifying residual disease but significantly improves the PPV of PET with fewer false positive results. These findings are in line with the recommendations made by an expert panel at the recent 4th International Workshop on PET in Lymphoma [24], and the consensus guidelines from the 11th International Conference on Malignant Lymphoma in Lugano. The Lugano consensus guidelines recommend using a five-point scale (based on Deauville-c) for end of therapy response assessment for nodal and extranodal sites of disease, with scores 1–3 considered negative for residual disease, unless treatment de-escalation is being considered in setting of response-adapted trial, where a more cautious approach may be preferred [29]. A concern using a five-point scale may be the introduction of greater variability in interpretation amongst readers. However, as previously shown for HL and diffuse large B-cell lymphoma and as also supported by our data, the clinical significance of minor disagreements amongst readers may not be high and agreement in defining positive and negative scans is excellent [25, 30].

The consensus of the imaging subcommittee of the IHP in lymphoma was that visual assessment alone is adequate for assessment at the conclusion of therapy, and that a quantitative approach (e.g., using SUV measurements) does not seem necessary [3]. The results from the current trial concur with this recommendation, as qualitative assessment was sufficient and in fact out-performed quantitative measures. There may be a few explanations for this. SUV measurements suffer from several limitations. For regions less than 3 cm, partial volume effect due to image noise and limited resolution of PET may lead to measurement errors, especially for,  $SUV_{mean}$  [31]. Although  $SUV_{mean}$  may not be as reproducible as  $SUV_{max}$ ,  $SUV_{max}$  may be dependent on the value of a single pixel and therefore prone to inaccuracies due to image noise. In clinical practice, this limitation may be at least partially controlled by noise correlations introduced during image reconstruction [32]. Combination of two semiquantitative measurements, for the residual mass and reference tissue may amplify these measurement inaccuracies.

Recent data suggest that CT response may play a complimentary role in patients with HL who have either a positive interim or post-treatment PET-CT. In a prospective trial including 729 patients with HL using IHP-c, over 26 % of patients had positive end of therapy PET. The subgroup of patients with a positive scan and a relative reduction of <40 % in long-axis diameter of the residual mass had a 23.1 % risk of progression or relapse within the 1st year, compared to only 5.3 % for patients with a larger size reduction [33]. Similarly, findings from the current study show that combination of size reduction of  $\geq 50$  % in short-axis diameter of a residual nodal mass to a positive PET scan by IHP-c improved overall specificity of PET/CT ( $p=0.008$ ). Using a lower threshold to define PET positivity (mediastinal blood pool, as per IHP-c; Tables 3 and 4) reduces the specificity of PET. A possible explanation may be the inclusion of resolving tumor-related or therapy-induced inflammatory changes when using a lower threshold. Inflammation may be attributed to release of intracellular proteins and nucleic acids to the extracellular space by necrotic cells, or to therapy-induced recruitment of immune cells [34, 35]. However, when using liver as the cutoff for a positive PET (Deauville score  $\geq 4$ ), size did not further improve specificity.

There were limitations to this study. First, the same scans were interpreted retrospectively using two different response criteria. However, studies were interpreted at a time interval to minimize recall bias. Furthermore, when defining Deauville score 3 as positive, performance was similar to IHP-c, indicating consistency between the two reads. Second, size of patient population and those who had portal venous phase contrast-enhanced CT at baseline and end of therapy may have limited assessment of other morphological parameters, and especially percentage decrease in density. It remains to be determined in a larger prospective cohort of patients, whether the addition of multiple morphological criteria to PET would further improve the performance of PET and prognostication including predicting risk of relapse with PET/CT. This would impact patient care as it would increase the diagnostic confidence in identifying residual viable disease, and may obviate the need for further workup in certain instances.

In conclusion, using liver as the visual reference to determine PET positivity for patients being assessed for residual masses at end of therapy improves specificity, yet maintains the high sensitivity of PET in identifying residual lymphoma. The addition of change in size after therapy improves specificity of PET when using IHP-based interpretation.

**Acknowledgments** This manuscript has not been published before, is not under consideration for publication anywhere else and has been approved by all co-authors.

**Conflict of Interest** All authors (Ur Metser, Ravi Mohan, Vaughan Beckley, Hadas Moshonov, David Hodgson, and Grainne Murphy) declare that they have no conflict of interest.

**Ethical Statement** This retrospective study was approved by the institution's Research Ethics Board (approval no. 13-6496-CE) and written informed consent from patients was waived. The study was performed in accordance with the ethical standards of the institution. The institutional research ethics board operates in compliance with the Tri-Council Policy Statement and the ICH guidelines for the Good Clinical Practice E6(R1).

## References

- Cheson BD, Horning SJ, Coiffier B, Shipp MA, Fisher RI, Connors JM, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. *J Clin Oncol.* 1999;17:1244–53.
- Cheson BD, Pfister B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007;25:579–86.
- 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.
- Juweid M, 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 (IWC) and 18F-fluorodeoxyglucose positron emission tomography (PET). *J Clin Oncol.* 2005;23:4652–61.
- Bangerter M, Moog F, Buchmann I, Kotzerke J, Griesshammer M, Hafner M, et al. Whole-body 2-[18F]-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for accurate staging of Hodgkin's disease. *Ann Oncol.* 1998;9:1117–22.
- Spaepen K, Stroobants S, Dupont P, Van Steenweghen S, Thomas J, Vandenberghe P, et al. Prognostic value of positron emission tomography (PET) with fluorine-18 fluorodeoxyglucose ( $[^{18}\text{F}]$ FDG) after first-line chemotherapy in non-Hodgkin's lymphoma: Is  $[^{18}\text{F}]$ FDG-PET a valid alternative to conventional diagnostic methods? *J Clin Oncol.* 2001;19:414–9.
- Spaepen K, Stroobants S, Dupont P, Vandenberghe P, Maertens J, Bormans G, et al. Prognostic value of pretransplantation positron emission tomography using fluorine 18-fluorodeoxyglucose in patients with aggressive lymphoma treated with high-dose chemotherapy and stem cell transplantation. *Blood.* 2003;102:53–9.
- Spaepen K, Stroobants S, Dupont P, Vandenberghe P, Thomas J, de Groot T, et al. Early restaging positron emission tomography with 18F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin's lymphoma. *Ann Oncol.* 2002;13:1356–63.
- Jerusalem G, Beguin Y, Fassotte MF, Najjar F, Paulus P, Rigo P, et al. Whole-body positron emission tomography using 18F-fluorodeoxyglucose compared to standard procedures for staging patients with Hodgkin's disease. *Haematologica.* 2001;86:266–73.
- Jerusalem G, Beguin Y, Fassotte MF, Najjar F, Paulus P, Rigo P, et al. Whole-body positron emission tomography using 18F-fluorodeoxyglucose for posttreatment evaluation in Hodgkin's disease and non-Hodgkin's lymphoma has higher diagnostic and prognostic value than classical computed tomography scan imaging. *Blood.* 1999;94:429–33.
- Jerusalem G, Beguin Y, Najjar F, Hustinx R, Fassotte MF, Rigo P, et al. Positron emission tomography (PET) with 18F-fluorodeoxyglucose (18F-FDG) for the staging of low-grade non-Hodgkin's lymphoma (NHL). *Ann Oncol.* 2001;12:825–30.
- Jerusalem G, Warland V, Najjar F, Paulus P, Fassotte MF, Fillet G, et al. Whole-body 18F-FDG PET for the evaluation of patients with Hodgkin's disease and non-Hodgkin's lymphoma. *Nucl Med Commun.* 1999;20:13–20.
- Zinzani PL, Magagnoli M, Chierichetti F, Zompatori M, Garraffa G, Bendandi M, et al. The role of positron emission tomography (PET) in the management of lymphoma patients. *Ann Oncol.* 1999;10:1141–3.
- Naumann R, Vaic A, Beuthien-Baumann B, Bredow J, Kropp J, Kittner T, et al. Prognostic value of positron emission tomography in the evaluation of post-treatment residual mass in patients with Hodgkin's disease and non-Hodgkin's lymphoma. *Br J Haematol.* 2001;115:793–800.
- Munker R, Glass J, Griffith LK, Sattar T, Zamani R, Heldmann M, et al. Contribution of PET imaging to the initial staging and prognosis of patients with Hodgkin's disease. *Ann Oncol.* 2004;15:1699–704.
- Mikhael NG, Hutchings M, Fields PA, Nunan T, Timothy AR. FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol.* 2005;16:1514–23.
- Juweid M, Cheson BD. Positron emission tomography (PET) in post-therapy assessment of cancer. *N Engl J Med.* 2006;354:496–507.
- Haioun C, Itti E, Rahmoui A, Brice P, Rain JD, Belhadj K, et al.  $[^{18}\text{F}]$ fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood.* 2005;106:1376–81.
- Hutchings M, Loft A, Hansen M, Pedersen LM, Berthelsen AK, Keiding S, et al. Positron emission tomography with or without

- computed tomography in the primary staging of Hodgkin's lymphoma. *Haematologica*. 2006;91:482–9.
20. Olsen K, Sohi J, Abraham T, Juweid M. Initial validation of standardized qualitative (visual) criteria for FDG-PET assessment of residual masses following lymphoma therapy. *Radiological Society of North America 92nd Scientific Assembly and Annual Meeting Program*, 2006, pp 323 (abstract 55:E23-02).
  21. Brepoels 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.
  22. Brepoels 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.
  23. Gallamini A, Rigacci L, Merli F, Nassi L, Bosi A, Capodanno I, et al. Predictive value of positron emission tomography performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. *Haematologica*. 2006;91:475–81.
  24. Meignan M, Barrington S, Itti E, Gallamini A, Haioun C, Polliack A. Report on the 4th International Workshop on Positron Emission Tomography in Lymphoma held in Menton, France, 3–5 October 2012. *Leuk Lymphoma*. 2014;55(1):31–7.
  25. Itti E, Meignan M, Berriolo-Riedinger A, Biggi A, Cashen AF, Vera P, et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and  $\Delta$ SUVmax. *Eur J Nucl Med Mol Imaging*. 2013;40:1312–20.
  26. Dupuis J, Berriolo-Riedinger A, Julian A, Brice P, Tychyj-Pinel C, Tilly H, et al. Impact of [(18)F]fluorodeoxyglucose positron emission tomography response evaluation in patients with high-tumor burden follicular lymphoma treated with immunochemotherapy: a prospective study from the Groupe d'Etudes des Lymphomes de l'Adulte and GOELAMS. *J Clin Oncol*. 2012;30:4317–22.
  27. Manohar K, Mittal BR, Raja S, Bhattacharya A, Malhotra P, Varma S. Comparison of various criteria in interpreting end of therapy F-18 labeled fluorodeoxyglucose positron emission tomography/computed tomography in patients with aggressive non-Hodgkin lymphoma. *Leuk Lymphoma*. 2013;54:714–9.
  28. Kajáry K, Molnár Z, Györke T, Szakáll Jr S, Molnár P, Lengyel Z. Comparison of the International Harmonization Project, London and Gallamini criteria in the interpretation of 18F-FDG PET/CT examinations after first-line treatment in Hodgkin's lymphoma. *Nucl Med Commun*. 2014;35:169–75.
  29. 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.
  30. Barrington SF, Qian W, Somer EJ, Franceschetto A, Begni 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.
  31. Kinahan PE, Fletcher JW. PET/CT standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy. *Semin Ultrasound CT MR*. 2010;31:496–505.
  32. Doot RK, Scheuermann JS, Christian PE, Karp JS, Kinahan PE. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. *Med Phys*. 2010;37:6035–46.
  33. Kobe C, Kuhnert G, Kahraman D, Haverkamp H, Eich HT, Franke M, et al. Assessment of tumor size reduction improves outcome prediction of positron emission tomography/computed tomography after chemotherapy in advanced-stage Hodgkin lymphoma. *J Clin Oncol*. 2014;32:1776–81.
  34. Han HS, Escalon MP, Hsiao B, Serafini A, Lossos IS. High incidence of false-positive PET scans in patients with aggressive non-Hodgkin's lymphoma treated with rituximab-containing regimens. *Ann Oncol*. 2009;20:309–18.
  35. Shacter E, Williams JA, Hinson RM, Sentürker S, Lee YJ. Oxidative stress interferes with cancer chemotherapy: inhibition of lymphoma cell apoptosis and phagocytosis. *Blood*. 2000;96:307–13.