

# FDG PET/CT response in diffuse large B-cell lymphoma

## Reader variability and association with clinical outcome

Eun Ji Han, MD, PhD<sup>a</sup>, Joo Hyun O, MD, PhD<sup>b,\*</sup>, Hyukjin Yoon, MD<sup>b</sup>, Seung Eun Jung, MD, PhD<sup>b</sup>, Gyeongsin Park, MD, PhD<sup>c</sup>, Byung Ock Choi, MD, PhD<sup>d</sup>, Seok-Goo Cho, MD, PhD<sup>e</sup>

### Abstract

F-18-fluoro-2-deoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT) is essential for monitoring response to treatment in patients with diffuse large B-cell lymphoma (DLBCL) and qualitative interpretation is commonly applied in clinical practice. We aimed to evaluate the interobserver agreements of qualitative PET/CT response in patients with DLBCL and the predictive value of PET/CT results for clinical outcome.

PET/CT images were obtained for patients with DLBCL 3 times: at baseline, after 3 cycles of first-line chemotherapy (interim), and after completion of chemotherapy. Two nuclear medicine physicians (with 3 and 8 years of experience with PET/CT) retrospectively assessed response to chemotherapy blinded to the clinical outcome using International Harmonization Project (IHP) criteria and Deauville 5-point score. The associations between PET/CT results and progression-free survival (PFS) and overall survival (OS) were assessed using Cox regression analysis.

A total of 112 PET/CT images were included from 59 patients with DLBCL (36 male, 23 female; mean age  $53 \pm 14$  years). Using the IHP criteria, interobserver agreement was substantial (Cohen  $\kappa=0.76$ ) with absolute agreement consistency of 89%. Using the Deauville score, interobserver agreement was moderate (Cohen weighted  $\kappa=0.54$ ) and absolute consistency was 62%. The most common cause of disagreements was discordant interpretation of residual tumor uptake. With median follow-up period of 60 months, estimated 5-year PFS and OS were 81% and 92%, respectively. Neither interim nor posttreatment PET/CT results by both readers were significantly associated with PFS. Interim PET/CT result by the more experienced reader using Deauville score was a significant factor for OS ( $P=0.019$ ).

Moderate-to-substantial interobserver agreement was observed for response assessments according to qualitative PET/CT criteria, and interim PET/CT result could predict OS in patients with DLBCL. Further studies are necessary to further standardize the PET/CT-based response criteria for more consistent interpretation.

**Abbreviations:** CI = confidence interval, CR = complete remission, CT = computed tomography, D5PS = Deauville 5-point score, DLBCL = diffuse large B-cell lymphoma, FDG = F-18-fluoro-2-deoxyglucose, IHP = International Harmonization Project, LDH = lactate dehydrogenase, MRU = minimal residual uptake, NCCN = National Comprehensive Cancer Network, OS = overall survival, PET = positron emission tomography, PFS = progression-free survival, R-CHOP = rituximab, cyclophosphamide, hydroxydaunomycin, oncovin, and prednisone.

**Keywords:** DLBCL, FDG PET/CT, outcome, reader variability, response criteria

Editor: Ramya Dhandapani.

Funding/support: This study was supported by Research Fund of Seoul St. Mary's Hospital, The Catholic University of Korea.

The authors have no conflicts of interest to disclose.

<sup>a</sup> Department of Radiology, Daejeon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Daejeon, <sup>b</sup> Department of Radiology,

<sup>c</sup> Department of Pathology, <sup>d</sup> Department of Radiation Oncology, <sup>e</sup> Department of Hematology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea.

\* Correspondence: Joo Hyun O, Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222, Banpo-daero, Seocho-gu, Seoul 06591, Republic of Korea (e-mail: jooohyun@catholic.ac.kr).

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medicine (2016) 95:39(e4983)

Received: 12 July 2016 / Received in final form: 4 August 2016 / Accepted: 4 September 2016

<http://dx.doi.org/10.1097/MD.0000000000004983>

## 1. Introduction

F-18-fluoro-2-deoxyglucose (FDG) positron emission tomography (PET) and PET/computed tomography (CT) are used worldwide for evaluation and management of various malignancies.<sup>[1]</sup> Because PET/CT has high diagnostic accuracy, especially for aggressive types of lymphoma compared to conventional imaging studies such as CT or magnetic resonance imaging, and has predictive value for outcomes,<sup>[2,3]</sup> it is recommended as an essential staging tool of various Hodgkin and non-Hodgkin lymphomas including diffuse large B-cell lymphoma (DLBCL). In addition, PET/CT is applied for assessment of response to first-line chemotherapy and subsequent treatment planning in DLBCL.<sup>[4,5]</sup> Recent reports showed interim PET/CT, performed midway through the chemotherapy, to be useful for predicting treatment response and prognosis.<sup>[6]</sup> Thus, risk-adapted and response-adapted therapies based on interim PET/CT results have been actively investigated.<sup>[7,8]</sup>

Accurate response assessment directs the decision for either treatment continuance or regimen change and can provide predictive value for prognosis.<sup>[9]</sup> Various criteria are currently

used to assess treatment response by imaging studies in the clinical setting, and the most widely used Response Evaluation Criteria in Solid Tumors and World Health Organization criteria are based on anatomical, more specifically, size changes.<sup>[10–12]</sup> In malignant lymphoma, residual masses are frequently observed even after successful treatment because of fibrosis or necrosis. Residual masses do not always represent viable tumors and do not indicate poorer outcomes.<sup>[13,14]</sup> Therefore, FDG uptake-based approach is applied for evaluation of metabolic change in lymphoma.<sup>[9]</sup> FDG PET/CT images can be evaluated qualitatively or quantitatively, and the qualitative analysis is more commonly applied in response assessment of lymphoma. The current National Comprehensive Cancer Network (NCCN) guidelines recommended that FDG PET images should be interpreted via Lugano classification incorporating the Deauville 5-point score (D5PS), which is based on visual assessment.<sup>[5,15]</sup> Revised International Harmonization Project (IHP) response criteria accept simple binary decision for PET-based response.<sup>[16]</sup> The degree of interobserver variation and reproducibility are important in these qualitative analyses, which have pivotal role in the patient management.

The purpose of our study was to evaluate the interobserver agreements of qualitative PET/CT response criteria in patients with DLBCL and analyze the possible causes of disagreements in response interpretation between readers. In addition, we evaluated the predictive value of PET/CT for long-term outcome.

## 2. Methods

### 2.1. Patient populations

We retrospectively reviewed 112 FDG PET/CT images of 59 consecutive patients with DLBCL taken between December 2006 and March 2012. FDG PET/CT images were performed at baseline (baseline PET/CT); after 3 cycles of rituximab, cyclophosphamide, hydroxydaunomycin, oncovin, and prednisone (R-CHOP) chemotherapy (interim PET/CT); and after completion of chemotherapy (posttreatment PET/CT). Cases with low FDG uptake in lymphoma lesions, below the hepatic activity, in the baseline PET/CT images were considered to be unassessable by FDG PET/CT and excluded from this study. Clinicopathologic variables such as age, sex, Eastern Cooperative Oncology Group performance status, Ann Arbor stage, lactate dehydrogenase (LDH) titer, extranodal involvement, bone marrow involvement, splenomegaly, international prognostic index score at diagnosis, and survival outcomes were obtained from medical and imaging records.

This study was approved by the institutional review board of Seoul St. Mary's Hospital, The Catholic University of Korea. The ethical committee of our institution waived the need for requiring patient consent for this retrospective review of imaging studies and clinical data.

### 2.2. FDG PET/CT imaging

All patients fasted for at least 6 hours before the PET/CT studies. FDG (370–555 MBq) was injected intravenously, and scanning began 60 minutes later. No patient had blood glucose levels greater than 150 mg/dL before injection. No intravenous contrast agent was administered. Studies were acquired using combined PET/CT in-line systems, Biograph Duo or Biograph Truepoint (Siemens Medical Solutions, Knoxville, TN). All patients were in a supine position. CT began at the orbitomeatal line and progressed to the upper thigh using a standard protocol: 130

kVp, 30 mA, 5-mm slice thickness (Biograph Duo); 120 kV, 50 mA, 5-mm slice thickness (Biograph Truepoint). PET followed immediately over the same body region. Acquisition time was 2 to 3 minutes per bed position. CT data were used for attenuation correction, and images were reconstructed using standard ordered-subset expectation maximization.

### 2.3. Image analysis

All PET/CT images were retrospectively assessed for response to chemotherapy by 2 nuclear medicine physicians, blinded to the clinical data and to each other's assessments. Reader A was a nuclear medicine resident with 3 years of experience in interpretation of FDG PET/CT images. Reader B was a board-certified nuclear medicine physician with 8 years of experience in interpretation of FDG PET/CT images and had periodically participated in lymphoma multidisciplinary meetings. All images were viewed at a workstation with fusion software (Syngo; Siemens) that provided multiplanar reformatted images and displayed PET images after attenuation correction, CT images, and PET/CT fusion images.

Interim or posttreatment PET/CT images were compared with the baseline PET/CT images using 2 visual analysis methods: the IHP response criteria and Lugano classification. In the IHP response criteria, last revised in 2007, the 4 recommended response categories are complete remission (CR), partial remission, stable disease, and progressive disease, and PET-based response is based on simple binary decision (PET-positive or PET-negative). PET-negative is considered to be moderately sized or large residual lesions (more than 2 cm at the greatest transverse diameter) with FDG uptake isointense or less intense than mediastinal blood pool structures. For smaller lesions less than 2 cm in diameter, FDG uptake that cannot be differentiated from surrounding background activity is considered negative because of the partial volume effect.<sup>[16]</sup> The readers recorded the sites of residual lesions or newly detected lesions when they interpreted the case as PET-positive. Lugano classification for response assessment of PET/CT incorporating the D5PS is as follows: score 1, no FDG uptake above background; score 2, FDG uptake  $\leq$  mediastinum; score 3, FDG uptake  $>$  mediastinum but  $\leq$  liver; score 4, FDG uptake moderately  $>$  liver; and score 5, FDG uptake markedly higher than liver and/or new lesions.<sup>[15]</sup> The readers recorded the sites of increased FDG uptake lesions or new lesions in cases interpreted as having score of 2, 3, 4, or 5. Based on the current NCCN guidelines, the PET/CT results by D5PS were then dichotomized using hepatic uptake as the reference (score 1, 2, or 3 vs score 4 or 5).<sup>[5]</sup>

### 2.4. Statistical analysis

Categorical variables were expressed as an absolute number and percentage, and continuous variables were expressed as median or mean  $\pm$  standard deviation and range. Interobserver agreement was measured using McNemar test and Cohen  $\kappa$ . A  $\kappa$  value of 0.0 to 0.2 was considered to represent slight agreement; 0.21 to 0.4, fair; 0.41 to 0.6, moderate; 0.61 to 0.8, substantial; and 0.81 to 1.0, almost perfect.<sup>[17,18]</sup>

The primary endpoint was overall survival (OS), which was defined as the time from the date of baseline PET/CT to the date of death from DLBCL. Progression-free survival (PFS) was defined as the time from the date of baseline PET/CT to the date progression was detected, or date of last clinical follow-up. The Kaplan–Meier method and log-rank test were used to estimate

**Table 1****General characteristics of 59 patients.**

Variables		Number of patients
Sex	Male	36 (61%)
	Female	23 (39%)
Age	Mean $\pm$ SD (range)	53 $\pm$ 14y (17–79)
ECOG PS	0	9 (15%)
	1	39 (66%)
	2	11 (19%)
Ann Arbor stage	I	10 (17%)
	II	17 (29%)
	III	11 (19%)
	IV	21 (35%)
LDH titer	Normal	33 (56%)
	Elevated	26 (44%)
Extranodal involvement	No	21 (36%)
	Yes	38 (64%)
Bone marrow involvement	No	47 (80%)
	Yes	12 (20%)
Splenomegaly	No	54 (92%)
	Yes	5 (8%)
IPI score	Low risk (0–1 point)	31 (52%)
	Low-intermediate risk (2 points)	11 (19%)
	High-intermediate risk (3 points)	6 (10%)
	High risk (4–5 points)	11 (19%)

ECOG PS = Eastern Cooperative Oncology Group performance status, IPI = international prognostic index, LDH = lactate dehydrogenase, SD = standard deviation.

the 5-year PFS and OS. A Cox proportional hazards model provided univariate analysis for the identification of significant prognostic factors for OS. Statistical analysis was carried out using SPSS software version 21.0 (IBM, Armonk, NY). A 95% confidence interval (CI) was calculated, and a *P* value less than 0.05 was considered significant.

### 3. Results

#### 3.1. Patient characteristics

A total of 55 interim and 57 posttreatment PET/CT images of 59 patients with DLBCL (36 male, 23 female; mean age 53  $\pm$  14 years) were included. At baseline, 10 patients (17%) were Ann Arbor stage I, 17 (29%) were stage II, and 32 (54%) had advanced stage. Bone marrow involvement was confirmed by biopsy in 12 patients (20%), and 26 patients (44%) had elevated LDH titers. The general characteristics of patients are in Table 1.

#### 3.2. Interobserver agreement of PET/CT response

Of a total of 112 PET/CT images using the IHP criteria, reader A interpreted 77 as PET-negative and 35 as PET-positive; reader B interpreted 73 as PET-negative and 39 PET-positive (Table 2). There was substantial interobserver agreement (Cohen  $\kappa$  = 0.76) with absolute agreement consistency of 89% (100 of 112). No

significant differences were seen between readers in interpretations of interim (*P* = 0.250, McNemar test), posttreatment (*P* = 1.000), and both time point assessments (*P* = 0.388). Most of the disagreements (11 of 12) were due to inconsistencies in interpretation of residual FDG uptake. In 1 case, the disagreement was due to different interpretation of lung and pleural lesions in the posttreatment PET/CT.

The interpretation results by the 2 readers using the D5PS are given in Table 3. For a total of 112 PET/CT images, the interobserver agreement was moderate (Cohen  $\kappa$  = 0.43 and weighted  $\kappa$  = 0.54) and the absolute consistency was 62% (69 of 112). Of the 43 disagreeing cases, 1-grade disagreement was the most frequent (*n* = 30). The causes of disagreements were discordant interpretation of residual tumor uptake in areas with very low background activity such as cervical nodal chains, lungs, and mesentery (*n* = 18) or very high physiologic activity such as tonsils, bowel, and bone marrow (*n* = 17). In the remaining 8 disagreeing cases, there were concomitant benign conditions such as postinflammatory change, reactive lymph nodes, and postoperative change. When the results of D5PS were dichotomized as score 1, 2, or 3 versus score 4 or 5, there was 82% agreement for interim PET/CT (Cohen  $\kappa$  = 0.41), 88% agreement for posttreatment PET/CT (Cohen  $\kappa$  = 0.52), and 85% agreement for both time points (Cohen  $\kappa$  = 0.46). Interpretation results of a total of 112 PET/CT images showed significant difference between readers (*P* = 0.013, McNemar test).

#### 3.3. Survival prediction

The median follow-up period was 60 months (range 7–103); the median observation time was 60 months (range 24–103) for the surviving patients, and the median survival time was 33 months (range 7–86) for the deceased patients. Of 59 patients, 13 patients presented with tumor progression during follow-up, and the estimated 5-year PFS was 81%. The prognostic performances of PET/CT results are given in Table 4. Neither interim nor posttreatment PET/CT results were significantly associated with PFS, although the tendency was noted for the interim PET/CT response by reader B using Deauville score (*P* = 0.068). Of 50 patients with negative posttreatment PET/CT (Deauville score 1–3), 12 (24%) had progression. Of 7 patients with positive posttreatment PET/CT (Deauville score 4–5), 1 patient had progression with PFS of 64.3 months, 2 patients had additional radiotherapy following positive posttreatment PET/CT findings and are presently in ongoing CR, and 4 patients have now been in CR without additional therapy.

During follow-up, 8 of the 59 patients died and the estimated 5-year OS was 92%. One patient's death was due to a cause unrelated to lymphoma. Forty-five patients remained alive without tumor progression. Four patients had tumor progression, but remained alive after salvage or palliative treatment. Estimating predictive power of PET/CT results by readers,

**Table 2****Response assessment using IHP response criteria.**

Response	Interim PET/CT (n = 55)		Posttreatment PET/CT (n = 57)	
	Reader A	Reader B	Reader A	Reader B
PET-negative	36 (65%)	33 (60%)	41 (72%)	40 (70%)
PET-positive	19 (35%)	22 (40%)	16 (28%)	17 (30%)
Cohen's $\kappa$	0.88		0.62	

IHP = International Harmonization Project, PET/CT = positron emission tomography/computed tomography.

**Table 3****Response assessment using D5PS.**

		Reader B score					Total
		1	2	3	4	5	
Reader A score	1	45	0	1	0	0	46 (41%)
	2	13	12	1	1	1	28 (25%)
	3	3	7	3	1	0	14 (12%)
	4	5	1	7	9	0	22 (20%)
	5	1	0	0	1	0	2 (2%)
Total		67 (60%)	20 (18%)	12 (11%)	12 (11%)	1 (1%)	112 (100%)

interim PET/CT result by reader B using Deauville score was the only significant factor for OS (Table 4). Of 8 deceased patients, only 1 patient had positive posttreatment PET/CT with Deauville score 4 and the other 7 patients had negative posttreatment PET/CT with Deauville score 1 or 2. The mean OS in patients with Deauville score 1, 2, or 3 in interim PET/CT images (95 months, 95% CI 86.715–102.507) was significantly longer than patients with Deauville score 4 or 5 (66 months, 95% CI 37.099–94.423;  $P=0.008$ , log-rank test) (Fig. 1).

#### 4. Discussion

One purpose of this study was to evaluate the interobserver agreements for 2 qualitative PET/CT response criteria widely used for patients with DLBCL. Substantial interobserver agreement was seen for the IHP criteria, whereas moderate interobserver agreement was seen for D5PS. When results of D5PS were dichotomized, the interobserver variability was slightly increased (Cohen  $\kappa$  from 0.43 to 0.46). Previous studies using mediastinal uptake as reference demonstrated Cohen  $\kappa$  ranging from 0.445 to 0.65.<sup>[19,20]</sup> In previous studies of response assessment using D5PS, few reported the interobserver agreement for all 5 points of the scoring method, and only a fair degree of agreement was observed with Cohen  $\kappa$  ranging from 0.13 to 0.387.<sup>[21,22]</sup> The dichotomization of the D5PS (score 1–3 vs score 4–5) produced higher agreement with Cohen  $\kappa$  ranging from 0.502 to 0.96.<sup>[19,21,23]</sup> Overall, the results of our study and previous studies show a wide range of interobserver agreement in qualitative PET analysis.

Mikhael et al<sup>[24,25]</sup> used the term of “minimal residual uptake (MRU)” for interim PET in patients with aggressive lymphoma. MRU is defined as residual low FDG uptake in previously involved site that is likely to represent inflammation from which small volume of residual viable tumor cannot be definitely

excluded and indicates neither positive nor negative PET findings. MRU designation is subjective and depends on the reader’s experience, reference point, and various elements of the imaging protocols.<sup>[26,27]</sup> In our study, most of the disagreements between PET-negative and PET-positive cases were due to inconsistencies in interpreting the residual FDG uptake. Using D5PS, 1- or 2-grade disagreements were observed in 18 cases due to discordant interpretation of the presence of residual tumor uptake in areas of very low background activity such as lungs or fat tissue, and most of them were disagreements between score 1 and score 2. Similarly, observer variability study by Zijlstra et al<sup>[28]</sup> reported that most common interpretation of equivocal for viable lymphoma was noted in the neck, around the abdominal aorta and iliac chains, and in the lungs. In the other end of the FDG uptake spectrum, discordant interpretations were in areas of high physiologic activity ( $n=17$ ) or concomitant benign condition ( $n=8$ ) (Fig. 2). Using the IHP criteria, 1 disagreeing case was due to discordant interpretation of increased extent of lung and pleural lesions. The lesion was from active inflammation when correlating with clinical history. Infectious or inflammatory lung lesions are frequently noted in patients with malignant lymphoma, and the IHP reported that new lung nodules are mostly benign in patients without history of pulmonary involvement.<sup>[16,29]</sup> In results of our study, interim PET/CT results by the less-experienced reader (reader A) were not significantly associated with PFS and OS. In PET/CT results using visual assessment, disagreeing interpretations can occur according to readers’ training and expertise.

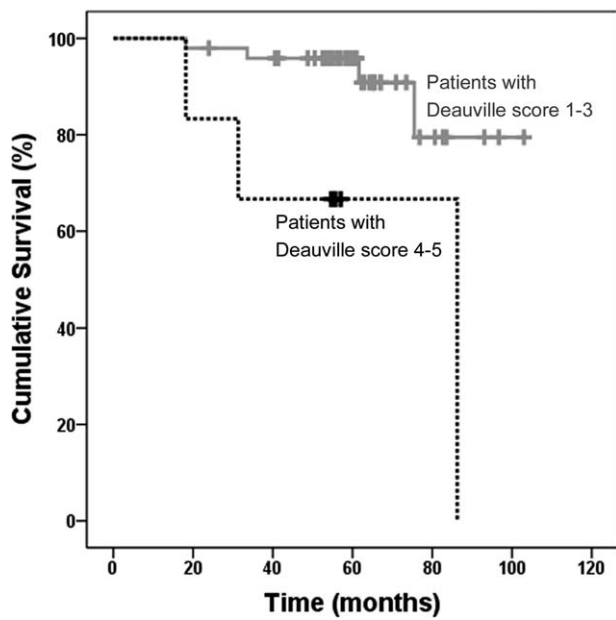
The second purpose of this study was to evaluate the predictive value of PET/CT results for clinical outcome. The OS was significantly different between interim PET/CT-positive and PET/CT-negative patients in our study, whereas PFS could not be predicted. Several studies reported that interim PET/CT has significant predictive value for PFS and OS in patients with

**Table 4****Survival prediction by PET/CT results.**

			PFS			OS		
			HR	95% CI	P	HR	95% CI	P
Interim PET/CT	IHP criteria	Reader A	1.938	0.623–6.032	0.253	1.224	0.271–5.533	0.793
		Reader B	2.229	0.704–7.057	0.173	1.060	0.236–4.759	0.940
	Deauville score	Reader A	1.732	0.520–5.775	0.371	2.518	0.559–11.344	0.229
		Reader B	3.402	0.915–12.639	0.068	6.346	1.356–29.701	0.019
Posttreatment PET/CT	IHP criteria	Reader A	2.567	0.857–7.694	0.092	1.330	0.312–5.673	0.700
		Reader B	2.023	0.675–6.060	0.208	0.615	0.122–3.103	0.556
	Deauville score	Reader A	1.525	0.419–5.556	0.522	0.619	0.076–5.052	0.654
		Reader B	0.551	0.072–4.243	0.567	0.894	0.108–7.383	0.917

CI = confidence interval, HR = hazard ratio, IHP = International Harmonization Project, OS = overall survival, PET/CT = positron emission tomography/computed tomography, PFS = progression-free survival.

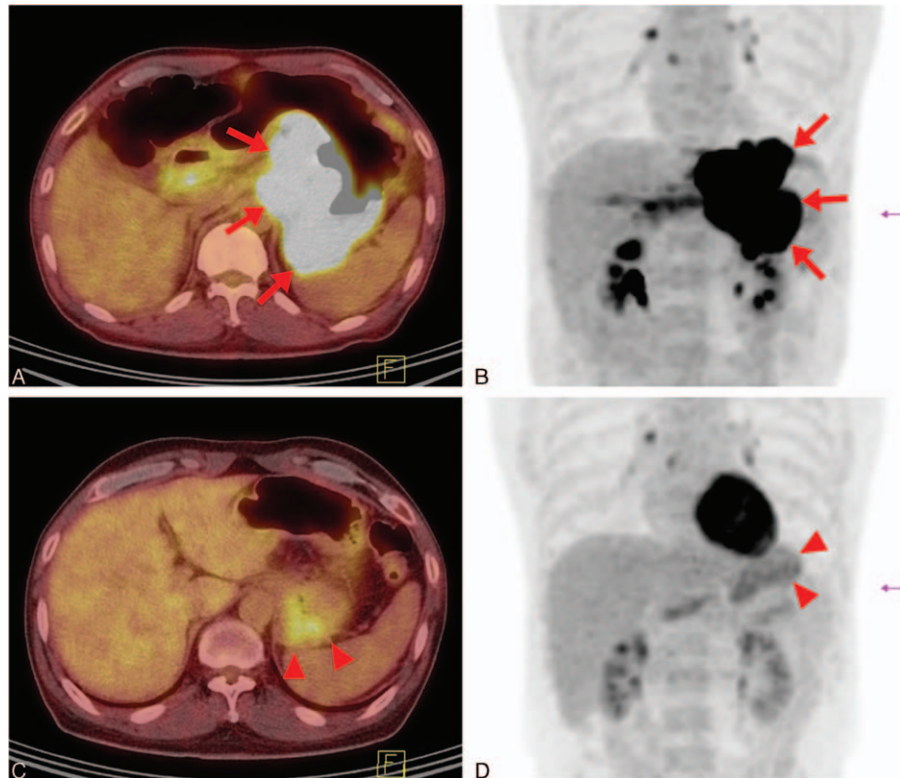




**Figure 1.** Kaplan–Meier curve of overall survival (OS). OS was significantly higher in patients with Deauville score 1, 2, or 3 than in those with score 4 or 5 in interim PET/CT images ( $P=0.008$ ). PET/CT = positron emission tomography/computed tomography.

DLBCL.<sup>[30,31]</sup> However, a systemic review including 7 studies of 311 DLBCL patients, the prognostic value of interim PET/CT was described as having conflicting results.<sup>[32]</sup> In a prospective multicenter study by Mamot et al,<sup>[34]</sup> interim PET/CT after 2 cycles of R-CHOP chemotherapy was significantly associated with event-free survival, but not with OS. Heterogeneities of patient populations, therapy strategies, and the timing of the scans might explain these differences. Interim PET/CT has been performed after 1, 2, 3, or 4 cycles of chemotherapy, and no consensus is formed on the optimal timing for interim assessment. In patients with DLBCL, interim PET in modern therapeutic era including rituximab tends to have lower positive predictive value than before. Some studies reported that increased false-positive rate of interim PET is a concern with the long half-life and unique mechanisms of cytotoxicity of rituximab.<sup>[20,33]</sup>

For posttreatment PET/CT as well, the reports on prognostic potential have been mixed.<sup>[6,35,36]</sup> Recent studies reported that PET/CT interpretation using Deauville score better predicts outcome than IHP criteria.<sup>[37,38]</sup> Our results demonstrated that posttreatment PET/CT applying the Deauville score was not significantly associated with PFS and OS and showed low positive predictive value of 14%. In addition, 24% of the patients had progression despite negative posttreatment PET/CT results. Our patients had longer follow-up period with median of 5 years compared to previous studies<sup>[35–37]</sup> and showed relatively longer PFS and OS with only 8 deaths during the follow-up period. Of 8



**Figure 2.** A 53-year-old man with gastric diffuse large B-cell lymphoma. In baseline positron emission tomography/computed tomography (PET/CT) transaxial (A) and PET maximum intensity projection (MIP) (B) images, intense F-18-fluoro-2-deoxyglucose (FDG) uptake was seen along the gastric wall from cardia to body, mainly along the lesser curvature side (maximum standardized uptake value 28.2) (arrows). In interim PET/CT transaxial (C) and PET MIP (D) images after 3 cycles of first-line chemotherapy, previously noted FDG uptake in the stomach showed markedly decreased intensity and extent (arrowheads). Remaining FDG uptake could not be differentiated from physiological gastric activity. Reader A interpreted this finding as Deauville score 4 and reader B interpreted it as score 1. After completion of first-line chemotherapy, the patient remains in clinical remission without additional therapy (overall survival 76.8 months). CT = computed tomography, FDG = F-18-fluoro-2-deoxyglucose, MIP = maximum intensity projection, PET = positron emission tomography.

deceased patients, 1 patient with positive posttreatment PET/CT result had the longest PFS of 64.3 months and OS of 86.3 months, whereas the 7 other patients with negative posttreatment PET/CT had relatively short PFS with median of 8.3 months (range 6.4–27.8). Small number of events and potent salvage therapy options might cause failure to assign significant association between posttreatment PET/CT and OS.

Although qualitative PET analysis has great advantage of easy application and is commonly used in evaluation of lymphoma, quantitative analysis is often used for evaluation of other solid tumors.<sup>[9,39]</sup> Quantitative assessment shows excellent intra-observer and interobserver reproducibility.<sup>[40,41]</sup> In an international confirmatory study by Itti et al,<sup>[19]</sup> interim PET/CT had higher interobserver reproducibility and prognostic value when using percentage change of maximum standardized uptake value than using Deauville score. Other previous studies reported that in DLBCL interim assessment, quantitative method has higher prognostic value compared to the visual assessment.<sup>[30,42]</sup> Lugano classification suggested that Deauville score 3 should be carefully interpreted depending on the timing of assessment, the clinical context, and the treatment choice.<sup>[15]</sup> Quantitative analysis could provide supplementary information to enhance the prognostic value and accuracy of PET/CT in response assessment of DLBCL.

In conclusion, an unsatisfactory interobserver agreement was observed for response assessment of PET/CT using visual analysis in patients with DLBCL who underwent first-line R-CHOP chemotherapy. Posttreatment PET/CT results using Deauville score had limited value for prediction of outcome. More standardized or detailed visual analysis method, or application of quantitative response assessment, should be considered for patients with DLBCL.

## References

- Von Schulthess GK, Hany TF. Imaging and PET-PET/CT imaging. *J Radiol* 2008;89:438–47. quiz 448.
- Gallamini A, Borra A. Role of PET in lymphoma. *Curr Treat Options Oncol* 2014;15:248–61.
- Kostakoglu L, Cheson BD. Current role of FDG PET/CT in lymphoma. *Eur J Nucl Med Mol Imaging* 2014;41:1004–27.
- Tilly H, Dreyling M, Group EGW. Diffuse large B-cell non-Hodgkin's lymphoma: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol* 2009;20(suppl 4):110–2.
- Non-Hodgkin's Lymphomas version 2.2016. National Comprehensive Cancer Network. <http://www.nccn.org>. [Accessed April 1, 2016].
- Zhu Y, Lu J, Wei X, et al. The predictive value of interim and final [18F] fluorodeoxyglucose positron emission tomography after rituximab-chemotherapy in the treatment of non-Hodgkin's lymphoma: a meta-analysis. *Biomed Res Int* 2013;2013:275805.
- Pardal E, Coronado M, Martin A, et al. Intensification treatment based on early FDG-PET in patients with high-risk diffuse large B-cell lymphoma: a phase II GELTAMO trial. *Br J Haematol* 2014;167:327–36.
- Kasamon YL, Wahl RL. FDG PET and risk-adapted therapy in Hodgkin's and non-Hodgkin's lymphoma. *Curr Opin Oncol* 2008;20:206–19.
- Wahl RL, Jacene H, Kasamon Y, et al. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med* 2009;50(suppl 1):1225–50S.
- Therasse P, Arbuuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92:205–16.
- Shanbhogue AK, Karnad AB, Prasad SR. Tumor response evaluation in oncology: current update. *J Comput Assist Tomogr* 2010;34:479–84.
- Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis* 2010;30:52–60.
- Tredaniel J, Brice P, Lepage E, et al. The significance of a residual mediastinal mass following treatment for aggressive non-Hodgkin's lymphomas. *Eur Respir J* 1992;5:170–3.
- Jochelson M, Mauch P, Balikian J, et al. The significance of the residual mediastinal mass in treated Hodgkin's disease. *J Clin Oncol* 1985;3:637–40.
- Cheson BD, Fisher RI, Barrington SF, 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, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007;25:579–86.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;10:37–46.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Itti E, Meignan M, Berriolo-Riedinger A, 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 DeltaSUVmax. *Eur J Nucl Med Mol Imaging* 2013;40:1312–20.
- Horning SJ, Juweid ME, Schoder H, 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.
- Oki Y, Chuang H, Chasen B, et al. The prognostic value of interim positron emission tomography scan in patients with classical Hodgkin lymphoma. *Br J Haematol* 2014;165:112–6.
- 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.
- 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.
- Mikhaeel NG, Timothy AR, O'Doherty MJ, et al. 18-FDG-PET as a prognostic indicator in the treatment of aggressive Non-Hodgkin's Lymphoma-comparison with CT. *Leuk Lymphoma* 2000;39:543–53.
- Mikhaeel NG, Hutchings M, Fields PA, et al. 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.
- Keyes JWJr. SUV: standard uptake or silly useless value? *J Nucl Med* 1995;36:1836–9.
- Povoski SP, Murrey DAJr, Smith SM, et al. 18F-FDG PET/CT oncologic imaging at extended injection-to-scan acquisition time intervals derived from a single-institution 18F-FDG-directed surgery experience: feasibility and quantification of 18F-FDG accumulation within 18F-FDG-avid lesions and background tissues. *BMC Cancer* 2014;14:453.
- Zijlstra JM, Comans EF, van Lingena A, et al. FDG PET in lymphoma: the need for standardization of interpretation. An observer variation study. *Nucl Med Commun* 2007;28:798–803.
- Juweid ME, Stroobants S, Hoekstra OS, 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.
- Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. Early interim PET scans in diffuse large B-cell lymphoma: can there be consensus about standardized reporting, and can PET scans guide therapy choices? *Curr Hematol Malig Rep* 2012;7:193–9.
- Kong Y, Qu L, Li Y, et al. Predictive significance of a new prognostic score for patients with diffuse large B-cell lymphoma in the interim-positron emission tomography findings. *Medicine (Baltimore)* 2016;95:e2808.
- Terasawa T, Lau J, Bardet S, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. *J Clin Oncol* 2009;27:1906–14.
- Han HS, Escalon MP, Hsiao B, et al. 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.
- Mamot C, Klingbiel D, Hitz F, et al. Final results of a prospective evaluation of the predictive value of interim positron emission tomography in patients with diffuse large B-cell lymphoma treated with R-CHOP-14 (SAKK 38/07). *J Clin Oncol* 2015;33:2523–9.

- [35] Cashen AF, Dehdashti F, Luo J, et al. 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.
- [36] Adams HJ, Nievelstein RA, Kwee TC. Prognostic value of complete remission status at end-of-treatment FDG-PET in R-CHOP-treated diffuse large B-cell lymphoma: systematic review and meta-analysis. *Br J Haematol* 2015;170:185–91.
- [37] Fallanca F, Alongi P, Incerti E, et al. 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). *Eur J Nucl Med Mol Imaging* 2016;43:1837–48.
- [38] Coughlan M, Elstrom R. The use of FDG-PET in diffuse large B cell lymphoma (DLBCL): predicting outcome following first line therapy. *Cancer Imaging* 2014;14:34.
- [39] Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999;35:1773–82.
- [40] Frings V, van Velden FH, Velasquez LM, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology* 2014;273:539–48.
- [41] Jacene HA, Lebloulloux S, Baba S, et al. Assessment of interobserver reproducibility in quantitative 18F-FDG PET and CT measurements of tumor response to therapy. *J Nucl Med* 2009;50:1760–9.
- [42] Lin C, Itti E, Haioun C, et al. Early 18F-FDG PET for prediction of prognosis in patients with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med* 2007;48:1626–32.