Improvement of Early ^{18}F-FDG PET Interpretation in Diffuse Large B-Cell Lymphoma: Importance of the Reference Background

Emmanuel Itti[1,2], Malik E. Juweid[3], Corinne Haioun[4], Imene Yeddes[1], Fatma Hamza-Maaloul[1], Intidhar El Bez[1], Eva Evangelista[1], Chieh Lin[1,5], Jehan Dupuis[4], and Michel Meignan[1,2]

[1]Department of Nuclear Medicine, H. Mondor Hospital, Assistance Publique-Hôpitaux de Paris, Créteil, France; [2]EAC CNRS 7054, Paris 12 University, Créteil, France; [3]Division of Nuclear Medicine, University of Iowa, Iowa City, Iowa; [4]Department of Clinical Hematology, H. Mondor Hospital, Assistance Publique-Hôpitaux de Paris, Créteil, France; and [5]Department of Nuclear Medicine and Molecular Imaging Center, Chang Gung Memorial Hospital and University, Taoyuan, Taiwan

This study investigated whether the reference background above which a residual mass is considered positive in the International Harmonization Project criteria should be modified for early ^{18}F-FDG PET evaluation. **Methods:** In 92 patients with newly diagnosed diffuse large B-cell lymphoma, the maximal standardized uptake value (SUVmax) was measured on post–cycle 2 PET in the most intense residual mass (or, in the case of negative PET findings, in the area of most intense tumor uptake before therapy), in the mediastinal blood pool (MBP) and the liver, as potential reference background tissues. **Results:** With MBP as a reference (SUVmax, 2.0 ± 0.6), PET was unable to distinguish early responders from nonresponders. In contrast, with liver as a reference (SUVmax, 2.5 ± 0.7), 2-y progression-free survival was significantly different between patients with PET-negative findings (81.8% [95% confidence interval, 71%–93%]) and patients with PET-positive findings (51.8% [95% confidence interval, 35%–69%], P = 0.003). **Conclusion:** When assessing early response, particularly in risk-adapted therapeutic trials, it seems preferable to refer to a background tissue (liver) with a higher level of uptake than that of current international criteria (MBP) which were designed for end-of-treatment evaluation.

Key Words: PET; lymphoma; prognosis; response

PET with ^{18}F-FDG has become a powerful tool for identifying—as early as after 2 cycles of first-line chemotherapy—which patients with diffuse large B-cell lymphoma (*1,2*) and classic Hodgkin lymphoma (*3,4*) are responders. Consequently, many ongoing clinical trials focus on tailoring therapeutic strategies based on early PET as

an indicator of chemosensitivity (*5*). Some use the International Harmonization Project interpretation criteria (*6*), which were current at the time of the trial design but are adapted to end-of-treatment evaluation. In diffuse large B-cell lymphoma, the GELA LNH2007-3B study (*7*) is evaluating a chemotherapy escalation scheme based on interim PET findings of metabolically active residual masses. However, the central-review readers have observed an unexpectedly high rate of positive findings and substantial interobserver variability using these criteria (*8*).

The purpose of the present study was to investigate whether the reference background above which a residual mass is considered positive in the International Harmonization Project criteria should be modified for interim PET evaluation in hopes of improving the positive predictive value and overall accuracy of interim scans in diffuse large B-cell lymphoma.

MATERIALS AND METHODS

Patients

The population consisted of 92 patients with newly diagnosed diffuse large B-cell lymphoma (*9*) and previously enrolled in a prospective multicenter trial designed to assess the prognostic value of early PET after 2 cycles of induction chemotherapy (*2*). Patient characteristics and treatment regimens have been previously described (*9*). Importantly, the treatment strategy was planned at inclusion according to age, International Prognostic Index, and anthracycline-based protocols currently active at that time and was not influenced by PET results. The study was approved by our institutional review board, and all patients gave informed written consent.

PET

All patients underwent ^{18}F-FDG PET before the onset of chemotherapy (PET0) and after 2 cycles (PET2). Images were acquired on a dedicated C-PET camera (ADAC) for the first 81 patients (at Tenon Hospital) and on a Gemini PET/CT system (Philips) for the last 11 patients (at H. Mondor Hospital). The image acquisition and reconstruction parameters have been previously described (*9*). All patients also underwent concurrent diagnostic CT of the chest, abdomen, and pelvis within a week of each PET scan and then every 6 mo for follow-up based on the International Workshop Criteria (*10*). Outcome analysis was masked to the PET results.

Standardized Uptake Value (SUV) Analysis of ¹⁸F⁻FDG Uptake

Experience has shown that visual interpretation of images through comparison with a certain tissue background, such as the mediastinal blood pool (MBP), is prone to considerable interobserver variability but is substantially improved by using a simple semiquantitative approach in which the residual lesion is considered positive when its maximal SUV (SUVmax) exceeds that of reference background by 25% (*11*).

For each attenuation-corrected PET dataset, the tumor with the most intense uptake was carefully identified, relying on a graded color-scale, with red indicating the maximal count value. A volume of interest (VOI) encompassing the entire tumor was drawn to ensure correct identification of this maximum. On PET2 images, in cases of residual tumor, the VOI was drawn on the most intense focus even though its location differed from the most intense tumor on PET0. When no lesions were identifiable, the VOI was drawn on PET2 in the area of most intense tumor uptake before therapy, with careful slice-to-slice comparison with PET0. In addition, a VOI was drawn on 5 contiguous slices in the MBP (at the level of the aortic arch) and in the liver parenchyma (at mid height), making sure that VOI outlines were restricted to areas of physiologic uptake and avoiding neighboring sites of residual disease (Fig. 1). Mean SUV (SUVmean) and SUVmax were calculated in each VOI after proper calibration and normalization to body weight (*9*).

PET2 images were considered positive when SUVmax in the residual tumor or area of initial tumor before therapy was 125% of the particular reference backgrounds investigated: SUVmax or SUVmean of the MBP or the liver. A 125% threshold was thought to best represent the mere visual impression of "higher uptake" than the background level by observer consensus because this threshold resulted in similar accuracy while minimizing the inherent interobserver variability of visual interpretation (*11*).

Statistical Analysis

To evaluate the early prognostic value of PET2, progression-free survival (PFS) and overall survival (OS) were chosen as endpoints. PFS was defined as the interval from the date of enrollment to the first evidence of progression or relapse. OS was defined as the interval from the date of enrollment to death from any cause. Continuous variables were compared using the Student t test, and significance was obtained when the 2-sided P value was less than 0.05. Survival curves according to visual and SUV analyses were obtained using Kaplan–Meier plots and compared using the Log rank test. In addition, receiver-operating-characteristic analysis was performed to identify the optimal threshold for predicting PFS for each category of reference background.

RESULTS

Patient Outcome

During a median follow-up period of 42 mo after inclusion, 63 patients were free from progression (PFS, 68.5%) whereas the remaining 29 progressed or relapsed, with a median delay of 6.7 mo; in addition, 71 patients survived (OS, 77.2%) whereas the remaining 21 died, with a median delay of 9.1 mo. Thirteen deaths were due to lymphoma progression, 3 to cardiac events, 3 to infectious complications, 1 to treatment toxicity, and 1 to a car accident.

Predictive Values of MBP-Based Interpretations

There was no statistically significant difference between SUVs computed from the C-PET system and those obtained from the Gemini PET/CT system for tumor, MBP, or liver on PET0 ($P = 0.2$–0.6) or PET2 ($P = 0.1$–0.4, respectively). SUVs are displayed in Table 1. SUVmax averaged 13.2 ± 4.8 on PET0 and decreased to 3.4 ± 2.7 on PET2 (4.8 ± 4.2 in patients who progressed and 2.7 ± 1.2 in those who did not, $P = 0.0004$).

When MBP was taken as the reference background for positivity, with the >125% SUVmax criterion (which may be considered close to a "quantitative International Harmonization Project criterion"), PET2 was unable to distinguish early responders from nonresponders. The 2-y estimate for PFS was 76.0% (95% confidence interval [CI], 62%–90%) in the PET-negative patients, compared with 65.1% (95% CI, 51%–79%) in the PET-positive patients ($P = 0.24$, Fig. 2A). The accuracies of this approach for predicting PFS and OS were 52.2% and 56.5%, respectively (Table 2). Accuracies increased to 70.7% and 72.8%, respectively, when the optimal criterion, determined by receiver-operating-characteristic analysis, was >168% SUVmax. The 2-y estimate for PFS was 80.2% (95% CI, 70%–91%) in the PET-negative patients, compared with 48.1% (95% CI, 29%–67%) in the PET-positive patients ($P = 0.001$, Fig. 2C).

With SUVmean, positive predictive values (PPVs) and accuracies were always lower than those obtained with the corresponding SUVmax thresholds (>125% or receiver-operating-characteristic–optimized criterion), for PFS and for OS prediction.

Predictive Values of Liver-Based Interpretations

When liver activity was used as a reference with the >125% SUVmax criterion, PET2 was a strong predictor of survival. The 2-y estimate for PFS was 81.8% (95% CI, 71%–93%) in PET-negative patients, compared with 51.8% (95% CI, 35%–69%) in PET-positive patients
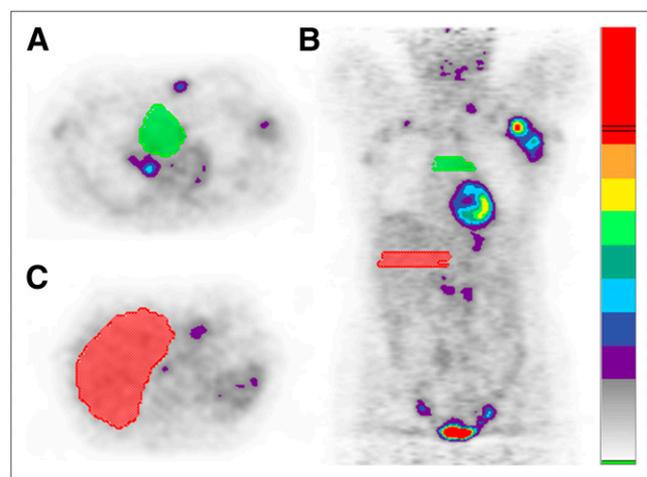

FIGURE 1. Example of VOI drawings in the 2 reference backgrounds: MBP, green outlines on axial slice (A), liver, red outlines on axial slice (C), and both on coronal slice (B).

TABLE 1
Mean and Maximal SUVs in Most Intense Tumor, MBP, and Liver at PET0 and PET2

Parameter	Tumor	MBP	Liver
PET0			
SUVmean	9.5 ± 3.6 (3.4–18.0)	1.3 ± 0.4 (0.5–2.4)	1.6 ± 0.3 (0.8–2.7)
SUVmax	13.2 ± 4.8 (4.8–24.8)	2.0 ± 0.5 (0.9–3.2)	2.4 ± 0.5 (1.7–3.7)
PET2			
SUVmean	3.4 ± 2.7 (0.9–20.8)	2.0 ± 0.6 (1.0–6.3)*	2.5 ± 0.7 (1.1–6.6)*
SUVmax	2.4 ± 2.1 (0.6–16.2)	1.4 ± 0.5 (0.6–4.8)*	1.7 ± 0.5 (0.8–4.8)*

*No statistically significant difference was found between PET0 and PET2.
Values are mean ± SD, with range in parentheses.

($P = 0.003$, Fig. 2B). The accuracies of this approach for predicting PFS and OS were 67.4% and 69.6%, respectively (Table 2). Accuracies increased to 72.8% and 75.0%, respectively, when the optimal criterion was >140% SUVmax. The 2-y estimate for PFS was 83.2% (95% CI, 73%–93%) in PET-negative patients, compared with 44.6% (95% CI, 26%–63%) in PET-positive patients ($P = 0.0001$, Fig. 2D).

Here again, with SUVmean, PPVs and accuracies were always lower than those obtained with the corresponding SUVmax threshold (>125% or receiver-operating-characteristic–optimized criterion), for PFS and for OS prediction.

Response Assessment with Conventional Diagnostic Methods

At the end of induction chemotherapy, 19 patients had a complete response according to CT, 50 had a complete response unconfirmed, 13 had a partial response, 2 had stable disease, 7 had progressive disease, and 1 did not undergo CT because the patient died during induction chemotherapy (Table 3).

DISCUSSION

The most predictive criterion proposed so far for early PET interpretation in diffuse large B-cell lymphoma was to consider as positive (i.e., predictive of relapse) a SUVmax
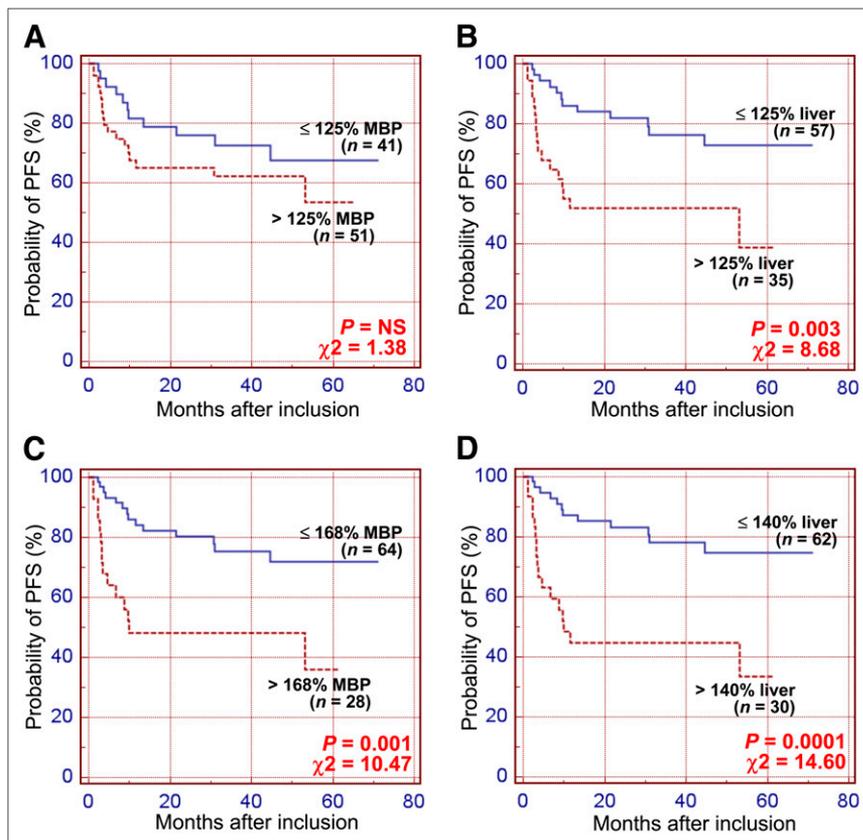

FIGURE 2. Kaplan–Meier estimates of PFS according to PET2 status: based on uptake > 125% of SUVmax in MBP (hazard ratio, 0.641; 95% CI, 0.311–1.340) (A), based on uptake > 125% of SUVmax in liver (hazard ratio, 0.348; 95% CI, 0.143–0.676) (B), based on uptake > 168% of SUVmax in MBP (hazard ratio, 0.322; 95% CI, 0.109–0.580) (C), and based on uptake > 140% of SUVmax in liver (hazard ratio, 0.264; 95% CI, 0.088–0.458) (D).

TABLE 2
Outcome Prediction Using Different Reference Backgrounds for PET Interpretation at 2 Cycles of Chemotherapy

Parameter	PFS (population $n = 92$, events $n = 29$)					OS (population $n = 92$, deaths $n = 21$)				
	Se	Sp	PPV	NPV	Acc	Se	Sp	PPV	NPV	Acc
MBP as reference background										
>125% SUVmean	82.8	11.1	30.0	58.3	33.7	90.5	14.1	23.8	83.3	31.5
>221% SUVmean	51.7	71.4	45.5	76.3	65.2	61.9	71.8	39.4	86.4	69.6
>125% SUVmax	62.1	47.6	35.3	73.2	52.2	76.2	50.7	31.4	87.8	56.5
>168% SUVmax	51.7	79.4	53.6	78.1	70.7	57.1	77.5	42.9	85.9	72.8
Liver as reference background										
>125% SUVmean	79.3	28.6	33.8	75.0	44.6	85.7	29.6	26.5	87.5	42.4
>201% SUVmean	55.2	76.2	51.6	78.7	69.6	61.9	74.6	41.9	86.9	71.7
>125% SUVmax	58.6	71.4	48.6	78.9	67.4	66.7	70.4	40.0	87.7	69.6
>140% SUVmax	58.6	79.4	56.7	80.6	72.8	66.7	77.5	46.7	88.7	75.0

Se = sensitivity; Sp = specificity; PPV = positive predictive value; NPV = negative predictive value; Acc = accuracy.
Data are percentages.

reduction of less than 66% between PET0 and PET2 in the most intense tumor (9). This approach dramatically reduced false-positive interpretations and, as such, is currently used in the PETAL study from the Essen group, Germany, to identify candidates for chemotherapy intensification (12). In the first 266 patients, with a median follow-up of less than 6 mo, lymphoma relapses across all International Prognostic Index risk groups have already occurred 3 times more frequently in PET2-positive than in PET2-negative patients, 23% (9/40) versus 8% (19/226), $P = 0.021$ (13). Importantly, this study will provide an external validation of the 66% SUVmax reduction criterion.

However, to be applicable, this criterion requires a baseline scan, which is not mandatory according to the International Harmonization Project criteria for response assessment of lymphoma and is, in fact, not routinely performed in most risk-adapted trials (6). An alternative approach investigated in the current study is to compare residual tumor uptake with an internal reference background, such as MBP or liver. SUV semiquantification was used to minimize subjectivity of PET2 interpretation as previously documented (11). Background intensities were stable between PET0 and PET2, indicating a rela-

tively low influence of treatment regimens on MBP and liver uptake of ^{18}F-FDG (Table 1). This finding is interesting, considering the observation that chemotherapy-induced steatosis and sinusoid obstruction affect liver imaging by CT, ultrasound, and MRI but apparently not, at least not to any significant extent, by ^{18}F-FDG PET (14). In addition, there was little variability in the SUVs of both the liver and MBP between patients. This has been previously shown for the liver (15) and is now also shown for the MBP.

Our data clearly demonstrate that the liver is a reference background superior to MBP for early PET interpretation in diffuse large B-cell lymphoma, even when using the conservative >125% SUVmax criterion, which should perform similarly to visual interpretation of liver activity but with lower interobserver variability. Indeed, it seems justified to use a higher level of reference background because, at 2 cycles, residual "tumor" uptake likely represents a significant component of posttherapy inflammatory changes at this relatively early time after treatment, with or without a minimal amount of viable tumor cells likely to be eradicated with further treatment (16,17). Corticosteroids have been shown to help reduce the contribution of inflammatory cells to the ^{18}F-FDG uptake and were included in our che-

TABLE 3
Prediction of Chemotherapy Response (End of Induction) Using Liver-Based Early PET Interpretation

Parameter	Cutoff = 125% SUVmax		Cutoff = 140% SUVmax	
	PET2-negative*	PET2-positive*	PET2-negative†	PET2-positive†
Complete response	15	4	15	4
Complete response unconfirmed	32	18	37	13
Partial response	9	4	9	4
Stable disease	0	2	0	2
Progressive disease	0	7	0	7
Not assessed (death during induction)	1	0	1	0

*$\chi^2 = 17.98$, $P = 0.003$.
†$\chi^2 = 21.25$, $P = 0.0007$.

motherapy regimens (*18*). With liver uptake as the reference background (>125% SUVmax), PPV reached 48.6% for PFS prediction, corresponding to a reduction of 15 false-positives, compared with MBP in the same setting. The difference between MBP and liver uptake was only 0.5 units in terms of SUVmax (2.0 ± 0.6 vs. 2.5 ± 0.7, respectively) but was sufficient to reduce false-positives. Interestingly, the >125% SUVmax liver criterion resulted in performance similar to our previously published custom visual analysis allowing minimal residual uptake (*9*).

Receiver-operating-characteristic analysis has identified optimal thresholds (e.g., >168% of MBP SUVmax, >140% of liver SUVmax) leading to higher predictive values and accuracies. Obviously, these figures cannot be transposed to a visual scale of interpretation, but they strengthen the message that liver appears a better reference organ than MBP for early PET interpretation. Although the positive predictive value of the >140% liver SUVmax criterion remains lower than those obtained by measuring the SUV reduction or by waiting 2 additional cycles, this criterion results in a better sensitivity and slightly higher negative predictive value (*9,19*). With the caveats of the somewhat different outcome measure (PFS vs. event-free survival) evaluated using the >140% liver SUVmax criterion, SUV reduction after 2 cycles and SUV reduction after 4 cycles, the difference in overall accuracy between these 3 approaches was relatively small (72.8%, 76.1%, and 77.5%, respectively). For OS prediction, the accuracies were 75.0%, 84.8%, and 80.0%, respectively, but here again the higher accuracy with the SUV reduction (due to higher PPV) was at the expense of substantially lower sensitivity (66.7% for the >140% liver SUVmax criterion vs. 52.4% and 52.9% for SUV reduction at 2 and 4 cycles, respectively).

The concept of minimal residual uptake has been used in many studies, but its definition varies. For Mikhaeel et al., minimal residual uptake was defined as a focus of low-grade uptake in an area of previously noted disease, likely to represent inflammation but for which small-volume malignancy could not be excluded (*20*). Therefore, minimal residual uptake was a "gray zone" with intermediate prognosis, such as the "complete remission unconfirmed" category of conventional assessment (*10*). The definition of minimal residual uptake was more stringent in Hodgkin lymphoma: for Gallamini et al., the definition was uptake equal to or slightly higher than the MBP, with an SUV between 2.0 and 3.5 (*3*). In this case, minimal residual uptake was considered a negative finding. Finally, an international consensus for early PET evaluation has recently been reached favoring the use of the London criteria, which emphasize comparison of residual uptake with liver uptake (*21*). The London criteria are currently being validated and have already shown moderate to fair interobserver agreement (*22,23*).

CONCLUSION

For assessment of early response, particularly in risk-adapted therapeutic trials, it seems preferable to refer to a background with a higher level of uptake (the liver) than that of the current international criteria (the MBP), which were designed for end-of-treatment evaluation. This approach warrants further validation in larger clinical trials.

REFERENCES

1. Mikhaeel NG, Hutchings M, Fields PA, O'Doherty MJ, 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;16:1514–1523.
2. Haioun C, Itti E, Rahmouni A, et al. [18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood.* 2005;106:1376–1381.
3. Gallamini A, Hutchings M, Rigacci L, et al. Early interim 2-[18F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *J Clin Oncol.* 2007;25:3746–3752.
4. 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;27:1906–1914.
5. Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. *J Nucl Med.* 2009;50(suppl 1):21S–30S.
6. 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;25:571–578.
7. ClinicalTrials.gov Web site. A Study of Two Associations of Rituximab and Chemotherapy, with a PET-Driven Strategy, in Lymphoma (LNH2007–3B). Available at: http://clinicaltrials.gov/ct2/show/NCT00498043. Accessed October 4, 2010.
8. Casasnovas O, Morschhauser F. The LNH07-3B study. 2nd International Workshop on Interim PET in Lymphoma. Available at: http://eitti.free.fr. Accessed October 14, 2010.
9. 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;48:1626–1632.
10. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol.* 1999;17:1244–1253.
11. Meignan M, Itti E, Bardet S, et al. Development and application of a real-time on-line blinded independent central review of interim PET scans to determine treatment allocation in lymphoma trials. *J Clin Oncol.* 2009;27:2739–2741.
12. ClinicalTrials.gov Web site. Positron Emission Tomography Guided Therapy of Aggressive Non-Hodgkin's Lymphomas (PETAL). Available at: http://clinicaltrials.gov/ct2/show/NCT00554164. Accessed October 4, 2010.
13. Hüttmann A, Müller S, Bockisch A, et al. Positron emission tomography guided therapy of aggressive non-Hodgkin's lymphomas (PETAL trial): correlation between clinical variables and PET parameters [abstract]. *Blood.* 2009;114:2695.
14. Robinson PJ. The effects of cancer chemotherapy on liver imaging. *Eur Radiol.* 2009;19:1752–1762.
15. Kanstrup IL, Klausen TL, Bojsen-Moller J, Magnusson P, Zerahn B. Variability and reproducibility of hepatic FDG uptake measured as SUV as well as tissue-to-blood background ratio using positron emission tomography in healthy humans. *Clin Physiol Funct Imaging.* 2009;29:108–113.
16. Spaepen K, Stroobants S, Dupont P, et al. [18F]FDG PET monitoring of tumour response to chemotherapy: does [18F]FDG uptake correlate with the viable tumour cell fraction? *Eur J Nucl Med Mol Imaging.* 2003;30:682–688.

17. Kasamon YL, Jones RJ, Wahl RL. Integrating PET and PET/CT into the risk-adapted therapy of lymphoma. *J Nucl Med.* 2007;48(suppl 1):19S–27S.

18. Brepoels L, Stroobants S, Vandenberghe P, et al. Effect of corticosteroids on ^{18}F-FDG uptake in tumor lesions after chemotherapy. *J Nucl Med.* 2007;48:390–397.

19. Itti E, Lin C, Dupuis J, et al. Prognostic value of interim ^{18}F-FDG PET in patients with diffuse large B-cell lymphoma: SUV-based assessment at 4 cycles of chemotherapy. *J Nucl Med.* 2009;50:527–533.

20. Mikhaeel NG, Timothy AR, O'Doherty MJ, Hain S, Maisey MN. 18-FDG-PET as a prognostic indicator in the treatment of aggressive Non-Hodgkin's lymphoma: comparison with CT. *Leuk Lymphoma.* 2000;39:543–553.

21. Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET scan in lymphoma. *Leuk Lymphoma.* 2009;50:1257–1260.

22. 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;115:775–777.

23. Barrington SF, Qian W, Somer EJ, 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;37:1824–1833.