



Interim PET-CT in the management of diffuse large B-cell lymphoma

Craig H. Moskowitz¹

¹Memorial Sloan-Kettering Cancer Center, New York, NY

Once the diagnosis of diffuse large B-cell lymphoma has been established, physicians and patients would like to know if a given treatment regimen is likely to succeed—if the patient can be cured or if at least a durable remission is achievable. This desire has led to efforts to use interim positron emission tomography (PET) scanning as part of risk-adapted therapeutic clinical trials. In general, these studies use a variable number of doxorubicin-based induction cycles with rituximab, followed by the interim PET. If the test is negative, treatment is continued, but if it is positive, therapy is changed to treatment that concludes with autologous stem cell transplantation. Results of studies for interim PET have yielded mixed and confusing results, with high negative predictive value but positive predictive value ranging from 20%-80%. To use interim PET scanning effectively, clinicians need simple (positive or negative) criteria that are easy to interpret, reproducible, and have a high positive and negative predictive value so that we can be certain that by not changing therapy if the test is positive, we are not doing the patient a disservice.

Introduction

In 2012, positron emission tomography (PET)-computed tomography (CT) with ¹⁸F-fluorescein di-β-D-galactopyranoside (¹⁸F-FDG; for the purposes of this review, the term PET-CT is used) is an important noninvasive diagnostic tool for management of patients with lymphoma, and most lymphoma clinical investigators are convinced that its use will have a tremendous impact on current National Comprehensive Cancer Network (NCCN) guideline recommendations.¹

PET-CT is the standard imaging modality for staging and determining remission status at the conclusion of therapy for the aggressive lymphomas. It is clearly superior to CT alone when attempting to identify active disease after therapy completion. In addition, PET-CT has led to the revision of response criteria, allowing the elimination of the complete remission/unconfirmed category.^{2,3}

Many clinicians order a routine, interim PET-CT (I-PET) after 2, 3, or 4 cycles of chemotherapy in diffuse large B-cell lymphoma (DLBCL) patients, although results are conflicting and many lymphoma experts believe that its use in this setting is investigational. With an early I-PET (after 1-2 cycles), response of the cells with the highest doubling time will be seen clearly and the clinician can identify early responders versus nonresponders. However, does the I-PET need to be negative this early? Do we really want to give patients only 50-100 mg/m² of the most active agent, doxorubicin? Early I-PET after 3 or 4 cycles clearly requires a committed definition of positive versus negative at this time point. A report such as “excellent response, minimal residual uptake in a small lymph node that could represent treated disease, and residual lymphoma cannot be excluded” is not acceptable. It is possible that lymphomas with high proliferation indices can start to regrow at this time point. Some advocate getting I-PET after cycles 2 and 4 of therapy, but who is going to pay for all of these imaging studies? And do we really want a patient to have 4 PET-CT scans in a 5-month time frame?

If PET-CT scans are to be helpful as an interim test, they need to be reproducible, with minimal intra- and interobserver variability, and

the results must be correlated with the desired clinical end point. I-PET must reliably identify patients who will fail and determine their needs to be considered effective alternate therapy. Lastly, outcome with early use of alternate therapy must be equivalent to or superior to the same therapy at time of treatment failure and, if not, why do the I-PET in the first place?

Finally, data from the pre-rituximab era are no longer applicable. Therefore, how to use I-PET scans effectively for potential risk-adapted therapy in patients receiving rituximab-based therapy remains problematic.

Definitions of a positive or negative PET-CT scan

Numerous definitions of a positive PET-CT exist, each with different advantages. The methods for interpreting PET-CT include visual assessment and semiquantitative and biopsy-proven methods.

Visual assessment

Five years ago, consensus criteria from the International Harmonization Project (IHP) were developed for interpreting posttreatment PET-CT (F-PET) scans.⁴ For each scan evaluated, a comparison was made visually between the baseline mediastinal blood pool activity and the lymph node sites of involvement. The investigators concluded that visual assessment alone was adequate for determining whether a scan was positive or negative at the conclusion of therapy, and quantitative or semiquantitative approaches (ie, standardized uptake value [SUV]) were not necessary. Separate recommendations were made for extranodal sites. The IHP criteria were developed to determine remission status posttherapy, not interim evaluation, although some studies have used these criteria with mixed results, as discussed below.

Another visual interpretation of the PET-CT scan uses a 5-point scale (ie, the Deauville criteria). Baseline and I-PET scans are scored according to uptake in sites initially involved by lymphoma as: (1) no uptake, (2) uptake ≤ mediastinum blood pool, (3) uptake ≤ liver, (4) moderately increased uptake > liver, or (5) markedly increased uptake > liver and/or new lesions. A score

Deauville Criteria

- **Score 1 no uptake**
- **Score 2 uptake \leq mediastinum**
- **Score 3 uptake $>$ mediastinum but \leq liver**
- **Score 4 uptake $>$ liver at any site**
- **Score 5 uptake $>$ liver and new sites of disease**
- **Score X: new areas of uptake unlikely to be related to lymphoma**

Figure 1. Deauville criteria.

of 1-3 was regarded as negative and 4 or 5 as positive (Figure 1). This scale has been used effectively in Hodgkin lymphoma (HL), but in DLBCL, the results once again are mixed.⁵⁻⁷

Semiquantitative method

The Groupe d'Etude des Lymphomes de l'Adulte (GELA) have led the effort to use a semiquantitative method to determine whether an I-PET scan should be considered positive or negative. In their opinion, the utility of I-PET is suspect based upon the heterogeneity of the visual criteria used and suboptimal interobserver reproducibility when interpreting PET-CT images on the basis of entirely visual criteria. Instead, their research strategy is based upon the differences of SUVs (δ SUV) from the initial scan to the interim scan.^{8,9} Unfortunately, the investigators observed that optimal cutoffs for δ SUV are dependent upon the timing of the imaging, and therefore their results may be difficult to reproduce. In addition, δ SUV may not be relevant for patients with low-level FDG avidity at baseline. Several of their reports using this technique have been described previously¹⁰ (Figure 2).

Biopsy-proven method

At Memorial Sloan-Kettering Cancer Center (MSKCC), we were concerned about the high rate of false-positive I-PET scans. Initially, we interpreted scans according to the IHP criteria, and the maximum SUV and δ SUV of the most FDG-avid lesions before treatment and at the interim scan were recorded. Our threshold for a positive scan was conservative and was considered abnormal if greater than the mediastinal blood pool activity. Our program is unique in that all patients with positive I-PET underwent biopsy of the FDG-positive site. Only if the I-PET was truly positive, as confirmed by histopathology showing active DLBCL, was therapy changed to a consolidative transplantation. If the I-PET was falsely positive, with biopsy not showing lymphoma, patients received 3 cycles of ICE alone (ie, the same treatment as given to patients with a negative I-PET).^{11,12} Although this strategy has merit, an open or core needle biopsy midtreatment is difficult outside of major academic centers and, as will be discussed, $> 80\%$ of these biopsies were negative, leading to many unnecessary procedures. We believe that nearly always changing therapy to a more aggressive approach should be based upon tissue confirmation of active DLBCL. If this premise is correct, then nuclear medicine physicians must reliably tell the clinician when a biopsy is warranted, not when activity on an I-PET is greater than mediastinal or liver uptake. The hematologist/oncologist can then change therapy if the biopsy is positive. However, it is critical that the clinician give as much information as possible to the nuclear medicine physician for an appropriate determination to be made for biopsy versus no biopsy (Figure 3).

Most importantly, clinicians need black-and-white criteria (positive or negative) that are simple to interpret, reproducible, and have a high positive and negative predictive value.

Selected I-PET studies using IHP criteria

In a prospective study reported by Cashen et al, patients with advanced-stage DLBCL were treated with R-CHOP-21 (rituximab + cyclophosphamide + hydroxydaunorubicin + vincristine + prednisone/prednisolone for 21 days) and I-PET was performed after cycle 2 or 3 and at the end of therapy (F-PET). Scans were interpreted in a manner similar to that described in the IHP criteria, but infradiaphragmatic disease was considered positive if greater than liver/spleen uptake and the maximum SUV of the most PET-avid lesions were recorded. At a median follow-up of 3 years, the positive predictive value of I-PET was 42% and the negative predictive value was 77%. Although the results were statistically significant, the investigators concluded that the high false-positive rate of the I-PET precluded risk-adapted therapy with R-CHOP-21. There was no difference in overall survival (OS) between patients with a positive or negative I-PET. Interestingly, the F-PET did predict for event-free survival (EFS), progression-free survival (PFS), and OS ($P < .001$).¹³

The Eastern Cooperative Oncology Group (ECOG) evaluated I-PET using both the IHP (slightly modified), which requires the intensity to be greater than average in the liver to be positive, and the 5-point scale criteria from a recent phase 2 advanced-stage DLBCL study. Patients underwent a baseline PET, which was repeated 2.5 weeks after 3 cycles of R-CHOP-21. Then there was a central PET review, and a fourth cycle of R-CHOP-21 was administered. Patients with a negative I-PET continued with R-CHOP-21 and, if positive, then therapy was changed to RICE (rituximab + ifosfamide + carboplatin + etoposide). Three external nuclear medicine experts then evaluated data on 38 patients. Unfortunately, but not unexpectedly, there was disagreement one-third of the time between the experts. Because the reproducibility was suboptimal, the investigators concluded that standardization of PET interpretation is critical before a change in patient management for DLBCL, based upon this interim test, can be part of standard practice.¹⁴

Δ SUV max for I-PET assessment

- **Baseline PET (PET0):**
 - SUV max in the most active lesion
- **Interim PET:**
 - If (+): SUV max in the most active lesion
 - If (-): SUV max in the area of PET0 tumor
- **Calculation of % of SUV max reduction**
(Lin et al *J. Nucl Med* 2007; 48: 1626-32)
- **Optimal cut-off determined by ROC**
 - 66% for Δ SUVmaxPET0-2
 - 72.9% for Δ SUVmaxPET0-4

Figure 2. δ -SUV assessment for I-PET assessment.

Filling out the Request for Interim FDG- PET

- How much information is necessary for the NM doc to help the oncologist determine risk?

Figure 3. Filling out the request for interim FDG-PET.

Selected I-PET studies using 5-point scale criteria

Pregno et al recently reported the use of I-PET as part of either a R-CHOP-21 (31 patients) or R-CHOP-14 (57 patients) chemotherapy program for 88 patients with DLBCL. Involved-field radiotherapy was also delivered if patients had bulky disease regardless of PET results. All patients had a baseline PET and the I-PET results were interpreted according to the 5-point score setting cutoff values of 1-3 as negative and 4 or 5 as positive. The median time of the I-PET was 13 days after chemotherapy cycle 2 (58 patients) or cycle 3 or 4 (30 patients). With a relatively short median follow-up of 2 years, OS and PFS were 91% and 77%, respectively. There was a statistically significant, but not clinically meaningful, difference in PFS based upon I-PET results: 85% for negative and 72% for positive patients. The positive predictive value of I-PET was only 36%. In that study, the F-PET also predicted 2-year PFS: 83% for negative and 64% for positive patients ($P = .001$), respectively.¹⁵

In 2011, Mikhael et al presented data at the international conference on I-PET¹⁶ on 3 sets of criteria for the interpretation of I-PET in a cohort of patients enrolled in the National Cancer Research Institute (NCRI) study of PET after 2 cycles in NHL. The 3 criteria were the 5-point scale, δ SUV ($< 66\%$ considered positive), and an internal NCRI study score using 5 categories: complete response, minimal residual uptake, partial response, stable disease, and progression. A total of 125 patients were evaluable at the time of the presentation. There were significant differences in the definition of good versus poor responders based upon the 3 scales: 57%, 36%, and 11% of patients were considered poor responders based upon the NCRI, 5-point scale, and δ SUV model, respectively. Furthermore, only 11 of 44 patients with positive scans using the 5-point scale (scores 4 or 5) were considered to be poor responders based upon δ SUV ($< 66\%$). It is clear from these data that the false-positive rate using the 5-point scale for DLBCL is unacceptably high.

Selected I-PET studies using δ SUV criteria

GELA has championed the semiquantitative method (δ SUV) to determine whether an I-PET scan should be considered positive or negative. Three of their reports are important to summarize.

Lin et al evaluated 92 patients with an I-PET after 2 cycles of therapy and calculated the δ SUV between the site of maximal pretreatment activity and the one at the time of the interim evaluation.¹⁰ They found that an optimal cutoff of a 65.7% SUVmax reduction yielded an accuracy of 76% to predict for EFS. Interestingly, they were able to reclassify 14 patients who were considered abnormal on visual analysis to a favorable group based upon a SUVmax reduction greater than 65.7%. Using this method, the number of false-positive scans was clearly reduced, and the 2-year EFS for patients with a favorable scan were 79% versus 21% for those with an unfavorable scan.¹⁰ However, only 16 of the 92 patients had an unfavorable scan and induction therapy was not standardized, making the interpretation of these results somewhat difficult.

In a second analysis of the above data, Itti et al determined that using the mediastinal blood pool as a benchmark for visual analyses is not

valid in DLBCL and that liver uptake is a better reference, although using this method, the false-positive I-PET rate still approaches 40%.¹⁷ Because many centers repeat an interim scan after 4 cycles of therapy, the second GELA report determined that the δ SUV value that should be used in this setting is 72.9%. However, this value was generated when induction therapy was R-ACVBP (doxorubicin + cyclophosphamide + vindesine + bleomycin + prednisone) and not R-CHOP-14 or R-CHOP-21; therefore, its applicability, especially for patients in North America, is unknown.

Recently, Casasnovas et al reported on a large study using either R-ACVBP or R-CHOP-14 induction.¹⁸ I-PET scans were done after cycles 2 and 4 and consolidation was dictated based upon central review for the results. In general, the results were excellent and the best predictor of a poor outcome was the I-PET 4 result.¹⁸

Selected I-PET studies using histologic confirmation of an abnormal interim scan

We previously reported on the use of a sequential treatment program R-CHOP-14 \times 4 followed by ICE \times 3, with an 80% 5-year PFS. I-PET did not predict outcome. Patients with a positive I-PET underwent biopsy (Figure 2), of which 85% were negative. PFS was identical for patients with I-PET-positive with a negative biopsy compared with I-PET-negative.¹¹

Other investigators have raised the issue of whether biopsy should be considered the gold standard to verify I-PET findings. Admittedly, there are several potential problems with biopsy.⁸ First, a false-negative biopsy is a possibility if the biopsy material is not obtained from the specific part of the node with persistent FDG uptake in a residual mass. Second, there may be methodological issues regarding how biopsies are analyzed; that is, qualitatively rather than quantitatively. Lastly, there needs to be a committed team of oncologists, interventional radiologists, and surgeons available to do the procedure in an expedited fashion. We used biopsy-proven persistent disease to justify treatment intensification and a conservative approach to identify residual FDG uptake and confirm its nature by biopsy. If we had accepted persistent FDG positivity as a true indicator of persistent disease, then 40% of patients would have undergone unnecessary treatment intensification. Nevertheless, if treatment-induced inflammation causing false-positive FDG uptake on interim scan could be avoided (eg, by slight modification of the therapeutic regimen without compromising patient outcome), it is conceivable that some SUV-related parameters might eventually emerge to distinguish metabolic responders from nonresponders with reasonable certainty.

In our current study, the induction treatment was altered in an attempt to reduce the rate of the false-positive I-PET. We also augmented consolidation for patients with a proliferative index $\geq 80\%$. Eligible patients were < 70 years of age with advanced-stage DLBCL or primary mediastinal large B-cell lymphoma. Pretreatment evaluation included CT with contrast, PET-CT, and ¹⁸F-fluorothymidine-PET scan. Induction consisted of R-CHOP-14 \times 3 and CHOP-21 \times 1 followed 17-20 days later by an I-PET; a

biopsy was done if the I-PET was positive. Consolidation was risk adapted, as with our last study, in which biopsy-proven positive disease was consolidated with high-dose therapy/autologous stem cell rescue. Use of ^{18}F fluorothymidine-PET was exploratory. Data on the first 50 patients were reported at the Lugano meetings in June 2011.¹⁹ At median follow-up of 18 months, the PFS and OS are 86% and 96%, respectively. Despite making changes to induction and only doing a biopsy on patients with scores of 4 or 5 in the 5-point scale, as in our previous study, an I-PET-positive scan was not correlated with an inferior PFS compared with those with a negative scan ($P = .27$). In addition, patients with negative interim biopsy had the same PFS as those with a negative I-PET ($P = .91$). Interestingly, F-PET predicted for PFS: 4 of 6 patients with positive scans have progressed ($P < .001$). This study is ongoing; however, we have already observed that despite altering the chemotherapy schedule and delaying interim restaging by 1 week, I-PET scans did not predict PFS. Further analysis of δSUV is planned and may be valuable.

Summary

FDG uptake in residual masses is determined by, among other factors, the number of viable tumor cells, number of inflammatory cells, and the degree of FDG retention. The contribution by each of these factors may vary with disease (eg, HL versus NHL), disease site (eg, mediastinum versus abdomen), and the treatment regimen. Therefore, many questions need to be addressed before any uniform number or cutoff parameters could be accepted in prospective trials. There are 2 large trials in progress: the US intergroup study (NCT00118209) is prospectively evaluating I-PET using visual criteria but not altering therapy based upon the result, and the PETAL study (NCT00554164) is using δSUV to change therapy to an acute lymphoblastic leukemia-type consolidation if the I-PET is deemed positive.

It is clearly not reasonable to subject 40% of DLBCL patients to open biopsies based upon the 5-point scale, so this scale seems to have limited utility for interpretation of I-PET in DLBCL. Various investigators have proposed specific SUV numbers or changes in SUV to differentiate between metabolic responders and nonresponders on I-PET in HL and NHL trials. However, there is still a significant false-positive rate when using δSUV .

What makes the most sense for the clinician? The oncologist must give the nuclear medicine physician the proper history and treatment administered (Figure 1). We need nuclear medicine physicians to reliably tell us that the FDG-PET is clearly abnormal and a biopsy is required. For this to happen, there needs to be standardization of reporting, specifically regarding criteria about what is positive and negative, as well as reproducibility among nuclear medicine physicians and machines. This will decrease the number of procedures and improve predictive value. For now, changing therapy based upon I-PET for patients with DLBCL should remain investigational.

Disclosures

Conflict-of-interest disclosure: The author is on the board of directors or an advisory committee for Genentech and Seattle Genetics; has received research funding from Seattle Genetics and Cephalon; and has been affiliated with the speakers' bureau for Genentech. Off-label drug use: None disclosed.

Correspondence

Craig H. Moskowitz, MD, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, New York, NY 10065; Phone: 212-639-2696; Fax: 646-422-2164; e-mail: moskowic@mskcc.org.

References

- Zelenetz AD, Abramson JS, Advani RH, et al. NCCN Clinical Practice Guidelines in Oncology: non-Hodgkin's Lymphomas. *J Natl Compr Canc Netw*. 2010;8(3):288-334.
- Cheson BD. Role of functional imaging in the management of lymphoma. *J Clin Oncol*. 2011;29(14):1844-1854.
- Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol*. 2007;25(5):579-586.
- 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(5):571-578.
- Barrington SF, Mackewn JE, Schleyer P, et al. Establishment of a UK-wide network to facilitate the acquisition of quality assured FDG-PET data for clinical trials in lymphoma. *Ann Oncol*. 2011;22(3):739-745.
- 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(10):1824-1833.
- Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. *J Nucl Med*. 2009;50(suppl 1):21S-30S.
- Meignan M, Gallamini A, Itti E, Barrington S, Haioun C, Polliack A. Report on the Third International Workshop on Interim Positron Emission Tomography in Lymphoma held in Menton, France, 26-27 September 2011 and Menton 2011 consensus. *Leuk Lymphoma*. 2012;53(10):1876-1881.
- Meignan M, Itti E, Gallamini A, Haioun C. Interim 18F-fluorodeoxyglucose positron emission tomography in diffuse large B-cell lymphoma: qualitative or quantitative interpretation—where do we stand? *Leuk Lymphoma*. 2009;50(11):1753-1756.
- 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(10):1626-1632.
- Moskowitz CH, Schoder H, Teruya-Feldstein J, et al. Risk-adapted dose-dense immunochemotherapy determined by interim FDG-PET in Advanced-stage diffuse large B-Cell lymphoma. *J Clin Oncol*. 2010;28(11):1896-1903.
- Moskowitz CH, Zelenetz A, Schoder H. An update on the role of interim restaging FDG-PET in patients with diffuse large B-cell lymphoma and Hodgkin lymphoma. *J Natl Compr Canc Netw*. 2010;8(3):347-352.
- Cashen AF, Dehdashti F, Luo J, Homb A, Siegel BA, Bartlett NL. 18F-FDG PET/CT for early response assessment in diffuse large B-cell lymphoma: poor predictive value of international harmonization project interpretation. *J Nucl Med*. 2011;52(3):386-392.
- 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(4):775-777; quiz 918.
- Pregno P, Chiappella A, Bello M, et al. Interim 18-FDG-PET/CT failed to predict the outcome in diffuse large B-cell lymphoma patients treated at the diagnosis with rituximab-CHOP. *Blood*. 2012;119(9):2066-2073.
- Mikhael NG, O'Doherty MJ, Barrington S. Comparison of quantitative and qualitative response criteria in the UK

- NCRI PET study in DLBCL: An interim analysis. International Workshop on Interim PET in Lymphoma. Menton, France; 2011.
17. Itti E, Juweid ME, Haioun C, et al. Improvement of early 18F-FDG PET interpretation in diffuse large B-cell lymphoma: importance of the reference background. *J Nucl Med.* 2010; 51(12):1857-1862.
 18. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. SUVmax reduction improves early prognosis value of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood.* 2011;118(1):37-43.
 19. Moskowitz CH. Interim FDG-PET scanning for DLBCL should remain investigational. International Conference on Malignant Lymphoma. Lugano, Switzerland; 2011.