^{18}F-FDG PET in Non-Hodgkin's Lymphoma: Qualitative or Quantitative?

Several studies have documented that ^{18}F-FDG PET is more accurate than CT for assessment of tumor response to chemotherapy in patients with Hodgkin's lymphoma (HL) and high-grade (aggressive) non-Hodgkin's lymphoma (NHL) (*1,2*). On the basis of these data, the International Workshop Criteria for assessment of tumor response in NHL have recently been revised to include ^{18}F-FDG PET (*3*). A complete response (CR) is now based primarily on the findings on ^{18}F-FDG PET. Patients with a residual mass of any size on CT are considered to be in CR as long as the mass is negative on ^{18}F-FDG PET. Visual assessment has been found to be sufficient to assess tumor response on ^{18}F-FDG PET after completion of therapy. To standardize interpretation of ^{18}F-FDG PET scans and to reduce the frequency of false-positive findings, the intensity of tumor ^{18}F-FDG uptake is compared with mediastinal blood pool (*4*). Regardless of their location, lesions are considered as "PET positive" only when their intensity is higher than that of mediastinal blood pool. An exception is made for lesions with a diameter of <2 cm. Because partial-volume effects in lesions of this size cause ^{18}F-FDG uptake to be underestimated on PET, any focal ^{18}F-FDG uptake above background is considered positive. In addition to these guidelines for image interpretation, the timing of ^{18}F-FDG PET scans has been standardized. It is recommended that ^{18}F-FDG PET be performed not earlier than 3 wk after completion of chemotherapy or chemoimmunotherapy and not earlier than 8–12 wk after completion of radiotherapy or chemoradiotherapy (*4*).

See page 1626

Although the use of ^{18}F-FDG PET for assessment of tumor response after completion of therapy is well established in HL and aggressive NHL, more recent studies have indicated that ^{18}F-FDG PET may also allow prediction of tumor response and patient outcome early in the course of therapy—that is, after 1 or 2 cycles of chemotherapy (*5–12*). The ability to predict tumor response early in the course of therapy would be very valuable clinically, as it would allow intensification of treatment in patients who are unlikely to respond to first-line chemotherapy. Conversely, treatment could potentially be shortened in patients who show a favorable response after 1 or 2 cycles of chemotherapy. This is of particular interest in HL, as chemotherapy combined with radiotherapy can cure most of the patients with HL, but it also puts them at increased risk for secondary malignancies and other serious long-term complications, such as infertility and cardiopulmonary toxicity. Therefore, there is great interest in using an early tumor response on ^{18}F-FDG PET for risk-adapted therapy of HL and aggressive NHL (*13,14*).

Several studies have now indicated that patients with HL or aggressive NHL and a "negative" PET scan after 2 cycles of chemotherapy are much more likely to remain event-free than patients with a "positive" scan. For example, Hutchings et al. (*11*) found that 2-y progression-free survival of patients with HL and a positive PET scan after 2 cycles of chemotherapy was 0%, whereas it was 96% in patients with a negative scan ($P < 0.001$). In aggressive NHL, Spaepen et al. (*6*) reported a 2-y progression-free survival of 4% in patients with a positive scan as compared with 85% in patients with a negative scan ($P < 0.001$). These data indicate that, in the hands of experienced investigators, visual assessment of tumor response provides a remarkably high accuracy for prediction of patient outcome.

However, visual assessment of tumor response to therapy is not without problems. In the studies reported so far, the positive predictive value (PPV) and negative predictive value (NPV) of early ^{18}F-FDG PET for progression-free survival are quite variable. Data for PPVs range from 44% to 100% (*5–12,15*). Conversely, NPVs ranging from 50% to 100% have been reported. Part of this wide variability is explained by differences in the patient populations. For example, in aggressive NHL, Haioun et al. (*9*) reported that the PPV of ^{18}F-FDG PET for tumor progression is only 44%, whereas the PPV was 71% in a study by Mikhaeel et al. (*15*). However, this difference in the PPV is likely explained by the overall frequency of tumor progression. In the study by Haioun et al., only 23% of the patients experienced tumor progression during the observation period, whereas the frequency of tumor progression was 40% (*15*) in the study by Mikhaeel et al. This almost 2 times higher pretest probability of recurrence resulted in the higher PPV of ^{18}F-FDG PET in the study by Mikhaeel et al., as sensitivity and specificity of ^{18}F-FDG PET for

Received Jul. 1, 2007; revision accepted Jul. 9, 2007.

For correspondence or reprints contact: Wolfgang A. Weber, MD, Abteilung Nuklearmedizin, University of Freiburg, Hugstetterstrasse 55, 79106 Freiburg, Germany.
E-mail: wolfgang.weber@uniklinik-freiburg.de

COPYRIGHT © 2007 by the Society of Nuclear Medicine, Inc.

DOI: 10.2967/jnumed.107.042911

prediction of tumor progression were quite comparable in both studies.

Differences in the studied patient populations explain some, but not all, of the variability of the predictive value of early ^{18}F-FDG PET. Sensitivity and specificity for prediction of progression-free survival also show considerable variability, although these 2 parameters are not dependent on the pretest probability of disease progression. For example, Torizuka et al. (*8*) reported a sensitivity and specificity of ^{18}F-FDG PET of 87% and 50%, respectively. In contrast, Jerusalem et al. (*5*) found a sensitivity of only 42% but a specificity of 100%. This suggests that criteria for interpreting PET scans as positive or negative might have varied among different studies. Most articles describe only briefly how scans were interpreted, but it is clear that some studies used any abnormal ^{18}F-FDG uptake as a criterion for a positive scan, whereas others compared the intensity of uptake with normal tissues and considered only lesions with a certain degree of uptake as positive. All of these issues raise concerns with regard to the interobserver variability of response assessment, especially when PET scans are read by physicians with limited experience in PET.

The International Workshop Criteria (*4*) now provide clear guidelines for scan interpretation, but these guidelines were developed for interpreting ^{18}F-FDG PET scans after completion of therapy—not after 1 or 2 cycles of chemotherapy. Accordingly, they basically describe normalization or almost complete normalization of the ^{18}F-FDG PET scan. However, one might expect that at least some patients with a negative scan after completion of chemotherapy still show focal ^{18}F-FDG uptake after 1 or 2 cycles of chemotherapy that resolves with continued therapy (typically, 4 or 5 additional cycles of chemotherapy). In fact, studies have shown that patients with only mildly positive lesions after 2 cycles of chemotherapy have a significantly better prognosis than patients with more intensely hypermetabolic lesions (*11,15*). Defining this "mild" or "moderate" ^{18}F-FDG uptake by visual assessment alone is difficult. Of course, one can visually compare the intensity of tumor ^{18}F-FDG uptake with various normal tissues, such as liver or brain, but there is no reason to assume that the intensity of ^{18}F-FDG uptake by these organs should provide the optimum threshold value for differentiation of patients with a favorable and unfavorable outcome.

Because of these limitations of visual assessment of tumor response, there is considerable interest in using quantitative parameters for monitoring response to therapy in lymphoma patients (*16*). Quantification of tumor ^{18}F-FDG uptake has the potential to provide less observer-dependent and less arbitrary criteria for a positive PET scan. In addition, it could capture more of the prognostic information of tumor metabolism, because ^{18}F-FDG uptake is measured as a continuous parameter and not as a binary (positive or negative) variable (*14*). Several studies have measured maximum standardized uptake values (SUVs) after 2 cycles of chemotherapy and found that the intensity of tumor ^{18}F-FDG uptake at this time is correlated with patient outcome (*8,11*). However, there are also concerns that tumor SUVs may be confounded by multiple factors, such as lesion size, image reconstruction, start of data acquisition, and so forth. (*17*). Although this is certainly correct, one should keep in mind that most of these factors will also affect visual assessment of tumor response. Because of partial-volume effects, smaller lesions will not only show lower SUVs on quantitative analysis but also appear less intense on visual assessment. Contrast between tumor and background will be lower when images are smoothed during reconstruction, whereas contrast will be higher when images are acquired later, because ^{18}F-FDG is cleared from the blood pool but retained by lymphoma cells. This will affect not only the measured SUVs but also visual assessment of ^{18}F-FDG uptake. Quantification of ^{18}F-FDG uptake by SUVs does, however, require some additional steps, such as a cross-calibration of the dose calibrator and the PET scanner as well as accurate measurement and decay correction of the injected dose (*17*).

The study by Lin et al. (*18*) provides interesting data with regard to the usefulness of quantitative parameters for monitoring tumor response in patients with aggressive NHL. A strength of the study is the inclusion of a homogeneous and rather large group of patients with diffuse large B-cell lymphoma (DLBCL). Ninety-two patients with DLBCL were studied before and after 2 cycles of chemotherapy. Tumor response was assessed visually and by various quantitative parameters. The authors then addressed the following questions: (a) Do quantitative parameters provide additional information to visual assessment of tumor response? (b) Which quantitative parameter should be used (maximum SUV, mean SUV, or a tumor-to-background ratio)? (c) Should tumor response be assessed by measuring SUV changes (from the baseline to the follow-up scan) or by measuring SUV in the follow-up scan?

Lin et al. (*18*) find that quantification of tumor ^{18}F-FDG uptake can markedly improve the accuracy of ^{18}F-FDG PET for prediction of patient outcome. On visual assessment, ^{18}F-FDG PET scans were classified as positive in 34 patients. However, 17 of these patients remained event free during follow-up. Thus, the PPV of a positive ^{18}F-FDG PET scan for disease progression or relapse was only 50%. The NPV was 74%. When the assessment of tumor response was based on a quantitative analysis of ^{18}F-FDG uptake, the PPV increased up to 92%, without a change in the NPV. The optimum threshold value for prediction of event-free survival was a maximum SUV of 5.0. This is considerably higher than the typical ^{18}F-FDG uptake of liver (SUV of 2–3) and indicates that a subgroup of patients with clearly positive PET scans after 2 cycles of chemotherapy is characterized by a favorable prognosis after continued chemotherapy. It would be difficult to identify this subgroup by visual assessment, as this would require estimation of whether tumor ^{18}F-FDG uptake is 1.67- to 2.5-fold higher than liver ^{18}F-FDG uptake.

All of the studied quantitative parameters demonstrated a similar accuracy

for prediction of progression-free survival, although there was a trend for a slightly lower accuracy for the tumor-to-muscle ratio. This might be related to the fact that dividing tumor ^{18}F-FDG uptake by muscle ^{18}F-FDG uptake increases statistical noise, as muscle ^{18}F-FDG uptake is low in fasted patients (SUV < 1). Normalization of SUVs to body surface area did not improve the accuracy of ^{18}F-FDG PET, likely indicating that there were no major changes in patient body weight at the time of the follow-up PET scan. Maximum tumor SUV and mean SUV performed equally well for prediction of event-free survival. This is encouraging because maximum SUVs are easier to determine and less operator dependent than mean SUVs.

Relative changes in tumor SUV (from the pretherapeutic scan to the scan after 2 cycles of chemotherapy) predicted overall survival with an accuracy similar to that of tumor SUVs measured on the scan after 2 cycles of chemotherapy. This is an important observation because SUV changes are less dependent on differences in image reconstruction and postprocessing than absolute SUVs. Therefore, SUV changes represent a more robust parameter for multicenter trials involving PET scanners from different manufacturers (*19*).

Finally, it is noteworthy that the authors used a straightforward approach to select lesions for quantitative analysis. ^{18}F-FDG uptake was measured for the lesion with the highest ^{18}F-FDG uptake in the baseline scan and the lesion with the highest ^{18}F-FDG uptake in the follow-up scan. In some cases this meant that a different lesion was analyzed in the follow-up than in the baseline scan. It is encouraging that this simple approach, which can be integrated easily into clinical practice, provided such a high accuracy for prediction of progression-free survival. Nevertheless, it will be interesting to evaluate in future studies whether the accuracy of ^{18}F-FDG PET can be further improved by averaging ^{18}F-FDG uptake over several lesions.

In conclusion, the study by Lin et al. (*18*) indicates that quantitative analysis of tumor ^{18}F-FDG uptake can be used to monitor tumor response to chemotherapy in patients with DLBCL. Furthermore, the study suggests that quantification of tumor ^{18}F-FDG uptake has the potential to improve the PPV of ^{18}F-FDG PET for disease progression. However, it is important to note that the quantitative criteria for monitoring tumor response were not defined prospectively by Lin et al. but were derived from the studied patient population. Therefore, this study may overestimate the accuracy of quantitative assessment of tumor ^{18}F-FDG uptake for prediction of progression-free survival. Thus, future investigations are necessary to prospectively validate the quantitative response criteria identified in this study. Eventually, monitoring of tumor response should probably include a combination of quantitative analysis and visual assessment. Although objective criteria for tumor response could be based primarily on quantitative parameters, visual assessment would still be critical for diagnosis of tumor progression (development of new lesions) and for differentiation of lymphoma from benign causes of focal ^{18}F-FDG uptake, such as thymus hyperplasia. Visual assessment will also be important to check the plausibility of quantitative measurements—for example, by comparing the measured ^{18}F-FDG uptake of a lesion with the known ^{18}F-FDG uptake of normal organs. Thus, ^{18}F-FDG PET in lymphoma will likely become qualitative and quantitative.

Wolfgang A. Weber
University of Freiburg
Freiburg, Germany

REFERENCES

1. Kirby AM, Mikhaeel NG. The role of FDG PET in the management of lymphoma: what is the evidence base? *Nucl Med Commun.* 2007;28:335–354.
2. Zijlstra JM, Lindauer-van der Werf G, Hoekstra OS, Hooft L, Riphagen II, Huijgens PC. ^{18}F-Fluoro-deoxyglucose positron emission tomography for post-treatment evaluation of malignant lymphoma: a systematic review. *Haematologica.* 2006;91:522–529.
3. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007;25:579–586.
4. 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.
5. Jerusalem G, Beguin Y, Fassotte MF, et al. Persistent tumor ^{18}F-FDG uptake after a few cycles of polychemotherapy is predictive of treatment failure in non-Hodgkin's lymphoma. *Haematologica.* 2000;85:613–618.
6. Spaepen K, Stroobants S, Dupont P, et al. Early restaging positron emission tomography with ^{18}F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin's lymphoma. *Ann Oncol.* 2002;13:1356–1363.
7. Zijlstra JM, Hoekstra OS, Raijmakers PG, et al. 18FDG positron emission tomography versus ^{67}Ga scintigraphy as prognostic test during chemotherapy for non-Hodgkin's lymphoma. *Br J Haematol.* 2003;123:454–462.
8. Torizuka T, Nakamura F, Kanno T, et al. Early therapy monitoring with FDG-PET in aggressive non-Hodgkin's lymphoma and Hodgkin's lymphoma. *Eur J Nucl Med Mol Imaging.* 2004;31:22–28.
9. Haioun C, Itti E, Rahmouni A, et al. [^{18}F]Fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood.* 2005;106:1376–1381.
10. Gallamini A, Rigacci L, Merli F, et al. The predictive value of positron emission tomography scanning performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. *Haematologica.* 2006;91:475–481.
11. Hutchings M, Loft A, Hansen M, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood.* 2006;107:52–59.
12. Kostakoglu L, Goldsmith SJ, Leonard JP, et al. FDG-PET after 1 cycle of therapy predicts outcome in diffuse large cell lymphoma and classic Hodgkin disease. *Cancer.* 2006;107:2678–2687.
13. Dann EJ, Bar-Shalom R, Tamir A, et al. Risk-adapted BEACOPP regimen can reduce the cumulative dose of chemotherapy for standard and high-risk Hodgkin lymphoma with no impairment of outcome. *Blood.* 2007;109:905–909.
14. 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.
15. 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.
16. Kelloff GJ, Sullivan DM, Wilson W, et al. FDG-PET lymphoma demonstration project invitational workshop. *Acad Radiol.* 2007;14:330–339.
17. Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med.* 2005;46:983–995.
18. Lin C, Itti E, Haioun C, et al. Early ^{18}F-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.
19. Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging.* 2005;32:294–301.