

^{18}F -FDG PET as a Routine Test for Posttherapy Assessment of Hodgkin's Disease and Aggressive Non-Hodgkin's Lymphoma: Where Is the Evidence?

The recently published 2 reports of the International Harmonization Project (IHP) in Lymphoma recommended the routine use of ^{18}F -FDG PET for posttherapy assessment of patients with Hodgkin's disease (HD) and diffuse large B-cell lymphoma, the most common and potentially curable form of aggressive lymphoma (1,2). Data available to the IHP subcommittees seemed compelling for making this recommendation. A particularly useful document critical to the decision was the systematic review by Zijlstra et al. on ^{18}F -FDG

See page 13

PET in posttherapy evaluation of HD and, mostly, aggressive non-Hodgkin's lymphoma (NHL), similar to the topic of the systematic review by Terasawa et al. reported in the current issue of *The Journal of Nuclear Medicine* (3,4). Based on the metaanalysis by Zijlstra et al., pooled sensitivity and specificity of ^{18}F -FDG PET for detection of residual disease after completion of first-line therapy were 84% (95% confidence interval, 71%–92%) and 90% (95% confidence interval, 84%–94%), respectively, for HD and 72% (95% confidence interval, 61%–82%) and 100% (95% confidence interval, 97%–100%), respectively, for NHL (3).

Received Oct. 4, 2007; revision accepted Oct. 12, 2007.

For correspondence or reprints contact: Malik E. Juweid, MD, University of Iowa, 200 Hawkins Dr., Iowa City, IA 52242-1077.

E-mail: malik-juweid@uiowa.edu

COPYRIGHT © 2008 by the Society of Nuclear Medicine, Inc.

DOI: 10.2967/jnumed.107.046292

Although clearly acknowledging the review by Zijlstra et al., Terasawa's report highlights some of its shortcomings. Perhaps the most justifiable critique pertains to lack of consideration of major methodologic variability between the original studies included in the Zijlstra review, such as inclusion of patients with the incurable indolent non-Hodgkin's lymphoma (NHL) rather than limiting of the analysis to patients with aggressive lymphoma or, better yet, diffuse large B-cell lymphoma in some studies assessing the accuracy of posttherapy ^{18}F -FDG PET in NHL (3,4). The combined analysis of post-salvage and post-first-line-therapy ^{18}F -FDG PET scans included in a few studies represents another flaw of the Zijlstra review (3,4). Admittedly, the systematic review by Zijlstra's group does not seem to exhibit the degree of attention to detail apparent in the review of Terasawa et al., who thoroughly interrogated the data reported in the original studies and made every effort to obtain specific data pertaining to the subsets of patients of interest to this analysis (i.e., HD and aggressive NHL after first-line chemotherapy). This effort was particularly evident by their contacting the authors of the various original reports whenever necessary to obtain such data (4). Another strength of the current review is the conducting of separate analyses for ^{18}F -FDG PET use for "residual mass" versus "posttherapy" evaluation, irrespective of the presence of residual masses on conventional imaging (i.e., CT or MRI). Although there are still a few, generally minor, deficiencies in the systematic

review of Terasawa et al., it might be more useful to focus this perspective on common conclusions that can be drawn from both reviews and determine whether these conclusions can support the routine use of ^{18}F -FDG PET at the completion of treatment in patients with HD and aggressive NHL.

Despite the remarkable heterogeneity and suboptimal methodologic quality of the included original studies in both systematic reviews, in aggregate these studies appear to show only a moderate positive predictive value (PPV) for ^{18}F -FDG PET in posttherapy evaluation of HD. Zijlstra's group reported a range of 60%–100% for PPV in 5 studies exclusively assessing ^{18}F -FDG PET in posttherapy evaluation of HD (3). The 10 posttherapy evaluations of HD included in the Terasawa review exhibit a wider range of PPVs (13%–100%), with a "weighted average" of 62% and all but one study showing a PPV of at least 50% (Table 1). On the other hand, both reviews show a very high, somewhat less variable, negative predictive value (NPV) for ^{18}F -FDG PET in posttherapy evaluation of HD, with the NPV ranging from 84% to 100% in the 5 studies reported by Zijlstra et al. and from 71% to 100% in the 10 studies included by Terasawa et al., with a weighted average of 94% for the latter studies (Table 1).

What are the implications of these findings on the overall diagnostic accuracy of ^{18}F -FDG PET and its utility as a routine test in the posttreatment evaluation of patients with HD? The implications might be clearer after one estimates the rate of ^{18}F -FDG PET

TABLE 1

Rate of ¹⁸F-FDG PET Positivity and Negativity, PPV, and NPV in Original Studies Included in the Terasawa Review on Posttherapy Evaluation of Patients with HD and Aggressive NHL

Study and Year*	Total no. of patients	% PET-positive	% PET-negative	PPV (%)	NPV (%)
HD (n = 399)					
Filmont et al., 2004	32	44	56	78	100
Friedberg et al., 2004	32	25	75	50	96
Guay et al., 2003	48	25	75	92	92
Jerusalem et al., 1999; 2003	31	13	87	100	92
Mikosch et al., 2003	31	61	39	89	100
Mocikova et al., 2004	71	32	68	13	100
Rigacci et al., 2005	28	29	71	50	100
Schaefer et al., 2004	18	22	78	100	71
Spaepen et al., 2001	60	8	92	100	91
Wickmann et al., 2003	48	52	48	60	91
Aggressive NHL (n = 236)					
Jerusalem et al., 1999	35	14	86	100	77
Juweid et al., 2005	54	35	65	74	83
Mikhaeel et al., 2000	45	20	80	100	83
Mikosch et al., 2003	24	50	50	83	75
Schaefer et al., 2004	5	20	80	100	50
Spaepen et al., 2001	73	30	70	100	82

*Please refer to Terasawa et al. (3) for complete citations.

scans with positive and negative findings in the various studies included in both reviews: The rate of ¹⁸F-FDG PET scans with positive findings after treatment in the 10 studies included in the Terasawa review ranges from 8% to 61%, with a weighted average of approximately 30% and all but 2 studies having a positive scan rate of between 22% and 52% (Table 1). Expectedly, the rate of ¹⁸F-FDG PET scans with negative findings in these studies is about 70%, with a range of 39%–92%; all but 2 studies had a negative scan rate of between 48% and 78%. Similar findings are noted in the report by Zijlstra et al. (3).

With a positive ¹⁸F-FDG PET scan rate of about 30% and a PPV of 62%, it is easy to calculate that misclassification of disease status because of positive posttherapy ¹⁸F-FDG PET findings would, on average, affect only about 11% of all patients (38% of PET scans in 30% of patients would have false-positive findings). Nevertheless, the relatively high rate of false-positive ¹⁸F-FDG PET findings necessitates that a biopsy be

performed of the PET-positive finding, which is typically at the site of residual mass, before any salvage therapy is contemplated—a clear-cut recommendation of the IHP (1,2). It cannot be over-emphasized, however, that in this case, an unnecessary biopsy, defined here as one with false-positive results, would be performed on only 11% of patients, a rate that is still reasonable and quite acceptable to virtually all hematologists or oncologists who treat HD. On the other hand, a 70% frequency of negative PET findings combined with an NPV of 94% translates into a misclassification of disease status because of a negative ¹⁸F-FDG PET finding in only about 4% of all patients (6% false-negative findings in 70% of patients). This remarkably low false-negative rate in patients with negative PET findings, which does not seem significantly different from that in patients with negative CT findings (i.e., no residual mass by CT), explains the high prognostic power and clinical utility of negative PET findings in patients with HD. HD patients with negative posttherapy PET findings, therefore, do not

require biopsy even in the face of a large residual mass and can safely be observed until there is clinical or radiologic evidence of relapse. In this context, it is noteworthy that the Terasawa review provides compelling evidence that the PPV and NPV of ¹⁸F-FDG PET after treatment are similar irrespective of the presence or absence of a residual mass as evident by the similar summary receiver operating characteristic curves and confidence regions for summary sensitivity and specificity for the “posttherapy” and “residual mass” evaluations presented in Figures 1 and 2 of Terasawa et al. (4).

The situation in patients with aggressive NHL does not seem fundamentally different, albeit generally higher PPVs combined with somewhat lower NPVs have been reported in patients with aggressive NHL than in patients with HD (3,4). Zijlstra et al. reported a PPV of 100% in the 2 studies exclusively assessing ¹⁸F-FDG PET in posttreatment evaluation of, mostly, aggressive NHL. The 6 posttherapy evaluation studies in patients with aggressive NHL included in the Terasawa review show a PPV ranging from 74% to 100%, with a weighted average of 90% (Table 1). The NPV of posttherapy ¹⁸F-FDG PET in aggressive NHL was about 84% in the 2 studies included in the Zijlstra review and 50%–83%, with a weighted average of 80%, in the 6 studies included in the Terasawa review, with all but one study showing an NPV of at least 75% (4). The implications of these findings on the utility of PET as a routine test in the posttreatment evaluation of patients with NHL are similar: the rates of positive and negative ¹⁸F-FDG PET scans after treatment in the 6 studies included in the Terasawa review were approximately 30% (range, 14%–50%) and 70% (range, 50%–86%), respectively, with similar findings reported in the Zijlstra review. With a positive ¹⁸F-FDG PET scan rate of 30% and a PPV of 90%, the misclassification of disease status because of a positive posttherapy ¹⁸F-FDG PET finding is estimated to affect only 3% of all patients (10% of PET scans had false-positive findings in 30% of patients). Although the false-positive rate is, on average, relatively low at 10%,

the reported variability in this rate between various studies (with a false-positive rate of approximately 25% in some studies) again necessitates that PET-positive findings undergo biopsy before any salvage therapy is contemplated (1,2,5). Here again, it should be noted that even with a false-positive rate of 25%, unnecessary (i.e., false-positive) biopsies would be performed on only 7.5% of patients—a rate that is clearly acceptable to hematologists or oncologists who treat aggressive NHL. A 70% rate of negative ^{18}F -FDG PET findings, combined with an NPV of 80%, translates into a misclassification of disease status because of negative ^{18}F -FDG PET findings in 14% of all patients (20% false-negative PET findings in 70% of patients), an acceptable false-negative rate that is not significantly different from that in patients with negative CT findings (5). The prognostic power and clinical utility of a negative PET finding is, therefore, maintained in patients with aggressive NHL (e.g., diffuse large B-cell lymphoma). Here again, a patient with a negative posttherapy ^{18}F -FDG PET finding does not require biopsy regardless of the presence or absence of residual masses and can safely be observed until there is clinical or radiologic evidence of relapse. The study by Terasawa et al. also shows here similar ranges of sensitivities, specificities, PPVs, and NPVs for the use of ^{18}F -FDG PET in posttherapy and residual-mass evaluations. Obviously, even a somewhat limited PPV of PET in the posttherapy or residual-mass assessment of lymphoma is still superior to that of conventional imaging with CT or MRI, which cannot reliably distinguish between necrosis or fibrosis and viable tumor (6). In fact, in one of the few studies in which CT and PET were compared in the same patients with aggressive NHL who underwent PET and CT within 1 mo of each other, Juweid et al. showed that the PPV of CT was only 43%, compared with 74% for PET ($P = 0.02$) (5,7). The data favor PET over CT even more in patients with HD, for whom the PPV of CT is only about 20%, compared with 60%–70% for PET (8).

Despite the clearly superior performance of PET or PET/CT to conventional imaging in posttherapy evaluation of lymphoma (6–8), it is disheartening to note such an investigational heterogeneity in the posttherapy ^{18}F -FDG PET studies included in the various systematic and unsystematic reviews published to date (3,4,7,9). Although the Terasawa review failed to identify “clinical or ^{18}F -FDG PET test characteristics...or any items that assessed the quality and applicability of each study” that could explain such heterogeneity, it might be interesting to speculate about at least some of the most likely ones. For example, whether radiation therapy was part of the treatment given before PET may have affected PET performance. It is conceivable that the generally lower PPV of PET (PET/CT) in patients with HD than in those with aggressive NHL may be related to the substantial fraction of HD patients who received radiation therapy, either alone or combined with chemotherapy, before undergoing PET (7,8). Another factor that should be considered is whether the posttherapy PET study was performed or interpreted with or without attenuation correction. It is conceivable that the PPV of PET was higher in the non-attenuation-corrected than in the attenuation-corrected studies simply because mild ^{18}F -FDG PET uptake, particularly in deep-seated residual masses, was not visualized on non-attenuation-corrected scans, thereby resulting in a negative scan interpretation which, in fact, turned out to be more often a true- than a false-negative interpretation (7). Apparently, without attenuation correction, only lesions with “substantial,” “prognostically significant” uptake were visualized, resulting in most of them proving to be true-positive. This could explain the higher PPVs seen in the earlier studies by Jerusalem et al. and Spaepen et al. (10,11), who used non-attenuation-corrected scans in their posttherapy PET studies in patients with lymphoma, compared with the PPV seen in the later study by Juweid et al. using attenuation-corrected scans (5). Interestingly, the NPV of ^{18}F -FDG PET was

strikingly similar in those 3 studies (Table 1). This observation, among others, provided the rationale for the IHP recommendation that, in interpretations of attenuation-corrected scans, mild diffuse ^{18}F -FDG PET uptake in ≥ 2 -cm residual masses that is equal to or less than that of the mediastinal blood pool structures should be considered a negative finding. Investigational heterogeneity might also partially be explained by differences in ^{18}F -FDG PET scan interpretation (although such differences were generally minor) and by other factors not yet explored (3).

In conclusion, similar to the previous review of Zijlstra et al. (3), the review by Terasawa et al., in fact, supports the use of ^{18}F -FDG PET for posttherapy evaluation of patients with HD and aggressive NHL. The key point here is to recognize both the strengths and the limitations of ^{18}F -FDG PET in these settings (6–9). The data for the prognostic power of negative PET findings for both disease types are compelling, regardless of the presence or absence of a residual mass, with no evidence that the predictive power of negative PET results found at therapy completion is inferior to the predictive power of negative CT or MRI results (5). On the other hand, the predictive power of positive PET findings is somewhat limited, dictating the necessity of biopsying PET-positive findings, whenever feasible, before salvage treatment is contemplated. The fact that the rate of PET-positive scans at therapy completion in both diseases is on the order of 30% and that the PPV in most studies exceeds 60% suggests that the incidence of unnecessary false-positive biopsies will remain relatively low and quite acceptable. The rate of positive posttherapy ^{18}F -FDG PET findings in patients with aggressive NHL is likely to decline further, with the widespread use of the more effective chemotherapy—rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP)—expected to result in a higher rate of negative ^{18}F -FDG PET scans approaching 75%, as preliminary results from an ongoing ^{18}F -FDG PET trial after 4 cycles of

R-CHOP indicate (written communication, L. Sehn). On the other hand, strict adherence to the IHP guidelines is likely to reduce variability in the interpretation of ^{18}F -FDG PET scans in the posttherapy setting and also decrease the rate of false-positive interpretations of ^{18}F -FDG PET findings at the site of residual masses (1). Of course, none of these factors negate the usefulness of additional prospective studies with a more rigorous design, conduct, and reporting to clearly establish the prognostic power and clinical utility of ^{18}F -FDG PET in the posttherapy setting to a point at which absolutely no doubt remains.

Malik E. Juweid
University of Iowa
Iowa City, Iowa

REFERENCES

1. 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.
2. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007;25:579–586.
3. Zijlstra JM, Lindauer-van der Werf G, Hoekstra OS, Hooft L, Riphagen II, Huijgens PC. ^{18}F -Fluorine-fluoro-deoxyglucose positron emission tomography for post-treatment evaluation of malignant lymphoma: a systematic review. *Haematologica.* 2006;91:522–529.
4. Terasawa T, Nihashi T, Hotta T, Nagai H. ^{18}F -FDG PET for posttherapy assessment of Hodgkin's disease and aggressive non-Hodgkin's lymphoma: a systematic review. *J Nucl Med.* 2007;49:13–21.
5. Juweid ME, Wiseman GA, Vose JM, et al. Response assessment of aggressive non-Hodgkin's lymphoma by integrated International Workshop Criteria and fluorine-18-fluorodeoxyglucose positron emission tomography. *J Clin Oncol.* 2005;23:4652–4661.
6. Juweid ME, Cheson BD. Role of positron emission tomography in lymphoma. *J Clin Oncol.* 2005;23:4577–4580.
7. Seam P, Juweid ME, Cheson BD. The role of FDG-PET scans in patients with lymphoma. *Blood.* 2007;110:3507–3516.
8. Juweid ME. Utility of positron emission tomography (PET) scanning in managing patients with Hodgkin lymphoma. *Hematology Am Soc Hematol Educ Program.* 2006:259–265, 510–511.
9. Juweid ME, Cheson BD. Positron emission tomography and assessment of cancer therapy. *N Engl J Med.* 2006;354:496–507.
10. Jerusalem G, Beguin Y, Fassotte MF, et al. Whole-body positron emission tomography using ^{18}F -fluorodeoxyglucose for posttreatment evaluation in Hodgkin's disease and non-Hodgkin's lymphoma has higher diagnostic and prognostic value than classical computed tomography scan imaging. *Blood.* 1999;94:429–433.
11. Spaepen K, Stroobants S, Dupont P, et al. Prognostic value of positron emission tomography (PET) with fluorine 18 fluorodeoxyglucose (^{18}F]FDG) after first-line chemotherapy in non-Hodgkin's lymphoma: is [^{18}F]FDG-PET a valid alternative to conventional diagnostic methods? *J Clin Oncol.* 2001;19:414–419.