# $^{18}$F-FDG PET for Evaluation of Bone Marrow Infiltration in Staging of Lymphoma: A Meta-Analysis

Emilios E. Pakos, MD[1]; Andreas D. Fotopoulos, MD[2]; and John P.A. Ioannidis, MD[1,3,4]

[1]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece;* [2]*Department of Nuclear Medicine, University Hospital of Ioannina, School of Medicine, Ioannina, Greece;* [3]*Biomedical Research Institute, Foundation for Research and Technology–Hellas, Ioannina, Greece; and* [4]*Department of Medicine, Institute for Clinical Research and Health Policy Studies, Tufts–New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts*

The ability of PET with $^{18}$F-FDG to evaluate bone marrow infiltration in patients with lymphoma has been a matter of extensive investigation with controversial results. Therefore, we aimed to evaluate systematically, with a meta-analysis, the diagnostic performance of $^{18}$F-FDG PET in this setting. **Methods:** Relevant studies were identified with MEDLINE and EMBASE searches (last update, August 2004). Data on the diagnostic performance of $^{18}$F-FDG PET were combined quantitatively across eligible studies. We estimated weighted summary sensitivities and specificities, summary receiver-operating-characteristic (SROC) curves, and weighted summary likelihood ratios. We also conducted separate analyses according to various subgroups. Bone marrow biopsy (BMB) was used as the reference standard. **Results:** Thirteen eligible nonoverlapping studies, which enrolled a total of 587 patients, were included in the meta-analysis. The independent random-effects weighted estimates of sensitivity and specificity against BMB were 51% (95% confidence interval [CI], 38%–64%) and 91% (95% CI, 85%–95%), respectively. Results were consistent in the SROC curve: a sensitivity of 51% corresponds to a specificity of 92%, whereas a specificity of 91% corresponds to a sensitivity of 55%. The weighted positive likelihood ratio (LR+) was 5.75 (95% CI, 348–9.48) and the negative likelihood ratio (LR−) was 0.67 (95% CI, 0.55–0.82). Six of 12 patients with positive $^{18}$F-FDG PET and negative initial biopsy were found to have bone marrow involvement when biopsy was performed at the sites with positive imaging signals. Subgroup analyses showed better sensitivity in patients with Hodgkin's disease and in aggressive histologic types of non-Hodgkin's lymphoma than in patients with less aggressive histologic types and in studies using unilateral BMB compared with those using bilateral biopsy. **Conclusion:** This meta-analysis showed that $^{18}$F-FDG PET has good, but not excellent, concordance with the results of BMB for the detection of bone marrow infiltration in the staging of patients with lymphoma. $^{18}$F-FDG PET may complement the results of BMB and its performance may vary according to the type of lymphoma.

**Key Words:** $^{18}$F-FDG PET; staging; lymphoma; bone marrow; biopsy

**J Nucl Med 2005; 46:958–963**

PET with the radiolabeled glucose analog $^{18}$F-FDG has been increasingly used in the evaluation of several malignant tumors, including lymphoma (*1–4*). One of the most promising applications in lymphoma is to determine the clinical stage of the disease at initial presentation or recurrence (*5,6*). The ability of $^{18}$F-FDG PET to evaluate both nodal and extranodal sites such as spleen, liver, and bone marrow has been a matter of extensive investigation. In particular, bone marrow infiltration is of crucial importance in staging of lymphoma, since it signifies advanced-stage disease and, thus, may affect both treatment and prognosis. Bone marrow biopsy (BMB) is the established method for the detection of bone marrow infiltration. However, BMB is a painful procedure. Moreover, sometimes only a small sample can be obtained, which may be inconclusive. Several studies have been conducted to date addressing the ability of $^{18}$F-FDG PET to evaluate bone marrow infiltration in staging of lymphoma (*7–19*). However, these studies have been ineffective in evaluating the diagnostic accuracy of $^{18}$F-FDG PET due to small sample sizes. A quantitative synthesis using rigorous methods would be important to perform. Therefore, we undertook a meta-analysis of all available studies to address the diagnostic performance of $^{18}$F-FDG PET in evaluating bone marrow infiltration in the staging of patients with primary lymphoma or recurrent lymphoma after complete remission.

## MATERIALS AND METHODS

### Identification and Eligibility of Relevant Studies

We considered studies examining the performance of $^{18}$F-FDG PET as a diagnostic test for detecting bone marrow infiltration in the initial staging or staging of recurrent disease before treatment

Received Dec. 27, 2004; revision accepted Jan. 27, 2005.

For correspondence contact: John P.A. Ioannidis, MD, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece.

E-mail: jioannid@cc.uoi.gr

in lymphoma. We considered all relevant studies that included patients with and without bone marrow infiltration according to biopsy results and that had a total sample size of at least 5 patients. Studies were included in our meta-analysis regardless of the type of lymphoma (Hodgkin's disease [HD], non-Hodgkin's lymphoma [NHL]). We excluded studies using $^{18}$F-FDG PET for evaluation of recurrences after treatment. We also excluded studies in which the selection to perform one test (BMB or $^{18}$F-FDG PET) was based on the results of the other test, since this would entail clear verification bias.

We conducted MEDLINE and EMBASE searches (last update, August 2004). The search strategy was based on the combination of the terms (a) PET, positron emission tomography, $^{18}$F-FDG, or fluorodeoxyglucose; (b) lymphoma, Hodgkin disease, or non-Hodgkin lymphoma; and (c) diagnosis or staging. Searches were limited to human subjects.

References of retrieved articles were also screened for additional studies. Investigators of eligible studies were contacted and asked to supplement additional data, when key information relevant to the meta-analysis was missing. Whenever reports pertained to overlapping patients, we retained only the largest study to avoid duplication of information. We set no language restrictions.

### Data Extraction

Two investigators extracted data from eligible studies independently, discussed discrepancies, and reached consensus for all items with the help of a third investigator. We extracted data on characteristics of studies and patients, measurements performed, and results. In each report, we recorded author names, journal and year of publication, country of origin, years of patient enrollment, number of eligible patients, number of patients analyzed, reasons for exclusions from the analysis, study design (prospective, retrospective, or unclear), type of lymphoma (HD, NHL), histologic type, disease status (primary or recurrent), inclusion and exclusion criteria, demographic characteristics of patients, stage of lymphoma, location of BMB and whether it was unilateral or bilateral, time of $^{18}$F-FDG PET (before or after biopsy), technical characteristics of $^{18}$F-FDG PET, definition of positive $^{18}$F-FDG PET test (qualitative or quantitative methods), and number of experts who assessed and interpreted the results of $^{18}$F-FDG PET and biopsy. We also recorded whether there was any mention on blinding of $^{18}$F-FDG PET measurements to the BMB results and vice versa and whether any data were given on inter- or intraobserver variability.

For each report, we recorded the number of true-positive, false-positive, true-negative, and false-negative findings for $^{18}$F-FDG PET in diagnosing bone marrow infiltration, using BMB as the reference standard. These terms are used for convention to denote the concordance of the 2 diagnostic tests, since it is unlikely that BMB is a perfect gold standard. We also recorded whether a new local BMB had been performed at a site with positive $^{18}$F-FDG PET, whenever the initial BMB was negative. Rebiopsy results were not considered in the main analysis but were analyzed separately. We also recorded separate data for HD and NHL and for primary and recurrent lymphoma, whenever these data were available.

### Statistical Analysis

Data on the diagnostic performance of $^{18}$F-FDG PET were combined quantitatively across eligible studies. Three approaches were used. First, we combined independently sensitivities and specificities across studies. Between-study heterogeneity was assessed with the Fisher exact test. We estimated the weighted sensitivities and specificities using a random-effects model that incorporated between-study heterogeneity. Second, we constructed summary receiver-operating-characteristic (SROC) curves. Third, we estimated the weighted positive and negative likelihood ratio (LR+, LR−) across studies using random-effects calculations.

For a diagnostic or predictive test, the sensitivity (true-positives) and specificity (1 − false-positive) are related to each other; therefore, it is not totally correct to estimate these 2 quantities independently. To bypass this problem, one may use the SROC method. The SROC curve is estimated by the regression $D = a + bS$, where D is the difference of the logits of the true-positive and false-positive rate and S is the sum of these logits (*20*). Both weighted and unweighted regressions were estimated. The SROC curve shows the trade-off between sensitivity and specificity across the included studies.

Likelihood ratios are also metrics that combine both sensitivity and specificity in their calculation. LR+ is defined as the ratio of sensitivity over 1 − specificity, whereas LR− is defined as the ratio of 1 − sensitivity over specificity. When there is absolutely no discriminating ability for a diagnostic test, both likelihood ratios equal 1. The discriminating ability is better with higher LR+ and lower LR−. Although there is no absolute cutoff, a good diagnostic test may have LR+ above 5 and LR− below 0.2. Between-study heterogeneity in the likelihood ratios was assessed with the Q statistic (*21*) and was considered significant for $P < 0.10$ (*22*). We also estimated whether the LR+ and LR− were significantly different in small versus larger studies.

The main analysis combined all data regardless of the definition of $^{18}$F-FDG PET positivity, type of lymphoma (HD or NHL), disease status (primary or recurrent), type of biopsy (unilateral or bilateral), study design (prospective or retrospective), and blinding of each diagnostic test to the results of the other. However, subgroup analyses were also performed for each of these parameters.

Analyses were conducted in SPSS (SPSS, Inc.), Meta-Test (Joseph Lau, Boston, MA), and StatXact 3.0 (Cytel Inc.). *P* values are 2-tailed.

## RESULTS

### Eligible Studies

Twenty-six potentially eligible reports were retrieved. Of those, 5 were excluded because, although the authors stated that patients had undergone both $^{18}$F-FDG PET and BMB, the results of the biopsy were not reported and the authors did not respond to our attempts to contact them (*23–27*). Two studies were excluded because all patients had positive $^{18}$F-FDG PET results (*28,29*). One study (*30*) was excluded because all BMBs that were performed were negative. Five reports pertained to overlapping patients (*8,31–34*). We accepted the report with the largest sample size (*8*) and the remaining 4 were excluded from our analysis. Another report (*35*) also overlapped with a larger study (*15*) and was excluded. Of 2 reports from the same institution (*18,19*), where only one reported results on NHL (*19*) while both gave data on HD, we excluded the data on the smaller population of HD patients to avoid overlap (*18*). One study (*11*), including 21 patients at initial staging and 9 patients at restaging after treatment without providing separate results,

was considered eligible for the meta-analysis. However, analyses excluding this study were also done and showed the same results (not shown).

Finally, 13 eligible non-overlapping studies, which enrolled a total of 587 patients, were included in the meta-analysis (Table 1). The mean age of patients varied from 13 to 65 years across eligible studies. Seven studies included patients with primary disease (*8,10,11,13,14,17,19*), whereas the others included mixed populations with primary and recurrent lymphoma (*7,9,12,15,16,18*). Four studies recruited patients with HD (*7,13,16,18*), 3 studies had patients with NHL (*10,15,19*), and 6 studies had mixed populations (*8,9,11,12,14,17*). Iliac crest BMB was performed in 6 studies (*7–9,11,17,19*), whereas in 1 study the biopsy was either from the sternum or the iliac crest (*12*), and 6 studies did not report the location of the biopsy (*10,13–16,18*). The $^{18}$F-FDG dose ranged considerably across studies (Table 1). The vast majority of studies ($n$ = 11) used qualitative methods to evaluate the $^{18}$F-FDG PET scans (Table 1). Two studies (*7,9*) used quantitative methods, with standardized uptake values (SUVs) of 2.0 (*7*) and cut-offs for positivity of 2.5 (*9*). Nine studies reported blinding of $^{18}$F-FDG PET or BMB measurements to each other (*8,10–12,15–19*).

### Data Synthesis

The sensitivity rates of $^{18}$F-FDG PET for identifying bone marrow infiltration ranged from 0% to 100% across the eligible studies ($P$ = 0.014 for heterogeneity). The respective specificity rates ranged from 72% to 100% ($P < 0.001$ for heterogeneity). When all studies were considered, there were 50 patients with bone marrow infiltration and positive $^{18}$F-FDG PET findings, 53 patients with bone marrow infiltration identified as negative by $^{18}$F-FDG PET, 449 patients without bone marrow infiltration and negative $^{18}$F-FDG PET findings, and 35 patients without bone marrow infiltration identified as positive by $^{18}$F-FDG PET. The independent random-effects summary estimates of sensitivity and specificity were 51% (95% confidence interval [CI], 38%–64%) and 91% (95% CI, 85%–95%), respectively. In the SROC curve, the results were consistent with those obtained in the independent weighting of sensitivity and specificity: a sensitivity of 51% corresponded to a specificity of 92%, whereas a specificity of 91% corresponded to a sensitivity of 55% (Fig. 1). The slope of the regression of the SROC curve was negligible and nonsignificant, suggesting that the overall diagnostic performance was similar at different parts of the curve, after allowing for the trade-off between sensitivity and specificity. Likelihood ratio syntheses gave a weighted LR+ of 5.75 (95% CI, 3.48–9.48) and weighted LR− of 0.67 (95% CI, 0.55–0.82) without any statistically significant between-study heterogeneity for either metric ($P > 0.10$ for both). There was no evidence that the LR+ differed in small versus larger studies ($\tau$ correlation coefficient between the natural logarithm of the LR+ and the weight of each study = −0.03, $P$ = 0.90). Conversely, there

**TABLE 1**
Characteristics of Eligible Studies

| Reference | Year | Design | Sample ($n$) | Primary ($n$) | Type of lymphoma | Mean or median age (y) | Male (%) | BMB site | $^{18}$F-FDG dose (MBq) | Measures | Any blinding stated |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Naumann et al. (*7*) | 2004 | Prospective | 88 | 77 | HD | 34 | 65 | NR | 300–370 | SUV | No |
| Elstrom et al. (*9*) | 2003 | Retrospective | 105 | NR | HD, NHL | NR | NR | NR | 2.516/kg | SUV | No |
| Hoffmann et al. (*10*) | 2003 | Retrospective | 21 | 21 | NHL | NR | NR | NR | 300–380 | Qualitatively | Yes |
| Hong et al. (*11*)* | 2003 | Prospective | 30 | 21 | HD, NHL | 49 | 63 | Bilateral | 370 | Qualitatively | Yes |
| Sasaki et al. (*12*) | 2002 | NR | 30 | NR | HD, NHL | 60 | 61 | NR | 270.6, mean | Qualitatively | Yes |
| Montravers et al. (*13*) | 2002 | NR | 7 | 7 | HD | 13 | NR | NR | 2–3/kg | Qualitatively | No |
| Wirth et al. (*14*) | 2002 | Retrospective | 39 | NR | HD, NHL | 48 | 62 | NR | NR | Qualitatively | No |
| Jerusalem et al. (*15*) | 2001 | Prospective | 42 | 26 | NHL | 62 | NR | Bilateral | 200–300 | Qualitatively | Yes |
| Jerusalem et al. (*16*) | 2001 | Prospective | 33 | 24 | HD | 33 | 39 | Bilateral | 200–300 | Qualitatively | Yes |
| Buchmann et al. (*17*) | 2001 | Prospective | 52 | 52 | HD, NHL | 41 | 54 | Unilateral | 390, mean | Qualitatively | Yes |
| Partridge et al. (*18*) | 2000 | Retrospective | 24 | NR | HD | 38 | 48 | Unilateral | 350 | Qualitatively | Yes |
| Moog et al. (*8*) | 1998 | Prospective | 78 | 78 | HD, NHL | 38 | 46 | Bilateral, unilateral | 270, mean | Qualitatively | Yes |
| Carr et al. (*19*) | 1998 | Prospective | 38 | 38 | NHL | NR | NR | Unilateral | 350 | Qualitatively | Yes |

*Includes 21 patients at initial staging and 9 patients at restaging after treatment.
NR = not reported; SUV = standardized uptake value.

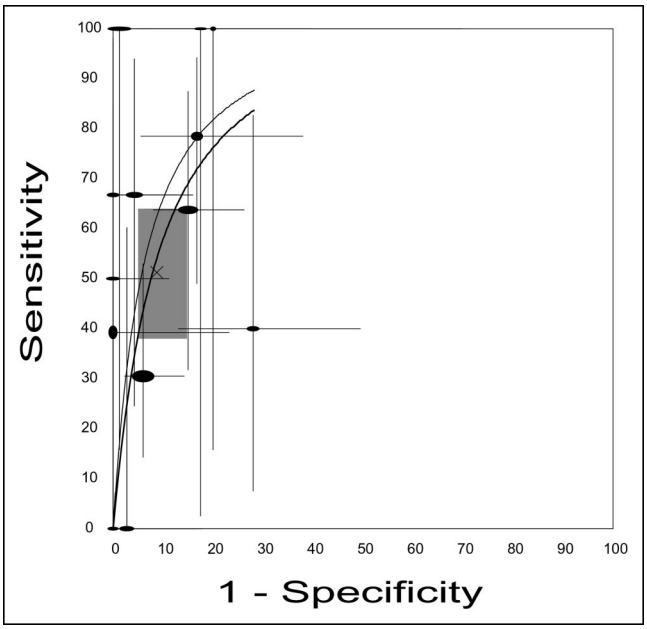


**FIGURE 1.** Summary SROC curve analysis shows ability of $^{18}$F-FDG PET to evaluate bone marrow infiltration in patients with lymphoma. Each study is shown by an ellipse demonstrating sensitivity and specificity of $^{18}$F-FDG PET along with lines extending to respective 95% CIs. BMB is reference standard. Ellipse axes are proportional to weight of study in specificity and sensitivity dimensions. Data are summarized with weighted (thick line) and unweighted (thin line) SROC curves.

was some evidence that the LR− was less favorable in larger studies (τ correlation coefficient between the natural logarithm of the LR− and the weight of each study = 0.69; $P = 0.001$).

Two studies ($n = 130$) reported secondary biopsy in seemingly false-positive patients (*8,17*). Six of 12 rebiopsied patients (50%; 95% binomial CI: 21%–79%) were actually found to have bone marrow involvement at the sites with positive PET signals. Including the results of the secondary biopsy in the reference standard, the weighted sensitivity and specificity of PET against BMB in these 2 studies became 74% (95% CI, 53%–88%) and 95% (95% CI, 72%–99%), respectively, and the weighted LR+ and LR− were 14.5 (95% CI, 2.15–98.1) and 0.29 (95% CI, 0.15–0.56). Replacing these 2 studies in the main analysis yielded summary sensitivity and specificity of 54% (95% CI, 40%–68%) and 92% (95% CI, 86%–96%), respectively, and the weighted LR+ and LR− became 6.45 (95% CI, 3.71–11.2) and 0.62 (95% CI, 0.49–0.79), respectively. Note that the 2 studies with available rebiopsy results already had a sensitivity of 65% for the PET against the first BMB results.

### Subgroup Analyses

The weighted rates showed significantly better sensitivity in studies with HD than in those with NHL patients. However, there were only 11 patients with positive BMB among cases with HD. For NHL, there was a clear difference in the sensitivity depending on the histologic type. On the basis of the available data, $^{18}$F-FDG PET identified 16 of 21 cases of bone marrow involvement (76.2%) from large lymphocytic, large B-cell, Burkitt, and centroblastic lymphocytic lymphomas, whereas it detected only 16 of 53 cases with bone marrow involvement (30.2%) from less aggressive histologic types (follicular, mantle cell, marginal zone, small lymphocytic lymphomas and mucosa-associated lymphoid tissue) ($P < 0.001$). There was also significantly better sensitivity in studies using unilateral BMB compared with those using bilateral biopsy, but this was also based on relatively sparse data (Table 2). No major subgroup differences were observed for prospective versus retrospective studies, studies with versus without reported blinding, and studies with qualitative versus quantitative PET measurements (Table 2).

## DISCUSSION

This meta-analysis including data from 587 patients showed that $^{18}$F-FDG PET has moderately good, but not excellent, concordance with the results of BMB for the detection of bone marrow infiltration in the staging of patients with lymphoma. Only about half of the patients with bone marrow infiltration detected in BMB were detected as positive by $^{18}$F-FDG PET. On the other hand, >90% of patients with a negative BMB will also have negative $^{18}$F-FDG PET. In fact, positive $^{18}$F-FDG PET in the presence of negative BMB often indicated missed bone marrow involvement that could be documented with a second BMB directed at the site of positive PET signal. On the basis of these findings, $^{18}$F-FDG PET cannot yet be recommended for replacing BMB routinely in the staging of lymphoma because many cases of bone marrow involvement would be missed. However, $^{18}$F-FDG PET could complement BMB and could occasionally identify additional cases of focal bone marrow involvement that would be missed by the BMB. It is essential to establish in future research whether this complementary information may have considerable impact on the prognosis of these patients. Also, the current meta-analysis did not address the accuracy of PET in restaged patients.

The differences that were observed in subgroup analyses could be possibly due to chance. However, $^{18}$F-FDG PET showed considerable variable sensitivity for the evaluation of bone marrow infiltration depending on the histologic type of lymphoma. Sensitivity was very good for HD, but very few patients with HD had bone marrow involvement in our accumulated sample, so this encouraging finding has to be verified in a larger number of patients with various levels of bone marrow involvement. Conversely, the aggregate sensitivity was modest in NHLs. Overall, the rates of bone marrow involvement are reported to be higher in NHL compared with HD (*36–38*). Scrutiny of the available data showed that sensitivity was actually very good for detection

**TABLE 2**
Subgroup Analyses for Diagnostic Performance of $^{18}$F-FDG PET of Bone Marrow Infiltration in Lymphoma

| Analysis | No. of studies (patients) | Independent estimates (95% CI) | | Likelihood ratio (95% CI) | |
|---|---|---|---|---|---|
| | | Sensitivity (%) | Specificity (%) | LR+ | LR− |
| Prospective design | | | | | |
| Yes | 7 (361) | 57 (42–70) | 91 (81–96) | 7.26 (3.15–16.7) | 0.56 (0.44–0.72) |
| No/not specified | 6 (226) | 39 (20–62) | 92 (83–96) | 4.65 (2.45–8.85) | 0.81 (0.67–0.97) |
| Type of lymphoma | | | | | |
| HD | 5 (191)* | 76 (47–92) | 92 (79–97) | 9.02 (3.52–23.2) | 0.33 (0.14–0.77) |
| NHL | 6 (239) | 43 (28–60) | 88 (75–94) | 3.53 (1.88–6.63) | 0.68 (0.57–0.81) |
| Both/not separable | 3 (121) | 52 (24–79) | 97 (91–99) | 13.3 (4.02–44.4) | 0.61 (0.29–1.26) |
| Disease status | | | | | |
| Primary | 7 (297) | 72 (57–83) | 93 (85–97) | 8.93 (4.31–18.5) | 0.41 (0.25–0.68) |
| Recurrent | NA | NA | NA | NA | NA |
| Both/not separable | 6 (270) | 38 (27–50) | 90 (79–96) | 3.91 (1.97–7.75) | 0.15 (0.06–0.37) |
| Type of BMB | | | | | |
| Unilateral | 3 (114) | 75 (53–89) | 87 (74–95) | 5.73 (2.84–11.5) | 0.30 (0.14–0.61) |
| Bilateral† | 4 (183) | 46 (32–60) | 87 (70–95) | 4.07 (1.50–11.1) | 0.61 (0.47–0.80) |
| Not specified | 6 (290) | 42 (20–67) | 95 (90–98) | 8.39 (3.04–23.1) | 0.07 (0.02–0.26) |
| Blinding | | | | | |
| Yes | 9 (348) | 54 (41–66) | 89 (80–94) | 5.06 (2.97–8.63) | 0.60 (0.48–0.75) |
| No | 4 (239) | 44 (15–79) | 94 (86–98) | 7.23 (2.09–25.1) | 0.80 (0.62–1.05) |
| PET measurement | | | | | |
| Qualitative | 11 (394) | 54 (41–67) | 89 (82–93) | 4.63 (3.00–7.14) | 0.64 (0.50–0.82) |
| Quantitative | 2 (193) | 36 (19–57) | 96 (92–99) | 14.1 (1.54–1.30) | 0.62 (0.24–1.59) |

*Thirty-two patients from Elstrom et al. (*9*) and 4 patients from Hong et al. (*11*) with HD were excluded from this subgroup analysis since all BMBs in these patients with HD were negative.
†Moog (*8*) includes 8 patients with unilateral BMB, but these could not be separated.
NA = not available.
All analyses are based on random-effects calculations.



of bone marrow disease when aggressive types of NHL were involved. This is probably due to the high metabolic activity, and possibly most extensive bone marrow involvement, of these tumors. On the contrary, $^{18}$F-FDG PET detected less than a third of bone marrow involvement by more indolent histologic types of NHLs. These cases might have had mostly limited involvement of the bone marrow (*8*).

Another challenging finding in this meta-analysis was that the sensitivity of $^{18}$F-FDG PET was significantly lower in studies using bilateral BMB compared with those using unilateral biopsy as the reference standard. BMB removes a small core of marrow and, therefore, is subject to sampling errors. The patchy nature of some lymphomas may lead to discordant findings between the 2 cores in bilateral biopsies. The reported rates of unilateral involvement in bilateral biopsies range from 10% to 50% (*8,19,38*), clear proof of the limitations of BMB as a proposed gold standard. Cases with bone marrow infiltration missed by unilateral biopsies might be mostly those with less extensive bone marrow infiltration. Therefore, cases detected on bilateral, but not unilateral, BMB may be less likely to be identified by $^{18}$F-FDG PET.

Some limitations of this meta-analysis should be acknowledged. First, the overall sample size was limited. However, we tried to be all-inclusive and, to our knowledge, the cumulative sample size of the meta-analysis was about 6 times larger than the largest single study published to date. We tried to retrieve additional data, but it is possible that some missing data may still exist. It is unknown whether publication bias may operate in this field against the publication of small studies with less-promising results. Second, as already acknowledged, the biopsy reference standard is not perfect for the evaluation of bone marrow involvement. However, this might lead mostly to underestimation of the diagnostic performance of $^{18}$F-FDG PET. Finally, in the vast majority of studies, the interpretation of $^{18}$F-FDG PET scans was performed by qualitative methods. The qualitative interpretation of $^{18}$F-FDG PET scans was largely based on subjective evaluation and the results were given after consensus between experts. Quantitative methods were used by only 2 studies. Future studies should focus more on quantitative indices.

## CONCLUSION

Allowing for these caveats, the meta-analysis suggests that $^{18}$F-FDG PET has overall good diagnostic performance for detecting bone marrow involvement, but this may depend also on the type of lymphoma. $^{18}$F-FDG PET may

complement BMB in the staging of primary or recurrent lymphoma.

## ACKNOWLEDGMENTS

We thank Drs. Ralph Naumann, Martha Hoffman, and Masayuki Sasaki for providing additional data and clarifications of their studies.

## REFERENCES

1. Kalff V, Hicks RJ, Ware RE, Hogg A, Binns D, McKenzie AF. The clinical impact of $^{18}$F-FDG PET in patients with suspected or confirmed recurrence of colorectal cancer: a prospective study. *J Nucl Med.* 2002;43:492–499.
2. Vansteenkiste J, Fischer BM, Dooms C, Mortensen J. Positron-emission tomography in prognostic and therapeutic assessment of lung cancer: systematic review. *Lancet Oncol.* 2004;5:531–540.
3. Ioannidis JP, Lau J. $^{18}$F-FDG PET for the diagnosis and grading of soft-tissue sarcoma: a meta-analysis. *J Nucl Med.* 2003;44:717–724.
4. Reske SN. PET and restaging of malignant lymphoma including residual masses and relapse. *Eur J Nucl Med Mol Imaging.* 2003;30(suppl 1):89–96.
5. Stumpe KD, Urbinelli M, Steinert HC, Glanzmann C, Buck A, von Schulthess GK. Whole-body positron emission tomography using fluorodeoxyglucose for staging of lymphoma: effectiveness and comparison with computed tomography. *Eur J Nucl Med.* 1998;25:721–728.
6. De Wit M, Bumann D, Beyer W, Herbst K, Clausen M, Hossfeld DK. Whole-body positron emission tomography (PET) for diagnosis of residual mass in patients with lymphoma. *Ann Oncol.* 1997;8(suppl 1):57–60.
7. Naumann R, Beuthien-Baumann B, Reiss A, et al. Substantial impact of FDG PET imaging on the therapy decision in patients with early-stage Hodgkin's lymphoma. *Br J Cancer.* 2004;90:620–625.
8. Moog F, Bangerter M, Kotzerke J, Guhlmann A, Frickhofen N, Reske SN. 18-F-fluorodeoxyglucose-positron emission tomography as a new approach to detect lymphomatous bone marrow. *J Clin Oncol.* 1998;16:603–609.
9. Elstrom R, Guan L, Baker G, et al. Utility of FDG-PET scanning in lymphoma by WHO classification. *Blood.* 2003;101:3875–3876.
10. Hoffmann M, Kletter K, Becherer A, Jager U, Chott A, Raderer M. $^{18}$F-Fluorodeoxyglucose positron emission tomography ($^{18}$F-FDG-PET) for staging and follow-up of marginal zone B-cell lymphoma. *Oncology.* 2003;64:336–340.
11. Hong SP, Hahn JS, Lee JD, Bae SW, Youn MJ. $^{18}$F-Fluorodeoxyglucose-positron emission tomography in the staging of malignant lymphoma compared with CT and $^{67}$Ga scan. *Yonsei Med J.* 2003;44:779–786.
12. Sasaki M, Kuwabara Y, Koga H, et al. Clinical impact of whole body FDG-PET on the staging and therapeutic decision making for malignant lymphoma. *Ann Nucl Med.* 2002;16:337–345.
13. Montravers F, McNamara D, Landman-Parker J, et al. [$^{18}$F]FDG in childhood lymphoma: clinical utility and impact on management. *Eur J Nucl Med Mol Imaging.* 2002;29:1155–1165.
14. Wirth A, Seymour JF, Hicks RJ, et al. Fluorine-18 fluorodeoxyglucose positron emission tomography, gallium-67 scintigraphy, and conventional staging for Hodgkin's disease and non-Hodgkin's lymphoma. *Am J Med.* 2002;112:262–268.
15. Jerusalem G, Beguin Y, Najjar F, et al. Positron emission tomography (PET) with $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) for the staging of low-grade non-Hodgkin's lymphoma (NHL). *Ann Oncol.* 2001;12:825–830.
16. Jerusalem G, Beguin Y, Fassotte MF, et al. Whole-body positron emission tomography using $^{18}$F-fluorodeoxyglucose compared to standard procedures for staging patients with Hodgkin's disease. *Haematologica.* 2001;86:266–273.
17. Buchmann I, Reinhardt M, Elsner K, et al. 2-(Fluorine-18)fluoro-2-deoxy-D-glucose positron emission tomography in the detection and staging of malignant lymphoma: a bicenter trial. *Cancer.* 2001;91:889–899.
18. Partridge S, Timothy A, O'Doherty MJ, Hain SF, Rankin S, Mikhaeel G. 2-Fluorine-18-fluoro-2-deoxy-D glucose positron emission tomography in the pretreatment staging of Hodgkin's disease: influence on patient management in a single institution. *Ann Oncol.* 2000;11:1273–1279.
19. Carr R, Barrington SF, Madan B, et al. Detection of lymphoma in bone marrow by whole-body positron emission tomography. *Blood.* 1998;91:3340–3346.
20. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* 1993;12:1293–1316.
21. Pettiti DB. *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis.* 2nd ed. New York, NY: Oxford University Press; 1999.
22. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med.* 1997;127:820–826.
23. Nuutinen J, Minn H, Bergman J, et al. Uncoupling of fatty acid and glucose metabolism in malignant lymphoma: a PET study. *Br J Cancer.* 1999;80:513–518.
24. Tatsumi M, Kitayama H, Sugahara H, et al. Whole-body hybrid PET with $^{18}$F-FDG in the staging of non-Hodgkin's lymphoma. *J Nucl Med.* 2001;42:601–608.
25. Hueltenschmidt B, Sautter-Bihl ML, Lang O, et al. Whole body positron emission tomography in the treatment of Hodgkin disease. *Cancer.* 2001;91:302–310.
26. Lapela M, Leskinen S, Minn HR, et al. Increased glucose metabolism in untreated non-Hodgkin's lymphoma: a study with positron emission tomography and fluorine-18-fluorodeoxyglucose. *Blood.* 1995;86:3522–3527.
27. Weihrauch MR, Re D, Bischoff S, et al. Whole-body positron emission tomography using $^{18}$F-fluorodeoxyglucose for initial staging of patients with Hodgkin's disease. *Ann Hematol.* 2002;81:20–25.
28. Jerusalem G, Warland V, Najjar F, et al. Whole-body $^{18}$F-FDG PET for the evaluation of patients with Hodgkin's disease and non-Hodgkin's lymphoma. *Nucl Med Commun.* 1999;20:13–20.
29. Hwang K, Park CH, Kim HC, et al. Imaging of malignant lymphomas with F-18 FDG coincidence detection positron emission tomography. *Clin Nucl Med.* 2000; 25:789–795.
30. Dobert N, Menzel C, Berner U, et al. Positron emission tomography in patients with Hodgkin's disease: correlation to histopathologic subtypes. *Cancer Biother Radiopharm.* 2003;18:565–571.
31. Bangerter M, Moog F, Buchmann I, et al. Whole-body 2-[$^{18}$F]-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for accurate staging of Hodgkin's disease. *Ann Oncol.* 1998;9:1117–1122.
32. Moog F, Bangerter M, Diederichs CG, et al. Extranodal malignant lymphoma: detection with FDG PET versus CT. *Radiology.* 1998;206:475–481.
33. Kotzerke J, Guhlmann A, Moog F, Frickhofen N, Reske SN. Role of attenuation correction for fluorine-18 fluorodeoxyglucose positron emission tomography in the primary staging of malignant lymphoma. *Eur J Nucl Med.* 1999;26:31–38.
34. Moog F, Kotzerke J, Reske SN. FDG PET can replace bone scintigraphy in primary staging of malignant lymphoma. *J Nucl Med.* 1999;40:1407–1413.
35. Najjar F, Hustinx R, Jerusalem G, Fillet G, Rigo P. Positron emission tomography (PET) for staging low-grade non-Hodgkin's lymphomas (NHL). *Cancer Biother Radiopharm.* 2001;16:297–304.
36. Brunning RD, Bloomfield CD, McKenna RW, Peterson LA. Bilateral trephine bone marrow biopsies in lymphoma and other neoplastic diseases. *Ann Intern Med.* 1975;82:365–366.
37. Rosenberg SA. Hodgkin's disease of the bone marrow. *Cancer Res.* 1971;31: 1733–1736.
38. Coller BS, Chabner BA, Gralnick HR. Frequencies and patterns of bone marrow involvement in non-Hodgkin lymphomas: observations on the value of bilateral biopsies. *Am J Hematol.* 1977;3:105–119.