Lymphoma is the most common form of hematological malignancy, or “blood cancer,” in the developed world. Subtypes of lymphoma differ in molecular characteristics and biologic behavior. Compared with Western regions, Asian countries have been reported to have higher rates of non-Hodgkin’s lymphoma (NHL) and a low incidence of Hodgkin’s lymphoma.¹ On the basis of the clinical characteristics, this entity is divided into aggressive and indolent types. The most important factors influencing therapeutic decisions and prognosis are histologic subtype and extent of disease.²

Bone marrow biopsy (BMB) is an important part of the routine staging of lymphoma. Bone marrow (BM) involvement by lymphoma confers advanced-stage disease and may affect both treatment and prognosis. Histologic evidence of lymphoma in the BM is found in approximately 50% to 80% of patients with low-grade and 25% to 40% of high-grade NHL.³

¹8F-fluorodeoxyglucose (¹⁸F-FDG) positron emission tomography (PET) in the staging and restaging of patients with lymphoma has a median sensitivity of 90.3% and a median specificity of 91.1%, respectively.⁴ In another meta-analysis, a good, but not excellent correlation was demonstrated between ¹⁸F-FDG PET focal uptakes and BMB in the detection of BM involvement in the staging of patients with malignant HL and NHL lymphoma.⁵ However, it is still under discussion whether ¹⁸F-FDG PET (or PET/CT, computed tomography) can reduce the need for staging iliac BMB.⁶,⁷ So, we further analyzed the accuracy of ¹⁸F-FDG PET or PET/CT in detecting BM infiltration in aggressive (high grade) and indolent (low grade) NHL.

MATERIALS AND METHODS

Data Search

A search of the biomedical literature was performed by 2 researchers (Y.K.C. and C.H.K.) working independently, using the PubMed/MEDLINE and EBM Review search engines to identify studies involving human subjects (Fig. 1). Each researcher used searches F1 with last update of May 2010. They used the combination of search terms “lymphoma” “bone marrow,” and “positron emission tomography.” There was no language restriction. Additional studies were manually searched using the references of the retrieved articles. A total of 163 potential studies were retrieved from these searches.

Data Selection

Studies were eligible for inclusion based on the following criteria: (1) they evaluated lymphoma staging, non-Hodgkin’s lymphoma, including aggressive and/or low-grade (indolent) lymphoma, (2) bone involvement and/or BM infiltration, and (3) FDG PET and/or PET/CT images. Studies were excluded based on the following criteria: (1) only Hodgkin’s lymphoma, (2) non-Hodgkin’s lymphoma, without further description of subtype, (3) totals of true positives, false positives, true negatives, and false negatives were not provided, and (4) no data from a subanalysis were provided. Unpublished data and conference proceedings were not included. On the basis of these criteria, 8 studies were eligible for this study.

Data Extraction

Two reviewers independently assessed the methodological quality of the selected studies. The criteria list recommended by the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests was used. Some items on the list...
were modified for this specific review. The complete criteria list used is presented in Table 1. Internal validity criteria (IV) were scored as “positive” (adequate methods), “negative” (inadequate methods, potential bias), or “unclear” if insufficient information had been provided on a specific item. External validity criteria (EV) were assessed to evaluate generalizability. Standard performance of FDG PET or PET/CT was scored positive when the type of PET or PET/CT camera, the dose of FDG, the time between injection and scanning, and the method of reconstruction were described. The criteria for external validity were scored positive if sufficient information was provided to judge generalizability of findings. After the consensus meeting, we decided to score unclear scores as negative. Agreement between both reviewers was quantified by Cohen’s $\kappa$. Quality scores were expressed as a percentage of the maximum score. Subtotals were calculated for internal (maximum 6) and external (maximum 6) validity separately.

Statistics Analysis

Data on sensitivity, specificity, positive predictive value, and negative predictive value of FDG PET or PET/CT in the detection of BM infiltration were calculated from the original numbers given in the publications. The datasets were pooled by adding the TP, FP, TN, and FN results from all relevant studies and finding the sensitivity and specificity for the combined data. A 95% confidence interval was constructed for these estimates by assuming that each of the sensitivity and specificity results was a simple proportion from a normal distribution. Overall weighted average for sensitivity and specificity was calculated for comparison with the results of the pooled data using random effect model. When estimation of sensitivities and specificities for an individual study was a least one zero cell, a correction of 1/2 was added to every cell for that study to make the estimators defined. Exploring heterogeneity other than threshold effect was performed using I-square which measures the degree of heterogeneity between studies. Figures 2 and 3 show the moderate inconsistency level of I-squares of sensitivities and specificities in aggressive and indolent non-Hodgkin’s lymphoma, respectively. In addition, exploring heterogeneity due to threshold effect was performed using Spearman correlation coefficient. In aggressive non-Hodgkin’s studies, the threshold effect was not existent ($P = 0.872$). There was a threshold effect among indolent group studies. We attempted to fit each set of data to a summary receiver operating characteristic (sROC) curve and the area under sROC curve was calculated. The maximum joint sensitivity and specificity (Q* index) that measured the overall diagnostic accuracy was estimated. Q* is the point where the sensitivity and specificity are equal. An sROC curve is used when the slope of the linear regression is within a prespecified range ($0.5–0.5$). When applicable, the mean threshold for each group of studies was determined, and the sensitivity and specificity at that point on the curve were

![FIGURE 1. Selection of studies. Inclusion criteria: (1) lymphoma staging, non-Hodgkin’s lymphoma, including aggressive and/or low grade (indolent) lymphoma, (2) bone involvement and/or BM infiltration, (3) FDG PET and/or PET/CT images. Exclusion criteria is as follows: (1) only Hodgkin’s lymphoma, (2) non-Hodgkin’s lymphoma, not further describe subtype, (3) totals of true positives, false positives, true negatives, and false negatives were not provided, and (4) no data from a subanalysis were provided.](image)

![TABLE 1. Criteria List Used to Assess the Methodological Quality of the Studies](image)
provided. Overall values were also obtained by pooling of datasets, along with determining weighted averages for each of these sets of data. Statistical analyses were executed using Meta-Disc, a free statistical software package, version 1.4 (Unit of Clinical Biostatistics, Ramón y Cajal Hospital, Madrid, Spain).

RESULTS

Literature Search

A total of 163 studies about initial staging of lymphoma with FDG PET and associated with bone involvement were identified (Fig. 1). After reviewing the titles and abstracts, 155 studies were excluded. These studies included Hodgkin’s lymphoma, reviews, case reports, studies reporting on the use of FDG PET for response evaluation to chemotherapy. Of the remaining 12 studies, data of one study did not differentiate between Hodgkin’s and non-Hodgkin’s lymphoma, data of 2 studies did not classify non-Hodgkin’s lymphoma into high grade and low grade, and 1 study was excluded because of insufficient information to construct a 2 × 2 table. Eight studies met the inclusion criteria (Table 1). The characteristics of the included studies are presented in Table 2.
of aggressive non-Hodgkin’s lymphoma, the median sensitivity was 24.5% (range, 0%–100%) (Table 3). The summary specificity was 100% (range, 88%–100%) (Table 3). The summary (pooled) true-positive rate (sensitivity) was 46% (Table 4) and the summary of false-positive rate was 11.2%. The maximum joint sensitivity and specificity, a global measure of diagnostic accuracy, was 81.3%. Among the studies with patient-based data of indolent non-Hodgkin’s lymphoma, the median sensitivity was 24.5% (range, 0%–100%) (Table 3). The summary (pooled) true-positive rate (sensitivity) was 74% (range, 0%–100%) (Table 3). The summary specificity was 94% (range, 0%–100%) (Table 3). The summary sensitivity was found to be higher in patients with non-Hodgkin’s aggressive lymphoma compared with patients with non-Hodgkin’s indolent lymphoma. The summary specificity was 74% and the summary specificity was 84%. The summary sensitivity was found to be slightly higher in patients with non-Hodgkin’s indolent lymphoma compared with patients with non-Hodgkin’s aggressive lymphoma.

**Methodological Quality Assessment**

Methodological quality was assessed by 12 items for each of the 12 selected studies. There was disagreement in 40 of 144 scores with a Cohen’s k of 0.70. Main disagreement was in the questions IV3 and IV5. Disagreements were caused by reading errors and differences in interpretation. The scores for internal and external validity of the 12 selected studies are presented in Table 2. All studies had a valid reference test, but some studies (37.5%) did not describe whether the reference test was interpreted without knowledge of the FDG PET findings. In 6 (75%) of the 8 studies, verification bias was avoided because patients were selected for assessment by the reference test independently of the FDG PET results (IV4). Four studies were prospective (50%), and in 4 studies (50%), patients entered the study consecutively. In all of the selected studies (100%), primary stage of disease was included. In all studies (100%), the inclusion criteria were described, and only in a minority of studies were the exclusion criteria described. The total score for the combined internal and external validity, expressed as a fraction of the maximum score, ranged from 58% to 75%, with a median of 68.9%. Seven of the 8 studies had a total score above 60%.

**Diagnostic Accuracy of FDG PET or PET/CT**

The data of each study and the results of the statistical pooling are presented in Table 3. Among the studies with patient-based data of aggressive non-Hodgkin’s lymphoma, the median sensitivity was 79% (range, 46%–100%) and the median specificity was 94% (range, 0%–100%) (Table 3). The summary (pooled) true-positive rate (sensitivity) was 74% (Table 4) and the summary of false-positive rate was 11.2%. The maximum joint sensitivity and specificity, a global measure of diagnostic accuracy, was 81.3%. Among the studies with patient-based data of indolent non-Hodgkin’s lymphoma, the median sensitivity was 24.5% (range, 0%–90%) and the median specificity was 100% (range, 88%–100%) (Table 3). The summary (pooled) true-positive rate (sensitivity) was 74% (Table 4) and the summary false-positive rate was 4.5%. The maximum joint sensitivity and specificity, a global measure of diagnostic accuracy, was 75.6%.
The meta-analysis by Pakos et al. reported 13 eligible non-overlapping studies: 4 studies recruited patients with HD, 3 studies had patients with NHL, and 6 studies had mixed populations. The weighted rates showed significantly better sensitivity in studies with HD than in those with NHL patients. However, in NHL, there was a clear difference in sensitivity depending on the histologic type. On the basis of the available data, $^{18}$F-FDG PET identified 16 of 21 cases of BM involvement (76.2%) from large lymphocytic, large B-cell, Burkitt, and centroblastic lymphocytic lymphomas, whereas it detected only 16 of 53 cases with BM involvement (30.2%) from less aggressive histologic types (follicular, mantle cell, marginal zone, small lymphocytic lymphomas, and mucosa-associated lymphoid tissue). Otherwise, the study of Paone et al. revealed that 21 patients with diffuse large B-cell lymphoma had BM involvement. Only 10 patients (45%) had abnormal BM FDG uptake, 6 of the 7 with a prominent component of large transformed lymphoid cells, and 4 of the 14 with lymphoid infiltrates composed of small cells. The study of Ngew et al. showed maximum standardized uptake value >10 may predict for an aggressive histology. In a patient with an indolent lymphoma, sites with standardized uptake value >10 suggest the possibility of transformation or the possibility of presence an aggressive component in addition to what is suggested by the histology. Except for indolent B-NHL, PET scans have a good overall negative predictive value in excluding lymphomatous BM involvement.

Unilateral BMB is the standard approach in the staging of the bone marrow. It has been recognized as an imperfect tool for a long time. Several large studies have consistently shown that a unilateral iliac crest trephine biopsy is an unreliable method of detecting marrow lymphoma, especially in high-grade NHL. Studies examining the yield of a bilateral biopsy have shown that a unilateral biopsy would miss 20% of cases compared with a bilateral biopsy, showing the yield of a bilateral biopsy have shown that a unilateral biopsy would miss 20% of cases compared with a bilateral biopsy.

The majority of the studies included a mix of patients with Hodgkin disease, non-Hodgkin lymphoma, and non-Hodgkin lymphoma with different cell types. Studies reported on B/BM lesions together and did not try to make a clear-cut distinction between bony involvement and BM involvement in every patient and for each lesion. Furthermore, due to the nature of this disease, biopsy results were available in only a few studies; the majority had to rely on clinical follow-up, including a variety of imaging modalities and clinical examinations, not all of which were performed in the same manner in all the studies. The use of an imperfect reference standard, together with variability in the quality of the primary studies, introduces important limitations for the interpretation of the results. The majority of the studies included a mix of patients with Hodgkin disease, non-Hodgkin lymphoma, and non-Hodgkin lymphoma with different cell types. Studies reported on B/BM lesions together and did not try to make a clear-cut distinction between bony involvement and BM involvement.

### TABLE 4. Meta-analysis of Sensitivity and Specificity Data

<table>
<thead>
<tr>
<th>Type of Scan</th>
<th>Type of NHL</th>
<th>No.</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>Pooled Sensitivity (95% CI)</th>
<th>Pooled Specificity (95% CI)</th>
<th>Accuracy (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PET</td>
<td>Aggressive</td>
<td>134</td>
<td>37</td>
<td>7</td>
<td>77</td>
<td>13</td>
<td>0.74 (0.62–0.86)</td>
<td>0.92 (0.86–0.98)</td>
<td>0.85 (0.79–0.91)</td>
</tr>
<tr>
<td>PET/CT</td>
<td>Aggressive</td>
<td>237</td>
<td>67</td>
<td>29</td>
<td>117</td>
<td>24</td>
<td>0.74 (0.65–0.83)</td>
<td>0.80 (0.74–0.87)</td>
<td>0.78 (0.72–0.83)</td>
</tr>
<tr>
<td>PET or PET/CT</td>
<td>Aggressive</td>
<td>321</td>
<td>67</td>
<td>36</td>
<td>194</td>
<td>24</td>
<td>0.74 (0.65–0.83)</td>
<td>0.84 (0.80–0.89)</td>
<td>0.81 (0.77–0.86)</td>
</tr>
<tr>
<td>PET or PET/CT</td>
<td>Indolent</td>
<td>156</td>
<td>26</td>
<td>7</td>
<td>92</td>
<td>31</td>
<td>0.46 (0.33–0.59)</td>
<td>0.93 (0.88–0.98)</td>
<td>0.76 (0.69–0.82)</td>
</tr>
</tbody>
</table>

We compared the different accuracy rate between Aggressive and Indolent using PET or PET/CT scan. There were no statistically significant differences in 2 groups ($P = 0.1507$).

False-positive BM involvement on the $^{18}$F-FDG PET scan due to chemotherapy, granulocyte colony-stimulating factor administration, infection/inflammation, and hyperplastic marrow must be excluded as they may increase the $^{18}$F-FDG uptake and lead to a false-positive $^{18}$F-FDG PET scan. False-negative BM involvement on the $^{18}$F-FDG PET scan may be due to relatively low FDG uptake per cell or to diffuse, low-density marrow involvement by tumor. In patients with diffuse large B-cell lymphoma, the lack of FDG uptake in patients with lymphoid infiltrates composed of small cells can be attributed to a lack of uptake by the cells of these infiltrates which are small atypical lymphocytes with only rare large transformed lymphoid cells.

Routine reading of CT provided the correct anatomic localization of FDG-avid lesions and has a low yield in detecting bone/born marrow lesions, because criteria for disease involvement by CT scan are usually based on the size of a lesion. Schaefer et al. examined a selected population of 50 lymphoma patients (28 NHL) with $^{18}$F-FDG avid bone lesions on PET/CT. On CT, only 32 of the 193 lesions (16.6%) were detected without the PET information. In 161 lesions (83.4%), only focal increased FDG uptake in the bone was observed on PET/CT, without morphologic alteration of osseous structures on CT images. In patients with positive FDG PET/CT and negative BMB, CT-guided BMBs at the involvement sites detected by the FDG PET/CT scan were recommended.

There are several potential limitations to conducting a meta-analysis of diagnostic tests. The presence of clinical heterogeneity (heterogeneity originated by the inclusion of patients at different stages of disease and other clinical characteristics) affects the generalizability of the results and it is not necessarily ruled out by the lack of statistical heterogeneity. It is important to note that the majority of the studies included a mix of patients with Hodgkin disease, non-Hodgkin lymphoma, and non-Hodgkin lymphoma with different cell types. Studies reported on B/BM lesions together and did not try to make a clear-cut distinction between bony involvement and BM involvement. Furthermore, due to the nature of this disease, biopsy results were available in only a few studies; the majority had to rely on clinical follow-up, including a variety of imaging modalities and clinical examinations, not all of which were performed in the same manner in all the studies. The use of an imperfect reference standard, together with variability in the quality of the primary studies, introduces important limitations for the interpretation of this meta-analysis. In addition, the verification bias potentially present in the primary studies cannot be fully addressed in a meta-analysis. Nevertheless, despite these limitations, meta-analytic techniques have been very useful for demonstrating the significant role of FDG PET or PET/CT imaging in the diagnosis and staging of several malignancies.

The results from this literature review and meta-analysis suggest that the diagnostic accuracy of FDG PET or PET/CT is slightly higher but without significantly statistical difference ($P = 0.1507$) in patients with non-Hodgkin’s aggressive lymphoma (accuracy: 81%) than in those with non-Hodgkin’s indolent lymphoma (accuracy: 76%) (Table 4). The overall high specificity of FDG PET
or PET/CT in patients with non-Hodgkin’s aggressive lymphoma and indolent lymphoma were 84% and 93%, respectively. FDG PET or PET/CT scan shows potential to detect BM involvement in non-Hodgkin’s aggressive lymphoma, which would otherwise be missed by iliac crest BMB. Furthermore, FDG PET or PET/CT can be used to directly guide the site of the biopsy, when PET demonstrates BM involvement in a different site. However, the overall sensitivity of FDG PET or PET/CT in patients with non-Hodgkin’s indolent lymphoma was only 46%. In FDG-negative cases of indolent lymphoma, a BMB is probably still warranted.

REFERENCES


 AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1—Please provide the month, and volume and issue numbers for this manuscript.

AQ2—Please expand the following terms: TP, FP, TN, and FN.

AQ3—Please note that references have been renumbered to make their text citations sequential per journal style.

AQ4—Please note that Figs. 4 and 5 are not cited in the manuscript. Please cite them at appropriate places in the manuscript.

AQ5—Please check whether the affiliations are OK as typeset.