End-Therapy Positron Emission Tomography for Treatment Response Assessment in Follicular Lymphoma: A Systematic Review and Meta-analysis

Junhee Pyo1, Kyung Won Kim2,4,5, Heather A. Jacene2,4, Christopher G. Sakellis2,4, Jennifer R. Brown3, and Annick D. Van den Abbeele2,4

Abstract

**Purpose:** Use of 2[18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in post-chemotherapy response assessment in follicular lymphoma is still a controversial issue. Here, we conducted the first systematic review and meta-analysis to determine the predictive value of FDG-PET in predicting outcome after chemotherapy of follicular lymphoma.

**Experimental Design:** Comprehensive literature search in Ovid-MEDLINE and EMBASE databases was performed to identify studies which evaluate predictive value of end-therapy PET and/or computed tomography (CT) in patients with follicular lymphoma. To quantitatively compare the predictive value of PET and CT, pooled hazard ratios (HRs) comparing progression-free survival (PFS) between patients with positive and negative results were adopted as the primary indicators for meta-analysis. To explore the efficiency in determining complete remission (CR), pooled CR rates of PET- and CT-based response criteria were calculated. Pooling of these parameters was based on the random-effects model.

**Results:** Review of 285 candidate articles identified eight eligible articles with a total of 577 patients for qualitative review and meta-analysis. The pooled HRs of end-therapy PET and CT were 5.1 [95% confidence interval (CI), 3.7–7.2] and 2.6 (95% CI, 1.2–5.8), respectively, which implies that PET is more predictive of PFS after chemotherapy than CT. The pooled CR rates of PET- and CT-based response criteria were 75% (95% CI, 70–79%) and 63% (95% CI, 53–73%), respectively, which implies that PET is more efficient in distinguishing CR (without residual disease) from other states with residual disease. In addition, qualitative systematic review indicates the same findings.

**Conclusions:** Consistent evidence favoring PET-based treatment assessment should be considered in the management of patients with follicular lymphoma. Clin Cancer Res; 19(23); 6566–77. ©2013 AACR.

Introduction

Follicular lymphoma, a common histologic type of non-Hodgkin lymphoma, is indolent with variable clinical behavior. It is regarded as an incurable disease, because follicular lymphoma almost inevitably relapses even after achieving good response to chemotherapy and patients with residual disease or early relapse after therapy show worse prognosis. Recently, Bachy E. et al reported that patients with a better response to first-line chemotherapy for follicular lymphoma (especially, complete response) show improved survival (1). Therefore, accurate response assessment at the end of first-line chemotherapy is very important to predict outcome of each patient and to identify patients with residual disease who may benefit from additional treatment.

The use of 2[18F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) or FDG-PET/computed tomography (FDG-PET and FDG-PET/CT are hereafter referred to as PET) for postchemotherapy response assessment has been the standard of care in the management of Hodgkin lymphomas and diffuse large B-cell lymphomas (2). Based on the strong evidence that negative results of an end-therapy PET scan indicate a better outcome and are highly predictive of progression-free survival (PFS) and overall survival (OS; ref. 3, 4), PET is particularly helpful to differentiate true residual disease from a treated fibrotic lesion when there is residual mass on CT after chemotherapy for Hodgkin lymphoma or diffuse large B-cell lymphomas.
Role of End-Therapy PET for Follicular Lymphoma

Translational Relevance

In this meta-analysis, the current evidence in the literature consistently shows that end-therapy PET is more predictive of PFS than CT and more efficient in determining complete remission and detecting residual disease in patients of follicular lymphoma treated with chemotherapy. PET is particularly helpful to differentiate metabolically active residual disease (true lesion) from residual mass, which is actually treated disease with fibrotic change (false-positive lesion). In follicular lymphoma, which is an indolent incurable lymphoma, although it is still unclear whether or not end-therapy PET is predictive of long-term overall survival, the consistent evidence favoring PET-based response criteria strongly suggests potential implications of PET in the management of patients with follicular lymphoma. These implications will need to be further explored in prospective clinical trials, particularly in the context of novel therapies such as oral kinase inhibitors.

Materials and Methods

Literature search strategy

A computerized search of Ovid-MEDLINE and EMBASE databases was conducted to find relevant studies published prior to February 28, 2012. We did not restrict the beginning date. The following search terms were used: ("follicular lymphoma") AND (response OR prognosis OR predict) AND (positron emission tomography OR computed tomography). Our search was restricted to human subjects and studies published in English. To expand the search, the bibliographies of articles that remained after the selection process were screened for other potentially suitable articles.

Inclusion criteria

Studies were eligible for inclusion if they evaluated the predictive value of PET after chemotherapy to predict outcome in patients with follicular lymphoma. Studies that satisfied all of the following criteria were included:

(i) Population: Patients with follicular lymphoma who underwent end-therapy PET after chemotherapy. We comprehensively included studies with various chemotherapeutic strategies.

(ii) Reference standard: Clinical follow-up with all available information, with or without pathologic confirmation of the suspected residual/relapsed/progressed lesion.

(iii) Study designs: All observational studies (retrospective or prospective) and clinical trials.

(iv) Outcomes: Survival analysis data such as HR or Kaplan–Meier curve of PFS comparing patients with positive and negative results of end-therapy PET.

Exclusion criteria

The exclusion criteria were: (i) case reports, case series, review articles, editorials, letters, comments, and conference abstracts; (ii) studies with insufficient survival analysis data to perform meta-analysis, (iii) studies that did not have data specific to follicular lymphoma, (iv) studies whose topics were outside the predictive value of PET in patients with follicular lymphoma; and (v) studies with overlapping patients and data. Two reviewers (K.W. Kim and J. Pyo) independently selected literature using a standardized protocol. Disagreements were resolved by discussion.

Quality assessment

The methodologic quality of the included studies was assessed by using Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) criteria, because it was developed for studies evaluating prognostic/predictive tumor markers or clinical variables by a time-to-event outcome such as PFS (22). Quality assessment was performed by consensus between two reviewers (K.W. Kim and J. Pyo).

Terminology

The term "chemotherapy" is too broad and nonspecific. Therefore, in this article, the term "chemotherapy" implies...
any type of systemic therapy for treatment of follicular lymphoma. We specified different types of systemic therapy as follows: (i) "conventional chemotherapy" stands for conventional regimens without rituximab or radionuclide-labeled antibody such as CHOP (cyclophosphamide + doxorubicin + vincristine + prednisone) and CVP (cyclophosphamide + vincristine + prednisone); (ii) "chemoimmuno-therapy" (CIT) stands for combinations of conventional chemotherapy and rituximab (R) such as R-CHOP or R-CVP; and (iii) "radioimmunotherapy" (RIT) stands for treatment with radionuclide-labeled antibody alone and/or conventional chemotherapy or CIT.

Slightly different definitions of positive and negative PET results after chemotherapy have been used across studies. In studies that used their own qualitative visual assessment, we adopted the definition of positive PET scan and negative PET scan of each individual study. In studies that used the revised 2007-IWG criteria, which is primarily based on qualitative assessment, "negative PET result" implies complete remission (CR) and "positive PET result" implies partial remission (PR), stable disease, or progressive disease (PD). Regarding CT-based response criteria, all studies used the 1999-IWG criteria, except a study by Le Dortz and colleagues (17), which used the 2007-IWG criteria. For the meta-analysis in this study, "negative CT result" stands for CR and/or complete remission/unconfirmed (CRu) and "positive CT result" stands for PR, stable disease, or PD. In this study, "metabolic CR" implies CR in PET-based response criteria and "anatomic CR" implies CR and/or CRu in CT-based response criteria.

End-therapy imaging could potentially serve as either a prognostic or predictive biomarker. In our study, the term "prognostic" means that a marker gives information on a patient’s final outcome regardless of therapy, whereas the term "predictive" means that a marker gives information on the outcome of a therapy (22). Because PFS is the main outcome of interest in our study, and is directly related to prior therapy, the term "predictive" is adopted for the role of end-therapy PET or CT throughout the article.

**Data extraction**

We extracted the available data for systematic review and meta-analysis from the included studies: (i) study characteristics: authors, year of publication, the location and period of studies, population size, and study design; (ii) demographic and clinical characteristics of the patients: age, sex, histologic grade of follicular lymphoma, and chemotherapeutic strategies; (iii) imaging characteristics: types of PET scanner, response criteria of PET and CT; and (iv) outcome characteristics: survival analysis data including HR or Kaplan–Meier curve of PFS comparing patients with positive and negative results of PET and/or CT were extracted for the primary outcome of interest. In addition, we extracted number of pairs of PET results and CT results including PET−/CT−, PET+/CT+, PET−/CT+, and PET+/CT+ cases to calculate metabolic CR rate, anatomic CR rate, and discordance rates of PET-based and CT-based response criteria, as detailed below in "Data synthesis and statistical analysis.

Two reviewers (K.W. Kim and J. Pyo) extracted data from the studies in consensus.

**Data synthesis and statistical analysis**

To quantitatively compare the predictive value of PET and CT, pooled HRs comparing PFS between patients with positive and negative results were adopted as the primary indicators of meta-analysis to predict the outcome of patients after chemotherapy. Pooled HRs were calculated to show how many times higher the probability of progression was in patients with positive results on end-therapy PET or CT as compared with those who had negative results. If the original literature did not report the HR for PFS and its variance, we estimated HRs and their 95% confidence intervals (CI) in each study using all available information from reported statistics according to the methods proposed by Tierney et al. (23) by (i) using the observed and expected events of the patients with positive and negative results and (ii) using the extracted survival probabilities in the Kaplan–Meier curves at specific time points. Funnel plots were used for visual assessment of publication bias, and Beg and Egger tests were used to detect asymmetry (24).

In addition, we pooled data of metabolic CR rate of PET-based response criteria and anatomic CR rate of CT-based response criteria to explore which modality is more efficient in differentiating CR (i.e., absence of residual disease) and PR/stable disease/PD (i.e., residual disease). In order to explore the cause of the discrepancy between the rates of metabolic CR and anatomic CR, we pooled data on the discordance rate of PET- and CT-based response criteria. There were two types of discordant pairs: (i) probability of a PET-positive test result, given that CT shows a negative result (referred as PET+/CT− rate), and (ii) probability of a PET-negative test result, given that CT shows a positive result (referred as PET−|CT+ rate). We constructed a 2 × 2 table, which contained cases of PET−/CT−, PET+/CT+, PET−/CT+, and PET+/CT+ results. When 0 counts occurred in any of the cells in the 2 × 2 table, 0.5 was added to all the cell values for a continuity correction. Then, the rates of metabolic and anatomic CR, PET+/CT−, and PET−|CT+ were calculated using the following equations:

\[
\text{Metabolic CR rate} = \frac{(\text{PET−/CT− cases} + \text{PET+/CT+ cases})}{\text{all cases}}
\]

\[
\text{Anatomic CR rate} = \frac{(\text{PET−/CT− cases} + \text{PET+/CT− cases})}{\text{all cases}}
\]

\[
\text{PET+/CT− rate} = \frac{(\text{PET+/CT− cases})}{(\text{PET+/CT− cases} + \text{PET−/CT− cases})}
\]

\[
\text{PET−/CT+ rate} = \frac{(\text{PET−/CT+ cases})}{(\text{PET−/CT+ cases} + \text{PET+/CT+ cases})}
\]

The pooled HR with 95% CIs was calculated using the DeSimonean–Liard random-effects model (23). Heterogeneity was assessed by using the Cochran Q test and the I² statistics (25). An I² value greater than 50% was considered to indicate substantial heterogeneity. P value less than 0.10
in a Q test was considered significant for heterogeneity. In order to test the robustness of the meta-analysis results, a sensitivity analysis was performed by recalculating the pooled estimates after excluding each individual study. For the statistical analysis, Stata version 10 (StataCorp LP) was used.

Results

Literature search

Our literature search process is illustrated in Fig. 1. In the literature search in Ovid-MEDLINE and EMBASE databases, after removing duplicates, 236 articles were screened for eligibility. Of these, 206 articles were excluded after review of the titles and abstracts: 58 review articles; 34 case reports/series; 55 letters, editorials, or conference abstracts; and 59 original articles whose topics are outside the prognostic/predictive value of PET in follicular lymphoma. Full texts of the remaining 30 articles were retrieved. Search of bibliographies of these articles did not find additional eligible studies. The following 22 articles were further excluded after reviewing the full texts: 13 articles whose topics were outside the field of interest of our study (three studies evaluating efficacy of chemotherapeutic strategy or stem-cell transplantation, two studies evaluating role of end-therapy CT, five studies evaluating diagnostic role of PET in staging or surveillance, and three studies that do not have data specific to follicular lymphoma), eight studies that had no or insufficient survival data for the meta-analysis (9, 10, 26–31), and one study with patient overlap (32). The remaining eight studies were used for qualitative and quantitative review (14–21).

Study characteristics and quality

The basic characteristics of the eight studies are summarized in Table 1. Two studies were prospective multicenter trials (15, 20), and six studies were retrospective observational studies. Most of the included patients had untreated follicular lymphoma, although one study recruited patients with relapsed/refractory follicular lymphoma (18). Five studies included patients with follicular lymphoma grades 1, 2, and/or 3A, one study included all grades, and two studies did not mention the histologic grades for inclusion. The majority of included patients underwent CIT such as R-CHOP or R-CVP according to the policy of institutions or oncologists, whereas participants in one study underwent RIT (18). In terms of reference standard, all included studies reported that they used all available clinical information including imaging follow-up as proof of disease recurrence/progression, whereas only two studies specifically mentioned histologic assessment for evaluation of recurrence/progression (14, 19).

With regard to the type of PET scanner, five studies used combined PET/CT (15, 17–20), one study used standalone PET (21), and two studies used both (14, 16). The criteria of treatment response assessment of PET results somewhat varied across the studies, as summarized in Table 1. However, there was no substantial difference in clinical decision making, because all response criteria were basically qualitative assessments and generally followed the consensus guidelines of the Imaging Subcommittee of the International Harmonization Project in Lymphoma (2).
Table 1. Characteristics of the included studies

<table>
<thead>
<tr>
<th>First author (reference number)</th>
<th>Location (study period)</th>
<th>No. of patients analyzed</th>
<th>Study design</th>
<th>Clinical setting</th>
<th>Histologic grade</th>
<th>Chemotherapeutic regimen</th>
<th>Type of PET scanner</th>
<th>PET response criteria</th>
<th>CT response criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bishu et al. (14)</td>
<td>Nebraska, USA (1999–2006)</td>
<td>16</td>
<td>Retrospective</td>
<td>Untreated follicular lymphoma</td>
<td>1, 2</td>
<td>1, 2 Conventional CTx or CIT</td>
<td>PET or PET/CT</td>
<td>Qualitative assessment</td>
<td>1999-IWG criteria</td>
</tr>
<tr>
<td>Dupuis et al. (15)</td>
<td>Multicenters in Italy and France (2007–2009)</td>
<td>104</td>
<td>Prospective</td>
<td>Unresected follicular lymphoma</td>
<td>1, 2, and 3A</td>
<td>CIT</td>
<td>PET/CT</td>
<td>Deauville 5-point scale</td>
<td>PET/CT</td>
</tr>
<tr>
<td>Janikova et al. (16)</td>
<td>Brno, Czech Republic (2002–2007)</td>
<td>93</td>
<td>Retrospective</td>
<td>Untreated and relapsed follicular lymphoma</td>
<td>1, 2, and 3A</td>
<td>CIT</td>
<td>PET/CT</td>
<td>Qualitative assessment and SUV &gt; 2.5</td>
<td>1999-IWG criteria</td>
</tr>
<tr>
<td>Le Dortz et al. (17)</td>
<td>Rennes, France (NA) (2007–2008)</td>
<td>45</td>
<td>Retrospective</td>
<td>Untreated follicular lymphoma</td>
<td>1, 2, and 3A</td>
<td>CIT</td>
<td>PET/CT</td>
<td>2007-IWG criteria</td>
<td>NA</td>
</tr>
<tr>
<td>Lopci et al. (18)</td>
<td>Multicenters in Italy (NA) (2002–2007)</td>
<td>59</td>
<td>Retrospective</td>
<td>Relapsed or refractory FL</td>
<td>All grades</td>
<td>RIT</td>
<td>PET/CT</td>
<td>2007-IWG criteria</td>
<td>NA</td>
</tr>
<tr>
<td>Trotman et al. (20)</td>
<td>Multicenters in 25 countries (2004–2007)</td>
<td>122</td>
<td>Prospective</td>
<td>Relapsed follicular lymphoma</td>
<td>1, 2, and 3A</td>
<td>CIT</td>
<td>PET/CT</td>
<td>2007-IWG criteria</td>
<td>NA</td>
</tr>
<tr>
<td>Zinzani et al. (21)</td>
<td>Bologna, Italy (2002–2004)</td>
<td>45</td>
<td>Retrospective</td>
<td>Untreated follicular lymphoma</td>
<td>NA</td>
<td>NA</td>
<td>PET/CT</td>
<td>Qualitative assessment and greater liver uptake</td>
<td>1999-IWG criteria</td>
</tr>
</tbody>
</table>

NOTE: Information obtained by communication with authors of the original literature.

Abbreviations: CTx, chemotherapy; CIT, chemoimmunotherapy; RIT, radioimmunotherapy; IWG, international working group; NA, not available.
The quality of included studies as assessed by the REMARK guidelines was overall moderate, satisfying 12 to 19 items out of the total 20 items (median 16, interquartile range 15–19). Reporting was especially insufficient in the items of analysis and presentation of the guidelines, which emphasized detailed reporting of methodology and results of multivariate analyses. Only three studies performed univariate and multivariate survival analysis by the Cox hazard regression model (18–20), whereas other studies performed only univariate analyses using Kaplan–Meier curves or the rate of PFS. In addition, one study focused on the diagnostic accuracy of PET and CT for residual disease detection without performing dedicated survival analysis, but it provided raw survival data of each patient within the article that enabled estimation of the HR (14).

Qualitative review

The eight included studies presented predictive parameters of PET- and/or CT-based response criteria and diagnostic accuracy of end-therapy PET and/or CT for residual disease detection, as summarized in Table 2. The studies differ substantially in population size, outcome, and interpretation, but all studies favored PET for the treatment response assessment of follicular lymphoma compared with CT. All studies that directly compared rates of PFS and relapse between PET and CT reported that PET was more efficient in predicting PFS of patients and relapse of disease compared with CT. In two studies that evaluated the rate of OS, patients with a negative PET scan after completion of CT also had a significantly higher OS than those with a positive PET scan, but they did not provide OS data in relation to CI-based response criteria (15, 20). With regard to detection of residual disease, the diagnostic value of PET was evaluated in three studies, and consistently reported high sensitivity (94%–100%), specificity (93%–97%), positive predictive value (PPV, 89%–94%), and negative predictive value (NPV, 93%–100%; refs. 14, 17, 19). However, the diagnostic value of CT was hampered by low specificity (51%) and PPV (43%) despite high sensitivity (100%) and NPV (100%; ref. 17).

Survival analysis: meta-analysis

In our meta-analysis evaluating the predictive value of end-therapy PET based on eight studies with a total 577 patients, the pooled HR comparing PFS between patients with positive and negative results was 5.1 (95% CI, 3.7–7.2), which implies that PFS in patients with a positive PET result was significantly shorter than those with a negative result (Fig. 2). The meta-analysis evaluating the predictive value of end-therapy CT, based on four studies with a total of 276 patients, also demonstrates that PFS in patients with a positive result (PR/stable disease/PD) was significantly shorter than those with a negative result (CR), but with a lesser degree of pooled HR of 2.6 (95% CI, 1.2–5.8) compared with the estimates of PET data (Fig. 2).

No substantial heterogeneity was found in the pooled HR synthesis of PET data ($I^2 = 0\%$, $P = 0.438$, Q test), whereas heterogeneity was present in pooling HR of CT data ($I^2 = 60\%$, $P = 0.060$, Q test). This heterogeneity could be attributed to the result of the study by Janikova and colleagues (HR, 5.7; 95% CI, 2.8–11.7) wherein the HR was much higher than the rest of the other studies that ranged from 1.4% to 2.6%. Among the four studies analyzed, only the study by Janikova et al. included patients with untreated and relapsed follicular lymphoma, whereas the other three studies included patients with untreated follicular lymphoma, which might explain the heterogeneity. In the sensitivity analysis when excluding the study by Janikova et al., heterogeneity was resolved with recalculated pooled HRs of 1.6 (95% CI, 0.9–2.9; $I^2 = 0\%$). The recalculated pooled HRs of CT data (1.6–3.8) were still lower than those of the PET data (4.6–6.0), which is concordant with the direction of the original study result (Supplementary Table S1). There was no publication/reporting bias as the funnel plots were symmetric, and the results of the Begg and Egger tests were not significant (Fig. 3).

CR rates and discordance rates: meta-analysis

The meta-analysis results of six studies that reported direct comparisons of PET-based response criteria and CT-based response criteria are illustrated in Fig. 4. Pooled metabolic CR rate based (75%–95% CI, 70%–79%) was higher than the pooled rate of anatomic CR (63%; 95% CI, 53%–73%), which could be attributed to discordance between PET-based response criteria and CT-based response criteria. Between two types of discordant pairs, pooled rate of PET−|CT+ (57%; 95% CI, 49%–65%) was higher than the pooled rate of PET+|CT− (15%; 95% CI, 11%–19%). These results imply that PET might be more efficient in differentiating CR (i.e., absence of residual disease) from PR/stable disease/PD (i.e., residual disease), mainly because 57% of residual masses detected on CT were negative on PET (i.e., false-positive lesion on CT).

No substantial heterogeneity was found in synthesizing rates of metabolic CR, PET−|CT+, and PET+|CT− in the $I^2$ statistics and Cochran Q test, whereas heterogeneity was present in pooling rate of anatomic CR ($I^2 = 76\%$, $P = 0.001$, Q test). This heterogeneity could be attributed to the result of the study by Le Dortz et al. (38%–95% CI, 24%–53%) wherein the rate of anatomic CR was much lower than the rest of the other studies that ranged from 56% to 73%. Among the six studies analyzed, only the study by Le Dortz et al. used 2007-IWG criteria, whereas the other five studies used the 1999-IWG criteria for CT-based response assessment, which might explain the heterogeneity. Sensitivity analysis revealed that recalculating pooled estimates of rates of metabolic CR, anatomic CR, and discordance after excluding each study were robust and stable in the direction of the original study results (Supplementary Table S2).

Discussion

The goals of imaging response assessment after completion of chemotherapy for lymphoma are several-fold,
Table 2. Qualitative review of outcomes of the included studies

<table>
<thead>
<tr>
<th>First author (reference number)</th>
<th>Predictive value of imaging response criteria</th>
<th>Relapse rate</th>
<th>Residual disease detection</th>
<th>Analyzed parameters in meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bishu et al. (14)</td>
<td>None</td>
<td>NA</td>
<td>PET: (7%, 1/14) vs. PET+: (100%, 2/2)</td>
<td>Pooled HR of PET and CT, CR rates, and discordance rates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT: (11%, 1/9) vs. CT+: (29%, 2/7)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dupuis et al. (15)</td>
<td>K-M curve for PET data</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Two-year PFS rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PET: (87%) vs. PET+: (51%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CT: (83.2%) vs. CT+: (67.7%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Two-year OS rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PET: (100%) vs. PET+: (88%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CT (NA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Janikova et al. (16)</td>
<td>K-M curve for PET and CT data</td>
<td>NA</td>
<td>PET: (26%, 19/73) vs. PET+: (70%, 14/20)</td>
<td>Pooled HR of PET and CT, CR rates, and discordance rates</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CT: (29%, 21/72) vs. CT+: (57%, 12/21)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Le Dortz et al. (17)</td>
<td>K-M curve for PET data</td>
<td>NA</td>
<td>PET: SEN 100%, SPE 97%, PPV 92%, NPV 100%</td>
<td>Pooled HR of PET, CR rates, and discordance rates</td>
</tr>
<tr>
<td></td>
<td>• Median PFS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PET: (48 mo) vs. PET+: (17.2 mo)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lopci et al. (18)</td>
<td>K-M curve with uni- and multivariate HR for PET data</td>
<td>NA</td>
<td></td>
<td>Pooled HR of PET</td>
</tr>
<tr>
<td></td>
<td>• Three-year PFS rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PET: (40%) vs. PET+: (10%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lopci et al. (19)</td>
<td>K-M curve with uni- and multivariate HR for PET data</td>
<td>NA</td>
<td></td>
<td>Pooled HR of PET</td>
</tr>
</tbody>
</table>

(Continued on the following page)
Table 2. Qualitative review of outcomes of the included studies (Cont'd)

<table>
<thead>
<tr>
<th>First author (reference number)</th>
<th>Predictive value of imaging response criteria</th>
<th>Residual disease detection</th>
<th>Analyzed parameters in meta-analysis</th>
<th>Relapse rate</th>
<th>PFS or OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trotman et al. (20)</td>
<td>K-M curve with univariate HR and multivariate HR for PET and CT data</td>
<td>NA</td>
<td>Pooled HR of PET and CT CR rates and discordance rates</td>
<td>NA</td>
<td>42-month PFS rate</td>
</tr>
<tr>
<td>Zinzani et al. (21)</td>
<td>None</td>
<td>NA</td>
<td>Pooled HR of PET and CT CR rates and discordance rates</td>
<td>NA</td>
<td>Two-year PFS rate</td>
</tr>
</tbody>
</table>

NOTE: *Estimated from data provided within the article.
Abbreviations: K-M, Kaplan-Meier; PFS, progression-free survival; OS, overall survival; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; NA, not available.
including: to predict a patient’s outcome after chemotherapy; to quantify treatment response of tumor; and to detect residual disease (22). Therefore, the studies evaluating the role of end-therapy PET in patients with follicular lymphoma differ: some studies focused on its predictive value whereas other studies focused on the adequate treatment response assessment.

Despite this heterogeneity of imaging response assessment, the preexisting literature included in our quantitative meta-analysis consistently present the same findings, as follows: (i) PET is more predictive of patient outcome after chemotherapy than CT, as evidenced by a higher pooled HR of positive PET scan than positive CT scan (5.1 vs. 2.6), and (ii) end-therapy PET is more efficient in distinguishing CR (without residual disease) from PR/stable disease/PD (with residual disease) than CT, as evidenced by a higher pooled metabolic CR rate than pooled rate of anatomic CR (75% vs. 63%). In addition, qualitative systematic review indicates the same findings, as evidenced by (i) higher relapse rates in patients with positive PET scan compared with those with a positive CT scan (70%–100% vs. 29%–57%) and (ii) higher PPVs of PET for residual disease detection than those of CT (89%–94% vs. 43%) despite high NPVs of both PET and CT. The result of pooled PET/CT+ rate (57%, 95% CI, 49%–65%), which indicates that 57% of residual masses on CT should be considered as negative lesions on PET (i.e., false-positive lesion on CT), also

Figure 2. Forest plots to show the pooled HR of PET- and CT-based response criteria.

Figure 3. Funnel plots for visual assessment of the publication bias in synthesizing pooled HR. PET-based response criteria (A) and CT-based response criteria (B) show symmetrical plots, which indicate absence of substantial publication bias.
supports the superiority of PET in differentiating CR from PR/stable disease/PD compared with CT.

Despite these findings, no consensus guidelines have adopted PET for postchemotherapy response assessment in follicular lymphoma, and further therapeutic decisions solely based on end-therapy PET results are not accepted as standard of care because end-therapy PET has not been shown to be more predictive of OS than CT, in a disease with an indolent relapsing and remitting course (7, 11, 33). It is still unclear whether subsequent intervention to eradicate residual disease in patients with a positive result on end-therapy PET would achieve OS gain compared with watchful observation until progression occurs, given that follicular lymphoma has an indolent nature and tends to relapse frequently during its natural history (33).

Nevertheless, the vast majority of literature indicates the superiority of PET-based response criteria compared with CT-based response criteria in predicting PFS and determining the presence of residual disease. Moreover, experts recommended end-therapy PET for prognostic/predictive purposes based on recent studies in the 4th International Workshop on PET in Lymphoma (12). A main drawback to the use of PET compared with CT is high cost, warranting a further investigation of cost-effectiveness of PET-based response criteria. Another concern with PET is false-positive lesions. If PET is performed too early after completion of chemotherapy, an inflammatory process around the tumor cells can cause overestimation of glucose metabolism and misinterpretation, a phenomenon that justifies further investigation of the optimal timing of end-therapy PET (5, 34).

Considering the current evidence on the advantages and drawbacks of PET, the use of end-therapy PET for treatment response assessment might be justifiable for prognostic/
predictive purposes. However, the data suggesting the value of end-therapy PET certainly justify the design of studies of subsequent treatment for PET-positive residual disease after chemotherapy. With the advent of highly effective well-tolerated oral kinase inhibitors, the options for such therapy have become much more palatable for patients. Still, studies are required to determine if this type of intervention would ultimately benefit OS or, perhaps, have other benefits, such as avoiding transformation into high-grade lymphoma.

The strengths of our study include validated systematic review methods and the reporting of results according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (35). In addition, we extracted the maximum information from the included studies by thorough qualitative review and quantitative meta-analysis.

Our study had several limitations. Firstly, we used estimated HR from the limited data of the original studies using methods proposed by Tierney et al. (23). The estimated HR might be slightly different from the true value of HR. However, it was the best way to synthesize reliable meta-analytic data from limited original data, and the same methods were applied to the PET and CT data so that it should not affect the comparability between them. Secondly, we included as many of the relevant studies as possible, including studies that focused on the prognostic/predictive value of end-therapy PET and/or CT as well as studies that focused on the diagnostic accuracy of residual disease. In addition, the chemotherapy regimens, included patient characteristics (untreated vs. relapsed/refractory, pathologic grading, and disease burden), and the criteria of treatment response assessment of PET and CT varied to some degree across the studies. This variety of studies does mean that the clinical heterogeneity of the included studies could be an issue. Thus, we adopted the random-effects models for synthesizing meta-analytic data to deal with the heterogeneity of patients with follicular lymphoma. Statistically, heterogeneity was found only in pooling HR and rate of anatomic CR of CT data, and we did not perform further statistical methods to explore the source of heterogeneity such as meta-regression or subgroup analysis due to the small number of included studies. However, heterogeneity was resolved in the sensitivity analysis, which excluded each individual study and demonstrated robust and stable pooled estimates.

In conclusion, the current evidence in the literature consistently shows that PET-based response criteria are more predictive of PFS than CT-based criteria and more efficient in determining CR and detecting residual disease in patients with follicular lymphoma treated with chemotherapy. In order to adopt PET as a standard method for treatment response assessment, further evidence is warranted to determine whether or not end-therapy PET is predictive of long-term OS and whether PET-based therapeutic decisions can improve OS. Despite this unsolved controversy, the consistent evidence favoring PET-based response criteria should be considered in the management of patients with follicular lymphoma.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed by all authors.

Authors’ Contributions
Conception and design: J. Pyo, K.W. Kim
Development of methodology: J. Pyo
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Pyo, K.W. Kim, J.R. Brown, A.D. Van den Abbeele
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Pyo, K.W. Kim, J.R. Brown
Writing, review, and/or revision of the manuscript: J. Pyo, J.R. Brown, H. A. Jacene, C.G. Sakellis, J.R. Brown, A.D. Van den Abbeele
Study supervision: A.D. Van den Abbeele

Grant Support
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 1, 2013; revised September 9, 2013; accepted September 17, 2013; published OnlineFirst September 19, 2013.

References

Lopci E, Santi I, Derenzini E, Fonti C, Savelli G, Bertagna F, et al. FDG- \(15\)


Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. Eur J Cancer 2008;44:946–53. \(23\)

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8:16. \(24\)


Trials 2007;8:16. \(32\)

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 2007;8:16. \(33\)


End-Therapy Positron Emission Tomography for Treatment Response Assessment in Follicular Lymphoma: A Systematic Review and Meta-analysis

Junhee Pyo, Kyung Won Kim, Heather A. Jacene, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-13-1511

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2013/09/19/1078-0432.CCR-13-1511.DC1

Cited articles
This article cites 33 articles, 14 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/19/23/6566.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/19/23/6566.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.