Interim positron emission tomography (PET) scans in diffuse large B-cell lymphoma: an independent expert nuclear medicine evaluation of the Eastern Cooperative Oncology Group E3404 study

Short Title: Expert review of interim PET in DLBCL

Sandra J. Horning¹, Malik E. Juweid², Heiko Schöder³, Gregory Wiseman⁴, Alex McMillan⁵, Lode J. Swinnen⁶, Ranjana Advani¹, Randy Gascoyne⁷, Andrew Quon⁸

¹Department of Medicine, Stanford University, Stanford, CA, ²Department of Radiology, University of Iowa, Iowa City, IA, ³Department of Radiology, Memorial Sloan-Kettering Cancer Center, New York, NY, ⁴Department of Radiology, Mayo Clinic, Rochester, MN, ⁵Department of Statistics, Health Research and Policy, Stanford University, Stanford, CA, ⁶Department of Oncology, Johns Hopkins University, Baltimore, MD, ⁷Department of Pathology, British Columbia Cancer Agency, Vancouver, BC, ⁸Department of Radiology, Stanford University, Stanford, CA.

Corresponding Author:
Sandra J. Horning, MD
875 Blake Wilbur Dr Suite 2338
Stanford, CA 94304
650-725-6456, 650-725-8222 (fax)
sandra.horning@stanford.edu

Presented in part at the 2008 American Society of Hematology Meeting, San Francisco, CA.

Scientific Category: Clinical Trials and Observations

Supported by the Eastern Cooperative Oncology Group Research and Foundation and, in part, by Public Health Service Grants CA21115, CA23318, CA66636, CA13650, CA16116 and from the National Cancer Institute, National Institutes of Health and the Department of Health and Human Services.
Abstract
Positive interim-PET scans are thought to be associated with inferior outcomes in diffuse large B-cell lymphoma (DLBCL). In the E3404 DLBCL study, PET scans at baseline and after 3 R-CHOP were centrally reviewed by a single reader. To determine reproducibility of interim PET interpretation, an expert panel of 3 external nuclear medicine physicians visually scored baseline and interim-PET scans independently and blinded to clinical information. The binary ECOG study criteria were based on modifications of the Harmonization Criteria; the London criteria were also applied. Of 38 interim scans, agreement was complete in 68% and 71% by ECOG and London criteria, respectively. The range of PET+ interim scans was 16% to 34% (p=NS) by reviewer. Moderate consistency of reviews was observed: kappa statistic 0.445 using ECOG and 0.502 using London criteria. These data, showing only moderate reproducibility among nuclear medicine experts, indicate the need to standardize PET interpretation in research and practice. This trial has been registered on http://www.clinicaltrials.gov as NCT00274924.

Introduction
Remarkable predictive accuracy with mid-treatment 18F fluorodeoxyglucose positron emission tomography (PET) scans has been reported in diffuse large B-cell lymphoma (DLBCL), based on the concept that tumor burden above or below the threshold of detection after 1-3 chemotherapy cycles results in treatment failure or success. Although guidelines for PET interpretation in clinical trials have been issued, their reproducibility has not been studied carefully. During conduct of the DLBCL E3404 study, the rate of PET-positive interim scans was lower than projected, and we therefore convened an expert panel to blindly review baseline and interim PET scans from the first ~third of participants to assess reproducibility.

Materials and Methods
Following a baseline PET scan, bulky or advanced DLBCL patients received 3 cycles of R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone) followed by a PET scan 14-20 days later. During the central PET review by a single reader, a fourth cycle of R-CHOP was given and patients continued R-CHOP if PET-negative or changed to R-ICE (rituximab, ifosfamide, carboplatin, etoposide) if PET-positive. Scans were obtained on dedicated high-resolution PET or PET/CT scanners according to protocol and quality control standards at participating ECOG sites. Centralized PET review of baseline and interim scans was performed via file transfer or compact disc with DICOM images. The protocol specified a binary visual interpretation, which the central reviewer based on modifications of the International Harmonization Project, customized for E3404 interim scans, and deemed the “ECOG criteria”: 1) only sites of abnormality at baseline are evaluated, 2) abnormal activity requires both a focal appearance and intensity greater than average liver, 3) all positive nodal sites must have an anatomic correlate, 4) activity in bone marrow and spleen is considered abnormal only if focal and clearly discernible, 5) symmetric abnormal foci in the mediastinum and hilum are considered abnormal only if the remainder of the scan is positive, 6) new foci are considered positive only if the remainder of the scan is positive or a new lesion is focal, very intense and associated with a lesion on CT. The scan interpretation was binary: positive or negative.

Three external nuclear medicine experts independently applied, without dedicated training, the ECOG study criteria as well as the “London” criteria to visually score every baseline lesion at mid-treatment for the first 38 cases (76 scans) from the E3404 study.
Neither the central reviewer nor the experts had access to any clinical information. The London criteria score scans 0-3 as “negative” if uptake is less than liver and 4-5 as “positive” for uptake that is moderately or markedly increased relative to liver. The 245 individual baseline lesions were identified by anatomical site and provided on a worksheet for the external experts, who applied the ECOG and London criteria to each lesion on interim PET. Each case was scored as negative or positive and agreement among external experts was analyzed by Fleiss’ kappa test to determine a p value for differences in proportion of positive scans. The kappa statistic was used to correct for chance in the agreement among the external experts.

Results and Discussion
The proportions of positive interim scans by reader were 16%, 34% and 26%, p=0.206 for ECOG criteria and 16%, 34% and 29%, p=0.263 for London criteria with only reader 3 scoring differently between criteria (Figure 1). With three experts scoring 38 interim scans (representing 1-25 baseline lesions/case), agreement was 68% for ECOG and 71% for London criteria. The kappa statistic was 0.445 using ECOG criteria, indicating only moderate (typical range 0.4-0.6) agreement per case, and 0.502 for London criteria, also in the moderate range. Table 1 details discordance among experts in 12 cases. Reviewer 2 was more likely to interpret interim PET scans as positive, reader 1 less likely, and reviewer 3 in between. In 5 cases, two reviewers considered the interim scan to be positive and in 7 cases, a single reviewer considered the interim scan positive.

Using ECOG criteria, there were 26 cases with complete agreement among experts and these cases were also in complete agreement with the central review. Each expert considered a single case of residual bone disease positive using London criteria but negative by ECOG criteria.

Among the 12 discordant cases, the number of baseline nodal sites ranged from 0-16 (median 5) and 5 cases had extranodal sites at baseline. A single site of disagreement was observed in each case, including para-aortic nodes (n=5), bone (n=3), spleen (2), and one each of iliac/inguinal and supra-clavicular nodes. A definitive CT correlate was present in one case, absent in 8 cases, debatable in 2 cases, and CT was not available in 1 case. Following independent review, the 12 cases were reviewed together to determine if consensus could be achieved. There was agreement in 3 cases, with two cases becoming negative and the other considered positive (Table 1).

The fact that agreement of mid-treatment PET among expert nuclear medicine physicians using standardized criteria was only moderate on a per case basis has important implications as decisions are being made regarding treatment efficacy in practice as well as in clinical trials. More recently, some investigators have raised concern about the false positive rate of interim PET in modern DLBCL treatment, which includes rituximab with its long half-life and unique mechanisms of cytotoxicity, use of dose dense chemotherapy with scans obtained within 2 weeks of treatment, and use of G-CSF. Indeed, the positive predictive value of interim PET scans appears to be lower in the current therapeutic era, ~60%, versus the prior 80% likelihood of failure with chemotherapy alone. The predictive value of interim PET-positive scans has been positively correlated with the international prognostic index and with the international working classification response criteria. Equivocal or indeterminate dictated reports of interim PET scans, which pose challenges for clinicians, appear to predict treatment success rather than failure. The literature is inconsistent with regard to the predictive value of PET scans at the conclusion of R-CHOP, suggesting real differences in interpretation. Lin et al. have proposed that changes in standard uptake value may
improve the predictive accuracy of interim FDG-PET. In sum, broader use of interim PET scans in the modern therapeutic era has not reproduced the dichotomous results previously reported, although progression-free survival is generally consistently inferior for interim PET-positive patients.

Using our study criteria, the proportion of PET-positive scans was relatively low and, the current report relates solely to the reproducibility of interpretation using standardized criteria. Agreement among external experts would likely have been higher if the study had been preceded by a training exercise using the two study criteria, neither of which is well validated (no such criteria, in fact, exist for interim PET scan). It is interesting that there was essentially no difference in agreement with either ECOG or the London criteria, which are being applied in a phase III Hodgkin lymphoma trial. Our results indicate that, among multiple involved sites at diagnosis, single sites, particularly para-aortic, spleen and bone were the source of disagreement on interim PET, and CT correlates of residual positive sites were frequently absent or debatable. We conclude that greater harmonization of PET interpretation is indicated for research and practice and this will require training of nuclear physicians using consistent, validated interpretive criteria and standardized reporting.
Acknowledgement: John Allen and Patrick Pringle, supported by Stanford University, provided technical and data management support. The study was supported by the Eastern Cooperative Oncology Group Research and Foundation. The E3404 clinical trial study was conducted by the Eastern Cooperative Oncology Group (Robert L. Comis, M.D., Chair) and supported in part by Public Health Service Grants CA21115, CA23318, CA66636, CA13650, CA16116 and from the National Cancer Institute, National Institutes of Health and the Department of Health and Human Services. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

Authorship Contributions
Sandra J. Horning: Designed research; collected, analyzed, and interpreted data, wrote the manuscript.
Malik Juweid: Performed research, participated in analysis and interpretation of data and manuscript preparation.
Heiko Schoder: Performed research, participated in analysis and interpretation of data and manuscript preparation.
Gregory Wiseman: Performed research, participated in analysis and interpretation of data and manuscript preparation.
Alex McMillan: Conducted statistical analysis and participated in interpretation of data and manuscript preparation.
Lode Swinnen: Principal investigator of the clinical trial; facilitated central review of PET scans, participated in manuscript review.
Ranjan Advani: Co-principal investigator of the clinical trial; participated in manuscript review.
Randy Gascoyne: Reviewed diagnostic pathology for the clinical trial: participated in manuscript review.
Andrew Quon: Designed and performed research; provided central PET review for the clinical trial; participated in analysis and interpretation of data and manuscript preparation.

Disclosure of Conflicts of Interest
Sandra J. Horning: None
Malik Juweid: None
Heiko Schoder: None
Gregory Wiseman: None
Alex McMillan: None
Lode Swinnen: None
Ranjan Advani: None
Randy Gascoyne: None
Andrew Quon: None
References


3. Radford J, O'Doherty M, Barrington S, et al.: Results of the 2nd Planned Interim Analysis of the RAPID Trial (involved field radiotherapy versus no further treatment) in Patients with Clinical Stages 1A and 2A Hodgkin Lymphoma and a 'negative' FDG-PET Scan after 3 Cycles ABVD. Blood 2008; 112(11): 369.

4. Moskowitz C, Hamlin PA, Horwitz SM, et al.: Phase II Trial of Dose-Dense R-CHOP Followed by Risk-Adapted Consolidation with Either ICE or ICE and ASCT, Based upon the Results of Biopsy Confirmed Abnormal Interim Restaging PET Scan, Improves Outcome in Patients with Advanced Stage DLBCL. Blood 2006; 108(11): 532.


Table 1. Twelve Cases of Expert Reviewer Disagreement in Interim PET Scans

<table>
<thead>
<tr>
<th>Case</th>
<th># Concordant Lesions</th>
<th># Discordant Lesions</th>
<th>Reviewer 1</th>
<th>Reviewer 2</th>
<th>Reviewer 3</th>
<th>Consensus Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1 (PA)</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1 (Bo)</td>
<td>0</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>1 (Sp)</td>
<td>0</td>
<td>0</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1 (PA)</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1 (PA)</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>1 (PA)</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1 (Ce)</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>Y*</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1 (Bo)</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>Y*</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>1 (IL)</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>1 (Sc)</td>
<td>0</td>
<td>X</td>
<td>X</td>
<td>Y**</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1 (Sp)</td>
<td>X</td>
<td>0</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1 (Bo)</td>
<td>0</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
</tbody>
</table>

0= negative scan, X= positive scan
PA=para-aortic, Bo=bone, Sp=spleen, Ce=cervical, IL=iliac, Sc=supra-clavicular.
*Consensus “negative;” ** consensus “positive.”
Figure 1: Proportion of interim-PET cases interpreted as positive by reader, according to the ECOG and London criteria (see text). Error bar indicates one standard error for the proportion.
Interim positron emission tomography (PET) scans in diffuse large B-cell lymphoma: an independent expert nuclear medicine evaluation of the Eastern Cooperative Oncology Group E3404 study

Sandra J. Horning, Malik E. Juweid, Heiko Schoder, Gregory Wiseman, Alex McMillan, Lode J. Swinnen, Ranjana Advani, Randy Gascoyne and Andrew Quon

Information about reproducing this article in parts or in its entirety may be found online at: http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at: http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at: http://www.bloodjournal.org/site/subscriptions/index.xhtml

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036. Copyright 2011 by The American Society of Hematology; all rights reserved.