CME article

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

Sandra J. Horning,1 Malik E. Juweid,2 Heiko Schöder,3 Gregory Wiseman,4 Alex McMillan,5 Lode J. Swinnen,6 Ranjana Advani,1 Randy Gascoyne,7 and Andrew Quon8

1Department of Medicine, Stanford University, CA; 2Department of Radiology, University of Iowa, Iowa City; 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, CA; 6Department of Oncology, Johns Hopkins University, Baltimore, MD; 7Department of Pathology, British Columbia Cancer Agency, Vancouver, BC; and 8Department of Radiology, Stanford University, CA

Positive interim positron emission tomography (PET) scans are thought to be associated with inferior outcomes in diffuse large B-cell lymphoma. In the E3404 diffuse large B-cell lymphoma study, PET scans at baseline and after 3 cycles of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone were centrally reviewed by a single reader. To determine the reproducibility of interim PET interpretation, an expert panel of 3 external nuclear medicine physicians visually scored baseline and interim PET scans independently and were blinded to clinical information. The binary Eastern Cooperative Oncology Group (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 = not significant) by reviewer. Moderate consistency of reviews was observed: \( \kappa \) statistic = 0.445 using ECOG criteria, and \( \kappa \) statistic = 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 was registered at www.clinicaltrials.gov as #NCT00274924. (Blood. 2010;115:775-777)

MedscapeCME Continuing Medical Education online

This activity has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint sponsorship of Medscape, LLC and the American Society of Hematology. Medscape, LLC is accredited by the ACCME to provide continuing medical education for physicians. Medscape, LLC designates this educational activity for a maximum of 0.25 AMA PRA Category 1 credits™. Physicians should only claim credit commensurate with the extent of their participation in the activity. All other clinicians completing this activity will be issued a certificate of participation. To participate in this journal CME activity: (1) review the learning objectives and author disclosures; (2) study the education content; (3) take the post-test and/or complete the evaluation at http://cme.medscape.com/cme/blood; and (4) view/print certificate. For CME questions, see page 918.

Disclosures

The authors, the Associate Editor Martin S. Tallman, and the CME questions author Charles P. Vega, University of California, Irvine, CA, declare no competing interests.

Learning objectives

Upon completion of this activity, participants should be able to:
1. Identify study procedures in the current research
2. Specify the interobserver agreement in regard to interim positron emission tomographic (PET) scans in the current study
3. Describe the current therapeutic approach to diffuse large B-cell lymphoma
4. List common anatomic sites of disagreement between nuclear medicine specialists in the current study

Introduction

Remarkable predictive accuracy with midtreatment 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 to 3 chemotherapy cycles results in treatment failure or success.1 Although guidelines for PET interpretation in clinical trials have been issued, their reproducibility has not been studied carefully.2 During conduct of the DLBCL E3404 study, the


An Inside Blood analysis of this article appears at the front of this issue.

Presented in part at the 50th annual meeting of the American Society of Hematology, San Francisco, CA, December 8, 2008.

© 2010 by The American Society of Hematology
rate of PET⁺ interim scans was lower than projected, and we therefore convened an expert panel to blindly review baseline and interim PET scans from approximately the first one-third of participants to assess reproducibility.

Methods

After a baseline PET scan, bulky or advanced DLBCL patients received 3 cycles of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP), followed by a PET scan 14 to 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⁻ or changed to rituximab, ifosfamide, carboplatin, and etoposide if PET⁺. Scans were obtained on dedicated high-resolution PET or PET/computed tomography (CT) scanners according to protocol and quality control standards at participating Eastern Cooperative Oncology Group (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; and (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. Translation of the “London criteria” to interim PET scans was achieved by the central PET reader, which was based on modifications of the ECOG and London criteria. The resulting criteria were applied by three expert reviewers, who independently reviewed each scan according to the methods described above. There was a fourth reader, the nonblind, who reviewed all scans to determine whether consensus could be achieved. There was complete agreement among the external experts, and these cases were also included in the review.

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 midtreatment 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 to 3 as “negative” if uptake is less than liver and 4 or 5 as “positive” for uptake that is moderately or markedly increased relative to liver. Scans were obtained on dedicated high-resolution PET or PET/computed tomography (CT) scanners according to protocol and quality control standards at participating Eastern Cooperative Oncology Group (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 used 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; and (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. Interpretation of the “London criteria” to interim PET scans was achieved by the central PET reader, which was based on modifications of the ECOG and London criteria. The resulting criteria were applied by three expert reviewers, who independently reviewed each scan according to the methods described above. There was a fourth reader, the nonblind, who reviewed all scans to determine whether consensus could be achieved. There was complete agreement among the external experts, and these cases were also included in the review.

The fact that agreement of midtreatment 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 granulocyte colony-stimulating factor. 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⁺ scans has been positively correlated with the international prognostic index and with the international working classification response criteria. Equivocal or indeterminate dichotomous 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, the 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⁺ patients. Using our study criteria, the proportion of PET⁺ scans was relatively low, and the current report relates solely to the reproducibility of interpretation using standardized criteria. Agreement among external experts would probably have been higher if the study had been preceded by a training exercise using the study criteria, neither of which is well validated (no such criteria 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 3 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.

Acknowledgments

The authors thank John Allen and Patrick Pringle, supported by Stanford University, who provided technical and data management support. This work was supported by the Eastern Cooperative Oncology Group Research and Foundation. The E3404 clinical trial study was conducted by the Eastern Cooperative Oncology Group (Dr Robert L. Comis, Chair) and supported in part by the National Cancer Institute, National Institutes of Health and the Department of Health and Human Services (Public Health Service grants CA21115, CA23318, CA66636, CA13650, and CA16116). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

Table 1. Twelve cases of expert reviewer disagreement in interim PET scans

<table>
<thead>
<tr>
<th>Case no.</th>
<th>Concordant lesions, n</th>
<th>Discordant lesions, n</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>+</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1 (Bo)</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>1 (Sp)</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1 (PA)</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1 (PA)</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>1 (PA)</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>1 (Ce)</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>Y*</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1 (Bo)</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>Y*</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>1 (IL)</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>1 (Sc)</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>Y†</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1 (Sp)</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1 (Bo)</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>N</td>
</tr>
</tbody>
</table>

0 indicates negative scan; +, positive scan; PA, para-aortic; Bo, bone; Sp, spleen; Ce, cervical; IL, iliac; and Sc, supraclavicular.
"Consensus "negative."
†Consensus "positive."

References


Authorship

Contribution: S.J.H. designed research, collected, analyzed, and interpreted data, and wrote the manuscript; M.E.J., H.S. and G.W. performed research and participated in analysis and interpretation of data and manuscript preparation; A.M. conducted statistical analysis and participated in interpretation of data and manuscript preparation; L.J.S. was the principal investigator of the clinical trial, facilitated central review of PET scans, and participated in manuscript review; R.A. was the coprincipal investigator of the clinical trial and participated in manuscript review; R.G. reviewed diagnostic pathology for the clinical trial and participated in manuscript preparation; and A.Q. designed and performed research, provided central PET review for the clinical trial, and participated in analysis and interpretation of data and manuscript preparation.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Sandra J. Horning, 875 Blake Wilbur Dr, Suite 2338, Stanford, CA 94304; e-mail: sandra.horning@stanford.edu.
Interim positron emission tomography 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 Schöder, Gregory Wiseman, Alex McMillan, Lode J. Swinnen, Ranjana Advani, Randy Gascoyne and Andrew Quon