

Revised Response Criteria for Malignant Lymphoma

Bruce D. Cheson, Beate Pfistner, Malik E. Juweid, Randy D. Gascoyne, Lena Specht, Sandra J. Horning, Bertrand Coiffier, Richard I. Fisher, Anton Hagenbeek, Emanuele Zucca, Steven T. Rosen, Sigrid Stroobants, T. Andrew Lister, Richard T. Hoppe, Martin Dreyling, Kensei Tobinai, Julie M. Vose, Joseph M. Connors, Massimo Federico, and Volker Diehl

A B S T R A C T

Purpose

Standardized response criteria are needed to interpret and compare clinical trials and for approval of new therapeutic agents by regulatory agencies.

Methods

The International Working Group response criteria (Cheson et al, *J Clin Oncol* 17:1244, 1999) were widely adopted, but required reassessment because of identified limitations and the increased use of [¹⁸F]fluorodeoxyglucose-positron emission tomography (PET), immunohistochemistry (IHC), and flow cytometry. The International Harmonization Project was convened to provide updated recommendations.

Results

New guidelines are presented incorporating PET, IHC, and flow cytometry for definitions of response in non-Hodgkin's and Hodgkin's lymphoma. Standardized definitions of end points are provided.

Conclusion

We hope that these guidelines will be adopted widely by study groups, pharmaceutical and biotechnology companies, and regulatory agencies to facilitate the development of new and more effective therapies to improve the outcome of patients with lymphoma.

J Clin Oncol 25:579-586. © 2007 by American Society of Clinical Oncology

INTRODUCTION

Standardized response criteria provide uniform end points for clinical trials, allowing for comparisons among studies, facilitating the identification of more effective therapies, and aiding the approval process for new agents by regulatory agencies. Before 1999, response criteria for malignant lymphomas varied widely among study groups and cancer centers with respect to the size of a normal lymph node, the frequency of assessment and the time point the response assessment was made, the methods used to assess response, whether response was assessed prospectively or retrospectively, the percentage increase required for disease progression, and many other factors.¹ Even relatively minor differences in the definition of normal size of a lymph node can have a major influence on response rates.²

In 1999, an international working group (IWG) of clinicians, radiologists, and pathologists with expertise in the evaluation and management of patients with non-Hodgkin's lymphoma (NHL) published guidelines for response assessment and outcomes measurement.¹ These recommendations were adopted rapidly

and widely by clinicians and regulatory agencies, and were used in the approval process for a number of new agents. However, they were subject to considerable inter- and intraobserver variation and recommended technologies, such as gallium scans, are no longer considered state-of-the-art. Several points were subject to misinterpretation, notably the application of the complete remission/unconfirmed (CRu), and the recommendations did not include assessment of extranodal disease. The widespread use of positron emission tomography (PET) scans and immunohistochemistry warranted a reassessment of the prior response criteria. Since the Hodgkin's lymphoma study groups had adopted these IWG criteria, any new recommendations needed to account for those patients as well. As a result, an International Harmonization Project was initiated by the German Competence Network Malignant Lymphoma to develop recommendations that were consistent across study groups.³ Subcommittees were organized on Response criteria, End Points for Clinical Trials, Imaging, Clinical Features, and Pathology/Biology, and the recommendations are reflected in this report.

From the Division of Hematology/Oncology, Georgetown University Hospital, Washington, DC; University of Cologne, Cologne; Department of Nuclear Medicine, University of Iowa, Iowa City, IA; Department of Pathology, British Columbia Cancer Agency and the University of British Columbia, Vancouver, British Columbia, Canada; Department of Oncology and Hematology, Rigshospitalet, Copenhagen University Hospital, Denmark; Division of Oncology and Department of Radiation Oncology, Stanford University, Stanford, CA; Department of Hematology, Hospices Civils de Lyon and Université Claude Bernard, Lyon, France; James P. Wilmot Cancer Center, University of Rochester, Rochester, NY; Academic Medical Center, Department of Hematology, Amsterdam, the Netherlands; Lymphoma Unit, Department of Medical Oncology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland; Lurie Cancer Center, Northwestern University, Chicago, IL; Department of Nuclear Medicine, University Hospital Gasthuisberg, Leuven, Belgium; Cancer Research UK Medical Oncology Unit, St Bartholomew's Hospital, London, United Kingdom; Department of Medicine III, University of Munich, Hospital Grosshadern, Munich, Germany; Hematology and Stem Cell Transplantation Division, National Cancer Center Hospital, Tokyo, Japan; Section of Hematology/Oncology, University of Nebraska Medical Center, Omaha, NE; and Dipartimento di Oncologia ed Ematologia, Università di Modena e Reggio Emilia, Modena, Italy.

Submitted September 18, 2006; accepted December 20, 2006; published online ahead of print at www.jco.org on January 22, 2007.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to Bruce D. Cheson, MD, Georgetown University Hospital, 3800 Reservoir Rd NW, Washington, DC 20007; e-mail: bdc4@georgetown.edu.

© 2007 by American Society of Clinical Oncology

0732-183X/07/2505-579/\$20.00

DOI: 10.1200/JCO.2006.09.2403

MODIFICATIONS OF THE IWG CRITERIA

PET

PET using [¹⁸F]fluorodeoxyglucose (FDG), has emerged as a powerful functional imaging tool for staging, restaging, and response assessment of lymphomas.^{4-24,25} The advantage of PET over conventional imaging techniques such as computed tomography (CT) or magnetic resonance imaging is its ability to distinguish between viable tumor and necrosis or fibrosis in residual mass(es) often present after treatment.^{9,11,26-28} This information may have important clinical consequences. Juweid et al²⁰ evaluated the impact of integrating PET into the IWG criteria in a retrospective study of 54 patients with diffuse large B-cell NHL who had been treated with an anthracycline-based regimen. PET increased the number of complete remission (CR) patients, eliminated the CRu category, and enhanced the ability to discern the difference in progression-free survival (PFS) between patients experiencing CR and partial remission (PR). Such findings provided rationale for incorporating PET into revised criteria.

However, a number of issues with PET need to be considered. The technique for performing and interpreting PET has only recently been standardized.²⁹ There is variability among readers and equipment. PET is also associated with false-positive findings due to rebound thymic hyperplasia, infection, inflammation, sarcoidosis, or brown fat. Diffusely increased bone marrow uptake is often observed after treatment or administration of hematopoietic growth factors.^{19,29,33,34} There are also false-negative results with PET relating to the resolution of the equipment, technique, and variability of FDG avidity among histologic subtypes.^{10,29-32} These and other considerations regarding interpretation of PET scans have recently been addressed.²⁹

Recommendations for the use of PET or PET/CT. Current recommendations for the use of PET scans reflect the FDG avidity of the lymphoma subtype, and the relevant end points of the clinical trial (Table 1).

1. PET is strongly recommended before treatment for patients with routinely FDG-avid, potentially curable lymphomas (eg, diffuse large B-cell lymphoma [DLBCL], Hodgkin's lymphoma) to better delineate the extent of disease; however, currently it is not mandated because of limitations imposed by cost and availability. For incurable,

routinely FDG-avid, indolent, and aggressive histologies (eg, follicular lymphoma and mantle-cell lymphoma), and for most variably FDG-avid lymphomas, the primary end points for clinical trials generally include PFS, event-free survival, and overall survival. PET is not recommended before treatment unless response rate is a major end point of the trial.

2. Numerous studies have demonstrated that PET performed after one to four cycles of multiagent chemotherapy predicts therapeutic outcome^{5-7,21,24,35,36}; however, no currently available data demonstrate improvement in results by altering treatment based on this information. Until such data exist, this practice should be restricted to clinical trials evaluating PET in this context.

3. PET is essential for the post-treatment assessment of DLBCL and Hodgkin's lymphoma because a complete response is required for a curative outcome. However, PET is recommended in the other, incurable histologies only if they were PET positive before treatment and if response rate is a primary end point of a clinical study.

4. Current data are inadequate to recommend routine surveillance PET scans after the restaging study.

Timing of PET scans after therapy. Post-therapy inflammatory changes may persist for up to 2 weeks after chemotherapy alone in lymphoma patients and for up to 2 to 3 months or longer after radiation therapy or chemotherapy plus radiation. To minimize the frequency of these potentially confounding interpretation finding, PET scans should not be performed for at least 3 weeks, and preferably 6 to 8 weeks, after completion of therapy.²⁹

Definition of a positive PET scan. Visual assessment currently is considered adequate for determining whether a PET scan is positive, and use of the standardized uptake value is not necessary.²⁹ A more extensive description of interpretation of PET scans is provided in the consensus guidelines of the Imaging Subcommittee.²⁹ In brief, a positive scan is defined as focal or diffuse FDG uptake above background in a location incompatible with normal anatomy or physiology, without a specific standardized uptake value cutoff.²⁹ Other causes of false-positive scans should be ruled out. Exceptions include mild and diffusely increased FDG uptake at the site of moderate- or large-sized masses with an intensity that is lower than or equal to the mediastinal blood pool, hepatic or splenic nodules 1.5 cm with FDG uptake lower than the surrounding liver/spleen

Table 1. Recommended Timing of PET (PET/CT) Scans in Lymphoma Clinical Trials

| Histology | Pretreatment | Mid-Treatment | Response Assessment | Post-Treatment Surveillance |
|---------------------------|--------------|----------------|---------------------|-----------------------------|
| Routinely FDG avid | | | | |
| DLBCL | Yes* | Clinical trial | Yes | No |
| HL | Yes* | Clinical trial | Yes | No |
| Follicular NHL | No† | Clinical trial | No† | No |
| MCL | No† | Clinical trial | No† | No |
| Variably FDG avid | | | | |
| Other aggressive NHLs | No† | Clinical trial | No†‡ | No |
| Other indolent NHLs | No† | Clinical trial | No†‡ | No |

Abbreviations: PET, positron emission tomography; CT, computed tomography; FDG, [¹⁸F]fluorodeoxyglucose; DLBCL, diffuse large B-cell lymphoma; HL, Hodgkin's lymphoma; NHL, non-Hodgkin's lymphoma; MCL, mantle-cell lymphoma; ORR, overall response rate; CR, complete remission.

*Recommended but not required pretreatment.

†Recommended only if ORR/CR is a primary study end point.

‡Recommended only if PET is positive pretreatment.

uptake, and diffusely increased bone marrow uptake within weeks after treatment. Specific criteria for lung nodules based on lesion size have been developed.²⁹

Bone Marrow Assessment

Restaging bone marrow examinations are commonly used to assess response to therapy. The determination of involvement may be difficult, given that no universally accepted standards exist. The usual approach to response determination relies on morphologic assessment of the bone marrow biopsy, and clot section if adequate and available, whereas ancillary studies using immunohistochemistry, flow cytometry, and polymerase chain reaction methodology are largely ignored or underused. Moreover, a direct comparison of these studies and their respective sensitivity and specificity for the detection of occult but clinically meaningful involvement are lacking. Thus, recommendations regarding the use of these strategies and their interpretation are largely empiric at this time.

The recommendation for bone marrow response is that histologically normal bone marrows with a small (< 2%) clonal B-cell population detected by flow cytometry should be considered normal, given that definitive clinical studies that demonstrate an inferior outcome are lacking. Immunohistochemistry has a clear role in the assessment of the bone marrow at diagnosis and restaging after therapy. When antibodies are used to detect CD20 and CD3 expression, morphologically normal bone marrows can often be shown to harbor disease. Sensitivity can be increased with the use of subtype-specific antibody panels directed at CD5, cyclin D1, CD23, CD10, DBA44, and kappa and lambda light chains. Less common lymphoma subtypes with occult bone marrow disease are particularly well suited to this approach, including splenic marginal zone B-cell lymphomas and a number of subtypes of DLBCL (ie, intravascular large B-cell lymphoma and HIV-related DLBCL). Indolent B-cell lymphomas and chronic lymphocytic leukemia are more difficult to assess, given that the distinction from reactive lymphoid aggregates and nodular partial remissions in the bone marrow can be difficult to assess because of the frequent admixture of reactive T cells in these diseases. Immunohistochemistry using anti-CD5 and anti-CD23 can be helpful in this setting, as are stains for kappa and lambda light chains that can detect surface membrane immunoglobulin in paraffin sections. Similarly, antibodies to cyclin D1 and CD10 are useful for recognizing subtle bone marrow involvement in mantle-cell lymphoma and follicular lymphoma, respectively. In the future, antibodies to Bcl-6 may improve detection of occult follicular lymphoma in the bone marrow; however, technical problems preclude their general use at this time. In fact, many routinely used immunohistochemical reagents can be difficult to apply consistently to the evaluation of bone marrow samples, largely due to subtleties in fixation methods and decalcification techniques.

Caution is recommended when interpreting biopsies post-therapy for residual disease. The use of rituximab may lead to a false-negative interpretation of residual B-cell disease, despite the fact that the widely used commercial anti-CD20 (L26) recognizes a cytoplasmic epitope of CD20, in contrast to the surface epitope recognized by rituximab. The judicious use of another pan-B-cell antibody, CD79a, is strongly recommended when evaluating post-treatment samples. Similar caution is required when interpreting CD20 flow cytometric data for several months after therapy with rituximab, given that surface epitopes may be blocked. The availability of clot sections

allows for immunohistochemical analysis without the influence of decalcification and may be useful for the post-treatment evaluation of bone marrow involvement.

Lastly, the role of molecular genetic analyses in the determination of response to therapy is difficult to resolve. Assay techniques and sensitivity vary enormously between laboratories, making systematic recommendations impossible. Residual clonal disease may exist without morphologic evidence of lymphoma (ie, gastric mucosa-associated lymphoid tissue [MALT] lymphoma after therapy). In aggregate, these data suggest that the disappearance of the molecular clone may lag behind the disappearance of morphologic evidence of disease. Alternatively, these findings may represent the persistence of residual disease or potentially repopulating lymphoma stem cells in biopsies lacking morphologic evidence of lymphoma. These distinctions need to be reconciled before molecular testing can be considered routine, particularly when the findings affect treatment decisions.

Sensitive and sophisticated diagnostic approaches such as flow cytometry and/or molecular genetic analyses should be incorporated into clinical trials to determine their relevance and potential utility for directing therapy. However, for routine practice we do not recommend that clinical decision making be based solely on flow cytometry and/or molecular genetic analyses that indicate a residual small (< 2% of gated or live events) B-cell clone in the absence of other supportive findings from morphology and immunohistochemistry. We strongly encourage investigators to collect these data together with clinical correlative data that might eventually support their routine use for the assessment of response criteria for lymphoid malignancies.

REVISED RESPONSE CRITERIA

CR

The designation of CR requires the following (Table 2):

1. Complete disappearance of all detectable clinical evidence of disease and disease-related symptoms if present before therapy.
 - 2a. Typically FDG-avid lymphoma: in patients with no pretreatment PET scan or when the PET scan was positive before therapy, a post-treatment residual mass of any size is permitted as long as it is PET negative.
 - 2b. Variably FDG-avid lymphomas/FDG avidity unknown: in patients without a pretreatment PET scan, or if a pretreatment PET scan was negative, all lymph nodes and nodal masses must have regressed on CT to normal size (≤ 1.5 cm in their greatest transverse diameter for nodes > 1.5 cm before therapy). Previously involved nodes that were 1.1 to 1.5 cm in their long axis and more than 1.0 cm in their short axis before treatment must have decreased to ≤ 1.0 cm in their short axis after treatment.
3. The spleen and/or liver, if considered enlarged before therapy on the basis of a physical examination or CT scan, should not be palpable on physical examination and should be considered normal size by imaging studies, and nodules related to lymphoma should disappear. However, determination of splenic involvement is not always reliable because a spleen considered normal in size may still contain lymphoma, whereas an enlarged spleen may reflect variations in anatomy, blood volume, the use of hematopoietic growth factors, or causes other than lymphoma.

Table 2. Response Definitions for Clinical Trials

| Response | Definition | Nodal Masses | Spleen, Liver | Bone Marrow |
|------------------------|---|---|--|--|
| CR | Disappearance of all evidence of disease | (a) FDG-avid or PET positive prior to therapy; mass of any size permitted if PET negative (b) Variably FDG-avid or PET negative; regression to normal size on CT | Not palpable, nodules disappeared | Infiltrate cleared on repeat biopsy; if indeterminate by morphology, immunohistochemistry should be negative |
| PR | Regression of measurable disease and no new sites | $\geq 50\%$ decrease in SPD of up to 6 largest dominant masses; no increase in size of other nodes (a) FDG-avid or PET positive prior to therapy; one or more PET positive at previously involved site (b) Variably FDG-avid or PET negative; regression on CT | $\geq 50\%$ decrease in SPD of nodules (for single nodule in greatest transverse diameter); no increase in size of liver or spleen | Irrelevant if positive prior to therapy; cell type should be specified |
| SD | Failure to attain CR/PR or PD | (a) FDG-avid or PET positive prior to therapy; PET positive at prior sites of disease and no new sites on CT or PET (b) Variably FDG-avid or PET negative; no change in size of previous lesions on CT | | |
| Relapsed disease or PD | Any new lesion or increase by $\geq 50\%$ of previously involved sites from nadir | Appearance of a new lesion(s) > 1.5 cm in any axis, $\geq 50\%$ increase in SPD of more than one node, or $\geq 50\%$ increase in longest diameter of a previously identified node > 1 cm in short axis Lesions PET positive if FDG-avid lymphoma or PET positive prior to therapy | $> 50\%$ increase from nadir in the SPD of any previous lesions | New or recurrent involvement |

Abbreviations: CR, complete remission; FDG, [^{18}F]fluorodeoxyglucose; PET, positron emission tomography; CT, computed tomography; PR, partial remission; SPD, sum of the product of the diameters; SD, stable disease; PD, progressive disease.

4. If the bone marrow was involved by lymphoma before treatment, the infiltrate must have cleared on repeat bone marrow biopsy. The biopsy sample on which this determination is made must be adequate (with a goal of > 20 mm unilateral core). If the sample is indeterminate by morphology, it should be negative by immunohistochemistry. A sample that is negative by immunohistochemistry but that demonstrates a small population of clonal lymphocytes by flow cytometry will be considered a CR until data become available demonstrating a clear difference in patient outcome.

CRu

The use of the above definition for CR and that below for PR eliminates the category of CRu.

PR

The designation of PR requires all of the following:

1. At least a 50% decrease in sum of the product of the diameters (SPD) of up to six of the largest dominant nodes or nodal masses. These nodes or masses should be selected according to all of the following; they should be clearly measurable in at least 2 perpendicular dimensions; if possible they should be from disparate regions of the body; and they should include mediastinal and retroperitoneal areas of disease whenever these sites are involved.

2. No increase should be observed in the size of other nodes, liver, or spleen.

3. Splenic and hepatic nodules must regress by $\geq 50\%$ in their SPD or, for single nodules, in the greatest transverse diameter.

4. With the exception of splenic and hepatic nodules, involvement of other organs is usually assessable and no measurable disease should be present.

5. Bone marrow assessment is irrelevant for determination of a PR if the sample was positive before treatment. However, if positive, the cell type should be specified (eg, large-cell lymphoma or small neoplastic B cells). Patients who achieve a CR by the above criteria, but

who have persistent morphologic bone marrow involvement will be considered partial responders.

When the bone marrow was involved before therapy and a clinical CR was achieved, but with no bone marrow assessment after treatment, patients should be considered partial responders.

6. No new sites of disease should be observed.

7. Typically FDG-avid lymphoma: for patients with no pretreatment PET scan or if the PET scan was positive before therapy, the post-treatment PET should be positive in at least one previously involved site.

8. Variably FDG-avid lymphomas/FDG-avidity unknown: for patients without a pretreatment PET scan, or if a pretreatment PET scan was negative, CT criteria should be used.

In patients with follicular lymphoma or mantle-cell lymphoma, a PET scan is only indicated with one or at most two residual masses that have regressed by more than 50% on CT; those with more than two residual lesions are unlikely to be PET negative and should be considered partial responders.

Stable Disease

Stable disease (SD) is defined as the following:

1. A patient is considered to have SD when he or she fails to attain the criteria needed for a CR or PR, but does not fulfill those for progressive disease (see Relapsed Disease [after CR]/Progressive Disease [after PR, SD]).

2. Typically FDG-avid lymphomas: the PET should be positive at prior sites of disease with no new areas of involvement on the post-treatment CT or PET.

3. Variably FDG-avid lymphomas/FDG-avidity unknown: for patients without a pretreatment PET scan or if the pretreatment PET was negative, there must be no change in the size of the previous lesions on the post-treatment CT scan.

Relapsed Disease (after CR)/Progressive Disease (after PR, SD)

Lymph nodes should be considered abnormal if the long axis is more than 1.5 cm regardless of the short axis. If a lymph node has a long axis of 1.1 to 1.5 cm, it should only be considered abnormal if its short axis is more than 1.0. Lymph nodes $\leq 1.0 \times \leq 1.0$ cm will not be considered as abnormal for relapse or progressive disease.

1. Appearance of any new lesion more than 1.5 cm in any axis during or at the end of therapy, even if other lesions are decreasing in size. Increased FDG uptake in a previously unaffected site should only be considered relapsed or progressive disease after confirmation with other modalities. In patients with no prior history of pulmonary lymphoma, new lung nodules identified by CT are mostly benign. Thus, a therapeutic decision should not be made solely on the basis of the PET without histologic confirmation.

2. At least a 50% increase from nadir in the SPD of any previously involved nodes, or in a single involved node, or the size of other lesions (eg, splenic or hepatic nodules). To be considered progressive disease, a lymph node with a diameter of the short axis of less than 1.0 cm must increase by $\geq 50\%$ and to a size of 1.5×1.5 cm or more than 1.5 cm in the long axis.

3. At least a 50% increase in the longest diameter of any single previously identified node more than 1 cm in its short axis.

4. Lesions should be PET positive if observed in a typical FDG-avid lymphoma or the lesion was PET positive before therapy unless the lesion is too small to be detected with current PET systems (< 1.5 cm in its long axis by CT).

Measurable extranodal disease should be assessed in a manner similar to that for nodal disease. For these recommendations, the spleen is considered nodal disease. Disease that is only assessable (eg, pleural effusions, bone lesions) will be recorded as present or absent only, unless, while an abnormality is still noted by imaging studies or physical examination, it is found to be histologically negative.

In clinical trials where PET is unavailable to the vast majority of participants, or where PET is not deemed necessary or appropriate for use (eg, a trial in patients with MALT lymphoma), response should be assessed as above, but only using CT scans. However, residual masses should not be assigned CRu status, but should be considered partial responses.

Primary CNS Lymphomas

Recommendations of the International Workshop on Evaluation of Primary Central Nervous System Lymphomas were adopted in their entirety.³⁷

Primary Gastric Lymphoma

Evaluation of patients with primary gastric lymphomas, especially MALT lymphomas, is difficult and confounded by the observation that prolonged clinical remissions may be associated with transient histologic and molecular relapses, and persistence of monoclonal B cells after histologic regression.^{38,39} Repeated biopsies remain a fundamental follow-up procedure, despite problems with reproducibility.

Interpretation of residual lymphoid infiltrates in post-treatment gastric biopsies can be difficult, with no uniform criteria for the definition of histologic remission. Older assessment systems have not been adopted uniformly.^{40,41} A histologic grading system proposed by the Groupe d'Etude des Lymphomes de l'Adulte may be an improvement over prior schemes, but will require additional validation.^{42,43}

Follow-Up Evaluation

The manner in which patients are evaluated after completing treatment may vary according to whether treatment was administered in a clinical trial or clinical practice, or whether treatment was delivered with curative or palliative intent. Good clinical judgment and a careful history and physical examination are the most important components of monitoring patients after treatment. Additional testing at follow-up visits should include CBC and serum chemistries, including lactate dehydrogenase and other blood tests and imaging studies for relevant clinical indications. There is no evidence to support regular surveillance CT scans, given that the patient or physician identifies the relapse more than 80% of the time without the need for imaging studies.⁴⁴⁻⁴⁷ Data with PET are also insufficient to recommend routine procedures at this time.⁴⁸

In a clinical trial, uniformity of reassessment is necessary to ensure comparability among studies with respect to the major end points of event-free survival, disease-free survival, and PFS. It is obvious, for example, that a protocol requiring re-evaluation every 2 months will produce different results compared with one requiring the same testing annually, even if the true times to events are the same. One recommendation has been to assess patients on clinical trials after completion of treatment at a minimum of every 3 months for 2 years, then every 6 months for 3 years, and then annually for at least 5 years.¹ Few recurrences occur beyond that point for patients with diffuse large-cell NHL or Hodgkin's lymphoma. However, the risk of relapse for patients with follicular and other indolent histologies is continuous. These intervals may vary with specific treatments, duration of treatment, protocols, or unique drug characteristics. Recently, the National Comprehensive Cancer Network published recommendations for follow-up of patients with Hodgkin's and NHL:^{49,50} for patients with Hodgkin's lymphoma in an initial CR, an interim history and physical examination every 2 to 4 months for 1 to 2 years, then every 3 to 6 months for the next 3 to 5 years, with annual monitoring for late effects after 5 years. For follicular or other indolent histology lymphoma patients in a CR, the recommendation for follow-up was every 3 months for a year then every 3 to 6 months. For diffuse large B-cell NHL, the guidelines proposed follow-up every 3 months for 24 months then every 6 months for 36 months.^{49,50}

Patients with a follicular or low-grade NHL who are being managed with a so-called watch and wait approach should be monitored for the development of disease-related symptoms or signs of organ involvement. No consensus regarding the frequency of follow-up of such patients exists and the interval should be specified in the protocol. Otherwise, imaging studies should be individualized based on the location of the disease and informed by the behavior of palpable disease.

END POINTS

The major end points of clinical trials should reflect the histology, clinical situation (eg, initial treatment *v* salvage), and objectives of the study (Table 3). It is important that consistent definitions of end points are used, and we hope that this document will harmonize the use of those definitions.

End points based on tumor measurements are greatly influenced by response criteria. Overall and complete response rates usually can be assessed accurately in single-arm as well as randomized trials.

Table 3. Efficacy End Points

| End Point | Patients | Definition | Measured From |
|----------------------------|-------------|---|---------------------------|
| Primary | | | |
| Overall survival | All | Death as a result of any cause | Entry onto study |
| Progression-free survival | All | Disease progression or death as a result of any cause | Entry onto study |
| Secondary | | | |
| Event-free survival | All | Failure of treatment or death as a result of any cause | Entry onto study |
| Time to progression | All | Time to progression or death as a result of lymphoma | Entry onto study |
| Disease-free survival | in CR | Time to relapse or death as a result of lymphoma or acute toxicity of treatment | Documentation of response |
| Response duration | In CR or PR | Time to relapse or progression | Documentation of response |
| Lymphoma-specific survival | All | Time to death as a result of lymphoma | Entry onto study |
| Time to next treatment | All | Time to new treatment | End of primary treatment |

Abbreviations: CR, complete remission; PR, partial remission.

However, response rates do not necessarily influence other measures of overall clinical benefit or outcome in patients with lymphoma,⁵¹ and are not considered as important as other end points. Exceptions are phase II trials of novel new agents, in which identification of biologic activity is of interest. Durable complete responses, if associated with measures of clinical benefit, may also be relevant.

Overall Survival

Overall survival is the least ambiguous end point, although it usually is not optimal to use for a lymphoma clinical trial. Overall survival is defined as the time from entry onto the clinical trial (random assignment in a phase III study) until death as a result of any cause. Survival, as well as other time-dependent variables (PFS, event-free survival) should be measured in a randomized trial because data derived from historical controls are unreliable and subject to bias. Survival should be measured in the intent-to-treat population, including all patients even if they did not fulfill the eligibility criteria. A per-protocol analysis includes all patients who received the treatment to which they were assigned. A treatment-given analysis includes all patients who received a particular treatment. Both of these types of analyses should be interpreted with caution because they are subject to considerable bias.

PFS

PFS is defined as the time from entry onto a study until lymphoma progression or death as a result of any cause. PFS is often considered the preferred end point in lymphoma clinical trials, especially those involving incurable histologic subtypes (eg, follicular, other low-grade lymphoma, or mantle cell lymphoma). PFS reflects tumor growth, and therefore is interpretable earlier than the end point of overall survival. In addition, PFS is not confounded by the administration of subsequent therapy. However, in studies in which failure to respond without progression is considered an indication for another therapy, such patients should be censored at that point for the progression analysis. Whether a prolongation of PFS represents direct clinical benefit or is an acceptable surrogate for clinical benefit depends on the magnitude of the effect and the risk-benefit ratio of the therapy under investigation. Unlike survival, the precise date of progression is generally unknown. It may be defined as the first date of documentation of a new lesion or enlargement of a previous lesion, or the date of the scheduled clinic visit immediately after radiologic

assessment has been completed. When there is missing information, censoring of the data may be defined as the last date at which progression status was assessed adequately or the first date of unscheduled new antilymphoma treatment.

Event-Free Survival

Event-free survival (time to treatment failure) is measured from the time from study entry to any treatment failure including disease progression, or discontinuation of treatment for any reason (eg, disease progression, toxicity, patient preference, initiation of new treatment without documented progression, or death). This composite end point is generally not encouraged by regulatory agencies because it combines efficacy, toxicity, and patient withdrawal. However, it may be useful in the evaluation of some therapies such as those that are highly toxic.

Time to Progression

Time to progression (TTP) is defined as the time from study entry until documented lymphoma progression or death as a result of lymphoma. In TTP, deaths from other causes are censored either at the time of death or at an earlier time of assessment, representing a random pattern of loss from the study. TTP is not as useful as PFS unless the majority of deaths on a study are unrelated to the lymphoma due to the toxicity of the treatment and/or prolonged follow-up.

Disease-Free Survival

Disease-free survival is measured from the time of occurrence of disease-free state or attainment of a CR to disease recurrence or death as a result of lymphoma or acute toxicity of treatment. This definition may be complicated by deaths that occur during the follow-up period that are unrelated to the lymphoma, and there is controversy about whether such deaths should be considered as events or censored at the time of occurrence. Although it is often possible to identify those deaths related to the lymphoma, there is the potential for bias in the attribution of deaths.

Response Duration

Response duration is from the time when criteria for response (ie, CR or PR) are met, for which the event is the first documentation of relapse or progression.

Lymphoma-Specific Survival

Lymphoma-specific survival (eg, disease-specific survival, cause-specific survival) is defined as time from study entry to death as a result of lymphoma. This end point is potentially subject to bias because the exact cause of death is not always easy to ascertain. To minimize the risk of bias, the event should be recorded as death as a result of lymphoma, or as a result of toxicity from the drug. Death as a result of unknown causes should be attributed to the therapy.

Time to Next Treatment

For certain trials, time to next lymphoma treatment may be of interest, and is defined as time from the end of primary treatment until the institution of the next therapy.

Clinical Benefit

One of the most important end points for patients as well as for drug approval by regulatory agencies has been evidence of clinical benefit. Clinical benefit may reflect improvement in quality of life, or reduction in patient symptoms, transfusion requirements, frequent infections, or other parameters. Time to reappear-ance or progression of lymphoma-related symptoms can also be used in this end point.

We hope that these revised guidelines will improve comparability among studies, and facilitate new agent development leading to improved therapies for patients with lymphoma.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The authors indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Bruce D. Cheson, Beate Pfistner, Volker Diehl
Administrative support: Beate Pfistner, Volker Diehl
Collection and assembly of data: Bruce D. Cheson, Malik E. Juweid, Randy D. Gascoyne, Sandra J. Horning
Data analysis and interpretation: Bruce D. Cheson, Malik E. Juweid, Randy D. Gascoyne, Lena Specht, Sandra J. Horning, Bertrand Coiffier, Richard I. Fisher, Anton Hagenbeek, Sigrid Stroobants, T. Andrew Lister, Martin Dreyling, Joseph M. Connors, Massimo Federico, Volker Diehl
Manuscript writing: Bruce D. Cheson, Beate Pfistner, Malik E. Juweid, Randy D. Gascoyne, Lena Specht, Sandra J. Horning, Bertrand Coiffier, Richard I. Fisher, Anton Hagenbeek, Emanuele Zucca, Steven T. Rosen, Sigrid Stroobants, T. Andrew Lister, Richard T. Hoppe, Martin Dreyling, Kensei Tobinai, Julie M. Vose, Joseph M. Connors, Massimo Federico, Volker Diehl
Final approval of manuscript: Bruce D. Cheson, Beate Pfistner, Malik E. Juweid, Randy D. Gascoyne, Lena Specht, Sandra J. Horning, Bertrand Coiffier, Richard I. Fisher, Anton Hagenbeek, Emanuele Zucca, Steven T. Rosen, Sigrid Stroobants, T. Andrew Lister, Richard T. Hoppe, Martin Dreyling, Kensei Tobinai, Julie M. Vose, Joseph M. Connors, Massimo Federico, Volker Diehl

REFERENCES

- Cheson BD, Horning SJ, Coiffier B, et al: Report of an International Workshop to standardize response criteria for non-Hodgkin's lymphomas. *J Clin Oncol* 17:1244-1253, 1999
- Grillo-López AJ, Cheson BD, Horning SJ, et al: Response criteria for NHL: Importance of "normal" lymph node size and correlations with response rates. *Ann Oncol* 11:399-408, 2000
- Pfistner B, Diehl V, Cheson B: International harmonization of trial parameters in malignant lymphoma. *Eur J Haematol Suppl* July:53-54, 2005
- Bangerter M, Moog F, Buchmann I, et al: Whole-body 2-[18F]-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for accurate staging of Hodgkin's disease. *Ann Oncol* 9:1117-1122, 1998
- Spaepen K, Stroobants S, Dupont P, et al: Prognostic value of positron emission tomography (PET) with fluorine-18 fluorodeoxyglucose (¹⁸F)FDG) after first-line chemotherapy in non-Hodgkin's lymphoma: Is [¹⁸F]FDG-PET a valid alternative to conventional diagnostic methods? *J Clin Oncol* 19:414-419, 2001
- Spaepen K, Stroobants S, Dupont P, et al: Prognostic value of pretransplantation positron emission tomography using fluorine 18-fluorodeoxyglucose in patients with aggressive lymphoma treated with high-dose chemotherapy and stem cell transplantation. *Blood* 102:53-59, 2003
- Spaepen K, Stroobants S, Dupont P, et al: Early restaging positron emission tomography with 18F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin's lymphoma. *Ann Oncol* 13:1356-1363, 2002
- Jerusalem G, Beguin Y, Fassotte MF, et al: Whole-body positron emission tomography using 18F-fluorodeoxyglucose compared to standard procedures for staging patients with Hodgkin's disease. *Haematologica* 86:266-273, 2001
- Jerusalem G, Beguin Y, Fassotte MF, et al: Whole-body positron emission tomography using 18F-fluorodeoxyglucose for posttreatment evaluation in Hodgkin's disease and non-Hodgkin's lymphoma has higher diagnostic and prognostic value than classical computed tomography scan imaging. *Blood* 94:429-433, 1999
- Jerusalem G, Beguin Y, Najjar F, et al: Positron emission tomography (PET) with 18F-fluorodeoxyglucose (18F-FDG) for the staging of low-grade non-Hodgkin's lymphoma (NHL). *Ann Oncol* 12:825-830, 2001
- Jerusalem G, Warland V, Najjar F, et al: Whole-body 18F-FDG PET for the evaluation of patients with Hodgkin's disease and non-Hodgkin's lymphoma. *Nucl Med Commun* 20:13-20, 1999
- Zinzani PL, Magagnoli M, Chierichetti F, et al: The role of positron emission tomography (PET) in the management of lymphoma patients. *Ann Oncol* 10:1141-1143, 1999
- Wehrauch MR, Re D, Scheidhauer K, et al: Thoracic positron emission tomography using ¹⁸F-fluorodeoxyglucose for the evaluation of residual mediastinal Hodgkin disease. *Blood* 98:2930-2934, 2001
- Naumann R, Vaic A, Beuthien-Baumann B, et al: Prognostic value of positron emission tomography in the evaluation of post-treatment residual mass in patients with Hodgkin's disease and non-Hodgkin's lymphoma. *Br J Haematol* 115:793-800, 2001
- Kostakoglu L, Leonard JP, Kuji I, et al: Comparison of fluorine-18 fluorodeoxyglucose positron emission tomography and Ga-67 scintigraphy in evaluation of lymphoma. *Cancer* 94:879-888, 2002
- 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* 90:620-625, 2004
- Munker R, Glass J, Griffeth LK, et al: Contribution of PET imaging to the initial staging and prognosis of patients with Hodgkin's disease. *Ann Oncol* 15:1699-1704, 2004
- Mikhaeel NG, Hutchings M, Fields PA, et al: FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol* 16:1514-1523, 2005
- Juweid M, Cheson BD: Positron emission tomography (PET) in post-therapy assessment of cancer. *N Engl J Med* 354:496-507, 2006
- Juweid M, Wiseman GA, Vose JM, et al: Response assessment of aggressive non-Hodgkin's lymphoma by integrated International Workshop criteria (IWC) and 18F-fluorodeoxyglucose positron emission tomography (PET). *J Clin Oncol* 23:4652-4661, 2005
- Haicion C, Itti E, Rahmouni A, et al: [¹⁸F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: An early prognostic tool for predicting patient outcome. *Blood* 106:1376-1381, 2005
- Hutchings M, Loft A, Hansen M, et al: Positron emission tomography with or without computed tomography in the primary staging of Hodgkin's lymphoma. *Haematologica* 91:482-489, 2006
- Querellou S, Valette F, Bodet-Milin C, et al: FDG-PET/CT predicts outcome in patients with aggressive non-Hodgkin's lymphoma and Hodgkin's disease. *Ann Hematol* 15:759-767, 2006
- Gallamini A, Rigacci L, Merli F, et al: Predictive value of positron emission tomography performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. *Haematologica* 91:475-481, 2006
- Hutchings M, Loft A, Hansen M, et al: FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood* 107:52-59, 2006
- 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* 91:889-899, 2001

27. Stumpe KD, Urbinelli M, Steinert HC, et al: Whole-body positron emission tomography using fluorodeoxyglucose for staging of lymphoma: Effectiveness and comparison with computed tomography. *Eur J Nucl Med* 25:721-728, 1998

28. 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* 112:262-268, 2002

29. Juweid ME, Stroobants S, Hoekstra OS, et al: Use of positron emission tomography for response assessment of lymphoma: Consensus recommendations of the Imaging Subcommittee of the International Harmonization Project in Lymphoma. *J Clin Oncol* 10.1200/JCO.2006.08.2305

30. Hoffmann M, Kletter K, Diemling M, et al: Positron emission tomography with fluorine-18-2-fluoro-2-deoxy-D-glucose (F18-FDG) does not visualize extranodal B-cell lymphoma of the mucosa-associated lymphoid tissue (MALT)-type. *Ann Oncol* 10:1185-1189, 1999

31. Elstrom R, Guan L, Baker G, et al: Utility of FDG-PET scanning in lymphoma by WHO classification. *Blood* 101:3875-3876, 2003

32. Karam M, Novak L, Cyriac J, et al: Role of fluorine-18 fluoro-deoxyglucose positron emission tomography scan in the evaluation and follow-up of patients with low-grade lymphomas. *Cancer* 107:175-183, 2006

33. Lewis PJ, Salama A: Uptake of fluorine-18-fluorodeoxyglucose in sarcoidosis. *J Nucl Med* 35:1647-1649, 1994

34. Castellucci P, Nanni C, Farsad M, et al: Potential pitfalls of ¹⁸F-FDG PET in a large series of patients treated for malignant lymphoma: Preva-

lence and scan interpretation. *Nucl Med Commun* 26:689-694, 2005

35. Kostakoglu L, Coleman M, Leonard JP, et al: PET predicts prognosis after 1 cycle of chemotherapy in aggressive lymphoma and Hodgkin's disease. *J Nucl Med* 43:1018-1027, 2002

36. Zinzani PL, Tani M, Fanti S, et al: Early positron emission tomography (PET) restaging: A predictive final response in Hodgkin's disease patients. *Ann Oncol* 17:1296-1300, 2006

37. Abrey LE, Batchelor TT, Ferreri AJ, et al: Report of an international workshop to standardize baseline evaluation and response criteria for primary CNS lymphoma. *J Clin Oncol* 23:5034-5043, 2005

38. Bertoni F, Conconi A, Capella C, et al: Molecular follow-up in gastric mucosa-associated lymphoid tissue lymphomas: Early analysis of the LY03 cooperative trial. *Blood* 99:2541-2544, 2002

39. Thiede C, Wundisch T, Alpen B, et al: Long-term persistence of monoclonal B cells after cure of *Helicobacter pylori* infection and complete histologic remission in gastric mucosa-associated lymphoid tissue B-cell lymphoma. *J Clin Oncol* 19:1600-1609, 2001

40. Wotherspoon AC, Dogliani C, Diss TC, et al: Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of *Helicobacter pylori*. *Lancet* 342:575-577, 1993

41. Neubauer A, Thiede C, Morgner A, et al: Cure of *Helicobacter pylori* infection and duration of remission of low-grade mucosa-associated lymphoid tissue lymphoma. *J Natl Cancer Inst* 89:1350-1355, 1997

42. Copie-Bergman C, Gaulard P, Lavergne-Slove A, et al: Proposal for a new histological grading system for post-treatment evaluation of gastric MALT lymphoma. *Gut* 52:1656, 2003

43. Copie-Bergman C, Capella C, Motta T, et al: Validation of the GELA scoring system for evaluating

gastric biopsies from patients with MALT lymphoma following eradication of *Helicobacter pylori*. *Ann Oncol* 16:v94, 2005 (suppl 5; abstr 194)

44. Weeks JC, Yeap BY, Canellos GP, et al: Value of follow-up procedures in patients with large-cell lymphoma who achieve a complete remission. *J Clin Oncol* 9:1196-1203, 1991

45. Oh YK, Ha CS, Samuels BI, et al: Stages I-III follicular lymphoma: Role of CT of the abdomen and pelvis in follow-up studies. *Radiology* 210:483-486, 1999

46. Foltz LM, Song KW, Connors JM: Who actually detects relapse in Hodgkin lymphoma: Patient or physician. *Blood* 104 (part 1):853a-854a, 2004 (abstr 3124)

47. Liedtke M, Hamlin PA, Moskowitz CH, et al: Surveillance imaging during remission identifies a group of patients with more favorable aggressive NHL at time of relapse: A retrospective analysis of a uniformly-treated patient population. *Ann Oncol* 17:909-913, 2006

48. Jerusalem G, Beguin Y, Fassotte MF, et al: Early detection of relapse by whole-body positron emission tomography in the follow-up of patients with Hodgkin's disease. *Ann Oncol* 14:123-130, 2003

49. Hoppe RT, Advani RH, Bierman PJ, et al: Hodgkin disease/lymphoma: Clinical practice guidelines in oncology. *J Natl Comp Cancer Net* 4:210-230, 2006

50. Zelenetz AD, Advani RH, Buadi F, et al: Non-Hodgkin's lymphoma: Clinical practice guidelines in oncology. *J Natl Comp Cancer Net* 4:258-310, 2006

51. Kimby E, Björkholm M, Gahrton G, et al: Chlorambucil/prednisone vs. CHOP in symptomatic low-grade non-Hodgkin's lymphomas: A randomized trial from the Lymphoma Group of Central Sweden. *Ann Oncol* 5:67-71, 1994 (suppl 2)

Acknowledgment

We thank our other colleagues who provided input into these guidelines: Lauren Abrey, Ralph Meyer, Otto S. Hoekstra, Gregory Wiseman, Markus Dietlein, Sven Reske, Ali Guermazi, Markus Schwaiger, Mary Gospodarowicz, Michael Pfreundschuh and the German High-Grade Lymphoma Study Group, Myriam Mendila, David Schenkein, Nancy Valente, Daphne de Jong, the EORTC Lymphoma Group, and the Nordic Lymphoma Study Group, Josée Zijlstra, Michinori Ogura, and the JCOG Lymphoma Study Group, A.J. Ferreri, and C. Copie-Bergmann.