Staging and response assessment in lymphomas: the new Lugano classification

Bruce D. Cheson

Division of Hematology/Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Hospital, Washington DC 20007, USA

Correspondence to: Bruce D. Cheson, MD. Professor of Medicine; Deputy Chief, Hematology-Oncology; Head of Hematology, Division of Hematology/Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Hospital, Washington DC 20007, USA.
Email: bdc4@georgetown.edu.

Abstract: Staging and response criteria were initially developed for Hodgkin lymphoma (HL) over 60 years ago, but not until 1999 were response criteria published for non-HL (NHL). Revisions to these criteria for both NHL and HL were published in 2007 by an international working group, incorporating PET for response assessment, and were widely adopted. After years of experience with these criteria, a workshop including representatives of most major international lymphoma cooperative groups and cancer centers was held at the 11th International Conference on Malignant Lymphoma (ICML) in June, 2011 to determine what changes were needed. An Imaging Task Force was created to update the relevance of existing imaging for staging, reassess the role of interim PET-CT, standardize PET-CT reporting, and to evaluate the potential prognostic value of quantitative analyses using PET and CT. A clinical task force was charged with assessing the potential of PET-CT to modify initial staging. A subsequent workshop was held at ICML-12, June 2013. Conclusions included: PET-CT should now be used to stage FDG-avid lymphomas; for others, CT will define stage. Whereas Ann Arbor classification will still be used for disease localization, patients should be treated as limited disease [I (E), II (E)], or extensive disease [III-IV (E)], directed by prognostic and risk factors. Since symptom designation A and B are frequently neither recorded nor accurate, and are not prognostic in most widely used prognostic indices for HL or the various types of NHL, these designations need only be applied to the limited clinical situations where they impact treatment decisions (e.g., stage II HL). PET-CT can replace the bone marrow biopsy (BMBx) for HL. A positive PET of bone or bone marrow is adequate to designate advanced stage in DLBCL. However, BMBs can be considered in DLBCL with no PET evidence of BM involvement, if identification of discordant histology is relevant for patient management, or if the results would alter treatment. BMBx remains recommended for staging of other histologies, primarily if it will impact therapy. PET-CT will be used to assess response in FDG-avid histologies using the 5-point scale, and included in new PET-based response criteria, but CT should be used in non-avid histologies. The definition of PD can be based on a single node, but must consider the potential for flare reactions seen early in treatment with newer targeted agents which can mimic disease progression. Routine surveillance scans are strongly discouraged, and the number of scans should be minimized in practice and in clinical trials, when not a direct study question. Hopefully, these recommendations will improve the conduct of clinical trials and patient management.

Keywords: Staging; response; PET-CT; lymphoma
Introduction

With newer and more effective therapies for the treatment of lymphomas the need for universally accepted, standardized criteria for staging and response becomes even more critical. Such guidelines permit the reporting of uniform endpoints, facilitate comparisons amongst studies, help identify promising regimens, and facilitate evaluation and approval by regulatory agencies. This manuscript describes the evolution of staging and response criteria leading to the most recent Lugano classification (1).

Staging is used to define the anatomic distribution of the disease for purposes of prognosis and treatment planning. The first such system in lymphoma was a three stage classification published by Peters in 1950 (2) and was followed by the Rye classification of 1966 (3), and then the Ann Arbor classification of 1971 (4), which has remained the most widely used, until the present day. These were all designed for the initial evaluation of patients with Hodgkin lymphoma (HL) to assist radiation oncologists in planning their radiation delivery, as radiation was the only effective treatment at the time, and did not seem to be applicable to non-HL (NHL) at the time (5). Chemotherapy for HL was reserved for patients with advanced disease because of its toxicity and unknown efficacy. The Ann Arbor classification subdivided patients into four stages and further subdivided them based on the absence of (A) or presence of (B) disease-related symptoms: fevers >38 °C, drenching night sweats, unexplained fevers, and unintentional weight loss of >10% over the prior 6 months. The term “E” was used to designate proximal/contiguous extranodal disease. Imaging studies included intravenous pyelogran, ultrasound, liver-spleen scan, and the torture of the lymphangiogram. Staging laparotomy was performed for all but those with obviously advanced disease. As a result of that procedure, patients were given an additional designation based on the presence of disease involving the spleen, liver, bone, lung, and other sites. As technology and treatment improved, the Ann Arbor classification underwent modification. The Cotswold revision incorporated CT scans for staging (6). This improved imaging technique, along with the front-line use of effective, systemic chemotherapy rendered surgical exploration unnecessary (7,8). Bulky disease was designated “X”, and the term complete remission unconfirmed (CRu) was created for those patients with a residual mass posttreatment thought more likely to be fibrosis than actual tumor.

In 1999, an international working group made the first recommendations for response assessment for NHL (9), which were also adopted for HL. The terms complete remission (CR), partial remission (PR), CRu, stable disease (SD), relapsed disease (RD) and progressive disease (PD) were codified. These guidelines became universally accepted by clinicians, the pharmaceutical industry, and regulatory agencies. However, as the recommendations were more widely used, it became apparent that a number of issues needed to be addressed. Some terms were unclear or misinterpreted (e.g., CRu), assessment was based largely on physical examination and standard laboratory tests, with imaging limited to chest X-ray, CT scans or MRI, gallium scan, and visual bone marrow evaluation. PET scans were invented in 1987 and first applied to lymphoma in 1990. The advantage of PET scans over CT was that PET was able to distinguish viable tumor from scar and fibrosis. PET could, therefore, eliminate the term CRu and clarify other issues in prior recommendations for both NHL and HL (10). PET, in conjunction with the availability of FDG-PET, immunohistochemistry and flow cytometry of the bone marrow justified the opportunity to update these recommendations with the international harmonization project recommendations (11). These guidelines published in 2007 incorporated PET scans for response became the international standard, and have been validated by other groups (12).

Following extensive experience with these criteria, and recognizing the progress made following their publication, particularly in imaging techniques, a workshop was held at the 11th International Conference on Malignant Lymphoma (ICML) in Lugano, Switzerland, June 2011. This workshop was attended by leading hematologists, oncologists, radiation oncologists, pathologists, radiologists, and nuclear medicine physicians, representing major lymphoma clinical trials groups and cancer centers in North America, Europe, Japan, and Australasia. The aims were to develop universally accepted, unambiguous, improved staging and response criteria for HL and NHL, relevant for community physicians, investigator-led trials, cooperative group and registration trials that would permit improved lymphoma patient evaluation, enhance comparisons amongst studies, and simplify evaluation of new therapies. Two task forces were formed, one to identify clinical issues such as the current relevance of the Ann Arbor staging system, whether a simplification was possible, to evaluate the role of bone marrow biopsies and chest X-rays, improve organ assessment, redefine PD, to take into account novel consequences of new drugs, including tumor flare reactions; and standardize follow-up. The charge to the imaging task
force was to update the relevance of existing imaging for staging and evaluating bulk and bone marrow involvement, to clarify the role of contrast enhanced CT scans (CeCT) versus low dose CT with PET-CT, discuss the role of interim PET-CT, and to evaluate the potential prognostic value of quantitative analyses using PET-CT and CT. A series of meetings and conference calls followed, and a subsequent workshop was held at the 12th ICML in Lugano, Switzerland in 2013. The result of those deliberations was the new Lugano classification (1).

**Initial patient evaluation**

The essential factor in patient management is making an accurate diagnosis. Fine needle aspiration of lymph nodes has a high likelihood of a false negative result, obtaining a nonrepresentative sample resulting in misdiagnosis, and not providing sufficient tissue for molecular and genetic studies, and, therefore, is strongly discouraged (13). An excisional biopsy is preferred, although a core biopsy may be acceptable when the former is not possible.

The next step is to perform a complete clinical evaluation which includes a careful history and physical examination, recording disease-related symptoms, measuring nodes and spleen size. Necessary laboratory tests include a complete blood count with differential, liver and renal function tests, uric acid, lactate dehydrogenase, and others as relevant to diagnosis (e.g., HIV), treatment (e.g., hepatitis B and C), or prognosis based on the currently used scoring systems (e.g., IPI, FLIPI, F-2, MIPI, IPS, etc.) (14-18).

**Staging**

A large number of studies have definitively demonstrated that PET-CT is the most sensitive of current imaging studies, and is highly specific, not only for response assessment, but for pretreatment determination of disease localization (19-23). Moreover, since it was accepted as the standard for restaging, it was reasonable to accept it for pretreatment staging (11). Thus, the new classification formally includes PET-CT as the standard imaging study for staging of FDG-avid lymphomas, which include all except for CLL/SLL, mycosis fungoides, and marginal zone (unless local radiation is considered as the sole modality of treatment) (24). Whereas, the goal of more sensitive staging techniques is that fewer patients undertreated or overtreated, a consequence is stage migration which can limit the ability to use historical comparisons. A contrast enhanced CT scan is also recommended if measuring nodes is important, or for radiotherapy planning.

As noted above, the four stage Ann Arbor system was developed primarily to direct radiation therapy. Although in some situations, stages I and II NHL may be approached differently, they are often treated the same. Those with stages III and IV patients are almost always treated similarly. Therefore, there is rationale to reclassify patients as Limited or Advanced disease for treatment purposes, factoring in important risk factors, rather than simply by Ann Arbor stage. In HL, stages I and II may be approached differently, but III and IV are managed in a similar manner. Therefore I and II can be considered as limited disease, III and IV as advanced, and treatment based on various prognostic factors. Stage IIB can be approached as either limited or advanced, directed by disease setting and other risk factors (Table 1). In addition, the presence or absence of the disease-related symptoms of fevers, unexplained weight loss or drenching night sweats does not appear to correlate with outcome in any of the commonly used prognostic scores in NHL (e.g., IPI, FLIPI, F-2, MIPI) and, thus, A and B do not need to be applied to NHL as they do not impact patient approach. In contrast A and B are still used to make some treatment decisions in early stage HL, and, therefore, are retained in that setting. Thus, the designation “X” for bulky is no longer used, but, instead, the greatest diameter of the largest mass should be recorded. The standard 10 cm or a third the transverse diameter of the chest was retained for HL, although as a single mass rather than a collection of smaller nodes with surrounding connective tissue. No consistent definition for bulky disease has been determined for NHL and, therefore, none was provided, with hopes that future studies would generate a data based definition, perhaps using tumor volume.

It is clear that certain components of staging have persisted more for historical than scientific reasons. For example, numerous studies have questioned the role of the bone marrow biopsy (BMBx) in staging of HL and diffuse large B-cell NHL (25-32). Recently, El-Galaly et al. (29) reported on 454 patients with newly diagnosed HL. In 18%, focal bone/bone marrow lesions were noted on PET-CT, but in only 8% by trephine biopsy. No patient with stage I-II by PET had a positive biopsy, and patients with a positive biopsy also had other evidence of advanced disease. All patients who were upstage by biopsy went from stage III-IV, with no alteration in treatment plan. Khan et al. (32) identified bone marrow involvement in 27% of 130 patients with DLBCL; 33 by PET and 14 by BM biopsy. PET

---

© Chinese Clinical Oncology. All rights reserved. www.thecco.net Chin Clin Oncol 2015:4(1):5
identified all patients with a positive biopsy, and cases with a positive PET had an outcome comparable to other patients with stage IV disease without a positive bone marrow. Based on these data and others, the new recommendation is that a BMBx is no longer needed for the routine staging of patients with HL, and only for those with DLBCL with a negative PET for whom identification of occult discordant histology is clinically important (33). However, a bone marrow remains a standard test for staging other lymphomas.

A chest X-ray has also been a traditional test in staging, especially for HL. However, CT scans are the superior test (34), and there no longer appears to be a need for the radiograph in the routine assessment of patients with HL.

**Response assessment**

Restaging scans are generally performed 6-8 weeks following completion of treatment in the setting of regimens with a fixed number of cycles. However, a different time point may be needed for regimens where treatment is continuous or if maximum response is expected to be delayed. If a contrast enhanced CT was performed at baseline and showed no additional findings beyond those identified by PET, then a lower dose CT during restaging is adequate.

Although PET-CT was considered standard for staging in the 2007 guidelines (11), visual interpretation was used with the mediastinal blood pool as the comparator. Interobserver variability was a problem with this approach (35). The Deauville 5-point scale has been validated as a reliable means of interpreting scans in FDG-avid histologies (Table 1) and improves consistency of interpretation (36). Therefore, it will now be the standard for interpretation of response (Table 2). CT scans should still be used for the variably avid or negative histologies.

**Patient follow-up**

Once response has been assessed, further imaging studies should be performed judiciously and prompted by clinical indications. Surveillance scans are not justified strongly discouraged after that point, especially in HL and DLBCL. They are not cost-effective, and are associated with a high likelihood of false-positive results, leading to unnecessary imaging and biopsies. Several studies have shown that 70-80% of the time it is the patient or the physician who identifies recurrence, and additional scans lead to false positive results and are not cost-effective (37,38). A repeat study may be needed if the posttreatment scan was equivocal, or conservatively in patients with an indolent NHL and intraabdominal or retroperitoneal disease.

**Conclusions**

These new criteria have a number of features that distinguish them from the 2007 recommendations (11) (Table 3). FDG-PET-CT is the new standard for staging of all FDG-avid histologies. A modified Ann Arbor classification has been retained for extent of disease. However, patients should be treated more on the basis of risk factors. Vestiges of the past, including routine chest X-rays, the use of “X”, and bone marrow biopsies in HL and DLBCL are no
Table 2 Revised response criteria for lymphoma

<table>
<thead>
<tr>
<th>Response criteria</th>
<th>PET-CT-based response</th>
<th>CT-based response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complete remission (CR)</strong></td>
<td><strong>Lymph nodes and extralymphatic sites</strong> Score 1, 2, or 3* with or without a residual mass on 5-PS** It is recognized that in Waldeyer’s ring or extranodal sites with high physiological uptake or with activation within spleen or marrow, e.g., with chemotherapy or myeloid colony stimulating factors, uptake may be greater than normal mediastinum and/or liver. In this circumstance, CMR may be inferred if uptake at sites of initial involvement is no greater than surrounding normal tissue even if the tissue has high physiological uptake</td>
<td>Target nodes/nodal masses must regress to ≤1.5 cm in LDi No extralymphatic sites of disease</td>
</tr>
<tr>
<td><strong>Non-measured lesion</strong></td>
<td>Not applicable</td>
<td>Absent</td>
</tr>
<tr>
<td><strong>Organ enlargement</strong></td>
<td>Not applicable</td>
<td>Regress to normal</td>
</tr>
<tr>
<td><strong>New lesions</strong></td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td><strong>Bone marrow</strong></td>
<td>No evidence of FDG-avid disease in marrow</td>
<td>Normal by morphology; if indeterminate, IHC negative</td>
</tr>
<tr>
<td><strong>Partial remission (PR)</strong></td>
<td><strong>Lymph nodes and extralymphatic sites</strong> Score 4 or 5** with reduced uptake compared with baseline and residual mass(es) of any size At interim these findings suggest responding disease At end of treatment these findings indicate residual disease</td>
<td>≥50% decrease in SPD of up to 6 target measurable nodes and extranodal sites When a lesion is too small to measure on CT, assign 5 mm × 5 mm as the default value When no longer visible, 0 mm × 0 mm For a node &gt;5 mm × 5 mm, but smaller than normal, use actual measurement for calculation</td>
</tr>
<tr>
<td><strong>Non-measured lesions</strong></td>
<td>Not applicable</td>
<td>Absent/normal, regressed, but no increase Spleen must have regressed by &gt;50% in length beyond normal</td>
</tr>
<tr>
<td><strong>Organ enlargement</strong></td>
<td>Not applicable</td>
<td>Spleen must have regressed by &gt;50% in length beyond normal</td>
</tr>
<tr>
<td><strong>New lesions</strong></td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td><strong>Bone marrow</strong></td>
<td>Residual uptake higher than uptake in normal marrow but reduced compared with baseline (diffuse uptake compatible with reactive changes from chemotherapy allowed). If there are persistent focal changes in the marrow in the context of a nodal response, consideration should be given to further evaluation with MRI or biopsy, or an interval scan</td>
<td>Not applicable</td>
</tr>
<tr>
<td><strong>No response or stable disease (SD)</strong></td>
<td><strong>Target nodes/nodal masses, extranodal lesions</strong> No response: score 4 or 5 with no significant change in FDG uptake from baseline, at interim or end of treatment <strong>Non-measured lesions</strong> Not applicable <strong>Organ enlargement</strong> Not applicable <strong>New lesions</strong> None <strong>Bone marrow</strong> No change from baseline</td>
<td>Stable disease: &lt;50% decrease from baseline in SPD of up to 6 dominant, measurable nodes and extranodal sites; no criteria for PD are met No increase consistent with progression No increase consistent with progression None Not applicable</td>
</tr>
</tbody>
</table>

Table 2 (continued)
Table 2 (continued)

<table>
<thead>
<tr>
<th>Response criteria</th>
<th>PET-CT-based response</th>
<th>CT-based response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progressive disease (PD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual target nodes/nodal masses, extranodal lesions</td>
<td>Score 4, 5 with an increase in intensity of uptake from baseline and/or new FDG-avid foci consistent with lymphoma at interim or end of treatment assessment</td>
<td>PPD progression:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>An individual node must be abnormal with:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LDi &gt;1.5 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Increase by ≥50% from PPD nadir</td>
</tr>
<tr>
<td></td>
<td></td>
<td>An increase in LDi or SDi from nadir</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 0.5 cm for lesions ≤2 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 1.0 cm for lesions &gt;2 cm</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In the setting of splenomegaly, the splenic length must increase by &gt;50% of the extent of its prior increase beyond baseline (e.g., a 15 cm spleen must increase to &gt;16 cm). If no prior splenomegaly, must increase by at least 2 cm from baseline</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New or recurrent splenomegaly</td>
</tr>
<tr>
<td>Non-measured lesions</td>
<td>None</td>
<td>New or clear progression of pre-existing non-measured lesions</td>
</tr>
<tr>
<td>New lesions</td>
<td>New FDG-avid foci consistent with lymphoma rather than another etiology, e.g. infection, inflammation. If uncertain regarding etiology of new lesions, biopsy or interval scan may be considered</td>
<td>Regrowth of previously resolved lesions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A new node &gt;1.5 cm in any axis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A new extranodal site &gt;1.0 cm in any axis if less than 1.0 cm in any axis, its presence must be unequivocal and must be attributable to lymphoma</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Assessable disease of any size unequivocally attributable to lymphoma</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>New or recurrent FDG avid foci</td>
<td>New or recurrent involvement</td>
</tr>
</tbody>
</table>

Measured dominant lesions: up to six of the largest dominant nodes, nodal masses and extranodal lesions selected to be clearly measurable in 2 diameters. Nodes should preferably be from disparate regions of the body, and should include, where applicable, mediastinal and retroperitoneal areas. Non-nodal lesions include those in solid organs, e.g., liver, spleen, kidneys, lungs, etc., gastrointestinal involvement, cutaneous lesions of those noted on palpation. Non-measured lesions: any disease not selected as measured, dominant disease and truly assessable disease should be considered not measured. These sites include any nodes, nodal masses, and extranodal sites not selected as dominant, measurable or which do not meet the requirements for measurability, but are still considered abnormal. As well as truly assessable disease which is any site of suspected disease that would be difficult to follow quantitatively with measurement, including pleural effusions, ascites, bone lesions, leptomeningeal disease, abdominal masses and other lesions that cannot be confirmed and followed by imaging. In Waldenyer’s ring or in extranodal sites, e.g., gastrointestinal tract, liver, and bone marrow, FDG uptake may be greater than mediastinum with CMR, but should be no higher than surrounding normal physiologic uptake, e.g., with marrow activation due to chemotherapy or myeloid growth factors. SPD, sum of the product of the perpendicular diameters for multiple lesions; LDi, longest transverse diameter of a lesion; SDi, shortest axis perpendicular to the LDi; PPD, cross product of the LDi and perpendicular diameter. *, score 3 in many patients indicates a good prognosis with standard treatment, especially if at the time of an interim scan. However in trials involving PET where de-escalation is investigated, it may be preferable to consider score 3 as inadequate response (to avoid undertreatment); **, PET five point scale (5-PS): 1, no uptake above background; 2, uptake ≤ mediastinum***; 3, uptake >mediastinum, but ≤ liver; 4, uptake moderately > liver; 5, uptake markedly higher than liver and/or new lesions; X, new areas of uptake unlikely to be related to lymphoma. Reprinted from reference (1).
longer indicated, and the designation A and B primarily for limited stage HL.

Nevertheless, a number of issues remain to be clarified. For example, the size of a nodal mass to be considered “bulky” disease with a clinically distinct approach and/or outcome remains to be elucidated. Whether tumor volume can be reliably incorporated into current clinical practice is unclear. Studies are evaluating more quantitative measures of assessing response to improve the predictability of PET, including the percent of reduction in the standard uptake to (SUV), the percent decrease in the size of the mass, alone or in combination (39-41). Until these questions and others are resolved, and newer, more powerful molecular and genetic prognostic factors are identified that impact treatment, these new staging and response criteria should serve to improve patient management.

**Acknowledgements**

The author expresses his most deep appreciation to his colleagues who were largely responsible for the original Lugano Classification manuscript: Richard I. Fisher, Sally F. Barrington, Emmanuele Zucca, Franco Cavalli, Lawrence B. Schwartz, and T. Andrew Lister.

**Disclosure:** The author declares no conflicts of interest.

**References**


1993;329:987-94.


