Response Assessment Criteria and their Applications in Lymphoma: Part 1

Mateen C. Moghbel¹, Lale Kostakoglu², Katherine Zukotynski³, Delphine L. Chen⁴, Helen Nadel⁵, Ryan Niederkohr⁶, Erik Mittra¹

Affiliations
1. Stanford University Medical Center, Stanford, CA, USA
2. Mount Sinai Medical Center, New York, NY, USA
3. McMaster University, Hamilton, ON, Canada
4. Washington University, St. Louis, MO, USA
5. University of British Columbia, Vancouver, BC, Canada
6. Kaiser Permanente, Santa Clara, CA, USA

First Author: Mateen Moghbel

Corresponding Author: Erik Mittra
Department of Radiology, Division of Nuclear Medicine
Stanford University Medical Center
300 Pasteur Dr., Rm. H2200
Stanford, CA 94305-5281
P: 650.721.2024
F: 650.498.5047

Counts:
Abstract: 135 words
Printed Text: 6139 words
Supplemental Text: 623 words
Tables: 7
Supplemental Tables: 4
ABSTRACT

The effectiveness of cancer therapy, both in individual patients and across populations, requires a systematic and reproducible method for evaluating response to treatment. Early efforts to meet this need resulted in the creation of numerous guidelines for quantifying post-therapy changes in disease extent, both anatomically and metabolically. Over the past few years, criteria for disease response classification have been developed for specific cancer histologies. To date, the spectrum of disease broadly referred to as lymphoma is perhaps the most common pathology for which disease response classification is used. This review article provides an overview of the existing response assessment criteria for lymphoma, while highlighting their respective methodologies and validities. Concerns over the technical complexity and arbitrary thresholds of many of these criteria, which have impeded the long-standing endeavor of standardizing response assessment, are also discussed.

KEYWORDS
Lymphoma, PET, CT, RECIST, PERCIST
INTRODUCTION

Lymphoma comprises a heterogeneous collection of lymphoproliferative malignancies with varying clinical behavior and response profiles. These disorders are commonly categorized as either Hodgkin’s lymphoma (HL) or non-Hodgkin’s lymphoma (NHL), with the latter group constituting the vast majority of cases. HL tends to be less aggressive in nature and carries a relatively high 5-year survival rate of 85.3% (1). In 2015, this subtype of lymphoma was diagnosed in an estimated 9,050 patients and caused 1,150 deaths in the United States (2). By comparison, NHL includes dozens of distinct conditions with varying etiologies and prognoses. Together, these conditions accounted for approximately 71,850 new cases and 19,790 deaths in the United States in 2015 (2), with a 5-year survival rate of 69.3% (1). The World Health Organization (WHO) guidelines subdivide NHL according to cell lineage into mature B-cell neoplasms and mature T-cell and NK-cell neoplasms (3). Diffuse large B-cell lymphoma (DLBCL), which falls into the first classification, represents approximately 40% of all cases of NHL, making it the most common form of the disease (4).

The nodular enlargements characteristic of lymphoma were noted in the medical literature as early as 1661 (5), but the constellation of “lymph node and spleen enlargement, cachexia and fatal termination” was first described by Thomas Hodgkin in 1832 (6). The development of modern treatments occurred over a century later, when the discovery of marked lymphoid and myeloid suppression in soldiers exposed to mustard gas during the Second World War led Louis S. Goodman and Alfred Gilman to test the effects of a related compound—nitrogen mustard—on patients with lymphoma and other hematological diseases (7).

Even these early chemotherapeutic agents required an objective means of evaluating their in vivo effectiveness in human subjects. Initially, standardized methods for the manual measurement of tumor size pre- and post-therapy were proposed for this purpose. But as anatomical medical imaging techniques, most notably computed tomography (CT), became available, an array of novel guidelines for response assessment were developed. More recently, functional information from positron emission tomography (PET) has been integrated to complement the anatomical information of CT. Currently, numerous response assessment criteria that rely on CT and PET individually, as well as a handful of criteria that combine these imaging modalities, have been reported for assessing treatment response in both solid tumors and hematologic malignancies (Supplemental Table 1). Although recent progress has been made
towards the standardization of response assessment, the clinical and research communities remain somewhat fragmented in their use of these various criteria. This review article outlines the available criteria and highlights what differentiates them in an attempt to facilitate a more uniform approach to response assessment.

HISTORICAL REVIEW OF RESPONSE ASSESSMENT IN SOLID TUMORS

From the development of the first chemotherapeutic agents in the 1940s to the advent of modern imaging techniques in the 1970s, objective and systematic assessment of treatment response depended largely on physical examination (8). However, palpation as a method of assessing response was imprecise, as demonstrated by a 1976 study by Moertel and Hanley in which sixteen oncologists palpated and measured twelve simulated tumor masses using “variable clinical methods” (9). The authors found that criteria that defined response as 25% and 50% reductions in the perpendicular diameters of these palpated tumors resulted in false positive readings in 19-25% and 6.8-7.8% of cases, respectively.

With the goal of achieving “the standardization of reporting results of cancer treatment,” the WHO held a series of meetings between 1977 and 1979 that culminated in the publication of a handbook outlining response assessment criteria that were widely publicized and rapidly adopted (10,11). The criteria called for bi-dimensional tumor measurements to be obtained prior to and following therapy, and the product of these bi-dimensional measurements to be calculated and summed across several sites of disease to form a single parameter with which to assess response. The changes in these parameters over time classified patients into one of four response groups: complete response (CR), partial response (PR), no response (NR), and progressive disease (PD) (Supplemental Table 2).

Although these guidelines made strides toward standardization of response assessment, they did not explicitly specify critical factors, including the number of masses to be measured and the minimum measurable size of a tumor (12). As a result of these ambiguities, as well as the introduction of imaging modalities such as CT, the WHO criteria eventually became the subject of reinterpretation by various research organizations and clinical groups, thus undermining the standardization it was designed to promote.

In order to address the gradual divergence of response assessment, institutions such as the National Cancer Institute (NCI) and the European Organisation for Research and Treatment of
Cancer (EORTC) began revisiting the WHO criteria throughout the 1990s with the goal of developing new guidelines that would re-standardize the practice of evaluating response to therapy. In 1999, the EORTC released its own recommendations for patient preparation prior to imaging, image acquisition and analysis, tumor sampling, and classification of tumor response (13). These were among the first guidelines to employ a functional imaging modality, namely PET, as a means of assessing treatment response (Supplemental Table 3). The PET radiotracer $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) was used to measure metabolic activity and tumor aggressiveness. Moreover, $^{18}$F-FDG was shown to delineate the metabolically active tumors borders providing insight into individual tumor biology. These metabolic classifications of treatment response laid the groundwork for similar $^{18}$F-FDG-based criteria in the years that followed.

The incorporation of PET imaging helped to address the issue of residual masses detected after therapy, which frequently comprise inflammatory, necrotic, and fibrotic tissue rather than residual disease (14-16). This phenomenon proved especially problematic for response assessment criteria for lymphoma that relied solely on anatomic imaging. Approximately 40% of NHL patients and 20% of HL patients continue to exhibit residual mediastinal or abdominal masses on CT following therapy (17,18). In studies where such patients were restaged via laparotomy, between 80% and 95% of residual masses were shown to be non-malignant on pathology (17,19). Moreover, the presence of residual masses on imaging was not found to be associated with time to relapse or survival (18). Therefore, by shedding light on the metabolic activity and thereby viability of these masses, PET overcame a significant limitation of CT-based response assessment for lymphoma (20).

In 2000, shortly after the EORTC devised its PET-based criteria, a collaboration between the NCI and EORTC provided a new set of CT-based guidelines called Response Evaluation Criteria In Solid Tumors (RECIST) (21). Unlike earlier anatomic criteria (11,22), RECIST assessed tumor response on the basis of unidimensional measurements made on CT along the tumor’s longest axis, rendering the process more reproducible and applicable to the clinical setting. RECIST also defined the parameters that had been the source of disagreement between groups implementing the WHO criteria; the maximum number of lesions to be measured was set at ten, with a maximum of five per organ, and the minimum size of a lesion to be measured was set at one cm. Finally, RECIST redefined the response categories that were established in the
WHO criteria (Table 1). These reformulated classifications were conservative relative to the WHO criteria, placing fewer patients in the progressive disease category (21,23,24).

However, RECIST was not without shortcomings. It was widely reported to be less suitable for particular cancers, such as mesothelioma and pediatric tumors (23,25,26). Furthermore, the arbitrary number of tumor foci to be measured according to the criteria and the relatively narrow definition of progressive disease were points of contention (27). It was also suggested that the routine clinical implementation of RECIST would significantly increase the workload of radiologists (28).

To address these limitations, the RECIST Working Group set out to amend the criteria, publishing “RECIST 1.1” in 2008 (29). There were a handful of significant changes made to both simplify and clarify the criteria, as well as to allow for its application in additional cancers and modalities. First, the maximum number of measured tumors was reduced to five tumors with a maximum of two per organ. This amendment was based on data showing that such a reduction in the number of measured lesions did not result in a significant loss of information (30). Second, the definition of progressive disease now required a minimum absolute increase of five mm in the sum of the tumor diameters, thereby preventing changes in individual small lesions from leading to unnecessary classifications of progression. Third, specific guidelines were established for the assessment of lymph node involvement, defining nodes spanning $\geq 15$ mm on their short axis as assessable target lesions and nodes shrinking to $<10$ mm on their short axis as normal. Finally, the criteria paved the way for the incorporation of information from functional imaging modalities such as PET.

The following year, in 2009, Wahl et al published a paper outlining “PET Response Criteria In Solid Tumors” (PERCIST) (12). This criteria followed a number of earlier guidelines for response assessment that utilized PET, namely those proposed by the EORTC in 1999 (31), Hicks et al in 2001 (32), and Juweid et al in 2005 (33,34). PERCIST uses similar criteria to those developed for RECIST, but incorporation of the metabolic information to anatomic information sets it apart. The authors stated that CT alone possesses “poor predictive ability” because the residual masses that are detected by this modality often reflect scarring that is mistaken for active tumor. As a PET-based criteria for response assessment, PERCIST was “designed to facilitate trials of drug development, but, if sufficiently robust, could be applied to individual patients” (12).
In their report outlining PERCIST, Wahl et al specified a host of parameters that would facilitate the standardization of PET-based response assessment once the criteria was widely adopted. Among these suggestions was a proposed maximum of 5 tumor foci of the highest 18F-FDG avidity, with up to 2 foci per organ, to be measured for comparison before and after therapy. It was also recommended that patients undergo 18F-FDG-PET scans at least 10 days after an early cycle of chemotherapy in order to maximize the prognostic value of the scan and minimize the effect of 18F-FDG-avid inflammation caused by chemotherapy and radiation. Moreover, the authors called for the standardized uptake values (SUVs) derived from a PET scan to be corrected for lean body mass (SUL) and compared to reference uptake in the liver or, if necessary, background blood pool (12). Finally, PERCIST retained the same four response classifications that were established in RECIST, but amended their respective specifications (Table 2). Although not yet fully validated, the PERCIST criteria are increasingly used in clinical trials for assessing therapy response in cancer (Bai et al, Theranostics. 2013 Oct 7;3(10):787-801). Such data will potentially help support its more widespread clinical application.

MODERN RESPONSE ASSESSMENT CRITERIA IN LYMPHOMA

While the guidelines included in the WHO criteria, RECIST, and PERCIST are generalizable to a wide array of cancers, a number of specialized criteria have also been proposed specifically for the spectrum of hematological malignancies. As early as the late-1980s, as guidelines began to be developed for response assessment in chronic lymphocytic leukemia (CLL) (35), Hodgkin lymphoma (HL) (36), and acute myelogenous leukemia (AML) (37), there were calls for similar efforts toward standardization in NHL (38). However, in the decade that followed, various organizations simply adapted existing criteria to create their own guidelines for response assessment in NHL, thereby hindering the ability to compare data across different groups.

At meetings sponsored by the NCI in February and May 1998, an international working group, which comprised both American and European experts, reached a consensus on response assessment criteria specifically for NHL by (39). The resulting International Working Criteria (IWC) defined anatomic parameters, obtained by clinical or radiological exam, that could be used to group patients into the traditional classifications of complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD), as well as a new classification of
“unconfirmed complete response” (CRu) (Table 3). To support these anatomically-based criteria, the IWC defined the upper limit of a normal lymph node size as one cm along its short axis on the basis of several prior studies (40-42). In the years following its publication, the IWC were also adopted for HL (43).

In 2005, Juweid et al integrated the originally CT-based IWC with 18F-FDG-PET to create the IWC+PET criteria, which was initially designed and validated for NHL (33) but was subsequently validated for HL as well (44). Citing the prevalence of post-therapy residual masses and the unique ability of PET to accurately predict tumor viability in these masses, the investigators sought to establish a standardized approach that would join the anatomical information of CT with the functional information of PET. The IWC+PET criteria retained the classifications of the original IWC criteria but amended the guidelines to incorporate PET findings (Supplemental Table 4). Juweid et al found that IWC+PET was a better predictor of progression-free survival than IWC in NHL.

Two publications authored by Cheson et al and Juweid et al in 2007 amended the existing IWC+PET criteria and made recommendations for its clinical utilization in both HL and NHL as the International Harmonization Project (IHP) (34,45). To avoid false positive results on PET as a result of therapy-induced inflammation, which can persist for as long as 2 weeks following chemotherapy and 3 months following radiation therapy, both reports recommended that PET acquisition occur at least 3 weeks, and preferably 6 to 8 weeks, after chemotherapy and 8 to 12 weeks after radiation therapy. Cheson et al also addressed the possibility of false-positive PET findings due to “rebound thymic hyperplasia, infection, inflammation, sarcoidosis, and brown fat,” as well as “[spatial] resolution…technique, and variability of 18F-FDG avidity among histologic subtypes” (45). For evaluating the tumor viability of residual masses larger than 2 cm in their greatest transverse diameters, mediastinal blood pool activity was recommended as a reference. On the other hand, for residual masses smaller than 2 cm, background activity was the recommended reference. Residual hepatic and splenic lesions larger than 1.5 cm detected on CT were deemed positive if their metabolic activity was higher than that of the liver and spleen. These amendments permitted the elimination of the CRu category of tumor response, returning the classification scheme to the classic tetrad of CR, PR, SD, and PD (Table 4).

An international workshop that first met in Deauville, France in 2009 conceived of novel criteria for both HL and NHL that signaled a significant change on multiple fronts (46-49). In
contrast to the predominantly quantitative guidelines proposed previously, the Deauville five-point scoring system (D5PS) assessed treatment response qualitatively, specifically in the form of a 5-point scale that graded the intensity of $^{18}$F-FDG uptake relative to the reference activity of the mediastinal blood pool and liver (Table 5) ($^{50}$). The technical simplicity of this classification system facilitated its widespread clinical adoption. Moreover, the D5PS became a standard-bearer for the rising trend of interim response assessment, which enabled improved determinations of prognosis and earlier treatment modifications during the course of therapy.

The D5PS have since been modified by a comprehensive set of recommendations developed at the 11th International Conference on Malignant Lymphomas (ICML) in 2011 and presented at the 4th International Workshop on PET in Lymphoma, held in Menton, France in 2012, and at the 12th ICML, convened in Lugano, Switzerland in 2013 ($^{51,52}$). The consensus revision of both the criteria for staging and the 2007 IWG response criteria led to the development of Lugano classification. In the Lugano classification, separate sets of response criteria were proposed for PET and CT imaging, although the former is generally preferred for $^{18}$F-FDG-avid lymphomas. The PET-based criteria built on the 5-point categorical scale established by the D5PS by adding considerations for new or recurring involvement of lymph nodes and bone marrow as well as organomegaly (Table 6) ($^{53}$). Stand-alone CT-based guidelines were also included, despite the known limitations of anatomic response assessment in $^{18}$F-FDG-avid lymphoma, for use in instances where PET/CT imaging was unavailable or in lymphomas with low or variable $^{18}$F-FDG avidity.

**METHODOLOGICAL COMPARISON OF EXISTING RESPONSE ASSESSMENT CRITERIA**

The various therapy response criteria discussed in this review employ varying approaches in the use of imaging modalities such that the RECIST and 1999 IWC criteria primarily utilize CT, while EORTC and PERCIST rely on PET, and the 2007 IWC and Lugano classifications make use of both modalities with the former using IHP and the latter the D5PS criteria for PET interpretation. The assorted definitions of the response classifications across these criteria are shown in Table 7, which presents a simplified and standardized scheme comprising four groupings: CR, PR, SD, and PD. Although there are identifiable trends across criteria, even those employing the same modality demonstrate considerable variability in their thresholds for each
response classification. For example, progressive disease is defined as a \( \geq 20\% \) increase in tumor size by RECIST, a \( \geq 25\% \) increase by the WHO criteria, and a \( \geq 50\% \) increase by IWC.

In recent years, the relative simplicity of the D5PS and the associated Lugano classification have distinguished them from their quantitative predecessors, whose technical demands and complexity often precluded their widespread clinical use. However, questions of the reproducibility of the simplified qualitative criteria remain. The literature includes a number of comparisons between the D5PS and other guidelines using functional imaging for response assessment. A 2010 study of DLBCL patients by Horning et al compared interobserver agreement in the D5PS and the IHP-based Eastern Cooperative Oncology Group criteria, reporting \( \kappa \) values of 0.502 and 0.445, respectively (54). However, this study was limited to a small study population. Another DLBCL study by Itti et al found lower interobserver agreement with the D5PS (\( \kappa = 0.66 \)) as compared to a semiquantitative counterpart based on SUVmax (\( \kappa = 0.83 \)) (55). In larger standardized studies of HL that utilized the D5PS, Barrington et al, Furth et al, and Gallamini et al reported \( \kappa \) values of 0.79-0.85, 0.748, and 0.69-0.84, respectively, proving its superiority (56-58). The implications of these findings on the reproducibility and clinical applicability of the Lugano classification have yet to be determined in prospective studies with large data sets.

**FUTURE TRENDS**

The recent advent and adoption of the D5PS and Lugano classification have marked a step towards standardization of interpretation and brought a relatively more objective system. However, this qualitative response assessment system should also be tested against quantitative criteria to determine their relative effectiveness in patient management. In an earlier study, the tradeoff between simplicity and reproducibility was studied by Lin et al, who compared the prognostic ability of qualitative and quantitative PET analysis in patients with diffuse large B-cell lymphoma (DLBCL) (59). Visual analysis was able to predict event-free survival with an accuracy of 65.2\%, whereas SUV-based analysis did so with an accuracy of 76.1\%. A reduction of 65.7\% in the maximal SUV of an interim PET scan was found to be the optimal cutoff value in differentiating between favorable and unfavorable responses to therapy. These earlier results suggest that, if optimized for clinical use, standardized PET criteria employing a quantitative method may be more adept at assessing tumor response. However, various study biases and
suboptimal technical methodology inherent in retrospective design of prior studies, make it difficult to arrive at a firm conclusion. Moreover, multiple factors, including differences in instrumentation, scanner calibration, variable human biology, complicates obtaining reliable PET measurements across medical centers in future studies [Bai et al, Theranostics. 2013 Oct 7;3(10):787-801]. Thus, the most reproducible and accurate method for PET quantification remains to be determined. As techniques for automated segmentation and quantification continue to improve, these advancements will likely be more readily implemented in the clinical setting and facilitate the use of quantitative response assessment.

CONCLUSION

Over the past six decades, the techniques employed for evaluating the efficacy of cancer therapies have steadily increased both the precision and intricacy, moving from crude manual measurement towards more complex structural and functional data acquisition, with many more advanced techniques such as heterogeneity measures, and advanced parametric mapping and kinetic acquisitions/modeling on the way. The integration of CT and PET in particular greatly enhanced the ability to assess disease progression, adjust therapeutic regimens, and form an accurate prognosis. However, the vast array of interpretative guidelines that were introduced, each with their own protocols and thresholds, created stifling methodological variability and sparked calls for “harmonization” that have rung out since the early days of response assessment and continue to echo to this day (60).

With respect to lymphoma in particular, recent PET/CT-based criteria have made significant strides toward standardization, and their simplified qualitative guidelines have remedied the technical complexity and time intensity that impeded the clinical application of prior quantitative criteria. However, their suitability for certain scenarios, such as in patients with lymphomas of low or variable 18F-FDG avidity or in those receiving immunochemotherapy or biologic therapy remains to be determined. Moreover, as technological advances ease the use of quantitative criteria, continued efforts in maintaining harmonization in response assessment will likely be necessary to avoid renewed fragmentation.

ACKNOWLEDGEMENT
This work is sponsored by the PET Center of Excellence of the Society of Nuclear Medicine and Molecular Imaging.
REFERENCES


Table 1: Tumor response classifications of RECIST (2000)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response (CR)</td>
<td>Disappearance of all tumors for at least four weeks</td>
</tr>
<tr>
<td>Partial response (PR)</td>
<td>Shrinkage of tumor by at least 30% for over four weeks</td>
</tr>
<tr>
<td>Stable disease (SD)</td>
<td>Less than 30% reduction and 20% increase in tumor size</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>Increase in tumor size by at least 20%</td>
</tr>
</tbody>
</table>
Table 2: Tumor response classifications of PERCIST (2009)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response (CR)</td>
<td>Decline of metabolic activity of measurable lesions to below mean liver activity and on par with background blood pool activity</td>
</tr>
<tr>
<td>Partial response (PR)</td>
<td>Reduction of peak SUL values of at least 30% and 0.8 SUL units in measurable lesions</td>
</tr>
<tr>
<td>Stable disease (SD)</td>
<td>Less than a 30% reduction and 30% increase in peak SUL values in measurable lesions</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>Increase in peak SUL values by at least 30% and 0.8 SUL units in measurable lesions</td>
</tr>
</tbody>
</table>
Table 3: Tumor response classifications of International Workshop Criteria (1999)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response (CR)</td>
<td>Total disappearance of all signs and symptoms of disease on clinical and radiographic evaluations; Regression of nodal masses to a normal size in their greatest transverse diameter ($\leq 1.5$ cm in nodes $&gt;1.5$ cm pre-therapy, $\leq 1$ cm in nodes between 1.1 and 1.5 cm pre-therapy); Regression of spleen to a normal size so as not to be palpable on physical examination; Clear infiltrate on bone marrow aspirate and biopsy in sites that were previously involved</td>
</tr>
<tr>
<td>Unconfirmed complete response (CRu)</td>
<td>Fulfill the requirements of a complete response, except: There are residual lymph nodes $&gt;1.5$ cm in size that have regressed by over 75% in the sum of the product of their greatest diameters (SPD); Bone marrow aspirate or infiltrate are indeterminate</td>
</tr>
<tr>
<td>Partial response (PR)</td>
<td>Regression of the six largest lymph nodes masses by at least 50% in SPD; No increase in the size of other nodes, liver, or spleen; Regression of splenic and hepatic nodules by at least 50% in SPD; No new sites of disease</td>
</tr>
<tr>
<td>Stable disease (SD)</td>
<td>Does not satisfy criteria for partial response or progressive disease</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>Increase of at least 50% in SPD of lymph node mass; Appearance of any new lesions</td>
</tr>
</tbody>
</table>
Table 4: Tumor response classifications of the International Harmonization Project (2007)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response (CR)</td>
<td>PET-negative nodes of any size or PET-positive nodes that have regressed to normal size</td>
</tr>
<tr>
<td>Partial response (PR)</td>
<td>Reduction in SPD of 6 largest masses by at least 50% (and no growth in other nodes), PET-positive at previously uninvolved site or PET-negative</td>
</tr>
<tr>
<td>Stable disease (SD)</td>
<td>PET-positive at sites of previous involvement but not at new sites, no change in size of nodes on CT</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>Appearance of new lesions larger than 1.5 or 50% growth in SPD of existing nodes, PET-positive nodes</td>
</tr>
</tbody>
</table>
Table 5: Tumor response classifications of the Deauville five-point scoring system (D5PS) (2009)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No uptake above background activity</td>
</tr>
<tr>
<td>2</td>
<td>Uptake equal to or lower than mediastinal blood pool activity</td>
</tr>
<tr>
<td>3</td>
<td>Uptake between mediastinal blood pool and liver activity</td>
</tr>
<tr>
<td>4</td>
<td>Uptake moderately higher than liver activity</td>
</tr>
<tr>
<td>5</td>
<td>Uptake markedly higher than liver activity</td>
</tr>
<tr>
<td>Classification</td>
<td>PET/CT-Based Criteria</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>Complete response (CR)</td>
<td>1, 2, or 3 points on D5PS; No new lesions; No bone marrow involvement</td>
</tr>
<tr>
<td>Partial response (PR)</td>
<td>4 or 5 points on D5PS; Reduced uptake compared to baseline; No new lesions; Residual bone marrow uptake that is reduced from baseline</td>
</tr>
<tr>
<td>Stable disease (SD)</td>
<td>4 or 5 points on D5PS; Unchanged uptake compared to baseline; No new lesions; Unchanged bone marrow involvement</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>4 or 5 on D5PS; Increased uptake compared to baseline; New or recurrent involvement in nodes and bone marrow demonstrated by 18F-FDG avidity</td>
</tr>
</tbody>
</table>
### Table 7: Comparison of simplified classifications of various response criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete response</td>
</tr>
<tr>
<td>WHO</td>
<td>100% reduction in tumor size</td>
</tr>
<tr>
<td>EORTC</td>
<td>Reduction of 18F-FDG uptake to background levels</td>
</tr>
<tr>
<td>RECIST</td>
<td>100% reduction in tumor size</td>
</tr>
<tr>
<td>PERCIST</td>
<td>Reduction of 18F-FDG uptake to level of background blood pool</td>
</tr>
<tr>
<td>IWC</td>
<td>Reduction of nodes to normal size</td>
</tr>
<tr>
<td>IWC+PET</td>
<td>CR by IWC + negative PET scan</td>
</tr>
<tr>
<td>IHP</td>
<td>PET- nodes or PET+ nodes of normal size</td>
</tr>
<tr>
<td>Deauville*</td>
<td>18F-FDG uptake at background level</td>
</tr>
<tr>
<td>Lugano PET/CT</td>
<td>Normalized 18F-FDG uptake (1-3 on Deauville scale)</td>
</tr>
<tr>
<td>CT</td>
<td>Reduction of nodes/organs to normal size</td>
</tr>
</tbody>
</table>

* The D5PS was converted from a five-point scale to the four categories of CR, PR, SD, and PD (1=CR, 2=PR, 3=SD, 4/5=PD)