

Are your **MRI contrast agents** cost-effective?

Learn more about generic **Gadolinium-Based Contrast Agents**.



**FRESENIUS  
KABI**

caring for life

**AJNR**

**RESISTing the Need to Quantify: Putting  
Qualitative FDG-PET/CT Tumor Response  
Assessment Criteria into Daily Practice**

J.G. Peacock, C.T. Christensen and K.P. Banks

*AJNR Am J Neuroradiol* 2019, 40 (12) 1978-1986

doi: <https://doi.org/10.3174/ajnr.A6294>

<http://www.ajnr.org/content/40/12/1978>

This information is current as  
of April 19, 2024.

# RESISTING the Need to Quantify: Putting Qualitative FDG-PET/CT Tumor Response Assessment Criteria into Daily Practice

J.G. Peacock, C.T. Christensen, and K.P. Banks



## ABSTRACT

**SUMMARY:** Tumor response assessments are essential to evaluate cancer treatment efficacy and prognosticate survival in patients with cancer. Response criteria have evolved over multiple decades, including many imaging modalities and measurement schema. Advances in FDG-PET/CT have led to tumor response criteria that harness the power of metabolic imaging. Qualitative PET/CT assessment schema are easy to apply clinically, are reproducible, and yield good prognostic results. We present 3 such criteria, namely, the Lugano classification for lymphoma, the Hopkins criteria, and the Neck Imaging Reporting and Data Systems criteria for head and neck cancers. When comparing baseline PET/CTs with interim or end-of-treatment PET/CTs, radiologists can classify the tumor response as complete metabolic response, partial metabolic response, no metabolic response, or progressive disease, which has important implications in directing further cancer management and long-term patient prognosis. The purpose of this article is to review the progression of tumor response assessments from CT- and PET/CT-based quantitative and semi-quantitative systems to PET/CT-based qualitative systems; introduce the classification schema for these systems; and describe how to use these rapid, powerful, and qualitative PET/CT-based systems in daily practice through illustrative cases.

**ABBREVIATIONS:** CMR = complete metabolic response; D5P5 = Deauville 5-point scale; NI-RADS = Neck Imaging Reporting and Data Systems; PD = relapsed/progressive disease; PERCIST = PET Response Criteria in Solid Tumors; RECIST = Response Evaluation Criteria in Solid Tumors; WHO = World Health Organization; HNSCCa = Head and Neck squamous cell carcinoma; ACR = American college of radiology; AUC = appropriate use criteria; SCCa = squamous cell carcinoma; PPD = product of the diameters; SPD = sum of the product of the diameters

## INTRODUCTION

Cancer is second only to cardiac disease with regard to leading causes of mortality in the United States.<sup>1</sup> Fortunately, the rate of cancer deaths is declining, with the prevalence of individuals surviving and living with cancer increasing.<sup>1</sup> New surgical and medical treatments, including immunotherapies and targeted molecular therapies, which are more and more frequently tailored to an individual's specific tumor, are to credit for much of these improvements in patient survival. Essential to any effective cancer treatment in this dawning era of personalized medicine is an

understanding of how each unique neoplasm responds to its customized therapy. As such, radiology plays an increasingly vital role in helping clinicians determine treatment success or failure; guiding decisions with regard to whose therapy may be de-escalated, which therapy reduces toxicity but preserves efficacy, and how to identify treatment failures in patients who would benefit from timely modification of their regimen to improve the likelihood of a positive outcome or determination of a course that leads to palliation and hospice care. In this article, we describe the development of tumor response assessments in radiology and review 3 easy-to-use, qualitative, FDG-PET/CT tumor response assessments used in head and neck tumors, including lymphoma.

### *Evolution of Solid Tumor Response Evaluation*

In 1976, the World Health Organization (WHO) introduced the concept of imaging-guided cancer therapy response assessment by using CT and quantitative tumor measurements.<sup>2</sup> This technique measures the longest axial dimension of a tumor and its perpendicular dimension, then calculates the product. The sum of the product of the diameters of multiple lesions were compared with previous results to quantify the overall response. The WHO system also included parameters for then classifying responses as complete remission/response,

Received February 9, 2019; accepted after revision September 11.

From the Department of Radiology (J.G.P., K.P.B.), Brooke Army Medical Center, San Antonio, Texas; Department of Radiology (C.T.C.), Wilford Hall Ambulatory Surgical Center, San Antonio, Texas; and Department of Radiology (K.P.B.), Uniformed Services University of the Health Sciences, Bethesda, Maryland

The views expressed herein are those of the authors and do not reflect the official policy or position of Brooke Army Medical Center, the U.S. Army Medical Department, the Department of the Air Force and Department of Defense or the U.S. Government.

Please address correspondence to Justin G. Peacock, MD, PhD, Brooke Army Medical Center, 3551 Roger Brooke Dr, San Antonio, TX 78234; e-mail: justin.g.peacock@gmail.com; @PeacockImaging

Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

<http://dx.doi.org/10.3174/ajnr.A6294>

partial remission/response, no response, or relapsed/progressive disease (PD).<sup>2</sup> Although a step in the right direction, the WHO assessment had a degree of ambiguity because it did not specify the number or minimum size of lesions to be measured.<sup>2,3</sup> These limitations, among other factors, restricted its utility and curbed widespread use.

To better standardize the treatment response criteria, a new quantitative CT assessment tool called Response Evaluation Criteria in Solid Tumors (RECIST 1.0) was developed.<sup>2</sup> In RECIST 1.0, all lesions that were at least 1.0 cm in size were measured, with the maximum number capped at 10 lesions and no more than 5 lesions per organ. Also, lesion size was now quantified solely according to the longest dimension. RECIST underwent a slight modification, RECIST 1.1, to further simplify the technique by reducing the maximum number of lesions that needed to be characterized to 5 and no more than 2 per organ.<sup>3,4</sup> Although superior to previous assessment classifications in terms of ease of use and standardization, studies found a wide variation in sizes measured by different physicians.<sup>5,6</sup> Also, the criteria only looked at anatomic measurements, not at the metabolism of lesions. Therefore, it is unclear if a residual mass represents posttreatment inflammatory changes, tissue fibrosis, or viable tumor.<sup>6</sup>

FDG-PET/CT permits assessment not only of the anatomy but also the metabolism of lesions. This can provide valuable information with regard to the true efficacy of treatments, often much earlier than can be discerned by CT changes alone.<sup>7-10</sup> The first tumor response assessment to use PET/CT was the PET Response Criteria in Solid Tumors (PERCIST) tool.<sup>11</sup> PERCIST 1.0 incorporates FDG avidity in classifying tumor responses. PERCIST 1.0 involves calculating the standardized uptake value relative to the standardized uptake value of lean body mass for the liver and tumor target lesions. If the PET avidity is too low or not measurable, the classification defers to RECIST 1.1. The standardized uptake value relative to the lean body mass was chosen to reduce variability of metabolic measurements related to fluctuations in body weight and habitus. In PERCIST 1.0, disease response is classified as complete metabolic response (CMR), partial metabolic response, stable metabolic disease or no metabolic response, and progressive metabolic disease.

Clinical research demonstrated that PERCIST 1.0 is more sensitive and accurate than RECIST 1 for nonsmall cell lung cancer, malignant solid tumors, and colorectal cancer.<sup>6</sup> As with measurements, different patient factors, scanners, protocols, and PET software algorithms can all contribute to this variability.<sup>6</sup> Consequently, there was a need to develop an assessment classification that was easy to use, consistent, and prognostically valid.

### **Unique Lymphoma Response Evaluation**

In 1999, lymphoma treatment was classified according to the International Workshop Group.<sup>12,13</sup> The initial criteria were based on physical examination, CT, and gallium-SPECT findings. Quantification of size depended on the sum of the product of the diameters, with tumor response according to the WHO classes. An additional complete remission unconfirmed classification was applied when there was a significant decrease in lesion size but a residual mass. PET/CT improved the interpretation of these

complete remission unconfirmed lesions, so the International Workshop Group criteria were updated to include PET data, which led to the International Harmonization Project.<sup>14</sup> The complete remission unconfirmed classification was eliminated, with residual masses characterized as complete remission/response or partial remission/response based on the FDG avidity. Challenges in this International Harmonization Project classification again included interpreter variability in sum of the product of the diameters measurements, along with variability in PET standardized uptake value measurements and the need for a bone marrow biopsy.

### **Qualitative Tumor Response Evaluation**

The previously mentioned challenges in response assessment and the increased use of FDG-PET/CT have led to the current tools that are easy to use and useful for both prognostics and guiding therapy. The criteria rely on qualitative metrics that can be rapidly performed during the clinical interpretation of PET/CT examinations, often by analysis of the PET MIP views alone. Lymphoma, head and neck tumors, esophageal cancer, lung cancer, pancreatic cancer, rectal cancer, prostate cancer, and cervical cancer have all been assessed by using qualitative FDG-PET/CT tumor response evaluation techniques.<sup>15-22</sup> Three major response assessment criteria in neuroradiology that use this simplified approach are the Lugano criteria for lymphoma and the Hopkins and Neck Imaging Reporting and Data Systems (NI-RADS) criteria for solid tumors of the head and neck.<sup>15,16,23-26</sup> The Lugano and Hopkins criteria use relatively stable internal reference standards for metabolism found on every examination, blood pool, and liver intensities, which minimizes the variation in response assessments related to differences in patients, examination protocols, scanner characteristics, and readers.<sup>15,25</sup> The NI-RADS criteria use a combination of contrast-enhanced CT, together with the PET-CT findings, to provide anatomic evaluation in areas of increased metabolic activity.<sup>26</sup> In this article, we describe the development of these 3 criteria and their use in daily practice, and provide illustrative examples of the utility of these tools in the modern evaluation of tumor response.

## **THE LUGANO CLASSIFICATION FOR LYMPHOMA TREATMENT RESPONSE EVALUATION**

### **Development of the Lugano Classification**

In 2011, leaders in the field of malignant lymphoma met in Lugano, Switzerland, to create more effective treatment response classification guidelines based on previous trials, clinical experience, and research groups.<sup>23</sup> The results were published in 2014 as the Lugano classification for lymphoma staging and response assessment.<sup>15,23,27</sup> The novel classification included quantitative CT parameters for non-FDG-avid lymphomas and independent PET/CT parameters for FDG-avid lymphomas.<sup>15,23,27</sup> Although FDG-PET/CT was first included in the International Harmonization Project clinical response criteria, the Lugano system is the first to define the role of FDG-PET/CT for response assessment in all FDG-avid lymphomas.<sup>15,23,27</sup> In addition, the criteria eliminated the cumbersome need for a bone marrow biopsy in FDG-avid lymphomas, which allows

**Table 1: Lugano classification for CT-based lymphoma response<sup>a</sup>**

Technique	Complete Response	Partial Response	No Response	PD
CT	Complete disappearance of disease or decrease in the lymph node long axis to <1.5 cm	Multiple: $\geq 50\%$ decrease in SPD of up to 6 lesions. Single: $\geq 50\%$ decrease in PPD	No criteria for PD and <50% decrease in SPD of up to 6 lesions.	(1) New or increased lymphadenopathy (long axis >1.5 cm): new nodes, $\geq 50\%$ increase in PPD, or >0.5 cm increase in diameter for $\leq 2$ cm or >1.0 cm for >2 cm; (2) splenic volume increase: with splenomegaly, >50% increase in length, or without splenomegaly, $\geq 2$ cm increase in length; (3) new or larger lesions; (4) recurrent previously resolved lesions; (5) new extranodal lesion >1 cm (or <1 cm if unequivocally lymphoma)
FDG-PET/CT	D5PS score of 1, 2, or 3 in disease sites, with or without residual mass	D5PS score of 4 or 5, with reduced uptake compared with baseline	D5PS score of 4 or 5, with no significant change in FDG uptake	D5PS score of 4 or 5, with increased uptake intensity compared with baseline and/or new FDG-avid foci

**Note:**—SPD indicates sum of the product of the diameters (PPD).

<sup>a</sup> For non-FDG avid lymphomas, the following CT-based measurement criteria are used to determine tumor response classification; adapted from Ref. 15.

**Table 2: D5PS<sup>a</sup>**

D5PS Score <sup>b</sup>	Lesion FDG Uptake
1	At or below background
2	Less than or equal to mediastinal blood pool
3	Greater than or equal to mediastinal, but less than hepatic blood pool
4	Greater than hepatic blood pool
5	Markedly greater than hepatic blood pool
X	Designator for nonlymphomatous lesion

<sup>a</sup> Adapted from Ref. 23.

<sup>b</sup> The D5PS score is used in the Lugano classification for grading lymphoma response.

response based solely on the single most metabolically active lesion.

**Using the Quantitative CT Lugano Parameters to Grade Lymphoma Response**

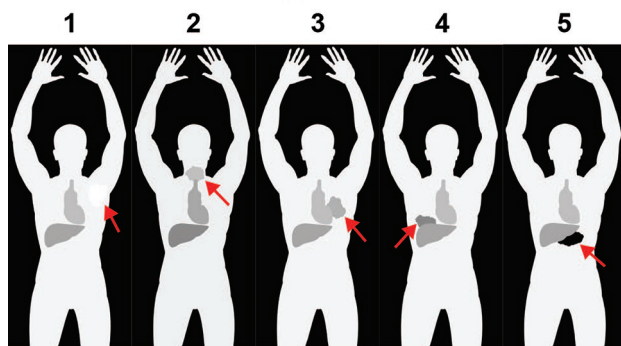
The Lugano criteria for CT lymphoma staging and response assessment are quantitative and reserved for non-FDG-avid lymphomas.<sup>23</sup> CT response assessment categories for interpretation of an interim

or end-of-treatment examination include complete remission/re-sponse, partial remission/response, no response, and PD.<sup>15,23</sup> CT criteria are based on measurement of lymph nodes, any extranodal lesions, and splenic sizes. Lymph nodes must have a long-axis measurement of >1.5 cm, whereas extranodal disease must be at least 1.0 cm. The bi-dimensional diameter product is calculated for single lesions and the sum of the product of the diameters for up to 6 nodal and/or extranodal target lesions. Complete remission/response occurs when all lymph node long-axis diameters are  $\leq 1.5$  cm and there is no residual extranodal disease. PD occurs with new lymphadenopathy or extranodal lesions, splenic size increase, or increased size of pre-existing lesions. The complete response criteria are detailed in Table 1. The CT parameters, principally derived from the previous WHO schema, have similar challenges in interpreter measurement reliability.

**Deauville 5-Point Scale for Scoring FDG Avidity**

The Lugano classification incorporates the Deauville 5-point scale (D5PS) for grading FDG avidity.<sup>23,28</sup> A score is assigned based on the single most intense focus of FDG-avid lymphomatous disease, relative to mediastinal blood pool and hepatic activity (Table 2). A D5PS score of 1 indicates that the lesion does not demonstrate FDG uptake greater than background activity. A score of 2

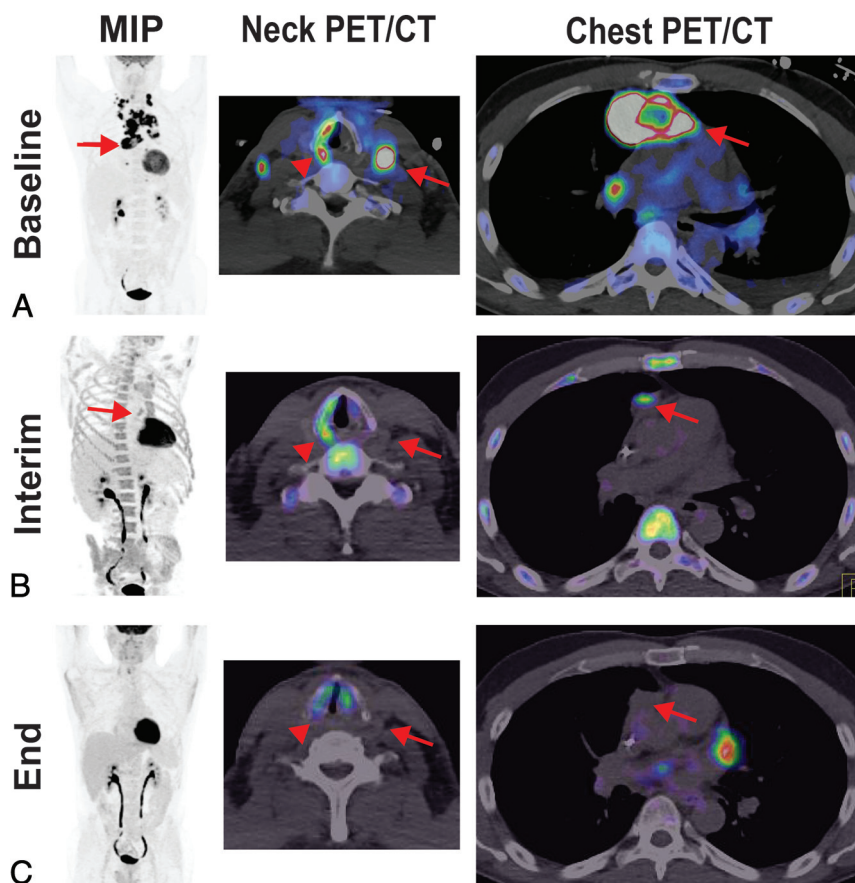
**Deauville Scores**



**FIG 1.** The D5PS scores of lesions qualitatively based on the FDG uptake relative to the mediastinal blood pool (MBP) and hepatic parenchymal FDG activity. The figure demonstrates hypothetical masses (arrows) and their FDG uptake relative to the MBP and liver activity. D5PS score of 1 for a left axillary mass with FDG uptake no greater than background activity. D5PS score of 2 for cervical mass with FDG uptake above background but less than MBP or liver. D5PS score of 3 for hilar mass with FDG uptake greater than MBP but less than or equal to the hepatic activity. A D5PS score of 4 for a mass in right lung base with FDG uptake greater than both MBP and liver. A D5PS score of 5 for a midabdominal mass with FDG uptake markedly greater than that of the liver.

for PET to serve as a surrogate for determining marrow involvement.<sup>23</sup> The FDG-PET/CT criteria also simplify interpretation by providing a qualitative assessment of the lymphoma treatment





**FIG 2.** A partial metabolic response and CMR based on Lugano criteria. A 47-year-old man with diffuse large B-cell lymphoma demonstrates (A) baseline intense FDG avidity in mediastinal (MIP and PET/CT of the chest [arrows denote a large pericardial lymph node conglomerate]) and bilateral levels III and IV lymphadenopathy (PET/CT of the neck [arrow denotes a large level III node]), a D5PS score of 5. Incidentally, on the PET/CT of the neck, the arrowhead denotes a dysfunctional right vocal cord, likely due to disease impacting the ipsilateral recurrent laryngeal nerve. B, Interim imaging demonstrates reduced but persistent FDG uptake (2 times greater than the liver) in the retrosternal, pericardial mass (MIP and PET/CT of the chest [arrows]), consistent with a D5PS score of 5 but a Lugano designation of a partial metabolic response. There was interval resolution of the cervical lymphadenopathy. C, End-of-treatment imaging shows a residual pericardial mass on CT (arrow), without FDG uptake above background, consistent with a D5PS score of 1 and a Lugano designation of CMR.

indicates that the FDG uptake is less than or equal to mediastinal blood pool, whereas a score of 3 indicates that the lesion's FDG uptake is greater than the mediastinal blood pool and less than or equal to liver activity. D5PS scores of 4 and 5 indicate that the lesion's FDG uptake is moderately and markedly greater than the liver activity, respectively, with markedly greater considered to be at least 2–3 times more intense (Fig 1). A D5PS "X" designation may be used in conjunction with the 5-point scale to describe an FDG-avid nonlymphomatous lesion, such as sarcoid related hilar lymph nodes or focal thyroid uptake attributed to a primary thyroid neoplasm. By using these qualitative D5PS scores, rapid and reproducible assessment of posttreatment lymphoma response can be performed.

#### Using the Lugano Classification in Daily Practice

According to the 2019 International Workshop on Interim-PET scan in lymphoma, the D5PS score of the interim or end-of-

treatment examination should be compared with the score assigned to the most recent comparison,<sup>13,29</sup> which then lead to CMR, partial metabolic response, no metabolic response, or PD designations. Scores 1 and 2 on the interim or end-of-treatment examination denote a CMR (Fig 2). A D5PS score of 3 on follow-up imaging also likely signifies a CMR but may be interpreted as an inadequate response to avoid undertreating patients being considered for de-escalation of therapy.<sup>23</sup> In a separate article, by Mikhael et al,<sup>30</sup> these patients were found to have an intermediate overall survival and progression-free survival compared with patients with CMR and no metabolic response or PD. A D5PS score of 4 or 5 can indicate a partial metabolic response, no metabolic response, or PD designation, depending on whether the interim or end D5PS score is decreased, unchanged, or increased, respectively (Figs 2 and 3). In addition, any new FDG-avid lesions on an examination are classified as PD.<sup>15,23,27</sup>

The benefit of the D5PS and Lugano classification is that it provides well-defined guidelines for rapid qualitative tumor response assessment. The scores are based on internal standardized uptake value references, relatively similar from patient to patient and from examination to examination, which

means that individual variability in patients and PET/CT technology do not impact the assessment (Fig 1 and Table 1). Razek et al<sup>31</sup> reported excellent interobserver agreement (95.8% agreement,  $\kappa = 0.91$ ) in assigning D5PS scores and Lugano posttreatment responses. Burggraaff et al<sup>32</sup> reported treatment response assessments in diffuse large B-cell lymphoma by using dichotomization of D5PS 1–3 as negative and D5PS 4–5 as positive. The interim and end-of-treatment positive, negative, and overall agreements were 73.7%, 92.0%, and 87.7%, and 76.3%, 95.0%, and 91.7%, respectively.<sup>32</sup> By contrast, Kluge et al<sup>33</sup> (42% agreement,  $\kappa = 0.24$ ), Sawan et al<sup>34</sup> ( $\kappa = 0.082$ ), and Ceriani et al<sup>35</sup> ( $\kappa = 0.35$ – $0.72$ ) all reported poorer interobserver agreement for assigning individual D5PS to random scores. Kluge et al<sup>33</sup> reported that interobserver agreement improved when interpreters assigned a simple binary positive or negative result, such as reporting D5PS scores of 1, 2, or 3 as negative for recurrence and 4 or 5 as positive for disease or recurrence (86% agreement,

$\kappa = 0.56$ ). Sawan et al,<sup>34</sup> reported improved interobserver agreement as well as accuracy when second-opinion reports were obtained by oncologic radiologists who had training and experience in evaluating these studies ( $\kappa = 0.86$  and radiology-pathology concordance = 78%). Ceriani et al,<sup>35</sup> reported improved interobserver agreement after training ( $\kappa = 0.77$ – $0.87$ ). These studies demonstrate that simple binary scoring systems and training in the D5PS and Lugano criteria can improve the precision and accuracy in posttreatment response assessment.

Multiple studies demonstrate the significant prognostic value of the Lugano classification and D5PS score in interim and end-of-treatment PET/CTs for a variety of FDG-avid lymphomas, including Burkitt, Hodgkin, non-Hodgkin, mantle cell, follicular, natural killer, and T-cell lymphomas in pediatric and adult

patients.<sup>8,13,27,29,36–50</sup> These studies indicate that D5PS outperforms other indicators in overall survival and progression-free survival, including simple changes in standardized uptake values or CT-only measurements.<sup>23,29,46</sup> Specific test metrics depend on the lymphoma subtype and grade; Hodgkin lymphoma demonstrates high (90%–100%) positive predictive values and negative predictive values, and non-Hodgkin lymphoma demonstrates high negative predictive values (80%–100%), but lower positive predictive values (50%–100%).<sup>23,29</sup> Metabolically active disease in these non-Hodgkin lymphoma cases warrants further evaluation with imaging or biopsy because the activity may simply represent posttreatment inflammation.

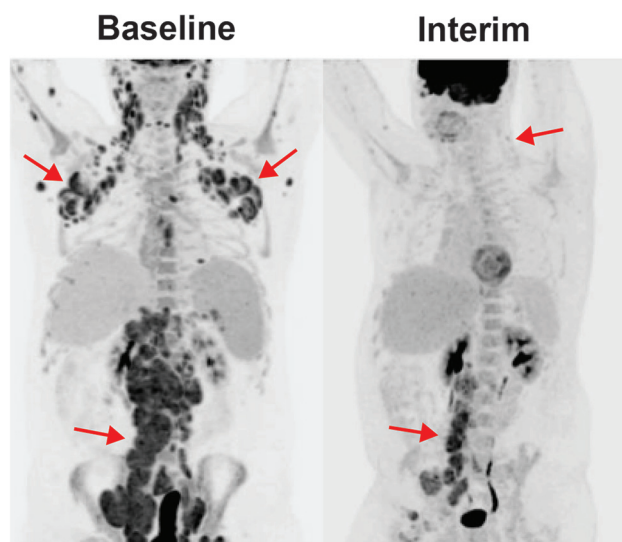
## HOPKINS CRITERIA FOR HEAD AND NECK TUMOR RESPONSE EVALUATION

### Development of the Hopkins Criteria for Head and Neck Tumors

Initially, the Hopkins criteria were developed to qualitatively assess head and neck tumor treatment response.<sup>16</sup> The criteria harnessed the well-documented power of FDG-PET/CT in head and neck tumors, which supersedes anatomic size changes in prognostic value.<sup>6,16</sup> Similar to the D5PS, lesions are classified into 5 scores relative to FDG avidity in the liver and the internal jugular vein blood pool, used in place of the mediastinum (Table 3). Lesions that demonstrate focal avidity less than the internal jugular vein have a score of 1, whereas those with focal avidity greater than internal jugular vein have a score of 2. In contrast to the D5PS, the Hopkins criteria address the presence of diffuse FDG avidity greater than internal jugular vein or liver, which is frequently seen after effective treatment and gives it a score of 3. Lesions with focal FDG avidity greater than the liver are given a score of 4, and focal and intense avidity greater than the liver are scored as 5.<sup>16</sup> These areas of FDG avidity are scored in the original tumor site, right neck, and left neck, and the overall score is the highest score. The scores correspond with the posttreatment response from CMR to posttreatment inflammation to residual tumor with overall scores of 1, 2, and 3 considered negative for residual disease, whereas scores of 4 or 5 are considered positive (Table 3).

Marcus et al<sup>16</sup> demonstrated a high interpreter reliability for scoring head and neck squamous cell carcinoma (HNSCCA) follow-up PET/CTs by using the Hopkins Criteria. They also demonstrated a robust specificity and negative predictive value of 92.2% and 91.1%, respectively, with an overall accuracy of 86.9%.

Importantly, the criteria demonstrated a significant clinical value by reversing management for approximately 64% of patients. Van den Wyngaert et al<sup>51</sup> in the European ECLYPS trial demonstrated similarly high specificity and negative predictive values, of 91.2% and 92.1%, respectively, by using the Hopkins criteria to grade head and neck squamous cell carcinoma tumor response. The sensitivity for residual



**FIG 3.** A posttreatment D5PS score of 5 but partial metabolic response based on Lugano criteria. Baseline imaging demonstrates diffuse lymphadenopathy, including cervical levels II–IV, axillary, mediastinal, abdominal periaortic/pericaval, and right greater than left iliac chains, with intense FDG avidity (arrows), consistent with a D5PS score of 5. Interim imaging during treatment demonstrates significantly reduced size and FDG avidity of the cervical, axillary, and mediastinal lymphadenopathy (arrows), now predominantly limited to the periaortic/pericaval regions and right greater than left iliac chains. The lymph nodes still have FDG avidity markedly greater than mediastinal blood pool and liver, consistent with a D5PS score of 5, but the reduction in size, number, and intensity results in a Lugano designation of a partial metabolic response.

**Table 3: Hopkins criteria scores for head and neck cancers<sup>a</sup>**

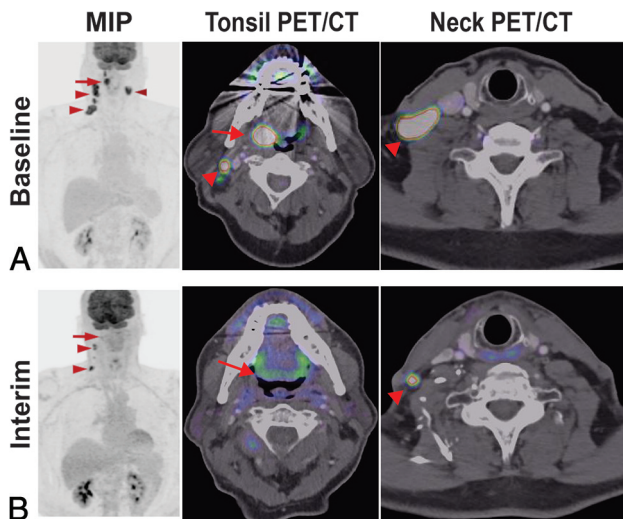
Hopkins Criteria Score <sup>b</sup>	Lesion FDG Uptake	Therapy Response
1	Focal uptake less than IJV	CMR
2	Focal uptake greater than IJV, but less than liver	Likely CMR
3	Diffuse uptake greater than IJV and liver	Likely posttreatment inflammation
4	Focal uptake greater than liver	Likely residual tumor
5	Focal uptake markedly greater than liver	Residual tumor

Note:—IJV indicates internal jugular vein.

<sup>a</sup> Adapted from Ref. 16.

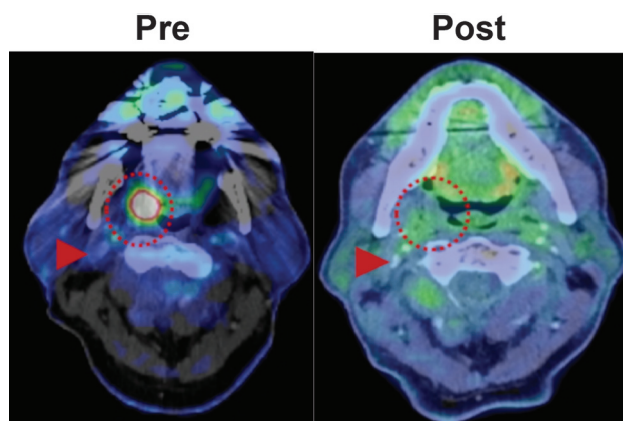
<sup>b</sup> Five-point scale Hopkins criteria scores used to classify head and neck cancer response after treatment; new lesions are considered to be PD.



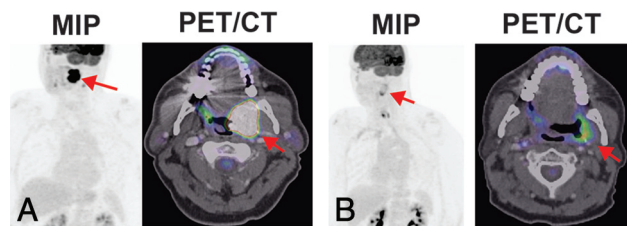


**FIG 4.** Hopkins criteria score of 5 in right tonsillar squamous cell carcinoma, consistent with residual tumor. A 71-year-old man with right tonsillar squamous cell carcinoma. MIP from baseline PET/CT demonstrates a Hopkins criteria score of 5 in the primary right tonsillar tumor (*arrow*) and right greater than left level II–IV cervical lymph nodes (*arrowheads*). Examination after radiation and chemotherapy shows resolution of the right tonsillar FDG uptake (*arrow*) and some of the cervical lymph nodes but persistent, focal, and intense FDG uptake in 2 ipsilateral level III and IV lymph nodes (*arrowheads*); a Hopkins criteria score of 5 is consistent with residual tumor.

disease was time dependent, and follow-up surveillance imaging was recommended at 1 year after therapy; however, guidelines with regard to posttreatment surveillance imaging are not well established, with some experienced Radiologists suggesting shorter intervals of as little as 12 weeks. Wray et al<sup>25</sup> demonstrated that the Hopkins criteria and FDG-PET/CT



**FIG 5.** Hopkins criteria score of 2 and NI-RADS score of 1 in right tonsillar squamous cell carcinoma, consistent with a CMR. A 71-year-old man with an intensely FDG-avid right tonsillar squamous cell carcinoma (*dotted circle*) on pretreatment PET/CT. After treatment, the mass has resolved, with FDG uptake in the region just above the adjacent right internal jugular vein (*arrowhead*) but similar to the surrounding oropharyngeal soft tissues, consistent with a Hopkins criteria score of 2 and an NI-RADS score of 1 at the primary site. By using Hopkins or NI-RADs criteria, the findings are consistent with an overall CMR.



**FIG 6.** An NI-RADS score of 2a in left tonsillar squamous cell carcinoma (SCCa) after chemoradiation, consistent with a Hopkins criteria score of 3 and posttreatment inflammation. A 63-year-old man with a history of left tonsillar SCCa. A, Baseline oblique MIP shows markedly hypermetabolic primary left tonsillar SCCa (*arrow*). After completion of chemoradiation therapy, repeated FDG PET/CT was obtained. B, Oblique MIP and axial PET/CT show residual hypermetabolic mucosal activity of moderate intensity (*arrow*) throughout the tonsil bed. The findings are compatible with NI-RADS 2a. Linked management recommendations in the NI-RADS criteria suggest correlation with direct visualization given that such a finding typically represents non-neoplastic FDG uptake. As a comparison, the FDG uptake is greater than the liver and much greater than the internal jugular vein, this is consistent with a Hopkins criteria score of 3. The findings suggest benign posttreatment inflammation. Lesions with this score can be false-negatives and demonstrate intermediate overall survival and progression-free survival compared with scores of 1–2 and 4–5. Due to the false-negative probability, these lesions should be biopsied before altering the treatment plan.

assessment after head and neck squamous cell carcinoma chemoradiotherapy was significantly better than residual neck node size in predicting overall survival and progression-free survival. In an external validation study, Kendi et al<sup>24</sup> investigated the use of the Hopkins criteria in patients with head and neck squamous cell carcinoma after radiation therapy. They found similar degrees of interpreter reliability and overall test statistics, including a specificity of 87.3% and a negative predictive value of 96.5%. These studies demonstrate the widespread applicability of the Hopkins criteria in assessing posttreatment head and neck squamous cell carcinoma disease response.

#### Using the Hopkins Criteria for Evaluating Head and Neck Tumor Treatment Response

The Hopkins criteria in the studies described above were generally performed approximately 5 to 24 weeks after chemoradiotherapy or surgical treatment for head and neck squamous cell carcinoma. The timing and type of treatment seems to have a significant impact on the positive predictive value and sensitivity of the Hopkins criteria. Wray et al<sup>25</sup> indicate that posttreatment inflammatory changes or false-positives can mostly be avoided if the PET/CT is performed  $\geq 12$  weeks after the completion of radiation therapy. Taghipour et al<sup>52</sup> indicate that postsurgical patients may not have as much posttreatment inflammation as postradiation patients, so scans could potentially be performed earlier.

The beauty of qualitative therapy response assessment is again found in the ease and rapid assessment that can be performed, even on MIP images. The tumor response assessments, including CMR, partial metabolic response, no metabolic response, and

**Table 4: NI-RADS risk stratification guidelines<sup>a</sup>**

Description	Category	Imaging Findings	Management
Incomplete	0	New baseline study or unavailable previous imaging	Previous imaging
No evidence of recurrence	1	Posttreatment changes without mass or abnormal FDG uptake	Routine surveillance
Low suspicion (superficial)	2a	Enhancement without FDG uptake or mild/moderate FDG uptake without a mass	Direct visual inspection
Low suspicion (deep)	2b	Enhancement without FDG uptake or mild/moderate FDG uptake without a mass	Short imaging follow-up
High suspicion	3	New or enlarged mass or lymph node, enhancement, and intense FDG uptake	Biopsy
Definitive recurrence	4	Pathologically proved or definite progression	Clinical management

<sup>a</sup>The NI-RADS risk stratification guidelines grade head and neck tumor treatment response; adapted from Ref. 26.

PD, are based on FDG uptake before and after treatment (Figs 4 and 5). The challenge in the Hopkins criteria can often be scores of 3, which demonstrate diffuse areas of FDG avidity consistent with posttreatment inflammatory changes (Fig 6). Without the application of the Hopkins criteria, these areas can often lead to false-positive interpretation of persistent metabolically active tumor. However, even with appropriate use of the Hopkins criteria, false-negatives may occur, and thus Hopkins criteria scores of 3 demonstrate intermediate overall survival and progression-free survival compared with scores of 1–2 and 4–5. Consequently, Hopkins criteria scores of 3 may necessitate biopsy to confirm their posttreatment status.

## NI-RADS GUIDELINES FOR HEAD AND NECK TUMORS

### Development of the NI-RADS Criteria

In 2016, the American College of Radiology (ACR) convened the NI-RADS committee to formulate a reporting and management system with risk stratification for head and neck tumors.<sup>53,54</sup> The guidelines, modeled after BI-RADS, define 6 categories of posttreatment findings for head and neck tumors, which range from an incomplete study to definite disease recurrence, and recommend appropriate steps for follow-up.<sup>26</sup> The imaging findings are based on combined CT of the neck with contrast and FDG-PET/CT findings (Table 4).<sup>26</sup> The findings of the CT of the neck characterize disease recurrence by soft-tissue masses or enhancement, whereas the PET/CT findings characterize the recurrence by PET avidity of lesions relative to background (Table 4). The benefit of NI-RADS is that it combines the morphologic findings from the CT of the neck, including CT appearance, size, and enhancement, and the metabolic PET/CT findings (Figs 5 and 6).<sup>26</sup>

In 2017, Krieger et al<sup>54</sup> analyzed the accuracy of NI-RADS in follow-up scans for varying head and neck cancers. In analyzing 618 head and neck lesions, 85.4% were scored NI-RADS 1 and demonstrated a 3.8% recurrence, 9.4% were scored NI-RADS 2 and demonstrated a 17.2% recurrence, and 5.2% were scored NI-RADS 3 and demonstrated a 59.4% recurrence.<sup>54</sup> The overall AUC accuracy curve for the NI-RADS criteria gave a value of 0.787 (a perfect test would give a value of 1.0). They also demonstrated that the combination of a CT of the neck and PET/CT functioned better than either technique alone. In 2018, Wangaryattawanich

et al<sup>55</sup> demonstrated the ability of the NI-RADS criteria to rule out disease recurrence in head and neck cancers after treatment. When using 2-year disease-free survival as the reference standard, the negative predictive value for patients with NI-RADS 2 scores on the first posttreatment PET/CT was 85%, compared with 91% for patients with NI-RADS 1 scores.

### Comparison of the Hopkins and NI-RADS Criteria for Head and Neck Cancers

Similar to the Hopkins criteria, the NI-RADS guidelines use qualitative mechanisms for characterizing disease recurrence. Unlike the Hopkins criteria, the NI-RADS guidelines use both CT and PET/CT findings. Although the Hopkins criteria reference FDG uptake to internal jugular vein and liver uptake, the NI-RADS reference uptake is not standardized, and it is up to interpreters to determine what constitutes moderate versus intense FDG avidity relative to background, a significant factor in determining a score of NI-RADS 2 (low suspicion) versus 3 (high suspicion) (Figs 5 and 6).<sup>26</sup> The NI-RADS guidelines do incorporate recurrence risk rates and definitive management guidelines.<sup>26</sup> In addition, in using the morphologic or anatomic features on CT, NI-RADS can improve overall accuracy in interpreting non-neoplastic patterns of FDG uptake.<sup>56</sup>

## CONCLUSIONS

With the widespread institution of FDG-PET/CT, qualitative, tumor response classification systems have been developed that allow for simple, yet accurate, early evaluations of posttreatment response. The Lugano classification uses PET/CT to characterize the posttreatment lymphoma response. The Hopkins criteria were developed to similarly characterize the posttreatment response in head and neck tumors. The NI-RADS criteria use contrast CT and PET/CT to characterize posttreatment response. These 3 treatment response assessments use qualitative means of comparison that can be performed readily in the busy clinical setting with standard PET/CT equipment and software. The response assessments, particularly in well-trained interpreters, can accurately predict overall survival and progression-free survival. In addition, the response assessments can improve communication with clinicians to drive management decisions. We recommend that radiologists and nuclear medicine physicians



consider using these classifications when reporting posttreatment response on follow-up PET/CTs.

## REFERENCES

1. US Cancer Statistics Working Group. US Cancer Statistics Data Visualizations Tool. 2018 <https://www.cdc.gov/cancer/uscs/dataviz/index.htm>. Accessed October 16, 2019
2. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92:205–16 [CrossRef Medline](#)
3. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47 [CrossRef Medline](#)
4. Chalian H, Töre HG, Horowitz JM, et al. Radiologic assessment of response to therapy: comparison of RECIST versions 1.1 and 1.0. *Radiographics* 2011;31:2093–105 [CrossRef Medline](#)
5. Patel CN, Goldstone AR, Chowdhury FU, et al. FDG PET/CT in oncology: “raising the bar”. *Clin Radiol* 2010;65:522–35 [CrossRef Medline](#)
6. Sheikhbahaei S, Mena E, Pattanayak P, et al. Molecular imaging and precision medicine: PET/computed tomography and therapy response assessment in oncology. *PET Clin* 2017;12:105–18 [CrossRef Medline](#)
7. Eugene T, Corradini N, Carlier T, et al. <sup>18</sup>F-FDG-PET/CT in initial staging and assessment of early response to chemotherapy of pediatric rhabdomyosarcomas. *Nucl Med Commun* 2012;33:1089–95 [CrossRef Medline](#)
8. MacManus MP, Seymour JF, Hicks RJ. Overview of early response assessment in lymphoma with FDG-PET. *Cancer Imaging* 2007; 7:10–8 [CrossRef Medline](#)
9. Passero VA, Branstetter BF, Shuai Y, et al. Response assessment by combined PET-CT scan versus CT scan alone using RECIST in patients with locally advanced head and neck cancer treated with chemoradiotherapy. *Ann Oncol* 2010;21:2278–83 [CrossRef Medline](#)
10. Tiseo M, Ippolito M, Scarlattei M, et al. Predictive and prognostic value of early response assessment using <sup>18</sup>FDG-PET in advanced non-small cell lung cancer patients treated with erlotinib. *Cancer Chemother Pharmacol* 2014;73:299–307 [CrossRef Medline](#)
11. Wahl RL, Jacene H, Kasamon Y, et al. From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. *J Nucl Med* 2009;50(Suppl 1):122S–50S [CrossRef Medline](#)
12. Younes A, Hilden P, Coiffier B, et al. International Working Group consensus response evaluation criteria in lymphoma (RECIL 2017). *Ann Oncol* 2017;28:1436–47 [CrossRef](#)
13. Meignan M, Gallamini A, Meignan M, et al. Report on the First International Workshop on interim-PET scan in lymphoma. *Leuk Lymphoma* 2009;50:1257–60 [CrossRef Medline](#)
14. Juweid ME, Stroobants S, Hoekstra OS, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol* 2007;25:571–78 [CrossRef Medline](#)
15. Johnson SA, Kumar A, Matasar MJ, et al. Imaging for staging and response assessment in lymphoma. *Radiology* 2015;276:323–38 [CrossRef Medline](#)
16. Marcus C, Ciarallo A, Tahari AK, et al. Head and neck PET/CT: therapy response interpretation criteria (Hopkins criteria)—inter-reader reliability, accuracy, and survival outcomes. *J Nucl Med* 2014;55:1411–16 [CrossRef Medline](#)
17. Valkema MJ, Noordman BJ, Wijnhoven BPL, et al. Accuracy of <sup>18</sup>F-FDG PET/CT in predicting residual disease after neoadjuvant chemoradiotherapy for esophageal cancer. *J Nucl Med* 2019;pii: jnumed.118.224196 [Epub ahead of print] [CrossRef Medline](#)
18. Sheikhbahaei S, Mena E, Marcus C, et al. <sup>18</sup>F-FDG PET/CT: therapy response assessment interpretation (Hopkins criteria) and survival outcomes in lung cancer patients. *J Nucl Med* 2016;57:855–60 [CrossRef Medline](#)
19. Sheikhbahaei S, Wray R, Young B, et al. <sup>18</sup>F-FDG-PET/CT therapy assessment of locally advanced pancreatic adenocarcinoma: impact on management and utilization of quantitative parameters for patient survival prediction. *Nucl Med Commun* 2016;37:231–38 [CrossRef](#)
20. Sole CV, Calvo FA, Alvarez E, et al. Metabolic and molecular relative percentage coreduction in patients with locally advanced rectal cancer treated with neoadjuvant therapy. *Eur J Nucl Med Mol Imaging* 2016;43:1444–52 [CrossRef Medline](#)
21. Rodado-Marina S, Coronado-Poggio M, Garcia-Vicente AM, et al. Clinical utility of <sup>18</sup>F-fluorocholine positron-emission tomography/computed tomography (PET/CT) in biochemical relapse of prostate cancer after radical treatment: results of a multicentre study. *BJU Int* 2015;115:874–83 [CrossRef Medline](#)
22. Khiewvan B, Torigian DA, Emamzadehfard S, et al. Update of the role of PET/CT and PET/MRI in the management of patients with cervical cancer. *Hell J Nucl Med* 2016;19:254–68
23. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol* 2014;32:3059–68 [CrossRef Medline](#)
24. Kendi AT, Brandon D, Switchenko J, et al. Head and neck PET/CT therapy response interpretation criteria (Hopkins criteria) - external validation study. *Am J Nucl Med Mol Imaging* 2017;7:174–80 [Medline](#)
25. Wray R, Sheikhbahaei S, Marcus C, et al. Therapy response assessment and patient outcomes in head and neck squamous cell carcinoma: FDG PET Hopkins criteria versus residual neck node size and morphologic features. *AJR Am J Roentgenol* 2016;207:641–47 [CrossRef](#)
26. Aiken AH, Rath TJ, Anzai Y, et al. ACR Neck Imaging Reporting and Data Systems (NI-RADS): A White Paper of the ACR NI-RADS Committee. *J Am Coll Radiol* 2018;15:1097–108 [CrossRef Medline](#)
27. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol* 2014;32:3048–58 [CrossRef Medline](#)
28. Barrington SF, Qian W, Somer EJ, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging* 2010;37:1824–33 [CrossRef Medline](#)
29. Barrington SF, Kluge R. FDG PET for therapy monitoring in Hodgkin and non-Hodgkin lymphomas. *Eur J Nucl Med Mol Imaging* 2017;44:97–110 [CrossRef Medline](#)
30. 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* 2005;16:1514–23 [CrossRef Medline](#)
31. Razek A, Shamaa S, Lattif MA, et al. Inter-observer agreement of whole-body computed tomography in staging and response assessment in lymphoma: The Lugano Classification. *Pol J Radiol* 2017; 82:441–47 [CrossRef](#)
32. Burggraaff CN, Cornelisse AC, Hoekstra OS, et al. Interobserver agreement of interim and end-of-treatment <sup>18</sup>F-FDG PET/CT in diffuse large B-cell lymphoma: impact on clinical practice and trials. *J Nucl Med* 2018;59:1831–36 [CrossRef Medline](#)
33. Kluge R, Chavdarova L, Hoffmann M, et al. Inter-reader reliability of early FDG-PET/CT response assessment using the Deauville scale after 2 cycles of intensive chemotherapy (OEPA) in Hodgkin's lymphoma. *PLoS One* 2016;11:e0149072 [CrossRef](#)
34. Sawan P, Rebeiz K, Schoder H, et al. Specialized second-opinion radiology review of PET/CT examinations for patients with diffuse

- large B-cell lymphoma impacts patient care and management. *Medicine (Baltimore)* 2017;96:e9411 [CrossRef](#)
35. Ceriani L, Barrington S, Biggi A, et al. Training improves the inter-observer agreement of the expert positron emission tomography review panel in primary mediastinal B-cell lymphoma: interim analysis in the ongoing International Extranodal Lymphoma Study Group-37 study. *Hematol Oncol* 2017;35:548–53 [CrossRef](#) [Medline](#)
  36. Cheson DB. Staging and response assessment in lymphomas: the new Lugano classification. *Chin Clin Oncol* 2015;4:5
  37. Mylam KJ, El-Galaly TC, Hutchings M, et al. Prognostic impact of clinician-based interpretation of <sup>18</sup>F-fluorodeoxyglucose positron emission tomography/computed tomography reports obtained in patients with newly diagnosed diffuse large B-cell lymphoma. *Leuk Lymphoma* 2014;55:1563–69 [CrossRef](#) [Medline](#)
  38. Cottreau AS, El-Galaly TC, Becker S, et al. Predictive value of PET response combined with baseline metabolic tumor volume in peripheral T-cell lymphoma patients. *J Nucl Med* 2018;59:589–95 [CrossRef](#) [Medline](#)
  39. Albano D, Bosio G, Re A, et al. Metabolic behavior and prognostic value of early and end of treatment 18F-FDG PET/CT in adult Burkitt's lymphoma: the role of Deauville and IHP criteria. *Leuk Lymphoma* 2019;60:326–33 [CrossRef](#) [Medline](#)
  40. Qin C, Yang S, Sun X, et al. 18F-FDG PET/CT for prognostic stratification of patients with extranodal natural killer/t-cell lymphoma. *Clin Nucl Med* 2019;44:201–08 [CrossRef](#) [Medline](#)
  41. Gallamini A, Barrington SF, Biggi A, et al. The predictive role of interim positron emission tomography for Hodgkin lymphoma treatment outcome is confirmed using the interpretation criteria of the Deauville five-point scale. *Haematologica* 2014;99:1107–13 [CrossRef](#) [Medline](#)
  42. Jiang C, Su M, Kosik RO, et al. The Deauville 5-Point Scale improves the prognostic value of interim FDG PET/CT in extranodal natural killer/T-cell lymphoma. *Clin Nucl Med* 2015;40:767–73 [CrossRef](#) [Medline](#)
  43. Kong Y, Qu L, Li Y, et al. Predictive significance of a new prognostic score for patients with diffuse large B-cell lymphoma in the interim-positron emission tomography findings. *Medicine (Baltimore)* 2016;95:e2808. [CrossRef](#)
  44. Bakhshi S, Bhethanabhotla S, Kumar R, et al. Posttreatment PET/CT rather than interim PET/CT using Deauville criteria predicts outcome in pediatric Hodgkin lymphoma: A prospective study comparing PET/CT with conventional imaging. *J Nucl Med* 2017; 58:577–83 [CrossRef](#) [Medline](#)
  45. Lombion N, Robin P, Tempescul A, et al. Prognostic value of interim FDG PET-CT in patients older than 60 years with diffuse large B-cell lymphoma treated by PMitCEBO plus rituximab: comparison between Deauville 5-point scale and International Harmonization Project criteria. *Q J Nucl Med Mol Imaging* 2016 [Epub ahead of print] [Medline](#)
  46. Kim J, Song YS, Lee JS, et al. Risk stratification of diffuse large B-cell lymphoma with interim PET-CT based on different cutoff Deauville scores. *Leuk Lymphoma* 2018;59:340–47 [CrossRef](#) [Medline](#)
  47. Baratto L, Davidzon GA, Moghbel M, et al. Comparison between different PET and CT-based imaging interpretation criteria at interim imaging in patients with diffuse large B-cell lymphoma. *Clin Nucl Med* 2018;43:1–8 [CrossRef](#) [Medline](#)
  48. Yuan L, Kreissl MC, Su L, et al. Prognostic analysis of interim <sup>18</sup>F-FDG PET/CT in patients with diffuse large B cell lymphoma after one cycle versus two cycles of chemotherapy. *Eur J Nucl Med Mol Imaging* 2019;46:478–88 [CrossRef](#) [Medline](#)
  49. Isik EG, Kuyumcu S, Kebudi R, et al. Prediction of outcome in pediatric Hodgkin lymphoma based on interpretation of <sup>18</sup>FDG-PET/CT according to  $\Delta$ SUV<sub>max</sub>, Deauville 5-point scale and IHP criteria. *Ann Nucl Med* 2017;31:660–68 [CrossRef](#) [Medline](#)
  50. Kostakoglu L, Goy A, Martinelli G, et al. FDG-PET is prognostic and predictive for progression-free survival in relapsed follicular lymphoma: exploratory analysis of the GAUSS study. *Leuk Lymphoma* 2017;58:372–81 [CrossRef](#) [Medline](#)
  51. Van den Wyngaert T, Helsen N, Carp L, et al. Fluorodeoxyglucose-positron emission tomography/computed tomography after concurrent chemoradiotherapy in locally advanced head-and-neck squamous cell cancer: The ECLYPS Study. *J Clin Oncol* 2017; 35:3458–64 [CrossRef](#) [Medline](#)
  52. Taghipour M, Sheikhabaei S, Wray R, et al. FDG PET/CT in patients with head and neck squamous cell carcinoma after primary surgical resection with or without chemoradiation therapy. *AJR Am J Roentgenol* 2016;206:1093–1100 [CrossRef](#) [Medline](#)
  53. Aiken AH, Farley A, Baugnon KL, et al. Implementation of a novel surveillance template for head and neck cancer: Neck Imaging Reporting and Data System (NI-RADS). *J Am Coll Radiol* 2016; 13:743–746.e1 [CrossRef](#) [Medline](#)
  54. Krieger DA, Hudgins PA, Nayak GK, et al. Initial performance of NI-RADS to predict residual or recurrent head and neck squamous cell carcinoma. *AJNR Am J Neuroradiol* 2017;38:1193–99 [CrossRef](#) [Medline](#)
  55. Wangaryattawanich P, Branstetter BF, Hughes M, et al. Negative predictive value of NI-RADS category 2 in the first posttreatment FDG-PET/CT in head and neck squamous cell carcinoma. *AJNR Am J Neuroradiol* 2018;39:1884–88 [CrossRef](#) [Medline](#)
  56. Rangaswamy B, Fardanesh MR, Genden EM, et al. Improvement in the detection of locoregional recurrence in head and neck malignancies: F-18 fluorodeoxyglucose-positron emission tomography/computed tomography compared to high-resolution contrast-enhanced computed tomography and endoscopic examination. *Laryngoscope* 2013;123:2664–69 [CrossRef](#) [Medline](#)