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; DSFS = Deauville 5-point scale; Nl-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.1 Fortunately, the rate of cancer deaths is declining, with the prevalence of individuals surviving and living with cancer increasing.1 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.2 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,
partial remission/response, no response, or relapsed/progres-
sive disease (PD).² 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 meas-
ured.²,³ 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.² In
RECIST 1.0, all lesions that were at least 1.0 cm in size were meas-
ured, with the maximum number capped at 10 lesions and no
more than 5 lesions per organ. Also, lesion size was now quanti-
fied 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.³,⁴ 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.⁵,⁶ 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.⁶

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.⁷–¹⁰ The
first tumor response assessment to use PET/CT was the PET
Response Criteria in Solid Tumors (PERCIST) tool.¹¹ 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 cho-
sen 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), par-
tial 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.⁶ As
with measurements, different patient factors, scanners, proto-
cols, and PET software algorithms can all contribute to this
variability.⁶ 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.¹²,¹³ 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.¹⁴ The
complete remission unconfirmed classification was eliminated,
with residual masses characterized as complete remission/re-
response or partial remission/response based on the FDG avidity.
Challenges in this International Harmonization Project classifica-
tion 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 can-
cer, pancreatic cancer, rectal cancer, prostate cancer, and cervical
cancer have all been assessed using qualitative FDG-PET/CT
tumor response evaluation techniques.¹⁵–²² 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.¹⁵,¹⁶,²³–²⁶ The
Lugano and Hopkins criteria use relatively stable internal
reference standards for metabolism found on every examina-
tion, blood pool, and liver intensities, which minimizes the
variation in response assessments related to differences in
patients, examination protocols, scanner characteristics, and
readers.¹⁵,²⁵ The NI-RADS criteria use a combination of con-
trast-enhanced CT, together with the PET-CT findings, to
provide anatomic evaluation in areas of increased metabolic
activity.²⁶ In this article, we describe the development of
these 3 criteria and their use in daily practice, and provide il-
lustrative 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 experi-
ence, and research groups.²³ The results were published in 2014
as the Lugano classification for lymphoma staging and response
assessment.¹⁵,²³,²⁷ The novel classification included quantitative
CT parameters for non–FDG-avid lymphomas and independ-
ent PET/CT parameters for FDG-avid lymphomas.¹⁵,²³,²⁷
Although FDG-PET/CT was first included in the Inter-
national 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.¹⁵,²³,²⁷
In addition, the criteria eliminated the cumbersome need for a
bone marrow biopsy in FDG-avid lymphomas, which allows
for PET to serve as a surrogate for determining marrow involvement. The FDG-PET/CT criteria also simplify interpretation by providing a qualitative assessment of the lymphoma treatment 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. 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. CT criteria are based on measurement of lymph nodes, any extranodal lesions, and splenic sizes. Lymph nodes must have a long-axis measurement of \( \leq 1.5 \text{ cm} \), whereas extranodal disease must be at least \( 1.0 \text{ cm} \). 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 \( \leq 1.5 \text{ cm} \) and there is no residual extranodal disease. PD occurs with new lymphadenopathy (long axis >1.5 cm); new nodes, \( \geq 50\% \) increase in PPD, or >0.5 cm increase in diameter for \( \leq 2 \text{ cm} \) or >1.0 cm for >2 cm; (2) splenic volume increase: with splenomegaly, \( >50\% \) increase in length, or without splenomegaly, \( \geq 2 \text{ cm} \) increase in length; (3) new or larger lesions; (4) recurrent previously resolved lesions; (5) new extranodal lesion >1 cm (or \( <1 \text{ cm} \) if unequivocally lymphoma).

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

The Lugano classification incorporates the Deauville 5-point scale (D5PS) for grading FDG avidity. 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, 3, 4, and 5 indicate increasing levels of FDG avidity compared with baseline activity. For non-FDG avid lymphomas, the following CT-based measurement criteria are used to determine tumor response classification; adapted from Ref. 15.

### Table 1: Lugano classification for CT-based lymphoma response

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

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

*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

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

*Adapted from Ref. 23.*

*The D5PS score is used in the Lugano classification for grading lymphoma response.*

### 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.
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-mean that individual variability in patients and PET/CT technology do not impact the assessment (Fig 1 and Table 1). Razek et al\(^{31}\) reported excellent interobserver agreement (95.8% agreement, $\kappa = 0.91$) in assigning D5PS scores and Lugano posttreatment responses. Burgraff et al\(^{32}\) 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.\(^{32}\) By contrast, Kluge et al\(^{33}\) (42% agreement, $\kappa = 0.24$), Sawan et al\(^{34}\) ($\kappa = 0.082$), and Ceriani et al\(^{35}\) ($\kappa = 0.35–0.72$) all reported poorer interobserver agreement for assigning individual D5PS to random scores. Kluge et al\(^{33}\) 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, treatment examination should be compared with the score assigned to the most recent comparison,\(^{23,24}\) which then lead to CMR, partial metabolic response, no metabolic response, or PD designations.

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.
κ = 0.56). Sawan et al.34 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 (κ = 0.86 and radiology-pathology concordance = 78%). Ceriani et al.,35 reported improved interobserver agreement after training (κ = 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.8,13,27,29,36-50 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.23,29,36 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%).23,29 Metabolically active disease in these non-Hodgkin lymphoma cases warrants further evaluation with imaging or biopsy because the activity may simply represent posttreatment inflammation.

### Table 3: Hopkins criteria scores for head and neck cancersa

<table>
<thead>
<tr>
<th>Hopkins Criteria Scoreb</th>
<th>Lesion FDG Uptake</th>
<th>Therapy Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Focal uptake less than IJV</td>
<td>CMR</td>
</tr>
<tr>
<td>2</td>
<td>Focal uptake greater than IJV, but less than liver</td>
<td>Likely CMR</td>
</tr>
<tr>
<td>3</td>
<td>Diffuse uptake greater than IJV and liver</td>
<td>Likely posttreatment inflammation</td>
</tr>
<tr>
<td>4</td>
<td>Focal uptake greater than liver</td>
<td>Likely residual tumor</td>
</tr>
<tr>
<td>5</td>
<td>Focal uptake markedly greater than liver</td>
<td>Residual tumor</td>
</tr>
</tbody>
</table>

*Note—IJV indicates internal jugular vein.

Adapted from Ref. 16.

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

![Baseline vs Interim](image-url)

**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 peri-aortic/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 peri-aortic/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.

### 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.16 The criteria harness the well-documented power of FDG-PET/CT in head and neck tumors, which supersedes anatomic size changes in prognostic value.6,16 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 5.16 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.16 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.51 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
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 al25 demonstrated that the Hopkins criteria and FDG-PET/CT 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 al24 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 al25 indicate that posttreatment inflammatory changes or false-positives can mostly be avoided if the PET/CT is performed $\approx 12$ weeks after the completion of radiation therapy. Taghipour et al52 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 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 al24 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.
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 can 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. 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. The imaging findings are based on combined CT of the neck with contrast and FDG-PET/CT findings (Table 4). 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. 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).

In 2017, Krieger et al 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. 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 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). The NI-RADS guidelines do incorporate recurrence risk rates and definitive management guidelines. 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.

**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

![Table 4: NI-RADS risk stratification guidelines](https://example.com/table4.png)

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

*The NI-RADS risk stratification guidelines grade head and neck tumor treatment response; adapted from Ref. 26."
consider using these classifications when reporting posttreatment response on follow-up PET/CTs.

REFERENCES


34. Sawan P, Rebeiz K, Schofer 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:e9111 CrossRef