

Original Article

Head and neck PET/CT therapy response interpretation criteria (Hopkins criteria) - external validation study

Ayse Tuba Kendi^{1,2}, David Brandon², Jeffrey Switchenko³, Jeffery Trad Wadsworth⁴, Mark W El-Deiry⁴, Nabil F Saba⁵, David M Schuster², Rathan M Subramaniam^{6,7}

¹Department of Radiology, Mayo Clinic, Rochester, MN, USA; ²Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA, USA; ³Department of Biostatistics & Bioinformatics, Emory University, Atlanta, GA, USA; ⁴Otolaryngology, Head and Neck Surgery, Emory University, Atlanta, GA, USA; ⁵Hematology Oncology, Winship Cancer Institute of Emory University, Atlanta, GA, USA; ⁶Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins School of Medicine, Baltimore, MD, USA; ⁷Department of Radiology, UT Southwestern Medical Center, Dallas, TX, USA

Received June 14, 2017; Accepted August 17, 2017; Epub September 1, 2017; Published September 15, 2017

Abstract: Qualitative assessment of PET/CT results in post therapy is very important to provide a reproducible and systemic reporting. A recently introduced response criteria, known as the Hopkins criteria showed promising results. Our aim is to externally validate the Hopkins interpretation system to assess therapy response in head and neck squamous cell cancer (HNSCC). The study included 69 biopsy proven HNSCC patients who underwent post therapy PET/CT between 5-24 weeks after completion of therapy. PET/CT images were interpreted by one nuclear medicine physician and one nuclear radiologist, independently. The studies were scored according to the Hopkins criteria for right neck, left neck, primary tumor site, and overall assessment. Scores 1, 2, 3 were considered as negative and scores 4 and 5 were considered as positive for tumors. Inter-reader variability was assessed using percent agreement and Kappa statistics. Progression-free survival (PFS) was estimated using the Kaplan-Meier method and analyzed using Cox proportional hazards regression. Of the 69 patients, 59 (85.5%) were males, with a mean age of 62.8 years. The percent agreement between readers for overall, right neck, left neck, and primary tumor site were 91.3%, 97.6%, 97.6%, 91.3% respectively. The sensitivity, specificity, positive predictive value, and negative predictive value of the overall therapy assessment were 66.7%, 87.3%, 33%, 96.5% respectively. Cox univariate regression analysis showed positive primary tumor site scores and overall scores were associated with a higher risk of progression ($p < 0.05$). External validation of Hopkins criteria showed excellent inter-reader agreement and prediction of PFS in HNSCC patients.

Keywords: Head and neck carcinoma, PET/CT, Hopkins criteria

Introduction

Therapy of head and neck squamous cell carcinoma (HNSCC) often entails a multimodality approach, including chemotherapy, radiation therapy and surgery. A large number of patients are treated with concurrent chemoradiation (CCRT) [1, 2]. Despite advances in therapy, the recurrence rate is still high, with locoregional recurrence between 15-50% and distant metastasis of 9% [1, 3, 4]. Early identification of locoregional recurrence will change the clinical management completely. Some patients with local treatment failure may undergo salvage surgery [2, 5]. Patients with a complete radiologic response to therapy will in general not

require surgery. Hence, accurate radiological identification of responders and non responders to therapy is of clinical interest [1, 2, 6]. Structural imaging such as computed tomography (CT) and magnetic resonance imaging (MRI) maybe inadequate for response assessment due to tissue distortion, post therapy fibrosis and scar formation. Positron emission tomography (PET) combined with CT, using ¹⁸F-Fluorodeoxyglucose (PET/CT) has been shown useful in evaluation of therapy, if performed at an appropriately timed distance from CCRT [1, 5].

Although the potential benefits of PET/CT for the assessment of therapy response is well recognized, there has been an increased need to

Therapy response criteria for head and neck pet/ct

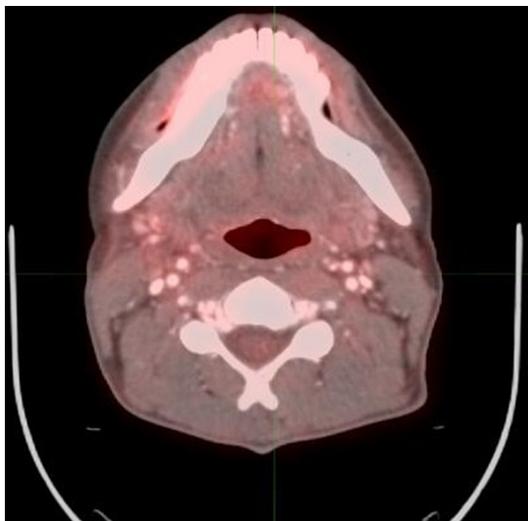


Figure 1. 40 year old male with oropharyngeal carcinoma. PET/CT performed 4 months after completion of therapy shows no evidence of abnormal FDG activity, consistent with score 1, complete response to therapy.

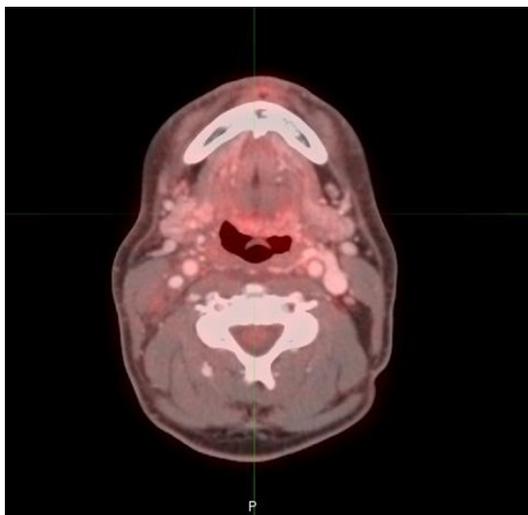


Figure 2. 53 year old male with oropharyngeal carcinoma. PET/CT performed 3 months after completion of therapy shows mild FDG activity at the left base of tongue region, consistent with score of 2.

establish a systemic and reproducible interpretation system to better classify post therapy PET/CT findings of patients with HNSCC [1, 5]. The recently introduced Hopkins criteria addressed this problem and showed a substantial inter-reader agreement with an excellent negative predictive value, prediction of overall survival (OS) and progression free survival (PFS) in patients with HPV-positive and HPV-negative HNSCC. The objective of this study is

to externally validate the reproducibility of the Hopkins criteria at a different institution.

Materials and methods

Patients

After institutional review board approval for this retrospective study, we performed a chart review to identify patients who were treated at our institution between January 2008 and January 2013. Sixty nine patients (59 men and 10 women; mean age \pm SD, 62.8 ± 9.25 y) with primary HNSCC were included. Histopathology was confirmed and patients underwent a baseline ^{18}F -FDG PET/CT and a post therapy ^{18}F -FDG PET/CT between 5-24 weeks after completion of concurrent or radiation therapy. Patients without a baseline ^{18}F -FDG PET/CT, or with no biopsy proven recurrence were excluded; in addition patients were excluded if the ^{18}F -FDG PET/CT was performed 24 weeks after completion of therapy. We considered a PET/CT confirmed within 6 months after completion of therapy as therapy assessment PET/CT and more than 6 months as follow-up PET/CT.

Image analysis

Head and Neck PET/CT Interpretation Criteria (Hopkins Criteria): The studies were scored using a qualitative 5-point scale, known as the Hopkins criteria [1]. Scoring was performed for the primary tumor, right and left neck, and for overall assessment. Maximum standardized uptake value (SUVmax) was used for analysis to get the score. The activity in the internal jugular vein (IJV) was taken as background blood pool for reference. Focal FDG uptake less than IJV was scored 1, consistent with complete metabolic response (**Figure 1**). Focal FDG uptake more than IJV, but less than liver was scored as 2, likely complete metabolic response (**Figure 2**). Diffuse FDG uptake greater than IJV or liver was scored as 3 (**Figure 3**), likely inflammatory changes. Focal FDG uptake more than liver was scored as **Figure 4**, likely residual tumor and intense FDG uptake greater than liver was scored as **Figure 5**, consistent with residual tumor [1].

Definition of Positive and Negative PET/CT Studies: On the basis of Hopkins criteria, the studies were grouped as positive or negative for primary tumor, right neck, left neck, and overall assessment. The scores less than or

Therapy response criteria for head and neck pet/ct

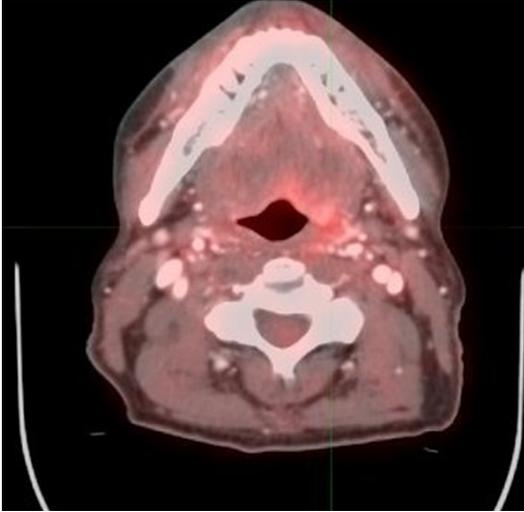


Figure 3. 63 year old male with oropharyngeal carcinoma. PET/CT performed 7 weeks after completion of therapy shows diffuse uptake at the left base of tongue region with maximum SUV higher than the liver. This is most likely post therapy changes, a score of 3.

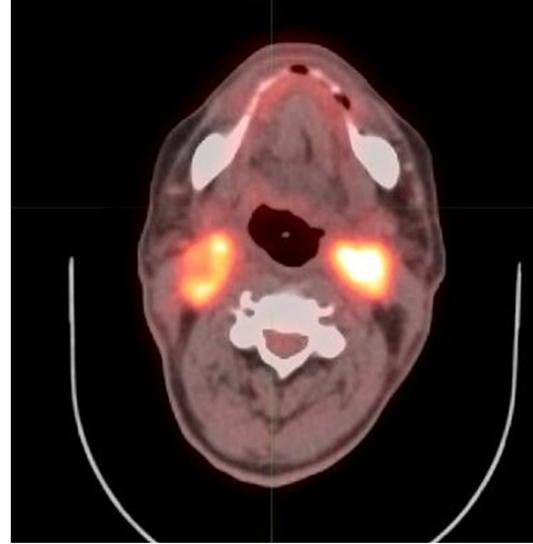


Figure 5. 68 year old male with oropharyngeal carcinoma. PET/CT performed 5 months after completion of therapy shows intense FDG activity at the bilateral mid cervical nodal regions, consistent with residual disease, score of 5.

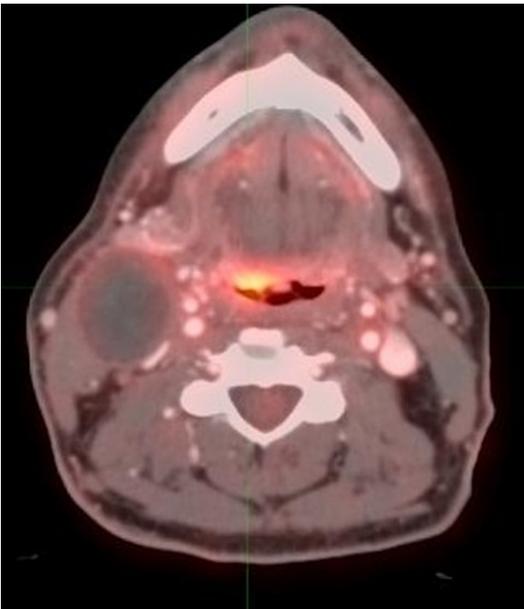


Figure 4. 55 year old male with oropharyngeal carcinoma. PET/CT performed 2 months after completion of therapy shows focal FDG activity with maximum SUV higher than the liver, consistent with score of 4, most likely representing residual disease. Please also note that the large cystic neck adenopathy on the right side.

equal to 3 were considered negative for residual tumor. Any score of 4 or 5 were considered positive for residual tumor (**Figures 4 and 5**).

Reader qualifications: All PET/CT studies were retrieved from the electronic archival system and reviewed on a MIM Software workstation (version 6.1; MIM Software Inc., Cleveland, Ohio) by two reviewers independently (reader 1 and reader 2). Reader 1 was a board certified radiologist with subspecialty trainings in nuclear radiology and neuroradiology (ATK), and the reader 2 was a board certified nuclear medicine physician with expertise in head and neck oncologic imaging (DB). After independent review, two readers reviewed cases together to reach a consensus as well.

Statistical analysis: PET scores were reported for each reader (2 total readers), and for consensus interpretation performed by both readers. Reader scores and consensus scores less than or equal to 3 were considered negative, and positive if greater than 3. Inter-reader variability was assessed using percent agreement and Kappa statistics. Kappa values yield the following interpretation: 0.01-0.2 (Slight agreement), 0.21-0.4 (Fair agreement), 0.41-0.60 (Moderate agreement), 0.61-0.8 (Substantial agreement), 0.81-0.99 (Almost perfect agreement). The ability of PET score to predict progression was assessed by estimating the percent of cases where the PET score matched progression (high PET and progression or low PET and non-progression were considered cor-

Therapy response criteria for head and neck pet/ct

Table 1. Patient Characteristics

Variable	Level	N=69	%
Sex	Female	10	14.5
	Male	59	85.5
Alcohol use	No	35	50.7
	Yes	34	49.3
Smoking status	No	34	49.3
	Yes	35	50.7
HPV status	Negative	8	16.0
	Positive	42	84.0
	Missing	19	-
Primary site	Tonsil	26	37.7
	Larynx	9	13.0
	BOT	27	39.1
	Other (Post oropharynx, hypopharynx, soft palate)	7	10.1
Therapy type	CRT	67	97.1
	XRT	2	2.9
Age	Mean	62.81	-
	Median	63	-
	Minimum	40	-
	Maximum	84	-
	Std Dev	9.25	-
	Missing	0	-

Table 2. Accuracy Statistics

Variable	Level	N=69	%
Accuracy prT score	No	10	14.5
	Yes	59	85.5
Accuracy R-neck score	No	7	16.7
	Yes	35	83.3
	Missing	27	-
Accuracy L-neck score	No	2	4.9
	Yes	39	95.1
	Missing	28	-
Accuracy Overall score	No	10	14.5
	Yes	59	85.5

rect matches). Sensitivity, specificity, positive predicted value (PPV), and negative predicted value (NPV) were also reported. These values were reported for PET scores which were averaged and dichotomized as positive or negative as noted above.

Categorical patient characteristics were compared across averaged PET scores, which were dichotomized as described above, using chi-squared tests or Fisher's Exact tests, where appropriate, and numeric characteristics were

compared using ANOVA. Progression-free survival (PFS) was defined as time from imaging to either progression or last follow-up, and was estimated using the Kaplan-Meier method. Patient characteristics and averaged dichotomized PET scores were compared across survival using log-rank tests, and univariate Cox proportional hazards models were fit. PFS curves were generated for all cases, as well as stratified by HPV status. Additionally, Firth's penalized maximum likelihood estimation was used for each survival model in order to handle empty cells and reduce bias in the confidence intervals and parameter estimates [6]. Hazard ratios, confidence intervals, and *p*-values are reported, and the proportional hazards assumption was checked. The statistical analysis was performed in SAS 9.4, and significance was assessed at the 0.05 level.

Results

Patient characteristics and follow-up

Sixty-nine patients (59 men, 10 women) met the eligibility and inclusion criteria. The mean age \pm SDD was 62.8 ± 9.25 . A history of alcohol use was present in 34 patients (49.3%), and a history of smoking was present in 35 patients (50.7%). HPV was positive in 42 patients (84%), was negative in 8 patients (16%). The primary site of tumor was classified as tonsil (37.7%), base of tongue (39.1%), larynx (13%), other sites (10.1%). Patient demographics and characteristics are listed in **Table 1**. All patients were followed up until death or September 2015.

Time interval of post therapy PET/CT

All 69 PET/CT studies were performed between 5 and 24 weeks after completion of therapy. The mean of months between completion of therapy and post-therapy imaging was 3.05 (median 2.99, maximum was 5.55 months).

Reader classification of PET/CT studies

The diagnostic accuracies of the scoring system were calculated for consensus reading. **Table 2** summarized the accuracy of each score. If the score matches progression, then the score was accurate. Accuracy ranged from 83.3% to 95.1%. Overall, we found moderate to almost perfect agreement between readers when assessing score variables (**Table 3**).

Therapy response criteria for head and neck pet/ct

Table 3. Agreement statistics

Variable	Percent agreement	Kappa statistic
prT score	63/69 (91.3%)	0.576
L-neck score	40/41 (97.6%)	0.788
R-neck score	41/42 (97.6%)	0.844
Overall score	63/69 (91.3%)	0.650

We found sensitivity ranges across primary tumor, right and left neck, and overall assessment between 20% to 66.7%, specificity ranging between 87.3% and 94.4%. Positive predictive value (PPV) of primary tumor score was 30%, and negative predictive value (NPV) was 94.9%. We found PPVs 25-33%, and NPVs ranging from 89.5% to 96.5%.

Kaplan-Meier survival curves: therapy assessment score and survival outcome in all patients (n=69)

Six of the 69 patients were found to have disease recurrence. None of the patients died within the period of the study. Due to small number of events in HPV+ and HPV-subsets, some of survival curves could not be estimated, as a result only survival curves of all patients reported.

Positive primary tumor scores and overall scores were associated with a higher risk of progression ($p < 0.05$).

Discussion

The primary aim of this study was to externally validate Hopkins criteria for FDG PET/CT therapy response assessment of HNSCCs. We aimed to establish its reader reliability, accuracy, predictive values for PFS. Our study showed that Hopkins criteria for post therapy response assessment has a moderate to almost perfect inter-reader agreement, has a NPV of 94.9% for primary tumor and 96.5% for nodal disease.

Identification of residual or recurrent disease after initial therapy of HNSCC is of value as it may affect the clinical management [1, 2, 7]. PET/CT's potential to assess response to therapy has been widely accepted in the recent years and have shown a more accurate assessment of treatment response in comparison to enhanced CT alone [7-9]. The presence of lack of abnormal FDG uptake in a post therapy PET/CT can therefore provide a higher degree of reassurance of a successful treatment [2].

Currently there is no consensus on how to assess and report therapeutic response by PET/CT. The use of semi-quantitative ways of tumor metabolism has been widely accepted [7]. The standardized uptake values (SUV), especially maximum SUV (SUVmax) are in use. SUV is known to have limited value for assessing treatment response [1, 7, 10, 11].

Qualitative assessment of treatment response by using visual inspection and identifying relative difference between the tumor site and background surrounding tissue is usually adequate to assess therapy response [2, 6]. However there is still lack of studies supporting value of visual inspection as well as validation of proposed criteria. In order to overcome this problem, different Likert scales has been introduced, the most widely known is the Deauville criteria that is used to assess therapy response in patients with lymphoma [6]. The Deauville criteria was based on relative metabolic activity of the tumor compared to the mediastinal blood pool and liver activity. The sensitivity, specificity, PPV, NPV, and accuracy of the Deauville criteria were 73%, 94%, 73%, 94%, and 91%, respectively [11]. The Hopkins interpretation criteria showed relatively lower sensitivity and PPV, likely related to CCRT [1]. This was expected as most of lymphoma patients were treated primarily with chemotherapy [11].

Krabbe et al [12] used a five point scale in a serial of PET evaluation and demonstrated PPV of 51% and NPV of 100%. Marcus et al [1] achieved substantial inter-reader reliability and showed a PPV of 71.1% and a high NPV of 91.1%. Sjo et al [6] combined SUVmax and Deauville criteria. This study showed a PPV of 68.7% and NPV of 86.4%. Porceddu et al [13] performed qualitative assessment of post therapy neck in a prospective study. They compared the focal FDG uptake at the site of nodal disease to the adjacent background or liver uptake. In this study, NPV was as high as 97.1% [13].

Recently a head and neck imaging reporting system has been introduced for contrast enhanced CT associated with a PET/CT, known as neck imaging reporting and data systems (NI-RADS) [14]. NI-RADS integrated PET/CT with contrast enhanced CT results and categorized PET uptake as mild/intermediate uptake or as intense uptake. Compared to NI-RADS,

Therapy response criteria for head and neck pet/ct

Hopkins criteria is a more detailed Likert scale dedicated for PET/CT.

Our study's limitations include a small sample; assessment of PFS was limited as there were a small number of events. We were also not able to assess PFS for HPV+ and HPV- due to small number of events. Another limitation was to have only two readers, although excellent inter-reader agreement was achieved. The patient population was mainly HPV (+) oropharyngeal carcinoma (OPC). This likely resulted in higher number of male patients as incidence of HPV (+) OPC is more common in white males [15].

In conclusion, the introduction of cancer specific Likert scales like Hopkins criteria should be encouraged and also externally validated by different institutions. Production of PET/CT reports by using Likert scales like the Hopkins criteria will significantly reduce the equivocal reports. The consistent use of Hopkins criteria may result in better categorization of the results, which will be easily interpreted by the referring oncologists and head and neck surgeons. The Hopkins criteria has a moderate to almost perfect inter-reader agreement and a very high NPV.

Acknowledgements

Research reported in this publication was supported in part by the Biostatistics and Bioinformatics Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of conflict of interest

None.

Address correspondence to: Ayse Tuba Kendi, Department of Radiology, Mayo Clinic, 200 Second Street SE, 55905, Rochester, MN, USA. Tel: 612-408-9737; E-mail: kendi.ayse@mayo.edu

References

[1] Marcus C, Ciarallo A, Tahari AK, Mena E, Koch W, Wahl RL, Kiess AP, Kang H, Subramaniam RM. Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-inter-

reader reliability, accuracy, and survival outcomes. *J Nucl Med* 2014; 55: 1411-1416.

- [2] Sherriff JM, Ogunremi B, Colley S, Sanghera P, Hartley A. The role of positron emission tomography/CT imaging in head and neck cancer patients after radical chemoradiotherapy. *Br J Radiol* 2012; 85: e1120-1126.
- [3] Bourhis J, Le Maitre A, Baujat B, Audry H, Pignon JP. Individual patients' data meta-analyses in head and neck cancer. *Curr Opin Oncol* 2007; 19: 188-194.
- [4] Garavello W, Ciardo A, Spreafico R, Gaini RM. Risk factors for distant metastases in head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg* 2006; 132: 762-766.
- [5] Basu S, Kumar R, Ranade R. Assessment of treatment response using PET. *PET Clin* 2015; 10: 9-26.
- [6] Sjövall J, Bitzén U, Kjellén E, Nilsson P, Wahlberg P, Brun E. Qualitative interpretation of PET scans using a Likert scale to assess neck node response to radiotherapy in head and neck cancer. *Eur J Nucl Med Mol Imaging* 2016; 43: 609-616.
- [7] Passero VA, Branstetter BF, Shuai Y, Heron DE, Gibson MK, Lai SY, Kim SW, Grandis JR, Ferris RL, Johnson JT, Argiris A. 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: 2278-2283.
- [8] Sjövall J, Chua B, Pryor D, Burmeister E, Foote MC, Panizza B, Burmeister BH, Porceddu SV. Long-term results of positron emission tomography-directed management of the neck in node-positive head and neck cancer after organ preservation therapy. *Oral Oncol* 2015; 51: 260-266.
- [9] Moeller BJ, Rana V, Cannon BA, Williams MD, Sturgis EM, Ginsberg LE, Macapinlac HA, Lee JJ, Ang KK, Chao KS, Chronowski GM, Frank SJ, Morrison WH, Rosenthal DI, Weber RS, Garden AS, Lippman SM, Schwartz DL. Prospective risk-adjusted [¹⁸F]Fluorodeoxyglucose positron emission tomography and computed tomography assessment of radiation response in head and neck cancer. *J Clin Oncol* 2009; 27: 2509-2515.
- [10] Gourin CG, Boyce BJ, Williams HT, Herdman AV, Bilodeau PA, Coleman TA. Revisiting the role of positron-emission tomography/computed tomography in determining the need for planned neck dissection following chemoradiation for advanced head and neck cancer. *Laryngoscope* 2009; 119: 2150-2155.
- [11] Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, Meignan M, Malkowski B, Hofman MS, Barrington SF. Interna-

Therapy response criteria for head and neck pet/ct

- tional validation study for interim PET in ABVD-treated, advanced-stage hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. *J Nucl Med* 2013; 54: 683-690.
- [12] Krabbe CA, Pruim J, Dijkstra PU, Balink H, van der Laan BF, de Visscher JG, Roodenburg JL. 18F-FDG PET as a routine posttreatment surveillance tool in oral and oropharyngeal squamous cell carcinoma: a prospective study. *J Nucl Med* 2009; 50: 1940-1947.
- [13] Porceddu SV, Pryor DI, Burmeister E, Burmeister BH, Poulsen MG, Foote MC, Panizza B, Coman S, McFarlane D, Coman W. Results of a prospective study of positron emission tomography-directed management of residual nodal abnormalities in node-positive head and neck cancer after definitive radiotherapy with or without systemic therapy. *Head Neck* 2011; 33: 1675-1682.
- [14] Aiken AH, Farley A, Baugnon KL, Corey A, El-Deiry M, Duszak R, Beitler J, Hudgins PA. 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.
- [15] Saba NF, Goodman M, Ward K, Flowers C, Ramalingam S, Owonikoko T, Chen A, Grist W, Wadsworth T, Beitler JJ, Khuri FR and Shin DM. Gender and ethnic disparities in incidence and survival of squamous cell carcinoma of the oral tongue, base of tongue, and tonsils: a surveillance, epidemiology and end results program-based analysis. *Oncology* 2011; 81: 12-20.