



This information is current as  
of August 23, 2025.

**Using FDG-PET to Measure Early Treatment  
Response in Head and Neck Squamous Cell  
Carcinoma: Quantifying Intrinsic Variability  
in Order to Understand Treatment-Induced  
Change**

J.K. Hoang, S.K. Das, K.R. Choudhury, D.S. Yoo and D.M.  
Brizel

*AJNR Am J Neuroradiol* 2013, 34 (7) 1428-1433

doi: <https://doi.org/10.3174/ajnr.A3412>

<http://www.ajnr.org/content/34/7/1428>

# Using FDG-PET to Measure Early Treatment Response in Head and Neck Squamous Cell Carcinoma: Quantifying Intrinsic Variability in Order to Understand Treatment-Induced Change

J.K. Hoang, S.K. Das, K.R. Choudhury, D.S. Yoo, and D.M. Brizel

## ABSTRACT

**BACKGROUND AND PURPOSE:** Quantification of both baseline variability and intratreatment change is necessary to optimally incorporate functional imaging into adaptive therapy strategies for HNSCC. Our aim was to define the baseline variability of SUV on FDG-PET scans in patients with head and neck squamous cell carcinoma and to compare it with early treatment-induced SUV change.

**MATERIALS AND METHODS:** Patients with American Joint Committee on Cancer stages III-IV HNSCC were imaged with 2 baseline PET/CT scans and a third scan after 1–2 weeks of curative-intent chemoradiation.  $SUV_{max}$  and  $SUV_{mean}$  were measured in the primary tumor and most metabolically active nodal metastasis. Repeatability was assessed with Bland-Altman plots. Mean percentage differences (% $\Delta$ SUV) in baseline SUVs were compared with intratreatment % $\Delta$ SUV. The repeatability coefficient for baseline % $\Delta$ SUV was compared with intratreatment % $\Delta$ SUV.

**RESULTS:** Seventeen patients had double-baseline imaging, and 15 of these patients also had intratreatment scans. Bland-Altman plots showed excellent baseline agreement for nodal metastases  $SUV_{max}$  and  $SUV_{mean}$ , but not primary tumor SUVs. The mean baseline % $\Delta$ SUV was lowest for  $SUV_{max}$  in nodes ( $7.6\% \pm 5.2\%$ ) and highest for  $SUV_{max}$  in primary tumor ( $12.6\% \pm 9.2\%$ ). Corresponding mean intratreatment % $\Delta$ SUV<sub>max</sub> was  $14.5\% \pm 21.6\%$  for nodes and  $15.2\% \pm 22.4\%$  for primary tumor. The calculated RC for baseline nodal  $SUV_{max}$  and  $SUV_{mean}$  were 10% and 16%, respectively. The only patient with intratreatment % $\Delta$ SUV above these RCs was 1 of 2 patients with residual disease after CRT.

**CONCLUSIONS:** Baseline SUV variability for HNSCC is less than intratreatment change for SUV in nodal disease. Evaluation of early treatment response should be measured quantitatively in nodal disease rather than the primary tumor, and assessment of response should consider intrinsic baseline variability.

**ABBREVIATIONS:** CRT = chemoradiation; HNSCC = head and neck squamous cell carcinoma; ICC = intraclass correlation coefficient; RC = repeatability coefficient; SUV = standardized uptake value

FDG-PET is the most widely used functional imaging technique in head and neck squamous cell carcinoma. Pretreatment imaging has a significant role in initial staging, prognosis assessment, and target delineation.<sup>1</sup> Posttreatment FDG-PET has become an important tool for the assessment of residual disease in cervical

lymph nodes.<sup>2,3</sup> Another area of active investigation is the use of PET to monitor therapy response during treatment. PET performed early in treatment (intratreatment PET) could detect favorable or unfavorable metabolic changes before anatomic changes are evident and could help determine whether a particular therapeutic strategy should be maintained or changed. This approach could enhance the choice of initial treatment and facilitate the use of adaptive radiation therapy strategies, including dose escalation, selection of nonresponding patients for new molecularly targeted therapies, or discontinuation in favor of primary surgery, among other options.<sup>4</sup>

Early response assessment with FDG-PET has been evaluated in lymphoma, soft-tissue sarcoma, and esophageal and lung cancers.<sup>5–8</sup> Findings that early treatment changes in glucose metabolism can predict histopathologic response or survival have led to proposals of using standardized uptake value cutoff values to stratify patients by outcome.<sup>5–7</sup> One of the largest studies of neo-

Received August 16, 2012; accepted after revision September 29.

From the Departments of Radiology (J.K.H., K.R.C.) and Radiation Oncology (J.K.H., S.K.D., D.S.Y., D.M.B.), Duke University Medical Center, Durham, North Carolina.

This work was supported by GE-AUR Research Grant to Jenny K. Hoang.

Paper previously presented at: Annual Meeting of the American Society of Neuroradiology and the Foundation of the ASNR Symposium, April 21–26, 2012, New York, New York. Awards: ASTRO travel grant for Head and Neck Symposium 2012 for the abstract and ASNR best scientific paper in the Head and Neck Symposium for oral presentation.

Please address correspondence to Jenny K. Hoang, MD, Department of Radiology, Division of Neuroradiology, Duke University Medical Center, Box 3808, Erwin Rd, Durham NC, 27710; e-mail: jennykh@gmail.com

<http://dx.doi.org/10.3174/ajnr.A3412>

adjuvant chemotherapy for esophageal cancer identified responders with high sensitivity by using 0% SUV decrease as a cutoff (ie, any decrease in SUV),<sup>5</sup> and the authors concluded that a decrease in SUV of any magnitude would indicate an early treatment response. In practice, using such small changes to signify treatment response should be viewed with caution, for it is known that PET scans repeated days or even hours apart without intervening treatment can vary considerably in terms of SUV.<sup>9-13</sup>

The significance of this phenomenon is that change observed during the course of treatment must be greater than inherent baseline variability to correctly attribute the observed change to the treatment itself. Intrinsic variability of SUV in the absence of treatment reflects biologic, technical, and observer variation. This fluctuation was observed recently in HNSCC in a study that evaluated change in SUV<sub>max</sub> on pretreatment PET/CT scans that were performed on different scanners.<sup>13</sup> The authors warned about the need to account for variability in PET biomarkers in clinical protocols. Wahl et al,<sup>14</sup> who proposed criteria for the Positron Emission Response Criteria in Solid Tumor (PERCIST) trial, also stated that more studies were needed to address questions concerning the reproducibility of baseline quantitative readings and PET response during the initial phases of treatment. Data are sparse, but baseline tumor PET metabolic activity for tumors outside the head and neck can vary by 10%–16% in single-center studies<sup>9-12</sup> and up to 39% in multicenter studies.<sup>11</sup>

Quantification of both baseline variability and intratreatment change is necessary to optimally incorporate functional imaging into adaptive therapy strategies for HNSCC. The aim of this prospective study was to define the intrinsic (pretreatment) variability of tumor SUV and compare it with early treatment-induced (intratreatment) change in patients with HNSCC. We hypothesized that intratreatment changes in HNSCC would be larger than the intrinsic variability in metabolic activity in patients responding favorably to treatment. A secondary aim was to determine whether the magnitude of intrinsic variability differed between primary tumor and nodal metastases or according to the parameter used to describe SUV, namely SUV maximum (SUV<sub>max</sub>) or SUV<sub>mean</sub>.

## MATERIALS AND METHODS

### Patient Selection and Imaging Protocol

Patients with newly diagnosed American Joint Committee on Cancer stages III-IV head and neck squamous cell carcinoma scheduled to undergo curative-intent chemoradiotherapy were prospectively enrolled between September 2009 and August 2011. Exclusion criteria included patients younger than 18 years of age, the presence of a synchronous second malignancy, and diabetes mellitus. To avoid data contamination by image noise, we excluded patients with both tumor and node SUV<sub>max</sub> of  $\leq 4$ . The Cancer Center Protocol Review Committee and the institutional review board of our institution approved the trial. Written informed consent was obtained from all patients before enrollment.

The imaging protocol specified the performance of 2 baseline pretreatment FDG-PET/CT scans (PET1 and PET2) separated by 1 week. The second scan was to be obtained just before the initiation of therapy. The third PET/CT was to be obtained after completion of the first week of CRT (PET3) to assess early treatment-

induced change. All patients were scheduled to receive a standard institutional regimen of CRT, consisting of intensity-modulated radiation therapy, 2 Gy once daily to 70 Gy. Chemotherapy consisted of 2 cycles of cisplatin during weeks 1 and 5 of intensity-modulated radiation therapy (20 mg/m<sup>2</sup>/day  $\times$  5 days per each cycle).

### PET/CT Scanning Technique

All acquisitions were performed by using 1 of 2 integrated PET/CT scanners, the Discovery STE with 16-section CT (GE Healthcare, Milwaukee, Wisconsin) and the Biograph mCT PET/CT System with 128-detector CT (Siemens Medical Solutions, Erlangen, Germany). All scans for any given patient were obtained on the same scanner. Patients fasted for at least 4 hours before intravenous administration of FDG (5.92 MBq/kg of body weight, with a minimum of 296 MBq and maximum of 555 MBq). Serum glucose concentrations were obtained in all patients and were  $<200$  mg/dL (11.1 mmol/L) (normal range, 70–115 mg/dL) in all patients. After an uptake phase of 60 minutes, patients were positioned in a head and neck immobilizer device, and unenhanced CT from the midcranium to the thoracic inlet was performed with the arms down (3.75-mm-thick contiguous images with 30-cm FOV).

The CT scan was followed by dedicated PET/CT neck images obtained during 1 or 2 bed positions (position-emission scan, 68 minutes/bed), with the patient's arms down. Both PET scanners had a resolution of 5-mm full width at half maximum and yielded PET sections with 3.27-mm center-to-center spacing. PET images were reconstructed with corrections for attenuation, scatter, random events, and dead time by using ordered subsets expectation maximization, resulting in a  $128 \times 128$  matrix. The FOV was 30 cm with 3 iterations of ordered subsets expectation maximization. Contrast-enhanced CT scans were obtained separately if they had not been obtained with the baseline PET/CT, to optimize target delineation to facilitate treatment planning.

### Image Analysis

PET/CT images were analyzed by a fellowship-trained, board-certified neuroradiologist (10 years' experience reading CT, without a Certificate of Added Qualification, 50% practice in head and neck imaging) who knew the site of the primary tumor but was blinded to the tumor and nodal staging and clinical treatment response. All PET studies were analyzed quantitatively with a software platform capable of deformable registration of multimodality images (VelocityAI; Velocity Medical Solutions, Atlanta, Georgia) in axial, coronal, and sagittal planes. On the PET scans, metabolic volumes were manually delineated in 2 tumor sites: the primary tumor and the most metabolically active nodal metastasis. Correlation to CT images was made to ensure accurate delineation.

The SUV was calculated by using the following formula:

$$\text{SUV} = c_{dc}/(d_i/w),$$

where  $c_{dc}$  is the decay-corrected tracer tissue concentration (in becquerels per gram),  $d_i$  is the injected dose (in becquerels), and  $w$  is the patient's body weight (in grams).

SUV was measured as SUV<sub>max</sub> and SUV<sub>mean</sub>. These param-

**Table 1: Patient characteristics**

Patient No.	Sex	Age (yr)	Primary Site and Stage				HPV Status	Double-Baseline PET	Intratreatment PET	Days between PET2 and PET3
1	M	65	Base of tongue	T1	N2b	M0	+	Yes	Yes	17
2	M	50	Tonsil	T1	N2a	M0	+	Yes	No	
3	M	44	Tonsil	T2	N2b	M0	Unknown	Yes	Yes	12
4	M	60	Tonsil	T1	N2b	M0	+	Yes	Yes	14
5	M	38	Nasopharynx	T2	N2	M0	Unknown	Yes	Yes	17
6	M	55	Base of tongue	T4	N0	M0	Unknown	Yes	Yes	13
7	M	53	Base of tongue	T4	N2c	M0	—	Yes	Yes	13
8	M	48	Tonsil	T2	N2b	M0	+	Yes	Yes	11
9	M	48	Tonsil	T2	N2b	M0	+	Yes	Yes	25
10	M	55	Tonsil	T1	N2b	M0	+	Yes	No	
11	M	50	Tonsil	T1	N2c	M0	+	Yes	Yes	13
12	M	60	Tonsil	T2	N2b	M0	+	Yes	Yes	21
13	M	39	Tonsil	T3	N2b	M0	+	Yes	Yes	14
14	M	49	Soft palate	T2	N2c	M0	—	Yes	Yes	23
15	M	42	Tonsil	T4	N0	M0	+	Yes	Yes	13
16	M	68	Base of tongue	T2	N2b	M0	+	Yes	Yes	11
17	M	49	Base of tongue	T3	N0	M0	+	Yes	Yes	8

**Note:**— + indicates positive; —, negative; HPV, human papillomavirus.

ters were obtained from a VOI that was generated by contouring a region of interest onto all axial images covering the metabolically active tumor. Large photopenic areas in the center of the nodal disease were excluded.  $SUV_{max}$  was defined as the highest pixel value in the VOI.  $SUV_{mean}$  was defined by the average pixel value for the VOI. This approach was used for the 2 baseline and the single intratreatment scans.

### Statistical Analysis

Differences in metabolic activity among the 2 scans were described as SUV unit differences and SUV percentage differences. If  $SUV_1$ ,  $SUV_2$ , and  $SUV_3$  are the respective measurements of SUV on PET1, PET2, and PET 3, then the formulas for SUV unit differences ( $\Delta$ ) and SUV percentage differences ( $\%\Delta$ ) are as follows:

$$\Delta \text{ Baseline SUV} = \text{Absolute } (SUV_1 - SUV_2),$$

$$\Delta \text{ Intratreatment SUV} = [(SUV_1 + SUV_2)/2] - SUV_3,$$

$$\%\Delta \text{ Baseline SUV} = \text{Absolute } (SUV_1 - SUV_2) \times 100 / [(SUV_1 + SUV_2)/2],$$

$$\%\Delta \text{ Intratreatment SUV} = [(SUV_1 + SUV_2)/2] - SUV_3 \times 100 / [(SUV_1 + SUV_2) / 2].$$

Repeatability of baseline measurements was examined graphically for each patient by using a Bland-Altman plot, which displays the mean of  $SUV_1$  and  $SUV_2$  against the difference between  $SUV_1$  and  $SUV_2$  ( $SUV_1 - SUV_2$ ).<sup>15</sup> Baseline repeatability was quantified with the intraclass correlation coefficient.<sup>10</sup>

Intratreatment change in the SUV was compared with baseline differences in SUV. The paired  $t$  test was used to determine whether there were statistically significant differences between these 2 measurements. A  $P$  value of  $<.05$  was statistically significant. The ICC for  $SUV_1$  and  $SUV_2$  was also compared with the ICC for  $SUV_2$  and  $SUV_3$ . A repeatability coefficient was also derived from the baseline  $\%\Delta SUV$  and was compared with the intratreatment  $\%\Delta SUV$ . The RC has been applied to measure intrinsic baseline variability<sup>10</sup> and is defined as the SD of baseline change multiplied by 1.96. This implies that the difference between repeated test results can be expected to be greater than RC only 5% of the time. Thus, intratreatment  $\%\Delta SUV$  would have to

be greater than the RC to be confident that the change represented more than baseline variability.

Finally, the SUVs of the 3 PET scans were examined for significant treatment changes by using a 2-factor ANOVA with subject and treatment status (pre- or post-) being the 2 factors.

## RESULTS

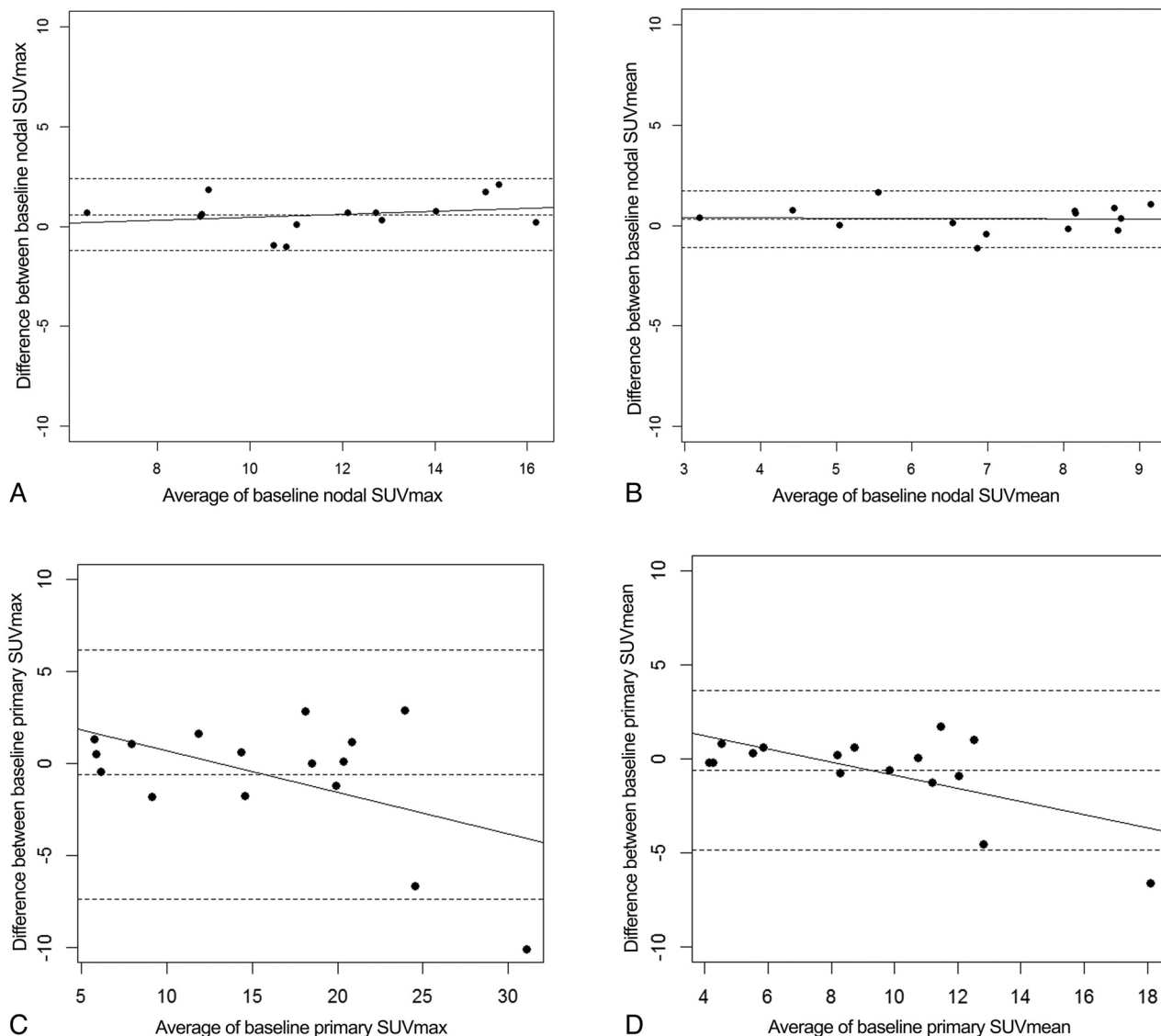
### Patients and Treatment Outcome

Nineteen patients were enrolled in the study. Two patients were excluded because the  $SUV_{max}$  in both the primary tumor and node at baseline was  $\leq 4$ . The 17 remaining patients were all men with a mean age of  $51 \pm 8.5$  years. Table 1 summarizes patient demographics and disease characteristics. All patients had double-baseline PET scans. Two of 17 patients did not have intratreatment scans because they did not wish to have the third scan. The median interval between the 2 baseline scans was 9 days (interquartile range [IQR], 7–13 days). The median interval between the intratreatment and second baseline scan was 13 days (IQR, 12.5–17 days). The median radiation dose to the clinical target volume at the time of the intratreatment scan was 12 Gy (IQR, 10–14 Gy). The timing in days between PET2 and PET3 is shown on Table 1. Three patients were scanned in the third intratreatment week because of conflicts in scheduling scanning and treatment or technical issues with the scanner.

Fifteen patients had no residual disease after CRT by clinical examination and/or on a PET/CT performed 12 weeks posttreatment. The posttreatment PET scan was obtained per clinical routine but not explicitly as part of this study. Two patients (7 and 14) had pathologically confirmed residual disease in the cervical lymph nodes.

### Baseline Repeatability

The baseline repeatability for primary tumor and nodes are displayed as Bland-Altman plots in Fig 1. There was excellent agreement between the 2 baseline studies for nodes measured as  $SUV_{max}$  and  $SUV_{mean}$ . The mean difference between baseline nodal SUV was  $0.58 \pm 0.90$  for  $SUV_{max}$  and  $0.33 \pm 0.71$  for  $SUV_{mean}$ . Figure 1A shows that all points for nodal  $SUV_{max}$  were within the mean  $\pm 1.96 \times SD$ . The solid black line in Fig 1 is the



**FIG 1.** Bland-Altman plots for baseline tumor SUV<sub>max</sub> and SUV<sub>mean</sub>. Graphs show the difference between the 2 baseline SUV measurements plotted against their average for SUV<sub>max</sub> in the node (A), SUV<sub>mean</sub> in the node (B), SUV<sub>max</sub> in primary tumor (C), and SUV<sub>mean</sub> in primary tumor (D). The solid line is the mean of differences, and dotted lines indicate the limits of agreement (mean of difference,  $\pm 1.96 \times \text{SD}$ ).

**Table 2: Means (SD, range) and intraclass coefficients for baseline and intratreatment SUV differences**

	Baseline $\Delta\text{SUV}$	Intratreatment $\Delta\text{SUV}$	Baseline $\%\Delta\text{SUV}$	Intratreatment $\%\Delta\text{SUV}$	<i>P</i> Value <sup>a</sup>	ICC of SUV1 and 2 <sup>b</sup>	ICC of SUV2 and 3 <sup>b</sup>
Primary tumor							
SUV <sub>max</sub>	2.13 (2.65)	-2.92 (4.43)	12.6 (9.2)	15.2 (22.4)	.56	0.94 (0.76–0.97)	0.67 (0.27–0.87)
SUV <sub>mean</sub>	1.27 (1.77)	-1.06 (2.24)	11.4 (10.5)	9.9 (22.6)	.81	0.91 (0.76–0.97)	0.67 (0.26–0.87)
Nodal metastasis							
SUV <sub>max</sub>	0.87 (0.61)	-1.84 (2.62)	7.6 (5.2)	14.5 (21.6)	.25	0.95 (0.86–0.98)	0.57 (0.11–0.83)
SUV <sub>mean</sub>	0.61 (0.46)	-1.29 (1.47)	9.4 (8.0)	15.1 (21.8)	.30	0.93 (0.81–0.97)	0.46 (-0.03–0.78)

<sup>a</sup> *P* value for the paired test of whether the baseline  $\%\Delta\text{SUV}$  is different from the intratreatment  $\%\Delta\text{SUV}$ .

<sup>b</sup> SUV1 and 2 represent baseline-baseline ICC. SUV 2 and 3 represent baseline-intratreatment ICC.

line of least-squares best fit for the association of the difference with the average. The slope of the line is close to zero, suggesting that the size of the difference in SUV does not depend on the magnitude of the SUV. The Bland-Altman plots for primary tumor revealed poorer agreement compared with nodal disease (Fig 1C, -D). In particular, at higher SUVs for primary tumor, there were larger differences between the 2 baseline SUV measurements.

The ICC for SUV1 and 2 (baseline SUVs) was high for SUV<sub>max</sub> and SUV<sub>mean</sub> in primary tumor and nodes (Table 1), but it was lowest for primary tumor SUV<sub>mean</sub> (0.91) and highest for nodal SUV<sub>max</sub> (0.95) (Table 2).

#### Intratreatment Change

The difference between mean baseline SUV variability and mean intratreatment change was larger for nodes than for primary tu-



mor as seen in Table 2. This was due to poorer repeatability in primary tumor compared with nodes. The differences between baseline and intratreatment % $\Delta$ SUV did not reach statistical significance by the paired *t* test, but we did find a nonoverlapping 95% confidence interval for the ICC of SUV1 and 2 (baseline-baseline) and the ICC of SUV2 and 3 for both nodal SUV<sub>max</sub> and nodal SUV<sub>mean</sub>. This suggested that the ICC of SUV2 and 3 was significantly different from the ICC of SUV1 and 2. In contrast, primary tumor SUV had an overlapping 95% confidence interval for the 2 sets of ICCs (not statistically significant).

The calculated repeatability coefficients for baseline nodal SUV<sub>max</sub> and SUV<sub>mean</sub> were 10% and 16%, respectively. Any values less than RC for % $\Delta$  intratreatment SUV could be due to intrinsic baseline variability rather than true intratreatment change. Only 1 patient had an intratreatment increase in both nodal SUV<sub>max</sub> and SUV<sub>mean</sub> above the respective RCs. This was patient 7, who was one of the patients with residual disease after completion of CRT. The other patient who was a nonresponder (patient 14) did not have a rise in the SUV<sub>max</sub> or SUV<sub>mean</sub>, but his intratreatment PET was performed later (after 22 Gy of radiation) compared with after 12 Gy for patient 7 (group median, 12 Gy).

ANOVA showed a significant effect of treatment on nodal SUV<sub>max</sub> and SUV<sub>mean</sub> ( $P < .002$ ). For the primary tumor, the effect of treatment was significant for SUV<sub>max</sub> ( $P = .006$ ) and marginal for SUV<sub>mean</sub> ( $P = .06$ ).

## DISCUSSION

The proper interpretation of treatment-induced changes in tumor glucose use with PET requires a quantitative understanding of the inherent variation of this metabolic parameter in the absence of treatment. Metabolic processes are dynamic and fluctuate as opposed to anatomic parameters such as tumor volume, which are relatively stable and static. The current study is the first to address this problem for HNSCC via the performance of double-baseline scans and an intratreatment scan in the same patient on the same scanner. Changes in glucose metabolism during the early phases of CRT were greater than the intrinsic variation in nodal metastases but not the primary tumor.

Temporal variability in pretreatment metabolic activity for HNSCC was recently reported by Chu et al<sup>13</sup> in a retrospective study using diagnostic PET/CT and planning PET/CT. They reported serial change as mean composite SUV velocity (SUV<sub>max</sub> change divided by time in weeks between scans). Factors that contributed to a mean composite SUV velocity of  $-0.1/\text{week}$  and a wide SD of 2.0 were a longer interval between scans (median interval of 3 weeks) and use of different PET/CT scanners.

The current study attempted to control for these additional factors affecting repeatability. The magnitude of pretreatment SUV variability in nodes and primary tumor ranged from 8% to 13%, which is consistent with the literature for lung and gastrointestinal malignancies with similar short time intervals between baseline scans. These tumors have mean baseline % $\Delta$ SUV ranging from 3% to 16%, with repeatability coefficients of 15%–20%.<sup>9–12</sup> The implication of this finding is that changes observed on PET scans obtained during the early portions of treatment that fall within these ranges should be interpreted with caution because they may only represent intrinsic fluctuation in metabolic activ-

ity. This concept is fundamental and must also be considered in the use of other non-FDG-PET isotopes and other functional imaging modalities, including CTP, dynamic contrast-enhanced MR imaging, and DWI.

Greater baseline variability in the primary tumor than in lymph nodes was an unexpected finding. The exact causes are unknown, but 2 explanations are possible: First, the primary tumor had overall higher SUV compared with nodes, and there may be poorer repeatability as the metabolic activity increases beyond an SUV threshold. This pattern was best appreciated on the Bland-Altman plots for primary tumor (Fig 1). A second factor could be tumor morphology. The infiltrative nature of primary tumors causes their boundaries with normal tissue to be more poorly defined than for lymph nodes, which are often surrounded by fat. Moreover, contouring the margins of metabolically active primary tumor without including surrounding inflamed or reactive mucosa is challenging. Smaller baseline variability in nodal disease compared with the primary head and neck tumor was also a finding in the study by Chu et al.<sup>13</sup> The repeatability was not quantified, but it was graphically displayed in plots showing the absolute change in SUV<sub>max</sub> with time in primary tumor and nodes for serial PET/CT scans.

The purpose of defining baseline variability was to better interpret intratreatment change. Early treatment change in SUV for HNSCC has not been reported, but Geets et al<sup>16</sup> did study FDG-PET intratreatment changes in metabolic gross tumor volume in HNSCC. This group reported a mean metabolic gross tumor volume reduction of 34% after 14 Gy (range, 2%–100%). We chose to evaluate change in SUV rather than gross tumor volume, and we separated changes in nodal SUV from those in the primary tumor. The rationale for using SUV is that it reflects the magnitude of tumor metabolic activity rather than size and adheres to PERCIST criteria.<sup>14</sup> PERCIST advocates viewing PET tumor response as a continuous SUV variable, and it defines a metabolic partial response as  $>30\%$  decrease in SUV after completion of therapy. PERCIST does not offer criteria for intratreatment change, however. The data from our study suggest that in the early stages of treatment using a relative change of  $>10\%$  in nodal SUV<sub>max</sub> and  $>16\%$  in nodal SUV<sub>mean</sub> in any direction could be indicative of a true increase or decrease in tumor glucose metabolism.

There are limitations to this study. First, it was a small single-institution pilot study, which limits the generalizability of the results. A larger trial will be required to validate the percentage cutoff criteria for true treatment-induced change for nodal SUV<sub>max</sub> and SUV<sub>mean</sub>. The results do, however, provide insight into the need to interpret small changes observed during the early phases of treatment with caution, given baseline variability. It is also obviously necessary to follow our study patients to determine the correlations between intratreatment changes and disease recurrence. However, very few patients may recur in our group since most had human papillomavirus positive oropharyngeal cancer.

Second, SUV<sub>max</sub> and SUV<sub>mean</sub> are summary parameters of tumor FDG uptake, while a malignant mass is heterogeneous in composition and metabolic activity. They represent “low-hanging fruit” and may not be the most robust parameters for charac-

terizing metabolic activity of a tumor. A segmented or voxelwise approach to evaluating functional imaging of tumor should also be considered, and such analyses will be a focus of our future effort. These alternative image-analysis methods may provide additional information about the 2 patients with low-metabolic-activity tumors at baseline who were excluded from this study. Additionally, there may be interobserver and intraobserver variability because SUV measurements were made by 1 radiologist. However, this is less likely to affect  $SUV_{max}$  and is minimized by deriving SUV from the VOI rather than from the region of interest.

Finally, the optimal time for assessing intratreatment metabolic response is unknown. This matter should be investigated in future studies. If imaging is performed too early, the effects of therapy may be small and partially masked by acute inflammatory changes. If imaging is performed during the latter phases of treatment, then the opportunity to modify therapy on the basis of an unfavorable response may be lost. This study suggests that the best timing would be when relative change is at least greater than the baseline variability.

## CONCLUSIONS

The intrinsic variability of HNSCC FDG-PET SUV is less than that in early treatment-induced change only in lymph node metastases. There is inherent variability in SUV measurements, which is greater for primary tumor than for nodes. Evaluation of early treatment response should be measured quantitatively in nodal disease rather than the primary tumor, and assessment of a positive response should account for the intrinsic baseline variability.

Disclosures: Jenny K. Hoang—RELATED: GE Radiology Research Academic Fellowship (GERRAF), Comments: GE provides funds the GERRAF grant. The grant is awarded by the Association of University Radiologists (AUR), Support for Travel to Meetings for the Study or Other Purposes: American Society of Radiation Oncology (ASTRO), UNRELATED: Payment for Lectures (including service on Speakers Bureaus): American College of Radiology. David M. Brizel—UNRELATED: Consultancy: Siemens Molecular Imaging, Comments: consultant to Siemens Molecular Imaging, which is pursuing a registration strategy for HX-4, which is an 18F PET hypoxia imaging agent, Employment: National Cancer Institute (NCI), Comments: Co-Chairman of the NCI Head and Neck Steering Committee, which works with the Cancer Therapy Evaluation Program and provides peer review of all phase II and III clinical trial concepts that emanate from the cooperative groups and, if approved, are then developed into clinical trials, Royalties: Up-to-Date Oncology, Comments: Section Co-Editor for *Head and Neck Cancer*, Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: ASTRO, American Radium Society (ARS), Comments: 1) ASTRO: travel expenses as invited faculty member for society-sponsored head and neck cancer educational programs in India and Brazil in 2012; 2) ARS: travel expenses for invited participation in debate on oropharynx cancer at 2012 annual meeting, Other: ARS, ASTRO, Comments: honoraria for invited presentations at annual meetings of these 2 societies in 2012.

## REFERENCES

- Higgins KA, Hoang JK, Roach MC, et al. Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: tumor SUVmean has superior prognostic value. *Int J Radiat Oncol Biol Phys* 2012;82:548–53
- Kapoor V, Fukui MB, McCook BM. Role of 18F-FDG PET/CT in the treatment of head and neck cancers: principles, technique, normal distribution, and initial staging. *AJR Am J Roentgenol* 2005;184:579–87
- Porceddu SV, Pryor DI, Burmeister E, et al. 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;33:1675–82
- Yoo DS, Kirkpatrick JP, Craciunescu O, et al. Prospective trial of synchronous bevacizumab, erlotinib, and concurrent chemoradiation in locally advanced head and neck cancer. *Clin Cancer Res* 2012;18:1404–14
- van Heijl M, Omluo JM, van Berge Henegouwen MI, et al. Fluorodeoxyglucose positron emission tomography for evaluating early response during neoadjuvant chemoradiotherapy in patients with potentially curable esophageal cancer. *Ann Surg* 2011;253:56–63
- Markova J, Kobe C, Skopalova M, et al. FDG-PET for assessment of early treatment response after four cycles of chemotherapy in patients with advanced-stage Hodgkin's lymphoma has a high negative predictive value. *Ann Oncol* 2009;20:1270–74
- Benz MR, Czernin J, Allen-Auerbach MS, et al. FDG-PET/CT imaging predicts histopathologic treatment responses after the initial cycle of neoadjuvant chemotherapy in high-grade soft-tissue sarcomas. *Clin Cancer Res* 2009;15:2856–63
- Hoekstra CJ, Stroobants SG, Smit EF, et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. *J Clin Oncol* 2005;23:8362–70
- Weber WA, Ziegler SI, Thodtmann R, et al. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med* 1999;40:1771–77
- Frings V, de Langen AJ, Smit EF, et al. Repeatability of metabolically active volume measurements with 18F-FDG and 18F-FLT PET in non-small cell lung cancer. *J Nucl Med* 2010;51:1870–77
- Velasquez LM, Boellaard R, Kolli G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med* 2009;50:1646–54
- Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by 18F-FDG PET in malignant tumors. *J Nucl Med* 2008;49:1804–08
- Chu KP, Murphy JD, La TH, et al. Prognostic value of metabolic tumor volume and velocity in predicting head-and-neck cancer outcomes. *Int J Radiat Oncol Biol Phys* 2012;83:1521–27
- 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
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60
- Geets X, Tomsej M, Lee JA, et al. Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: impact on target volume delineation and dose distribution using helical tomotherapy. *Radiother Oncol* 2007;85:105–15