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$_{\text{max}}$ and SUV$_{\text{mean}}$ 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$_{\text{max}}$ and SUV$_{\text{mean}}$, but not primary tumor SUVs. The mean baseline $%\Delta$SUV was lowest for SUV$_{\text{max}}$ in nodes (7.6% ± 5.2%) and highest for SUV$_{\text{max}}$ in primary tumor (12.6% ± 9.2%). Corresponding mean intratreatment $%\Delta$SUV$_{\text{max}}$ was 14.5% ± 21.6% for nodes and 15.2% ± 22.4% for primary tumor. The calculated RC for baseline nodal SUV$_{\text{max}}$ and SUV$_{\text{mean}}$ 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. Posttreatment FDG-PET has become an important tool for the assessment of residual disease in cervical lymph nodes. 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.

Early response assessment with FDG-PET has been evaluated in lymphoma, soft-tissue sarcoma, and esophageal and lung cancers. 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. One of the largest studies of neo-
adjuvant chemotherapy for esophageal cancer identified responders with high sensitivity by using 0% SUV decrease as a cutoff (ie, any decrease in SUV), 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.

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 max on pretreatment PET/CT scans that were performed on different scanners. The authors warned about the need to account for variability in PET biomarkers in clinical protocols. Wahl et al, who proposed criteria for the Positron Emission Response Criteria in Solid Tumor (PERCIST) trial, also included patients with both tumor and node SUV max of mellitus. To avoid data contamination by image noise, we prospectively enrolled between September 2009 and August 2011. Cancer stages III-IV head and neck squamous cell carcinoma of any treatment and node SUV max 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 enhanced CT from the midcranium to the thoracic inlet was performed with the arms down (3.75-mm-thick contiguous images with 30-cm FOV).

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 enhanced 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 × 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:

\[ \text{SUV} = \frac{c_d}{d_{j/w}}, \]

where \( c_d \) 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 max and SUV mean. These paramet-
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\textsubscript{max} was defined as the highest pixel value in the VOI. SUV\textsubscript{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\textsubscript{1}, SUV\textsubscript{2}, and SUV\textsubscript{3} are the respective measurements of SUV on PET\textsubscript{1}, PET\textsubscript{2}, and PET\textsubscript{3}, then the formulas for SUV unit 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\textsubscript{1} and SUV\textsubscript{2} was also compared with the ICC for SUV\textsubscript{2} and SUV\textsubscript{3}. A repeatability coefficient was also derived from the baseline %ΔSUV and was compared with the intratreatment %ΔSUV. The RC has been applied to measure intrinsic baseline variability\textsuperscript{10} 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 %ΔSUV 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\textsubscript{max} in both the primary tumor and node at baseline was ≤4. The 17 remaining patients were all men with a mean age of 51 ± 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 is displayed as Bland-Altman plots in Fig 1. There was excellent agreement between the 2 baseline studies for nodes measured as SUV\textsubscript{max} and SUV\textsubscript{mean}. The mean difference between baseline nodal SUV was 0.58 ± 0.90 for SUV\textsubscript{max} and 0.33 ± 0.71 for SUV\textsubscript{mean}. Figure 1A shows that all points for nodal SUV\textsubscript{max} were within the mean ± 1.96 × SD. The solid black line in Fig 1 is the
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$_{\text{max}}$ and SUV$_{\text{mean}}$ in primary tumor and nodes (Table 1), but it was lowest for primary tumor SUV$_{\text{mean}}$ (0.91) and highest for nodal SUV$_{\text{max}}$ (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-
marginal for SUVmean (compared with after 12 Gy for patient 7 (group median, 12 Gy).

SUVmax and SUVmean (patient 14) did not have a rise in the SUVmax or SUVmean, but his completion of CRT. The other patient who was a nonresponder patient 7, who was one of the patients with residual disease after nodal SUVmax and SUVmean above the respective RCs. This was intrinsically baseline variability rather than true intratreatment change. Only 1 patient had an intratreatment increase in both nodal SUVmax and SUVmean 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 SUVmax or SUVmean, 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 SUVmax and SUVmean (P < .002). For the primary tumor, the effect of treatment was significant for SUVmax (P = .006) and marginal for SUVmean (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. in a retrospective study using diagnostic PET/CT and planning PET/CT. They reported serial change as mean composite SUV velocity (SUVmax change divided by time in weeks between scans). Factors that contributed to a mean composite SUV velocity of −0.1/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 %ΔSUV ranging from 3% to 16%, with repeatability coefficients of 15%–20%. 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 activity. 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. The repeatability was not quantified, but it was graphically displayed in plots showing the absolute change in SUVmax 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. 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. 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 SUVmax and >16% in nodal SUVmean 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 SUVmax and SUVmean. 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, SUVmax and SUVmean 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, 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
1. 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
2. Kapoor V, Fukui MB, McCook BM. Role of 18FFDG PET/CT in the treatment of head and neck cancers: principles, technique, normal distribution, and initial staging. AIR Am J Roentgenol 2005;184:579–87
3. 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
4. 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
5. van Heijl M, Omloo JM, van Berge Henegouwen MI, et al. Fluorodeoxoglucose positron emission tomography for evaluating early response during neoadjuvant chemoradiotherapy in patients with potentially curable esophageal cancer. Ann Surg 2011;253:56–63
6. 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
7. 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 sarcoma. Clin Cancer Res 2009;15:2856–63
8. 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
9. Weber WA, Ziegler SI, Thodmann R, et al. Reproducibility of metabolic measurements in malignant tumors using FDG PET. J Nucl Med 1999;40:1771–77
10. 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
11. Velasquez LM, Boellaard R, Kollia G, et al. Reproducibility of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. J Nucl Med 2009;50:1646–54
12. 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
13. 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
14. 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):i22S–50S
15. Bland JM, Altman DG. Measuring agreement in method comparison studies, Stat Methods Med Res 1999;8:135–60
16. Geets X, Tomsej M, Lee JA, et al. Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngeo-laryngeal tumors: impact on target volume delineation and dose distribution using helical tomotherapy, Radiother Oncol 2007;85:105–15