Evaluation of Time-Phase Effect on $^{18}$F-FDG PET/CT Delineation Methods for Treatment Planning of Nasopharyngeal Carcinoma

Yang-Zong Chen, MS,†‡ Wen-Feng Li, MD,§ Jing-Ying Wang, MS, Jin-Meng Wang, MS,†‡ Rong-Ying Ou, MD, † Xiang-Wu Zheng, MD,†‡ Yun-Sheng Xu, MD,§ and Liang Zhao, MS†‡

**Purpose:** Tumor boundary delineation using $^{18}$F-FDG PET/CT is a promising tool for radiotherapy applications, but no consensus has been established regarding the optimal delineation method. Time-phase variability of $^{18}$F-FDG PET/CT imaging frequently affects metabolically active volumes and treatment planning for nasopharyngeal carcinoma (NPC). This study aimed to evaluate the time-phase robustness of 8 methods commonly used for tumor volume delineation in NPC.

**Patients and Methods:** Twenty patients with biopsy-proven NPC were included and underwent multiple time-phase $^{18}$F-FDG PET/CT imaging. Gross tumor volumes (GTVs), absolute SUV, gradient-based watershed segmentation (GTV-GWT), and anatomic biologic contouring (GTV-ABC) values were determined. The volume of overlap between GTV-CT and the 8 PET-based GTVs was enclosed and the overlap fraction (OF-CT) calculated. Color matrix was used to semiquantify the time-phase differences. Gross tumor volume values obtained with different methods were recorded and compared using paired $t$ test. Time-phase differences of GTVs and SUV\textsubscript{max} were compared among groups by analysis of variance with Tukey honest significance tests. The coefficients of variation were computed to assess intrapatient time-phase variability. Similarity coefficient was calculated to evaluate similarity.

**Results:** Differences were observed between GTVs obtained at different time points using various delineation procedures. Nonsignificantly higher percentages were obtained for GTV-GWT (88.17%) and GTV-ABC (86.98%) compared with other methods, showing their robustness. GTV-40% (0.81–0.88) and GTV-ABC (0.82–0.88) indicated higher similarity with GTV-MRI than the other methods.

**Conclusions:** PET/CT-based GTV-ABC between 35 and 55 minutes should be the first choice for NPC treatment planning.

**Key Words:** $^{18}$F-FDG PET, PET/CT imaging, GTV, segmentation, nasopharyngeal carcinoma, time-phase variation

**ORIGINAL ARTICLE**

Received for publication September 21, 2015; revision accepted December 22, 2015.

From the †Laboratory for Advanced Interdisciplinary Research, Institutes of Translational Medicine, ‡Institute of Intelligent and Molecular Imaging, §Division of PET/CT, Department of Radiology, and †Division of Radiation Oncology, the First Affiliated Hospital, Wenzhou Medical University, Wenzhou, China.

Yang-Zong Chen and Wen-Feng Li contributed equally to this study.

Conflicts of interest and sources of funding: Supported by the Science Technology Department of Zhejiang Province of China (2013C33171) and a program of Wenzhou Science and Technology Bureau (Y20130235).

Correspondence to: Liang Zhao, MS, Laboratory for Advanced Interdisciplinary Research, Institutes of Translational Medicine, the First Affiliated Hospital, Wenzhou Medical University, Wenzhou 325000, China. E-mail: Zhao45671@hotmail.com; Yun-Sheng Xu, MD, Laboratory for Advanced Interdisciplinary Research, Institutes of Translational Medicine, the First Affiliated Hospital, Wenzhou Medical University, Wenzhou 325000, China. E-mail: 1435393932@qq.com.

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.nuclearmed.com).

Copyright © 2016 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0363-9762/16/0000-0000

DOI: 10.1097/RLU.0000000000001161

INTRODUCTION

Intensity-modulated radiation therapy (IMRT) and other modern radiotherapy (RT) techniques rely heavily on high-quality medical imaging. Indeed, imaging-based target volume delineation is critical in treatment planning and impacts clinical outcomes. In head and neck cancers, IMRT increases the dose delivered to the primary tumor while reducing that reaching normal tissues. Therefore, identifying the locations and boundaries of the primary tumor and cervical lymph node metastases is of prime importance in IMRT.

$^{18}$F-FDG PET with the glucose analog $^{18}$F-FDG provides biological information complementary to the anatomical data obtained by CT or MRI. Indeed, PET/CT has been widely proposed for the detection and staging of head and neck cancers, and tumor boundary delineation using $^{18}$F-FDG PET is a promising tool for RT applications. However, $^{18}$F-FDG PET has drawbacks such as limited spatial resolution, the lack of a standardized method for signal segmentation, and false-positive readings caused by inflammation. Furthermore, relatively high costs and undefined clinical outcomes have made clinicians wary of this option, causing limited clinical applicability.

The delineation of gross tumor volumes (GTVs) using PET/CT for head and neck cancers varies considerably partly because of the disagreement among institutions in defining the threshold of malignant disease based on physiologic imaging. The GTV derived from $^{18}$F-FDG PET is highly dependent upon the method employed for PET signal segmentation, with distinct volumes obtained with different thresholds. Although multiple studies have compared GTVs reported with currently available methods, no individual delineation method is favored unanimously by institutions, either nationally or internationally.

Studies have shown that $^{18}$F-FDG uptake continues to increase in malignant tumors for several hours after injection, in agreement with our preliminary data in patients with nasopharyngeal carcinoma (NPC). Such time-phase variations could result in inconsistencies with RT planning. The aim of this study was to evaluate the time-phase robustness of 8 methods commonly used for tumor volume delineation in NPC. Interestingly, these 8 methods were not always robust enough for treatment planning. Taking time-phase fluctuation into account, GTV-anatomic biologic contouring (GTV-ABC) and GTV-40% between 40 and 60 minutes after injection had higher similarity with GTV-MRI compared with the other methods. In addition, PET/CT-based GTV-ABC between 35 and 55 minutes turned out to be the first choice for NPC treatment planning.

**PATIENTS AND METHODS**

**Patients**

Twenty patients were recruited between November 2011 and July 2013 from our hospital. Fourteen of them were male and 6 were female; the mean (SD) age was 58.3 (10.7) years (range, 35–77 years). The primary tumor site was the nasal cavity. The disease stages comprised T2 to T4, N0 to N2, and M0. All the patients were imaged 72 hours before treatment. Inclusion criteria...
were (1) biopsy-proven NPCs and the (2) ability to participate cooperatively in the procedures. Exclusion criteria were (1) low $^{18}$F-FDG uptake in target volume and the (2) inability to participate cooperatively in the procedures. All 20 patients were eligible and finally included in the study.

All the evaluated patients were systematically staged using direct pharyngoscopy and tissue biopsy diagnosis with plain chest radiography, serum chemistry and liver function parameters, contrast-enhanced CT, and MRI scans of the head and neck. The seventh edition of the Union for International Cancer Control and American Joint Committee on Cancer TNM Staging System was used for staging. The study was approved by the ethics committee of our hospital, and all patients provided written informed consent.

Detailed patient characteristics are shown in Table 1.

### PET/CT Image Acquisition Protocol

Patients were required to have a normal blood glucose level (<120 mg/dL) and fast for 6 hours before FDG PET/CT imaging. Head and neck image acquisition was initiated 30 minutes after injection of a weight-adjusted standard dose of $^{18}$F-FDG (3.7 MBq/kg) using a full-ring LYSO-based hybrid PET/CT scanner (Philips Gemini 64 TF; Philips Medical Systems, Cleveland, Ohio). Low-dose CT was acquired from the basis cranii through the clavicle in an imaging procedure using 120 kVp, 80 mA, a $16 \times 1.25$-mm detector collimation, a beam pitch of 1.6, a section thickness of 3.75 mm, and a reconstruction pitch of 5 mm (to match the PET section thickness). After the CT procedure, 3-dimensional PET data were acquired using a 5-minute loop, that is, 2 and 3 minutes for scanning and acquisition initialization, respectively. In total, 13 scanning loops were obtained in the same region (Fig. 1), with the whole scanning lasting 65 minutes. All scans were performed with the patient in supine position and immobilized with an individual head support and a rigid customized mask covering the head and neck area to reduce movement artifacts during image acquisition. A custom-designed face mask (Openface; Klarity Medical Products, Newark, Ohio) attached to a head-and-neck cradle was used for patient immobilization. General settings of PET/CT followed the protocol used in our center.

### MRI and CT Acquisition

Transverse MRI scans were acquired using a 3-T superconducting system (SIGHA HDX; GE Medical System, Milwaukee, Wis). The U-shaped elastic pillow was used for patient immobilization during the long scanning periods (PET/CT and MRI). CT sessions were performed with a dual-detector spiral CT scanner (Brilliance CT Big Bore; Philips Medical Systems, Best, the Netherlands). All scans were carried out from the vertex to 5 cm below the clavicle in the treatment position after Omnipaque contrast administration. A custom-designed face mask (Openface; Klarity Medical Products, Newark, Ohio) attached to a head-and-neck cradle was used for patient immobilization. The MRI and CT protocols used for NPC’s IMRT planning in our hospital were previously described.

### Delineation of GTV

Eight GTV delineation methods in 4 segmentation strategies were used in this study: (1) threshold method, with a fixed threshold range of 40% to 70% of the maximum voxel value within the tumor (GTV-40%, GTV-50%, GTV-60%, GTV-70%); (2) absolute SUV (GTV-SUV2.5 and GTV-SUV3); (3) gradient-based watershed segmentation method (GWT) and (4) anatomic biologic contour (ABC).
For reference, an experienced radiologist and a radiation oncologist contoured the GTV from the CT and MRI data sets. The GTV-CT volumes included anatomically abnormal and contrast-enhancing regions. GTV-MRI volumes were initially determined from the T2-weighted images, but comparisons with precontrast and postcontrast T1-weighted images were performed when necessary. The criteria for malignancy on T2-weighted images were unilateral change in anatomy compared with the normal contralateral side, mass effect, nodular or infiltrative abnormal tissue, fat replacement, and hypointensity. All GTV delineations were blinded to results of other experiments.

Overlaps (or mismatches) in the volumes derived from different GTV methods were evaluated using an image fusion program (EBW, Philips Medical Systems) based on anatomic landmarks. Volumes of the overlaps between GTV-CT (or GTV-MRI) and the 8 PET-based GTVs were enclosed, respectively, and the overlap fraction (OF-CT or OF-MRI) was calculated as the overlap GTV-PET volume relative to the GTV-CT (or GTV-MRI). Similarity coefficients \(^{16,37}\) were calculated to measure the similarity. GTV-MRI is considered the “silver standard,” and PET/CT GTVs with an OF-MRI larger than 95% and an MIS-MRI smaller than 5% considered to be the “acceptable PET/CT GTVs” for specific patients. The similarity coefficient of acceptable PET/CT GTVs should be greater than 0.95.

**Statistical Analysis**

A color matrix was used to semiquantify the time-phase difference. A unit labeled by a nonsignificant difference resulted in a score of 1; otherwise, the score was 0. A total score was calculated for each GTV method. Nonsignificant difference ratio was defined as the ratio of the total score to total number of units.

Gross tumor volume values were recorded for each method. The absolute and relative differences between the methods' planned treatment volumes were computed. Mann-Whitney \(U\) tests were used for GTV comparison. Tukey honest significance tests were used for analysis of variance with color matrix. Coefficients of
variation were computed to measure intrapatient time-phase variability. For numeric parameters, data are presented as data mean (SD). $P < 0.05$ was considered statistically significant. All statistical analyses were carried out using SPSS 19.0.

**RESULTS**

**Time-Phase Difference of FDG Uptake in NPC**

The FDG uptake distribution was continuously changing (Fig. 2), with SUV$_{\text{max}}$ increasing and the activity boundary changing irregularly. The mean (SD) SUV$_{\text{max}}$ in the primary region of interest changed from 7.20 (4.07) to 11.04 (5.58) during the period of 30 to 90 minutes after injection (Supplementary Figure 1, http://links.lww.com/CNM/A34). The peak value of SUV$_{\text{max}}$ appeared at different time points in various patients: 80, 85, and 90 minutes in 3, 11, and 6 patients, respectively (Table 1).

Figure 3 displays the mean volumes obtained at different time points using the various delineation procedures. GTV-CT and GTV-GWT yielded similar volumes between 30 and 80 minutes after injection. Meanwhile, GTV-CT and GTV-2.5 showed similar volumes between 40 and 90 minutes. The other 7 PET-based GTVs (GTV-40%, GTV-50%, GTV-60%, GTV-70%, GTV-SUV2.5, GTV-SUV3, and GTV-ABC) all yielded values smaller compared with that of GTV-CT (all $P < 0.001$). On the other hand, GTV-MRI and GTV-40% yielded similar volumes between 30 and 60 minutes, whereas similar volumes were obtained for GTV-MRI and GTV-ABC within the same time frame. Each of the percentage threshold methods (GTV-40%, $P = 0.01$; GTV-50%, $P = 0.001$; GTV-60%, $P = 0.001$; and GTV-70%, $P = 0.0005$) showed significantly smaller values compared with GTV-SUV2.5, indicating that these methods resulted in GTVs with significantly different sizes. Finally, 3 percentage threshold methods (GTV-50%, $P < 0.0001$; GTV-60%, $P < 0.0001$; and GTV-70%, $P < 0.0001$) had smaller values than the GTV-ABC method, indicating that they resulted in GTVs with significantly different sizes. There was no significant difference between GTV-40% and GTV-ABC ($P = 0.72$). For each of the percentage threshold methods (GTV-40%, GTV-50%, GTV-60%, and GTV-70%), a clear trend was observed, with GTVs decreasing from 30 to 90 minutes after injection. The opposite trend was observed for the absolute SUV methods (GTV-SUV2.5 and GTV-SUV3), where an overtly increasing trend was observed for GTVs from 30 to 90 minutes. A rapid decrease in GTV-ABC value was observed before 50 minutes, after which the volume stabilized. A rapid decrease was observed in GTV-GWT after 80 minutes, but the volume was stable beforehand (Fig. 3).

**Robustness of GTV Delineation Methods**

To assess time-phase differences in a semiquantitative manner and evaluate the robustness of various methods, GTV results were compared by analysis of variance with Tukey honest significance tests, and a color matrix was filled (Fig. 4). Nonsignificantly different percentages for GTV-GWT (88.17%, 149/169) and GTV-ABC (86.98%, 147/169) were obtained, and these values were higher than those recorded for the other methods (GTV-40%, 60.95%, 103/169; GTV-50%, 78.70%, 133/169; GTV-60%, 83.43%, 141/169; GTV-70%, 81.07%, 137/169; GTV-SUV2.5, 57.40%, 97/169; and GTV-SUV3, 51.48%, 87/169).

To further evaluate the effects of time-phase variations, coefficients of variation were computed as a measure of the intrapatient time-phase variability (Table 2; Supplementary Figure 2, http://links.lww.com/CNM/A35). Except for GTV-60% and GTV-70%, all mean (SD) coefficients of variation were less than 10% (GTV-40%, 8.00% [4.38%]; GTV-50%, 7.20% [5.25%]; GTV-SUV2.5, 6.85% [3.93%]; GTV-SUV3, 6.08% [4.31%]; GTV-GWT,
7.41% [3.09%]; GTV-ABC, 8.14% [6.46%]). There was no significant difference among these 6 methods (all P > 0.2).

**Similarity Analysis of PET/CT GTVs Versus GTV-CT and GTV-MRI**

Figure 5 shows the similarity coefficient indexes of PET/CT GTVs and GTV-MRI. The average similarity coefficient of GTV-40% was always greater than 0.8 (range, 0.81–0.88); idem for GTV-ABC (0.82–0.88). However, the similarity coefficient of GTV-50% was greater than 0.8 only at the very beginning (30–40 minutes), and decreased with time (range, 0.72–0.82). The other 5 PET/CT GTVs had similarity coefficient all below 0.8: GTV-60% ranged from 0.58 to 0.71, GTV-70% from 0.40 to 0.50, GTV-2.5 from 0.60 to 0.72, GTV-3 from 0.51 to 0.75, and GTV-GWT from 0.55 to 0.57.


Figure 6 depicts the similarity coefficient of PET/CT GTVs and GTV-CT. The average similarity coefficient of all PET/CT GTVs were less than 0.7: GTV-40% ranged from 0.46 to 0.57, GTV-50% from 0.33 to 0.44, GTV-60% from 0.23 to 0.31, GTV-70% from 0.13 to 0.20, GTV-2.5 from 0.49 to 0.70, GTV-3 from 0.41 to 0.60, GTV-GWT from 0.65 to 0.71, and GTV-ABC from 0.42 to 0.50. Data obtained for OF-MRI, OF-CT, MIS-MRI, and MIS-CT are shown in Supplementary Figures 3 to 6, http://links.lww.com/CNM/A36, http://links.lww.com/CNM/A37, http://links.lww.com/CNM/A38, http://links.lww.com/CNM/A39.
Assuming PET-based GTVs with OF-MRIs with values greater than 95% and MIS-MRIs less than 5% could be considered acceptable in specific patients; various segmentation methods offered quite different performances. Figure 7 shows the count-time distribution of acceptable points. GTV-ABC covered 57% (8/14) patients with acceptable similarity between 35 and 55 minutes. GTV-40% covered 57% (8/14) patients with acceptable similarity between 50 and 55 minutes. The other GTVs or on the other time points covered less patients with acceptable similarity.

DISCUSSION

Most of the proposed work dealing with GTV determination in PET relies upon thresholding, which is either based on a priori CT knowledge or uses a fixed threshold derived from phantom studies. Yet, Scarfone et al suggested that the threshold for PET images needs to be adjusted on a case-by-case basis to adequately depict FDG-avid disease relative to background. In this study, we have highlighted the time-phase variability of $^{18}$F-FDG PET/CT imaging in patients with NPC. When using a threshold method for GTV delineation, the estimated GTV volume will depend on the threshold level selected. Indeed, phantom and patients data presented by Ford et al revealed a high sensitivity dependence between these parameters. $^{18}$F-FDG PET/CT imaging is often acquired 1 hour after injection. However, no evidence supports this choice as the optimal time for treatment planning. As shown in Figure 7, GTV-ABC between 35 and 55 minutes after $^{18}$F-FDG injection might be the best option, assuming MRI to be a suitable silver standard and GTV-70% between 35 and 70 minutes after $^{18}$F-FDG injection the secondary options. To our knowledge, these findings are quite distinct from current clinical practices related to $^{18}$F-FDG PET/CT scanning, and no previous study has addressed this point in NPC.

In addition, there is no consensual criterion for estimating time-phase variation in the GTV. It is unclear whether such variation would ultimately affect clinical outcome. As shown previously,
GTG-GWT and GTV-ABC might reduce the risk of time-phase variation in treatment planning. Indeed, the threshold and absolute SUV methods would be sensitive to SUV changes. As shown in Figure 2, SUV changed minute by minute, and would result in time-phase variation. Since the GWT and ABC methods mainly depend on signal contrast and are less affected by SUV changes, they were more robust than the others.

Although GTV-GWT is robust enough in itself, it yielded much larger volumes compared with GTV-MRI, but still slightly smaller than GTV-CT. MRI in head and neck cancer treatment planning is considered an excellent tool in current clinical practice. Therefore, GTV-GWT was employed with abundant caution. A comprehensive performance comparison of GTV-GWT and GTV-MRI should be further studied.

It was difficult to obtain complete pathological NPC specimens, and we did not correlate GTVs with pathological characteristics. In addition, correlation of pathological specimens with imaging is problematic because of possible contraction during tissue fixation. MRI is currently recommended as the imaging modality for NPC because of superb soft tissue contrast and acceptable clinical outcomes. Therefore, we employed GTV-MRI as the silver standard in this study. However, it remains unclear whether MRI is superior to $^{18}$F-FDG PET/CT in NPC treatment planning. Based on the findings described previously, GTV-ABC was robust enough and showed high similarity with GTV-MRI. Because of reduced robustness or similarity with GTV-MRI, the use of GTV-40%, GTV-50%, GTV-2.5, GTV-3, and GTV-GWT remains debatable. In addition, both robustness and similarity with GTV-MRI were poor for GTV-60% and GTV-70%, which should not be used for NPC treatment planning.

Before reaching a consensus on PET-based treatment planning in RT, methods should be developed to ensure that proper parameter settings are maintained with sufficient robustness. Otherwise, the best acquisition time point should be determined from sufficient clinical evidence.

In summary, we recommend GTV-ABC between 35 and 55 minutes after injection for NPC treatment planning as the first choice.

**REFERENCES**


