

# Jacobs Journal of Radiation Oncology

## Research Article

### Interobserver and Intermodality Variability in GTV Delineation on Simulation CT, FDG- PET, and MR Images of Head and Neck Cancer

Carryn M. Anderson<sup>1</sup>, Wenqing Sun<sup>1</sup>, John M. Buatti<sup>1</sup>, Joan E. Maley<sup>2</sup>, Bruno Policeni<sup>2</sup>, Sarah L. Mott<sup>3</sup>, John E. Bayouth<sup>4</sup>

<sup>1</sup>Departments of Radiation Oncology, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA

<sup>2</sup>Radiology, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA

<sup>3</sup>Biostatistics, Holden Comprehensive Cancer Center, University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA

<sup>4</sup>Department of Human Oncology, University of Wisconsin, Madison, Wisconsin, USA

\*Corresponding author: Dr. Carryn M. Anderson, University of Iowa Hospitals and Clinics, 200 Hawkins Dr, Iowa City, IA 52242, 01614 PFPW, USA, Tel: 319-353-8836; Fax: 319-356-1530; E-mail: carryn-anderson@uiowa.edu

Received: 07-29-2014

Accepted: 09-05-2014

Published: 09-08-2014

Copyright: © 2014 Anderson

## Abstract

### Purpose

To compare the interobserver and intermodality differences in image-based identification of head and neck primary site gross tumor volumes (GTV). Modalities compared include: contrast-enhanced CT, F-18 fluorodeoxyglucose positron emission tomography (PET/CT) and contrast-enhanced MRI.

### Methods and Materials

Fourteen patients were simulated after immobilization for all 3 imaging modalities (CT, PET/CT, MRI). Three radiation oncologists (RO) contoured GTVs as seen on each modality. The GTV was contoured first on the contrast-enhanced CT (considered the standard), then on PET/CT, and finally on post-contrast T1 MRI. Interobserver and intermodality variability were analyzed by volume, intersection, union, and volume overlap ratio (VOR).

### Results

Analysis of RO contours revealed the average volume for CT-, PET/CT-, and MRI-derived GTVs were 45cc, 35cc and 49cc, respectively. In 93% of cases PET/CT-derived GTVs had the smallest volume and in 57% of cases MRI-derived GTVs had the largest volume. CT showed the largest variation in target definition (standard deviation amongst observers 35%) compared to PET/CT (28%) and MRI (27%). The VOR was largest (indicating greatest interobserver agreement) in PET/CT (46%), followed by MRI (36%), followed by CT (34%). For each observer, the least agreement in GTV definition occurred between MRI & PET/CT (average VOR = 41%), compared to CT & PET/CT (48%) and CT & MRI (47%).

### Conclusions

A non-significant interobserver difference in GTVs for each modality was seen. Among three modalities, CT was least consistent, while PET/CT-derived GTVs had the smallest volumes and were most consistent. MRI combined with PET/CT provided the least agreement in GTVs generated. The significance of these differences for head & neck cancer is important to explore as we move to volume-based treatment planning based on multi-modality imaging as a standard method for treatment delivery.

## Introduction

Volume-based intensity-modulated radiotherapy (IMRT) is the preferred radiation treatment technique for head and neck cancer. Target volume definition remains the most subjective and hence least consistent variable in the delivery of accurate radiotherapy. Gross tumor volume definition is influenced by the judgment of the individual doing the contouring, the type of imaging utilized, the resolution and slice thickness of the scan, contrast administration, registration of multiple imaging modalities, and patient positioning in co-registered scans, as well as other factors [1]. Planning tumor volumes have been most commonly defined by CT-based imaging. More recently, PET/CT and MR images have further refined our ability to define tumor extent. Treatment planning software has evolved to allow registration of these imaging data with the simulation CT. Review of the head and neck cancer literature shows substantial interest in enhancement of tumor definition with the addition of PET/CT to CT and MRI to CT, however intermodality comparison of all three modalities is scarce [2,3]. Our institution has both PET/CT and MRI simulators. Since 2007, we have utilized all three imaging modalities (contrast-enhanced CT, fluorodeoxyglucose (FDG) positron emission tomography (PET/CT) and contrast-enhanced MRI) to simulate selected patients for treatment planning. By imaging the patient in the head-and-shoulder aquaplast mask on all three modalities, registration error in the treatment planning software is minimized. To evaluate the utility and consistency of these imaging modalities in defining the primary site gross tumor volume (GTV), we compared tumor volumes contoured by three observers on CT, PET/CT, and MRI.

## Material and Methods

### Patient selection

Department records were reviewed and fourteen patients with advanced head and neck cancer who had undergone simulation with contrast-enhanced CT, FDG-PET/CT, and contrast-enhanced MRI scans were selected as a case study (see Table 1).

**Table 1. Patient Characteristics**

| Patient | Primary Site    | Clinical Stage | Clinical Notes           |
|---------|-----------------|----------------|--------------------------|
| 1       | oropharynx      | T3N2b          |                          |
| 2       | oropharynx      | T3N1           |                          |
| 3       | larynx          | T4aN0          |                          |
| 4       | oropharynx      | T1N2           |                          |
| 5       | oropharynx      | T2N2a          | post tonsillectomy       |
| 6       | oropharynx      | T2N2c          |                          |
| 7       | nasopharynx     | T2bN2          |                          |
| 8       | paranasal sinus | T4bN0M1        | neuroendocrine carcinoma |
| 9       | hard palate     | recurrent      | spindle cell sarcoma     |
| 10      | oropharynx      | recurrent      |                          |
| 11      | oropharynx      | T1N3           |                          |
| 12      | paranasal sinus | T4N1           |                          |
| 13      | oropharynx      | T3N2b          | post tonsillectomy       |
| 14      | oropharynx      | T4N2b          |                          |

## Imaging protocol and registration

Patients were immobilized with a head-and-shoulders S-Frame Aquaplast mask (CIVCO Medical Solutions, Kalona, IA) and simulated in the treatment position for all modalities. PET/CT simulation was performed on the Siemens LSO Biograph Duo PET/CT scanner (Siemens Medical Systems, Hoffman Estates, IL). In a single imaging session, an IV contrast-enhanced CT scan and a PET scan were acquired.

For the PET scan, the patient's fasting glucose was required to be 200 mg/dL or less. To decrease patient movement and agitation during the uptake period, 0.25–2.5mg of alprazolam was given by mouth, as our standard protocol. Imaging was completed 90min after injection of 10–15 mCi of FDG.

CT imaging was obtained from the vertex to 2cm below the carina. Isovue 250, 100 cc, was given by IV injection if kidney function was normal. If creatinine was elevated or glomerular filtration rate was decreased, the amount of Isovue was decreased or the contrast agent was changed to Visapaque. Slice thickness and spacing were 2 mm throughout imaging.

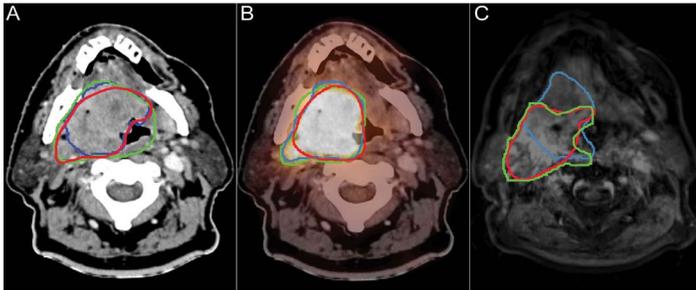
3.0-Tesla MR images were generated using a Siemens MAGNETOM Trio 3T MRI scanner (Siemens Medical Systems, Erlangen, Germany) with the use of head coil (3T Body MATRIX A TIM Coil, Siemens, Erlangen, Germany). Multihance (0.2 cc/kg body weight, max 20 cc) was injected IV prior to T1 contrast imaging.

The images were uploaded and PET and MRI were registered (fused) to the primary CT data set by an auto registration algorithm in Pinnacle3 version 8.0m (Philips, Fitchburg, WI) called "normalized mutual information." Physicians confirmed registration adequacy and all contouring were completed in the Pinnacle treatment planning software.

### Contouring

Using the planning system, three observers with different experience levels (one third-year radiation oncology resident and two head and neck radiation oncologists) contoured primary site GTVs on each modality. When distinguishable, adjacent abnormal lymph nodes were excluded from the tumor volume. All observers were informed of location of primary site and clinical stage. The observer was instructed to contour only what was seen as abnormal on that specific modality. Observers were allowed to adjust window width and level on CT to optimize soft tissue and bone GTV delineation. PET/CT-based GTVs were drawn on the default window width and level settings without the assistance of an SUV threshold or tumor-to-background ratio. When contouring on each image set, the contours were turned off after completing the task on one image set before beginning contours on the next modality

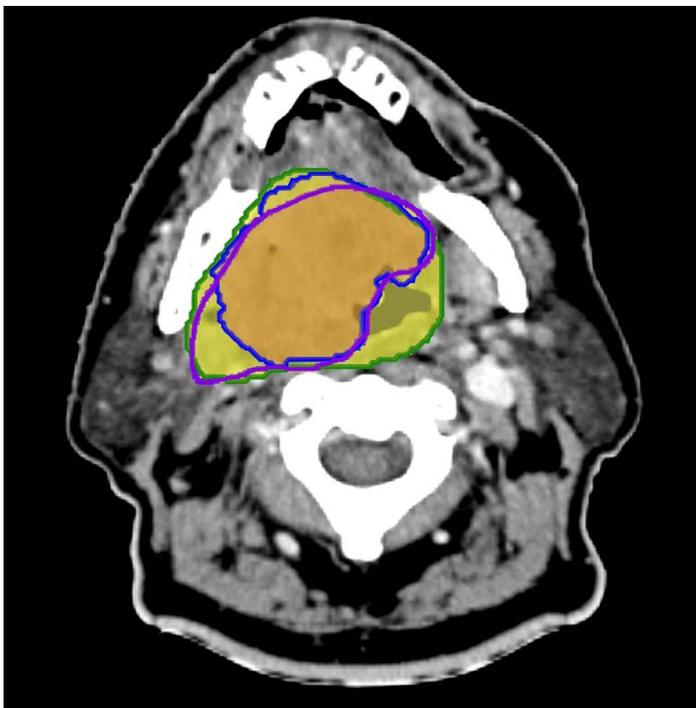
(to simulate a scenario in which we assume that the image set would be the only image set available for radiation treatment planning). Observers were blinded to the GTVs outlined by the other observers. In all cases, the GTV was contoured first on the contrast-enhanced CT, then on the PET/CT, and finally on the post-contrast T1 MRI (Figure 1).



**Figure 1.** Axial slice of patient #14 (T4 oropharynx cancer). Each observer contoured the primary tumor (GTV) as uniquely seen on the simulation CT with contrast (A), PET/CT (B), and MRI (C). Contours from the 3 radiation oncologists are shown in different colors.

### Data analysis

In this study, interobserver and intermodality variability were analyzed by volume, union, intersection, and volume overlap ratio (VOR, intersection divided by union) (Figure 2). The data analysis for this paper was generated using SAS software, Version 9.3(Cary, NC).



**Figure 2.** The inter-observer union (yellow colorwash) is defined as the volume encompassing the contours of all observers in an imaging modality. The inter-observer intersection (orange colorwash) is defined as the volume common to all observers in one imaging modality. The volume overlap ratio (VOR) is the ratio of the intersection to the union.

The inter-observer average volume was defined as the mean of all volumes outlined in a scan set by all observers for that modality. The inter-observer union volume was defined as the volume encompassing the GTVs delineated by all observers in a dataset, i.e. the union volume was the total GTV volume outlined by every observer. The inter-observer intersection volume was defined as the common volume designated by all observers as part of the GTV in a given modality dataset. The inter-observer VOR indicated the uncertainty in delineating the GTV in that scan set by different observers and was calculated as the intersection divided by the union.

The inter-modality average volume was defined as the mean GTV volume designated by a specific observer in all three modalities. The inter-modality union volume was defined as the volume encompassing the GTVs delineated by an observer in all datasets; i.e., the volume designated as the combination of all GTVs contoured by a specific observer in all three imaging modalities. The inter-modality intersection volume was the common volume designated by a specific observer as part of the GTV in all three imaging modalities. The inter-modality VOR indicated the uncertainty in delineating the GTV in all imaging modalities by a specific observer and was calculated as the intersection divided by the union.

A linear regression model was employed to compare the interobserver variability for each imaging modality. The coefficient of variation (defined as the percent standard deviation of volume) was utilized as the dependent variable. The main effect for imaging modality was included as the independent variable. A linear mixed effects regression model was used to compare the intermodality differences. To assess intermodality differences, the VOR was utilized as the dependent variable. The main effect for between modality agreements was included as the independent variable. To control for the potential confounding of observer differences, the observer was included as a random effect. All tests were two sided and tested at the 5% significance level.

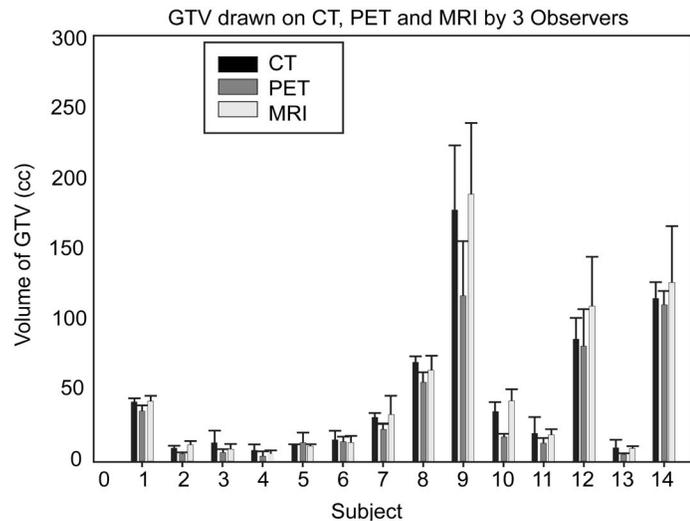
## Results

### Patient and tumor characteristics

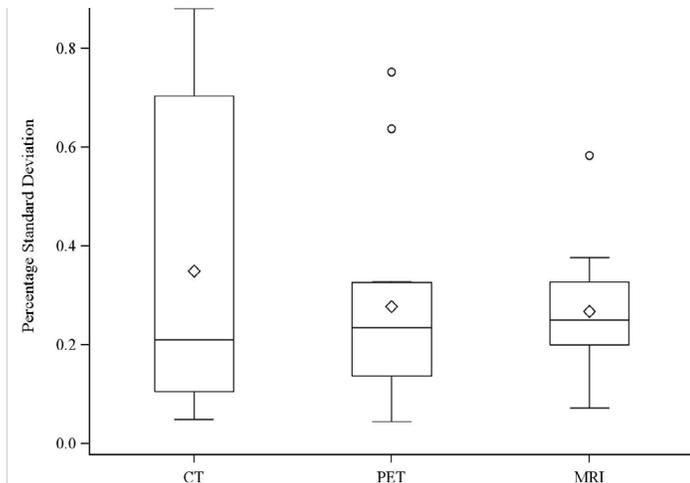
Patient characteristics are noted in Table 1. There were 2 T1 patients, 2 post-tonsillectomy patients, 2 intact T2 patients, 2 intact T3 patients, 4 intact T4 patients, and 2 gross recurrences. The distribution according to primary site was: tonsil, 6; oropharynx, 2; paranasal sinus, 2; larynx, 1; base of tongue, 1; nasopharynx, 1; hard palate, 1. One paranasal sinus tumor had neuroendocrine features and one recurrent tumor was a radiation-induced spindle cell sarcoma. The remainder of tumors were squamous cell carcinomas.

**Analysis of Contours**

The average volume for CT-, PET/CT-, and MRI-derived GTVs were 45cc, 35cc and 49cc, respectively. In 93% (13/14) of cases, PET/CT-derived GTVs had the smallest volume while MRI-derived GTVs had the largest volume in 57% (8/14) of cases (Figure 3). CT showed the largest standard deviation (variability of target definition) amongst observers (35%) compared to PET/CT (28%) and MRI (27%). However, the percent standard deviation for CT was not significantly different from the percent standard deviation for PET/CT and MRI (F2,39=0.56, p=0.58, Figure 4).



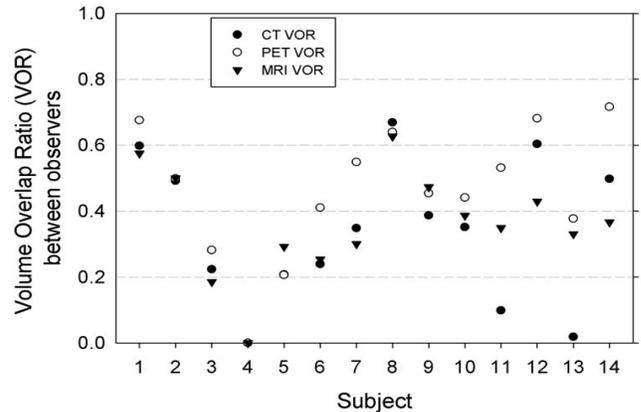
**Figure 3.** The average volume for CT-, PET/CT-, and MRI-derived GTVs were 45cc, 35cc and 49cc, respectively. In 93% (13/14) of cases, PET/CT-derived GTVs had the smallest volume while MRI-derived GTVs had the largest volume in 57% (8/14) of cases.



**Figure 4.** CT showed the largest standard deviation (variability of target definition) amongst observers (35%) compared to PET/CT (28%) and MRI (27%). However, the percent standard deviation for CT was not significantly different from the percent standard deviation for PET/CT and MRI (F2,39=0.56, p=0.58).

By imaging modality, the VOR was largest (indicating greatest interobserver agreement) in PET/CT. There was 46% average agreement, with more than 50% agreement in voxels included in 6/14 cases and more than 25% agreement in 12/14 cases. MRI had the second largest VOR (36% avg, >50% for 2 patients, >25% for 11 patients) then followed by CT (34% avg, >50% for 3 patients, > 25% for 8 patients), Figure 5. The CT VOR was lowest in the 2 T1 patients and the 2 post-tonsillectomy patients.

**GTV drawn on CT, PET, & MRI by 3 Observers**



**Figure 5.** The VOR was largest (indicating greatest interobserver agreement) in PET (46% avg, > 50% in 6 patients, > 25% in 12 patients), followed by MRI (36% avg, > 50% in 2 patients, >25% in 11 patients) then followed by CT (34% avg, > 50% for 3 patients, > 25% for 8 patients).

For all three RO contour sets, the least agreement in GTV definition occurred between MRI & PET/CT (average VOR = 41%), compared to CT & PET/CT (48%) and CT & MRI (47%). However, the VOR for MRI & PET/CT was not significantly different from the VOR for CT & PET/CT and CT & MRI (F2,121=2.27, p=0.11).

**Discussion**

Head and neck radiation therapy is currently a volume-based treatment modality defined by contours of tumor and normal structures. The potential to integrate functional, molecular and soft tissue imaging into planning is therefore of keen interest [4,5]. One of the factors limiting uniformity of treatment delivery is the variation in which the tumor target is defined. This work provides quantification of known human inconsistency as we subjectively interpret objective image-based information. The issues surrounding interobserver variability are eloquently reviewed by Weiss and Hess [1] and are an important topic as we consider future clinical trials and outcome data based on inconsistently defined (albeit by standard methods) targets.

This work explores how contouring consistency is influenced by multi-modality (CT, PET/CT, MR) imaging, and which, if any imaging modality, may result in more reproducible volumes by different observers. Unique to only a few centers in the country, our institution has PET/CT and MRI simulators in the department. This allows the patient's treatment position to be reproduced on each scanner, and thus significantly decreases registration error within the planning software.

Several descriptive findings are of interest: 1) CT showed the largest variability in target definition amongst observers despite being our standard simulation template. 2) PET/CT-derived volumes were the smallest and resulted in the least interobserver variation. 3) Low-volume GTVs were less consistently defined. 4) The least agreement in GTV definition occurred between MRI & PET/CT.

### **CT-derived volumes**

The analysis of our cohort of observers showed that the largest standard deviation occurred across CT-derived volumes. This is concerning, as this imaging modality is considered the standard for simulation, and is the modality most familiar to us. However, contouring on CT imaging is impacted by relatively less conspicuous boundaries between tumor and soft tissues, dental artifacts, and greater interobserver subjective interpretation [6,7]. The very heterogeneous patient population in this study, including 2 T1 patients and the 2 post-tonsillectomy patients, resulted in the finding that these smaller targets contributed to the largest standard deviation across observers. The danger of misinterpretation and hence missing smaller targets with unimodality imaging is suggested.

### **PET/CT-derived volumes**

The advantage of PET/CT-derived volumes in our analysis was improved consistency between observers. This finding is not universal amongst institutions who have studied CT versus PET/CT contours [8]. PET/CT-volumes were also the smallest, as they did not include subtle soft tissue/edema abnormalities suggested on CT or MRI. Several institutions have published on the utility of PET/CT as a complimentary modality to contrast-enhanced CT in radiation planning [8-17]. Recent literature has explored quantitative ways of defining PET/CT-derived volumes, as volumes defined by absolute SUV thresholds are not representative of CT-derived volumes [5,18,19]. The potential use of PET/CT to optimize consistency, then automating and/or consistently applying information from CT and MRI would be a possible benefit of this modality to simulation.

### **Low-volume GTVs**

For patient #4 (T1 intact tonsil) the VOR was zero for CT, PET/CT and MRI (Figure 5). Similarly, in patient #11 (T1 intact ton-

sil) and patient #13 (T3 post-tonsillectomy), VOR for CT was very low. Patient #5 was also post-tonsillectomy, but observers were able to define a volume with some reproducibility. These data indicate the difficulty in reproducibly defining a low-volume GTV radiographically and highlight the importance of physical examination and the need for better tools to correlate radiographic anatomy.

### **MRI versus PET/CT-derived volumes**

MRI-derived volumes were more consistent between observers than CT-derived volumes in our study. This finding is similar to Rasch, et al [20]. It is accepted that MRI enhances discrimination of extent of disease in nasopharyngeal cancer, especially in the presence of significant dental artifact and when contrast-enhanced CT is contra-indicated [21,22]. Enhanced MRI images can also be useful in defining perineural extension. The degree of variance amongst observers on MRI-derived volumes was not significantly different from CT-derived volumes in one series of twenty pharyngo-laryngeal tumors [23].

Our study provides new intermodality data comparing MRI versus PET/CT. The VOR for MRI versus PET/CT was smaller than either modality with CT, suggesting that observers are seeing unique tumor information on MRI versus PET/CT. The significance of this finding is not clear and will require detailed study as we attempt to both improve consistency and find the "true" gross tumor volume.

Overall, our study has similar findings to Daisne, et al., who compared multi-modality imaging to gross tumor specimens, and Thaigarajan, et al., who evaluated the impact of physical exam findings in addition to CT, PET/CT and MRI [2,3]. Daisne, et al. imaged patients immobilized in a thermoplastic mask prior to surgical resection. They found contours delineated on PET/CT were the smallest and most accurate compared to the gold standard of pathologically defined gross disease measured at the time of resection. There was no significant difference between the volume drawn on CT versus MRI in their study and no imaging modality captured the full extent of mucosal disease. Thaigarajan, et al. referenced multiple observers to the expert clinician who incorporated physical exam findings in their target volume [2]. Similar to our study, their group found a poor concordance between PET/CT and MRI/CT, suggesting that all three imaging modalities provide unique tumor information that could be complementary. As with Daisne, Thaigarajan found that GTVs based on imaging alone underestimated the mucosal extent identified on physical exam.

### **Study Limitations**

By having every observer use the same imaging modality ordering, a potential ordering bias could be imposed. When looking at the boxplot of the percent standard deviation (Figure 4),

we can see that CT has the greatest variability, then PET/CT followed by MRI. In the future to eliminate this potential bias, the order in which the modality images are contoured should be randomized. The small sample size and inclusion of patients with low-volume primary disease is also a likely contributor to the null findings. With a larger sample size we may be able to find significant differences, especially in the intermodality differences, since that analysis is approaching significance with the present data. However, intermodality differences should be re-analyzed with the inclusion of T1 non-contrast normal fat and T2 fast-suppressed MR images before conclusions can be made.

### Implications

Our clinical practice of GTV delineation has continued to include contrast-enhanced CT, PET/CT and MRI simulation, as well as reference to physical exam findings. Endoscopy performed by the treating radiation oncologist has proven invaluable for understanding full mucosal extent. PET/CT and MRI have been particularly valuable in defining deep soft tissue extent of disease and perineural involvement, especially in those with dental artifact on CT. We have also expanded the sequences of simulation MRI images obtained beyond T1 fat suppressed with contrast, to include T1 normal fat signal without contrast (which can reveal where fat planes between muscles are disrupted and bone marrow is involved) and T2 fat suppressed images (which may reveal peritumoral edema). Further research is needed to understand how these sequences contribute to GTV delineation. In addition, our institution's research is explaining the potential role of fluorothymidine (FLT) in predicting tumor response, with a potential role in target definition as well. As multi-modality imaging utilization for target delineation increases [24], further research may clarify how to standardize contouring on these modalities across observers. This will be critical as multimodality imaging is incorporated into head and neck cancer trial design.

### Conclusion

An interobserver difference in GTVs derived from each image modality was seen. Among three modalities, CT was least consistent, while PET/CT-derived GTVs had the smallest volumes and least interobserver variation. MRI combined with PET/CT provided the least overlap of GTVs. The significance of these differences for head and neck cancer is important to explore as we move to multimodality volume-based treatment planning as a standard method for treatment delivery. Poster presentation at Multidisciplinary Head and Neck Cancer Symposium, February 25-27, 2010, Chandler, AZ Supported by 5 U01 CA 140206

### Conflicts of interest

None of the authors have a conflict of interest.

### References

1. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol.* 2003, 179(1): 21-30.
2. Thiagarajan A, Caria N, Schöder H, Iyer NG, Wolden S et al. Target volume delineation in oropharyngeal cancer: impact of PET, MRI, and physical examination. *Int J Radiat Oncol Biol Phys.* 2012, 83(1): 220-227.
3. Daisne JF, Duprez T, Weynand B, Lonneux M, Hamoir M et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology.* 2004, 233(1): 93-100.
4. Leibfarth S, Mönnich D, Welz S, Siegel C, Schwenzer N et al. A strategy for multimodal deformable image registration to integrate PET/MR into radiotherapy treatment planning. *Acta Oncol.* 2013, 52(7): 1353-1359.
5. Chen H, Jiang J, Gao J, Liu D, Axelsson J et al. Tumor volumes measured from static and dynamic 18F-fluoro-2-deoxy-D-glucose positron emission tomography-computed tomography scan: comparison of different methods using magnetic resonance imaging as the criterion standard. *J Comput Assist Tomogr.* 2014, 38(2): 209-215.
6. O'Daniel JC, Rosenthal DI, Garden AS, Barker JL, Ahamad A et al. The effect of dental artifacts, contrast media, and experience on interobserver contouring variations in head and neck anatomy. *Am J Clin Oncol.* 2007, 30(2): 191-198.
7. Hermans R, Feron M, Bellon E, Dupont P, Van den Bogaert W et al. Laryngeal tumor volume measurements determined with CT: a study on intra- and interobserver variability. *Int J Radiat Oncol Biol Phys.* 1998, 40(3): 553-557.
8. Breen SL, Publicover J, De Silva S, Pond G, Brock K et al. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. *Int J Radiat Oncol Biol Phys.* 2007, 68(3): 763-770.
9. Berson AM, Stein NF, Riegel AC, Destian S, Ng T et al. Variability of gross tumor volume delineation in head-and-neck cancer using PET/CT fusion, Part II: the impact of a contouring protocol. *Med Dosim.* 2009, 34(1): 30-35.
10. Murakami R, Uozumi H, Hirai T, Nishimura R, Katsuragawa S et al. Impact of FDG-PET/CT fused imaging on tumor volume assessment of head-and-neck squamous cell carcinoma: intermethod and interobserver variations. *Acta Radiol.* 2008, 49(6): 693-699.

11. El-Bassiouni M, Ciernik IF, Davis JB, El-Attar I, Reiner B et al. [18FDG] PET-CT-based intensity-modulated radiotherapy treatment planning of head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2007, 69(1): 286-293.
12. Wang D, Schultz CJ, Jursinic PA, Bialkowski M, Zhu XR et al. Initial experience of FDG-PET/CT guided IMRT of head-and-neck carcinoma. *Int J Radiat Oncol Biol Phys.* 2006, 65(1): 143-151.
13. Vanderstraeten B, Duthoy W, De Gersem W, De Neve W, Thierens H. [18F]fluoro-deoxy-glucose positron emission tomography ([18F]FDG-PET) voxel intensity-based intensity-modulated radiation therapy (IMRT) for head and neck cancer. *Radiother Oncol* 2006, 79(3): 249-258.
14. Riegel AC, Berson AM, Destian S, Ng T, Tena LB et al. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *Int J Radiat Oncol Biol Phys.* 2006, 65(3): 726-732.
15. Ford EC, Kinahan PE, Hanlon L, Alessio A, Rajendran J et al. Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. *Med Phys.* 2006, 33(11): 4280-4288.
16. Paulino AC, Koshy M, Howell R, Schuster D, Davis LW. Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer. *Int J Radiat Oncol Biol Phys.* 2005, 61(5): 1385-1392.
17. Grosu AL, Piert M, Weber WA, Jeremic B, Picchio M et al. Positron emission tomography for radiation treatment planning. *Strahlenther Onkol.* 2005, 181(8): 483-499.
18. Kao CH, Hsieh TC, Yu CY, Yen KY, Yang SN et al. 18F-FDG PET/CT-based gross tumor volume definition for radiotherapy in head and neck cancer: a correlation study between suitable uptake value threshold and tumor parameters. *Radiat. Oncol.* 2010, 5: 76.
19. Yu H, Caldwell C, Mah K, Mozeg D. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans Med Imaging.* 2009, 28(3): 374-383.
20. Rasch C, Keus R, Pameijer FA, Koops W, de Ru V et al. The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. *Int J Radiat Oncol Biol Phys.* 1997, 39(4): 841-848.
21. Chung NN, Ting LL, Hsu WC, Lui LT, Wang PM. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. *Head Neck.* 2004, 26(3): 241-246.
22. Gardner M, Halimi P, Valinta D, Plantet MM, Alberini JL et al. Use of single MRI and 18F-FDG PET-CT scans in both diagnosis and radiotherapy treatment planning in patients with head and neck cancer: advantage on target volume and critical organ delineation. *Head Neck.* 2009, 31(4): 461-467.
23. Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. *Radiother Oncol.* 2005, 77(1): 25-31.
24. Simpson DR, Lawson JD, Nath SK, Rose BS, Mundt AJ et al. Utilization of advanced imaging technologies for target delineation in radiation oncology. *J Am Coll Radiol.* 2009, 6(12): 876-883.