

A new index to measure intra-observer variation in delineating GTV primary in head and neck cancer

Abstract

Background: Gross Tumor Volume (GTV) delineation is the most important step in radiotherapy treatment planning in head and neck cancer. The variation in GTV contouring can be of two types viz (I) Inter-observer variation and (II) Intra-observer variation. The goal of this study is to determine the intra-observer variation in GTV delineations in Head & Neck cancer on Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) images.

Material and methods: A total of 15 patients with histo-logically confirmed squamous cell carcinoma of oro pharynx, larynx and hypo pharynx requiring planning for radical radiotherapy with IMRT were enrolled retrospectively in this study for duration of two months. A database of 349 patients with laryngeal hypopharyngeal and oropharyngeal cancer were searched and only 15 patients underwent both CT scan and MRI for treatment planning and fulfilling the inclusion criteria. One observer contoured GTV primary on four occasions on CT and MRI within a gap of two weeks (CT₁, CT₂, MRI₁ and MRI₂ respectively). A new index is proposed to measure the intra-observer concordance with absolutely correct target volume (ACTV) and GTVs.

Results: The median volume of GTVs for CT₁, CT₂, MRI₁ and MRI₂ are observed with 19.17cc (33.55-12.19), 23.05cc (28.03-12.4), 18.93cc (32.05-11.84) and 16.77cc (32.16-11.07) respectively. The mean (SD) index values for CT₁, CT₂, MRI₁ and MRI₂ are 0.51(0.22), 0.53 (0.23), 0.54(0.13) and 0.52(0.12) respectively. It quantifies the measure of concordance of GTVs between different modalities.

Conclusion: The index study revealed that there is better concordance in delineating GTV primary on two different occasions in MR images as compared to CT images. There is no significant difference in median GTV volumes of CT scan and MRI. However, MRI-GTVs were slightly lower than GTVs in CT image.

Keywords: gtv, actv, intra-observer variations, head and neck cancer

Volume 4 Issue 4 - 2017

Tapesh Bhattacharyya,¹ Atanu Bhattacharjee²

¹Department of Radiation Oncology, Malabar Cancer Centre, India

²Department of Cancer Epidemiology, Research and Education in Cancer (ACTREC), India

Correspondence: Atanu Bhattacharjee, Department of Cancer Epidemiology, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), India, Tel +919846863789, Email atanustat@gmail.com

Received: July 10, 2017 | **Published:** August 24, 2017

Introduction

A critical step in radiotherapy treatment planning process is to accurately locate the tumor and determine its extent. Variability in gross tumor volume (GTV) delineation is a major source of systematic error in conformal radiotherapy. CT scan is the principle source of imaging data used for defining GTV for planning conformal therapy. CT is widely available less costly and compatible with radiotherapy treatment planning systems. It provides information on relative electron densities of various tissues used for those calculations algorithm. CT offers excellent bony details whereas MRI provides better soft tissue resolution, hence complimentary to each other. There is no doubt that MRI helps better visualization of soft tissue structures in head and neck cancer, however there is no agreement on impact of MRI for delineation of GTV primary tumor and its influence on intra-observer variation. It is well established in literature that there is a considerable inter-observer variability in GTV delineation.¹⁻⁸ Hermans et al.⁹ showed the inter-observer variability of CT based volume measurements of thirteen laryngeal tumors by five different observers. Both inter-observer variability and to some extent intra observer variability had a statistically significant effect on volume measurement and the most experienced observer obtained the most stable mean tumor volume over all sessions. Various modern imaging modalities like MRI and PET-CT are used to reduce the inter-observer variability and thus helping in greater agreement between different

observations of the individuals. The aim of this study is to compare the agreement between different intra-observer GTVs in head and neck cancers delineated on CT and MRI on different occasion. We propose a new index to measure the intra-observer variability in delineating GTV in head and neck malignancies.

Material and methods

It is a retrospective study with a study period of 2 months. A total of 15 patients with histo-logically confirmed squamous cell carcinoma of oro pharynx, larynx and hypo pharynx requiring planning for radical radiotherapy with IMRT were enrolled in this study. A database of 349 patients with laryngeal hypopharyngeal and oropharyngeal cancer were searched and we found 15 patients underwent both CT scan and MRI for treatment planning and fulfilling the inclusion criteria. Pre treatment workup and assessment of tumor stage was obtained through triple-endoscopy under anesthesia, contrast enhanced CT-scan (from base of skull to thoracic inlet) and indirect laryngoscope. The study group consisted of twelve male and three females with a median age of 59.3 years.

Image acquisition

All patients were scanned on a GE light speed RT₆ scanner. The patients were immobilized in a customized thermoplastic immobilization device and positioned using a reference vertical

baseline. Head and neck rests were adopted as necessary. The patients were scanned from the supra orbital ridge to entire lung. Images were acquired with the slices thickness and interval of 2.5mm. The intra venous iodinated contrast omipaque of 50ml was administered. This study was undertaken at 1.5T MRI. The patients were positioned, customized and scanned using the vertical baseline and setup as for planning CT scan. The data set were auto fused and then manually fine tuned using the visualization Box of MRI over CT scan with the correction made in the axial, sagittal, coronal or rotational directions if required.

Target volumes and organs at risk were delineated in accordance with the departmental Intensity modulated Radiotherapy (IMRT) protocol. The windowing level was adjusted to obtain the optimum view in both modalities of imaging. For each patients the primary gross tumor volume (GTV primary) and a separate GTV including macroscopic nodal disease (GTV node) was outlined. Delineated organs at risk included spinal cord, brain stem and parotid glands. The volumetric analysis of the difference between CT and MRI volumes were performed.

GTV delineation

The macroscopic (gross) extent of the primary tumor that is demonstrable on the imaging modality e.g, MRI-scan, CT-scan is defined as Gross tumor volume (GTV). The GTV primary was outlined by one observer on four occasions to access the intra observer variation. On week one day one GTV was outlined on planning CT (defined as CT₁) without access to MRI. On week one day seven GTV was outlined by the same observed on fused MRI (defined as MRI₁). After two weeks the same procedure was repeated by the same observer with blinding to the previously delineated volumes on CT (defined as CT₂) and MRI (defined as MRI₂) respectively. This blinding process and gap period between the contouring was implemented to reduce the recall bias. The volume of intersections of GTV obtained through contouring on four different occasions in CT and MRI was defined as absolutely correct target volume (ACTV). The uncommon volume in respect to ACTV observed was defined as intra observer variations for respective occasions.

Volumetric determination and intra-observer variation

A volumetric software tool has been applied to delineate all the GTVs through simultaneous visualization of multiple three dimensional data sets. The voxel size was used to determine the GTVs. The voxel size depends on the slice thickness and the resolution of the image reconstruction. The number of voxel contained within a contour was multiplied by the size of the voxel to determine the volume of the GTV. For four observations, the common volume was ACTV which is actually the intersecting volume of all four GTVs. The Boolean option of the Eclipse treatment planning system was used to determine the ACTVs. The proposed index detailed below was used as measuring index for this situation in which four delineated GTVs in two imaging modalities (CT and MRI) are compared. Proposed Index was defined as the difference between the GTV in specific imaging on particular occasion and ACTV divided by the GTV in specific imaging on that particular occasion. The mathematical formula of proposed index is given below.

Proposed index

The index I_{ijk} is defined as

$$I_{ijk} = \frac{GTV_{ijk} - ACTV_j}{GTV_{ijk}}$$

as I_{ijk} , $i = 1, 2$, $j = 1, \dots, 15$ and $k=1, 2$.

The term “i” denotes the type of imaging adopted for contouring of GTV (i.e. i=1 indicated for CT and i=2 indicated for MRI). Next the term “j” indicates the specific patient’s (jth patient) GTV detected by CT and MRI respectively. A total of 15 patient’s data is considered in this study. So the term “j” is having values of 1, 2, 3, ..., 15. The measurements of GTV detected separately by CT and MRI at two time points are denoted as CT₁, CT₂, MRI₁ and MRI₂ respectively. The repeated measurements of GTV in both the imaging is denoted through k (i.e. k=1 for CT₁ and MRI₁ and k=2 for CT₂ and MRI₂).

Further,

$$I_{1j1} = \frac{GTV_{1j1} - ACTV_j}{GTV_{1j1}} \text{ j}^{\text{th}} \text{ patient's index measured through CT}$$

scanner at week one day one.

$$I_{2j1} = \frac{GTV_{2j1} - ACTV_j}{GTV_{2j1}} \text{ j}^{\text{th}} \text{ patients index measured through MRI}$$

scanner at week one day seven.

$$I_{1j2} = \frac{GTV_{1j2} - ACTV_j}{GTV_{1j2}} \text{ j}^{\text{th}} \text{ patients index measured through CT}$$

scanner at two weeks.

$$I_{2j2} = \frac{GTV_{2j2} - ACTV_j}{GTV_{2j2}} \text{ j}^{\text{th}} \text{ patients index measured through MRI}$$

scanner at two weeks.

The range of index is between 1 (for complete disagreement between the ACTV and imaging measurements) and 0 (for perfect ACTV and imaging measurements). This method is sensitive for small variations in overlap as well as different volume sizes.

Data exploration and statistical analysis

The Shapiro-Wilk test was performed to check the non-normality of the sample data. The mean delineated GTV’s were compared through Wilcoxon signed-rank test. The coefficient of variation for all indices values from delineated GTV’s were measured and compared. Pearson correlation coefficient was performed between delineated GTVs and Indices. Student paired t-test was performed to compared the mean difference between CT₁ and CT₂ and MRI₁ and MRI₂ respectively. Statistical analyses were performed through R-studio using significance level of 0.05.

Results

Patient and tumor characteristics

A total of 15 patient’s data is considered in this study. The study group consisted of twelve males and three females with a median age of 59.3years. Most of the patients i.e. 9(60%) belong to Ca oropharynx followed by Ca larynx 3(20%), and Ca hypo pharynx 3(20%) respectively. Most of the patients had T3 disease 7(46.66%). All patients belong to locally advanced group (stage III and IVA).

Image quality

Image quality was reasonable good for both CT and MR images. However, for some patients image quality of CT scan was not up to the mark due to inadequate contrast or arte facts. But it was not difficult to perform contouring on those images.

Measurements of different volumes and indices:

The median (IQR) GTVs for CT₁, CT₂, MRI₁ and MRI₂ are observed with 19.17cc (33.55-12.19), 23.05cc (28.03-12.4), 18.93cc (32.05-11.84) and 16.77cc (32.16-11.07) respectively. The Pearson’s correlation between images are observed with 0.96 (between CT₁ & CT₂), 0.98(between MRI₁ & MRI₂) respectively. The t-test of mean value comparison observed between CT₁ & CT₂ and MRI₁ & MRI₂ are 0.40 and 0.48 respectively. No significant difference between CT images and MRI images are observed. However, difference between CT images was more in comparison to MR Images. The median (IQR) volume of ACTV observed was 7.63cc (14.64-4.33). The median (IQR) values of indices for images are computed from ACTV. The mean (SD) index values for CT₁, CT₂, MRI₁ and MRI₂ are 0.51(0.22), 0.53 (0.23), 0.54 (0.13) and 0.52 (0.12) respectively.

Performance of different images through measuring different indices and intra-observer variation

The performances of images were measured through their corresponding index value on different occasions. The minimum coefficient of variation is used to define the best performing index. The best performing index proposed the best performing imaging modality. In our case the coefficient of variations observed was 9.83, 10.36,

3.48 and 3.09 for CT₁, CT₂, MRI₁ and MRI₂ respectively. It showed that MRI₂ was the best performed contouring of GTV followed by MRI₁, CT₁ and CT₂ respectively. The indices values approaching near to 1 represent the large gap between ACTV and GTV delineated on specific imaging modality on specific occasion. However, the indices values are approaching 0 or close to zero represent the narrow or less distance between ACTV and GTV delineated on particular imaging on particular occasion. The consistency of given above indices were measured to find out the best performing observation obtain with reference to ACTV. The consistencies of those indices were ordered to point out the best observation obtained through specific imaging (CT vs MRI). The calculated table for fifteen individuals tumor volume measured in cc is detailed in Table 1. It shows that for patient’s number 1, 3,6,13 and 15 the best performing imaging is MRI₂ followed by MRI₁, CT₂ and CT₁. In contrast for patient number 4,5,7,8 and 9 CT performed better than MRI.

Figure 1 shows the GTV volume observed for each modality on different occasions. No specific pattern of GTV estimation has been observed in Figure 1. Figure 2 gives the observation considered to measured indices for each modality through ACTV. The ACTV values are considered same to measure the index values for each modality. The pattern of computed indices values are detailed in Figure 3. Plot 4 gives the box plot of indices for all imaging modalities. Figure 4 shows that there is a difference between the median (IQR) index values between CT₁ and CT₂ but that difference is quite less or negligible incase of MR₁ and MR₂ respectively. The CT₂ median index value was more as compare to other indices. This reflects its poor performance. The consistencies of performance between two MRI (i.e. MR₁ and MR₂) were better in comparison to CT (i.e. CT₁ and CT₂).

Table 1 Individuals tumor volume measured in cc

Subject	CT ₁	CT ₂	MR ₁	MR ₂	ACTV	I _{1j1}	I _{2j1}	I _{1j2}	I _{2j2}
1	34.15	23.05	18.93	12.89	8.54	0.749927	0.629501	0.548864	0.337471
2	12.13	11.56	12.92	10.69	6.12	0.495466	0.470588	0.526316	0.427502
3	32.96	23.29	47.9	34.87	14.72	0.553398	0.367969	0.692693	0.577861
4	18.89	17.89	23.69	28.03	17.31	0.083642	0.03242	0.269312	0.382447
5	25.43	24.22	32.22	33.89	14.56	0.427448	0.398844	0.548107	0.570375
6	16.04	13.24	5.78	5.49	2.32	0.855362	0.824773	0.598616	0.577413
7	45.36	47.65	62.39	63.77	31.76	0.299824	0.333473	0.490944	0.50196
8	8.96	10.13	11.39	11.78	6.22	0.305804	0.385982	0.453907	0.471986
9	12.26	23.62	31.89	30.43	7.63	0.377651	0.676969	0.76074	0.749261
10	2.57	7.51	4.47	3.8	1.62	0.36965	0.784288	0.637584	0.573684
11	28.67	31.85	27.35	28.29	13.49	0.529473	0.576452	0.506764	0.523153
12	34.38	37.41	13.89	16.77	3.35	0.90256	0.910452	0.758819	0.800239
13	19.17	13.58	12.3	11.46	4.8	0.749609	0.646539	0.609756	0.581152
14	7.6	11.32	7.2	7.47	3.86	0.492105	0.659011	0.463889	0.483266
15	123.5	90.91	98.3	104.7	64.74	0.475789	0.287867	0.341404	0.381662
Var						0.050284	0.05519	0.019052	0.016394
CVx 100%						9.83	10.36	3.48	3.09

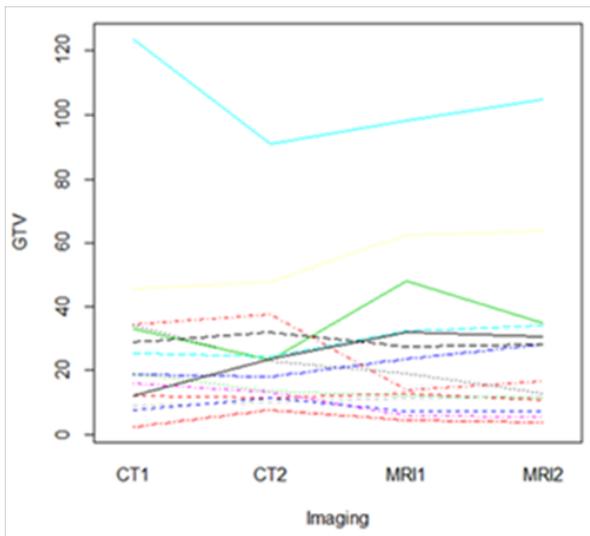


Figure 1 GTVs Obtained through different imaging.

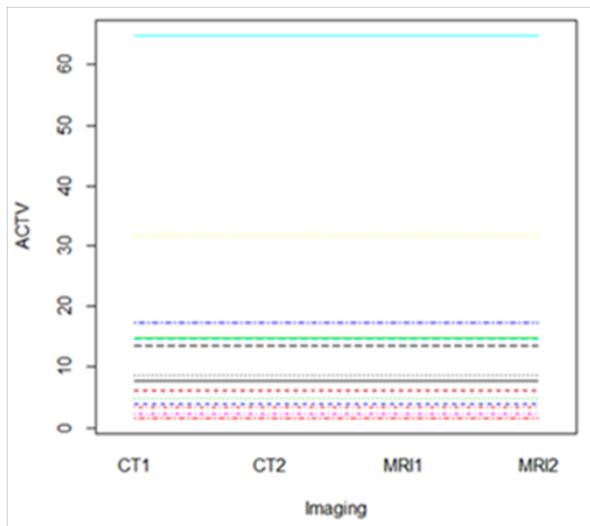


Figure 2 ACTVs obtained through different imaging.

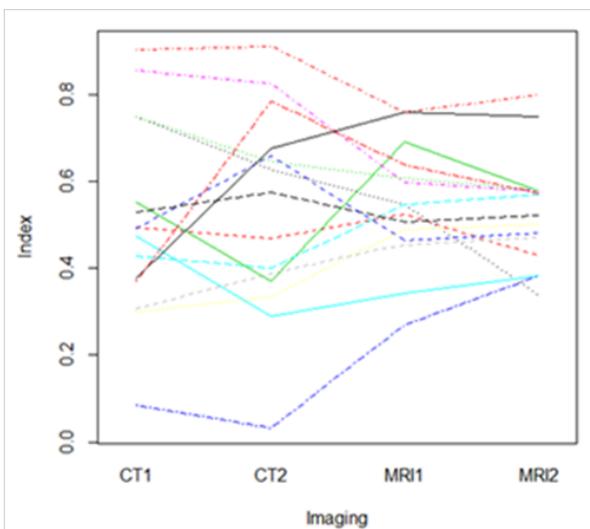


Figure 3 Estimated index obtains for different imaging.

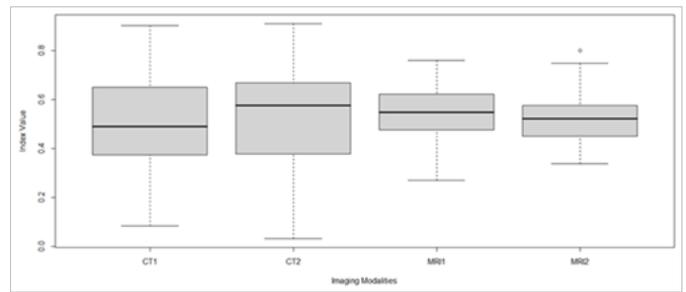


Figure 4 Comparing indices in CT and MRI.

Discussion

GTV delineation is the most important and crucial step in radiotherapy treatment planning. We add margins to GTV for clinical target volume (CTV) to cover the subclinical microscopic disease. The CTV is further expanded to create the planning target volume (PTV) encompassing the setup error and movement. If PTV margin is increased, the intra or inter observer variation in GTV can be taken care off. However in an era of high precision radiotherapy when tight margins are applied the inter or intra observer variation for delineating the GTV might result in inadequate dose coverage of the GTV. To optimize the benefits of high-precision radiation therapy GTV should be delineated as accurate as possible. Complementary information from alternative imaging modalities could help decrease the variability. Encountered for GTV delineation. MRI discriminates better between tumor and normal tissues. The use of MRI in treatment planning is limited because of geometrical distortions. Electron density information necessary for treatment dosimeter cannot be derived from MRI as it is proton based. Hence, fusion of distortion corrected imaging is required.

The index study revealed that there is better concordance in delineating GTV primary on two different occasions in MR images as compared to CT images. There is no significant difference in median GTV volumes of CT scan and MRI. However, MRI GTVs were slightly lower than GTVs in CT images. In contrast Ahmed et al. [10] showed that the GTV volume for base of tongue tumors on MRI was almost twice the volume on CT. They also reported a superior subjective visualization and delineation of base of tongue tumors on MRI scans relative to CT scans. A study performed by Geets et al.¹¹ revealed no clinical advantage of MRI over CT scan in terms of volume estimation and inter observer agreement for head and neck cancers. Same kind of results were obtained by Jager et al.¹² in delineating GTV in supra glottic laryngeal cancer and determining inter observer agreement of CT versus CT-MRI fused images. Jager et al.¹² calculated common volume(C) and encompassing volume (E) and C/E was determined for each pair of observers. The CI gen (conformity index general) was used to quantify the inter observer agreement Anderson et al.¹³ used volume overlap ratio (VOR) = Intersection/union to determine inter observer agreement. In this study the ratio, particularly the ratio between difference of GTV with ACTV and ACTV values were adopted to build the index value. The intention to build this index value is to address the relative distance measures between ACTV and GTVs (measured by concerned imaging) which helped us quantify the intra-observer concordance. Image quality was reasonably good for majority of CT and MRI images. In our Institute we have facilities of 1.5T MRI only. MRI image qualities would have been much better if we had access to 3 T MRI. The spatial difference between GTVs and ACTVs could not be addressed properly in this study.

Conclusion

The index study revealed that there is better concordance in delineating GTV primary on two different occasions in MR images as compared to CT images. There is no significant difference in median GTV volumes of CT scan and MRI. However, MRI-GTVs were slightly lower than GTVs in CT image.

Acknowledgements

The work was supported by the Department of Oral & Maxillofacial Surgery, Special Dental Care and Orthodontics of Erasmus Medical Centre in Rotterdam, the Netherlands.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors./All applicable international, national, and/or institutional guidelines for the care and use of animals were followed./All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

For this type of study, formal consent is not required. /Informed consent was obtained from all individual participants included in the study.

Conflict of interest

Authors declare that he has no conflict of interest. This is to intimate the partial work of this manuscript has been presented in AROICON-2015 Conference.

References

- Gao Z, Wilkins D, Eapen L, et al. A study of prostate delineation referenced against a gold standard created from the visible human data. *Radiother Oncol.* 2007;85(2):239–246.
- Tai P, Van Dyk J, Yu E, et al. Variability of target volume delineation in cervical esophageal cancer. *Int J Radiat Oncol Biol Phys.* 1998;42(2):277–288.
- Van de Steene J, Linthout N, de Mey J, et al. Definition of gross tumor volume in lung cancer: inter-observer variability. *Radiother Oncol.* 2002;62(1):37–49.
- Giraud P, Elles S, Helfre S, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol.* 2002;62(1):27–36.
- Cazzaniga LF, Marinoni MA, Bossi A, et al. Interphysician variability in defining the planning target volume in the irradiation of prostate and seminal vesicles. *Radiother Oncol.* 1998;47(3):293–296.
- McJury M, Dyker K, Nakielny R, et al. Optimizing localization accuracy in head and neck, and brain radiotherapy. *Br J Radiol.* 2006;79(944):672–680.
- Seddon B, Bidmead M, Wilson J, et al. Target volume definition in conformal radiotherapy for prostate cancer: quality assurance in the MRC RT-01 trial. *Radiother Oncol.* 2000;56(1):73–83.
- Fiorino C, Reni M, Bolognesi A, et al. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiother Oncol.* 1998;47(3):285–292.
- Hermans R, Feron M, Bellon E, et al. Laryngeal tumor volume measurements determined with CT: a study on intra and inter observer variability. *Int J Radiat Oncol Biol Phys.* 1998;40(3):553–557.
- Ahmed M, Schmidt M, Sohaib A, et al. The value of magnetic resonance imaging in target volume delineation of base of tongue tumors—A study using flexible surface coils. *Radiother Oncol.* 2010;94(2):161–167.
- Geets X, Daisne JF, Arcangeli S, 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.
- Jager Elise Anne, Kasperts Nicolien, Joana CM, et al. GTV delineation in supraglottic laryngeal carcinoma: inter observer agreement of CT versus CT-MR delineation. *Radiation Oncology.* 2015;10:26.
- Anderson CM, Sun Wenqing, Buatti JM, et al. Inter observer and inter modality variability in GTV delineation on simulation CT, FDG-PET and MR Images of Head and Neck Cancer. *Jacobs J Radiat Oncol.* 2014;1(1):006.