

## ORIGINAL ARTICLE

# Double reading for gross tumor volume assessment in radiotherapy planning

Ryuji Murakami<sup>1</sup>, Toshinori Hirai<sup>2</sup>, Ryo Toya<sup>3</sup>, Hideo Nakamura<sup>4</sup>, Yasuyuki Yamashita<sup>2</sup>

1. Department of Medical Imaging, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan. 2. Department of Diagnostic Radiology, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan. 3. Department of Radiation Oncology, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan. 4. Department of Neurosurgery, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan

**Correspondence:** Ryuji Murakami. Address: Department of Medical Imaging, Faculty of Life Sciences, Kumamoto University, 4-24-1 Kuhonji, Kumamoto 862-0976, Japan. Telephone: 81-963-735-463. Fax: 81-963-735-463. E-mail: murakami@kumamoto-u.ac.jp

**Received:** June 1, 2012

**Accepted:** July 16, 2012

**Published:** August 1, 2012

**DOI:** 10.5430/jst.v2n4p38

**URL:** <http://dx.doi.org/10.5430/jst.v2n4p38>

## Abstract

**Background/Objective:** The precise definition of the gross tumor volume (GTV) that takes into account intra- and interobserver variability is necessary for high-precision radiotherapy (RT) techniques. The purpose of this study was to demonstrate the practical GTV assessment by a “double reading” approach.

**Methods:** Pretreatment magnetic resonance (MR) imaging, including the post-contrast 3D magnetization-prepared rapid-gradient echo (MP-RAGE) sequence (section thickness 1.0 mm) was performed on a 3T superconducting imager in 50 patients with glioblastoma. MR images were transferred to a RT planning system (RTPS) that provides many opportunities for GTV contouring, e.g., at diagnosis, surgical navigation, and RT deliberations. Independent 2 observers preliminarily contoured the GTV on MR images. After planning-CT scanning, CT images with a 1.0 mm slice interval were transferred to the RTPS, registered with the diagnostic images, and then the preliminarily-contoured strictures were copied onto the CT images and used for GTV assessment. The practical GTV on the planning CT was determined by integrating the interpretations and adding information on postoperative changes. The interobserver variability in GTV contouring was assessed by Bland-Altman analysis and the concordance index.

**Results:** There was substantial interobserver variability in GTV contouring (95% limits of agreement: -29.4%, 16.8%). The mean interobserver concordance rate for the GTV was 82.1% (range 56.5-91.2%). The practical GTVs were significantly larger than the preliminarily-contoured GTVs by both observers ( $p < 0.01$ ).

**Conclusions:** Considering interobserver variability, “double reading” is necessary for practical GTV assessment. This approach for volume assessment may facilitate the standardization of treatments, not only of RT but also of surgery and chemotherapy.

## Key words

Radiation therapy planning, Gross tumor volume, Image registration, Interobserver variability, Glioblastoma

# 1 Introduction

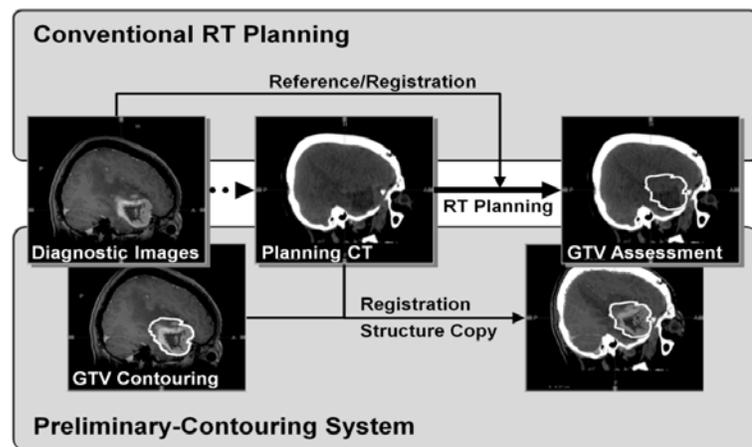
The precise definition of the gross tumor volume (GTV) is essential for high-precision radiotherapy (RT) techniques such as three-dimensional conformal RT (3D-CRT), intensity-modulated RT, particle-beam RT, and brachytherapy. With respect to the clinical target volume (CTV) and organs at risk (OARs) based on normal structures, in efforts to standardize contouring with consensus among radiation oncologists, radiologists, and clinicians, atlases are developed for several disease sites, i.e. brain-, head and neck-, breast-, prostate-, rectal-, and gynecologic malignancies<sup>[1, 2]</sup>. On the other hand, the GTV must be contoured consensually on a case-by-case basis. In spite of contouring protocols on multimodal registered images, e.g., contrast-enhanced computed tomography (CT), magnetic resonance (MR) imaging, and positron emission tomography (PET), previous studies revealed that intra- and interobserver variability in GTV assessment persists<sup>[3-9]</sup>. However, reported data demonstrated the experimental contouring but the practical GTV in RT planning.

Errors in diagnostic radiology are common, and dialogue among clinicians and radiologists, the evaluation of multimodal images, and iterative readings are thought to decrease the error<sup>[10-14]</sup>. Therefore, a practical consensus for GTV assessments, similar to the “double reading” approach, must be developed. In the routine clinical course, however, structures including GTV, CTV, and OARs are contoured on planning-CT images, and directly used for RT planning. Consequently, there are little opportunities for integration of interobserver variability.

We designed a preliminary-contouring system for “double reading” in the practical GTV assessment. It permits many opportunities for GTV contouring on diagnostic images, e.g., at diagnosis, surgical navigation, and RT deliberations (Figure 1). The purpose of this study was to demonstrate the practical GTV assessment by a “double reading” approach.

**Figure 1.** The preliminary-contouring system

In the conventional radiotherapy (RT) planning, the gross tumor volume (GTV) is contoured on planning-CT images and directly used for RT planning. The preliminary-contouring system permits many opportunities for GTV contouring on diagnostic images even before planning-CT scanning.



## 2 Patients and methods

### 2.1 Patients

Between March 2007 and January 2010, we performed RT planning using the preliminary-contouring system in 50 consecutive patients with newly-diagnosed glioblastoma, as part of an institutional review board-approved study. The subjects were 28 men and 22 women ranging in age from 32 to 85 years (mean 68 years). All underwent pretreatment MR imaging, surgery, and postoperative RT with a total dose of 60 Gy administered by using conventional fractionation. Surgery consisted of biopsy (n = 15) and partial- (n = 21) or gross-total resection (n = 14). Before RT planning, the appropriate treatment for each patient was discussed at a clinical conference attended by radiation oncologists, neuroradiologists, and neurosurgeons.

## 2.2 Diagnostic imaging

Pretreatment MR imaging was performed on a 3T superconducting imager (Magnetom Trio; Siemens AG, Erlangen, Germany) using an 8-channel phased-array head coil. Routine brain MR imaging at our hospital consisted of pre- and postcontrast T1-weighted spin-echo (SE)-(600/8.5 [repetition time msec/echo time msec]), T2-weighted fast SE-(3,600/96, echo train length 7), and fluid-attenuated inversion-recovery (FLAIR) (9,000/81/2,500[repetition time msec/echo time msec/inversion time msec], echo train length 15) sequences. These conventional images were acquired at a section thickness of 5 mm with a 1-mm intersection gap, a 256 - 512 matrix, and a 230-mm field of view (FOV). All patients underwent a post-contrast 3D magnetization-prepared rapid acquisition gradient-echo (MP-RAGE) <sup>[15]</sup> sequence (1,900/4.7/900, 9 flip angle) at a section thickness of 1.0 mm, 256 matrix, 256 mm FOV, and pixel size of 1.0 × 1.0 mm<sup>2</sup>. Post-contrast T1-weighted SE and 3D-MP-RAGE sequences were obtained after a bolus injection of 0.1 mmol/kg Gd-DTPA (Magnevist; Bayer HealthCare Pharmaceuticals, Osaka, Japan). For RT planning we determined the contrast-enhanced GTV areas on post-contrast T1-weighted SE- and 3D-MP-RAGE images and identified the high-intensity areas representing the CTV on T2-weighted fast SE- and FLAIR images.

## 2.3 Preliminary contouring

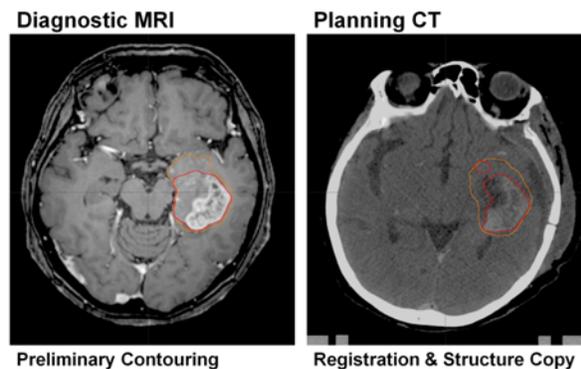
Diagnostic MR images were transferred to the 3D-RT planning system (3D-RTPS) (Eclipse version 7.5; Varian Medical Systems, Palo Alto, CA, USA) using the digital imaging and communications in medicine (DICOM) format. The system permits the preliminary contouring and recording of structures on diagnostic images. After a clinical conference addressing the appropriate RT for each patient, 2 observers, a neuroradiologist (observer A) and a radiation oncologist (observer B), independently interpreted diagnostic MR images and preliminarily contoured the GTV (GTV<sub>MRI</sub>) on 3D-MP-RAGE images.

## 2.4 Planning CT

Planning-CT images were obtained on a 4-row multidetector CT scanner (LightSpeed RT; GE Healthcare, Tokyo, Japan) at a slice thickness of 2.5 mm with a 1-mm slice interval, a 512 matrix, and a 350 mm FOV. The CT images were transferred to 3D-RTPS and registered with diagnostic MR images using a combination of automatic and manual methods. Then the GTV<sub>MRI</sub> determined by the 2 observers were copied onto planning-CT images (GTV<sub>COPY</sub>) and used to assess the target volume (Figure 2). The radiation oncologist (observer B) determined the practical GTV on planning CT by integrating the interpretations and adding information on postoperative changes.

**Figure 2.** GTV contouring in a 59-year-old woman with glioblastoma

The independent observers A and B preliminarily contoured the gross tumor volume (GTV, red and orange lines) on diagnostic MR images. After image registration between CT and MR images, the preliminarily-contoured structures were copied onto CT images and used for GTV assessment. There was a considerable difference in GTV contouring; the interobserver concordance rate was 56.5%.



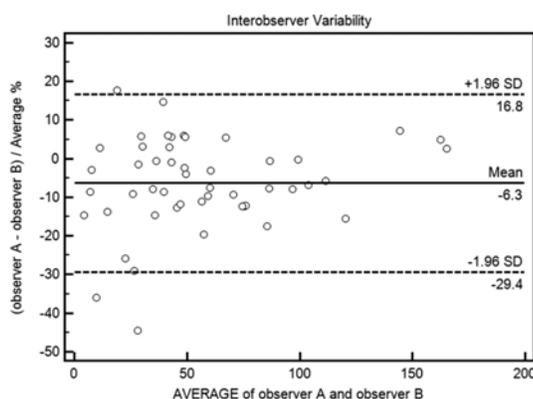
## 2.5 Statistical analysis

We used the repeated-measures analysis of variance (ANOVA) with Bonferroni adjustment to compare the GTVs. Additionally, differences in the GTV, i.e. between GTV<sub>MRI</sub> and GTV<sub>COPY</sub> (n = 100, system error) and in the GTV<sub>COPY</sub> recorded by the 2 observers (n = 50, interobserver variability) were assessed by Bland-Altman analysis. For geometric interobserver comparison of the GTVs we calculated the concordance index as the ratio of the intersection ( $A \cap B$ ) of the

GTVs to their union (AUB); this is the concordance rate =  $(A \cap B)/(A \cup B)$ . This rate ranges from 0% (complete disagreement, i.e.  $A \cap B = 0$ ) to 100% (perfect concordance) and is very sensitive to small variations in overlap because it is normalized to the union of the volumes. For example, if two equal volumes overlap by 50% each, the concordance index is 33%<sup>[5,6]</sup>. Statistical analyses were with statistical software (MedCalc version 9.2.1.0; MedCalc Software, Mariakerke, Belgium).

### 3 Results

Preliminary contouring was completed within 5 minutes by each observer in all patients. The median volume recorded as GTV<sub>COPY</sub> for the 50 tumors was 46.3 cm<sup>3</sup> (observer A, range 4.0 - 167.3 cm<sup>3</sup>) and 48.0 cm<sup>3</sup> (observer B, range 4.7 - 163.0 cm<sup>3</sup>). The practical GTVs (median 52.8 cm<sup>3</sup>, range 4.4-172.5 cm<sup>3</sup>) including information on postoperative changes were significantly larger than the preliminary GTVs by both observers ( $p < 0.01$ ). Bland-Altman plot revealed interobserver variability in GTV contouring (Figure 3). Furthermore, the 95% limits of agreement for the GTVs were much wider for interobserver variability (-29.4%, 16.8%) than system error (-1.8%, 0.8%) (Table 1). The mean interobserver concordance rate for GTV<sub>COPY</sub> was  $82.1\% \pm 7.6\%$  (range 56.5 - 91.2%).



**Figure 3.** Bland-Altman plot of interobserver differences in GTV contouring

**Table 1.** Differences in GTV contouring by Bland-Altman analysis

	Mean difference $\pm$ SD	95% limits of agreement
System error*		
Difference (cm <sup>3</sup> )	-0.2 $\pm$ 0.1	-0.3, 0.0
Percent difference	-0.5% $\pm$ 0.7%	-1.8%, 0.8%
Interobserver variability <sup>†</sup>		
Difference (cm <sup>3</sup> )	-2.8 $\pm$ 5.7	-13.9, 8.4
Percent difference	-6.3% $\pm$ 11.8%	-29.4%, 16.8%

Abbreviations: SD = standard deviation; \*Difference between GTV<sub>MRI</sub> and GTV<sub>COPY</sub> (n = 100); <sup>†</sup>Difference in the GTV<sub>COPY</sub> recorded by the 2 observers (n = 50)

### 4 Discussion

As patients with GBM are usually treated with surgery followed by a combination of RT and chemotherapy<sup>[16]</sup>, the GTV assessment is important for surgery and RT. Experimental studies that evaluated the consistency of GTV assessments demonstrated substantial intra- and interobserver variability even among experienced readers<sup>[3-9]</sup>. Therefore, we designed and attempted RT planning with a preliminary-contouring system in 50 glioblastoma patients. In our practical application, we took into account interobserver variability and integrated their GTV contouring results.

Cooper et al. [8], who compared GTVs delineated by 8 experienced observers on contrast-enhanced CT from 20 patients with supraglottic carcinoma, found that the average concordance rate between pairs of readers was 53%. Fox et al. [7] studied the GTV on FDG-PET/CT registered images of lung cancer; they documented intra- and interobserver concordance rates of 70% and 71%, respectively. Regarding brain tumors, the central tumor core, due to its visibility on contrast-enhanced MR images, is now accepted as an indicator of the GTV [17]. To utilize brain MR images for GTV contouring, previous studies suggested the value of the registration of MR images to planning-CT images; the use of MR/CT-registered images rather than hard copies of MR images reduced interobserver variability [3,4].

We used 3D-MP-RAGE images obtained on a 3T superconducting imager for GTV contouring in glioblastoma patients. These 3D images provide readily-acquired multiplanar reformations and depict more focal lesions than T1-weighted SE images [15]. Although our mean interobserver concordance rate should be higher than in earlier studies that employed conventional MR images [3-5], considerable interobserver differences for GTV contouring persisted; our lowest concordance rate was 56.5% (Figure 2). The “double reading” system presented here can be a valuable method for the integration of GTV contouring. Our system provides many opportunities for GTV contouring even before planning-CT scanning.

There were some limitations to our study. First, there were minor volume differences between  $GTV_{MRI}$  and  $GTV_{COPY}$  in our system. The system error was attributable to image registration and structure copying. Similar errors, depending on the imaging modality and software, should be encountered in other systems for image registration. We determined the practical GTV not only with using preliminarily-contoured structures but also with considering misregistration and postoperative changes. The re-interpretation may allow for the integration of intraobserver variability. Additionally, the current study demonstrated no histopathological confirmations but substantial interobserver variability in GTV contouring. Although previous studies suggested that the “double reading” can reduce errors in diagnostic radiology [10,11], efforts are underway in our institution to resolve these issues, especially the accuracy of GTV assessments in the “double reading” system. Nevertheless, the “double reading” system should provide the clinical consensus and the clinical training in GTV contouring.

## 5 Conclusion

In spite of up-to-date contouring protocols, there is considerable variability in the visual interpretation of oncologic images. Until optimal imaging protocols for the consistency of GTV contouring are developed, “double reading” is necessary for practical GTV assessment. This approach for volume assessment may facilitate the standardization of treatments, not only of RT but also of surgery and chemotherapy.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements and funding

This work was supported by a Grant-in-Aid for Scientific Research (22591388).

## References

- [1] Small W, Jr., Mell LK, Anderson P, Creutzberg C, De Los Santos J, Gaffney D, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy in postoperative treatment of endometrial and cervical cancer. *Int J Radiat Oncol Biol Phys.* 2008; 71: 428-34. <http://www.ncbi.nlm.nih.gov/pubmed/18037584>. PMID:18037584 <http://dx.doi.org/10.1016/j.ijrobp.2007.09.042>
- [2] Lim K, Small W, Jr., Portelance L, Creutzberg C, Jurgenliemk-Schulz IM, Mundt A, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. *Int J Radiat Oncol*

- Biol Phys. 2011; 79: 348-55. <http://www.ncbi.nlm.nih.gov/pubmed/20472347>. PMID:20472347  
<http://dx.doi.org/10.1016/j.ijrobp.2009.10.075>
- [3] Aoyama H, Shirato H, Nishioka T, Hashimoto S, Tsuchiya K, Kagei K, et al. Magnetic resonance imaging system for three-dimensional conformal radiotherapy and its impact on gross tumor volume delineation of central nervous system tumors. *Int J Radiat Oncol Biol Phys.* 2001; 50: 821-7. <http://www.ncbi.nlm.nih.gov/pubmed/11395252>.  
[http://dx.doi.org/10.1016/S0360-3016\(01\)01598-X](http://dx.doi.org/10.1016/S0360-3016(01)01598-X)
- [4] Cattaneo GM, Reni M, Rizzo G, Castellone P, Ceresoli GL, Cozzarini C, et al. Target delineation in post-operative radiotherapy of brain gliomas: Interobserver variability and impact of image registration of MR (pre-operative) images on treatment planning CT scans. *Radiother Oncol.* 2005; 75: 217-23. <http://www.ncbi.nlm.nih.gov/pubmed/15885831>.
- [5] Weltens C, Menten J, Feron M, Bellon E, Demaerel P, Maes F, et al. Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. *Radiother Oncol.* 2001; 60: 49-59. <http://www.ncbi.nlm.nih.gov/pubmed/11410304>. [http://dx.doi.org/10.1016/S0167-8140\(01\)00371-1](http://dx.doi.org/10.1016/S0167-8140(01)00371-1)
- [6] Murakami R, Uozumi H, Hirai T, Nishimura R, Katsuragawa S, Shiraishi 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: 693-9. <http://www.ncbi.nlm.nih.gov/pubmed/18568563>. PMID:18568563 <http://dx.doi.org/10.1080/02841850802027034>
- [7] Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys.* 2005; 62: 70-5. <http://www.ncbi.nlm.nih.gov/pubmed/15850904>. PMID:15850904  
<http://dx.doi.org/10.1016/j.ijrobp.2004.09.020>
- [8] Cooper JS, Mukherji SK, Toledano AY, Beldan C, Schmalfuss IM, Amdur R, et al. An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN protocol 6658). *Int J Radiat Oncol Biol Phys.* 2007; 67: 972-5. <http://www.ncbi.nlm.nih.gov/pubmed/17208386>. PMID:17208386 <http://dx.doi.org/10.1016/j.ijrobp.2006.10.029>
- [9] Geets X, Daisne JF, Arcangeli S, Coche E, De Poel M, Duprez T, 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: 25-31. <http://www.ncbi.nlm.nih.gov/pubmed/15919126>. PMID:15919126 <http://dx.doi.org/10.1016/j.radonc.2005.04.010>
- [10] Goddard P, Leslie A, Jones A, Wakeley C, Kabala J. Error in radiology. *Br J Radiol.* 2001; 74: 949-51. <http://www.ncbi.nlm.nih.gov/pubmed/11675313>. PMID:11675313
- [11] Singh H, Sethi S, Raber M, Petersen LA. Errors in cancer diagnosis: Current understanding and future directions. *J Clin Oncol.* 2007; 25: 5009-18. <http://www.ncbi.nlm.nih.gov/pubmed/17971601>. PMID:17971601  
<http://dx.doi.org/10.1200/JCO.2007.13.2142>
- [12] Murakami R, Uozumi H, Hirai T, Nishimura R, Shiraishi S, Ota K, et al. Impact of FDG-PET/CT imaging on nodal staging for head-and-neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys.* 2007; 68: 377-82. <http://www.ncbi.nlm.nih.gov/pubmed/17321066>. PMID:17321066 <http://dx.doi.org/10.1016/j.ijrobp.2006.12.032>
- [13] Kajihara H, Hayashida Y, Murakami R, Katahira K, Nishimura R, Hamada Y, et al. Usefulness of diffusion-weighted imaging in the localization of prostate cancer. *Int J Radiat Oncol Biol Phys.* 2009; 74: 399-403. <http://www.ncbi.nlm.nih.gov/pubmed/19019564>. PMID:19019564 <http://dx.doi.org/10.1016/j.ijrobp.2008.08.017>
- [14] Murakami R, Hirai T, Sugahara T, Fukuoka H, Toya R, Nishimura S, et al. Grading astrocytic tumors by using apparent diffusion coefficient parameters: Superiority of a one- versus two-parameter pilot method. *Radiology.* 2009; 251: 838-45. <http://radiology.rsna.org/content/251/3/838.full.pdf>. PMID:19318585 <http://dx.doi.org/10.1148/radiol.2513080899>
- [15] Brant-Zawadzki M, Gillan GD, Nitz WR. MP RAGE: A three-dimensional, T1-weighted, gradient-echo sequence -- Initial experience in the brain. *Radiology.* 1992; 182: 769-75. <http://www.ncbi.nlm.nih.gov/pubmed/1535892>. PMID:1535892
- [16] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med.* 2005; 352: 987-96. <http://www.ncbi.nlm.nih.gov/pubmed/15758009>. PMID:15758009 <http://dx.doi.org/10.1056/NEJMoa043330>
- [17] Chang EL, Akyurek S, Avalos T, Rebuena N, Spicer C, Garcia J, et al. Evaluation of peritumoral edema in the delineation of radiotherapy clinical target volumes for glioblastoma. *Int J Radiat Oncol Biol Phys.* 2007; 68: 144-50. <http://www.ncbi.nlm.nih.gov/pubmed/17306935>. PMID:17306935 <http://dx.doi.org/10.1016/j.ijrobp.2006.12.009>