Radiotherapy and Oncology 115 (2015) 407–411

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Auto-segmentation in NPC

# Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: A multi-institution clinical study

Chang-Juan Tao [a], Jun-Lin Yi [b], Nian-Yong Chen [c], Wei Ren [d], Jason Cheng [e], Stewart Tung [f], Lin Kong [g], Shao-Jun Lin [h], Jian-Ji Pan [h], Guang-Shun Zhang [a], Jiang Hu [a], Zhen-Yu Qi [a], Jun Ma [a], Jia-De Lu [d], Di Yan [i], Ying Sun [a,*]

[a] Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou; [b] Department of Radiation Oncology, Cancer Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing; [c] Department of Radiation Oncology, Cancer Center, West China Hospital, Sichuan University, Chengdu, China; [d] Department of Radiation Oncology, National University Cancer Institute, National University Health System, National University of Singapore, Republic of Singapore; [e] Department of Radiation Oncology, National Taiwan University Hospital, Taipei, Taiwan; [f] Department of Clinical Oncology, Tuen Mun Hospital, Hong Kong, China; [g] Department of Radiation Oncology, Fudan University, Shanghai Cancer Center; [h] Department of Radiation Oncology, Cancer Hospital of Fujian Medical University, Fuzhou, China; [i] Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, United States

ARTICLE INFO

*Article history:*
Received 9 September 2014
Received in revised form 12 May 2015
Accepted 12 May 2015
Available online 26 May 2015

*Keywords:*
Nasopharyngeal carcinoma
Organs at risk
Atlas-based auto segmentation (ABAS)
Contour variation
Dosimetric parameter

ABSTRACT

*Background and purpose:* To assess whether consensus guideline-based atlas-based auto-segmentation (ABAS) reduces interobserver variation and improves dosimetric parameter consistency for organs at risk (OARs) in nasopharyngeal carcinoma (NPC).

*Materials and methods:* Eight radiation oncologists from 8 institutes contoured 20 OARs on planning CT images of 16 patients via manual contouring and manually-edited ABAS contouring. Interobserver variation [volume coefficient of variation (CV), Dice similarity coefficient (DSC), three-dimensional isocenter difference (3D-ICD)] and dosimetric parameters were compared between the two methods of contouring for each OAR.

*Results:* Interobserver variation was significant for all OARs in manual contouring, resulting in significant dosimetric parameter variation ($P < 0.05$). Edited ABAS significantly improved multiple metrics and reduced dosimetric parameter variation for most OARs; brainstem, spinal cord, cochleae, temporomandibular joint (TMJ), larynx and pharyngeal constrictor muscle (PCM) obtained most benefit (range of mean DSC, volume CV and main ICD values was 0.36–0.83, 12.1–84.3%, 2.2–5.0 mm for manual contouring and 0.42–0.86, 7.2–70.6%, 1.2–3.5 mm for edited ABAS contouring, respectively; range of dose CV reduction: 1.0–3.0%).

*Conclusion:* Substantial objective interobserver differences occur during manual contouring, resulting in significant dosimetric parameter variation. Edited ABAS reduced interobserver variation and improved dosimetric parameter consistency, particularly for brainstem, spinal cord, cochleae, TMJ, larynx and PCM.



Intensity-modulated radiotherapy (IMRT), the main treatment modality for nasopharyngeal carcinoma (NPC), provides steep dose gradients to enable precise tumor target coverage and normal tissue sparing. Accurate target volume and organs at risk (OARs) delineation are required to obtain the therapeutic advantages of IMRT and minimize normal tissue irradiation. However, large interobserver OAR contouring variation has been reported in head and neck cancer [1] and may significantly affect dosimetric parameters, impeding study of late side-effects and establishment of a reliable normal tissue complication probability model [1,2].

Interobserver variation in OARs delineation mainly originates from different subjective interpretation of organ boundaries and objective contouring variation [3,4]. Standardized guidelines and anatomy atlases have reduced subjective diversity. More recently, atlas-based auto-segmentation (ABAS), a promising tool that automatically contours the OARs on CT simulation images, has gained popularity and is clinically acceptable, time-saving and potentially decreases interobserver variation [5–11]. Furthermore, a multi-subject atlas template, created using specialist contouring on a database of CT images based on recognized guidelines, was

* Corresponding author at: Department of Radiation Oncology, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou, Guangdong 510060, China.
*E-mail address:* sunying@sysucc.org.cn (Y. Sun).

http://dx.doi.org/10.1016/j.radonc.2015.05.012
0167-8140/

superior to a single-subject atlas [5]. However, manual editing is necessary to ensure the accuracy of ABAS [12] and it is questionable whether reduced interobserver variation in OAR volumes delineated using ABAS, converts into improved dosimetric parameter consistency.

We performed a multi-institution study to assess whether multi-subject ABAS auto-segmentation with manual editing can reduce objective interobserver variation and improve dosimetric parameter consistency for OARs in NPC.

## Materials and methods

### *Patients*

Volumetric CT datasets for 16 patients with newly-diagnosed, pathologically-confirmed stage I–IVB NPC (7th edition of AJCC staging system; Supplemental Table 1) treated with radical IMRT at one cancer center between October 2010 and October 2011 were retrospectively reviewed. IMRT was delivered at prescribed doses of 70 Gy (33 fractions) to the nasopharyngeal gross tumor volume (GTV), 60–66 Gy to involved nodal GTV, 60 Gy to high-risk clinical target volume (CTV), and 56 Gy to low-risk CTV and neck nodal regions. Neoadjuvant, concurrent or adjuvant platinum-based chemotherapy were recommended in stage III–IVB.

All patients were immobilized in the supine position using a head, neck and shoulder thermoplastic mask and CT images with and without contrast were obtained (3 mm slices from head to 2 cm below sternoclavicular joint; matrix size, $512 \times 512$; voxel resolution, $0.97 \times 0.97 \times 3.0$ mm in left–right, antero–posterior and cranio–caudal directions).

### *Manual contouring*

Eight radiation oncologists from 8 independent institutes manually contoured 20 OARs (Supplemental Table 2) on the fused enhanced and non-enhanced planning CT images of the 16 cases with reference to OARs Delineation Guidelines for the Head and Neck [3] using Focal (version 4.3.3; Elekta AB, Stockholm, Sweden), Pinnacle (Philips Medical Systems, Bothell, WA, USA) or Eclipse (Varian Medical Systems, Palo Alto, CA, USA). Clinical characteristics and staging information were provided; physicians were blinded to each other's contours.

### *Multi-subject atlas creation and edited ABAS OAR-contouring*

ABAS segments OARs on CT datasets using a single- or multi-patient atlas. Atlas patients were selected from (1) patients whose head and neck planning CT scans had no obvious artifacts and normal tissue tumor infiltration; (2) the circumference of the center plane of the nasopharyngeal cavity was ranked for 50 patients who fulfilled criteria 1 (Supplemental Table 3); then (3) one of every eight of these patients was selected as a representative population. An experienced radiation oncologist manually contoured OARs on the planning CT scans of the 7 patients with reference to consensus guidelines, before importing the datasets into ABAS (Version 2.01.00, Elekta AB). The Simultaneous Truth And Performance Level Evaluation (STAPLE) algorithm [13] was used to fuse the multiple single-subject atlas auto-segmentation sets into one multi-subject auto-segmentation set.

To minimize recall bias, a minimum of one month after manual contouring, the eight radiation oncologists reviewed and edited the final multi-subject auto-segmented OARs using consensus guidelines.

### *Quantitative analysis of interobserver variation in OAR contouring*

Three indices were used to assess inter-observer variation and were calculated on OARs contours using VODCA (v.5.3.4a; MSS; Hagendorn, Switzerland), which is well-validated [14].

Volume variation was evaluated using volume coefficient of variation (CV), which assesses observer-relative standard deviation in delineation of volume (ratio between standard deviation and mean OAR volume). The absolute volume differences between the two methods of contouring ($\Delta V$) were also calculated.

Similarity was assessed using the Dice similarity coefficient (DSC), calculated using:

$$\mathrm{DSC} = \frac{2|A \cap B|}{|A| + |B|}$$

where $A$ and $B$ are the two structures evaluated. DSC varies between 0 (no overlap) and 1 (perfect agreement). For multiple observers and multiple patients, DSC can be generalized by considering DSC value distribution with reference to pairs of observers [14].

Position variation was assessed using three-dimensional isocenter difference (3D-ICD), calculated as:

$$\mathrm{ICD}\ [\mathrm{mm}] = |((\mathrm{Max\ Range\ A} + \mathrm{Min\ Range\ A})/2 - (\mathrm{Max\ Range\ B} + \mathrm{Min\ Range\ B}/2))|$$

where MaxRange indicates the largest; and MinRange, the smallest coordinate value of each contour created by eight radiation oncologists in a certain direction in CT simulation images.

**Table 1**
Mean DSC values and mean volume coefficient of variance (CV ± SD) for the OARs.

| OAR | DSC | | | Volume CV | | | $P^*$ |
|---|---|---|---|---|---|---|---|
| | Manual | Edited ABAS | $P^*$ | Manual | Edited ABAS | $\triangle V$ (cm$^3$) # | |
| Brainstem | 0.83 ± 0.03 | 0.86 ± 0.04 | <0.01 | 12.1% ± 0.03 | 7.2% ± 0.02 | 1.58 | <0.01 |
| Spinal cord | 0.77 ± 0.04 | 0.82 ± 0.04 | 0.04 | 25.6% ± 0.06 | 21.6% ± 0.10 | −0.43 | <0.05 |
| TMJ_L | 0.49 ± 0.18 | 0.69 ± 0.07 | <0.01 | 60.6% ± 0.06 | 32.7% ± 0.09 | 0.37 | <0.01 |
| TMJ_R | 0.50 ± 0.18 | 0.71 ± 0.07 | <0.01 | 58.8% ± 0.07 | 36.0% ± 0.07 | 0.34 | <0.01 |
| Cochlea_L | 0.37 ± 0.10 | 0.43 ± 0.12 | <0.01 | 84.3% ± 0.18 | 70.6% ± 0.13 | 0.11 | 0.07 |
| Cochlea_R | 0.36 ± 0.11 | 0.42 ± 0.11 | <0.01 | 82.5% ± 0.19 | 71.8% ± 0.12 | 0.11 | 0.09 |
| PCM_S | 0.44 ± 0.07 | 0.63 ± 0.09 | <0.01 | 30.4% ± 0.09 | 24.1% ± 0.07 | 1.98 | <0.01 |
| PCM_M | 0.50 ± 0.08 | 0.64 ± 0.07 | <0.01 | 53.3% ± 0.22 | 24.4% ± 0.14 | −0.46 | <0.01 |
| PCM_I | 0.50 ± 0.09 | 0.65 ± 0.06 | <0.01 | 30.6% ± 0.11 | 21.7% ± 0.09 | 0.52 | <0.01 |
| Larynx_supraglottic | 0.60 ± 0.05 | 0.73 ± 0.04 | <0.01 | 36.4% ± 0.17 | 24.8% ± 0.09 | 0.73 | 0.02 |
| Larynx_glottic | 0.49 ± 0.09 | 0.64 ± 0.08 | <0.01 | 68.5% ± 0.12 | 48.2% ± 0.11 | −2.22 | 0.02 |

*Abbreviations*: SD: standard deviation; CV: coefficient of variation; L, left; R, right; TMJ: temporomandibular joint; PCM: pharyngeal constrictor muscle; S: superior; M: middle; I: inferior.

* $P$ value: the difference between manual contouring and edited ABAS contouring was tested using paired t-tests or the Wilcoxon rank test.

\# $\Delta V$: The absolute volume difference between manual contouring and edited ABAS contouring; a negative value indicates the mean volume for manual contouring was larger than that of edited ABAS contouring.