Clinical and Translational Radiation Oncology 12 (2018) 40–46

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

journal homepage: www.elsevier.com/locate/ctro

Original Research Article

A prospective in silico analysis of interdisciplinary and interobserver spatial variability in post-operative target delineation of high-risk oral cavity cancers: Does physician specialty matter?

Sweet Ping Ng [a,*], Brandon A. Dyer [b], Jayashree Kalpathy-Cramer [c], Abdallah Sherif Radwan Mohamed [a], Musaddiq J. Awan [d], G. Brandon Gunn [a], Jack Phan [a], Mark Zafereo [e], J. Matthew Debnam [f], Carol M. Lewis [e], Rivka R. Colen [f], Michael E. Kupferman [e], Nandita Guha-Thakurta [f], Guadalupe Canahuate [g], G. Elisabeta Marai [h], David Vock [i], Bronwyn Hamilton [j], John Holland [k], Carlos E. Cardenas [l], Stephen Lai [e], David Rosenthal [a], Clifton David Fuller [a]

[a] *Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA*
[b] *Department of Radiation Oncology, UC Davis Comprehensive Cancer Center, Sacramento, California, USA*
[c] *Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, Massachusetts, USA*
[d] *Department of Radiation Oncology, Case Western Reserve University, Cleveland, Ohio, USA*
[e] *Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA*
[f] *Department of Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA*
[g] *Department of Electrical & Computer Engineering, University of Iowa, Iowa City, Iowa, USA*
[h] *Department of Computer Science, University of Illinois at Chicago, Chicago, Illinois, USA*
[i] *Department of Biostatistics, University of Minnesota of Public Health, Minneapolis, Minnesota, USA*
[j] *Department of Radiology, Oregon Health & Science University, Portland, Oregon, USA*
[k] *Department of Radiation Oncology, Oregon Health & Science University, Portland, Oregon, USA*
[l] *Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA*

ARTICLE INFO

Article history:
Received 13 July 2018
Accepted 31 July 2018
Available online 2 August 2018

Keywords:
Oral cavity cancer
Target delineation
Post-operative
Interdisciplinary
Interobserver

ABSTRACT

Background: The aim of this study was to determine the interdisciplinary agreement in identifying the post-operative tumor bed.
Methods: Three radiation oncologists (ROs), four surgeons, and three radiologists segmented post-operative tumor and nodal beds for three patients with oral cavity cancer. Specialty cohort composite contours were created by STAPLE algorithm implementation results for interspecialty comparison. Dice similarity coefficient and Hausdorff distance were utilized to compare spatial differentials between specialties.
Results: There were significant differences between disciplines in target delineation. There was unacceptable variation in Dice similarity coefficient for each observer and discipline when compared to the STAPLE contours. Within surgery and radiology disciplines, there was good consistency in volumes. ROs and radiologists have similar Dice similarity coefficient scores compared to surgeons.
Conclusion: There were significant interdisciplinary differences in perceptions of tissue-at-risk. Better communication and explicit description of at-risk areas between disciplines is required to ensure high-risk areas are adequately targeted.

Introduction

Intensity modulated radiotherapy (IMRT) is a conformal radiation technique that enables the generation of steep dose gradients within complex geometries [1]. The widespread adoption of this modality has resulted in improved dose sparing of organs at risk, ultimately resulting in improved delivery of tumoricidal dose and dose-toxicity profiles. A shift from traditional two-dimensional (2D) treatment to use highly conformal IMRT treatment has greatly reduced concurrent and late-onset toxicity sequelae. However, this problem continues to be a challenge as even minor variability in treatment setup and/or execution may result in significant under

* Corresponding author at: Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, Texas 77030 USA.

E-mail address: spng@mdanderson.org (S.P. Ng).

https://doi.org/10.1016/j.ctro.2018.07.006
2405-6308/

dosage of at-risk areas and/or over dosage of surrounding normal tissues [2].

Modern evaluation and treatment of head and neck cancer (HNC) patients typically involves, and is dependent upon, the collective, coordinated expertise of multidisciplinary care teams with heavy input from radiologic, surgical, and radiation oncologic specialties. The multidisciplinary input in patient care has become widely accepted as "best practice," having demonstrated measureable improvements in clinical quality indicators [3,4]. The teams are continuously challenged with effectively communicating at all stages of the process (diagnosis, staging, treatment, support, rehabilitation, and follow-up) to maximize benefit to the patient. While communication itself can be challenging, the process may be further complicated by differing use (or understanding) of specialty-specific vernacular and/or therapeutic decision making algorithms.

Furthermore, the radiation oncologist, particularly in a non-academic setting, is dependent upon the descriptive language of the surgeons and/or radiologists in the post-operative setting if adequate pre-operative imaging is not available for comparison. Given the potential for significant adverse patient outcomes based on a variable interdisciplinary understanding of fundamental radiation oncology treatment paradigms we consequently sought to investigate the variation in delineation of target volumes in post-operative HNC patients recommended to receive adjuvant radiation therapy by all parties involved in a typical case. We evaluated whether any discrepancy in nomenclature, particularly 'post-operative tumor bed' and 'post-operative nodal bed', between disciplines was present, necessitating the need to establish a standardized set of definitions. Fundamentally, we sought to determine whether, when specialist head and neck surgeons, radiologists and radiation oncologists discuss the "post-operative tumor bed" they were actually talking about the same spatial region; furthermore, did the 'surgical', 'radiological' and 'radiation oncological' post-operative tumor bed mean the same within a specialty to differing physicians?

This study is a prospective *in silico* human performance evaluation to identify and quantify the intradisciplinary and interdisciplinary observer variability in post-operative target volume delineation using a standardized case set and accepted spatial metrics as a surrogate for shared understanding of where radiation should be directed in high-risk cases.

Materials and Methods

Imaging and contouring of cases

Three standardized cases of patients with resected oral cavity cancers and recommended to receive post-operative radiation therapy (PORT) were selected for this study. Patients were randomly chosen from an Institutional Review Board (IRB) approved dataset and real patient data were extracted from patient records. The patients had a planning CT scan acquired (CTAqSim, Philips Medical Systems) without IV contrast utilizing immobilization devices, including tongue-depressing oral stent, and head and neck thermoplastic mask [5,6]. CT imaging was obtained from the vertex to the carina with 2 mm thickness slice thickness reconstruction [7]. Clinical information and planning CT images were anonymized and used for target delineation. After receiving a standardized set of explicit instructions [Appendix A], radiation oncologists (RO) specializing in HNC, head and neck surgeons (HNS), and head and neck radiologists (NR) contoured the post-operative tumor and nodal beds, respectively, using DICOM images in a commercial treatment planning/segmentation software (Pinnacle v9.0, Philips Medical Systems); expertise levels by specialty and years of experience are listed in Table 1. For users who were unfamiliar with the software interface (e.g. HNS and NR physicians, who do not use segmentation software daily), a skilled segmentation software user [BD/ASRM] was present throughout the initial contouring process to answer software interface questions only. Physicians were allowed access to all pertinent anonymized patient records including pre-operative imaging and any operative, pathologic or clinical note relevant to their task, excluding the actual delivered radiation treatment plan or planning notes. One radiation oncologist and one radiologist did not complete a majority (>50%) of contouring. Therefore their volumes were excluded from the analysis.

Comparison of volumes

Contour information was subsequently exported and analyzed using the EvaluateSegmentation program and metrics as described by Taha and Hanbury [8]. Contours were compared for agreement. The Warfield's simultaneous truth and performance level estimation (STAPLE) algorithm was used to generate a consensus contour representing the 'ground truth' volume. Warfield's STAPLE is an algorithm which incorporates multiple unordered and assumed independent segmentations to create an estimate of the hidden true segmentation, enabling characterization of the performance level of each observer [9]. The STAPLE volume was compared to each observer's volumes, allowing direct comparison between observers' volumes and the 'ground truth' volume [9].

The following metrics were included in the analysis:

1. Dice similarity coefficient (overlap based) – measures the similarity between two sets of segmentations and is calculated using the formula $DSC = \frac{2(A \cap B)}{(A+B)}$, [10]

   where A represents the observer dataset and B represents the STAPLE dataset [11] (Fig. 5).

2. Sensitivity and specificity (Information theoretic based) – Sensitivity, also known as the True Positive Rate (TPR), measures the voxels that are labeled positive by both the observer and STAPLE and is calculated by the formula: $Sensitivity = TPR = \frac{TP}{TP+FN}$ [8]

where true positive (TP) represents positive voxels in STAPLE and observer and false negative (FP) represents the positive voxels in observer segmentation but not in STAPLE.

Similarly, specificity, also known as True Negative Rate (TNR), measures the voxels that are labeled negative by both observer and STAPLE: $Specifciity = TNR = \frac{TN}{TN+FP}$ [8], where TN represents true negative, and FP represents false negative.

3. Hausdorff distance (Spatial distance based) measures the maximum distance between contours and is measured from one point in one set of segmentation to the closest point in another set of segmentation [8].

Table 1
Observers and their corresponding specialty and years of experience.

Specialty	Observer	Years of Experience
Radiation Oncology	RadOnc_A	4
	RadOnc_B	22
	RadOnc_C	2
Head and Neck Surgery	Surgeon_A	3
	Surgeon_B	10
	Surgeon_C	9
	Surgeon_D	5
Radiology	Radiologist_A	12
	Radiologist_B	5
	Radiologist_C	8