Radiotherapy and Oncology 119 (2016) 166–171

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Atlas-based segmentation

Assessment of fully-automated atlas-based segmentation of novel oral mucosal surface organ-at-risk

Jamie A. Dean [a,*], Liam C. Welsh [b], Dualta McQuaid [a], Kee H. Wong [b], Aleksandar Aleksic [b], Emma Dunne [b], Mohammad R. Islam [b], Anushka Patel [b], Priyanka Patel [b], Imran Petkar [b], Iain Phillips [b], Jackie Sham [b], Kate L. Newbold [b,c], Shreerang A. Bhide [b,c], Kevin J. Harrington [b,c], Sarah L. Gulliford [a], Christopher M. Nutting [b,c]

[a] Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust; [b] The Royal Marsden NHS Foundation Trust; and [c] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK

ARTICLE INFO

Article history:
Received 27 November 2015
Received in revised form 5 February 2016
Accepted 9 February 2016
Available online 9 March 2016

Keywords:
Atlas-based segmentation
Contouring
Delineation
OAR
Oral mucosa
Oral cavity

ABSTRACT

Background and purpose: Current oral mucositis normal tissue complication probability models, based on the dose distribution to the oral cavity volume, have suboptimal predictive power. Improving the delineation of the oral mucosa is likely to improve these models, but is resource intensive. We developed and evaluated fully-automated atlas-based segmentation (ABS) of a novel delineation technique for the oral mucosal surfaces.

Material and methods: An atlas of mucosal surface contours (MSC) consisting of 46 patients was developed. It was applied to an independent test cohort of 10 patients for whom manual segmentation of MSC structures, by three different clinicians, and conventional outlining of oral cavity contours (OCC), by an additional clinician, were also performed. Geometric comparisons were made using the dice similarity coefficient (DSC), validation index (VI) and Hausdorff distance (HD). Dosimetric comparisons were carried out using dose-volume histograms.

Results: The median difference, in the DSC and HD, between automated-manual comparisons and manual-manual comparisons were small and non-significant (−0.024; $p = 0.33$ and −0.5; $p = 0.88$, respectively). The median VI was 0.086. The maximum normalised volume difference between automated and manual MSC structures across all of the dose levels, averaged over the test cohort, was 8%. This difference reached approximately 28% when comparing automated MSC and OCC structures.

Conclusions: Fully-automated ABS of MSC is suitable for use in radiotherapy dose–response modelling.

Oral mucositis is a common and important toxicity of head and neck radiotherapy. It impacts on patients' quality of life [1], potentially causing pain, dysphagia [2–4] and consequential "late" effects [5–8]. It is frequently the limiting toxicity in dose–escalation and accelerated fractionation regimens that aim to improve tumour control [9–11]. Currently, normal tissue complication probability models have limited predictive performance and are not routinely used to aid clinical decision-making. Additionally, further evidence is required to find an optimal strategy for dose–sparing of the oral mucosa to reduce the incidence of severe toxicity.

In an attempt to improve the performance of oral mucositis normal tissue complication probability (NTCP) models developed by our group [12], we devised a novel contouring approach, which characterises the dose delivered to the mucosal surfaces of the oral cavity (MSC), including the buccal mucosa, mucosa of the lips and mucosa of the oral tongue [13]. We believe that this offers an improvement over the previously used oral cavity contours (OCC) volume (equivalent to the "extended oral cavity" structure in international consensus guidelines detailed in [14]), which predominantly describes the dose distribution to the musculature of the tongue and floor of mouth and does not incorporate the dose delivered to the buccal mucosa or mucosa of the lips. Differences in delineation guidelines have been shown to lead to differences in reported dose metrics and corresponding NTCP estimates [15]. To test whether our novel contouring approach improves NTCP modelling of oral mucositis, we must apply it to a large cohort of patients for whom we have mucositis outcome data.

* Corresponding author at: Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, Downs Road, Sutton SM2 5PT, UK.
E-mail address: jamie.dean@icr.ac.uk (J.A. Dean).

http://dx.doi.org/10.1016/j.radonc.2016.02.022
0167-8140/

Organ-at-risk (OAR) segmentation is highly time- and resource-intensive. This has motivated the development and evaluation of algorithms for automatic OAR segmentation [16,17]. The burden of OAR contouring can limit the feasibility of performing dose–response studies that make use of a large enough patient cohort (of the order of hundreds or thousands of patients) to enable strong statistical inference. This is especially true when the OAR of interest is not contoured as part of routine clinical practice and is challenging to delineate. This is certainly the case for MSC due to the relatively poor image contrast on planning CT scans and its complex shape. Being able to automate the MSC segmentation process would thus be of great benefit to oral mucositis dose–response modelling. It could also be valuable for use in treatment plan optimisation and assessment.

The aim of this study was to assess the performance of fully-automated (with no *post hoc* editing) atlas-based segmentation (ABS) of the MSC, in terms of geometry and dosimetry, in order to ascertain its suitability for use in dose–response modelling. The primary endpoint of our study was defined, prior to commencing the work, by *pre hoc* acceptability criteria, as follows: (i) if the geometric differences between the ABS-generated MSC (MSC_{ABS}) and manually delineated MSC (MSC_{manual}) did not exceed the inter-clinician variability and (ii) if the dosimetric differences between MSC_{ABS} and MSC_{manual} were smaller than those between the MSC_{ABS} and OCC (which is the current international standard) structures, then the MSC_{ABS} approach would be deemed suitable for dose–response modelling.

Materials and methods

Atlas construction

An MSC atlas of 46 patients, treated in the phase III trial of parotid-sparing intensity-modulated versus conventional radiotherapy in head and neck cancer (PARSPORT) (CRUK/03/005) [18], was generated from MSC_{manual} structures delineated on contrast-enhanced computed tomography (CT) scans by six clinical oncologists (L.W., E.D., R.I., P.P., I.Ph. and J.S.) using the RayStation, research version 4.6.100.12 treatment planning system (RaySearch Laboratories AB, Stockholm, Sweden). We have previously described the structure and contouring technique in detail [13]. Briefly, it includes the "buccal mucosa, buccal gingiva, gingiva proper, lingual gingiva, lingual frenulum, alveolar mucosa, labial mucosa, labial gingiva, labial frenulum, mucosal surface of the floor of mouth, mucosal surface of the tongue anterior to the terminal sulcus, and the mucosal surface of the hard palate". As described previously, the structures added to the atlas were the lines representing the positions of mucosal surfaces rather than the expanded 3 mm thick mucosal walls. The clinical oncologists received training in the novel contouring approach (from J.D. and L.W.) prior to commencing the study. Once complete, the MSC structures in the atlas were reviewed and, where necessary (7 out of 46 patients; incorrect delineation of the surface of the tongue, likely due to streak artefact), edited (by J.D. and C.N.). Other structures mentioned in this study were neither reviewed prior to the analysis nor edited at any time.

Fully-automated atlas-based segmentation

Fully-automated ABS of the MSC structure was applied to contrast-enhanced CT scans of 10 patients (first 10 patients treated at our institution with all data available) treated as part of the cochlear-sparing intensity-modulated radiotherapy versus conventional radiotherapy in patients with parotid tumours (COSTAR) phase III trial (CRUK/08/004). An isotropic 1.5 mm expansion from the ABS contours was performed to form a 3 mm thick wall structure, as previously described [13].

Manual segmentation by multiple clinicians

MSC_{manual} contouring was performed for the same 10 patients by each of three clinical oncologists (A.A., A.P. and I.Pe.). When performing the contouring, the clinicians were blinded to the MSC_{manual} structures contoured by the other clinicians and the MSC_{ABS} structures. The contoured mucosal surface lines were expanded to a 3 mm thick wall as previously described [13]. The clinical oncologists received training in the contouring technique prior to commencing the study. Manual OCC segmentation was performed for the same 10 patients by a clinical oncologist (K.W.) to enable dosimetric comparison between the new MSC_{ABS} structure and the conventionally used OCC structure. The OCC structure is based on international consensus guidelines and is equivalent to the "extended oral cavity" OAR described in [14].

Comparison of automated and manual segmentation

In-house software was written to extract the structure coordinates from RayStation and perform comparisons of the different structures using the Python programming language version 2.7.9 [19] and the NumPy version 1.9.2 [20], SciPy version 0.16.0 [21], MatPlotLib version 1.4.3 [22] and PyDicom version 0.9.9 [23] modules.

A geometric comparison was performed using the dice similarity coefficient (DSC) [24], validation index (VI) [15] and Hausdorff distance (HD) [25]. The DSC describes the amount of agreement between two volumes, V and S, and is given by

$$\mathrm{DSC} = 2\frac{|V \cap S|}{|V| + |S|} \tag{1}$$

The VI is a recently designed measure, for geometric comparison of automated and multiple manually contoured structures, that attempts to account for uncertainties in the manual contouring [26].

$$\mathrm{VI} = \sum_{k=1}^{N}\left(\frac{k^{\alpha}|V_k|}{\sum_{j=1}^{N} j^{\alpha}|V_j|}\right)\left(\frac{k}{N}\right)\left(2\frac{|V_k \cap S_k|}{|V_k| + |S_k|}\right) \tag{2}$$

where V_k is the volume of overlap between k experts out of a total of N experts, S is the whole automated segmentation and α is a control parameter (allowing for the weighting term (first bracket in Eq. (2)) to be changed to meet specific radiotherapy treatment planning requirements in terms of how conservative the segmentation should be), which was set to 1. When $\alpha = 1$ the first bracket in Eq. (2) represents the normalised frequency at which the different proportions of agreeing clinicians for a volume (second bracket in Eq. (2)) occur. VI is 0 if the ABS has no overlap with the manual structures and 1 if the ABS and all manual structures perfectly overlap. The HD describes the maximum of all of the distances from each point in one structure to the closest point in the other structure.

The means of the DSC and HD values for the pairwise comparison between MSC_{ABS} and each of the three MSC_{manual} structures ($DSC_{pw,ABS}$ and $HD_{pw,ABS}$) were calculated for each patient. The means of the DSC and HD values for the pairwise comparisons between the different manually contoured structures ($DSC_{pw,man}$ and $HD_{pw,man}$) were also calculated for each patient and these value subtracted from the $DSC_{pw,ABS}$ and $HD_{pw,ABS}$ values for comparison ($DSC_{pw,diff}$ and $HD_{pw,diff}$). A two-tailed Wilcoxon signed-rank test was used to test for statistical significance.

A dosimetric comparison was carried out using fractional dose–volume histograms (DVHs). The differences in the normalised volumes receiving each dose level between the MSC_{ABS} and each