Published in final edited form as:
Radiother Oncol. 2016 April ; 119(1): 166–171. doi:10.1016/j.radonc.2016.02.022.

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], and **Christopher M Nutting**[b,c]

[a]Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, UK, SM2 5NG

[b]The Royal Marsden NHS Foundation Trust, Fulham Road, London, UK, SW3 6JJ

[c]Division of Radiotherapy and Imaging, The Institute of Cancer Research, Fulham Road, London, UK, SW3 6JJ

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.

Corresponding author: Jamie Dean, Mailing address: Joint Department of Physics at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, SM2 5PT, UK, Telephone: +442089156223, Fax: +442086433812, jamie.dean@icr.ac.uk.

Europe PMC Funders Author Manuscripts

Keywords

Atlas-based segmentation; contouring; delineation; organ-at-risk; OAR; oral mucosa; oral cavity; mucositis

Europe PMC Funders Author Manuscripts

Introduction

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][Dean et al. (submitted to Radiother Oncol)], 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.

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

A 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 (LW, ED, RI, PP, IPh and JS) 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 muocsa, 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 JD and LW) 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 JD and CN). 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 (AA, AP and IPe). 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 (KW) 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

$$DSC = 2\frac{|V \cap S|}{|V| + |S|} \quad \text{Equation 1}$$

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].

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

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 equation 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 equation 2 represents the normalised frequency at which the different proportions of agreeing clinicians for a volume (second bracket in equation 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 of the OCC and MSC_{manual} structures were measured and compared.

Results

In all cases the ABS was able to segment the MSC without any gross errors. Figure 1 shows an example of the ABS and manual MSC contours for patient 4. The greatest variation between the structures is in the position of the posterior border, the lateral extents of the buccal mucosa and the inferior extent of the mucosa where the lateral tongue base meets the mucosa overlying the floor of mouth. This was representative of the whole 10-patient cohort. The ABS performed poorest, as assessed by DSC, HD and VI, for patient 2. In this case the MSC_{ABS} structure extended too far inferiorly beyond where the lateral tongue border meets the floor of mouth.

The geometric comparisons are described in table 1. The DSC increases with increasing overlap between structures. The HD decreases with increasing proximity of structures. The mean values of $DSC_{pw,diff}$ and $HD_{pw,diff}$ were small in magnitude and not close to statistically significant at the $p = 0.05$ level. This indicates that the geometric differences between the MSC_{ABS} and MSC_{manual} structures did not exceed the geometric differences between MSC_{manual} structures delineated by different clinicians. Therefore, our *pre hoc* criterion (i) was met. The VI values were substantially lower than the $DSC_{pw,ABS}$ values. This is because $DSC_{pw,ABS}$ generally overestimates the agreement between ABS and manual structures due to not incorporating the amount of agreement between multiple clinician-delineated structures.

Figure 2 summarises the DVHs for the 10-patient cohort using the different structures. Figure 3 shows the pairwise differences in the DVHs between the different structures for each of the 10 patients. The median MSC_{ABS} DVH falls within the centre of the median MSC_{manual} DVH range across the three clinicians (AA, AP, IPe). The maximum normalised volume differences at any dose level were, on average across the cohort, within 8%. The dose distributions extracted using the OCC structure characterise larger volumes receiving low and intermediate doses than any of the automated or manual MSC structures. The maximum normalised volume difference between the MSC_{ABS} and OCC structure, average across the cohort, was 28%. This indicates that the dosimetric differences between the MSC_{ABS} and MSC_{manual} structures were small and substantially smaller than the differences between the MSC_{ABS} and OCC structures. Therefore, our *pre hoc* criterion (ii) and, hence, the primary endpoint of our study was met. Moreover, the MSC_{ABS} (and MSC_{manual}) structures capture information on the dose delivered to the buccal mucosa, which is not captured by the OCC structure (figure 1).

Discussion

The results of the geometric comparison between the MSC_{ABS} and MSC_{manual} structures indicate that the difference between fully-automated and manual segmentation of the MSC structure is within the inter-clinician variability of the manual delineation. The same is true of the dosimetric comparison between the MSC_{ABS} and MSC_{manual} structures. Furthermore, the dosimetric comparison demonstrates that the DVHs for the MSC_{ABS} structure are more similar to the “gold standard” MSC_{manual} structures than the OCC structure that was previously employed for dose-response modelling. Based on our *pre hoc* criteria, the

primary endpoint of our study was met. We, therefore, suggest that the MSC_{ABS} structure is suitable for use for oral mucosa dose-response modelling.

The DSC and VI values are low for the comparison of the MSC_{ABS} and MSC_{manual} structures and the inter-clinician comparisons. This is a result of the nature of the morphology of the structure being wall-like. DSC is highly sensitive to the volume of the structure. The same is also true of the VI metric. This can be illustrated using a “toy” example. Two cubes, of dimension 10 units, diagonally displaced, by 1 unit in each dimension, have a DSC of 0.729. Two hollow cubes, of dimension 10 units and thickness 1 unit, with the same displacement, have a substantially lower DSC of 0.098. This makes it challenging to determine whether an automated contouring approach is suitable for clinical use solely by using a threshold value of the DSC or VI. The current “gold-standard” for OAR contouring is manual delineation by a trained expert [27]. However, variability exists between delineations performed by different experts, even when following the same guidelines [28]. Therefore, there exists a rationale for deciding whether an automated approach is fit for clinical use based on whether or not it falls within or outside the variability of clinical oncologists who would perform the manual delineation for clinical use. Several studies evaluating the performance of automated structure segmentation simply use the DSC magnitude value to assess suitability for clinical use. We suggest that our method, in which we compared the differences in automated and manually segmented structures to inter-clinician variability in the manual segmentation, is, at least, equally valid.

The median $HD_{pw,diff}$ value provides further indication that the automated segmentation fell within the range of the inter-clinician variability in manual segmentation. Unlike the DSC and VI values, the HD values are not sensitive to the volumes of the structures. Using the same “toy” example, the two displaced cubes have a HD of 1.73 units and the two displaced hollow cubes also have a HD of 1.73 units. This suggests that HD is a more suitable metric for assessing the similarity between structure segmentations than DSC for small volume or wall-like structures.

The dosimetric comparison data (figures 2 and 3) indicate that the MSC_{ABS} structure provides a dose distribution to the oral mucosal surfaces that is closer to the MSC_{manual} structures than the previously used OCC structure. This provides further rationale for using the MSC_{ABS} structure, rather than the OCC structure, for dose-response modelling. Many other studies evaluating the performance of automated structure segmentation consider only geometric indices and do not directly measure the effects of geometric differences on the dose distribution to the segmented structure. We suggest that dosimetric comparisons can also provide useful, and often more relevant, information in performing such evaluations, particularly for small or wall-like structures.

A potential limitation of our study is the use of six different clinicians to perform the delineation of the structures included in the atlas. This may have led to increased variability in the atlas MSC structures compared with what might have been produced had a single clinician performed all of the delineations. The fact that the VI values were substantially lower than the $DSC_{pw,ABS}$ values indicates that there was uncertainty in the manual delineation of the MSC structure. This is likely due to the relatively poor CT image contrast

and complex shape of the structure. We trained the clinicians performing the delineation and independently reviewed and edited the contours in an attempt to minimise the effects of inter-clinician variability in the atlas. However, some variation inevitably remained. This could explain the poorer performance of the ABS for patient 2. The MSC for the atlas patient selected in the initial rigid registration comparison step of the ABS may also have extended too far inferiorly. Multi-atlas-based segmentation (using information from all of the available patients, rather than just the one that is the closest match, to perform the automated segmentation) might improve the robustness of the automated segmentation (at the expense of computational time) [29], but we were unable to assess this, as RayStation does not currently include a multi-atlas segmentation algorithm. We would expect this approach to reduce the sensitivity of the ABS to the atlas patient selected and, therefore, improve the ABS performance for patient 2.

Moreover, it should be noted that the patients included in this study did not have tumours of the oral cavity. Therefore, based on our data, we cannot guarantee that the performance of the automated or manual segmentation of the MSC for patients with oral cavity tumours will match that reported in this study. The cohort that we intend to apply this technique to for NTCP modelling does not contain patients with oral cavity tumours. Including patients for whom the MSC include tumour could confound investigations of the dose-response relationship of normal oral mucosa. Furthermore, it is unlikely that the MSC could be spared if it overlapped with the planning target volume, reducing the utility of segmenting the MSC.

When applying the MSC structure to NTCP modelling, it is important to consider that scoring of oral mucositis also takes into account mucosal surfaces not included as part of the MSC structure, particularly the oropharynx. For this reason, when applying the MSC approach to NTCP modelling of oral mucositis, we recommend that the pharyngeal mucosa also be considered as an OAR.

In conclusion, we performed a thorough geometric and dosimetric assessment of fully-automated ABS of the novel MSC OAR structure and demonstrated that it is suitable for use in radiotherapy dose-response studies. This represents the first evaluation of a method to fully segment the oral mucosal surfaces automatically. In the future, we aim to apply this contouring approach to a cohort of patients from six head and neck radiotherapy trials, previously described [Dean et al. (submitted to Radiother Oncol)], and establish whether it improves the predictive performance of NTCP modelling of severe acute mucositis.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council, Cancer Research UK Programme Grant A13407 and NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and ICR. The PARSPORT and COSTAR trials were supported by Cancer Research UK (trial reference numbers CRUK/03/005 and CRUK/08/004). We wish to thank Hannah Eyles, Emma Wells, Dr Emma Hall at The Institute of Cancer Research Clinical Trials and Statistics Unit, Dr Cornelis Kamerling, Dr Alex Dunlop, Dr Simeon Nill and Prof Uwe Oelfke for general support.

Europe PMC Funders Author Manuscripts

References

1. Kelly C, Paleri V, Downs C, Shah R. Deterioration in quality of life and depressive symptoms during radiation therapy for head and neck cancer. Otolaryngol Head Neck Surg. 2007; 136:108–11. DOI: 10.1016/j.otohns.2006.06.1278 [PubMed: 17210344]
2. Schwartz DL, Hutcheson K, Barringer D, Tucker SL, Kies M, Holsinger FC, et al. Candidate dosimetric predictors of long-term swallowing dysfunction after oropharyngeal intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys. 2010; 78:1356–65. DOI: 10.1016/j.ijrobp.2009.10.002 [PubMed: 20646872]
3. Sanguineti G, Gunn GB, Parker BC, Endres EJ, Zeng J, Fiorino C. Weekly dose-volume parameters of mucosa and constrictor muscles predict the use of percutaneous endoscopic gastrostomy during exclusive intensity-modulated radiotherapy for oropharyngeal cancer. Int J Radiat Oncol Biol Phys. 2011; 79:52–9. DOI: 10.1016/j.ijrobp.2009.10.057 [PubMed: 20418027]
4. Sanguineti G, Rao N, Gunn B, Ricchetti F, Fiorino C. Predictors of PEG dependence after IMRT ±chemotherapy for oropharyngeal cancer. Radiother Oncol. 2013; 107:300–4. DOI: 10.1016/j.radonc.2013.05.021 [PubMed: 23773408]
5. Denham JW, Peters LJ, Johansen J, Poulsen M, Lamb DS, Hindley A, et al. Do acute mucosal reactions lead to consequential late reactions in patients with head and neck cancer? Radiother Oncol. 1999; 52:157–64. [PubMed: 10577701]
6. Sonis ST. Mucositis: The impact, biology and therapeutic opportunities of oral mucositis. Oral Oncol. 2009; 45:1015–20. DOI: 10.1016/j.oraloncology.2009.08.006 [PubMed: 19828360]
7. Bentzen SM. Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. Nat Rev Cancer. 2006; 6:702–13. [PubMed: 16929324]
8. Martin M, Lefaix J-L, Delanian S. TGF-β 1 and radiation fibrosis: a master switch and a specific therapeutic target? Int J Radiat Oncol. 2000; 47:277–90. DOI: 10.1016/S0360-3016(00)00435-1
9. Jackson SM, Weir LM, Hay JH, Tsang VH, Durham JS. A randomised trial of accelerated versus conventional radiotherapy in head and neck cancer. Radiother Oncol. 1997; 43:39–46. [PubMed: 9165135]
10. Maciejewski B, Skladowski K, Pilecki B, Taylor JM, Withers RH, Miszczyk L, et al. Randomized clinical trial on accelerated 7 days per week fractionation in radiotherapy for head and neck cancer. Preliminary report on acute toxicity. Radiother Oncol. 1996; 40:137–45. [PubMed: 8884967]
11. Bourhis J, Overgaard J, Audry H, Ang KK, Saunders M, Bernier J, et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. Lancet. 2006; 368:843–54. DOI: 10.1016/S0140-6736(06)69121-6 [PubMed: 16950362]
12. Otter S, Schick U, Gulliford S, Franceschini D, Newbold K, Nutting C, et al. Evaluation of the risk of grade 3 oral and pharyngeal dysphagia using atlas-based methodology and multivariate analyses of individual patient dose distributions. Int J Radiat Oncol. 2015; doi: 10.1016/j.ijrobp.2015.07.2263
13. Dean JA, Welsh LC, Gulliford SL, Harrington KJ, Nutting CM. A novel method for delineation of oral mucosa for radiotherapy dose–response studies. Radiother Oncol. 2015; :10–3. DOI: 10.1016/j.radonc.2015.02.020 [PubMed: 26026485]
14. Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. Radiother Oncol. 2015; 117:83–90. DOI: 10.1016/j.radonc.2015.07.041 [PubMed: 26277855]
15. Brouwer CL, Steenbakkers RJHM, Gort E, Kamphuis ME, Van Der Laan HP, Van'T Veld AA, et al. Differences in delineation guidelines for head and neck cancer result in inconsistent reported dose and corresponding NTCP. Radiother Oncol. 2014; 111:148–52. DOI: 10.1016/j.radonc.2014.01.019 [PubMed: 24560759]
16. Hardcastle N, Tomé W a, Cannon DM, Brouwer CL, Wittendorp PW, Dogan N, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. Radiat Oncol. 2012; 7:90.doi: 10.1186/1748-717X-7-90 [PubMed: 22704464]

17. Zhu M, Bzdusek K, Brink C, Eriksen JG, Hansen O, Jensen HA, et al. Multi-institutional quantitative evaluation and clinical validation of Smart Probabilistic Image Contouring Engine (SPICE) autosegmentation of target structures and normal tissues on computer tomography images in the head and neck, thorax, liver, and male. Int J Radiat Oncol Biol Phys. 2013; 87:809–16. DOI: 10.1016/j.ijrobp.2013.08.007 [PubMed: 24138920]
18. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol. 2011; 12:127–36. DOI: 10.1016/S1470-2045(10)70290-4 [PubMed: 21236730]
19. Python Software Foundation. [accessed July 1, 2015] Python n.d. http://www.python.org
20. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Comput Sci Eng. 2011; 13:22–30. DOI: 10.1109/MCSE.2011.37
21. McKinney W. Data Structures for Statistical Computing in Python. Proc 9th Python Sci Conf. 2010; 1697900:51–6.
22. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007; 9:90–5. DOI: 10.1109/MCSE.2007.55
23. Mason, D. [accessed July 1, 2015] Pydicom n.d. https://code.google.com/p/pydicom/
24. Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology. 1945; 26:297–302. DOI: 10.2307/1932409
25. Dubuisson M, Jain AK. A Modified Hausdorff Distance for Object Matching. Int Conf Pattern Recognit. 1994; :566–8. DOI: 10.1109/ICPR.1994.576361
26. Juneja P, Evans PM, Harris EJ. The validation index: a new metric for validation of segmentation algorithms using two or more expert outlines with application to radiotherapy planning. IEEE Trans Med Imaging. 2013; 32:1481–9. DOI: 10.1109/TMI.2013.2258031 [PubMed: 23591482]
27. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. Med Phys. 2014; 41:050902.doi: 10.1118/1.4871620 [PubMed: 24784366]
28. Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol. 2012; 7:32.doi: 10.1186/1748-717X-7-32 [PubMed: 22414264]
29. Aljabar P, Heckemann Ra, Hammers a, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage. 2009; 46:726–38. DOI: 10.1016/j.neuroimage.2009.02.018 [PubMed: 19245840]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

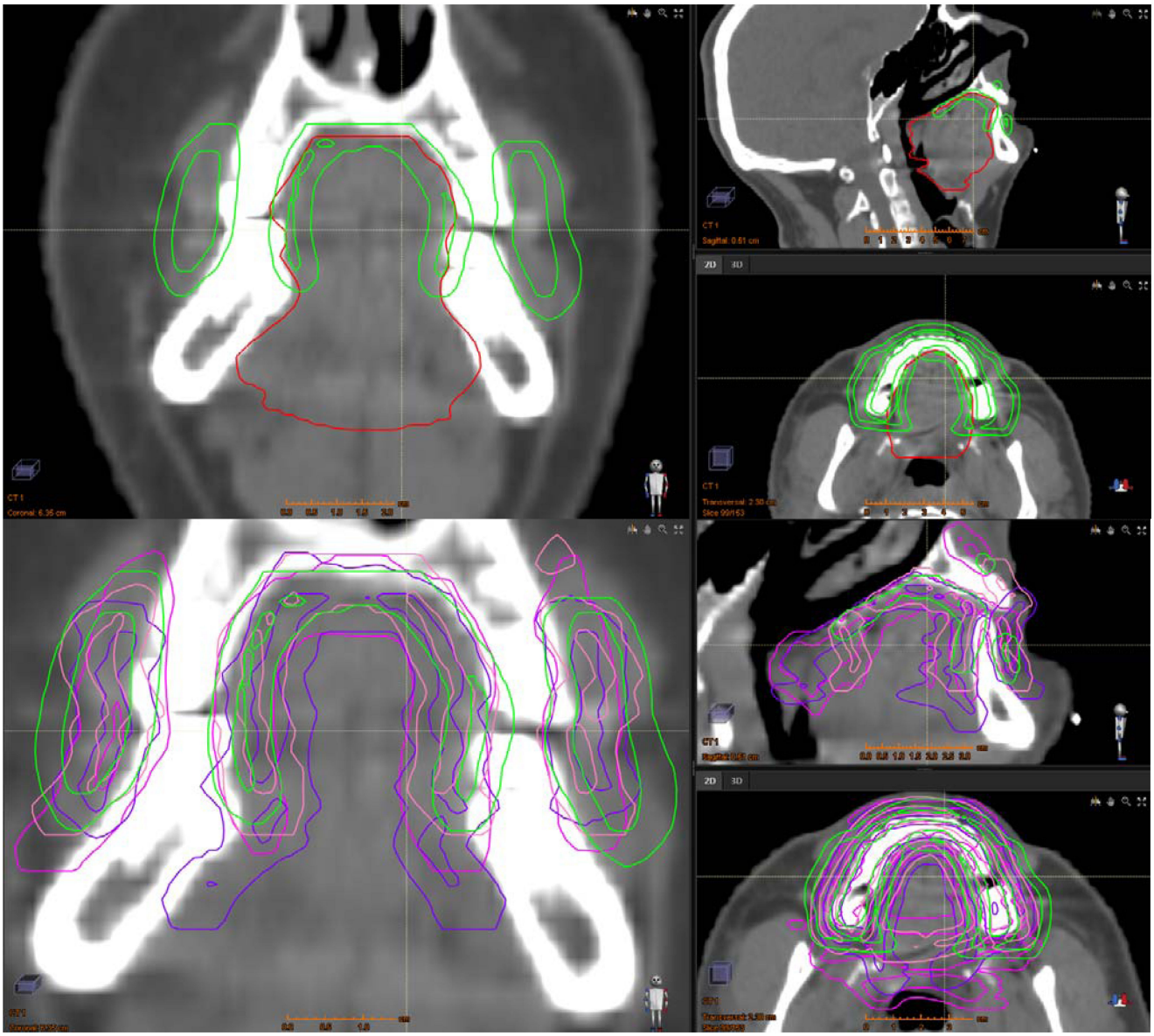

Figure 1.

Example (patient 4) of the fully automated ABS-generated MSC structure (top and bottom; green), the manually delineated OCC structure (top; red) and three different manually delineated MSC structures (bottom; pink, purple and coral). For each set of 3 images (top and bottom) the left, top right and bottom right images are coronal, sagittal and axial views, respectively.

Europe PMC Funders Author Manuscripts

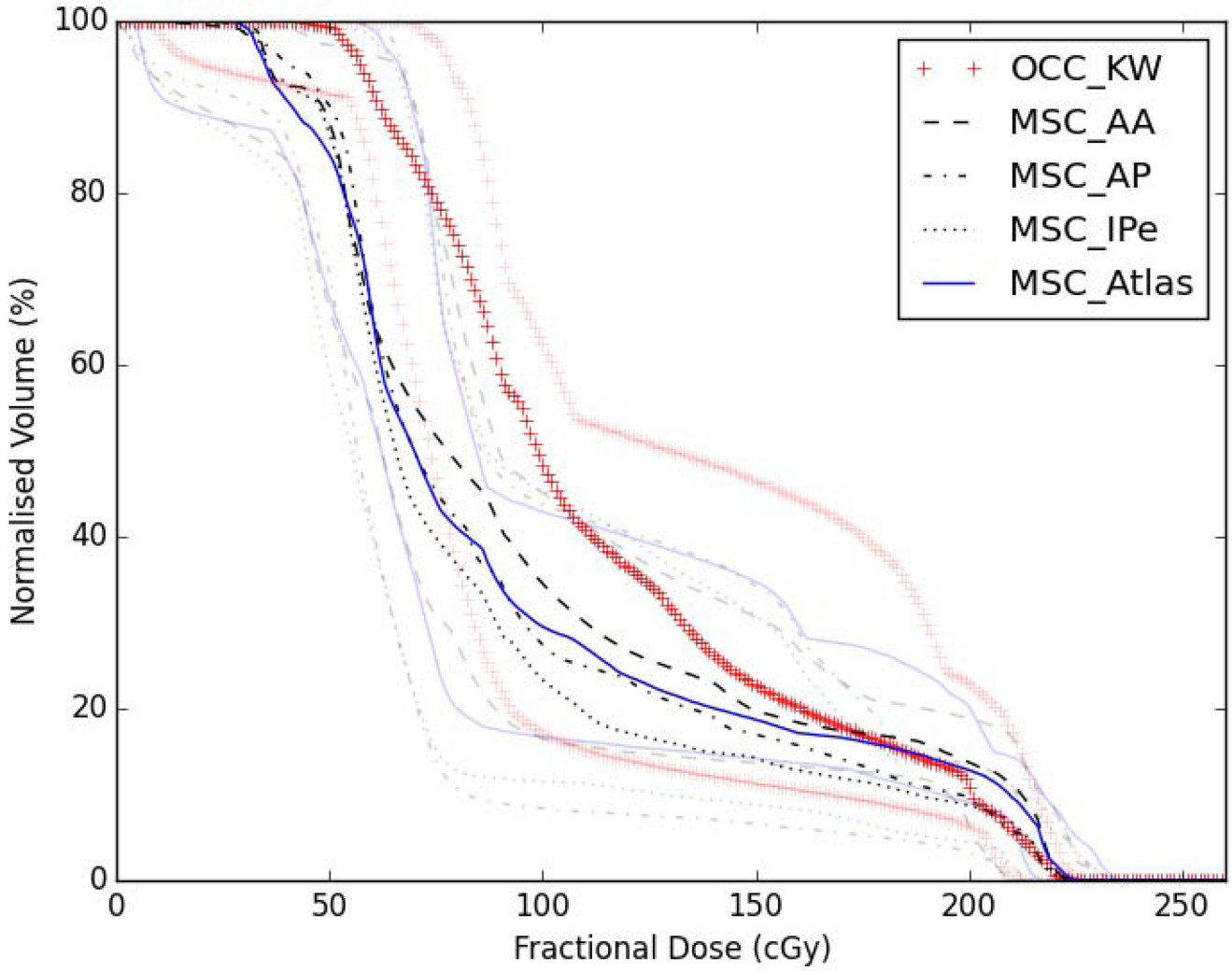

Figure 2.

Fractional dose-volume histograms. Black lines show group medians and grey lines show group ranges. OCC_KW – oral cavity contours (by KW); MSC_AA – manual mucosal surface contours (by AA); MSC_AP – manual mucosal surface contours (by AP); MSC_IPe – manual mucosal surface contours (by IPe); MSC_Atlas – atlas-based segmentation mucosal surface contours.

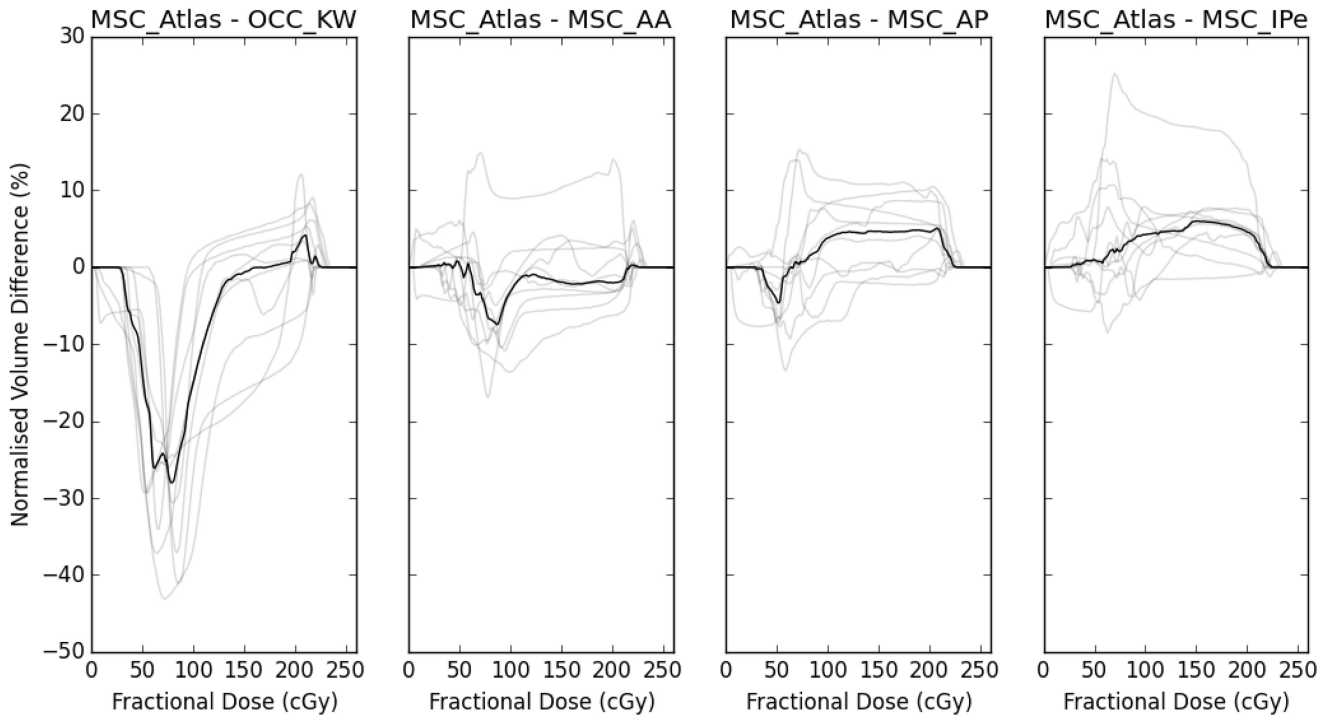

Figure 3.

Normalised volume differences, at each dose level, between atlas-based segmentation mucosal surface contours and oral cavity contours and each of the three sets of manually delineated mucosal surface contours. Black lines show medians and grey lines show the individual patients values. OCC_KW – oral cavity contours (by KW); MSC_AA – manual mucosal surface contours (by AA); MSC_AP – manual mucosal surface contours (by AP); MSC_IPe – manual mucosal surface contours (by IPe); MSC_Atlas – atlas-based segmentation mucosal surface contours.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Geometric comparison of the automated and manual MSC contours.

Patient	VI	$DSC_{pw,ABS}$	$DSC_{pw,man}$	$DSC_{pw,diff}$	$HD_{pw,ABS}$ (mm)	$HD_{pw,man}$ (mm)	$HD_{pw,diff}$ (mm)
1	0.033	0.070	0.134	−0.064	21.0	16.1	4.9
2	0.025	0.062	0.269	−0.207	21.3	14.8	6.5
3	0.089	0.210	0.159	0.052	13.3	14.4	−1.1
4	0.152	0.327	0.349	−0.022	13.7	13.6	0.1
5	0.115	0.270	0.352	−0.082	11.5	15.4	−3.9
6	0.124	0.283	0.309	−0.026	13.1	16.1	−3.0
7	0.063	0.162	0.111	0.051	12.5	17.2	−4.7
8	0.056	0.137	0.098	0.040	17.8	21.0	−3.2
9	0.105	0.243	0.272	−0.029	17.4	14.8	2.6
10	0.082	0.185	0.204	−0.018	16.7	13.4	3.3
Median	0.086	0.198	0.236	−0.024	15.2	15.1	-0.5
				($p^{*} = 0.332$)			($p^{*} = 0.881$)

VI = validation index; $DSC_{pw,ABS}$ = mean of the dice similarity coefficients for pairwise comparisons between the automatically segmented structure and each of the three manually delineated structures; $DSC_{pw,man}$ = mean of the dice similarity coefficients for pairwise comparisons between the three different manually delineated structures (3 different combinations of pairwise comparisons); $DSC_{pw,diff}$ = $DSC_{pw,ABS}$ minus $DSC_{pw,man}$; $HD_{pw,ABS}$ = mean of the Hausdorff distances for pairwise comparisons between the automatically segmented structure and each of the three manually delineated structures; $HD_{pw,man}$ = mean of the Hausdorff distances for pairwise comparisons between the three different manually delineated structures (3 different combinations of pairwise comparisons); $HD_{pw,diff}$ = $HD_{pw,ABS}$ minus $HD_{pw,man}$;

*p-value for two-tailed Wilcoxon signed-rank test.