

Performance evaluation of head and neck contour adaptation with cone beam CT using two commercial software systems

LOUISE S. H. BENDALL^{1,2,3}, MARIA NAJIM¹, RACHEL STENSMYR¹, EMILY FLOWER^{1,3}, SHAN YAU^{1,2,4}, DAVID I. THWAITES³ & JONATHAN R. SYKES^{1,2,3,4}

¹Department of Medical Physics, Crown Princess Mary Cancer Centre, Westmead Hospital, Westmead, Sydney, NSW, Australia, ²Department of Medical Physics, Nepean Cancer Care Centre, Nepean Hospital, Sydney, NSW, Australia, ³Institute of Medical Physics, School of Physics, University of Sydney, Sydney, NSW, Australia and ⁴Department of Medical Physics, Blacktown Cancer Care Centre, Blacktown Hospital, Sydney, NSW, Australia

Correspondence: L. Bendall, CR-UK/MRC Institute for Radiation Oncology, Department of Oncology, University of Oxford, Radiobiology Research Institute, Churchill Hospital, Oxford OX3 7LE, UK. E-mail: louise.bendall@oncology.ox.ac.uk

(Received 15 May 2015; accepted 26 June 2015)

ISSN 0284-186X print/ISSN 1651-226X online © 2015 Informa Healthcare
DOI: 10.3109/0284186X.2015.1068448

To the Editor,

Radiotherapy is one of the main treatment techniques for head and neck cancer and is generally delivered in daily fractions, five times a week over 6–7 weeks [1]. During this time, there is often a change in anatomy due to weight loss, tumour shrinkage or organ motion/deformation. Previous studies have found a reduction in parotid gland volumes of 16–24% [2,3]. Several studies have investigated the use of adaptive radiotherapy (ART), during which the treatment plan can be adapted to account for changes seen during the treatment [4,5]. However, ART can be a time consuming activity and as such difficult to implement into a busy clinical department due to increased workload [6]. Deformable image registration (DIR) is a tool which could potentially improve the efficiency of ART. DIR determines the spatial correspondence between points in a reference image [the planning computed tomography (CT)] and the target image (the replan CT). This allows the contours originally delineated on the planning CT to be deformed to the new anatomy in the rescan CT.

The use of additional CT scans for ART is well documented [4,5,7]. In future, cone beam CT (CBCT) scans which are acquired routinely for image guidance at the time of beam delivery could be used to complement, or replace, a second CT scan. This would involve the recalculation of dose and dose-volume histograms (DVH) on CBCTs as part of the process to decide if and when plan adaptation is required in order to meet initial plan goals [4,5,7]. The aim of this study was to measure the accuracy of DIR transfer of contours from a planning CT to both a rescan CT and a CBCT with two commercial DIR systems.

Material and methods

Twenty head and neck radiotherapy patients who were previously treated at our institution were randomly selected for this retrospective study. The treatment sites were predominantly early stage node positive oropharynx and nasopharynx (Supplementary Material, Supplementary Table I available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>). All patients were treated radically, with a prescribed dose of 66–70 Gy in 33–35 fractions.

Each patient had a planning (CT1) scan and a prospective rescan (CT2) approximately half way through their 30–33 fraction treatment, at an average of 46 days after CT1 as per clinical protocol (both CTs were acquired with IV contrast). Each patient underwent a CBCT scan (without IV contrast) approximately once a week throughout their treat-

ment. Scanning parameters and further information can be found in the Supplementary Material available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>.

The organs at risk (OAR) (listed in Table I) were contoured on the CT1, CT2 and CBCT scans by a single radiation oncologist (RO) in the EclipseTM Treatment Planning System (Varian Medical Systems, Palo Alto, CA, USA). The images and contours were then exported to the DIR software. The RO's contours were taken as the benchmark, allowing comparison with the deformed contours. To ensure consistency and avoid variability due to multiple observers, the same RO performed all contouring. To assess the intra-observer variability in contouring, five randomly selected patients were re-contoured by the same RO at least four months after the initial contouring took place.

Two commercially available DIR software packages were used in this study to determine the deformation between CT1 and CT2, and CT1 and CBCT and deform the structures accordingly. These were MIMTM Maestro (Version 6.4.3, MIM Software Inc., Cleveland, OH, USA) (MIM) and Velocity AITM (Version 3.1.0, Varian Medical Systems).

CT1 was registered deformably to CT2, and the RO contours from CT1 transferred to CT2 and compared to the RO's contours delineated on CT2. This was performed independently on both software systems and repeated using CBCT.

The RO and DIR produced contours were compared using the volume ratio (ratio of RO volume/DIR produced volume, VR), maximum, mean and standard deviation of Hausdorff distance (HD, which is an unsigned distance to agreement), signed distance to agreement (DTA) and normalised Dice similarity coefficient (nDSC) [8,9]. Details of DTA and nDSC

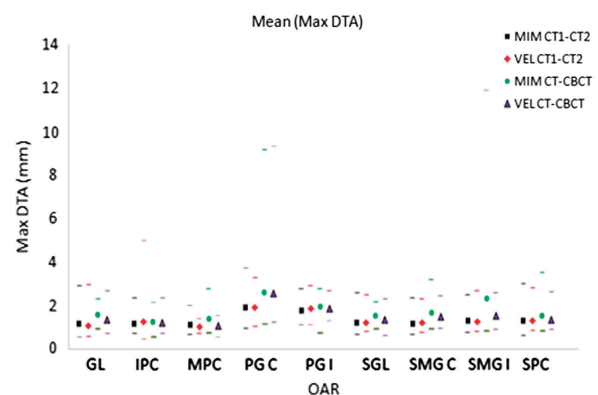


Figure 1. Mean of max DTA results for single DIR for each organ. Maximum and minimum results also shown by corresponding markers.

Table I. Calculated values of the standard deviation of DTA for intra-RO (σ_{RO}) and for DIR (σ_{DIR}) calculated from Equation 2 of Supplementary Material, for all patients. As σ_{DIR} was generally larger than σ_{RO} , this demonstrated that DIR was not considered a success.

OAR	DTA Result/mm					
	RO CT	MIM CT1-CT2	VEL CT1-CT2	RO CBCT	MIM CT-CBCT	VEL CT-CBCT
Glottic larynx (GL)	1.6	2.1	2.7	1.9	2.1	2.9
Inferior pharyngeal constrictors (IPC)	1.1	1.9	2.2	1.5	1.8	1.8
Middle pharyngeal constrictors (MPC)	2.0	1.6	2.1	1.2	2.2	2.4
Contralateral parotid gland (PG C)	2.1	2.7	2.9	1.6	3.8	3.5
Ipsilateral parotid gland (PG I)	2.3	2.2	3.1	2.5	2.4	2.5
Supraglottic larynx (SGL)	2.1	1.3	2.4	1.8	1.9	2.5
Contralateral submandibular gland (SMG C)	1.6	1.7	2.4	1.6	3.5	2.8
Ipsilateral submandibular gland (SMG I)	2.0	1.8	2.2	1.6	4.0	2.8
Superior pharyngeal constrictors (SPC)	1.6	1.4	2.1	1.3	2.1	2.0
Mean	1.8	1.9	2.5	1.6	2.6	2.6

Calculated values of σ_{RO} and σ_{DIR} calculated from Equation 2 (Supplementary Material) for all patients. As σ_{DIR} was generally larger than σ_{RO} , this demonstrated that DIR was not considered a success.

calculation can be found in the Supplementary Material available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>.

Results

Fifteen of the 20 patients lost 2 kg or more during their treatment (mean -5% , range of $-16 - +3\%$), leading to clinically significant anatomical changes (Supplementary Table II, available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>). Large, statistically significant ($p < 0.05$) decreases in volume were seen for both parotid glands and submandibular glands.

Variation between repeat OAR delineation by the same RO was small for most OARs (Supplementary Material, Supplementary Table III, available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>). The VR ranged from (0.5–1.4) for CT2 and (0.6–1.4) for CBCT. The val-

ues of HD ranged between 2.7 and 15.6 mm for CT2 and 3.0 and 17.4 mm for CBCT. While the increased range of VR and DTA for CBCT was expected due to the decrease in image quality inherent in CBCT compared to CT, a paired t-test showed there was no statistically significant difference in mean values ($p < 0.05$) between the CT2 and CBCT results for individual OARs.

The mean of the standard deviation of DTA (σ_{RO}) in RO contouring (overall OARs and all patients) was calculated to be 1.8 mm for the CT and 1.6 mm for CBCT (Table I).

The standard deviation of DTA for DIR-produced contours (σ_{DIR}) was similar to that seen for the variation in RO contouring (σ_{RO}), although slightly larger in most cases (Figure 1, Table I). Across the two software systems DTA (σ_{DIR}) for Velocity AI was found to be greater than for MIM with CT-CT ($p < 0.05$, paired Wilcoxon signed rank), but there was no significant difference with CT-CBCT.

The mean nDSC was found to be 0.94 and 0.86 for MIM and Velocity AI, respectively, for CT-CT DIR, with overall 34% and 7% of contours having $nDSC \geq 1$ (Supplementary Table IV, available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>). The percentage of individual OARs with $nDSC \geq 1$ was 26% and 29% for CT1-CT2 and CT1-CBCT respectively for MIM, whilst for Velocity AI this was 11% and 19%. MIM had a statistically significantly higher nDSC than Velocity AI ($p < 0.05$, paired Wilcoxon signed rank) for CT-CT but there was no significant difference for CT-CBCT (Figure 2).

Across all DIR techniques, the range of standard deviation of volume, DSC and DTA were 0.0–4.3 cm^3 , 0.0–0.2 and 0.0–7.5 mm respectively, for the five

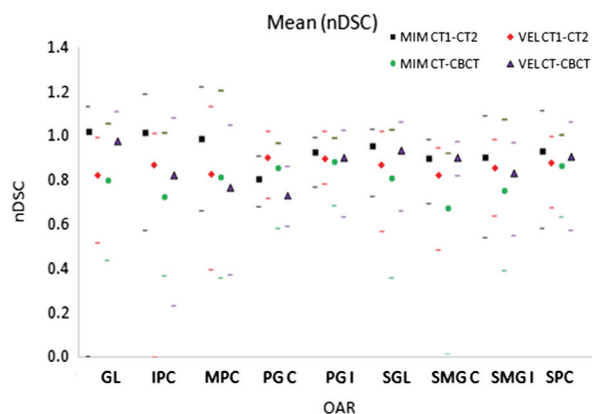


Figure 2. Mean of nDSC results for single DIR. Maximum and minimum results also shown by corresponding markers.

image sets that had DIR performed five independent times.

Discussion

In this study, manual contouring of OARs has been used to measure the uncertainties in the DIR for the MIM and Velocity AI systems using both CT and CBCT images. The DIR techniques were automated to investigate the implementation of automated DIR using a commercial registration system. To our knowledge, no other study has investigated this. Previous studies have discussed the errors associated with contouring by ROs [10,11] and efforts were taken to minimise this by using a single RO. In terms of the metrics used in this study, Zidenbos et al. found a DSC > 0.7 to indicate good agreement [12] while Walker et al. and Mattiucci et al. considered a DSC > 0.85 and DSC > 0.8 to be clinically useful [6,13]. For the pairs of RO contours on CT2 and CBCT, 84% and 91%, respectively, had a DSC of 0.7 or greater, whereas this dropped to just 36% at a DSC of 0.85 for both CT2 and CBCT. However, the calculation of nDSC which factors in the uncertainty in RO contouring aimed to give some clinical value to the metrics obtained.

Based on both nDSC and DTA measurements it is clear that DIR does not perform within RO contouring uncertainty for most OARs. Further investigation in terms of examination of contours by the RO would be required to determine whether the results of DIR using the two software packages would be clinically acceptable.

A limitation of this study was that only OAR contours were available to assess DIR accuracy as the study was performed as part of a clinical study only requiring the organs listed in Table I. Ideally, GTV and CTV contours would have also been available.

For adaptive planning with CT, the intended purpose of DIR would be to automatically contour structures on CBCT scans for which dose has been calculated using the current treatment plan and to then assess DVHs for acceptability against the plan objectives. With highly conformal dose distributions small contouring errors could have a large impact on the decision or conversely large errors may have only a small impact, as seen by Brouwer et al., who reported large NTCP differences exceeding 10% for a small minority of patients when investigating different contouring guidelines.

A limitation of this study was that the contour comparison metrics did not indicate whether there would have been a different decision on whether to adapt or not comparing the RO and DIR contours. This is the subject of planned further investigation.

In conclusion, for the two DIR systems evaluated in this study, automatic contours adapted from CT1 to CT2 using DIR were found to have an uncertainty of similar size to the inherent intra-clinician uncertainty, for the majority of OARs. MIM was found to have smaller variations than Velocity AI for CT1 to CT2 adaption. The MIM system was found to have increased variations for CT1-CBCT adaption which were similar in size to variations found with Velocity AI. These results suggest that automated DIR has the potential for use for transforming contours from CT1-CT2 with some editing required, but may be limited for determining whether replanning is required using CBCT images.

Acknowledgements

The authors would like to thank those in Sydney West Radiation Oncology Network for their practical help with using the DIR systems.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- [1] 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.
- [2] Castadot P, Lee JA, Geets X, Grégoire V. Adaptive radiotherapy of head and neck cancer. *Semin Radiat Oncol* 2010;20:84–93.
- [3] Hansen EK, Bucci MK, Quivey JM, Weinberg V, Xia P. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2006;64:355–62.
- [4] Schwartz DL, Garden AS, Shah SJ, Chronowski G, Seipal S, Rosenthal DI, et al. Adaptive radiotherapy for head and neck cancer – dosimetric results from a prospective clinical trial. *Radiation Oncol* 2013;106:80–4.
- [5] Castelli J, Simon A, Louvel G, Henry O, Chajon E, Nassef M, et al. Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia. *Radiat Oncol* 2015;10:1–10.
- [6] Mattiucci GC, Boldrini L, Chiloiri G, D'Agostino GR, Chiesa S, De Rose F, et al. Automatic delineation for replanning in nasopharynx radiotherapy: What is the agreement among experts to be considered as benchmark? *Acta Oncol* 2013;52:1417–22.
- [7] Bishop M, Greskovich J. Adaptive replanning of IMRT for head and neck cancer: A case report of replanning in a middle-aged patient with squamous cell carcinoma of the tonsil. *Appl Radiat Oncol* 2012;1:30–2.
- [8] Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26:297–302.
- [9] Ezhil M, Choi B, Starkschall G, Bucci MK, Vedam S, Balter P. Comparison of rigid and adaptive methods of propagating gross tumor volume through respiratory phases

- of four-dimensional computed tomography image data set. *Int J Radiat Oncol Biol Phys* 2008;71:290–6.
- [10] Parvathaneni U, Laramore GE, Liao JJ. Technical advances and pitfalls in head and neck radiotherapy. *J Oncol* 2012; 2012:1–13.
- [11] Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008;33: 136–40.
- [12] Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: Method and validation. *Med Imaging* 1994;13:716–24.
- [13] Walker GV, Awan M, Tao R, Koay EJ, Boehling NS, Grant JD, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. *Radiother Oncol* 2014;112:321–5.

Supplementary material available online

Supplementary Figure 1, Tables I–V available online at <http://informahealthcare.com/doi/abs/10.3109/0284186X.2015.1068448>
