

Prostate contour variability in an MR-only prostate SBRT workflow

M Davidson^{1,2}, A Shaaer³, J Stewart¹, D Vesprini^{1,2}, J Detsky^{1,2}, A Loblaw^{1,2}, W Chu^{1,2}, H Chung^{1,2}, S Liu^{1,2}, J Shahi, M Wronski^{1,2}, M Ruschin^{1,2} and C McCann^{1,2}

¹ Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto Ontario Canada

² Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada

³ Department of Physics, Ryerson University, Toronto, Ontario, Canada

INTRODUCTION

The advent of MR simulation (MRsim) and MR-Linac (MRL) delivery systems have made it possible to plan and guide radiotherapy treatments using MR only.

Historically, staff in a radiotherapy department are well versed in using CT images for contouring, treatment planning and image guidance. While MRI provides superior soft tissue visualization over CT, there remains a relative unfamiliarity in the interpretation of MRI images compared to CT.

With the installation of an MRL unit at our centre where daily online re-contouring on T2-weighted MR images is mandated for prostate SBRT, it is important to understand contouring variability within an expert team of genito-urinary (GU) radiation oncologists (RO).

AIM

Quantify inter-observer variability in prostate contouring on images from a 1.5T MRsim and two 1.5T MRL online adaptive prostate SBRT sequences, including 2-minute (MRL-2min) and 6-min scans (MRL-6min).

METHODS

- Prostate was contoured on 3 T2-weighted MRs (MRsim, MRL-2min, MRL-6min), for 5 patients by 7 GU-RO.
- For each image set, the STAPLE contour was generated for comparison. STAPLE is a probabilistic estimate of the “true” contour, derived from all individual RO contours.
- Inter-observer agreement was quantified against the STAPLE contour using:
 - Dice Similarity Coefficient (DSC)
 - Mean Distance-to-Agreement (MDA)
 - Hausdorff distance (HD)
 - Kappa coefficient (κ).
- Directional differences were quantified by computing 3D distance between centroids.
- Paired t-tests between image pair results done.

RESULTS

Across the cohort, there were no statistically significant differences in mean similarity metrics (summarize in Table 1):

- Mean κ was similar for all scans ($p>0.5$), with values approaching 1 indicating strong inter-observer agreement.¹
- Mean DSC values approaching 1 further indicate good concordance between RO and STAPLE contours regardless of scan ($p>0.05$ for each pairwise-comparison).

Since DSC can sometimes be insensitive to local discrepancies, mean (MDA) and maximum (HD) distances from STAPLE were also scored.

- No difference in mean MDA between MRsim & MRL scans ($p>0.05$)
- Small improvement in MDA favoring MRL-2min (1.5mm) over MRL-6min (1.9mm), $p<0.01$.
- HD similar between STAPLE and RO contours ($p>0.25$, with a 6.6mm overall average). However, HD values were variable (range: 2.6-25.7mm), indicating large local differences in some contours.

To better understand where local differences were occurring:

- 3D distance between individual RO and STAPLE contour centroids calculated (mean distance in each direction shown in Figure 1)
- Most variation seen in:
 - cradio-caudal direction (3.1mm averaged over all scan types)
 - anterior-posterior (2.4mm averaged over all scan types), with majority of differences occurring towards prostate base & apex.

Qualitative observations:

- Some observers systematically contoured larger or smaller prostate volumes (most obviously in the cranio-caudal direction).
- Contouring strategies may be rooted in CT-based contouring practice, where more generous contour boundaries are commonly drawn.
- Review of prostate MR-contouring consensus guidelines², with consultation from GU radiologists may help harmonize RO practice

Metric	MRsim	MRL-2min	MRL-6min
DSC	0.91 ± 0.04	0.89 ± 0.04	0.91 ± 0.06
MDA (mm)	1.7 ± 0.6	1.5 ± 0.7	1.9 ± 0.8
HD (mm)	6.3 ± 2.4	6.4 ± 2.1	7.1 ± 4.7
κ	0.90 ± 0.02	0.90 ± 0.02	0.89 ± 0.03

Table 1. Summary of similarity metric given as means \pm standard deviation

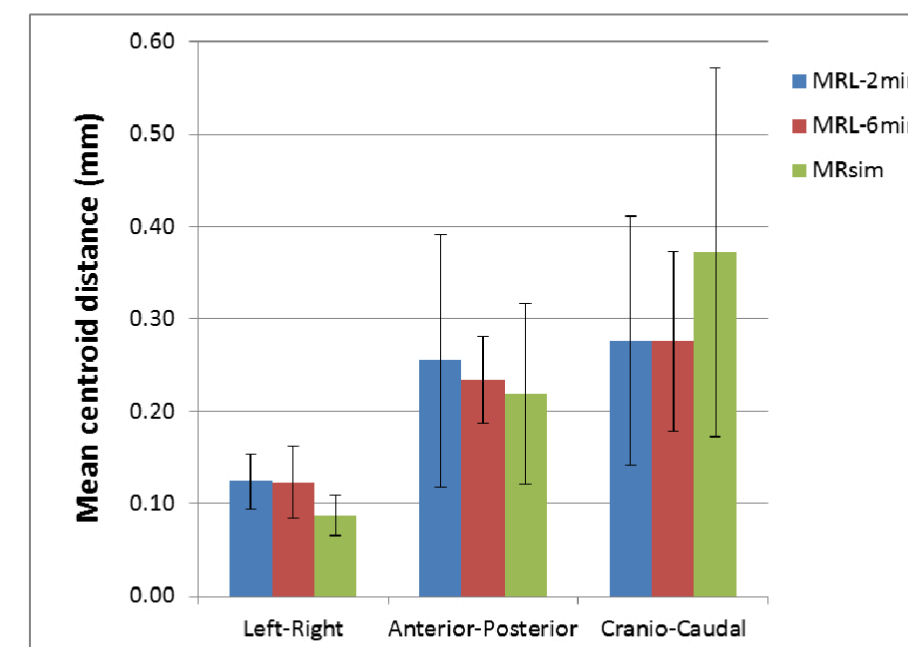


Figure 1. Mean distance between RO and STAPLE contour centroids in the 3 cardinal planes (error bars = standard deviation)

CONCLUSIONS

Similarity analyses indicate that the degree of contouring variability was comparable overall between all image sets. This suggests that:

- Contouring is similar on MRL and MRsim images
- The MRL-2min sequence is appropriate for use in an online adaptive radiotherapy workflow, with no benefit seen by using the longer MRL-6min scan.

Lack of uniformity amongst GU-RO participants, particularly towards the longitudinal extents of the prostate, indicates the need to for education structured around prostate contouring on MR that could include:

- Review of consensus guidelines² and other relevant literature
- Interactive review of prostate anatomy on MRI and practical contouring session with GU radiologists

Contouring variability in the MR-only workflow may have implications on planning margins:

- Larger local differences in contouring variability observed for individual patients > 4 mm (current PTV margin).

REFERENCES

- McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-282.
- Salembier C et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. Radiotherapy and Oncology 2018; 127:49-61.

CONTACT INFORMATION

melanie.davidson@sunnybrook.ca