



Contents lists available at ScienceDirect

Technical Innovations & Patient Support in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/technical-innovations-and-patient-support-in-radiation-oncology



Interobserver variation of clinical oncologists compared to therapeutic radiographers (RTT) prostate contours on T2 weighted MRI

Gillian Adair Smith^{a,*}, Alex Dunlop^c, Sophie E. Alexander^b, Helen Barnes^a, Francis Casey^c, Joan Chick^c, Ranga Gunapala^d, Trina Herbert^a, Rebekah Lawes^a, Sarah A. Mason^c, Adam Mitchell^c, Jonathan Mohajer^c, Julia Murray^b, Simeon Nill^c, Priyanka Patel^b, Angela Pathmanathan^b, Kobika Sritharan^b, Nora Sundahl^b, Rosalyn Westley^b, Alison C. Tree^b, Helen A. McNair^b

^a The Royal Marsden NHS Foundation Trust, London, United Kingdom

^b The Institute of Cancer Research/The Royal Marsden NHS Foundation Trust, London, United Kingdom

^c Joint Department of Physics at the Royal Marsden and The Institute of Cancer Research, United Kingdom

^d Clinical Trials and Statistic Unit, The Institute for Cancer Research, London, United Kingdom

ARTICLE INFO

Keywords:

Prostate radiotherapy
MR-Linac
Radiographer contouring

ABSTRACT

The implementation of MRI-guided online adaptive radiotherapy has enabled extension of therapeutic radiographers' roles to include contouring. An offline interobserver variability study compared five radiographers' and five clinicians' contours on 10 MRIs acquired on a MR-Linac from 10 patients. All contours were compared to a "gold standard" created from an average of clinicians' contours. The median (range) DSC of radiographers' and clinicians' contours compared to the "gold standard" was 0.91 (0.86–0.96), and 0.93 (0.88–0.97) respectively illustrating non-inferiority of the radiographers' contours to the clinicians. There was no significant difference in HD, MDA or volume size between the groups.

Introduction

The implementation of online MRI-guided adaptive radiotherapy requires an increase in number of staff to be present at time of treatment compared to conventional radiotherapy [1,2]. To utilise the staff effectively, some roles routinely performed by the clinical oncologist are being delegated to other staff members, for instance therapeutic radiographers (RTTs/radiographers) [3]. When evaluating contours outlined online on an MR-Linac, a comparison of the single online contour with a 'gold standard' or expert is undertaken. This process could include comparison with the expert contour performed offline or review offline of the online contour by the expert [4]. In either case the process is very dependent on the expert. Prior to undertaking online MR-Linac contouring, we evaluated the offline contouring of five clinicians and five radiographers, all of whom were experienced in MR-Linac treatment. The aim was to provide a baseline of an acceptable range of contours created by the clinicians to compare with radiographers. We hypothesized that if the radiographer contours were within the range of

the clinician contours, then online contouring would be acceptable.

Materials and methods

The interobserver variability of prostate and seminal vesicles (SV) contours was assessed offline. Five radiographers and five clinicians, with one to three years of MR-Linac experience, contoured the prostate and SV on a treatment planning system (Monaco treatment planning system, Version: 5.59.02 Research, Elekta, Stockholm, Sweden) on 10 T2-weighted MRIs acquired on an Elekta Unity MRL (Elekta, Stockholm, Sweden), from 10 patients. Two clinical target volumes (CTV) were created using each of the 100 contours:

High dose CTV: prostate plus 1 cm proximal SV.

Low dose CTV: prostate plus 2 cm proximal SV.

A simultaneous truth and performance level estimation (STAPLE) algorithm generated structures in ADMIRE (Version: Research 2.0 Elekta AB, Stockholm, Sweden) using the five clinicians' contours to create "gold standard" high dose CTV and low dose CTV structures [5]. Fifty

* Corresponding author at: The Royal Marsden, Downs Road, Sutton, Surrey SM2 5PT, United Kingdom.

E-mail address: gillian.smith1@rmh.nhs.uk (G. Adair Smith).

<https://doi.org/10.1016/j.tipsro.2022.12.007>

Received 14 September 2022; Received in revised form 22 November 2022; Accepted 21 December 2022

Available online 30 December 2022

2405-6324/© 2022 The Author(s). Published by Elsevier B.V. on behalf of European Society for Radiotherapy & Oncology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

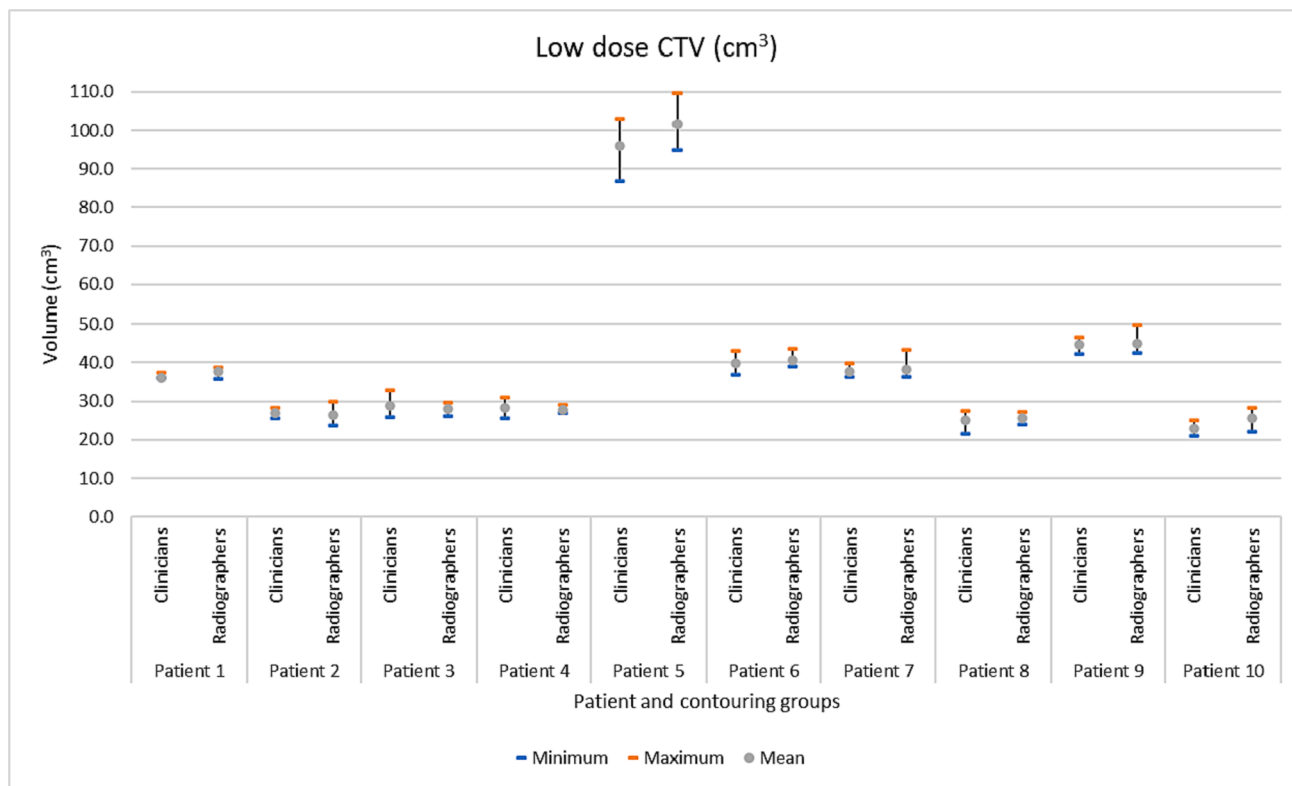
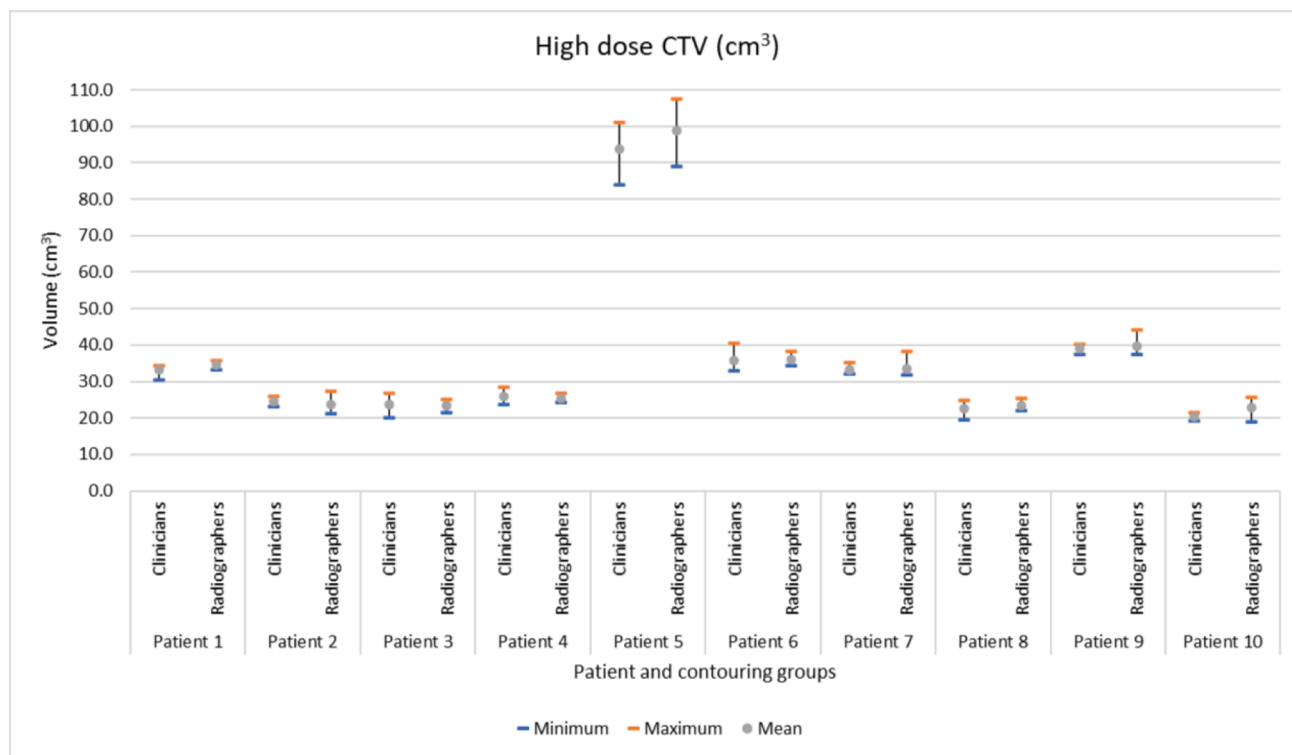


Fig. 1. (a). Summary of 100 volumes for the radiographer and clinician contoured high dose CTV for ten patients’ images (five radiographer and five clinician contours per patient). Fig. 1b. Summary of 100 volumes for the radiographer and clinician contoured low dose CTV for ten patients’ images (five radiographer and five clinician contours per patient).

high dose CTV and fifty low dose CTV structures were created from both clinicians’ and radiographers’ contours. Each radiographer and clinician structure was then compared with the gold standard structure using Dice similarity coefficient (DSC), mean distance to agreement (MDA) and Hausdorff distance (HD). DSC is an overlap metric, such that 0 equals no

overlap and 1 equals perfect overlap. MDA and HD are similarity metrics, with MDA measuring the mean distance comparable points on two surfaces would need to move to overlap and HD describing the maximum distance between two comparable points. Since we expect the radiographer contours to have slightly lower DSC than clinicians,

Table 1

Dice similarity coefficient (DSC), mean distance to agreement (MDA) and Hausdorff distance (HD) for radiographer and clinician contours compared to the 'gold standard' STAPLE contour. Results presented as median (range).

	DSC		MDA (mm)		HD (mm)	
	Clinicians	Radiographers	Clinicians	Radiographers	Clinicians	Radiographers
High dose CTV	0.93 (0.88–0.97)	0.91 (0.86–0.96)	0.8 (0.4–1.9)	0.9 (0.5–1.6)	5.3 (3.2–9.9)	4.8 (2.9–9.5)
Low dose CTV	0.92 (0.86–0.97)	0.91 (0.83–0.96)	0.9 (0.8–1.0)	0.9 (0.4–1.5)	5.3 (3.2–10.0)	5.1 (3.3–12.2)

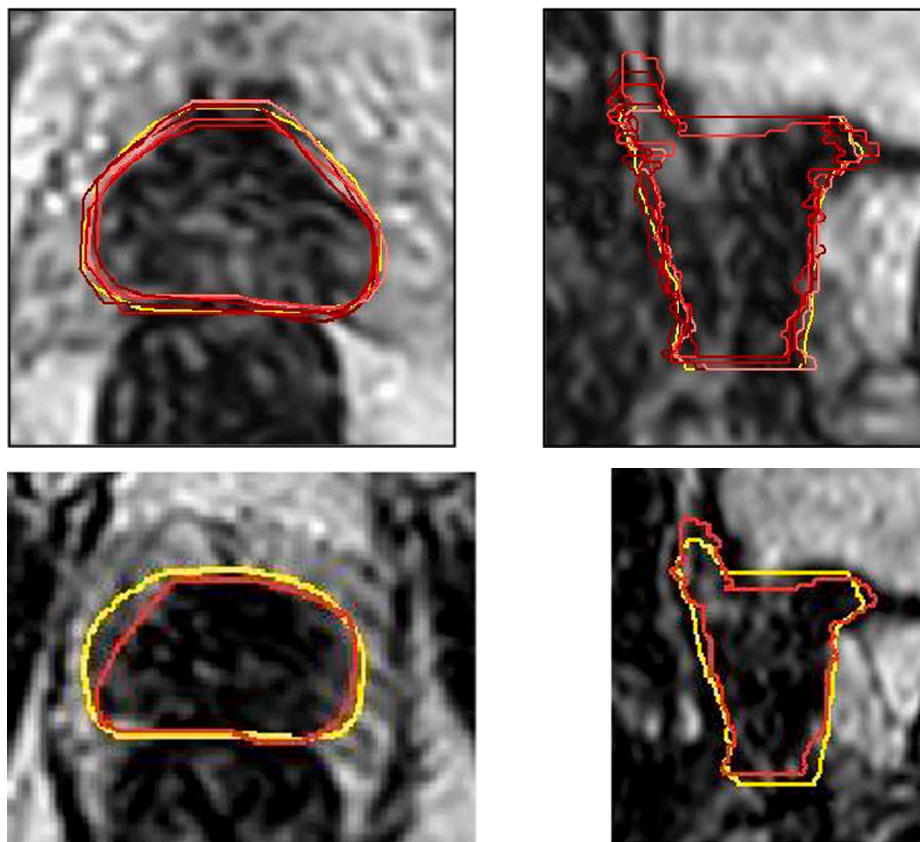


Fig. 2. (a) Five radiographers' contours (red) superimposed onto 'gold standard' (yellow) contour illustrating discrepancies at base and apex. Median Dice similarity coefficient (DSC) = 0.95. **Fig. 2b** - Worst case: Radiographer contour (red) superimposed onto 'gold standard' (yellow) contour illustrating discrepancies at base and apex DSC = 0.83.

because of the bias introduced by defining the gold standard as the average contour of the clinicians, we tested non-inferiority based on the radiographers being within 0.025 DSC of the clinicians. This corresponded to a reduction of the median clinicians' contour from 0.93 to 0.905, a DSC threshold we previously determined to result in acceptable plans [6]. The mean distance to agreement (MDA) and Hausdorff distance (HD) were compared using a Mann Whitney *U* test. The volumes of radiographer and clinician contours were compared using a Mann Whitney *U* test.

Results

High dose CTV; Of the 50 high dose CTV radiographer structures, 30 volumes were within the clinicians' volume range (Fig. 1a), six smaller (1% to 8%) and 14 larger (1% to 19%), and were not significantly different ($p = 0.60$). There was no significant difference in the HD or MDA, between the clinician and radiographers' contours when compared to gold standard ($p = 0.84$ and $p = 1$ respectively) (Table 1). The DSC when comparing the radiographers' contours to the gold

standard was significantly different (lower) than the DSC from clinicians' contours and gold standard ($p = 0.018$). However, the DSC from radiographers and gold standard was significantly higher than the DSC from the 'clinicians and gold standard - 0.025', indicating that the radiographers were non-inferior to the clinicians' contours ($p = 0.017$).

Low dose CTV; Of the 50 low dose CTV radiographer structures, 35 were within the clinicians' volume range (Fig. 1b). Three were smaller (1% to 7.5%) and 12 were larger (1% to 13%) and were not significantly different ($p = 0.67$).

The radiographers' DSC and gold standard were significantly higher than the DSC between the 'clinicians and gold standard - 0.025', indicating that the radiographers were non-inferior to the clinicians' contours ($p = 0.031$). There was no significant difference in the HD or MDA between the clinicians and radiographers when compared to gold standard ($p = 0.75$ and $p = 0.155$ respectively).

On visual inspection (see Fig. 2), variations at the base and apex of CTV contributed to the largest difference in contour volume and the lowest DSC was 0.83 overall (low dose CTV).

Discussion

Of the 200 high and low dose CTVs, 60% and 70% of the radiographer contours respectively were within the clinicians' range. The contours which were smaller and would result in a target miss, hence of greater concern, were < 8% smaller, equivalent to 4 cm³ of the high dose CTV and 2.6 cm³ of the low dose CTV. The larger contours, although ensuring the target was encompassed, may increase the dose to organs at risk (OARs). Although statistically significant inter-clinician variability in prostate contour volumes were found in a study investigating inter-observer variation of five clinicians, the outcome with respect to the irradiated volume of the rectum and bladder was not clinically relevant [7]. The greatest agreement in contours at the prostate/bladder and prostate/rectum interfaces was thought to be the reason. Although we observed discrepancies at the prostate base as well as the apex, we found no significant difference in the volumes of the contours by five radiographers and five clinicians. It has been noted when comparing plans generated using gold standard contours versus auto segmented contours [8] or when comparing 25 observers contouring from one CT image [9] that a poor DSC did not necessarily result in poor dose coverage. When comparing plans generated using gold standard contours versus auto segmented contours the mean \pm SD DSC was 0.87 \pm 0.03, but the D98 \pm 2% and V95% for prostate and its 3 mm expansion were within 2% (3 Gy) of each other, except for one case with a DSC of 0.82.

The variations at base and apex were the greatest and may have contributed to the significant difference found when comparing the DSC raw data (Fig. 2). There is an inherent bias towards the clinicians' contours when using a gold standard created from the clinicians' contours. By creating a threshold for an acceptable DSC of 0.905, a threshold we previously determined resulted in acceptable plans [6], we found that the DSC from the radiographers' contours were significantly greater, illustrating good agreement. Contouring on MRI is known to result in less inter-observer variability than CT [10] and our DSC threshold of 0.905 is higher than studies using CT.

Delineation errors are recognised as a potential source of error in the conventional radiotherapy pathway [11]. Incorporating this task into the online workflow with multiple observers has the potential to increase this risk [12]. Training programmes have been shown to improve consistency but had not been completed by the radiographers in this study and may improve the results further [13]. The effect of the variations on dose to the target and OARs has not been established in this study and may not be clinically relevant, as in the Livesey study [7], however this needs to be established.

Conclusion

The majority (>60%) of radiographers' contours were within the range of clinicians' contours. Although the DSC and gold standard comparisons between the two groups were significantly different there was no significant difference in other metrics used, including a non-inferiority test between DSC. Clinical relevance measures, for example dose, should be investigated alongside contour comparisons. Training programmes may improve agreement.

Notes

The author(s) confirm that written informed consent has been obtained from the involved patient(s) or if appropriate from the parent, guardian, power of attorney of the involved patient(s); and, they have given approval for this information to be published in this case report (series).

Declaration of Competing Interest

The authors declare the following financial interests/personal

relationships which may be considered as potential competing interests: Dr Helen A McNair reports financial support was provided by National Institute for Health Research and Health Education England. Alison Tree, Angela Pathmanathan, Rosalynne Westley reports a relationship with Elekta Ltd. Alison Tree reports a relationship with Accuray Inc. Alison Tree reports a relationship with Varian Medical Systems Inc. Alison Tree, Sophie Alexander reports a relationship with Cancer Research UK that. Research at The Institute of Cancer Research is also supported by Cancer Research UK under Programme C33589/A28284 and C7224/A28724.

Acknowledgements

Helen McNair, Senior Clinical Lecturer, ICA-SCL-2018-04-ST2-002 is funded by the NIHR for this research project. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care.

Research at The Institute of Cancer Research is also supported by Cancer Research UK under Programme C33589/A28284 and C7224/A28724.

The Institute of Cancer Research and The Royal Marsden Hospital receive funding for research from Elekta Ltd within the MOMENTUM trial and are a member of the Elekta MR Linac consortium.

Alison Tree acknowledges research funding from Elekta Ltd, Accuray and Varian and Cancer Research UK Radnet Grant C7224/A28724.

Angela Pathmanathan and Rosalynne Westley acknowledge research funding from Elekta Ltd.

Sophie Alexander is funded by Cancer Research UK Programme Grant C33589/A28284 - Adaptive Data-driven Radiation Oncology.

References

- [1] Lamb J, Cao M, Kishan A, Agazaryan N, Thomas DH, Shaverdian N, et al. Online adaptive radiation therapy: implementation of a new process of care. *Cureus* 2017; 9(8).
- [2] McNair HA, Joyce E, O'Gara G, Jackson M, Peet B, Huddart RA, et al. Radiographer-led online image guided adaptive radiotherapy: A qualitative investigation of the therapeutic radiographer role. *Radiography* 2021;27(4): 1085–93.
- [3] Willigenburg T, de Muinck Keizer DM, Peters M, Claes A, Legendijk JJ, de Boer HC. Evaluation of daily online contour adaptation by radiation therapists for prostate cancer treatment on an MRI-guided linear accelerator. *Clin Translat Radiat Oncol* 2021;27:50–6.
- [4] Willigenburg T, et al. Evaluation of daily online contour adaptation by radiation therapists for prostate cancer treatment on an MRI-guided linear accelerator. *Clin Translat Radiat Oncol* 2021;27:50–6. <https://doi.org/10.1016/j.ctro.2021.01.002>.
- [5] Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004;23(7):903–21.
- [6] Pathmanathan, A et al. Dosimetric comparison of propagated and radiographer contours for the MR-Linac' [Manuscript submitted for publication]; 2022.
- [7] Livesey JE, Wylie JP, Swindell R, Khoo VS, Cowan RA, Logue JP. Do differences in target volume definition in prostate cancer lead to clinically relevant differences in normal tissue toxicity? *Int J Radiat Oncol* Biol* Phys* 2004;60(4):1076–81.
- [8] Kawula M, Purice D, Li M, Vivar G, Ahmadi SA, Parodi K, et al. Dosimetric impact of deep learning-based CT auto-segmentation on radiation therapy treatment planning for prostate cancer. *Radiat Oncol* 2022;17(1):1–12.
- [9] Barghi A, Johnson C, Warner A, Bauman G, Battista J, Rodrigues G. Impact of contouring variability on dose-volume metrics used in treatment plan optimization of prostate IMRT. *Cureus* 2013 Nov 5;5(11).
- [10] Rasch C, Barillot I, Remeljeer P, Touw A, van Herk M, Lebesque JV. Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys* 1999; 43(1):57–66.
- [11] Van Herk M. Errors and margins in radiotherapy. *Semin Radiat Oncol* 2004;14(1): 52–64. <https://doi.org/10.1053/j.semradonc.2003.10.003>.
- [12] Garcia Schuler HI, Pavic M, Mayinger M, et al. Operating procedures, risk management and challenges during implementation of adaptive and non-adaptive MR-guided radiotherapy: 1-year single-center experience. *Radiat Oncol*. 2021;16 (1):217. Published 2021 Nov 14. doi: <https://doi.org/10.1186/s13014-021-01945-9>.
- [13] Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol* 2016;60(3):393–406.