Impact of a contouring atlas on radiographer inter-observer variation in male pelvis radiotherapy

DOI:
10.1016/j.jmir.2024.03.004

Document Version
Accepted author manuscript

Abstract

Purpose/Objective

To determine the impact of a MR-based contouring atlas for male pelvis radiotherapy delineation on inter-observer variation to support radiographer led real-time magnetic resonance image guided adaptive radiotherapy (MRgART).

Material/Methods

Eight RTTs contoured 25 MR images in the Monaco treatment planning system (Monaco 5.40.01), from 5 patients. The prostate, seminal vesicles, bladder, and rectum were delineated before and after the introduction of an atlas developed through multi-disciplinary consensus.

Inter-observer contour variations (volume), time to contour and observer contouring confidence were determined at both time-points using a 5-point Likert scale. Descriptive statistics were used to analyse both continuous and categorical variables. Dice similarity coefficient (DSC), Dice-Jaccard coefficient (DJC) and Hausdorff distance were used to calculate similarity between observers.

Results

Although variation in volume definition decreased for all structures among all observers post intervention, the change was not statistically significant. DSC and DJC measurements remained consistent following the introduction of the atlas for all observers. The highest similarity was found in the bladder and prostate whilst the lowest was the seminal vesicles. The mean contouring time for all observers was reduced by 50% following the introduction of the atlas (53 to 27 minutes, $p=0.01$). For all structures across all observers, the mean contouring confidence increased significantly from 2.3 to 3.5 out of 5 ($p \leq 0.02$).

Conclusion

Although no significant improvements were observed in contour variation amongst observers, the introduction of the consensus-based contouring atlas improved contouring confidence and speed; key factors for a real-time RTT-led MRgART.

Impact of a contouring atlas on radiographer inter-observer variation in male pelvis radiotherapy

Authors: A. Clough[1], Robert Chuter[1,2], Rosie B Hales[1], Jacqui Parker[1], John McMahon[1], Lee Whiteside[1], Louise McHugh[1], Lucy Davies[1], Josephine Sanders[1], Rebecca Benson[1], Claire Nelder[1] , Lisa McDaid[1], Ananya Choudhury[1,2], Cynthia L Eccles[1,2]

1. The Christie NHS Foundation Trust, Manchester, United Kingdom
2. Division of Cancer Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

Keywords: Radiotherapy, MR-Linac, MRI-guided adaptive radiotherapy, MRgART, prostate, contour

Acknowledgements: MOMENTUM data used as part of this study. The authors acknowledge support from the NIHR Manchester BRC, BRC-1215-20007, Cancer Research UK Manchester Centre award [CTRQQR-2021\100010, ARTNET and Manchester CRUK RADNET.

Corresponding Author - Cynthia Eccles cynthia.eccles1@nhs.net

Figure and table legend

Figure 1 – Image A is the STAPLE contour from the pre-atlas contours, B is the STAPLE contour from the post-atlas contours. Contours from all observers for rectum (brown), prostate (cyan), SVs (green) and bladder (yellow) are shown. All STAPLE contours are shown in red.

Figure 2 – Prostate contour for each observer pre- and post- atlas.

Table 1 – Summary of results including Inter-observer variability and Assessment of contouring efficiency for each specific contour and overall.

Table 2 - Levels of significance as measured by t test on variation in contour size from STAPLE contour to observers contours pre- and post-atlas introduction.

Figure 3 - Observer confidence levels pre- and post-atlas.

Introduction

Magnetic resonance linear accelerators (MRL) are a new hybrid technology combining a magnetic resonance imaging (MRI) scanner and a linear accelerator [1, 2]. MRI has superior soft tissue contrast compared to computed tomography (CT) which is conventionally used in radiotherapy [3]. Better visualisation of patient anatomy including malignant tumours and surrounding radiosensitive normal tissues (organs at risk [OAR]) could facilitate individualised adaptive treatment through online recontouring in MRI-guided online adaptive radiotherapy [2]. Better anatomical visualisation combined with daily adaptation allows greater confidence of target coverage which could facilitate treatment margins reductions and hypo-fractionation [1, 5].

A crucial component for successful intensity-modulated radiation therapy (IMRT) is accurate delineation of target volumes and OARs. IMRT involves inverse planning through defining Planning Target Volumes (PTVs) and OARs, stipulating constraints on the dose volume histograms (DVHs) of each of these structures and optimising the treatment plan to meet these constraints [6]. Delineation variability of target volumes has been extensively documented [7]. Inter-observer variation in the delineation of target volumes and OARs may result in under-dosing of tumours, overdosing of OARs or both [6]. Steenbakkers *et al*. (2003) [8] observed a reduction of mean dose to the rectal wall by 5.1 Gy and to the penile bulb by 11.6 Gy when reducing inter-observer variability by using MRI for delineation in prostate cancer radiotherapy.

The clinical application of MR image guided adaptive radiotherapy (MRgART) requires fast, accurate delineation of target volumes and OARs to mitigate the risk of both internal and external patient motion leading to contouring inaccuracies which could result in treatment errors [9]. Inaccurate contours could lead to excessive irradiation of OARs and increased radiotherapy toxicity [10, 11]. ~~Currently,~~ It is argued that inter-observer variability in treatment volume definition is the most significant contributor to the uncertainty in radiation treatment planning [7] and therefore, a 'gold standard' of delineation is rarely attainable. The focus of current research has consequently been on minimising inter-observer variability (IOV) with the presumption that the intersecting volume of a range of observer's contours will be closer to the ground truth [10]. To reduce IOV in contouring,

research has recommended the use of contouring atlases and guidelines with a combination of external peer reviewed guidelines and local protocols [10]. In this case, the term ‘atlas' refers to a model set of expertly contoured CT images for one (single-patient) or more (multi-patient) cases that can be used to guide the delineation of target volumes and OARs for similar cases [11].

Currently, at the authors’ institution, the activity of re-contouring remains primarily within the scope of practice of clinical oncologists. To facilitate a sustainable model of MRgART, agreed protocols need to be established for radiographer-led treatment pathways. This must include radiographer contouring if the service is to mirror that of a conventional linear accelerator in terms of workflow and efficiency [12]. Previous research has demonstrated that a ‘clinician-lite’ prostate treatment has been well established within the adapt-to-position (ATP) workflows, which is similar to a virtual couch shift [4]. Radiographers’ scope of practice includes being able to localise the target volume and identify OARs on images such as CT and MRI to provide information for radiotherapy treatment planning [13]. Despite this, on-treatment MRI-based contouring is a little practised skill in most radiotherapy centres, meaning RTT’s skills should be refreshed to allow not just competence, but also confidence, to facilitate online recontouring in MRI-guided online adaptive radiotherapy. This work describes the effectiveness of a contouring atlas to facilitate this training for adaptive prostate radiotherapy treatments on the MRL.

Materials and Methods

Patient Inclusion

This retrospective matched pairs study used imaging datasets previously acquired on patients recruited to an ethically approved research study (ClinicalTrials.gov ID: NCT02973828) at the authors’ institution. This study permits retrospective use of imaging data. All patients received radical radiotherapy to the prostate on the MRL in a single department.

Image Acquisition

T2-weighted three-dimensional turbo spin echo sequences, reconstructed axially (T2w 3D TSE) during the course of radiation treatment on a 1.5 Telsa Elekta Unity MRL. Patients were

in the treatment position with standard MRL immobilisation. No bladder or bowel preparation was used as outlined in local protocol.

Atlas Production

The contouring atlas was based on the 'Male Pelvis Normal Tissue RTOG Consensus Contouring' recommendations [14], local contouring guidelines and multi-disciplinary consensus on the contouring of the prostate, seminal vesicles, and associated OARs. The consensus team included 3 experienced MRL therapeutic radiographers, 2 clinical oncologists and 1 diagnostic radiographer.

Structure Delineation

Eight therapeutic radiographers with varying MRL experience (0-18 months) contoured 5 sequential MRI scans obtained during the MRgRT treatment process in the Monaco treatment planning system (Monaco 5.40.01, Elekta AB, Stockholm, Sweden), for 5 patients. Radiographers contoured the prostate, seminal vesicles, bladder, and rectum on each data set before and after the introduction of the atlas. The 'after' contours were generated with a gap of at least 21 days following the 'before' contours to minimise bias from of repetition. The observers have been separated into two groups; > 1 years' experience (3 observers) and < 1 years' experience (5 observers).

Contour Assessment

The Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm of the ADMIRE software (Version 3.23.0.0, Elekta) was used to generate consensus contours. The algorithm combined the multiple observer contours for each anatomical site to establish as a single "true contour" that represents a high-level of agreement amongst the observers [15, 16]. This process was completed to achieve consensus contours for both pre- and post-atlas **(Figure 1**). As variation is a well reported phenomena amongst clinician-based contouring it was decided to utilise a STAPLE contour from all the observer's contours pre-atlas as a true target volume was required for the study. All observers' contours were weighted equally and the consensus contours were reviewed by one oncologist. Each individual contour was then compared to the STAPLE contour to assess inter-observer variability.

The probabilistic estimate of the true segmentation was created using three factors: an estimate of the optimal combination of the segmentations, the weight of each segmentation depending on performance, and the incorporation of an a priori model for the spatial distribution of structures being contoured [17].

Inter-observer contour variations, time taken to contour, and observer contouring confidence was determined prior to and following the introduction of the atlas. The contouring confidence utilised a 5-point Likert scale (observers score confidence 1 *not confident* to 5 *extremely confident*).

Statistical Analysis

Data analysis and statistical tests were performed in Microsoft Excel 2010 (Microsoft Office 2010, Microsoft, US). Descriptive statistics were used to analyse both continuous and categorical variables.

To assess the inter-observer contour variations, a Dice similarity coefficient (DSC) metric was calculated for both pre-atlas to pre-atlas STAPLE and post-atlas to the post-atlas STAPLE using MIRADA.

DSC measures the similarity between two sets and was calculated using the following formula [18]:

$$DSC=\frac{2(A \cap B)}{(A+B)},$$

Where A is the first set, B is the second set, and ∩ is the intersection of the two sets. A closely related similarity metric called Dice-Jaccard coefficient (DJC) is also calculated in MIRADA [19]:

$$DJC=\frac{(A \cap B)}{(A \cup B)}.$$

Hausdorff distance is a metric to measure the maximum distance between of a set to the nearest point in another set. Informally, if every point of either set is close enough to some

point of the other sets, then the two sets are close in the Hausdorff distance. Formally, Hausdorff distance from set A to set B can be defined by:

$$\mathbf{H(A, B) = max\{h(A, B), h(B, A)\}}.$$

Paired T tests were used to assess variation in contour size from the STAPLE before and after use of the registration manual in each observer independently and as team. Statistical significance was defined as $p \leq 0.05$.

Results

Examples of radiographer prostate contours pre- and post-atlas are shown in **Figure 2**. The STAPLE contours pre- and post-atlas are shown in **Figure 1**.
For all eight observers there was no statistically significant change in the pre and post introduction of the contour atlas for any volumes (prostate or OAR).

Before the atlas was introduced, the rectum contour demonstrated the largest variability in 4 of the 5 patients. In two patients, the inter-observer variation improved. For example, for patient 3, the rectal volume contoured pre-atlas ranged from 24.74cm^3 to 60.86cm$^{3.}$ Post-altas, this was reduced to a range of 34.65cm^3 to 58.17cm^3. For two patients, the variation in rectal volumes pre- and post-atlas remained the same. For patient 1, the contour variation increased with the introduction of the atlas, from a range of 17.69cm^3 -31.01cm^3 to 20.13cm^3 – 38.05cm^3 respectively.

Although the variation in volume definition decreased for all structures among all observers post intervention, the greatest reduction was observed in the prostate contour (mean variation before atlas 7.37 mm^3 to 3.46 mm^3 after). The change was not statistically significant (**Table 1**). DSC and DJC measurements remained consistent post-atlas for all observers (**Table 1**). The highest similarity was found in the bladder and prostate whilst the lowest was the seminal vesicles. This is supported by the Hausdorff distances, with bladder having the smallest distance of 4 mm both pre- and post-atlas. Hausdorff distances did improve for both prostate and rectum from 15.92 mm to 10.86 mm demonstrating an improvement in variability between observers. The Hausdorff distance became larger for the seminal vesicles, 6.63mm to 7.48mm, indicating greater variation.

Through the analysis of each observer's variation from the STAPLE, 6 observers independently had significant less variation from the STAPLE for the bladder (p=0.03-0.04), prostate (p=0.01-0.03) and seminal vesicles (p=0.03-0.05) post-atlas when compared to pre-atlas. Similarly, 5 observers had significantly less variation from the STAPLE for the rectal contour (p=0.02-0.05) post atlas. When analysing all the observers combined, the only contour which had a significant reduction in variation from the STAPLE was the rectum (p=0.05) after the atlas when compared to pre-atlas. For observers with greater than 1 years' experience, the use of the atlas significantly reduce variation from the STAPLE for 3 of the 4 contours, bladder (0.04), rectum (0.02) and Seminal Vesicles' (0.04), when compared to pre-atlas variation from the STAPLE (**Table 2**). Observers with less than 1 years' experience also had similar results for rectum (p=0.05) and seminal vesicles (0.03) however, this was not found for bladder (p=0.29) but unlike those with > 1 years' experience there was a significant reduction in variation for prostate (p=0.03).

Time

The mean contouring time for all observers for all structures was reduced by nearly 50%, from 53 to 27 minutes (range for no atlas 33 to 82 minutes and with atlas 22 to 55 minutes, p=0.01) following the introduction of the atlas (**Table 1**).

Confidence levels

For all structures, the observer median contouring confidence increased from 2.13 (range = 1.0-3.1) to 3.5 (range= 2.8-4.4) out of 5 ($p \leq 0.02$) (**Table 3**). For all observers the median confidence increased with the introduction of the atlas across the observers for prostate (median=2, range=1.0-3.0 to median= 3.1, range= 2.6-4.0), for bladder (median=3.0, range=1-4 to median= 4.0, range= 3.2-4.0), rectum (median=1.87, range=1.0-3.0 to median= 3.0, range= 2.6-4.0) and seminal vesicles (median=2.12, range=1.0-3.1 to median= 3.53, range= 2.8-4.4).

Discussion

This work shows the introduction of a contouring atlas did not lead to a statistically significant reduction of the variation among observers. Independent observer variation

from the STAPLE was reduced significantly following the introduction of the atlas for the bladder, prostate and seminal vesicles for 6 observers and for the rectal volume for 5 observers. The time taken to contour was reduced and radiographer confidence was also significantly improved (p=0.01).

It has been reported within literature that the largest delineation variation occurs around the apex and the base of the prostate [20]. A limitation of using the DSC, is that it does not quantitatively describe where within the contoured volume the majority of the variation occurs [11]. This work demonstrated that although variation was reduced post-atlas in both the apex and the base, these did remain the areas with most variation (**Figure 2**).

Valicenti *et al*. [21] found that inter-observer variability is four-times larger for seminal vesicles delineation than prostate delineation. We report a similar result both pre- and post-atlas. Pre-atlas the range of DSC coefficient was 0.32 and post-atlas introduction was 0.19. Another limitation of DSC is sensitivity to variability for small volume structures, which may have impacted the results seen on seminal vesicles more than the other structures [11]. Hausdorff distance became larger for the seminal vesicles after the introduction of the atlas indicating greater variation. Further work could be required on the teaching of the contouring of this structure. DSC, DJC and Hausdorff were utilised within this study as currently there is no consensus within the literature as to the appropriate metric for the analysis and comparison of contours [22]. Future work could be conducted to specifically analyse observer variation at the apex and base of the prostate.

A survey conducted in 2007 within the UK for radiotherapy planning and delivery demonstrated a lack of formal education in target volume and OARs delineation [23]. Only 4% of NHS radiotherapy departments offered structured training on image interpretation for contouring purposes, while 6% offered informal sessions with radiologists. CT-based delineation guidelines however, have been published on a national or international level for several tumour sites both in external beam radiotherapy and brachytherapy [24, 25, 26, 27, 28, 29, 30]. Studies have reported that the use of site-specific anatomical atlases, consensus delineation guidelines and standardised contouring protocols reduce variability between observers in various tumour sites [31]. The introduction of a site-specific consensus atlas for rectal cancer significantly reduced inter-observer variability in a pilot study [32]. Results

were repeated in a large study, with inter-observer variability reducing in caudal CTV border (from 1.8 to 1.2 cm) and the size of average CTV volume by 25% (620 *vs*. 460 cc) [33]. Within this study the only intervention was the introduction of the atlas, although our sample size was small and results were not statistically significant, a reduction in contour variation was noted. Other studies which have evaluated anatomic atlas as part of contouring guidelines have found similar results for rectum [32] and head and neck regions [34]. Vinod et al. (2016) [10] demonstrated that group training or one-to-one teaching may be more beneficial when compared to atlas-based training. However, due to the impact of the coronavirus pandemic, this method was not feasible in the effort to reduce staff contact. Alternative methods including both a combination of directed and self-directed teaching material could be investigated in the future, which could be virtual or in person. With online recontouring for MR guided adaptive radiotherapy not requiring radiographers to contour afresh, but instead adapt the pre-existing contours to encompass the daily anatomical variation, disparities in interobserver variation may be reduced within this setting.

Fractionation time has been reported as a potential weakness of the MRL with contouring time being indicated as a key burden [5]. Reducing contour time is therefore a key area for improvement, and with the participants increased confidence levels and associated reduction in task time, hopefully the use of the atlas will help improve efficiency of the MRL.

STAPLE is advantageous as it provides sensitivity and specificity for each structure which can then be compared for each participant. This helps facilitate individual assessment, comparison, and the analysis of the impact of experience and inter-observer variability on overall delineation results [35]. Interestingly, within this research the STAPLE volume for some of the contours changed greatly before and after the atlas introduction, for example for the rectum contour one patient's STAPLE increased in size from 28.613cm^3 to 39.488 cm^3. Whereas, for the same patient the bladder STAPLE did not alter that greatly from pre to post atlas, 125cm^3 to 122.88 cm^3 respectively. The reasons for this are unknown and not studied as part of this research but could be hypothesised that observers were unsure of where sigmoid colon becomes rectum, and appearance of ano-rectum prior to the atlas introduction meaning a greater volume was contoured after. This could explain why before the atlas was introduced, the rectum contour demonstrated the largest variability in 4 of the

5 patients. In 2 patients post-atlas, the variation in rectal contour volume remained the same and for 1 patient variation increased. This could mean clarification on the rectal size within the atlas could be of benefit and should be explored in the future. It is important to highlight figure one demonstrates that the rectum contour around the prostate appears to have low variability which will be within the high dose PTV. Future work could assess the dosimetric impact in the variation in rectal volumes.

A limitation of STAPLE contours is that the true volume depends on the number of contours inputted and number of iterations used in the algorithm. The number of contours required for a robust STAPLE contour has not been established. However, as only 8 participants have been included within this research which could have impacted the statistical significance of the results. Repeating this process with more participants could be a future project to ensure robustness is established.

All participants despite being a minimum of 6 years post qualification, only had limited experience working on the MRL and with the Monaco planning system (0-18 months). All participants had gained urology CT-CBCT image registration competencies. No participants had previous experience with contouring on either MRI or CT. The atlas significantly reduced variation from the STAPLE for the prostate contours for those with less than a years' experience ($p=0.03$) but did not for those with great than a years' experience ($p=0.30$). The reasons for this are unknown, one suggestion could be those with experience already had a good idea of the prostate contour pre-atlas so the atlas did not reduce the variation. Alternatively, those with less experience were more cautious in following the atlas due to their lack of experience/confidence. A future project could repeat this research on the same observers now they have greater experience on the MRL to assess this impact on inter-observer variation.

One limitation of this methodology is the completing of the contour's pre-atlas could be training itself. This might be especially an influence for the more inexperienced raters. Furthermore, there can be repetition bias, and recall bias. Because of the biases it is difficult to tell whether the observer changes between the pre an post-atlas contours are a result of the contouring atlas.

In conclusion, contour variation across observers did reduce, but not significantly. However, the use of an atlas allowed radiographers to feel significantly more confident to contour male pelvis anatomy on MR faster. Confidence and time are both are key factors when contouring in real-time within a radiographer-led Adapt to Shape work flow.

References

1. Liney, G. Whelan, B. Oborn, B. Barton, M. Keall, P. (2018) 'MRI-Linear Accelerator Radiotherapy Systems', *Clinical Oncology*, 30(11), pp. 686-691. doi: 10.1016/j.clon.2018.08.003.

2. Chin, S. Eccles, C. McWilliam, A. Chuter, R. Walker, E. Whitehurst, P. Berresford, J. Van Herk, M. Hoskin, P. Choudhury, A. (2020), 'Magnetic resonance-guided radiation therapy: A review', *Journal of Medical Imaging and Radiation Oncology*, 64(1), pp. 163.177. doi:10.1111/1754-9485.12968.

3. Rai, R., Kumar, S., & Batumalai, V., et al. (2017). The integration of MRI in radiation therapy: collaboration of radiographers and radiation therapists. J Med Radiat Sci 64(1), 61–68. doi: 10.1002/jmrs.225.

4. Winkel, D. Bol, G. Kroon, P. Van Asselen, B. Hackett, S. Werensteijn-Honingh, A. Intven, M. Eppinga, W. Tijssen, R. Kerkmeijer, L. De Boer, H. Mook, S. Meijer, G. Hes, J. Willemsen-Bosman, M. De Groot-van Breugel, E. Jurgenliemk-Schulz, I. Raaymakers, B.(2019) 'Adaptive radiotherapy: The Elekta Unity MR-linac concept', *Clinical and Translational Radiation Oncology*, 18, pp. 54–59. Doi:10.1016/j.ctro.2019.04.001.

5. Van Herk, M. McWilliam, A. Dubec, M. Faivre-Finn, C. Choudhury, A. (2018) 'Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats analysis', *Int J Radiat Oncol Biol Phys*, 101(5), pp. 1057–1060. DOI: 10.1016/j.ijrobp.2017.11.009.

6. Loo, S.W., Martin, W.M., Smith, P., Cherian, S., Roques, T.W. (2012) Interobserver variation in parotid gland delineation: a study of its impact on intensity modulated

radiotherapy solutions with a systematic review of the literature, *Br J Radiol.,* 85(1016), pp.1070–1077. doi: 10.1259/bjr/32038456.

7. Weiss, E., & Hess, C.F. (2003) The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol,* 179, pp.21–30. doi: 10.1007/s00066-003-0976-5.

8. Steenbakkers, R.J.H., Deurloo, K.E., Nowak, P.J.C., Lebesque, J.V., van Herk, M., Rasch, C.R. (2003) Reduction of dose delivered to the rectum and bulb of the penis using MRI delineation for radiotherapy of the prostate. *Int J Radiat Oncol.* 57, pp. 1269–79. doi: 10.1016/s0360-3016.

9. Jameson, M.G., Kumar, S., Vinod, S.K., Metcalfe, P.E., Holloway, L.C. (2014) Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. *Radiother Oncol*, 112: pp. 332–6. doi: 10.1016/j.radonc.

10. Vinod, S. K. Min, M. Jameson, M. G. Holloway L.C. (2016) A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology, *Journal of Medical Imaging and Radiation Oncology*, 60, pp. 393-406. doi.org/10.1111/1754-9485.12462

11. Velker, V.M., Rodrigues, G.B., Dinniwell, R., Hwee, J., Louie, A.V. (2013) Creation of RTOG compliant patient CT-atlases for automated atlas based contouring of local regional breast and high-risk prostate cancers. *Radiat Oncol*. 8, pp. 188. doi:10.1186/1748-717X-8-188.

12. Hales, R. Rodgers, J. Whiteside, L. McDaid, L. Berresford, J. Budgell, G. Choudhury, A. Eccles, C. (2020) 'Therapeutic Radiographers at the Helm: Moving Towards Radiographer-Led MR-Guided Radiotherapy', *Journal of Medical Imaging and Radiation Science*, 51(3), pp. 364-372. Doi: 10.1016/j.jmir.2020.05.001.

13. Health & Care Professions Council (2013) *Standards of Proficiency: Radiographers.* Available at: https://www.hcpc-uk.org/globalassets/resources/standards/standards-

of-proficiency---radiographers.pdf?v=637106257920000000 (Accessed: 29 September 2022).

14. Radiation Therapy Oncology Group (2012) *MALE PELVIS Normal Tissue RTOG Consensus Contouring Guidelines,* Available at: https://www.nrgoncology.org/Portals/0/Scientific%20Program/CIRO/Atlases/Male%20RTOG%20Normal%20Pelvis%20Atlas.pdf (Accessed: 20 September 2022).

15. Allozi, R. Li, X.A. White, J. Apte, A. Tai, A. Michalski, J.M. Bosch, W.R. El Naqa, I. (2010) Tools for consensus analysis of experts' contours for radiotherapy structure definitions. *Radiotherapy and Oncology*, 97(3), pp. 572-8. doi: 10.1016/j.radonc.2010.06.009.

16. Lin, D. Lapen, K. Sherer, M.V. Kantor, J. Zhang, Z. Boyce, L.M. Bosch, W. Korenstein, D. Gillespie, E.F. (2020) A Systematic Review of Contouring Guidelines in Radiation Oncology: Analysis of Frequency, Methodology, and Delivery of Consensus Recommendations. *Int J Radiat Oncol Biol Phys*, 107(4), pp. 827-835. doi: 10.1016/j.ijrobp.2020.04.011.

17. Warfield, S.K. Zou, K.H. Wells, W.M. (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging,* 23(7), pp. 903-21. doi: 10.1109/TMI.2004.828354.

18. Van Mourik, A.M., Elkhuizen, P.H., Minkema, D., Duppen, J.C., Van Vliet-Vroegindeweij, C. (2010) Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines, *Radiother Oncol*, 94, pp.286–291. doi: 10.1016/j.radonc.2010.01.009.

19. Vorwerk, H., Beckmann, G., Bremer, M., et al. (2009) The delineation of target volumes for radiotherapy of lung cancer patients, *Radiother Oncol*, 91, pp. 455–460. doi: 10.1016/j.radonc.2009.03.014.

20. Villeirs, G.M., Van Vaerenbergh, K., Vakaet, L., Bral, S., Claus, F., De Neve, W.J., Verstraete, K.L., De Meerleer, G.O. (2005) Interobserver delineation variation using

CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer. *Strahlenther Onkol*. 181(7), pp. 424-30. doi: 10.1007/s00066-005-1383-x.

21. Valicenti, R.K., Sweet, J.W., Hauck, W.W., Hudes, R.S., Lee, T., Dicker, A.P. (1999) Variation of clinical target volume definition in three-dimensional conformal radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys.* 44, pp.931–5. doi: 10.1016/s0360-3016(99)00090-5.

22. Jameson, M.G., Holloway, L.C., Vial, P.J., Vinod, S.K., Metcalfe, P.E. (2010) A review of methods of analysis in contouring studies for radiation oncology. *J Med Imaging Radiat Oncol.* 54, pp. 401–410. doi: 10.1111/j.1754-9485.2010.02192.x.

23. Jefferies, S., Taylor, A., Reznek, R. (2007) Results of a national survey of radiotherapy planning and delivery in the UK in 2007. *Clin Oncol.* 21, pp. 204–17. doi: 10.1016/j.clon.2008.11.017.

24. Haie-Meder, C., Pötter, R., Van Limbergen, E., Briot, E., De Brabandere, M., Dimopoulos, J. Dumas, I., Hellebust, T.P., Kirisits, C., Lang, S., Muschitz, S., Nevinson, J., Nulens, A., Petrow, P., Wachter-Gerstner, N. (2005) Recommendations from Gynaecological (GYN) GECESTRO Working Group (I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. *Radiother Oncol.*74, pp. 235–45. doi: 10.1016/j.radonc.2004.

25. Small, W., Mell, L.K., Anderson, P., Creutzberg, C., De Los Santos, J., Gaffney, D. Jhingran, A., Portelance, L., Schefter, T., Iyer, R., Varia, M., Winter, K., Mundt, A.J.C. (2008) Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy in postoperative treatment of endometrial and cervical cancer. *Int J Radiat Oncol Biol Phys.* 2008, 71, pp. 428–34. doi: 10.1016/j.ijrobp.2007.09.042.

26. Lim, K., Small, W., Portelance, L., Creutzberg, C., Jürgenliemk-Schulz, I.M., Mundt, A. et al. (2011) Consensus guidelines for delineation of clinical target volume for

intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. *Int J Radiat Oncol.* 79, pp. 348–55. doi.org/10.1016/j.ijrobp.2009.10.075.

27. Nielsen, M.H., Berg, M., Pedersen, A.N., Andersen, K., Glavicic, V., Jakobsen, E.H. et al. (2013) Delineation of target volumes and organs at risk in adjuvant radiotherapy of early breast cancer: National guidelines and contouring atlas by the Danish Breast Cancer Cooperative Group. *Acta Oncol (Madr),* 52, pp. 703–10. doi: 10.3109/0284186X.2013.765064.

28. Grégoire, V., Ang, K., Budach, W., Grau, C., Hamoir, M., Langendijk, J.A. et al. (2014) Delineation of the neck node levels for head and neck tumors: A 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC, CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother Oncol.*110, pp. 172–81. DOI: 10.1016/j.radonc.2013.10.010.

29. Harris, V.A., Staffurth, J., Naismith, O., Esmail, A., Gulliford, S., Khoo, V. et al. (2015) Consensus guidelines and contouring atlas for pelvic node delineation in prostate and pelvic node intensity modulated radiation therapy. *Int J Radiat Oncol.* 92, pp. 874–83. doi: 10.1016/j.ijrobp.2015.03.021.

30. Wu, A.J., Bosch, W.R., Chang, D.T., Hong, T.S., Jabbour, S.K., Kleinberg, L.R. et al. (2015) Expert consensus contouring guidelines for intensity modulated radiation therapy in esophageal and gastroesophageal junction cancer. *Int J Radiat Oncol*. 92, pp. 911–20. doi.org/10.1016/j.ijrobp.2015.03.030.

31. Segedin, B., & Petric, P. (2016) Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them?. *Radiol Oncol*. 50 (3), pp. 254-262. doi:10.1515/raon-2016-0023.

32. Fuller, C.D., Nijkamp, J., Duppen, J., Rasch, C.R., Thomas, C.R., Wang, S.J. et al. (2011) Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. *Int J Radiat Oncol Biol Phys.*79, pp. 481–9. doi: 10.1016/j.ijrobp.2009.11.012.

33. Nijkamp, J., Swellengrebel, M., Hollmann, B., de Jong, R., Marijnen, C., van Vliet-Vroegindeweij, C., van Triest, B., van Herk, M., Sonke, J.J. (2012) Repeat CT assessed

CTV variation and PTV margins for short- and long-course pre-operative RT of rectal cancer. *Radiother Oncol*. 102(3), pp. 399-405. doi: 10.1016/j.radonc.2011.11.011.

34. Awan M, Kalpathy-Cramer J, Gunn GB et al. (2013) Prospective assessment of an atlas-based intervention combined with real-time software feedback in contouring lymph node levels and organs-at-risk in the head and neck: quantitative assessment of conformance to expert delineation. *Pract Radiat Oncol*, 3, pp. 186–93. doi: 10.1016/j.prro.2012.11.002.

35. Sandström, H., Toma-Dasu, I., Chung, C., GÅrding, J., Jokura, H., Dasu, A. (2021) Simultaneous Truth and Performance Level Estimation Method for Evaluation of Target Contouring in Radiosurgery. *Anticancer Res*. 41(1), pp. 279-288. doi: 10.21873/anticanres.14774. PMID: 3341982.

Figure 1

Click here to access/download;Figure;Figure 1.png

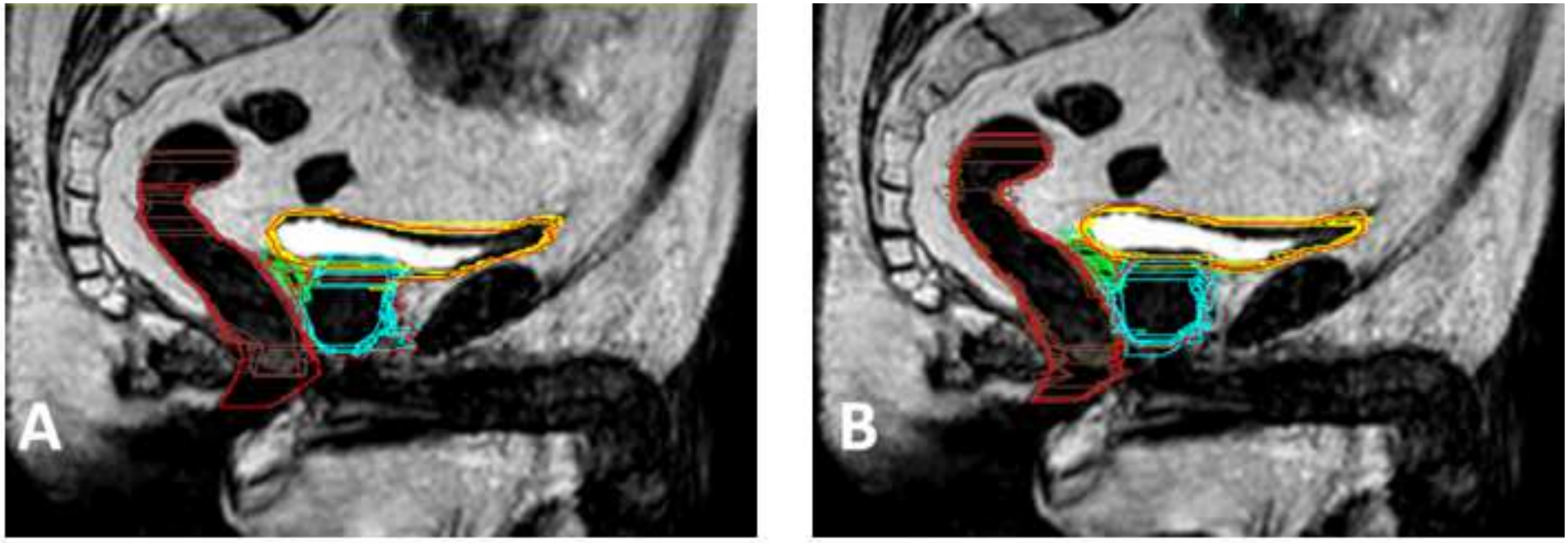


Figure 2

Click here to access/download;Figure;figure 2.png

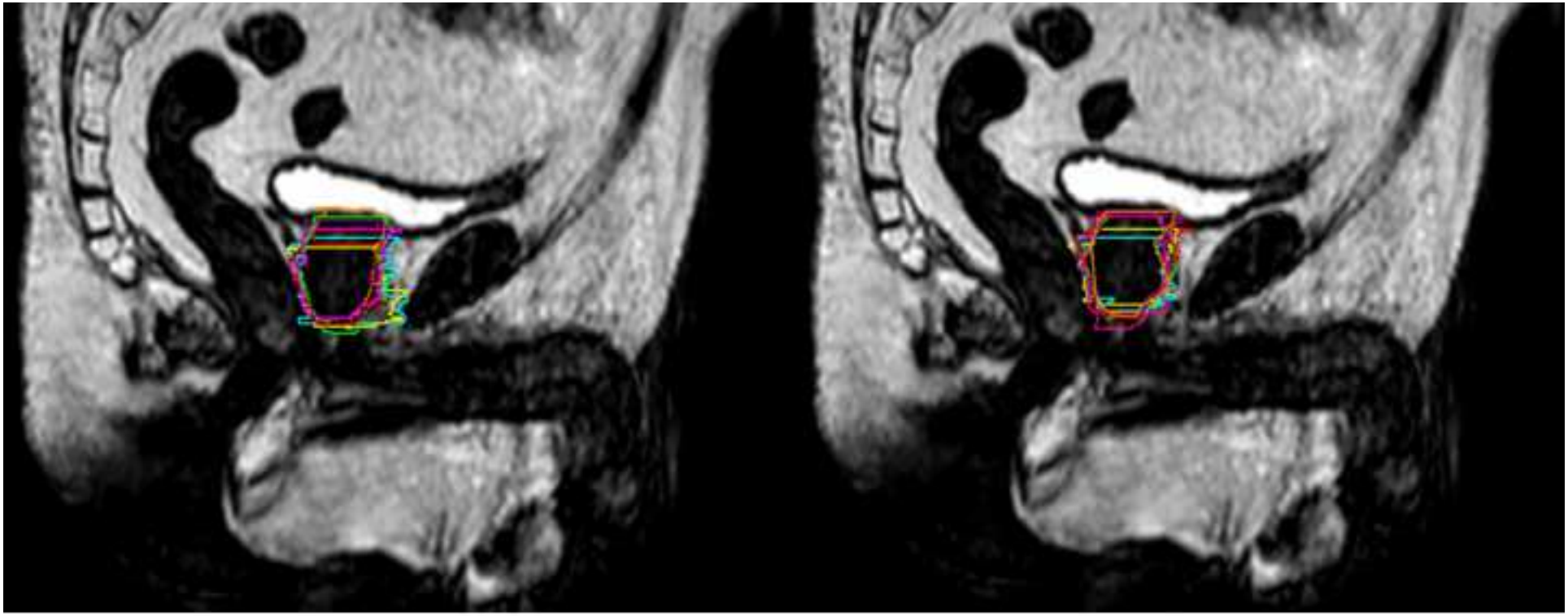
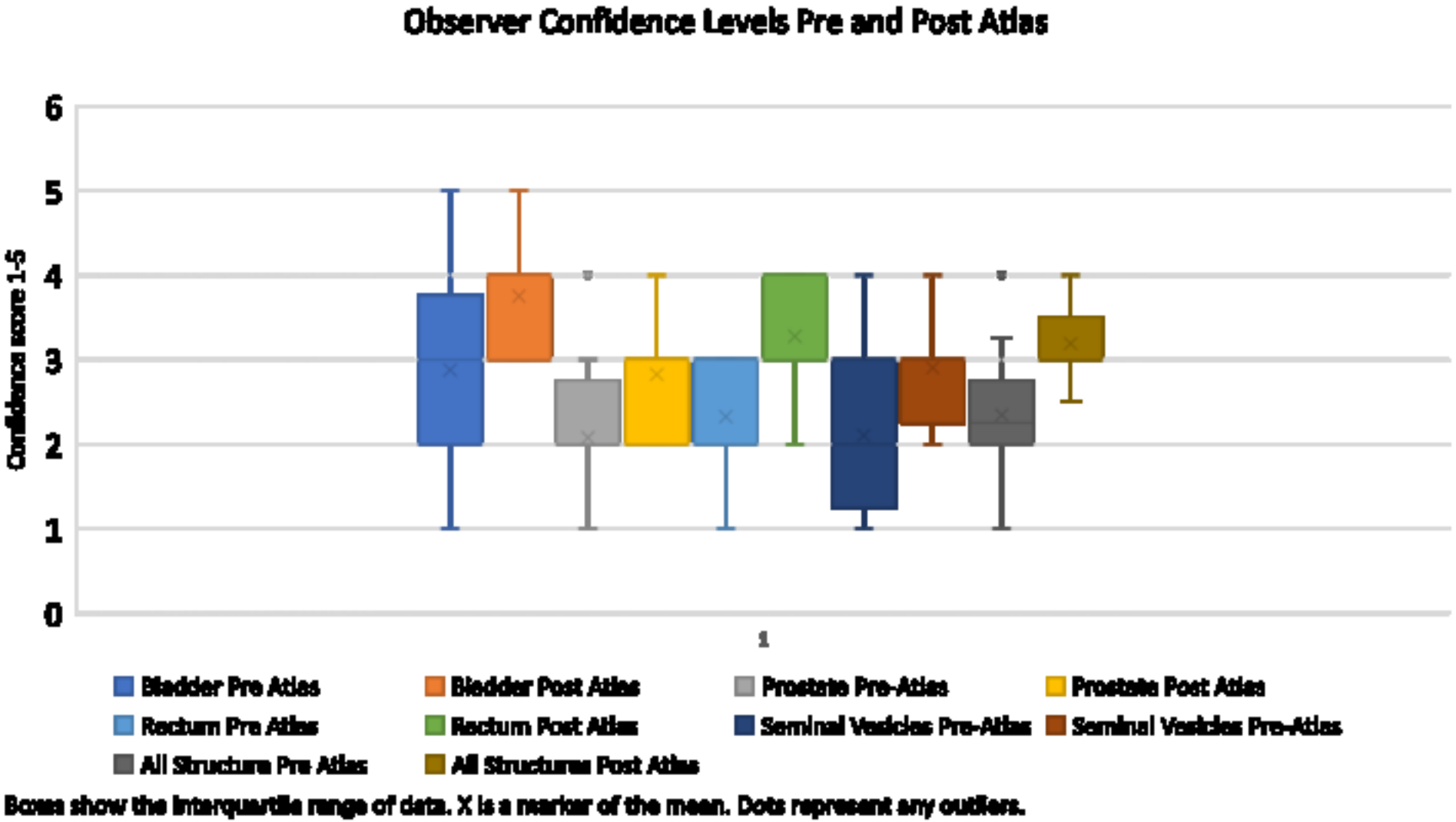


Figure 3

Click here to access/download;Figure;figure 3.1.png



Boxes show the interquartile range of data. X is a marker of the mean. Dots represent any outliers.

Table 1

			Pre Atlas (CI)	Post Atlas (CI)
Bladder	Inter-observer variability	DSC	0.95 (0.97-0.82)	0.95(0.99-0.86)
		Hausdorff D (mm)	4 (8.94-0.68)	4 (9.38—2.82)
		Jaccard	0.91 (0.95-0.65)	0.90 (0.95-0.25)
		Variance (cm^3)	5.83	4.99
	Assessment of contouring efficiency	Time taken to contour (mins)	14 (21-9)	7 (11-6)
Prostate	Inter-observer variability	DSC	0.86 (0.97-0.65)	0.88 (0.94-0.75)
		Hausdorff D (mm)	15.92 (38.42-3.46)	10.86 (28.7-2.82)
		Jaccard	0.76 (0.94-0.48)	0.81(0.88-0.6)
		Variance (cm^3)	7.37	3.47
	Assessment of contouring efficiency	Time taken to contour (mins)	14 (21-7)	9 (17-5)
Rectum	Inter-observer variability	DSC	0.86 (0.95-0.56)	0.88 (0.95-0.66)
		Hausdorff D (mm)	15.92(38.42-3.46)	10.86 (28.7-2.82)
		Jaccard	0.75(0.91-0.42)	0.78 (0.90-0.49)
		Variance (cm^3)	9.96	6.69
	Assessment of contouring efficiency	Time taken to contour (mins)	14 (25-7)	8 (17-4)
Seminal Vesicles	Inter-observer variability	DSC	0.79 (0.91-0.36)	0.78 (0.9-0.42)
		Hausdorff D (mm)	6.63 (24.74-0.69)	7.48 (15.62-3.46)
		Jaccard	0.65 (0.83-0.22)	0.64 (0.82-0.27)
		Variance (cm^3)	2.23	2.45
	Assessment of contouring efficiency	Time taken to contour (mins)	8 (13-4)	5 (12-2)

All Structures	Assessment of contouring efficiency	Time taken to contour all structures (mins)	53 (33 - 82)	27 (22 - 55)

Table 2

	Experience on MRL	Bladder	Prostate	Rectum	SV's
Observer 1	< 1 year	0.12	0.02	0.13	0.05
Observer 2	< 1 year	0.04	0.03	0.05	0.04
Observer 3	**> 1 year**	0.04	0.02	0.07	0.04
Observer 4	**> 1 year**	0.04	0.02	0.02	0.03
Observer 5	< 1 year	0.03	0.18	0.08	0.08
Observer 6	< 1 year	0.06	0.01	0.04	0.03
Observer 7	**> 1 year**	0.03	0.29	0.05	0.05
Observer 8	< 1 year	0.17	0.03	0.33	0.10
All observers		0.13	0.42	0.02	0.10
> 1 year		0.04	0.30	0.02	0.04
< 1 year		0.29	0.03	0.05	0.03