Received:
11 May 2020

Revised:
25 July 2020

Accepted:
28 July 2020

<https://doi.org/10.1259/bjr.20200543>

Cite this article as:

Hearn N, Bugg W, Chan A, Vignarajah D, Cahill K, Atwell D, et al. Manual and semi-automated delineation of locally advanced rectal cancer subvolumes with diffusion-weighted MRI. *Br J Radiol* 2020; **93**: 20200543.

FULL PAPER

Manual and semi-automated delineation of locally advanced rectal cancer subvolumes with diffusion-weighted MRI

^{1,2,3}NATHAN HEARN, BSc/MBBS (Hons), ⁴WILLIAM BUGG, BSc, MBBS, MRCS, FRCR, FRANZCR,
⁴ANTHONY CHAN, MBBS, FRANZCR, ^{1,2}DINESH VIGNARAJAH, MBBS, FRANZCR, ¹KATELYN CAHILL, BAppSc (MRT-RT),
^{1,2,3}DAISY ATWELL, BSc, MBBS, ^{3,5}JIM LAGOPOULOS, BSc, MBIomedE, PhD, FAINM and
^{1,2,3}MYO MIN, MBBS, GradDip (Biotech), PhD, FRANZCR

¹Department of Radiation Oncology, Sunshine Coast University Hospital, Birtinya, QLD, Australia

²CON Cancer Centre, Maroochydore, QLD, Australia

³University of the Sunshine Coast, Sippy Downs, QLD, Australia

⁴Department of Medical Imaging, Sunshine Coast University Hospital, Birtinya, QLD, Australia

⁵Sunshine Coast Mind and Neuroscience – Thompson Institute, University of the Sunshine Coast, Birtinya, QLD, Australia

Address correspondence to: Dr Nathan Hearn
E-mail: nathan.hearn@research.usc.edu.au

Objectives To evaluate interobserver agreement for T_2 weighted (T_2W) and diffusion-weighted MRI (DW-MRI) contours of locally advanced rectal cancer (LARC); and to evaluate manual and semi-automated delineations of restricted diffusion tumour subvolumes.

Methods 20 cases of LARC were reviewed by 2 radiation oncologists and 2 radiologists. Contours of gross tumour volume (GTV) on T_2W , DW-MRI and co-registered T_2W /DW-MRI were independently delineated and compared using Dice Similarity Coefficient (DSC), mean distance to agreement (MDA) and other metrics of inter-observer agreement. Restricted diffusion subvolumes within GTVs were manually delineated and compared to semi-automatically generated contours corresponding to intratumoral apparent diffusion coefficient (ADC) centile values.

Results Observers were able to delineate subvolumes of restricted diffusion with moderate agreement (DSC 0.666, MDA 1.92 mm). Semi-automated segmentation

based on the 40th centile intratumoral ADC value demonstrated moderate average agreement with consensus delineations (DSC 0.581, MDA 2.44 mm), with errors noted in image registration and luminal variation between acquisitions. A small validation set of four cases with optimised planning MRI demonstrated improvement (DSC 0.669, MDA 1.91 mm).

Conclusion Contours based on co-registered T_2W and DW-MRI could be used for delineation of biologically relevant tumour subvolumes. Semi-automated delineation based on patient-specific intratumoral ADC thresholds may standardise subvolume delineation if registration between acquisitions is sufficiently accurate.

Advances in knowledge This is the first study to evaluate the feasibility of semi-automated diffusion-based subvolume delineation in LARC. This approach could be applied to dose escalation or 'dose painting' protocols to improve delineation reproducibility.

BACKGROUND

Optimisation of pathological complete response (pCR) rates after concurrent neoadjuvant chemoradiotherapy (nCRT) for locally advanced rectal cancer (LARC) could improve the feasibility of surgery-sparing treatment paradigms. In previous meta-analyses, radiotherapy escalation (>54–60 Gy) was associated with high rates of pCR (20.4–25%),^{1,2} while a recent randomised trial demonstrated favourable toxicity profiles with dose-escalated treatment compared to intensified chemotherapy.³ However, an earlier dose-response modelling study predicted that the

administration of 92 Gy was required to achieve a 50% pCR rate⁴; therefore, there is a need to explore methods of targeted dose escalation which maintain a favourable therapeutic ratio.

MRI, including T_2 weighted (T_2W) and diffusion-weighted sequences (DW-MRI, or DWI), provides superior soft tissue differentiation for tumour delineation in image-guided radiotherapy workflows.⁵ DW-MRI has prognostic significance in LARC⁶ and may provide a more biologically relevant target volume for radiotherapy boost techniques

based on tumour cellularity.⁷ Specifically, radiotherapy boost to areas of restricted diffusion corresponding to hypercellular tumour regions could improve tumour regression.⁸

However, standardisation of tumour volumes deriving from functional imaging remains challenging due to resolution constraints, image artefact and the potential for inconsistent interpretation between observers. Alternatively, functional subregions of interest can be targeted within larger anatomically defined tumour volumes with co-registered images, such as with 'dose painting' methods.^{9,10} Automated region of interest definition based on quantitative functional imaging, such as the apparent diffusion coefficient (ADC) derived from DW-MRI, has also been proposed.¹¹ While some prognostication studies have explored automated approaches to DWI-based LARC gross tumour volume (GTV) delineation,^{12,13} no study to date has evaluated this approach in defining LARC subvolumes for radiotherapy planning purposes.

Thus, the aim of this paper was to evaluate the interobserver variability of MRI-based rectal cancer GTVs on T_2 W, DWI and co-registered MRI contouring sessions; and to compare the consistency of manual and semi-automated subvolume delineations based on areas of restricted diffusion within GTVs, in the context of focal radiotherapy dose escalation strategies.

METHODS

Study population

This retrospective study was approved by a local Human Research Ethics Committee. The institution's radiation oncology departmental database was searched for patients with histologically confirmed rectal adenocarcinoma who underwent neoadjuvant radiotherapy with or without chemotherapy from 2017 onwards. Disease staging was confirmed with pelvic MRI and CT scanning. Mucinous tumours were excluded. Other exclusions were performed for cases of missing ADC images, or poor image quality such as with hip implants or other artefacts.

Image acquisition

Images were acquired on a 3T (Siemens MAGNETOM Skyra) or 1.5T (Siemens MAGNETOM Aera) MRI scanner prior to commencement of therapy. Acquisition sequences included high-resolution T_2 W turbo spin echo (TSE) acquisition aligned with the tumour axis, along with DW-MRI with echoplanar imaging (EPI) acquired in anatomical-axial orientation. Diffusion sequences were acquired either with b-value sets of (50, 800) or (50, 400, 800). An antimuscarinic agent (hyoscine butylbromide 20mg) was administered intramuscularly prior to scanning. Bowel preparation was not used. (See [Supplementary Material 1 A: Image Acquisition](#) for further details).

Images were imported into MIM Maestro® (MIM Software Inc, Cleveland, OH) from the hospital imaging archiving system. True-axial T_2 W images were rigidly registered to b800/ADC sequences using inbuilt MIM software tools, then checked by the study radiation oncologists. Final minor registration adjustments were performed manually if required. CT images were not examined in this study.

Manual volume delineation

Two radiation oncologists (6 and 4 years of specialist rectal cancer experience respectively) and two radiologists (14 and 8 years of specialist abdominal imaging experience respectively) completed contours independently. Observers were provided with endoscopic and rectal examination findings. GTV contours encapsulating all visual tumour on each axial slice were completed in three different sessions for each patient on separate sittings: 1) T_2 W imaging only, 2) DWI (b800 and ADC) only, and 3) co-registered T_2 W and DWI ('CoReg'). Additionally, in the CoReg session, observers were instructed to delineate a smaller subvolume ('SubV') specifically corresponding to their visual assessment of greatest restricted diffusion within their CoReg GTV delineation on b800 or ADC sequences. Observers were allowed to adjust image windowing as desired.

Each observer was asked to complete a questionnaire evaluating image quality and confidence of contouring for each image set using a 4-point confidence scale (1 = poor confidence/quality, 2 = average confidence/quality, 3 = good confidence/quality, 4 = excellent confidence/quality) along with the time required to complete each task (see [Supplementary Material 1 B: Questionnaire](#)).

Quantitative analysis

Interobserver variability of delineated contour volumes, overlap and margins was evaluated within each contouring session for each subject using established metrics.¹⁴ The intraclass correlation coefficient (two-way mixed method, absolute agreement for single measures) evaluated agreement in contour volumetry. The generalised conformity index (CIgen), calculated as the sum of all interobserver contour-pairs' intersections divided by their unions,¹⁵ was used as generalised measure of observers' contour overlap for a given modality/subject session.

Additionally, a consensus contour for each session and subject was created using the simultaneous truth and performance level estimation (STAPLE) algorithm with equal weighting for all observers.¹⁶ The Dice similarity coefficient (DSC), Jaccard index (JI), maximum Hausdorff distance (HDmax) and mean distance to agreement (MDA) between each observer-STAPLE pair was calculated and averaged across observers for a given session/subject combination.

Semi-automated subvolume delineation

Further contour sets were created for each patient through quantitative thresholding of STAPLE CoReg contours, based on intratumoral ADC centiles. First, the STAPLE CoReg contour for each patient was propagated to the corresponding ADC image. Using inbuilt tools in MIM Maestro, ADC histogram centiles were extracted. New segmentations were then automatically generated by applying thresholds of $\leq n$ th centile ADC to the consensus CoReg contour, where n ranged from 25th to the 75th centile at 5-centile increments. Contours were cleaned to remove non-contiguous volumes and then compared with manually delineated consensus SubV contours for each patient using the same pairwise comparisons described above.

Table 1. Patient and tumour characteristics

Patients	20
Age (median, range)	68 (33–83)
Gender (male/female)	10/10
T stage, n (%)	
T2	2 (10)
T3	13 (65)
T4	5 (25)
N stage, n (%)	
N0	7 (35)
N1	7 (35)
N2	6 (30)
Metastatic, n (%)	3 (15)
Distance from anorectal junction	
Median [cm] (range)	3.2 (0–11)
Low (0–5 cm), n (%)	10 (50)
Mid (5–10 cm), n (%)	8 (40)
High (>10 cm), n (%)	2 (10)
Tumour length [cm] (median, range)	4.3 (2–7)
MRI availability, n (%)	
T ₂ W	20 (100)
ADC	20 (100)
DWI (b800)	18 (90)

ADC, apparent diffusion coefficient; DWI, diffusion weighted imaging; T₂W, T₂ weighted.

Statistical analysis

All statistical tests were carried out using the R Statistical Environment (v. 3.6.1, R Core Team). Between-modality comparisons for contour statistics and variability measures were performed using the Wilcoxon signed-rank test for repeated measures (two-tailed, $p < 0.05$). Agreement values were characterised as follows, in accordance with established literature^{7,17}: $\kappa = 1$: perfect agreement; >0.8 : excellent agreement; $0.6–0.8$: substantial agreement; $0.4–0.6$: moderate agreement; and <0.4 : poor agreement. Linear regression was used to evaluate the relationship between CIgen and volume of contours.

RESULTS

After local database search and screening, 20 patients were suitable for inclusion (Table 1). All patients had pre-treatment T₂W and ADC images available for contour delineation. 19 out of 20 patients had MRI acquired at 3T. Image quality of all scans was judged satisfactory or higher by the observers. The b800 sequence was not available for two patients, however all observers were able to complete contours on the corresponding ADC image for DWI and CoReg delineations.

Comparison of GTV volumes

Examples of individual and consensus STAPLE GTV contours for the T₂W, DWI and CoReg sessions are shown in Figure 1.

Figure 1. Individual observer contours and consensus STAPLE contours for A–C) T₂W imaging (axial, coronal, sagittal views), D–F) DWI (axial, coronal, sagittal views), and G–H) coregistered sessions (axial views T₂W and DWI). Contour colours: red – T₂W STAPLE; green – DWI STAPLE; cyan – CoReg STAPLE; dark blue, yellow, orange, pink – individual contours of observers 1–4.

Contour volumes for each observer and for consensus STAPLE contours are shown in Figure 2. Interobserver agreement statistics are summarised in Figure 3. Observer-reported confidence on the four-point scale was highest in T₂W GTV delineation (mean: 3.28), followed by CoReg (mean: 2.56) and DWI (mean: 1.98) sessions. CoReg delineations took longer to complete (mean: 8 min) compared to T₂W (mean: 5 min) and DWI (mean: 5 min).

ICC measures for T₂W, DWI and CoReg sessions were 0.91 (95% CI: 0.84–0.95), 0.74 (95% CI: 0.60–0.85) and 0.56 (95% CI: 0.38–0.74) respectively. Mean consensus tumour volumes for T₂W, DWI and CoReg delineations were 24.00 ml (SD: 14.98 ml), 41.86 ml (SD: 26.74 ml) and 36.11 ml (SD: 22.83 ml) respectively. DWI and CoReg volumes were significantly larger than T₂W volumes ($p < 0.001$ for both comparisons), while there was

Figure 2. Summary boxplot of individual observer and consensus (STAPLE) contour volumes across contouring sessions. Rad1: radiologist 1; Rad2: radiologist 2; RO1: radiation oncologist 1; RO2: radiation oncologist 2.

Figure 3. Boxplot and summary statistics of observer-STAPLE comparisons for A) DSC, B) JI, C) HDmax, and D) MDA across delineation sessions.

Figure 4. Scatterplot of generalised conformity index (CIgen) versus volume of delineations. Volume of the consensus STAPLE contour for each delineation is shown. CIgen significantly correlated with contour volume for T2w ($r = 0.575$, $R^2 = 0.330$, $p = 0.008$) and CoReg GTV delineations ($r = 0.532$, $R^2 = 0.283$, $p = 0.016$), but not for DWI or CoReg GTVs or subvolumes (SubV) ($p > 0.05$).

no significant difference between volume of DWI and CoReg contours ($p > 0.05$). CIgen significantly correlated with contour volume on T2w ($r = 0.550$, $R^2 = 0.303$, $p = 0.012$) and CoReg ($r = 0.532$, $R^2 = 0.283$, $p = 0.016$) but not on DWI delineations ($r = 0.205$, $R^2 = 0.042$, $p = 0.386$) (Figure 4).

By all three overlap metrics (CIgen, DSC, JI), CoReg contours demonstrated significantly lower interobserver agreement than T2w ($p < 0.001$ for all) and DWI ($p = 0.017$, $p = 0.021$, $p = 0.019$) contours, while there were no significant differences between T2w and DWI contours ($p > 0.05$ for all).

By both distance metrics (HDmax and MDA), T2w-based delineations performed significantly better than DWI ($p = 0.033$, $p = 0.021$) and CoReg ($p = 0.002$, $p < 0.001$) delineations; while no significant difference between DWI and CoReg sessions was found ($p > 0.05$).

Review of the data suggested the presence of a negative outlier case on DWI-based contouring. In this instance, two out of four observers included adjacent perirectal collections with restricted diffusion as tumour GTV (Figure 5), resulting in average to poor performance across all metrics (CIgen: 0.287, DSC: 0.638, JI: 0.521, HDmax: 26.33 mm, MDA: 5.81 mm). This error did not occur in the T2w contouring session, however one observer still included these areas in CoReg, resulting in poor CIgen (0.145).

Manual subvolume delineations

Both radiation oncologists and radiologists were able to delineate volumes of restricted diffusion within their CoReg GTV contours (Figures 6 and 7) with moderate confidence (mean:

Figure 5. Outlier case due to peri-rectal collections with restricted diffusion on A) T2w, B) ADC and C) b800 image sequences. This was misattributed as extensive tumour spread by two observers on the DWI contouring session, by one observer on CoReg contouring, and not by any observer on T2w. Red: T2w STAPLE contour; Green: DWI STAPLE contour; Cyan: CoReg STAPLE contour.

2.41). The mean STAPLE SubV contour volume was 9.90 ml (SD: 7.66 ml), with an average ratio of SubV:CoReg volume size of 28.4% (SD: 8.6%). ICC (0.32, 95% CI: 0.13–0.54) and CIgen (0.318, SD: 0.112) for SubV contour volumes were poor, and no correlation between CIgen and volume was observed ($p > 0.05$).

Observer-STAPLE comparisons of SubV contours generally showed greater variability than those of whole tumour GTVs. Specifically, DSC (0.666, SD: 0.126) and JI (0.531, SD: 0.126) of subvolume delineations were significantly lower than all GTV

Figure 6. Examples of semi-automatically generated contours based on ADC threshold of 40th centile, compared to manually-defined consensus STAPLE subvolumes. A-C) Axial, coronal and sagittal view of T2w MRI; D-F) axial, coronal and sagittal views of ADC. Cyan: whole tumour GTV (CoReg); purple: consensus (STAPLE) SubV delineations; shaded green: thresholded volume.

Figure 7. Examples of semi-automatically generated contours based on ADC threshold of 40th centile, compared to manually-defined consensus STAPLE subvolumes. A-C) Axial, coronal and sagittal view of T2w MRI; D-F) axial, coronal and sagittal views of ADC. Cyan: whole tumour GTV (CoReg); purple: consensus (STAPLE) SubV delineations; shaded green: thresholded volume

contour groups ($p < 0.001$ for all comparisons). The MDA of SubV contours (2.08 mm, SD: 1.21 mm) was significantly greater than that of T_2W -based GTV delineations ($p = 0.004$), however in other between-group comparisons of HDmax and MDA, no significant difference in SubV delineation performance was found ($p > 0.05$).

ADC centile-based contour comparisons

Pairwise comparisons between consensus SubV contours and automated ADC-centile based subregions are summarised in Table 2 and examples are shown in Figures 6 and 7. Subvolumes corresponding to the lowest 40th centile of intratumoral ADC values demonstrated best performance when comparing contour overlap with expert delineations, with mean DSC and JI of 0.580 (SD: 0.112) and 0.418 (SD: 0.116) respectively. Both HDmax and MDA tended to improve with lower cut-off values, to an optimal 14.5 mm (SD: 7.78 mm) at the 25th centile for HDmax and 2.31 mm (SD: 1.06 mm) at the 35th centile for MDA.

Contour cases were reviewed to evaluate causes for poor agreement. In several instances, CoReg consensus contours, while conforming to tumour boundaries on T_2W imaging, included small portions of luminal contents or gas on the co-registered DWI sequences. This occurred even in instances of clinically satisfactory rigid image registration, as checked by the study radiation oncologists. As a result, some luminal areas were

Table 2. ADC and interobserver variability measures of semi-automatically generated subvolume based on ADC nth-centile thresholding of consensus co-registered gross tumour volumes

Centile	ADC threshold (10^{-6} mm ² /s)		DSC		JI		HDmax (mm)		MDA (mm)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<75th	1371.0	92.5	0.492	0.125	0.335	0.112	19.27	9.09	3.82	1.95
<70th	1321.7	88.1	0.509	0.128	0.350	0.116	19.25	9.10	3.63	1.87
<65th	1276.6	86.6	0.523	0.127	0.364	0.118	19.05	8.87	3.48	1.83
<60th	1233.9	84.2	0.540	0.127	0.379	0.120	19.00	8.90	3.31	1.81
<55th	1192.6	80.6	0.560	0.130	0.400	0.129	16.66	6.99	2.91	1.52
<50th	1156.9	80.8	0.570	0.126	0.409	0.127	16.41	6.87	2.77	1.42
<45th	1119.6	81.5	0.578	0.120	0.416	0.122	15.81	6.71	2.58	1.30
<40th	1082.6	82.9	0.581	0.112	0.418	0.116	15.15	6.28	2.44	1.23
<35th	1044.9	83.7	0.577	0.105	0.413	0.109	14.84	6.43	2.31	1.06
<30th	1008.7	84.5	0.556	0.111	0.393	0.109	15.16	8.36	2.46	1.46
<25th	970.3	86.2	0.530	0.109	0.368	0.100	14.55	7.78	2.40	1.40

ADC: apparent diffusion coefficient; DSC: Dice Similarity Coefficient; JI: Jaccard Index; HDmax: maximum Hausdorff distance; MDA: mean distance to agreement

Values shown are the mean of all patients. Bold entries indicate best values for each given metric across the cohort.

erroneously included within the thresholded 'restricted diffusion' volume. Expert delineations of SubV did not make this error (Figure 8).

Validation set

The same pairwise comparisons were undertaken on a small validation set of four additional patients who underwent standardised planning MRI prior to treatment. Delineations were performed by the same radiation oncologists as previous. T_2W and DW-MRI ($b = 50, 400, 800$) sequences were both acquired in the anatomical axial plane at 1.5T with night-before bowel preparation. Agreement between manual and semi-automated subvolume delineations appeared to be improved with the new protocol, with DSC, JI, HDmax and MDA of 0.669, 0.507, 13.31 mm and 1.91 mm respectively for the 40th centile (Figure 9).

Figure 8. Example of poor agreement between consensus STAPLE subvolume (SubV) delineation and ADC thresholding technique. Slight misregistration and changes in bowel gas between T_2W and DWI scans result in inclusion of bowel lumen anteriorly as boost volume. Cyan: whole tumour GTV (CoReg); purple: consensus (STAPLE) SubV delineations; shaded green: 40th centile ADC thresholded volume.

Figure 9. A-C) T_2W and D-F) ADC images of a favourable case with dedicated planning MRI. Images acquired co-axially with night-before bowel preparation. Manually-delineated and semi-automated delineation based on 40th centile ADC demonstrated substantial agreement (DSC: 0.801, JI: 0.668, HD: 7.12 mm, MDA: 1.03 mm). Cyan: whole tumour GTV (CoReg); purple: consensus (STAPLE) SubV delineations; shaded green: thresholded volumes

Table 3. Comparison of interobserver variability of MRI-based rectal tumour GTV across existing rectal MRI delineation studies

Publication	Design	No. subjects	No. observers	T ₂ W	DWI	CoReg	SubV	Volume (mL) differences
Regini et al. 2014	Retro.	26	2	ICC: 0.889	ICC: 0.893	-	-	T ₂ W ≈ DWI
Burbach et al. 2016	Prosp.	24	3	ICC: 0.820 CIgen: 0.699 HDmax: 18.59 MDA: 1.80	ICC: 0.824 CIgen: 0.707 HDmax: 14.42 MDA: 1.49	ICC: 0.815 CIgen: 0.690 HDmax: 15.42 MDA: 1.68	-	T ₂ W > DWI T ₂ W > CoReg
Rosa et al. 2019	Retro.	50	2	ICC: 0.989 DSC: 0.80	ICC: 0.992 DSC: 0.77	-	-	T ₂ W > DWI
Present study	Retro.	20	4	ICC: 0.91 CIgen: 0.533 DSC: 0.818 JI: 0.704 HDmax: 9.76 MDA: 1.19	ICC: 0.74 CIgen: 0.510 DSC: 0.804 JI: 0.689 HDmax: 12.76 MDA: 1.66	ICC: 0.56 CIgen: 0.448 DSC: 0.768 JI: 0.644 HDmax: 12.94 MDA: 2.08	ICC: 0.32 CIgen: 0.318 DSC: 0.666 JI: 0.531 HDmax: 12.74 MDA: 1.92	T ₂ W < DWI T ₂ W < CoReg

CIgen, generalised conformity index; CoReg, delineations undertaken on co-registered T₂W/DWI; DSC, Dice Similarity Coefficient; DWI, diffusion weighted imaging/MRI; HDmax, maximum Hausdorff distance; ICC, intraclass correlation coefficient; JI, Jaccard Index; MDA, mean distance to agreement; Retro, retrospective; SubV, restricted diffusion subvolume; T₂W, T₂ weighted MRI; prosp, prospective.

Subvolume delineation performance in this study is also shown for comparison.

DISCUSSION

To our knowledge, this is the first study to evaluate the feasibility of semi-automated tumour subvolume delineations with DW-MRI for radiotherapy planning in LARC. Our results indicate that expert observers can delineate restricted diffusion subvolumes with moderate agreement and confidence; and that similar volumes may be created through thresholding of a conventional GTV based on intratumoral ADC values of a co-registered image. However, image registration and other issues identified here need to be optimised before such an approach could be clinically used.

Few studies have evaluated interobserver agreement of MRI-based contours for rectal cancers (Table 3). In contrast to existing literature,^{7,18,19} this study found that DW-MRI delineations were significantly larger than corresponding T₂W contours; and while overlap measures were similar between T₂W and DW-MRI sessions, delineations based on T₂W imaging demonstrated significantly better boundary agreement. DWI, specifically ADC images, may lack the necessary anatomical information to accurately define tumour margins due to resolution limitations or similar contrast with normal adjacent tissues. Correspondingly, observer confidence in DWI contours was also lower on average than T₂W or CoReg sessions. We identified an outlier case of note (Figure 5) where adjacent perirectal collections were mistaken by several observers to reflect tumour extension based on DW-MRI findings. Similar (or subtler) artefacts may impact delineation, particularly in the setting of EPI or non-coaxial acquisitions and variable observer familiarity with DW-MRI.

This boundary uncertainty may have persisted in co-registered delineations, which demonstrated similar HDmax and MDA to contours completed with DWI alone. Compared to the similar

‘Combi’ co-registered delineations undertaken by Burbach et al. (2016) with three radiation oncologists,¹⁸ our CoReg volumes had poorer CIgen (0.448 vs 0.690) and ICC (0.56 vs 0.815) but similar agreement in HDmax (12.94 vs 15.42mm) and MDA (2.08 vs 1.68mm). In part, this may be due to the smaller tumour volumes in this study, the inclusion of an additional observer, and variable contouring experience and technique with the inclusion of radiologists as observers. We note in particular that radiologist delineations tended to be smaller than those of the radiation oncologists (Figure 3). While all image registrations were checked by an experienced radiation oncologist, it is also possible that some residual alignment inaccuracy between tumour-axial T₂W and anatomical-axial DWI sequences contributed to poorer CoReg interobserver agreement: in cases of misregistration, different observer preference for which imaging sequence to follow may have increased variability.

We also examined whether restricted diffusion tumour subvolumes on DW-MRI could be identified and manually delineated within a larger GTV for the purposes of intratumoral radiotherapy boosting. The primary assumptions underlying this approach are that areas of restricted diffusion on DW-MRI correspond to histopathological areas of hypercellularity, poor differentiation or other correlates of radioresistance; and that specific targeting of these regions with higher radiotherapy doses would provide clinical benefit through optimising reductions in malignant cell count.⁸ In LARC, this could improve rates of tumour regression and pCR, which are associated with better oncological and surgical outcomes.²⁰

To date, few studies have addressed the underlying relationship between cellularity and ADC in primary rectal adenocarcinoma. Surov et al. (2017) describe trends towards lower mean ADC

value and cell count ($r = -0.334$, $p = 0.190$) and total nucleic area ($r = -0.453$, $p = 0.068$) of biopsy specimens,²¹ while a study of resected hepatic colorectal metastases found similar associations between ADC and markers of cellular proliferation.²² Further evidence must be extrapolated from histopathological studies of other malignancies.²³ Conversely, higher pre-treatment mean tumour ADC may be associated with poor response to neoadjuvant therapy due to the presence of necrosis²⁴; however, it is unclear to what extent aggregate measures of tumour ADC could be used to guide focal therapy. Targeting of necrotic areas would require means of reliable and reproducible differentiation from benign tissue with similar diffusion characteristics, which in turn may require more sophisticated analysis of the tumour ADC histogram skew and kurtosis.²⁵ Given these issues, and to provide a point of comparison with similar approaches taken in prostate cancer²⁶ and elsewhere, restricted diffusion subvolumes were targeted in this study.

Despite reasonable observer confidence in SubV delineations, these contours demonstrated significantly lower agreement than GTVs in measures of contour overlap. This would be expected in delineations of smaller volumes; however, we also identified significant variability between observers in the raw volume of SubV contours (ICC: 0.32) despite the use of a standardised instruction set for all observers.

While this could be better mitigated with additional image training sets and with greater observer experience in the task, subvolume delineation based on visual assessment of functional imaging would remain inherently arbitrary to a degree. Automated approaches to delineation may improve standardisation: Heeswijk *et al.* (2015) evaluated LARC GTV delineation based on voxel intensity on DW-MRI (b1000–1100 images) as an alternative to manual contouring.¹³ In a subsequent study, the authors found that thresholding based on the 45th ADC centile of non-specific volumes encapsulating the tumour and mesorectum could be used as an alternative to manual GTV delineation, describing significant time and efficiency benefits to doing so.¹² Similar centile- or absolute threshold ADC-based approaches have been suggested for dominant intraprostatic lesion (DIL) prostate radiotherapy boost^{11,27,28} given the known variability of DIL localisation with multiparametric MRI.^{29,30}

Here, a similar centile-based thresholding approach was applied within a consensus GTV given the intended focus on intratumoural dose escalation. Overall, the 40th centile of intratumoural ADC values demonstrated closest pairwise overlap with consensus SubV contours of restricted diffusion, presumed to be reflective of regions of greatest cellularity. While several cases demonstrated high degrees of agreement between expert and threshold-defined subvolumes (Figures 6 and 7), multiple instances of poor agreement due to inadvertent inclusion of gaseous bowel lumen in consensus CoReg contours (Figure 8) led to average to poor overall measures of agreement with SubV. Such errors would be amenable to manual correction; however, the issue of T_2W -DWI discordance and subsequent changes in the GTV ADC histogram would still affect replicability across patients.

Ideally, for both whole-tumour and automated/manual subvolume delineation methods, dedicated radiotherapy planning MRI should be performed with co-axial T_2W and DWI acquisition. Rapid acquisition sequences such as RESOLVE or DWI-SPLICE alongside antiperistaltic agents may reduce positional discrepancies between co-registered images.¹⁸ B-value choice should also be standardised across patients and optimised for repeatability.³¹ Finally, bowel preparation (enema or laxative) and could be considered to reduce susceptibility artefact and registration issues associated with luminal gas and bowel motion; however this may then need to be repeated prior to each fraction of treatment for consistency.

We undertook a brief evaluation of the effectiveness of some of these measures by performing a similar analysis on a set of four consecutive patients scanned more recently in our department after introduction of standardised planning MRI. Measures of interobserver agreement appeared to be improved for SubV contours, and registration errors as described previously were less common with co-axial acquisition (Figure 9). A planning study to evaluate the dosimetric feasibility of this ADC-derived subvolume boost approach will be undertaken using these improved acquisition practices in a larger cohort.

The main strengths of this study are its multifaceted evaluation of delineation concordance for both GTV and tumour subvolumes, along with the inclusion of two radiologists and two radiation oncologists with specialist experience. Limitations pertain to the source data available, small sample size and the retrospective study design. Most scans included in the main part of this study were performed for diagnostic purposes and thus were not specifically tailored for radiotherapy planning. Differences in acquisition sequence parameters may have introduced variability into the quantitative thresholding, though this was mitigated through the use of patient-specific tumour ADC centile values. While the 40th centile is suggested as most closely matching clinician-defined volumes, this requires further validation in a larger standardised patient/imaging cohort. Additionally, as is common to most studies of this type, the consensus STAPLE contour is taken to be the 'ground-truth', with observer-pair and observer-consensus comparisons indicative of accuracy. In lieu of histopathological validation, assessment of subvolume delineations is inherently limited to these measures and may be affected by observer expertise. In the present study, lack of observer experience in the task may have contributed to the lower reported confidence in DWI contours and larger DWI GTV volumes; additional imaging training sets would be beneficial for reducing this source of variability.

Finally, this study did not evaluate registration with planning-CT images. While MRI-based subvolume delineation may be a feasible approach for radiotherapy boost protocols, CT-MRI registration inconsistency and interfraction motion would need to be accounted for with appropriate expansions.³² The recent results of Kleijnen *et al.* (2018) suggest only marginal benefits with image-matching to rectal wall compared to standard bony anatomy.³³ Measures to improve online tumour

localisation would be beneficial in a dose escalation setting to small subvolumes, such as verification with pre-treatment MRI or the implantation of fiducial markers.^{34,35} Alternatively, MRI-linac treatment with online visualisation of restricted diffusion volumes, or highly targeted endoluminal techniques such as high dose-rate brachytherapy, may be more suited to clinical implementation. While there appears to be reasonable intrasubject consistency of tumour ADC across multiple time points,³⁶ further validation of subvolume or voxel-based ADC histogram repeatability between should also be undertaken.

CONCLUSION

DWI-based radiotherapy planning for rectal cancer may allow for targeted dose escalation to biologically relevant tumour subregions; however, high fidelity between co-registered T_2 -W- and DW-MRI, alongside optimised acquisition practices, is required for accurate and reproducible delineation. Contours corresponding to the 40th centile intratumoral ADC value demonstrated moderate agreement with consensus manual segmentations of restricted diffusion. A radiotherapy planning study evaluating this approach in a larger cohort with dedicated radiotherapy planning MRI will be undertaken.

REFERENCES

- Burbach JPM, den Harder AM, Intven M, van Vulpen M, Verkooijen HM, Reerink O. Impact of radiotherapy boost on pathological complete response in patients with locally advanced rectal cancer: a systematic review and meta-analysis. *Radiother Oncol* 2014; **113**: 1–9. doi: <https://doi.org/10.1016/j.radonc.2014.08.035>
- Hearn N, Atwell D, Cahill K, Elks J, Vignarajah D, Lagopoulos J, et al. Neoadjuvant radiotherapy dose escalation in locally advanced rectal cancer: a systematic review and meta-analysis of modern treatment approaches and outcomes. *Clin Oncol* 2020;12 Jul 2020. doi: <https://doi.org/10.1016/j.clon.2020.06.008>
- Valentini V, Gambacorta MA, Cellini F, Aristei C, Coco C, Barbaro B, et al. The INTERACT Trial: Long-term results of a randomised trial on preoperative capecitabine-based radiochemotherapy intensified by concomitant boost or oxaliplatin, for cT2 (distal)-cT3 rectal cancer. *Radiother Oncol* 2019; **134**: 110–8. doi: <https://doi.org/10.1016/j.radonc.2018.11.023>
- Appelt AL, Pløen J, Vogelius IR, Bentzen SM, Jakobsen A. Radiation dose-response model for locally advanced rectal cancer after preoperative chemoradiation therapy. *Int J Radiat Oncol Biol Phys* 2013; **85**: 74–80. doi: <https://doi.org/10.1016/j.ijrobp.2012.05.017>
- Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of image-guided radiotherapy will be Mr guided. *Br J Radiol* 2017; **90**: 20160667. doi: <https://doi.org/10.1259/bjr.20160667>
- Pham TT, Liney GP, Wong K, Barton MB. Functional MRI for quantitative treatment response prediction in locally advanced rectal cancer. *Br J Radiol* 2017; **90**: 20151078. doi: <https://doi.org/10.1259/bjr.20151078>
- Rosa C, Caravatta L, Delli Pizzi A, Di Tommaso M, Cianci R, Gasparini L, et al. Reproducibility of rectal tumor volume delineation using diffusion-weighted MRI: agreement on volumes between observers. *Cancer Radiother* 2019; **23**: 216–21. doi: <https://doi.org/10.1016/j.canrad.2018.10.004>
- Nahum AE, Uzan J. Radio)biological optimization of external-beam radiotherapy. *Comput Math Methods Med* 2012; **2012**: 1–13. doi: <https://doi.org/10.1155/2012/329214>
- van der Heide UA, Houweling AC, Groenendaal G, Beets-Tan RGH, Lambin P. Functional MRI for radiotherapy dose painting. *Magn Reson Imaging* 2012; **30**: 1216–23. doi: <https://doi.org/10.1016/j.mri.2012.04.010>
- Thorwarth D. Biologically adapted radiation therapy. *Z Med Phys* 2018; **28**: 177–83. doi: <https://doi.org/10.1016/j.zemedi.2017.08.001>
- Grönlund E, Johansson S, Nyholm T, Thellenberg C, Ahnesjö A. Dose painting of prostate cancer based on Gleason score correlations with apparent diffusion coefficients. *Acta Oncol* 2018; **57**: 574–81. doi: <https://doi.org/10.1080/0284186X.2017.1415457>
- van Heeswijk MM, Lambregts DMJ, Maas M, Lahaye MJ, Ayas Z, Slenter JMGM, et al. Measuring the apparent diffusion coefficient in primary rectal tumors: is there a benefit in performing histogram analyses? *Abdom Radiol* 2017; **42**: 1627–36. doi: <https://doi.org/10.1007/s00261-017-1062-2>
- van Heeswijk MM, Lambregts DMJ, van Griethuysen JJM, Oei S, Rao S-X, de Graaff CAM, et al. Automated and semiautomated segmentation of rectal tumor volumes on diffusion-weighted MRI: can it replace manual volumetry? *Int J Radiat Oncol Biol Phys* 2016; **94**: 824–31. doi: <https://doi.org/10.1016/j.ijrobp.2015.12.017>
- Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. *Radiother Oncol* 2016; **121**: 169–79. doi: <https://doi.org/10.1016/j.radonc.2016.09.009>
- Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. *Phys Med Biol* 2009; **54**: 2863–73. doi: <https://doi.org/10.1088/0031-9155/54/9/018>
- 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**: 903–21. doi: <https://doi.org/10.1109/TMI.2004.828354>
- Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. *Strahlenther Onkol* 2012; **188**: 160–7. doi: <https://doi.org/10.1007/s00066-011-0027-6>
- Burbach JPM, Kleijnen J-PJ, Reerink O, Seravalli E, Philippens MEP, Schakel T, et al. Inter-Observer agreement of MRI-based tumor delineation for preoperative radiotherapy boost in locally advanced rectal cancer. *Radiother Oncol* 2016; **118**: 399–407. doi: <https://doi.org/10.1016/j.radonc.2015.10.030>
- Regini F, Gourtsoyanni S, Cardoso De Melo R, Charles-Edwards GD, Griffin N, Parikh J, et al. Rectal tumour volume (GTV) delineation using T2-weighted and diffusion-weighted MRI: implications for radiotherapy planning. *Eur J Radiol* 2014; **83**: 768–72. doi: <https://doi.org/10.1016/j.ejrad.2014.02.007>
- Martin ST, Heneghan HM, Winter DC. Systematic review and meta-analysis of outcomes following pathological complete response to neoadjuvant chemoradiotherapy for rectal cancer. *Br J Surg* 2012; **99**: 918–28. doi: <https://doi.org/10.1002/bjs.8702>
- Surov A, Meyer HJ, Höhn A-K, Behrmann C, Wienke A, Spielmann RP, et al. Correlations

- between intravoxel incoherent motion (IVIM) parameters and histological findings in rectal cancer: preliminary results. *Oncotarget* 2017; **8**: 21974–83. doi: <https://doi.org/10.18632/oncotarget.15753>
22. Heijmen L, Ter Voert EEGW, Nagtegaal ID, Span P, Bussink J, Punt CJA, et al. Diffusion-Weighted MR imaging in liver metastases of colorectal cancer: reproducibility and biological validation. *Eur Radiol* 2013; **23**: 748–56. doi: <https://doi.org/10.1007/s00330-012-2654-4>
 23. Surov A, Meyer HJ, Wienke A. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis. *Oncotarget* 2017; **8**: 59492–9. doi: <https://doi.org/10.18632/oncotarget.17752>
 24. Xie H, Sun T, Chen M, Wang H, Zhou X, Zhang Y, et al. Effectiveness of the apparent diffusion coefficient for predicting the response to chemoradiation therapy in locally advanced rectal cancer: a systematic review and meta-analysis. *Medicine* 2015; **94**: e517. doi: <https://doi.org/10.1097/MD.0000000000000517>
 25. Enkhbaatar N-E, Inoue S, Yamamuro H, Kawada S, Miyaoka M, Nakamura N, et al. Mr imaging with apparent diffusion coefficient histogram analysis: evaluation of locally advanced rectal cancer after chemotherapy and radiation therapy. *Radiology* 2018; **288**: 129–37. doi: <https://doi.org/10.1148/radiol.2018171804>
 26. Casares-Magaz O, van der Heide UA, Rørvik J, Steenbergen P, Muren LP. A tumour control probability model for radiotherapy of prostate cancer using magnetic resonance imaging-based apparent diffusion coefficient maps. *Radiother Oncol* 2016; **119**: 111–6. doi: <https://doi.org/10.1016/j.radonc.2016.02.030>
 27. Zamboglou C, Thomann B, Koubar K, Bronsert P, Krauss T, Rischke HC, et al. Focal dose escalation for prostate cancer using ⁶⁸Ga-HBED-CC PSMA PET/CT and MRI: a planning study based on histology reference. *Radiat Oncol* 2018; **13**: 81. doi: <https://doi.org/10.1186/s13014-018-1036-8>
 28. Abedi I, Tavakkoli MB, Jabbari K, Amouheidari A, Yadegarfar G. Dosimetric and radiobiological evaluation of multiparametric MRI-guided dose painting in radiotherapy of prostate cancer. *J Med Signals Sens* 2017; **7**: 114–21.
 29. Rischke HC, Nestle U, Fechter T, Doll C, Volegova-Neher N, Henne K, et al. 3 Tesla multiparametric MRI for GTV-definition of Dominant Intraprostatic Lesions in patients with Prostate Cancer--an interobserver variability study. *Radiat Oncol* 2013; **8**: 183. doi: <https://doi.org/10.1186/1748-717X-8-183>
 30. van Schie MA, Dinh CV, Houdt PJvan, Pos FJ, Heijmink SWTJP, Kerkmeijer LGW, et al. Contouring of prostate tumors on multiparametric MRI: evaluation of clinical delineations in a multicenter radiotherapy trial. *Radiother Oncol* 2018; **128**: 321–6. doi: <https://doi.org/10.1016/j.radonc.2018.04.015>
 31. Chen L, Shen F, Li Z, Lu H, Chen Y, Wang Z, et al. Diffusion-Weighted imaging of rectal cancer on repeatability and cancer characterization: an effect of b-value distribution study. *Cancer Imaging* 2018; **18**: 43. doi: <https://doi.org/10.1186/s40644-018-0177-1>
 32. Kleijnen J-PJE, van Asselen B, Van den Begin R, Intven M, Burbach JPM, Reerink O, et al. Mri-Based tumor inter-fraction motion statistics for rectal cancer boost radiotherapy. *Acta Oncol* 2019; **58**: 232–6. doi: <https://doi.org/10.1080/0284186X.2018.1532598>
 33. Kleijnen J-PJE, van Asselen B, Intven M, Burbach JPM, Philippens MEP, Legendijk JJW, et al. Does setup on rectal wall improve rectal cancer boost radiotherapy? *Radiat Oncol* 2018; **13**: 61. doi: <https://doi.org/10.1186/s13014-018-1011-4>
 34. van den Ende RPJ, Rigter LS, Kerkhof EM, van Persijn van Meerten EL, Rijkmans EC, Lambregts DMJ, et al. Mri visibility of gold fiducial markers for image-guided radiotherapy of rectal cancer. *Radiother Oncol* 2019; **132**: 93–9. doi: <https://doi.org/10.1016/j.radonc.2018.11.016>
 35. Rigter LS, Rijkmans EC, Inderson A, van den Ende RPJ, Kerkhof EM, Ketelaars M, et al. EUS-guided fiducial marker placement for radiotherapy in rectal cancer: feasibility of two placement strategies and four fiducial types. *Endosc Int Open* 2019; **7**: E1357–64. doi: <https://doi.org/10.1055/a-0958-2148>
 36. Intven M, Reerink O, Philippens MEP. Repeatability of diffusion-weighted imaging in rectal cancer. *J Magn Reson Imaging* 2014; **40**: 146–50. doi: <https://doi.org/10.1002/jmri.24337>