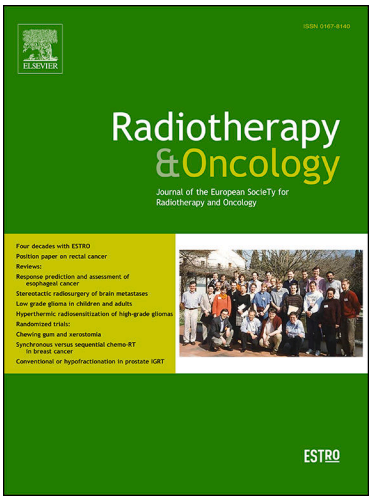
Journal Pre-proofs



Original Article

Gross tumour volume delineation in anal cancer on T2-weighted and diffusion-weighted MRI – reproducibility between radiologists and radiation oncologists and impact of reader experience level and DWI image quality

Lisa A. Min, Younan J.L. Vacher, Luc Dewit, Mila Donker, Carmelo Sofia, Baukelien van Triest, Paula Bos, Joost J.W. van Griethuysen, Monique Maas, Regina G.H. Beets-Tan, Doenja M.J. Lambregts

PII: S0167-8140(20)30335-2
DOI: https://doi.org/10.1016/j.radonc.2020.06.012
Reference: RADION 8376

To appear in: *Radiotherapy and Oncology*

Received Date: 13 March 2020
Revised Date: 6 June 2020
Accepted Date: 7 June 2020

Please cite this article as: Min, L.A., Vacher, Y.J.L., Dewit, L., Donker, M., Sofia, C., van Triest, B., Bos, P., van Griethuysen, J.J.W., Maas, M., Beets-Tan, R.G.H., Lambregts, D.M.J., Gross tumour volume delineation in anal cancer on T2-weighted and diffusion-weighted MRI – reproducibility between radiologists and radiation oncologists and impact of reader experience level and DWI image quality, *Radiotherapy and Oncology* (2020), doi: https://doi.org/10.1016/j.radonc.2020.06.012

TITLE:

Gross tumour volume delineation in anal cancer on T2-weighted and diffusion-weighted MRI – reproducibility between radiologists and radiation oncologists and impact of reader experience level and DWI image quality

AUTHORS

Lisa A. Min[a,b], Younan J.L. Vacher[a], Luc Dewit[c], Mila Donker[c], Carmelo Sofia[d], Baukelien van Triest[c], Paula Bos[a,b,e], Joost J.W. van Griethuysen[a, b], Monique Maas[a], Regina G.H. Beets-Tan[a, b], Doenja M.J. Lambregts[a].

a. Department of Radiology, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, The Netherlands
b. GROW School for Oncology and Developmental Biology – University of Maastricht, Universiteitssingel 40, Maastricht, The Netherlands
c. Department of Radiation Oncology, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, The Netherlands
d. Department of Biomedical Sciences and Morphologic and Functional Imaging, Policlinico Universitario G. Martino, University of Messina, Via Consolare Valeria 1, Messina, Italy
e. Department of Head and Neck Oncology and Surgery, The Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, The Netherlands

CORRESPONDING AUTHOR

Lisa A. Min, M.D. M.Sc.
POB 90203, 1006 BE, AMSTERDAM
l.min@nki.nl

DELARATION OF INTEREST

Conflicts of interest: none

ABSTRACT

Purpose

To assess how gross tumour volume (GTV) delineation in anal cancer is affected by interobserver variations between radiologists and radiation oncologists, expertise level, and use of T2-weighted MRI (T2W-MRI) vs. diffusion-weighted imaging (DWI), and to explore effects of DWI quality.

Methods and materials

We retrospectively analyzed the MRIs (T2W-MRI and b800-DWI) of 25 anal cancer patients. Four readers (senior&junior radiologist; senior&junior radiation oncologist) independently delineated GTVs, first on T2W-MRI only and then on DWI (with reference to T2W-MRI). Maximum Tumour Diameter (MTD) was calculated from each GTV. Mean GTVs/MTDs were compared between readers and between T2W-MRI vs. DWI. Interobserver agreement was calculated as intraclass correlation coefficient (ICC), Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD). DWI image quality was assessed using a 5-point artefact scale.

Results

Interobserver agreement between radiologists vs. radiation oncologists and between junior vs. senior readers was good-excellent, with similar agreement for T2W-MRI and DWI (e.g. ICCs 0.72-0.94 for T2W-MRI and 0.68-0.89 for DWI). There was a trend towards smaller GTVs on DWI, but only for the radiologists (P=0.03-0.07). Moderate-severe DWI-artefacts were observed in 11/25 (44%) cases. Agreement tended to be lower in these cases.

Conclusion

Overall interobserver agreement for anal cancer GTV delineation on MRI is good for both radiologists and radiation oncologists, regardless of experience level. Use of DWI did not improve agreement. DWI artefacts affecting GTV delineation occurred in almost half of the patients, which may severely limit the use of DWI for radiotherapy planning if no steps are undertaken to avoid them.

KEYWORDS (MeSh)

Anus Neoplasms, Magnetic Resonance Imaging, Diffusion Magnetic Resonance Imaging, Chemoradiotherapy, Artifacts

INTRODUCTION

In many centres Magnetic Resonance Imaging (MRI) is routinely used for the locoregional staging of anal cancer, as it offers excellent soft tissue contrast for detailed anatomical assessment of the anal canal and its surrounding structures. Recently, there has been a growing interest in the use of MRI for radiotherapy planning [1]. This is because MRI – instead or in combination with CT, and in addition to digital rectal examination – aids in identification and delineation of the radiotherapy target area, the Gross Tumour Volume (GTV) [2-5]. Accurate GTV delineation is essential for effective dose delivery and to avoid toxicity to surrounding tissues, and is also an important prerequisite when considering dose escalation strategies. Recently introduced hybrid MR-linear accelerator (MRL) systems allow visualization of tumour changes during the course of treatment that can be used as input to adapt and optimize the treatment plan. This offers new possibilities for boosting strategies targeting specific areas of residual tumour, provided that these areas can accurately be defined on MRI [6, 7].

Tumour delineation on anatomical MRI requires detailed knowledge of cross-sectional image anatomy and morphology. This can be challenging, in particular for less experienced readers. Addition of diffusion-weighted imaging (DWI) to anatomical MRI sequences has been suggested to improve lesion conspicuity and interobserver reproducibility [8, 9]. DWI is a functional MRI technique that highlights hypercellular tissues, making it an attractive tool to discriminate malignant tumour from the surrounding (normocellular) anatomy. As anal squamous cell carcinomas are usually hypercellular, these tumours will typically show restricted diffusion and appear bright on DWI [10].

A potential disadvantage of DWI is that the technique is prone to artefacts [11]. Evidence detailing how this affects the applicability of DWI for radiotherapy planning is however sparse. Further investigation is also required on the effect of interobserver variability, especially between readers of different specialties and experience levels. In several reports, GTV delineations were exclusively performed by radiologists [8, 12] while in clinical practice GTV delineations are usually performed by radiation oncologists, who tend to be less experienced in reading MRI. One previous study by Rosa et al. showed good interobserver agreement for GTV delineation in rectal cancer between an expert radiologist and expert radiation oncologist [9], but these results remain to be confirmed for anal cancer.

The aims of the current study are therefore (1) to assess the interobserver reproducibility for MRI-based GTV delineations in anal cancer between radiologists and radiation oncologists of varying

expertise levels, (2) to evaluate how the use of DWI compares to that of anatomical T2-weighted MRI, and (3) to explore effects of DWI image quality on interobserver reproducibility.

METHODS

The retrospective analysis of routinely acquired clinical imaging data for the purpose of this study was approved by the institutional review board. Informed consent was waived.

Patients

From our hospitals' database (2010-2018) we identified 39 patients with histologically proven anal squamous cell carcinoma who underwent pre-therapy MRI including standardized T2-weighted MRI and DWI at 3.0T. Six patients were excluded because the anal tumour was partially outside the MRI field of view, eight patients were excluded because no visible tumour could be identified on MRI (e.g. when MRI was performed after biopsy or diagnostic resection). The remaining 25 patients were included in the study.

MR imaging

MRIs were performed in supine position on a 3T MR scanner (Achieva; Philips Medical System) using a phased-array body coil. The protocol consisted of T2-weighted (T2W) fast spin-echo sequences in three orthogonal planes, and an axial echo planar imaging diffusion-weighted sequence (highest b-value: 800 s/mm2) angled in the same plane as the axial T2W-MRI. For detailed parameters see **Supplementary Table 1**. No spasmolytics or bowel preparation were given.

Gross Tumour Volume (GTV) delineation

MR images were analysed by four independent readers, two board-certified radiologists and two board-certified radiation oncologists:

- Senior radiologist, a pelvic MR expert working at an oncologic referral centre (± 9 years of dedicated pelvic MRI experience)
- Junior radiologist, a general radiologist working at a university hospital (± 5 years of pelvic MRI experience)
- Senior radiation oncologist, specialized in GI oncology, working at an oncologic referral centre (± 30 years of experience)
- Junior radiation oncologist, working at an oncologic referral centre (±1.5 years after completion of residency training)

MRIs were transferred to an offline workstation for tumour delineation, which was performed using the open-source segmentation software 3D Slicer (version 4.8.1, slicer.org). Each reader manually delineated the gross tumour volume (GTV) for each patient by free-hand tracing of the tumour

boundaries on each consecutive slice. GTV-delineation was first done on axial T2-weighted MRI only (blinded to DWI) and subsequently on the b800-DWI (with anatomical reference to T2W-MRI). The maximum tumour diameter (MTD) and whole tumour volume (GTV) were automatically calculated from the delineations using PyRadiomics (version 2.2.0, pyradiomics.readthedocs.io [13]). The MTD was not measured in a standard plane, but automatically extracted by the software from the 3D-GTVs as the maximum diameter within the tumour volume. Readers were blinded for each other's delineations.

DW image quality

The quality of each DWI exam was graded using a five-point scale (by the senior radiologist, the most experienced DWI reader), taking into account both geometrical distortions and pile-up artefacts occurring specifically in the anorectal region:

1 = no artefacts, optimal image quality

2 = minimal distortions and/or pile-up artefacts (no impact on GTV delineation)

3 = moderate distortions and/or pile-up artefacts (minor impact on GTV delineation)

4 = severe distortions and/or pile-up artefacts (major impact on GTV delineation)

5 = non-diagnostic quality

Statistical analysis

Statistical analysis was performed with IBM SPSS statistics (version 25). The GTVs and MTDs obtained from T2W- MRI and DWI by the different readers were compared using a Wilcoxon signed-rank test (for non-normally distributed data, paired samples). Inter-reader agreement was analysed using the intraclass correlation coefficient (ICC; two-way random effects model for absolute agreement, single measurements) and by calculating the Dice Similarity Coefficient (DSC) and Hausdorff distances, which were both calculated using a built-in module in 3D slicer. P-values <0.05 were considered statistically significant.

RESULTS

The final study cohort consisted of 15 female and 10 male patients with a median age of 61 years (range: 34-77). Clinical tumour stage was cT1 in 3 patients, cT2 in 14, cT3 in 3 and cT4 in 5 patients.

Table 1 summarizes the mean GTV and MTD measurements for T2W-MRI and DWI for the four readers. An example illustrating the GTV delineations on T2W-MRI and DWI by all different readers is shown in **Figure 1**. Overall, the mean GTVs and MTDs measured on T2W-MRI by the senior radiologist and senior

radiation oncologist were smaller than those measured by the two junior readers on T2W-MRI (P< 0.01 – 0.02). Mean GTVs ranged between 15.1-20.8 cm^3 for the senior readers versus 26.8-27.7 cm^3 for the junior readers. On DWI, a similar trend was observed towards smaller measurements for the senior readers, but the differences were only significant for the two radiologist-readers (both P<0.01) and not for the radiation oncologists (P=0.09 – 0.89). Mean GTVs on DWI ranged between 13.8-22.5 cm^3 for the senior versus 23.6-25.8 cm^3 for the junior readers. There was a trend towards smaller GTV measurements on DWI compared to T2W-MRI for the radiologists (P=0.03 and P=0.07; example shown in **Figure 2**), but not for the radiation oncologists (P=0.53 and P=0.82). For the MTD measurements, no significant differences were observed between T2W-MRI and DWI for any of the readers.

Table 2 summarises the agreement for the GTV delineations between the four readers. Although there were some variations in agreement between individual readers, overall inter-reader agreement was good to excellent with ICCs ranging between 0.72-0.94 for T2W-MRI and 0.68-0.89 for DWI, mean DSCs ranging between 0.63-0.70 for T2W-MRI and 0.67-0.70 for DWI (with the exception of moderate agreement on DWI between the senior radiologist and senior radiation oncologist; DSC 0.58) and mean Hausdorff distances ranging between 13.63-16.68 mm for T2W-MRI and between 14.57-19.94 mm for DWI. Use of DWI did not improve the ICC, DSC or Hausdorff distances for any of the inter-reader comparisons.

Figure 3 shows the interobserver agreement (expressed as DSCs) between the readers stratified for T stage. Overall, there was a subtle trend towards higher agreement with increasing T stage (i.e. in larger tumours), with the exception of T4 tumours, which tended to result in lower DSCs than T3 tumours. An example of a T4 tumour with substantial disagreement between readers is illustrated in **Fig 3E-F.**

Effects of DWI artefacts on inter-reader agreement are illustrated in **Figure 4.** In total, 14/25 (56%) of patients had a DWI image quality score of 1-2 (no/minor artefacts with no impact on GTV delineation). Severe artefacts (score 4; major impact on GTV delineation) were encountered in three patients (12%), the remaining eight patients had moderate artefacts (score 3; minor impact on GTV delineation). In total, moderate-severe artefacts were thus encountered in 11/25 patients (44%). Overall, there was a subtle trend towards lower agreement with higher artefact scores. **Fig 4E-F** shows an example in which artefacts negatively affected interobserver agreement.

DISCUSSION

The superior soft tissue resolution of MRI makes it the most accurate imaging modality for radiotherapy planning in anal cancer as it is the only technique that allows detailed visualisation of the tumour versus the surrounding anatomy. Accurate GTV delineation is essential for effective dose delivery, to avoid toxicity to surrounding tissues and when considering dose escalation strategies. The results of this study show an overall good interobserver agreement for GTV delineation on MRI between radiologists and radiation oncologists, as well as between readers of different experience levels. Moderate to severe DWI artefacts occurred in 44% of the patients and negatively impacted interobserver reproducibility. Although we observed a trend towards smaller GTV measurements for DWI compared to T2W-MRI, this difference was not statistically significant for the majority of readers and did not result in better interobserver agreement.

This overall limited effect of DWI is consistent with the findings of Regini et al. who compared T2W-MRI and DWI for GTV delineation in rectal cancer [12], as well as a diagnostic study on T2W-MRI versus DWI tumour volumetry in rectal cancer by Ha et al. [14], who both found no significant differences between T2W-MRI and DWI. When focussing specifically on the radiologist readers, we did see an effect with significantly smaller GTVs on DWI for the senior radiologist and a similar – albeit non-significant – trend for the junior radiologist. The latter is in line with previous studies on GTV delineation in both anal and rectal cancer that included mainly radiologist-readers showing smaller GTVs on DWI [8, 9]. These previous reports attributed the reduction in GTV on DWI to an improved tumour conspicuity. In the study by Prezzi et al, this statement was substantiated by a higher delineation confidence score for DWI vs. T2W-MRI, especially for a non-expert reader [8]. The more substantial effect of DWI for the radiologist-readers may be explained by the fact that these readers are more accustomed to reading DWI, while the radiation oncologists may gravitate towards the anatomical information from T2W-MRI when delineating DWI in conjunction with T2W-MRI.

In terms of interobserver agreement, we observed no benefit from the use of DWI. This is in line with three previous reports in rectal cancer that all found similar interobserver agreement for T2W-MRI and DWI [9, 12, 15]. The abovementioned study by Prezzi et al. in anal cancer did report an improved interobserver agreement for DWI, although only minor (ICC 0.90 for T2W vs. 0.92 for DWI) [8]. Interestingly, in our current study agreement between the senior radiologist and radiation oncologist even decreased with DWI. This could again be explained by the fact that radiation oncologists will be typically less familiar with DWI. Also, a lack of anatomical detail on DWI (as result of background signal suppression in the surrounding anatomy) and presence of artefacts may make interpretation of DWI more challenging for less-experienced or non-radiologist readers.

Our results suggest that artefacts on DWI also had a negative effect on interobserver agreement. This factor has so far been poorly documented in literature. The study by Regini et al. reported that one patient (out of 27) had to be excluded because of significant artefacts on DWI, but no specifics on DWI image quality were provided for the remaining included patients [12]. A study by Burbach et al. in rectal cancer used a turbo spin echo type of DWI sequence (DWI-SPLICE), which was specifically adapted to prevent geometrical distortions [15], acknowledging this as a relevant limitation of DWI for radiotherapy planning. Another study in anal cancer by Rusten et al. corrected for geometrical distortions on DWI using anatomical data from CT, although the authors did not specify the exact method and proportion of cases that required such corrections [16]. In our study, all DWI exams were performed using an echo planar imaging (EPI) DWI sequence. EPI-DWI allows fast image acquisition and is the most commonly used type of body DWI sequence. However, this technique is prone to artefacts caused by variations in magnetic susceptibility. The anorectal region is sensitive to these artefacts because of the large difference in magnetic susceptibility between any gas in the rectal lumen and adjacent soft tissues (especially in proximal anal tumours close to the gas-filled rectum) or at the border of the scan in distal tumours at the transition between the perineum and air outside the body. A previous report by van Griethuysen et al. in rectal cancer showed that in up to 24% of rectal EPI-DWI examinations, susceptibility-related artefacts were so severe that they hampered diagnostic interpretation of the images [11]. In the current report, the percentage of severe artefacts was 12%, but the total proportion of cases with any degree of interference from DWI artefacts on GTV delineation was higher (44%). These artefacts may not only negatively affect tumour delineation, but may also cause problems with co-registration of MR images with simulation CT scans. DWI image quality is a crucial factor to take into account when considering the use of DWI for radiotherapy planning. For the time being, T2W-MRI thus appears to be a more robust method for tumour delineation, with respect to image quality as well as reproducibility between readers of varying expertise. When using T2W-MRI, our results indicate that radiation oncologists perform anal tumour delineations with similar results as radiologists, suggesting that expert-radiologist input would generally not be required when performing MRI-based GTV delineation for radiotherapy planning.

Exceptions to this may be the smaller or more complex tumours. Although subgroups in our cohort were small and there were some variations between findings on T2W-MRI and DWI as well as between readers, we observed an overall trend towards better interobserver agreement with increasing T-stage (i.e., with increasing tumour size), with on average best DSC's in T3 tumours. This is consistent with a previous report by Burbach et al. who also found improved agreement for larger tumour volumes in rectal cancer [15]. This finding may not be surprising, as larger tumours will typically be easier to define and the DSC as a measure of interobserver agreement will generally be higher for larger volumes.

Interestingly, agreement for the T4 tumours in our cohort tended to be slightly lower than for the T3 tumours, although the numbers in the T-stage categories were obviously too small to draw firm conclusions. This observation is explainable though, since T4 stage in anal cancer is determined by invasion into adjacent organs and not by tumour size (as in T1-3 stage). Hence, T4 tumours are not necessarily larger in size than T3 tumours [17], while they are more often irregularly shaped and relatively ill-defined. This could make them more difficult to delineate, especially for less-trained readers (as is also illustrated in **Figure 3**). Based on these observations, it may thus be in particular for small and more complex (T4) tumours where guidance from expert radiologists may be of benefit for GTV delineation. Along this line, we believe that training and education of radiation oncologists (by radiologists) on how to interpret MRI is essential and should form an integrated part of the training programme. Overall, there is currently a lack of guidelines detailing how MRI-guidance should be implemented into daily workflows. The radiation oncology community should therefore make an effort to generate such guidelines, which should include detailed recommendations on MRI protocols, contouring atlases based on MRI anatomy as well as recommendations on education and training, and integration of digital rectal examination combined with good quality imaging as necessities for robust target identification.

Our study had some limitations in addition to the small study cohort and its retrospective nature. First, a common limitation in delineation studies such as the current study is the lack of a gold standard to determine the actual tumour boundaries. Second, due to the retrospective approach, the DWI protocols were not optimized for the purpose of this study and no specific measures were undertaken (at the time of acquisition and patient preparation) to avoid MRI artefacts. Also, some slight variations in acquisition protocol occurred over time, although we believe that overall image quality was similar during the inclusion period and will likely not have had a significant effect on the primary outcomes of this study. As a follow-up study it would be interesting to validate our findings prospectively with standardized and optimized DWI protocols. Also, it would be interesting to explore how the performance to discriminate tumour (and thus delineation reproducibility) is affected during and after treatment, which would be a potentially relevant issue when considering sequential boosting strategies (using MRL systems) in the future.

In conclusion, this study shows overall good interobserver agreement for GTV delineation on MRI in anal squamous cell carcinoma between radiologists and radiation oncologists, as well as between readers of different experience levels. As an important caveat, we have shown that – if no steps are undertaken to avoid these artefacts – susceptibility-related artefacts on DWI can have a considerable negative impact on interobserver reproducibility, thereby limiting the potential use of DWI for radiotherapy planning.

REFERENCES

[1] Dirix P, Haustermans K, Vandecaveye V. The value of magnetic resonance imaging for radiotherapy planning. Semin Radiat Oncol. 2014;24:151-9. https://doi.org/10.1016/j.semradonc.2014.02.003
[2] Chung NN, Ting LL, Hsu WC, Lui LT, Wang PM. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head Neck. 2004;26:241-6. https://doi.org/ 10.1002/hed.10378
[3] Schmid MP, Fokdal L, Westerveld H, Chargari C, Rohl L, Morice P, et al. Recommendations from gynaecological (GYN) GEC-ESTRO working group - ACROP: Target concept for image guided adaptive brachytherapy in primary vaginal cancer. Radiother Oncol. 2019;145:36-44. https://doi.org/10.1016/j.radonc.2019.11.005
[4] Salembier C, Villeirs G, De Bari B, Hoskin P, Pieters BR, Van Vulpen M, et al. ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer. Radiother Oncol. 2018;127:49-61. https://doi.org/10.1016/j.radonc.2018.01.014
[5] Owrangi AM, Greer PB, Glide-Hurst CK. MRI-only treatment planning: benefits and challenges. Phys Med Biol. 2018;63:05TR1. https://doi.org/ 10.1088/1361-6560/aaaca4
[6] Corradini S, Alongi F, Andratschke N, Belka C, Boldrini L, Cellini F, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. Radiat Oncol. 2019;14:92. https://doi.org/ 10.1186/s13014-019-1308-y
[7] Winkel D, Bol GH, Werensteijn-Honingh AM, Intven MPW, Eppinga WSC, Hes J, et al. Target coverage and dose criteria based evaluation of the first clinical 1.5T MR-linac SBRT treatments of lymph node oligometastases compared with conventional CBCT-linac treatment. Radiother Oncol. 2020;146:118-25. https://doi.org/10.1016/j.radonc.2020.02.011
[8] Prezzi D, Mandegaran R, Gourtsoyianni S, Owczarczyk K, Gaya A, Glynne-Jones R, et al. The impact of MRI sequence on tumour staging and gross tumour volume delineation in squamous cell carcinoma of the anal canal. Eur Radiol. 2018;28:1512-9. https://doi.org/10.1007/s00330-017-5133-0
[9] 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. https://doi.org/10.1016/j.canrad.2018.10.004
[10] Kochhar R, Plumb AA, Carrington BM, Saunders M. Imaging of anal carcinoma. AJR Am J Roentgenol. 2012;199:W335-44. https://doi.org/ 10.2214/AJR.11.8027
[11] van Griethuysen J, Bus E, Hauptmann M, Lahaye M, Maas M, ter Beek L, et al. Gas-induced susceptibility artefacts on diffusion-weighted MRI of the rectum at 1.5T - effect of applying a micro-enema to improve image quality. Eur J Radiol. 2018;99:131-7. https://doi.org/10.1016/j.ejrad.2017.12.020
[12] Regini F, Gourtsoyianni 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. https://doi.org/10.1016/j.ejrad.2014.02.007
[13] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77:e104-e7. https://doi.org/ 10.1158/0008-5472.CAN-17-0339
[14] Ha HI, Kim AY, Yu CS, Park SH, Ha HK. Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after

preoperative chemoradiation therapy. Eur Radiol. 2013;23:3345-53. https://doi.org/10.1007/s00330-013-2936-5
[15] Burbach JP, Kleijnen JP, Reerink O, Seravalli E, Philippens ME, 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. https://doi.org/10.1016/j.radonc.2015.10.030
[16] Rusten E, Rekstad BL, Undseth C, Al-Haidari G, Hanekamp B, Hernes E, et al. Target volume delineation of anal cancer based on magnetic resonance imaging or positron emission tomography. Radiat Oncol. 2017;12:147. https://doi.org/10.1186/s13014-017-0883-z
[17] Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. 2017;67:93-9. https://doi.org/10.3322/caac.21388

TABLES

Table 1. GTV and MTD measurements using T2-weighted MRI and diffusion-weighted imaging (DWI)

Gross tumour volume (GTV) in cm^3							
Reader	**T2W (mean ± sd)**	**DWI (mean ± sd)**	**P-value (T2W vs DWI)**	**Reader**	**T2W (mean ± sd)**	**DWI (mean ± sd)**	**P-value (T2W vs DWI)**
Senior Radiologist	15.1 ± 16.5	13.8 ± 15.6	0.03	Senior Radiation Oncologist	20.8 ± 20.9	22.5 ± 17.4	0.53
Junior Radiologist	26.8 ± 30.8	23.6 ± 30.4	0.07	Junior Radiation Oncologist	27.7 ± 32.9	25.8 ± 23.4	0.82
P-value (senior vs junior)	<0.001	<0.001		**P-value (senior vs junior)**	0.02	0.09	
Maximum tumour diameter (MTD) in cm							
Reader	**T2W (mean ± sd)**	**DWI (mean ± sd)**	**P-value (T2W vs DWI)**	**Reader**	**T2W (mean ± sd)**	**DWI (mean ± sd)**	**P-value (T2W vs DWI)**
Senior Radiologist	4.7 ± 1.9	4.7 ± 2.0	0.21	Senior Radiation Oncologist	5.2 ± 1.8	6.0 ± 2.2	0.07
Junior Radiologist	5.8 ± 2.1	5.8 ± 2.6	0.90	Junior Radiation Oncologist	5.7 ± 2.0	5.6 ± 1.9	0.82
P-value (senior vs junior)	<0.001	<0.001		**P-value (senior vs junior)**	0.02	0.88	

Table 2. Interobserver agreement for GTV delineations

	Intraclass correlation Coefficient (ICC)		**Mean Dice Similarity Coefficient (DSC)**		**Mean Hausdorff Distance (HD) in mm**	
	T2W (95% CI)	**DWI (95% CI)**	**T2W (±SD)**	**DWI (±SD)**	**T2W (±SD)**	**DWI (±SD)**
senior vs. junior readers						
senior radiologists vs. junior radiologist	0.72 (0.29-0.88)	0.68 (0.34-0.84)	0.63 (±0.15)	0.68 (±0.19)	16.68 (±7.38)	15.58 (±10.59)
senior radiation oncologist vs. junior radiation oncologist	0.85 (0.65-0.93)	0.89 (0.76-0.95)	0.68 (±0.19)	0.70 (±0.21)	16.12 (±10.24)	17.32 (±15.93)
radiologists vs. radiation oncologists						
senior radiologist vs. senior radiation oncologist	0.90 (0.55-0.96)	0.72 (0.21-0.89)	0.65 (±0.24)	0.58 (±0.25)	13.62 (±9.49)	19.94 (±18.67)
junior radiologist vs. junior radiation oncologist	0.94 (0.86-0.97)	0.78 (0.57-0.90)	0.70 (±0.18)	0.67 (±0.20)	15.11 (±7.88)	14.57 (±9.42)

Supplementary Table 1. MRI protocol

	T2-weighted fast spin echo	Single-shot EPI DWI
Repetition time (ms)	4250 – 6412	4365- 6380
Echo time (ms)	120	53,3- 66.0
Slice thickness (mm)	3	3-5
Slice gap (mm)	3-3.3	3.3-5.3
In plane resolution (mm x mm)	0.39x0.39 – 0.49x0.49	1.07x1.07 – 1.11x1.11
Echo train length	30	-
Echo planar imaging (EPI) factor	-	53-81
No. signal averages	1-2	1
b-values (s/mm^2)	-	0, 200, 800[a]

[a] Only the b800 images were used for tumour delineation on DWI.

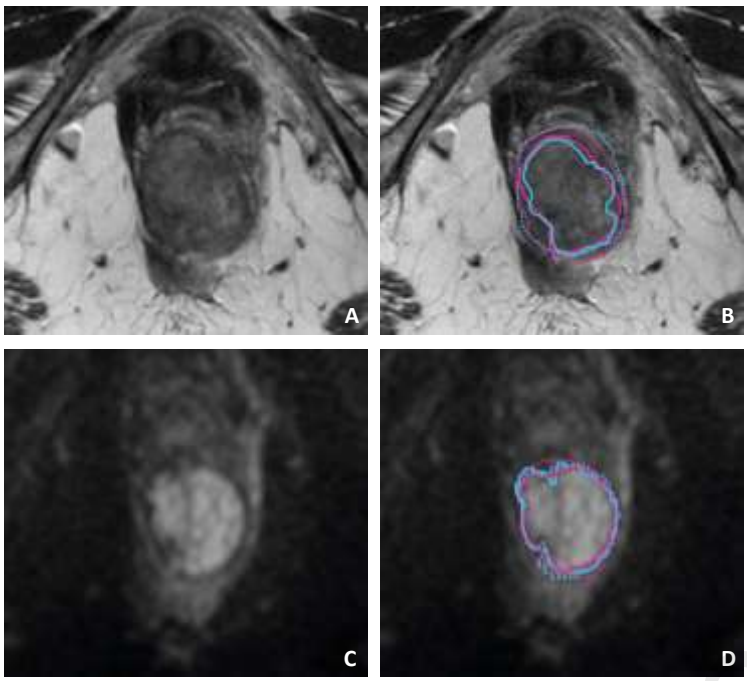
FIGURE CAPTIONS

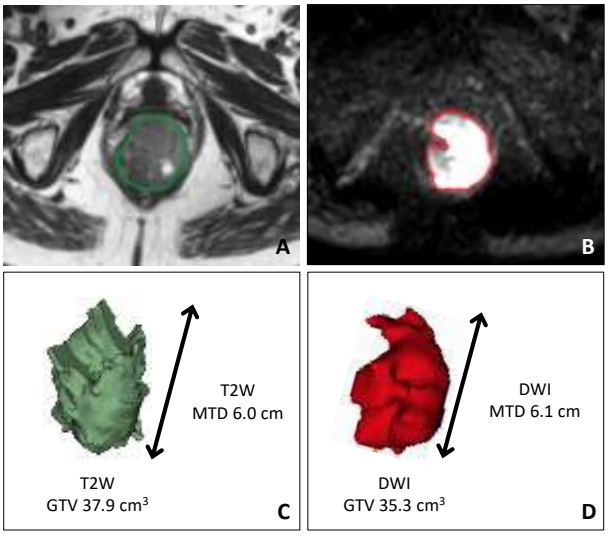
Figure 1. *Example of T2W and DWI images of a female patient with a bulky tumour in the anal canal presented without (A,C) and with (B, D) the GTV delineations performed by the four readers: Senior Radiologist (blue line), Junior Radiologist (blue dotted line), Senior Radiation oncologist (pink line) and Junior Radiation oncologist (pink dotted line). In this example, delineations by the senior readers (continuous lines) were smaller than those by the junior readers (dotted lines), in particular on T2W-MRI (B). Delineations on DWI (D) were smaller than on T2W-MRI, especially for the two junior readers and spatial overlap between the four readers was higher in this particular case.*

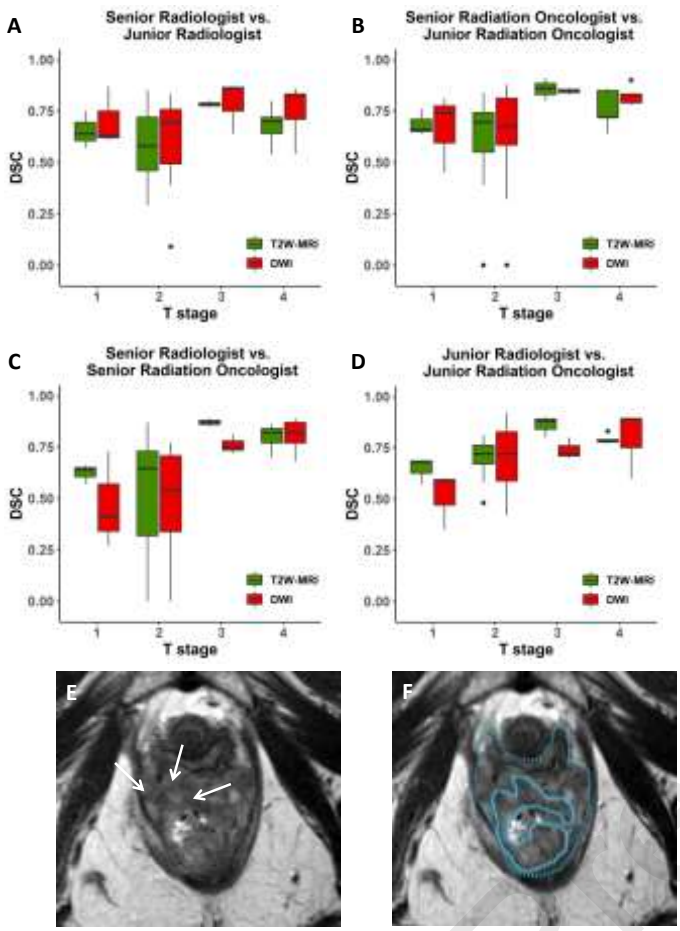
Figure 2. *Example of tumour delineation performed on T2W-MRI (A) and b800 DWI (B) by the Senior Radiologist in a female patient. The GTV and MTD results for T2W-MRI (C) and DWI (D) show that the GTV was considerably smaller on DWI, but that the corresponding MTD was similar for both techniques.*

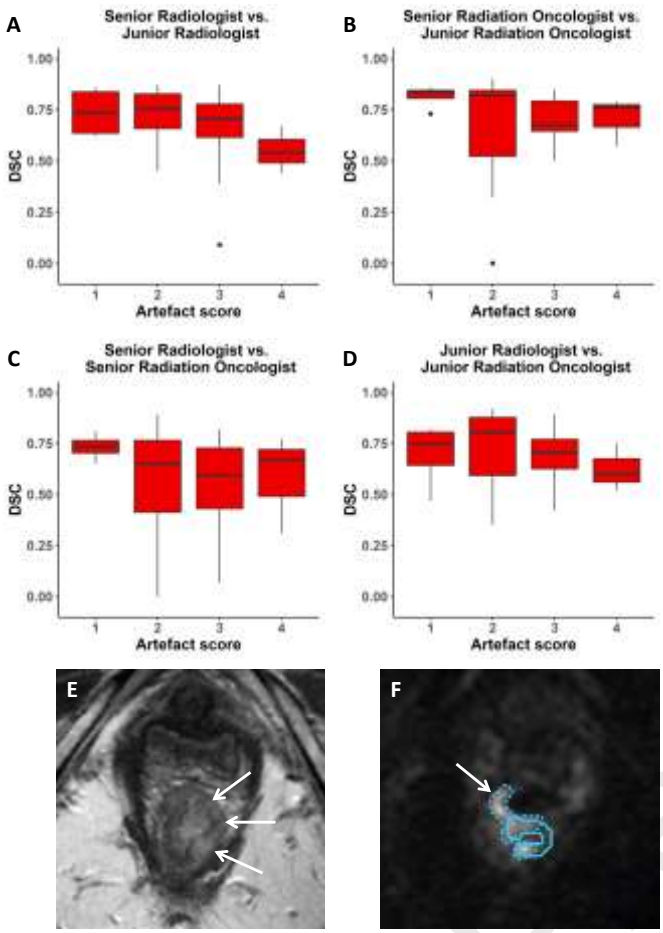
Figure 3. *Boxplots showing the distribution of DSCs for the GTV delineations between the two radiologists (A), the two radiation oncologists (B), the two senior readers (C) and the two junior readers (D) depending on the T-stage of the tumour. DSCs tended to improve with increasing T-stage, except for T4 tumours, which showed slightly lower DSCs. The scan images (E, F) shown an example of a complex, irregular shaped and ill-defined T4 tumour in a female patient. The tumour invades the posterior vaginal wall (arrows in E). Delineations of the senior radiologist (blue line in F) and junior radiologist (blue dotted line in F) are shown, illustrating that there was substantial disagreement between the two readers.*

Figure 4. *Boxplots showing the interobserver agreement (expressed as DSCs) on diffusion-weighted imaging between the two radiologists (A), the two radiation oncologists (B), the two senior readers (C) and the two junior readers (D) for each of the DWI artefact scores (with a higher score indicating more significant artefacts). The T2-weighted (E) and DWI (F) images show an example of a female patient with a tumour on the left side of the anal canal (arrows in E) whose DWI scan was assigned an artefact score of 4 due to severe geometrical distortions and a signal pile-up artefact on the right anterior side (arrow in F). These effects hampered visualization of the tumour which led to substantial disagreement in delineations between readers (blue line: senior radiologist; blue dotted line: junior radiologist).*









HIGHLIGHTS

- MRI is increasingly used for gross tumour volume (GTV) delineation in anal cancer
- GTV delineation is reproducible between radiologists and radiation oncologists
- Diffusion-weighted MRI (DWI) renders similar results as anatomical (T2-weighted) MRI
- Artefacts are common on DWI and may limit its use for radiotherapy planning

CONFLICT OF INTEREST STATEMENT

The authors declare that this work has not been published previously, and that it is not under consideration for publication elsewhere. All authors declare no conflict of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.