

# Diffusion-weighted MRI to assess response to chemoradiotherapy in rectal cancer: main interpretation pitfalls and their use for teaching

Doenja M. J. Lambregts<sup>1</sup> · Miriam M. van Heeswijk<sup>1,2,3,4</sup> · Andrea Delli Pizzi<sup>5</sup> · Saskia G. C. van Elderen<sup>6</sup> · Luisa Andrade<sup>7</sup> · Nicky H. G. M. Peters<sup>8</sup> · Peter A. M. Kint<sup>9</sup> · Margreet Osinga-de Jong<sup>8</sup> · Shandra Bipat<sup>10</sup> · Rik Ooms<sup>11</sup> · Max J. Lahaye<sup>1</sup> · Monique Maas<sup>1</sup> · Geerard L. Beets<sup>12,4</sup> · Frans C. H. Bakers<sup>2</sup> · Regina G. H. Beets-Tan<sup>1,4</sup>

Received: 17 November 2016 / Revised: 3 February 2017 / Accepted: 20 March 2017 / Published online: 13 April 2017  
© European Society of Radiology 2017

## Abstract

**Objectives** To establish the most common image interpretation pitfalls for non-expert readers using diffusion-weighted imaging (DWI) to assess response to chemoradiotherapy in patients with rectal cancer and to explore the use of these pitfalls in an expert teaching setting.

**Methods** Two independent non-expert readers (R1 and R2) scored the restaging DW MRI scans (*b*1,000 DWI, in conjunction with ADC maps and T2-W MRI scans for anatomical reference) in 100 patients for the likelihood of a complete response versus residual tumour using a five-point confidence score. The readers received expert feedback and the final response outcome for each case. The supervising expert documented any potential interpretation errors/pitfalls discussed for each case to identify the most common pitfalls.

**Results** The most common pitfalls were the interpretation of low signal on the ADC map, small susceptibility artefacts, T2 shine-through effects, suboptimal sequence angulation and collapsed rectal wall. Diagnostic performance (area under the ROC curve) was 0.78 (R1) and 0.77 (R2) in the first 50 patients and 0.85 (R1) and 0.85 (R2) in the final 50 patients.

**Conclusions** Five main image interpretation pitfalls were identified and used for teaching and feedback. Both readers achieved a good diagnostic performance with an AUC of 0.85.

## Key Points

- Fibrosis appears hypointense on an ADC map and should not be mistaken for tumour.
- Susceptibility artefacts on rectal DWI are an important potential pitfall.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00330-017-4830-z) contains supplementary material, which is available to authorized users.

✉ Doenja M. J. Lambregts  
d.lambregts@nki.nl

<sup>1</sup> Department of Radiology, The Netherlands Cancer Institute, PO Box 90203, 1006 BE Amsterdam, The Netherlands

<sup>2</sup> Department of Radiology, Maastricht University Medical Centre, Maastricht, The Netherlands

<sup>3</sup> Department of Surgery, Maastricht University Medical Centre, Maastricht, The Netherlands

<sup>4</sup> GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands

<sup>5</sup> Department of Neuroscience and Imaging, Gabriele d'Annunzio University, SS. Annunziata Hospital, Chieti, Italy

<sup>6</sup> Department of Radiology, Leiden University Medical Centre, Leiden, The Netherlands

<sup>7</sup> Department of Radiology, Hospitais Da Universidade De Coimbra, Coimbra, Portugal

<sup>8</sup> Zuyderland Medical Center, location Heerlen, Heerlen, The Netherlands

<sup>9</sup> Department of Radiology, Amphibia Hospital, Breda, The Netherlands

<sup>10</sup> Department of Radiology, Academic Medical Centre, Amsterdam, The Netherlands

<sup>11</sup> Department of Radiology, Maxima Medical Centre, Eindhoven-Veldhoven, The Netherlands

<sup>12</sup> Department of Surgery, The Netherlands Cancer Institute, Amsterdam, The Netherlands

- *T2 shine-through on rectal DWI is an important potential pitfall.*
- *These pitfalls are useful to teach non-experts how to interpret rectal DWI.*

**Keywords** Rectal neoplasms · Chemoradiotherapy · Magnetic resonance imaging · Diffusion magnetic resonance imaging · Diagnosis

## Introduction

In the last decade over 60 original studies have been published on the use of diffusion-weighted imaging (DWI) for rectal cancer assessment. The majority of these studies focused on the use of DWI for evaluation of response to chemoradiotherapy (CRT). This specific focus on tumour response evaluation can probably largely be attributed to recent developments in the treatment of rectal cancer. Studies have shown that patients who respond very well to a long course of CRT may be treated with organ-preserving treatments (local excision of the tumour remnant or watchful waiting) instead of surgical resection, making accurate response evaluation after CRT an increasingly important issue [1–3]. In this setting, imaging – in particular MRI – plays an important role. Although morphological MRI is beneficial for assessing tumour downsizing and downstaging, its use is associated with difficulties in determining the presence and extent of residual tumour within areas of postradiation fibrosis.

A recent meta-analysis has shown that the sensitivity for overall tumour restaging after CRT with MRI is only 50%, with even poorer results (sensitivity 19%) for selecting complete responders [4]. The overall sensitivity for restaging was considerably better (84%) in a subgroup analysis focusing on studies that used DWI. Various studies have also shown that the addition of DWI significantly improves the performance of MRI in differentiating viable tumour within areas of post-radiation fibrosis [5–7]. Moreover, a recent study has shown that out of a variety of MRI features (e.g. tumour location, signal intensity, T stage and N stage, tumour volume and volume reduction ratios), visual assessment of response on DWI was one of the best predictors of a complete tumour response on MRI [8].

In the majority of the published studies, DW images were read by expert radiologists typically with dedicated experience (2–13 years) in reading rectal MR images and previous experience in reading DW images [5–7, 9]. The performance of such readers may not necessarily reflect that of general radiologists. It is well known that radiological readers with more experience will have better diagnostic performance. Moreover, in different imaging settings (for example, reading of mammograms, CT colonography examinations and MRI for diagnosing endometriosis) it has been demonstrated that a learning curve is required before non-expert readers can

reach a certain diagnostic performance level [10–12]. Teaching is therefore an important issue.

It would be helpful to gain knowledge on the pitfalls and interpretation errors most commonly encountered by non-expert readers when assessing DWI after CRT. This information could then serve as a teaching reference for future readers who wish to improve their DWI reading skills. The aim of this study was therefore to establish the most common image interpretation errors and pitfalls encountered by non-expert readers when assessing DWI to discriminate between a complete tumour response and residual tumour after CRT for rectal cancer and to explore the use of these pitfalls in a teaching setting.

## Materials and methods

This study was a retrospective analysis of MR images acquired as part of routine diagnostic procedures. The study was approved by the local institutional review board and informed consent was waived.

## Patients

From a retrospective imaging dataset, 105 consecutive patients with rectal cancer were selected who were receiving a long course of neoadjuvant treatment and who had undergone a restaging MRI examination including a DWI sequence at Maastricht University Medical Centre between November 2011 and November 2014. Five patients were used as training cases and the other 100 patients constituted the study (test) dataset. Inclusion criteria were: (a) histopathologically proven rectal adenocarcinoma, (b) neoadjuvant treatment consisting of a long course of CRT (or radiotherapy with a prolonged waiting period), (c) availability of a good quality restaging MRI scan including a DWI sequence, and (d) data on final response outcome. Patients with a low-quality DWI examination (e.g. severe susceptibility artefacts due to metal implants) as well as patients with a mucinous tumour (as these are known to exhibit different signal characteristics on DWI) were excluded. The routine neoadjuvant treatment consisted of 50.4 Gy radiation + capecitabine 825 mg/m<sup>2</sup> twice daily during the radiation period. According to current guidelines, patients with locally advanced disease (cT3/4 stage, involved mesorectal fascia on MRI, and/or clinical node-positive disease) were routinely stratified for a long course of CRT.

## MR imaging

All MR examinations were performed at 1.5 T on an Intera Achieva or an Ingenia MR system (Philips Medical Systems, Best, The Netherlands) using a phased array surface coil with

patients in feet-first supine position. The routine interval between completion of CRT and restaging MRI examinations was 6–10 weeks. To reduce bowel motility, patients received 20 mg of scopolamine butylbromide (Buscopan; Boehringer Ingelheim, Ingelheim am Rhein, Germany) intravenously, either because of anticipated bowel movement artefacts on the sagittal planning scan (first part of the study period) or routinely (final part of the study period). From March 2014 patients also routinely received a microenema (Microlax®; McNeil Healthcare Ireland Ltd, Dublin, Ireland) about 15 minutes before the start of the examination to reduce the amount of air in the rectal lumen. The standard clinical MRI protocol at the institution (both for primary staging and restaging) included 2D T2-weighted fast spin echo sequences in the sagittal, axial and coronal planes and an axial echo planar imaging DWI sequence with  $b1,000$  being the highest  $b$ -value. The axial T2-weighted and DWI sequences were angled in an identical plane perpendicular to the rectal tumour axis. Apparent diffusion coefficient (ADC) maps were automatically generated by the operating system. Detailed sequence parameters are given in Appendix 1.

### Training

All images were independently read by two senior (5th year) radiology residents (S.G.C.v.E. and A.D.P.) with an interest in abdominal imaging but with no specific previous experience in reading DWI of rectal cancer. Before the start of the study both readers received a short (1–2 h) baseline training from an expert radiologist (D.M.J.L., with 8 years of specific expertise in reading rectal MRI and DWI) in how to read DWI and MRI scans of rectal cancer. Training consisted of discussion of various imaging examples and cases derived from

PowerPoint presentations and previous literature as well as a hands-on training session with the first five training cases.

### Scoring and feedback

The two readers were asked to independently assess the remaining 100 cases (the test dataset) and for each case to report the likelihood of a complete tumour response using a five-point confidence level score (Table 1; similar to scores described in the literature [13]). The readers based their score on the high  $b$ -value ( $b1,000$ ) DW images that were read in conjunction with the corresponding ADC maps and T2-weighted images (for anatomical reference). The ADC maps were mainly used to discriminate T2 shine-through effects from restricted diffusion in the presence of a high signal on  $b1,000$  DW images. The primary staging MR images (including DW images and ADC map) were also at the readers' disposal. The readers were blinded to each other's results as well as to the final patient outcome. In the first 30 cases, the two readers received immediate expert feedback (as well as the final response outcome) after the scoring of each single case. In the last 70 cases the readers received feedback after every ten cases.

### Documentation of interpretation pitfalls

For each single case the supervising expert reader documented any doubts, potential reasons for error and interpretation pitfalls encountered by the two non-expert readers discussed during the feedback sessions. Pitfalls were discussed separately with each reader and were recorded when discussed with either of the two readers regardless of whether or not they eventually resulted in false-positive or false-negative findings.

**Table 1** Confidence level scores to discriminate between tumour and complete response after CRT

Score	Likelihood of complete tumour response	Imaging report
0	Definitely complete response	Completely normalized rectal wall without any focal high signal on high $b$ -value DWI, <i>or</i> well-defined fibrotic wall thickening at the previous tumour site on T2-weighted MRI with no focal high signal on high $b$ -value DWI
1	Probably complete response	Fibrotic wall thickening with no or only minimal focal high signal on DWI, probably not related to the location of the former tumour bed
2	Possibly complete response/possibly residual tumour (equivocal)	Indeterminate/ambiguous focal high signal on DWI; questionable whether it corresponds to the location of the former tumour bed
3	Probably residual tumour	Focal high DWI signal corresponding to the location of the former tumour bed (i.e. within the fibrosis)
4	Definitely residual tumour	Marked mass-like high signal on DWI (corresponding to suspicious residual mass on T2-weighted MRI)

*Note:* In case of any high signal on DWI, the readers were instructed to refer to the ADC map to confirm that the high signal was caused by restricted diffusion (with corresponding low signal on ADC) and not by T2 shine-through effects (with corresponding high signal on ADC) or artefacts

## Reference standard

In 62 patients the final response outcome was based on the final tumour stage on histopathology after surgical resection (ypT stage). The remaining 38 patients had a complete clinical response and were intensively followed according to a watchful waiting strategy (including follow-up with endoscopy and imaging 3-monthly in the first year and 6-monthly in the 2nd to 5th years). In these patients a sustained complete clinical response (with repeated negative MRI examinations and endoscopy with or without biopsy) was considered a surrogate endpoint for a complete response (yT0), similar to methods used in previous studies [5, 13–15]. The follow-up in these patients was  $37 \pm 11$  months (mean  $\pm$  SD). The responses in the whole patient group were dichotomized as residual tumour (ypT1–4) or complete response (ypT0 after surgery or ycT0 with a sustained complete response during watchful waiting).

## Statistical analysis

The results were analysed using SPSS, version 22 (IBM Corp., Armonk, NY). Receiver operator characteristic curves were constructed to analyse diagnostic performance in assessing the presence of residual tumour, and areas under the curve with 95% confidence intervals were calculated. Two-way contingency tables were constructed to calculate diagnostic parameters (sensitivity, specificity, positive and negative predictive values, overall accuracy). For these calculations the confidence level scores were dichotomized as confidence level 0/1 or 2–4. Results were calculated separately for the first 50 and final 50 study patients. Interobserver agreement was analysed using kappa statistics with quadratic kappa weighting.

## Results

### Patient characteristics

Of the 100 test patients, 69 were men and 31 women, and their median age was 64 years (range 31–82 years). In total, 46 patients were complete responders (8 ypT0, and 38 ycT0 of those undergoing watchful waiting), and 54 patients had residual tumour (7 ypT1, 16 ypT2, 25 ypT3, 6 ypT4). Further patient characteristics are given in Table 2.

### Diagnostic performance and interobserver agreement

Table 3 shows the results for the two readers for the first 50 and final 50 study cases. The AUCs for the first 50 patients were 0.78 for reader 1 and 0.77 for reader 2. The AUCs for the final 50 patients were 0.85 for reader 1 and 0.85 for reader 2. The numbers of equivocal (uncertain) scores for reader 1 were 11 for the first 50 cases and 6 for the final 50 cases. The numbers of equivocal

**Table 2** Baseline characteristics of the 100 patients and their treatment

Variable	Value
Sex ( <i>n</i> )	
Male	69
Female	31
Age (years), median (range)	64 (31–82)
Primary cTN stage as assessed by MRI ( <i>n</i> )	
cT	
1/2	17
3	74
4	9
cN	
0	27
1	27
2	46
Surgical treatment ( <i>n</i> )	
Low anterior resection	42
Abdominoperianal resection microsurgery	12
Transanal endoscopic microsurgery	4
Pelvic exenteration	4
None (watchful waiting) <sup>a</sup>	38
Final yTN stage ( <i>n</i> )	
yT	
0 <sup>b</sup>	46
1	7
2	16
3	25
4	6
yN	
0 <sup>c</sup>	77
1	18
2	5
Final tumour response (yT stage)	
Complete response <sup>b</sup>	46
Residual tumour	54

<sup>a</sup> Follow-up (mean  $\pm$  SD)  $32 \pm 11$  months

<sup>b</sup> Based on histopathology after surgery in 8 patients; based on long-term follow-up in 38 patients with a sustained clinical complete response undergoing watchful waiting

<sup>c</sup> Based on histopathology after surgery in 35 patients; based on long-term follow-up without evidence of recurrence in 4 patients after transanal endoscopic microsurgery and in 38 patients with a sustained clinical complete response undergoing watchful waiting

scores for reader 2 were 4 for the first 50 cases and 0 for the final 50 cases. Interobserver agreement was moderate ( $\kappa$  0.58) for the first 50 cases and good ( $\kappa$  0.71) for the final 50 cases.

### Interpretation pitfalls

The five most common interpretation difficulties and pitfalls encountered by the non-expert readers that were discussed

**Table 3** Diagnostic accuracy figures, number of equivocal scores and interobserver agreement for the first and final 50 study cases respectively

	First 50 cases		Final 50 cases	
	Reader 1	Reader 2	Reader 1	Reader 2
AUC (95% CI)	0.78 (0.65–0.91)	0.77 (0.64–0.90)	0.85 (0.74–0.95)	0.85 (0.74–0.96)
Accuracy	70% (35/50)	72% (36/50)	80% (40/50)	82% (41/50)
Sensitivity	60% (15/25)	76% (19/25)	72% (21/29)	86% (25/29)
Specificity	80% (20/25)	68% (17/25)	90% (19/21)	76% (16/21)
Positive predictive value	75% (15/20)	70% (19/27)	91% (21/23)	83% (25/30)
Negative predictive value	67% (20/30)	74% (17/23)	70% (19/27)	80% (16/20)
No. of equivocal scores	11	4	6	0
Interobserver agreement ( $\kappa$ ) (95% CI)	0.58 (0.41–0.74)		0.71 (0.57–0.85)	

CI Confidence interval, AUC Area under the ROC curve

during the feedback sessions were as follows (and are summarized in Table 4):

1. Hypointense fibrosis on ADC map:

In patients with a complete response who showed a fibrotic residue, readers were taught not to erroneously interpret low signal on ADC as suspicious for tumour in the absence of a corresponding high signal on DWI (Fig. 1).

2. Susceptibility effects:

Readers were taught to recognize high signal caused by susceptibility effects and differentiate it from high signal caused by the presence of tumour (Fig. 2).

3. T2 shine-through:

A potential pitfall was the presence of high signal in the rectal lumen on *b*1,000 DWI caused by T2 shine-through effects of intraluminal fluid. Readers were taught to recognize these luminal shine-through effects and differentiate them from tumour-related high

signal by comparing the diffusion images with the ADC map (where fluid will result in high signal while tumour will show low signal) and by taking into account the shape of the signal (luminal shine-through effects are typically star-shaped while high signal caused by tumour is typically more nodular or tubular/U-shaped; Fig. 3).

4. Suboptimal sequence angulation:

In proximal rectal tumours, angulation perpendicular to the tumour results in coronal-like imaging planes, which were found to be more difficult to interpret. Moreover when the images before and after CRT were not angled in identical planes, the readers found it difficult to compare the images and correctly interpret the diffusion images after CRT (Fig. 4).

5. Collapsed rectal wall:

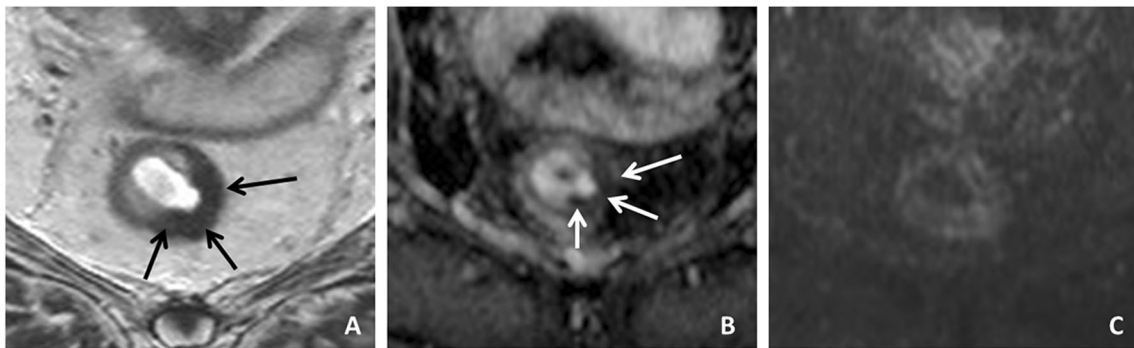
In patients with a collapsed rectal wall, the readers found it difficult to determine whether a high signal on DWI was caused by superposition of the two sides of the rectal wall or by the presence of tumour (Fig. 5).

**Table 4** Main potential causes of error (pitfalls)

Number	Potential pitfall	Illustrated in	Percentage of cases <sup>a</sup>
1	Misinterpretation of low signal on ADC map from fibrosis	Fig. 1	41
2	Susceptibility effects	Fig. 2	33
3	T2 shine-through of fluid in the rectal lumen	Fig. 3	12
4	Suboptimal sequence angulation	Fig. 4	6
5	Collapsed rectal wall	Fig. 5	2

ADC Apparent diffusion coefficient

<sup>a</sup> Cases in which a potential pitfall was present (i.e. identified by the expert reader after discussing the case with the two non-expert readers). It does not necessarily indicate that the case was in fact misinterpreted by one or both readers



**Fig. 1** A male patient treated for a tumour on the left dorsolateral side in the mid-rectum after CRT. **a** The restaging T2-weighted MRI image shows semicircular fibrotic wall thickening (*black arrows*). **b** On the ADC map the wall thickening is markedly hypointense (*white arrows*). **c** The *b*1,000 diffusion-weighted image shows no high signal. This patient showed a

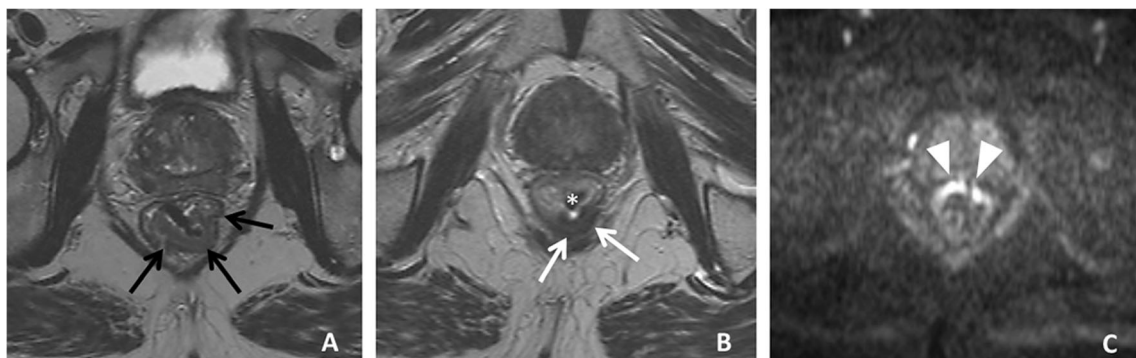
complete response. The low signal on the ADC map is caused by the fact that fibrotic tissue (containing many macromolecules) has a low T2 relaxation time and the low signal is not due to diffusion restriction, which is why there is no high signal present on the *b*1,000 diffusion image

## Discussion

The main aim of this study was to document the most common potential interpretation pitfalls encountered by non-expert readers when reading DWI for assessing response to CRT to serve as a teaching reference for future readers. An important potential pitfall encountered was the misinterpretation of low signal on the ADC map as being suspicious for residual tumour. When studying the most basic concepts of DWI, the typical instruction is to consider low signal on the ADC map as a sign of restricted diffusion. However, this is not always the case. Dense fibrosis contains a lot of extracellular matrix macromolecules (collagen), which typically have such short T2 relaxation times that at the time of image acquisition (with commonly used clinical pulse sequences) the signal will be very low or even zero. As a result fibrosis will be markedly hypointense on the ADC map due to lack of sufficient signal itself and not due to actual diffusion restriction. The same goes for several other tissues and structures such as calculi, tendons and ligaments, cortical bone and some blood products, which

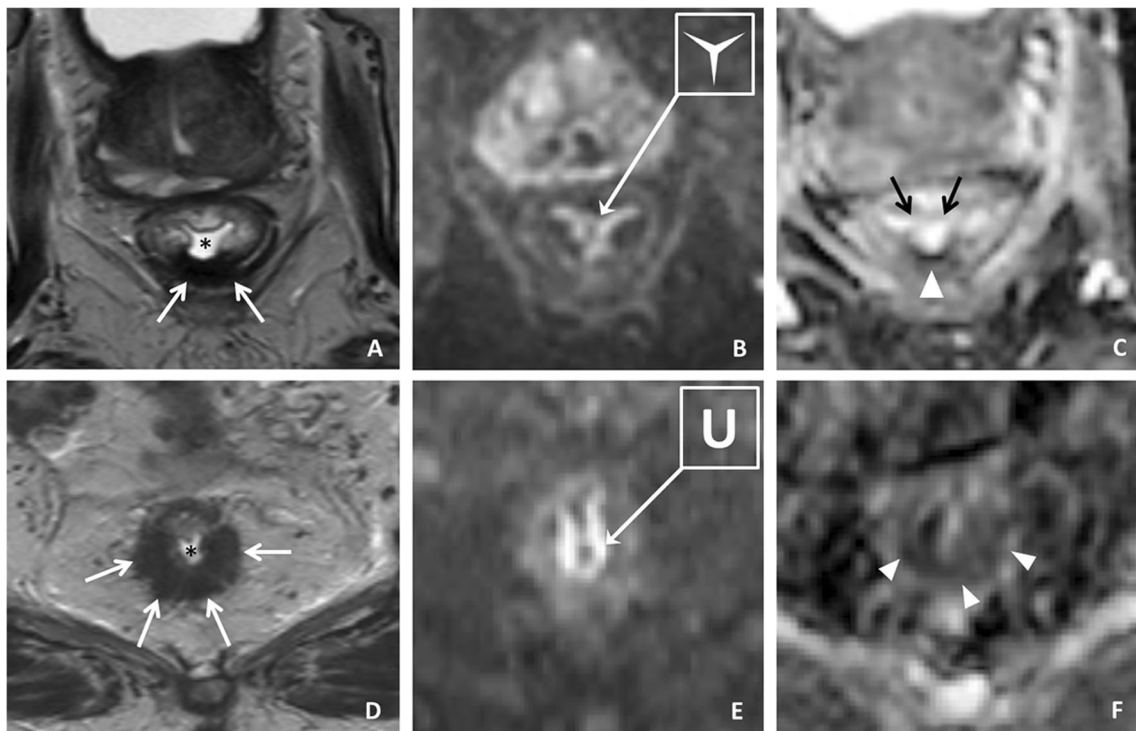
also have insufficient MRI signal and will typically be dark on all sequences, including the DWI images and the ADC map [16]. In contrast, tissues with true diffusion restriction (e.g. tumour) will show low signal on the ADC map, but will always show a corresponding high signal on high *b*-value DWI. This effect has also been documented, for example, as an important caveat for assessing prostate cancer on DWI [17]. Therefore, if low signal is seen on the ADC map residual tumour should only be diagnosed if there is corresponding high signal on DWI.

Referring to the ADC map is also important to differentiate T2 shine-through effects from tumoral signal. T2 shine-through is a well-known pitfall in diffusion imaging. Since a DWI sequence is an adaptation of a T2-weighted sequence, the signal intensity observed on DWI is dependent on both water diffusion and the T2 relaxation time. Structures with a very long T2 relaxation time (such as fluids) can therefore retain a high signal as a result of T2 effects, which may be mistaken for restricted diffusion. In rectal cancer, this pitfall will mainly be caused by small amounts of fluid causing high



**Fig. 2** A male patient with a tumour in the distal rectum on the dorsal side. T2-weighted images before treatment (**a** *black arrows*) and after CRT (**b**). **b** After CRT hypointense fibrotic wall thickening is seen (*white arrows*). **c** The corresponding *b*1,000 DW image shows hyperintensity in the rectal wall on the anterior side (*arrowheads*). This signal was misinterpreted as residual tumour

by one of the readers. It is, however, caused by signal pile up from susceptibility effects caused by a small amount of air in the rectal lumen (**b** *asterisk*). The main clue to recognizing this signal as an artefact is that it is located on the opposite side of the tumour bed which makes it very unlikely that it is actually a tumoral diffusion signal



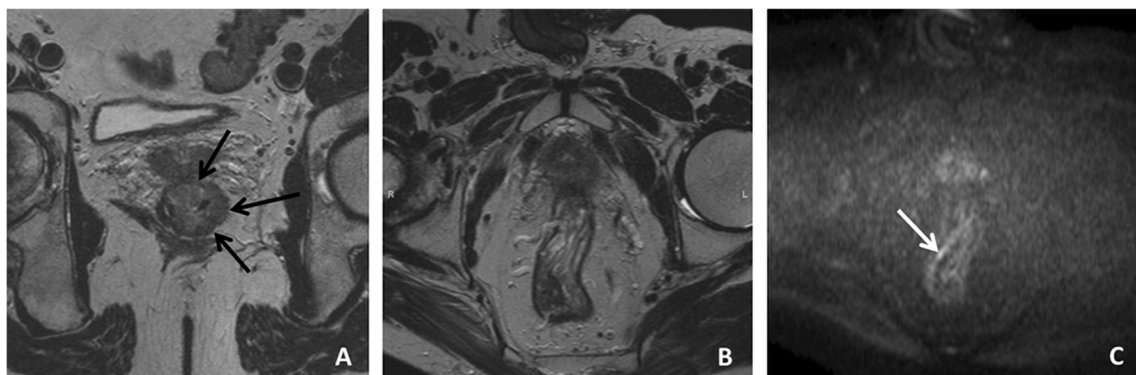
**Fig. 3** Two patients, both with fibrotic wall thickening on T2-weighted MR images on the dorsal side (**a, d white arrows**), after CRT. In both patients a high signal is seen on the corresponding  $b1,000$  DW images (**b, e**). **a–c** First patient. On the DW image (**b**) the signal is star-shaped and corresponds to T2 shine-through effects from fluid in the rectal lumen (**a asterisk**). On the ADC map (**c**), the signal in the lumen is also high (**black arrows**), indicating that there is no actual diffusion restriction. On the

dorsal side there is a markedly hypointense signal (**arrowhead**), caused by the short T2 relaxation times of the collagen in the fibrosis (see also Fig. 2). This patient showed a complete response. **d–f** Second patient. On the DW image (**e**) the high signal is U-shaped with corresponding low signal on the ADC map (**f arrowheads**). This is the typical shape of signal caused by residual tumour. Histopathology showed that this patient had a ypT2 tumour remnant

signal in the rectal lumen. The pitfall of ‘luminal shine-through’ can be corrected by looking at the ADC map where if T2 shine-through is present the signal will remain high (as opposed to structures with true diffusion restriction that will show low signal on DWI). Moreover, as demonstrated in Fig. 3, luminal shine-through effects will typically have a star-shaped configuration, while high signal caused by

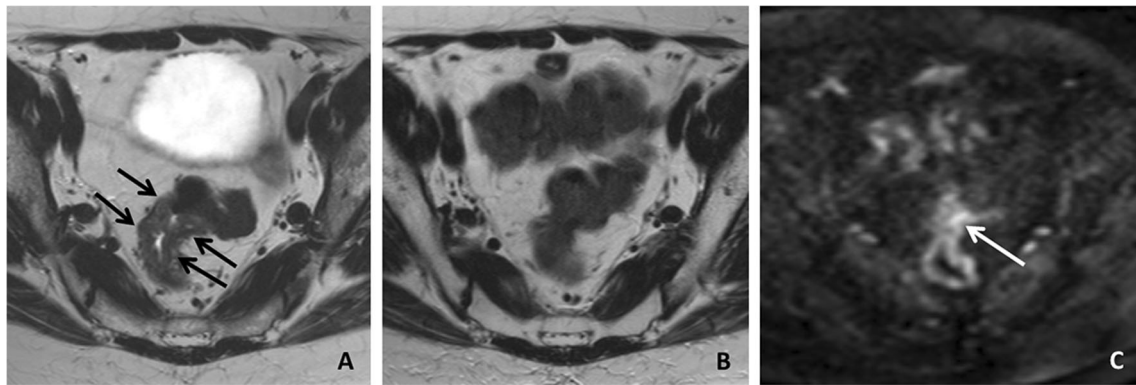
restricted diffusion will typically have a more nodular or tubular/U-shaped configuration. Critically looking at the shape of the signal is another feature that can therefore help differentiate residual tumour from luminal T2 shine-through.

A third potential cause of error was the presence of small artefacts related to susceptibility effects caused by air in the rectal lumen. While severe artefacts that result in large



**Fig. 4** A patient with a tumour in the lower third of the rectum, before treatment (**a arrows**) and after CRT (**b, c**). The T2-weighted (**b**) and diffusion-weighted (**c**) images after CRT are angled in a different plane from that of the primary staging MR image so that it is more difficult to compare the tumour before and after treatment. After CRT some

submucosal oedema is seen on the T2-weighted MR image but there is no clear tumour remnant. On the  $b1,000$  DW image a small focus of high signal is seen (**c arrow**), which was erroneously interpreted as suspicious for residual tumour by both readers. Histopathology showed that this patient had a complete response



**Fig. 5** A patient with a tumour in the upper third of the rectum on primary staging (**a** arrows) and after CRT (**b**, **c**). On the MR image (**b**), the rectal wall is collapsed at the site of the primary tumour, making it difficult to establish whether or not a tumour remnant is still present. On

the corresponding  $b1000$  DW image (**c**) some high signal is seen (arrow). This was, however caused by superposition of the two sides of the rectal wall. Histopathology showed that this patient had a complete response

geometrical distortions are easy to recognize, more subtle artefacts may lead to focal high signal projecting over the rectal wall, which may easily be mistaken for tumour. In these cases, the ADC map will not be of added value. It may be helpful to look at the location of the high signal and critically correlate this with the primary tumour location. If a tumour remnant is present, high signal will occur solely within the boundaries of the (former) tumour bed. Readers were therefore instructed to ignore any high signal occurring outside the tumoral ‘region of interest’ and consider it as nonsuspicious, as illustrated in Fig. 2.

However, given the potential interpretation difficulties caused by such air-induced artefacts, efforts should first be made to prevent them. Solutions advocated include the use of endorectal filling or the use of a small rectal enema, as used in our study. Endorectal filling might also have been beneficial in the small number of patients with tumours in the upper third of the rectum where the rectal wall was completely collapsed on the post-CRT images (Fig. 5). This made it difficult to differentiate high signal from a small tumour remnant from signal caused by superposition of the different layers of the adjacent sides of the rectal wall. Although rectal distension may be helpful in such individual cases, it is currently not routinely advised [18, 19]. From the acquisition point of view, the use of turbo spin echo DWI sequences (rather than the typically used echo planar sequences) may reduce air artefacts [20], although the use of such sequences within the abdomen has so far scarcely been studied. Finally, it is important to ensure optimal sequence angulation by well-trained personnel. This can prevent interpretation difficulties caused by sub-optimal sequence planning as observed in a small number of cases (six patients) in this study. It is mainly important to ensure similar angulation between the pretreatment and post-treatment scans to allow adequate comparison of the tumour before and after treatment.

The pitfalls described above were used as the main input during the expert feedback and teaching sessions. Although

this study was not designed as a formal learning curve study, and we therefore cannot draw any conclusions on the effects of this teaching from a statistical point of view, it is remarkable that in the final 50 study cases both readers achieved an AUC of 0.85, which is similar to that previously reported for expert readers [5–7]. Moreover, the two readers showed good inter-observer agreement in the final 50 cases ( $\kappa$  0.71) while they were only in moderate agreement in the first 50 cases ( $\kappa$  0.58). In the first 50 cases, equivocal (uncertain) scores were assigned to eleven cases by reader 1 and to four cases by reader 2. In the final 50 cases, equivocal scores were assigned to six and zero cases by the two readers, respectively. The difference between the two readers in assigning equivocal scores is remarkable. Although both readers were instructed to assign a confidence level 2 score whenever they felt uncertain about the diagnosis (see Table 1), reader 2 appeared to be more keen to give a conclusive outcome.

Our study had some limitations. First, as described above, in describing potential effects of expert feedback and teaching on diagnostic performance, this study was only a descriptive study. We fully acknowledge that it was not a formal learning curve study, which would require multiple readers and more advanced statistics. Second, there were some variations in the patient preparation and acquisition parameters of the DWI sequences used throughout the study. This reflects daily practice where protocols are subject to optimization/changes over time, which may have had some effect on the study results. However, we believe that this effect would probably have been limited since all scans were deemed to be of good diagnostic quality and because similar  $b1,000$  images (with comparable slice thickness and resolution) were consistently used. Scans with severe artefacts were excluded from the study.

Third, the number of complete responders in this study was very high (46%). This is because patient cases were derived from a database from a referral centre for watchful waiting to which patients with a suspected good response to CRT are referred for final response evaluation. Given the primary study



outcome (discrimination of complete responders) this may in fact have been beneficial, but it is not representative of the percentage of complete responders that will generally be encountered in daily clinics, which lies more in the range 10–24% [21]. The two readers in this study were aware of this specific ‘case mix’ at our institution. Finally, in 38 patients managed according to a watchful waiting policy, a clinical complete response (with a mean recurrence-free follow-up period of 37 months) was used as a surrogate endpoint of a complete response, according to methods previously reported [5, 13–15]. However, we acknowledge that this is a suboptimal standard of reference as very late recurrences (although reported to be rare after 2 years) may still occur in this group [1, 2, 22, 23].

In conclusion, this study showed that there are five important potential DWI interpretation pitfalls which were documented with imaging examples and may serve as a reference to teach future readers interested in the use of DWI for rectal tumour response evaluation. The study also showed that non-expert readers (when trained using these pitfalls) can achieve a diagnostic performance comparable to that previously reported for expert radiologists with AUCs of 0.85.

#### Compliance with ethical standards

**Guarantor** The scientific guarantor of this publication is Doenja Lambregts.

**Conflict of interest** The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

**Funding** The authors state that this work did not receive any funding.

**Statistics and biometry** No complex statistical methods were necessary for this paper.

**Ethical approval** Institutional Review Board approval was obtained.

**Informed consent** Written informed consent was waived by the Institutional Review Board.

**Overlap of study subjects or cohorts** Some of the study patients were included in previously reported studies, as follows:

#### Clinical studies

- The following study included the 38 patients with a clinical complete response after CRT with follow-up according to a watchful waiting strategy:

Martens MH, et al. Long-term outcome of an organ preservation program after neoadjuvant treatment for rectal cancer. *J Natl Cancer Inst.* 2016;108(12).

This study reports the clinical outcomes in these patients and did not focus on imaging.

#### Studies on diffusion-weighted imaging

- The following study included one of the patients in our study:

Curvo-Semedo L, et al. Rectal cancer: assessment of complete response to preoperative combined radiation therapy with chemotherapy – conventional MR volumetry versus diffusion-weighted MR imaging. *Radiology.* 2011;260(3):734–43.

This study focused on quantitative (ADC and volume) measurements and not on visual assessment of DWI.

- The following study included 15 of the patients in our study:

Lambregts DM, et al. MRI and diffusion-weighted MRI volumetry for identification of complete tumour responders after preoperative chemoradiotherapy in patients with rectal cancer: a bi-institutional validation study. *Ann Surg.* 2015;262(6):1034–9.

This study focused on quantitative (volume) measurements on DWI. It did not include visual DWI assessment, pitfalls or teaching effects.

- The following study included 12 of the patients in our study:

van Heeswijk MM, et al. Automated and semiautomated segmentation of rectal tumour volumes on diffusion-weighted MRI: can it replace manual volumetry? *Int J Radiat Oncol Biol Phys.* 2016;94(4):824–31.

This study was a technical study on automated software methods to measure tumour volumes on DWI. It did not include a visual assessment or comparison of DWI findings with treatment response.

- The following study included 21 of the patients in our study:

van Heeswijk MM, et al. DWI for assessment of rectal cancer nodes after chemoradiotherapy: is the absence of nodes on DWI proof of a negative nodal status? *AJR Am J Roentgenol.* 2017;208(3):W79–W84.

This study focused on DWI for lymph node assessment and not for tumour response evaluation.

#### Methodology

- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

#### References

1. Habr-Gama A, Gama-Rodrigues J, Sao Juliao GP et al (2014) Local recurrence after complete clinical response and watch and wait in rectal cancer after neoadjuvant chemoradiation: impact of salvage therapy on local disease control. *Int J Radiat Oncol Biol Phys* 88: 822–828
2. Maas M, Beets-Tan RG, Lambregts DM et al (2011) Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 29:4633–4640
3. Lezoche E, Guerrieri M, Paganini AM et al (2005) Transanal endoscopic versus total mesorectal laparoscopic resections of T2-N0 low rectal cancers after neoadjuvant treatment: a prospective

- randomized trial with a 3-years minimum follow-up period. *Surg Endosc* 19:751–756
4. van der Paardt MP, Zagers MB, Beets-Tan RG, Stoker J, Bipat S (2013) Patients who undergo preoperative chemoradiotherapy for locally advanced rectal cancer restaged by using diagnostic MR imaging: a systematic review and meta-analysis. *Radiology* 269:101–112
  5. Lambregts DM, Vandecaveye V, Barbaro B et al (2011) Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. *Ann Surg Oncol* 18:2224–2231
  6. Kim SH, Lee JM, Hong SH et al (2009) Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumor response to neoadjuvant chemo- and radiation therapy. *Radiology* 253:116–125
  7. Song I, Kim SH, Lee SJ, Choi JY, Kim MJ, Rhim H (2012) Value of diffusion-weighted imaging in the detection of viable tumour after neoadjuvant chemoradiation therapy in patients with locally advanced rectal cancer: comparison with T2-weighted and PET/CT imaging. *Br J Radiol* 85:577–586
  8. Sathyakumar K, Chandramohan A, Masih D, Jesudasan MR, Pulimood A, Eapen A (2016) Best MRI predictors of complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. *Br J Radiol* 89:20150328
  9. Sassen S, de Booi M, Sosef M et al (2013) Locally advanced rectal cancer: is diffusion weighted MRI helpful for the identification of complete responders (ypT0N0) after neoadjuvant chemoradiation therapy? *Eur Radiol* 23:3440–3449
  10. Miglioretti DL, Gard CC, Carney PA et al (2009) When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology* 253:632–640
  11. Boellaard TN, Nio CY, Bossuyt PM, Bipat S, Stoker J (2012) Can radiographers be trained to triage CT colonography for extracolonic findings? *Eur Radiol* 22:2780–2789
  12. Saba L, Guerriero S, Sulis R et al (2011) Learning curve in the detection of ovarian and deep endometriosis by using magnetic resonance: comparison with surgical results. *Eur J Radiol* 79:237–244
  13. Maas M, Lambregts DM, Nelemans PJ et al (2015) Assessment of clinical complete response after chemoradiation for rectal cancer with digital rectal examination, endoscopy, and MRI: selection for organ-saving treatment. *Ann Surg Oncol* 22:3873–3880
  14. Lambregts DM, Lahaye MJ, Heijnen LA et al (2016) MRI and diffusion-weighted MRI to diagnose a local tumour regrowth during long-term follow-up of rectal cancer patients treated with organ-preservation after chemoradiotherapy. *Eur Radiol* 26:2118–2125
  15. Curvo-Semedo L, Lambregts DM, Maas M et al (2011) Rectal cancer: assessment of complete response to preoperative combined radiation therapy with chemotherapy – conventional MR volumetry versus diffusion-weighted MR imaging. *Radiology* 260:734–743
  16. Gatehouse PD, Bydder GM (2003) Magnetic resonance imaging of short T2 components in tissue. *Clin Radiol* 58:1–19
  17. Weinreb JC, Barentsz JO, Choyke PL et al (2016) PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. *Eur Urol* 69:16–40
  18. Slater A, Halligan S, Taylor SA et al (2006) Distance between the rectal wall and mesorectal fascia measured by MRI: effect of rectal distension and implications for preoperative prediction of a tumor-free circumferential resection margin. *Clin Radiol* 61:65–70
  19. Beets-Tan RG, Lambregts DM, Maas M et al (2013) Magnetic resonance imaging for the clinical management of rectal cancer patients: recommendations from the 2012 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 23:2522–2531
  20. Mikayama R, Yabuuchi H, Kobayashi K et al (2016) Comparison of image quality for diffusion-weighted imaging in the head and neck between turbo spin-echo and echo-planar imaging. *Sci Poster Eur Congr Radiol*. doi:10.1594/ecr2016/C-1045
  21. Maas M, Nelemans PJ, Valentini V et al (2010) Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 11:835–844
  22. Martens M, Maas M, Heijnen LA et al (2016) Long term outcome of an organ preservation program after neoadjuvant treatment for rectal cancer. *J Natl Cancer Inst* 108(12). doi:10.1093/jnci/djw171
  23. Smith JD, Ruby JA, Goodman KA et al (2012) Nonoperative management of rectal cancer with complete clinical response after neoadjuvant therapy. *Ann Surg* 256:965–972