TITLE*:

Diffusion-weighted MRI to assess response to chemoradiotherapy in rectal cancer: learning curve, pitfalls, tips and tricks.

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

DOI 10.1007/s00330-017-4830-z

AUTHORS

Doenja M.J. Lambregts[1], Miriam M. van Heeswijk[2,3,4], Andrea Delli Pizzi[5], Saskia van Elderen[6], Luisa Andrade[7], Nicky Peters[8], Peter Kint[9], Margreet Osinga-de Jong[8], B Mearadji[10], Rik Ooms[11], Max Lahaye[1], Monique Maas[1, 2], Geerard Beets[12,4], Frans C.H. Bakers[2], Regina Beets-Tan[1,4]

AFFILIATIONS

1 Department of Radiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

2 Department of Radiology, Maastricht University Medical Centre, Maastricht, The Netherlands

3 Department of Surgery, Maastricht University Medical Centre, Maastricht, The Netherlands

4 GROW School for Oncology and Developmental Biology - Maastricht University, Maastricht, The Netherlands

5 Department of Neuroscience and Imaging, Gabriele d'Annunzio University, SS. Annunziate Hospital, Chieti, Italy

6 Department of Radiology, Leiden University Medical Centre, Leiden, The Netherlands

7 Department of Radiology, Hospitais Da Universidade De Coimbra, Coimbra, Portugal

8 Department of Radiology, Zuyderland Medical Center, Sittard/Heerlen, The Netherlands

9 Department of Radiology, Amphia Hospital, Breda, The Netherlands

10 Department of Radiology, Amsterdam Medical Centre, Amsterdam, The Netherlands

11 Department of Radiology, Maxima Medical Centre, Veldhoven, The Netherlands

12 Department of Surgery, The Netherlands Cancer Institute, Amsterdam, The Netherlands

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 papers focus on the use of DWI for evaluation of response to chemoradiotherapy. 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 undergo a very good response to long course chemoradiotherapy 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. Though morphological MRI is beneficial to assess tumour downsizing and downstaging, it has difficulties in determining the presence and extent of residual tumour within areas of post-radiation fibrosis. In a recent meta-analysis sensitivity for overall tumour restaging after CRT with MRI was only 50%, with even poorer results (sensitivity 19%) for the specific selection of complete responders (4). Overall sensitivity for restaging was considerably better (84%) in subgroup analysis focusing on studies that used DWI. This is in conjunction with various studies that showed that addition of DWI significantly improves the performance of MRI to differentiate viable tumour within areas of post-radiation fibrosis (5-7). Moreover, in a recent publication it was shown that out of a variety of MRI features (e.g. tumour location, signal intensity. T- and N-stage, tumour volume and volume reduction ratios), visual assessment of response on DWI was one of the best predictors to diagnose a complete tumour response on MRI (8).

In the majority of the published reports, DW images were read by expert-radiologists typically with dedicated experience (2-13 years) in reading rectal MRI and previous experience in reading diffusion-weighted images (5, 6, 9, 10). The performance of such readers may not necessarily reflect that of general readers from non-referral centers. 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 exams, and MRI for diagnosing endometriosis) it has been demonstrated that it requires a learning curve before non-expert readers can reach a certain diagnostic performance level (11-13). For the assessment of DWI for rectal tumour response evaluation, this effect has not previously been studied or documented. It would also be helpful to gain knowledge on which pitfalls and interpretation errors occur most common amongst non-expert readers when assessing DWI post-CRT so that these may serve as a teaching reference for readers who wish to improve their DWI reading skills. Hence the aim of this study was

twofold: [1] to assess the learning curve for assessing rectal tumour response on post-chemoradiotherapy DWI and [2] to explore which are the most common image interpretation errors and pitfalls.

MATERIALS AND METHODS

This study concerns 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 rectal cancer patients were selected who were treated with long course neoadjuvant treatment and underwent a restaging MRI including a DWI sequence at [BLINDED]. Five patients were used as training cases and the other 100 patients constituted the study (test) dataset to study learning curve effects and pitfalls. Inclusion criteria consisted of *a.* histopathological proven rectal adenocarcinoma, *b.* neoadjuvant treatment consisting of long course chemoradiotherapy (or radiotherapy only with a prolonged waiting interval), c*.* availability of a good quality restaging MRI including a DWI sequence, and *d.* data on final response outcome. Patients with low quality DWI exams (e.g. severe susceptibility artefacts due to metal implants) as well as patients with mucinous tumours (as these are known to exhibit different signal characteristics on DWI) were excluded. The routine neoadjuvant treatment consisted of 50.4 Gy radiation + capecitabine 2 x 825 mg/m^2/d during the radiation period.

MR imaging

The restaging MR examinations were all performed at 1.5T (Intera (Achieva) or 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 MRIs was 6-10 weeks. To reduce bowel motility, patients received 20 mg of scopolamine butylbromide (Buscopan, Boehringer Ingelheim, Germany) intravenously, either in case of aniticipated 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 micro-enema (Microlax ®, McNeil Healthcare (Ireland) Ltd, Dublin, Ireland) ±15 minutes before onset of the examination, to reduce the amount of air in the rectal lumen. The standard clinical MRI protocol included 2D-T2-weighted fast spin echo sequences in sagittal, axial and coronal plane and an axial echo planar imaging diffusion-weighted imaging sequence with b1000 being the highest b-value. The axial T2-

weighted and DWI sequences were angled in identical plane perpendicular to the rectal tumour axis. Apparent Diffusion Coefficient (ADC) maps were automatically generated at the operating system. Detailed sequence parameters are given in appendix 1.

Training

All images were read by two senior radiology residents ([BLINDED] and [BLINDED]) with an interest in abdominal imaging but with no specific previous experience in reading DWI of rectal cancer. Before onset of the study both readers received a short training from an expert radiologist ([BLINDED], with 8 years of specific expertise in reading rectal MRI and DWI) in how to read DWI-MRI of rectal cancer using powerpoint presentations as well as a hands on training session with the first 5 cases of the 105 patients in the dataset.

Scoring and feedback

The two readers were asked to assess the remaining n=100 cases (the test dataset) and for each case to report the likelihood of a complete tumour response using a 5-point confidence level score (0=definitely complete response; 1=probably complete response; 2=equivocal; 3=probably residual tumour; 4=definitely residual tumour). The readers based their score primarily on the high b-value (b1000) diffusion weighted images. The corresponding ADC maps and T2-weighted images, as well as the primary staging MR images (including DWI) were also at the readers disposal. In the first 30 cases, the two readers received expert feedback (as well as the final response outcome) after each single case. In the last 70 cases the readers received feedback after every ten cases. For each case the supervising expert reader documented any potential reasons of error and the main image interpretation pitfalls encountered by the two readers. Readers received specific feedback on these errors and pitfalls for each case.

Reference standard:

In 62 patients the final response outcome was based on the final tumour stage at histopathology after surgical resection (ypT-stage). The remaining 38 patients had strong clinical evidence of a complete response and were followed according to a watchful waiting strategy (including follow up with endoscopy and imaging in 3-monthly in the first year and 6-monthly in the second to fifth year). In these patients a sustained clinical complete response (with repeated negative MRIs and endoscopies +/- biopsies) was considered a surrogate endpoint for a complete response. Response

for the whole patient group was dichotomized between residual tumour (ypT1-4) and complete response (ypT0 after surgery or ycT0 with a sustained complete response during wait-and-see).

Statistical analysis

Results were analysed using the Statistical Package for the Social Sciences (SPSS, version 22, IBM® Corps., Armonk, NY, USA). Receiver operator characteristics (ROC) curves were constructed to analyze diagnostic performance to assess 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 (before onset of the study) between confidence level 0-2 and 3-4. Results for each readers were separately analyzed for each consecutive set of 20 patients.

RESULTS

Patient characteristics

Of the 100 test patients 69 were male, 31 female. Median age was 64 years (range 31-82). In total 46 patients were complete responders and 54 patients had residual tumour. The mean follow-up for the patients in the watchful waiting group with a sustained clinical complete response as a surrogate end-point was 32 (±11) months. Further patient characteristics are given in Table 1.

Diagnostic performance and learning curve

Figure 1 and Table 2 show the results for the two readers over time (calculated per each consecutive 20 cases). Overall accuracy improved from a minimum of 70% (R1) and 65% (R2) in the first 20 patients to a maximum of up to 85% for both readers. AUCs for the 5 consecutive sets of 20 patients were 0.75, 0.83, 0.83, 0.89 and 0.82 for reader 1 and 0.77, 0.61, 0.84, 0.87 and 0.88 for reader 2. For reader 1 the number of equivocal scores was 4 in the first 20 cases, and decreased to 2 and 3 in the last 2 sets of 20 cases. For reader 2 the number of equivocal scores decreased from 2 to 0.

Reasons for error and interpretation pitfalls:

The 5 most common pitfalls (as documented by the expert reader) are summarized in Table 3:

1. *Hypointense fibrosis on ADC map* (potential pitfall in 41% of the cases): 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 (Figure 2).
2. *Susceptibility effects* (potential pitfall in 33% of the cases): Readers were trained to recognize high signal caused by susceptibility effects and differentiate these from actual tumour signal (Figure 3).
3. *T2 shine through* (potential pitfall in 12% of the cases): A potential pitfall was the presence of high signal in the rectal lumen on b1000 DWI caused by T2 shine through effects of intraluminal fluid. Readers were instructed how to recognize these luminal shine through effects by comparing the diffusion images with the ADC map and by taking into account the shape of the signal, as luminal shine through effects are typically star-shaped while high signal caused by tumour is typically more nodular, or tubular/U-shaped (Figure 4).
4. *Suboptimal sequence angulation* (potential pitfall in 6% of the cases): In proximal rectal tumours, angulation perpendicular to the tumour will result in coronal-like imaging planes, which are typically more difficult to interpret. Moreover when the pre-CRT and post-CRT images are not angled in identical planes it will be more difficult to compare the images and interpret the post-CRT diffusion images (Figure 5).
5. *Collapsed rectal wall* (potential pitfall in 2% of the cases): In patients with a collapsed rectal wall it can be difficult to determine whether a high signal on DWI is caused by superposition of the two sides of the rectal wall or by actual tumoural signal (Figure 6).

DISCUSSION

First aim of this study was to assess the effect of expert teaching and feedback on the diagnostic performance of non-expert readers in reading diffusion-weighted MRI of rectal cancer to differentiate between complete responders and patients with residual tumour. The results of our study show that expert teaching and feedback indeed leads to a certain learning curve. Overall diagnostic performance gradually improved: at the end of study both readers achieved an area under the ROC curve of up to 0.88-0.89 (compared to AUC 0.61-0.83 in the first two sets of patients), which is concordant with that previously reported in literature for expert readers (5-7). Apparently, after intensive teaching, non-expert readers are able to achieve a level similar to that of experts. As can be appreciated in Figure 1 and Table 2, this level is reached after approximately 40-60 cases. Moreover,

the readers (particularly reader 2) became more confident over time, which is reflected by the decrease in the number of equivocal scores.

Second aim was to document the most common potential interpretation pitfalls to serve as a teaching basis for future readers. An important potential pitfall was the misinterpretation of low signal on the ADC map as being suspicious for residual tumour. When studying the most basic concepts of DWI, one is typically instructed to consider low signal on ADC 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 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 ADC map (14). 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 for example also been documented as an important caveat for assessing prostate cancer on DWI (15). Therefore, one should only diagnose residual tumour in case of low signal on the ADC map with corresponding high signal on DWI.

Referring to the ADC map is furthermore important to differentiate T2 shine through effects from tumoural 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 signal in the rectal lumen. The pitfall of 'luminal shine through' can be corrected by looking at the ADC map where in case of T2 shine through the signal will remain high (as opposed to structures with true diffusion restriction that will show low signal on DWI). Moreover, as demonstrated in Figure 4, 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 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. What may be of help, is to look at the location of the high signal and critically correlate this with the primary tumour location and the site of the residual changes (fibrosis) potentially suspicious for tumour on the post-CRT images. If a tumour remnant is present, a high signal will solely occur within the boundaries of the (former) tumor bed. Readers were therefore instructed to ignore any high signal occurring outside the tumoural 'region of interest' and consider it as non-suspicious as illustrated in Figure 3. However, given the potential interpretation difficulties caused by such air-induced artefacts, efforts should be undertaken to prevent them. Several solutions have been advocated for this purpose, including the use of endorectal filling and the use of a small rectal enema, such as used in our study. The use of endorectal filling might theoretically also have been beneficial in the 2% of patients with tumours in the upper third of the rectum in whom the rectal wall was completely collapsed on the post-CRT images (Figure 6). 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. From the acquisition point of view, use of turbo spin echo DWI sequences (rather than the typically used echo planar imaging DWI methods) may offer a potential solution to reduce air artefacts (16), 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 might have prevented the pitfalls caused by suboptimal sequence planning observed in a small number of cases (n=6). It is mainly important to ensure similar angulation between the pre- and posttreatment scans to allow for adequate comparison of the tumour before and after treatment.

Our study has some limitations. First, due to the highly time consuming nature of the study where readers received personal feedback from an expert for each study case, the study cases were only reviewed by two readers, whereas typically learning curve studies are performed with multiple readers. 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, but may have had some effect on our study results. We believe that this effect will, however, likely be limited since all scans were deemed to be of good diagnostic quality and because similar DWI protocols with b1000 being the highest b-value were consistently used. Scans with severe artefacts were excluded from the study. Finally, the number of complete responders in our study was very high (46%). This is because patient cases were derived 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 be beneficial, but it will not be representative for the percentage of complete responders that will generally be encountered in daily clinics, which lies

more in the range of 10-24% (17). The two readers in this study were aware of the 'case mix' at our institution.

In conclusion, results of this study suggest a certain learning curve for reading diffusion-weighted MRI of rectal cancer to discriminate between complete responders and patients with residual tumour after neoadjuvant chemoradiotherapy. After approximately 40-60 cases with intensive expert feedback and training, non-expert readers can obtain a diagnostic accuracy of around 85%, which is similar to that previously reported for experienced radiologists. Moreover, five important potential pitfalls were identified and documented with imaging examples, which may serve as reference to teach future readers interested in the use of DWI fore rectal tumour response evaluation.

REFERENCES

1. Habr-Gama A, Gama-Rodrigues J, Sao Juliao GP, Proscurshim I, Sabbagh C, Lynn PB, et al. 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. 2014;88(4):822-8.
2. Maas M, Beets-Tan RG, Lambregts DM, Lammering G, Nelemans PJ, Engelen SM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. J Clin Oncol. 2011;29(35):4633-40.
3. Lezoche E, Guerrieri M, Paganini AM, D'Ambrosio G, Baldarelli M, Lezoche G, et al. 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. 2005;19(6):751-6.
4. van der Paardt MP, Zagers MB, Beets-Tan RG, Stoker J, Bipat S. Patients who undergo preoperative chemoradiotherapy for locally advanced rectal cancer restaged by using diagnostic MR imaging: a systematic review and meta-analysis. Radiology. 2013;269(1):101-12.
5. Lambregts DM, Vandecaveye V, Barbaro B, Bakers FC, Lambrecht M, Maas M, et al. Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. Ann Surg Oncol. 2011;18(8):2224-31.
6. Kim SH, Lee JM, Hong SH, Kim GH, Lee JY, Han JK, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumor response to neoadjuvant chemo- and radiation therapy. Radiology. 2009;253(1):116-25.
7. Song I, Kim SH, Lee SJ, Choi JY, Kim MJ, Rhim H. 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. 2012;85(1013):577-86.
8. Sathyakumar K, Chandramohan A, Masih D, Jesudasan MR, Pulimood A, Eapen A. Best MRI predictors of complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. Br J Radiol. 2016;89(1060):20150328.

9. Sassen S, de Booij M, Sosef M, Berendsen R, Lammering G, Clarijs R, et al. Locally advanced rectal cancer: is diffusion weighted MRI helpful for the identification of complete responders (ypT0N0) after neoadjuvant chemoradiation therapy? Eur Radiol. 2013;23(12):3440-9.
10. Song I, Kim SH, Lee SJ, Choi JY, Kim MJ, Rhim H. 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. 2012;85(1013):577-86.
11. Miglioretti DL, Gard CC, Carney PA, Onega TL, Buist DS, Sickles EA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. Radiology. 2009;253(3):632-40.
12. Boellaard TN, Nio CY, Bossuyt PM, Bipat S, Stoker J. Can radiographers be trained to triage CT colonography for extracolonic findings? Eur Radiol. 2012;22(12):2780-9.
13. Saba L, Guerriero S, Sulis R, Pilloni M, Ajossa S, Melis G, et al. Learning curve in the detection of ovarian and deep endometriosis by using Magnetic Resonance: comparison with surgical results. Eur J Radiol. 2011;79(2):237-44.
14. Gatehouse PD, Bydder GM. Magnetic resonance imaging of short T2 components in tissue. Clin Radiol. 2003;58(1):1-19.
15. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. Eur Urol. 2016;69(1):16-40.
16. Mikayama R, Yabuuchi H, Kobayashi K, Sonoda S, Nagatomo S, Kimura M, et al. Comparison of image quality for diffusion-weighted imaging in the head and neck between turbo spin-echo and echo-planar imaging. Scientific poster European Congress of Radiology 2016 (DOI: 101594/ecr2016/C-1045). 2016.
17. Maas M, Nelemans PJ, Valentini V, Das P, Rodel C, Kuo LJ, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. Lancet Oncol. 2010;11(9):835-44.

TABLES

Table 1. Baseline patient and treatment characteristics	
Variable	**N° Patients (total n=100)**
Sex:	
Male	69
Female	31
Median age (range)	64 (31-82)
Primary cTN stage (as assessed with MRI)	
cT-stage	
cT1-2	17
cT3	74
cT4	9
cN-stage	
cN0	27
cN1	27
cN2	46
Surgical treatment	
Low Anterior Resection	42
Abdominoperianal Resection microsurgery	12
Transanal Endoscopic Microsurgery	4
Pelvic exenteration	4
None (watchful waiting)	38
Mean follow up in months (±SD)	32 (±11)
Final yTN stage	
yT-stage	
yT0*	46
yT1	7
yT2	16
yT3	25
yT4	6
yN-stage	
yN0^	77
yN1	18
yN2	5
Final tumour response (yT-stage)	
Complete response*	46
Residual tumour	54

* based on histopathology after surgery in n=8; based on long-term follow-up in n=38 patients with a sustained clinical complete response undergoing watchful waiting.
^ based on histopathology after surgery in n=35; based on long-term follow-up without evidence of recurrence in n=4 patients after TEM and n=38 patients with a sustained clinical complete response undergoing watchful waiting.

Table 2. Diagnostic accuracy figures and number of equivocal scores for each subset of 20 cases.					
Reader 1					
Pt. cases	1-20	21-40	41-60	61-80	81-100
AUC (95% CI)	0.75 (0.53-0.97)	0.83 (0.64-1.00)	0.83 (0.64-1.00)	0.89 (0.74-1.00)	0.82 (0.64-1.00)
Accuracy	70% (14/20)	70% (14/20)	75% (15/20)	85% (17/20)	75% (15/20)
Sensitivity	60% (6/10)	64% (9/14)	63% (5/8)	90% (9/10)	58% (7/12)
Specificity	80% (8/10)	83% (5/6)	83% (10/12)	80% (8/10)	100% (8/8)
PPV	75% (6/8)	90% (9/10)	71% (5/7)	82% (9/11)	100% (7/7)
NPV	67% (8/12)	50% (5/10)	77% (10/13)	89% (8/9)	62% (8/13)
N° equivocal scores	4	4	4	2	3
Reader 2					
Pt. cases	1-20	21-40	41-60	61-80	81-100
AUC (95% CI)	0.77 (0.55-0.98)	0.61 (0.34-0.89)	0.84 (0.67-1.00)	0.87 (0.69-1.00)	0.88 (0.72-1)
Accuracy	65% (13/20)	70% (14/20)	80% (16/20)	85% (17/20)	85% (17/20)
Sensitivity	60% (6/10)	86% (12/14)	75% (6/8)	90% (9/10)	92% (11/12)
Specificity	70% (7/10)	33% (2/6)	83% (10/12)	80% (8/10)	75% (6/8)
PPV	67% (6/9)	75% (12/16)	75% (6/8)	82% (9/11)	85% (11/13)
NPV	64% (9/11)	50% (2/4)	83% (10/12)	89% (8/9)	86% (6/7)
N° equivocal scores	2	1	1	0	0
Numbers (except AUCs and number of equivocal scores) are percentages. Unless otherwise indicated numbers in parentheses are absolute numbers					

Table 3. Overview of main potential causes of error (pitfalls)

Nº	**Potential pitfall**	**Illustrated in:**	**% of cases***
1	Misinterpretation of low signal on ADC map in case of fibrosis	Figure 2	41%
2	Susceptibility effects	Figure 3	33%
3	T2 shine through of fluid in the rectal lumen	Figure 4	12%
4	Suboptimal sequence angulation	Figure 5	6%
5	Collapsed rectal wall	Figure 6	2%

* Represents the percentage of cases in which a potential pitfall was present cq. identified by the expert reader after discussing the case with the readers. It does not necessarily indicate that the case was in fact misinterpreted by the two study readers.

FIGURES

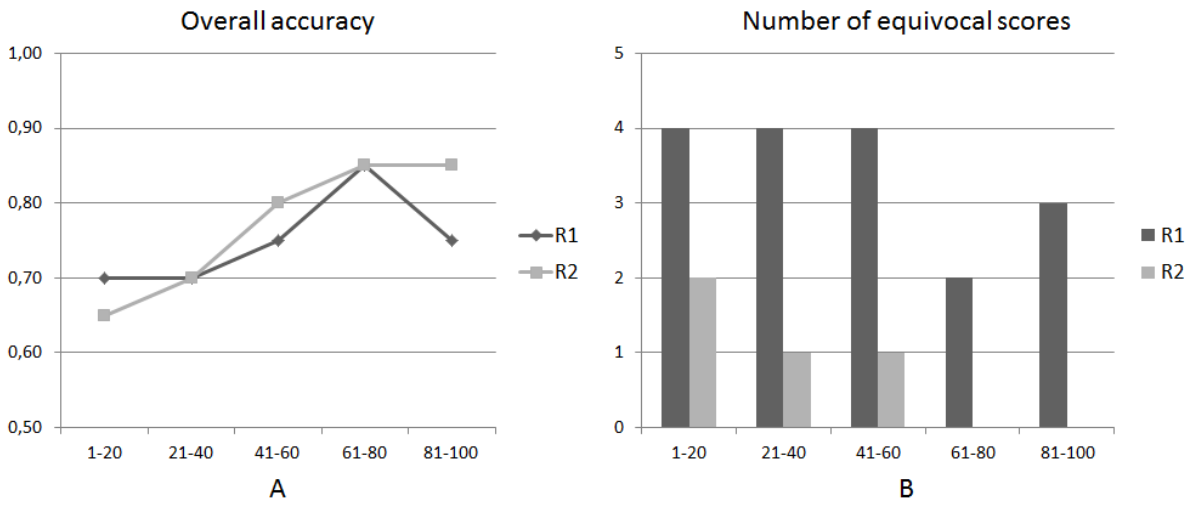

Figure 1.

Graphs illustrating the evolution in overall accuracy (A) and in the number of equivocal scores (B) per consecutive set of 20 cases.

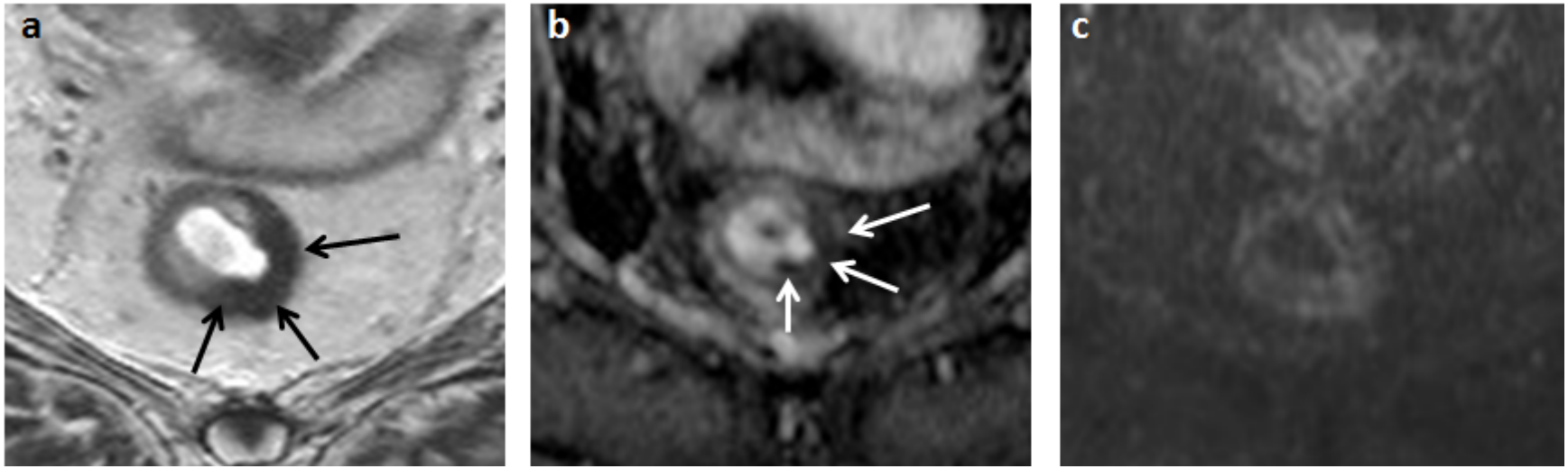

Figure 2. Example of the post-CRT images of a male patient with a tumour on the left dorsolateral side in the mid-rectum. On the restaging T2-weighted MRI (a) a semi-circular fibrotic wall thickening is visualized (black arrows). On the ADC map (b) the wall thickening is markedly hypointense (white arrows). On the b1000 diffusion-weighted image (c) no high signal is observed. This patient turned out to be a complete responder. 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 is not due to diffusion restriction, which is why there is no high signal present on the b1000 diffusion images.

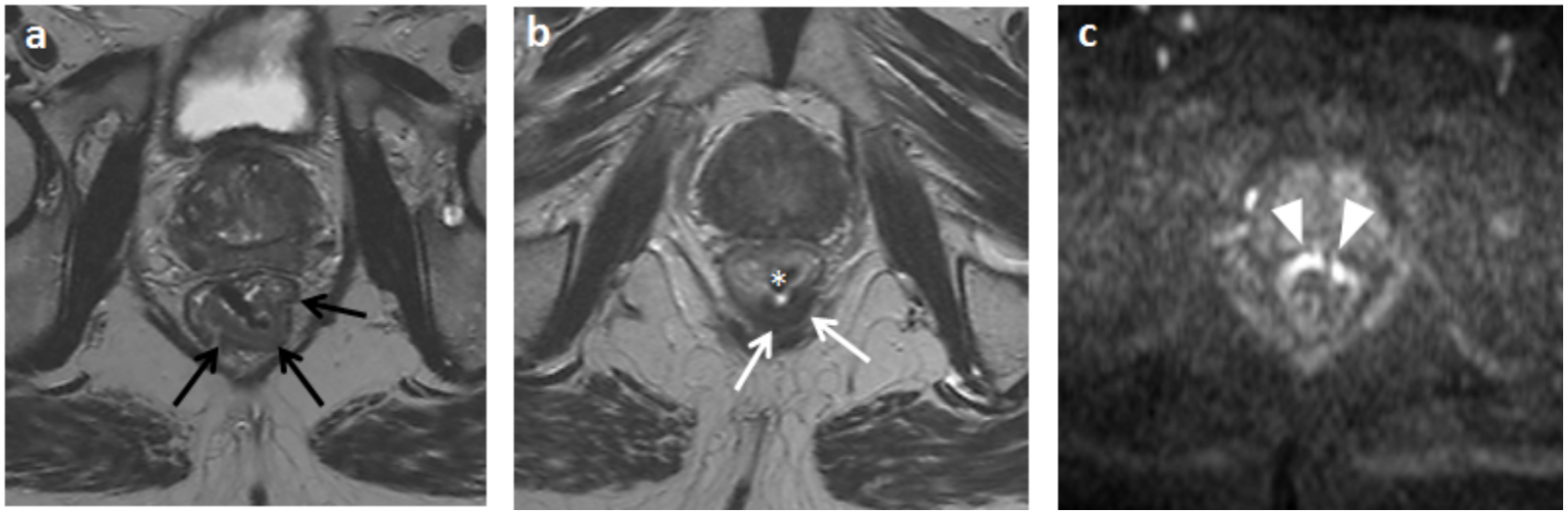

Figure 3. T2-weighted images of a male patient with a tumour in the distal rectum at the dorsal side before treatment (black arrows in a) and after chemoradiotherapy (b). After CRT a hypointense fibrotic wall thickening is visualized (white arrows). On the corresponding b1000 DWI (c) a hyperintensity is seen in the rectal wall at the anterior side (arrowheads). This signal was misinterpreted as residual tumour by one of the readers. It is, however, caused by pile up of signal due to susceptibility effects caused by a small amount of air in the rectal lumen (* in b). The main clue to recognize this signal as an artifact is that it is located at the opposite side of the tumour bed which makes it very unlikely that this signal is caused by actual tumoural diffusion signal.

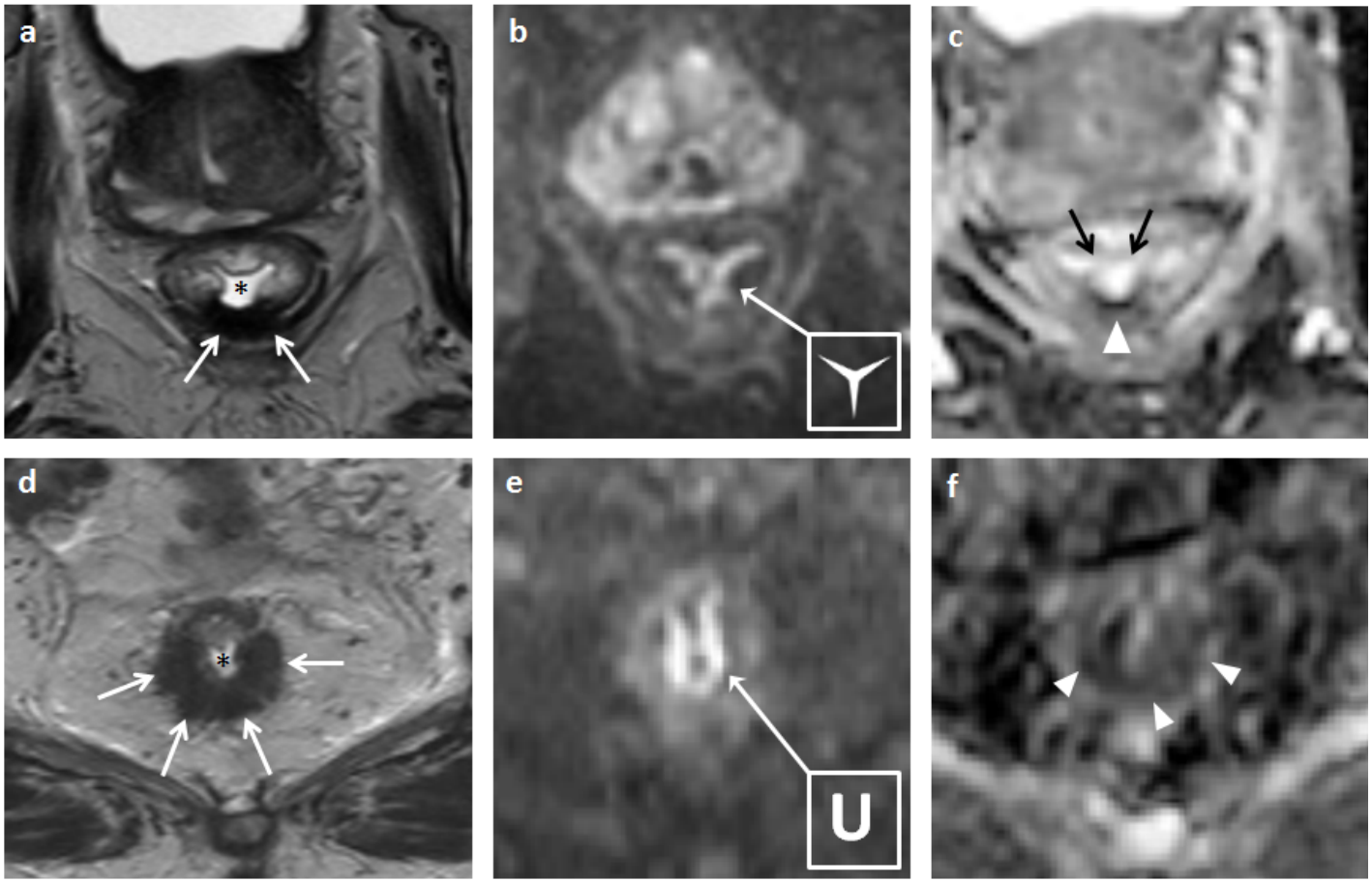

Figure 4. Example of the post-chemoradiotherapy images of two patients, both with a fibrotic wall thickening at the dorsal side (white arrows in a and d). In both patients a high signal is visualized on the corresponding b1000 DWI (b and e). In the upper patient the signal is star-shaped and corresponded to T2 shine through effects of fluid in the rectal lumen (* in a). On the ADC map, the signal in the lumen is also high (black arrows), indicating that there is no actual diffusion restriction. At the dorsal side there is a markedly hypointense signal on the ADC (arrowhead), caused by the short T2 relaxation times of the collagen in the fibrosis (see also Figure 2). This patient was confirmed to be a complete responder. In the lower patient, the high signal on DWI is U-shaped (e) with corresponding low signal on the ADC-map (arrowheads in f). This is the typical shape of signal caused by residual tumour. At histopathology this patient had a ypT2 tumour remnant.

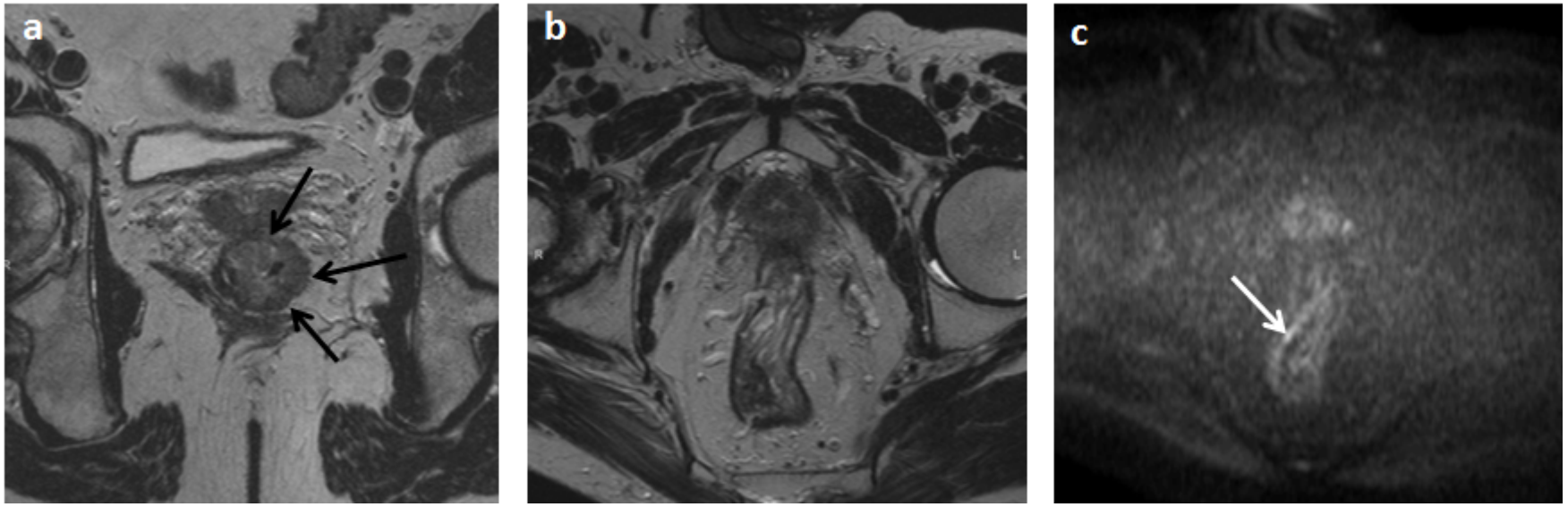

Figure 5. Example of a patient with a tumour in the lower third of the rectum, before treatment (arrows in a) and after chemoradiotherapy (b,c). The post-CRT T2-weighted (b) and diffusion-weighted (c) images were angled in a different plane than the primary staging images. As a result is is more difficult to compare the tumour before and after treatment. After CRT some submucosal edema is observed on T2-weighted MRI but no clear tumour remnant. On b1000 DWI a small focus of high signal was observed (arrow in c), which was erroneously interpreted to be suspicious for residual tumour by both readers. At histopathology this patient had a complete response.

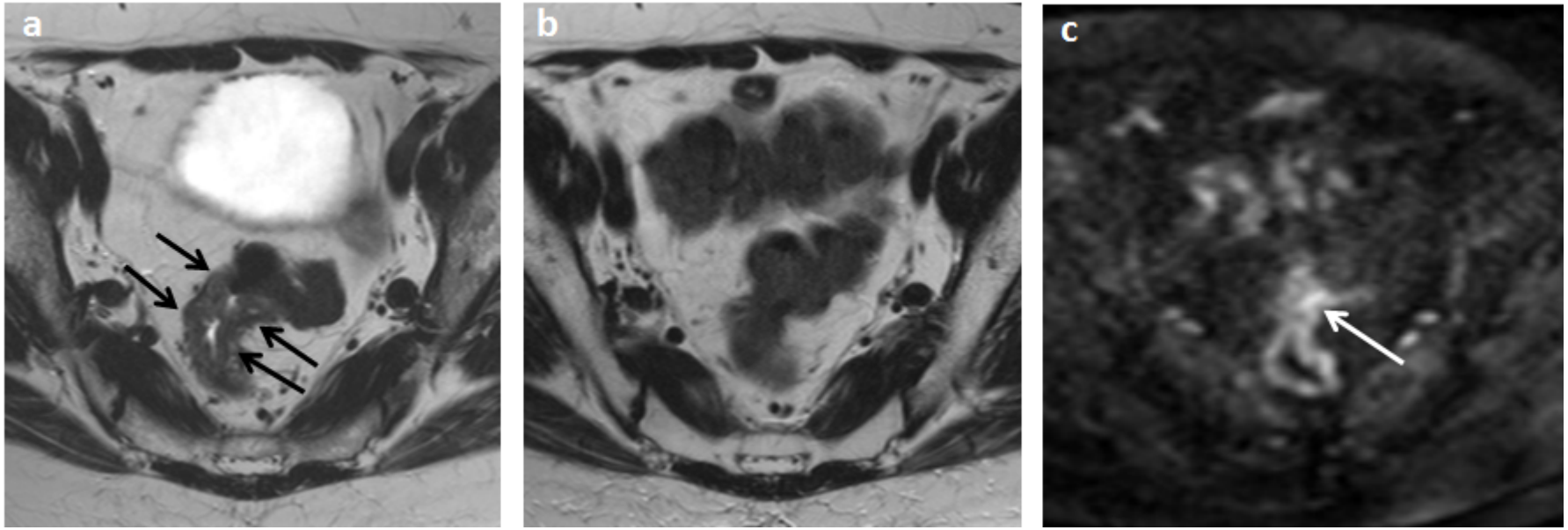

Figure 6. Example of a patient with a tumour in the upper third of the rectum at primary staging (arrows in a). After CRT the rectal wall is collapsed at the site of the primary tumour, making it difficult to establish whether or not a tumour remnant is still present. On corresponding b1000 DWI (c) some high signal is observed (arrow). This was, however caused by superposition of the two sides of the rectal wall: at histopathology this patient was confirmed to be a complete responder.