

# Novel MR techniques in rectal cancer

Citation for published version (APA):

van Heeswijk, M. M. (2017). *Novel MR techniques in rectal cancer: translating research into practice*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20171013mvh>

## Document status and date:

Published: 01/01/2017

## DOI:

[10.26481/dis.20171013mvh](https://doi.org/10.26481/dis.20171013mvh)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# **Novel MR techniques in rectal cancer**

- Translating research into practice -

**Miriam Maria van Heeswijk**

## **Novel MR techniques in rectal cancer Translating research into practice**

M.M. van Heeswijk

All rights reserved. No part of this publication may be reproduced, stored in a retrieval database or published in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise, without the prior written permission of the publisher.

ISBN: 978-94-92683-99-1

Cover Design: Richard Beerens | <http://tilt.amsterdam/>

Lay-out Design: Roy Sanders

Printed by: Optima Grafische Communicatie, Rotterdam, The Netherlands  
([www.ogc.nl](http://www.ogc.nl))

Printing of this thesis was financially supported by Philips, the Netherlands.

# **Novel MR techniques in rectal cancer**

- Translating research into practice -

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan Maastricht University  
op gezag van Rector Magnificus Prof. dr. R.M. Letschert,  
volgens het besluit van het College van Decanen,  
in het openbaar te verdedigen  
op 13 oktober 2017 om 14:00 uur

door

**Miriam Maria van Heeswijk**

Geboren op 28 mei 1989 te Waalwijk

**Promotores**

Prof. dr. R.G.H. Beets-Tan

Prof. dr. G.L. Beets

**Co-promotor**

Dr. D.M.J. Lambregts

**Beoordelingscommissie**

Prof. Dr. R.F.P.M. Kruitwagen (voorzitter)

Prof. dr. S.G.F. Robben

Prof. dr. L.P.S. Stassen

Prof. dr. U.A. van der Heide (Nederlands Kanker Instituut – Antoni van Leeuwenhoek, Amsterdam)

Dr. R.A. Dresen (UZ Leuven, België)



## CONTENTS

Chapter 1	General introduction	9
-----------	----------------------	---

### **PART I - Diffusion-weighted imaging for daily rectal staging**

Chapter 2	Measuring tumour height and length on MRI – effect of using diffusion-weighted MRI instead of T2-Weighted MRI.	19
-----------	--	----

Chapter 3	DWI for assessment of rectal cancer nodes after CRT: is the absence of nodes on DWI proof of a negative nodal status?	33
-----------	---	----

Chapter 4	Diffusion-weighted MRI to assess response to chemoradiotherapy in rectal cancer: main interpretation pitfalls and their use for teaching.	51
-----------	---	----

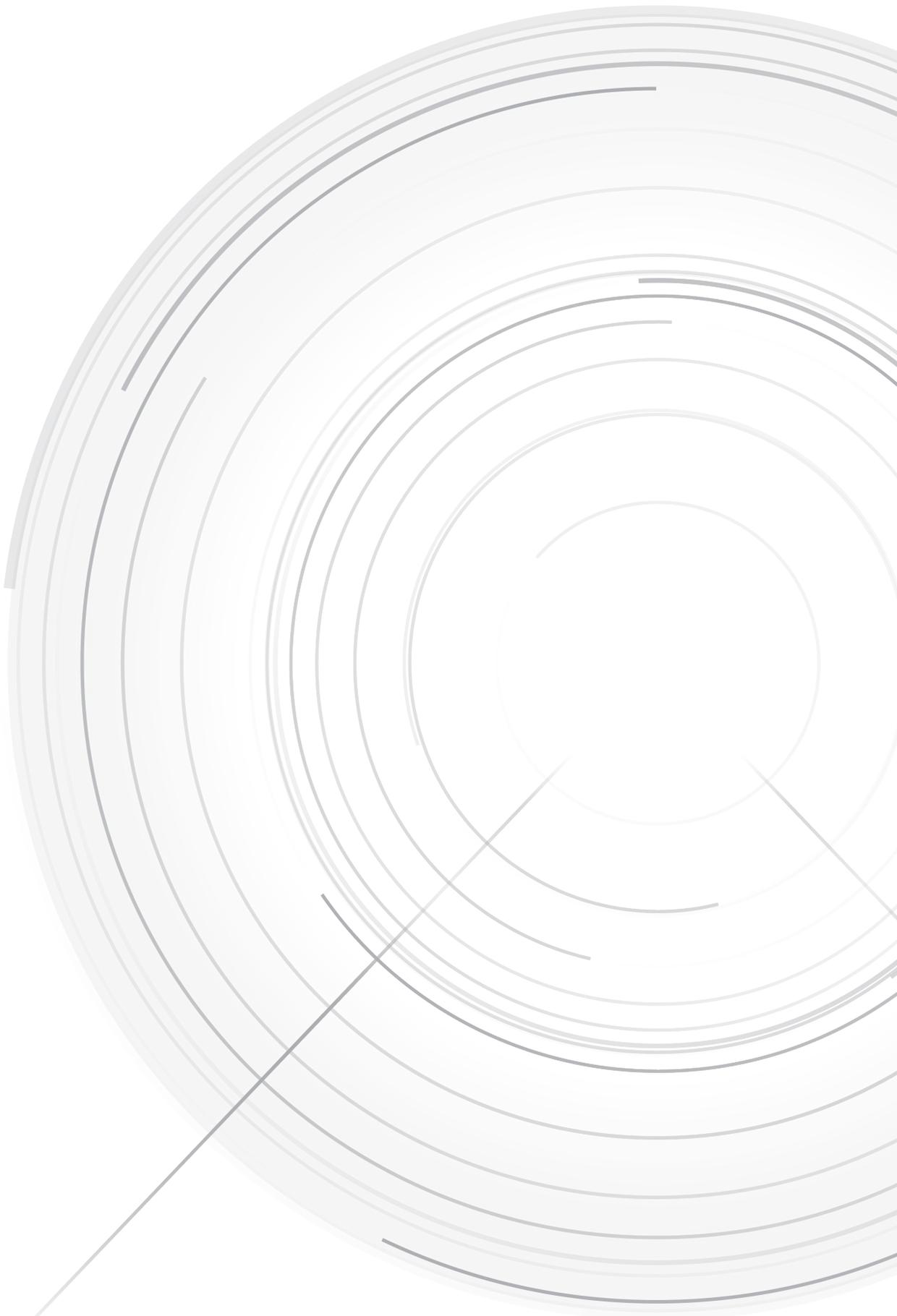
### **PART II - Quantitative analysis of rectal cancer imaging – a practical approach**

Chapter 5	Prospective multicenter validation study of MRI volumetry for response assessment after preoperative chemoradiation in rectal cancer: Can the results in literature be reproduced?	71
-----------	--	----

Chapter 6	Automated and semi-automated segmentation of rectal tumour volumes on diffusion-weighted MRI: can it replace manual volumetry?	89
-----------	--	----

Chapter 7	Measuring the apparent diffusion coefficient in primary rectal tumours – is there a benefit in performing histogram analyses?	107
-----------	---	-----

Chapter 8	General Discussion	129
Chapter 9	Valorisation	141
Chapter 10	Summary	147
	Samenvatting	153
	List of publications	159
	Dankwoord	163
	Curriculum Vitae	169





# Chapter 1

General Introduction



## ROLE OF IMAGING IN RECTAL CANCER TREATMENT

In 2015, over 4000 patients were diagnosed with rectal cancer in the Netherlands<sup>1</sup>. Imaging plays a fundamental role in assessing the extent (stage) of rectal cancer in order to determine the most appropriate treatment for each patient. Magnetic Resonance Imaging (MRI) is one of the most commonly used tools to stage rectal tumours. Based on the MRI assessment of risk factors, such as the local tumour extent (T-stage), the presence of nodal metastases (N-stage) and whether or not the tumour invades the anticipated surgical resection margin, tumours are classified as low risk (early), intermediate risk or high risk (locally advanced). According to current (Dutch) guidelines, early tumours can safely be treated with surgery only, while the intermediate and locally advanced tumour receive neoadjuvant treatment in the form of a short course of preoperative radiotherapy or a long course of chemoradiotherapy (CRT) followed by surgical resection<sup>1</sup>. The goal of long course neoadjuvant treatment is to downsize and downstage the tumour in order to increase the chance of complete surgical resection afterwards, thereby reducing the risk for local recurrence<sup>2</sup>. Until recently, the standard surgical treatment has been a total mesorectal excision (TME) which entails complete surgical removal of the mesorectal compartment using a standardized resection plane. Nowadays, however, we are witnessing a paradigm shift towards less invasive treatments for patients with very early tumours or patients that show a very good response after neoadjuvant treatment. Patients who primarily present with very small tumours may be managed with a local excision of the tumour (e.g. by transanal endoscopic microsurgery or 'TEM')<sup>3</sup>. Local excision may also be an option for patients who respond well to neoadjuvant treatment with only a small tumour remnant left after CRT<sup>4, 5</sup>. In patients in whom the tumour has completely disappeared after CRT, a watchful waiting approach without any surgery may be considered<sup>6, 7</sup>. Given the recent developments in these patient-tailored organ-preserving treatments, the role of imaging is now also evolving, pressing the need for an accurate response evaluation with imaging after CRT.

Overall, MRI has proven to be a powerful technique for rectal cancer staging and restaging, yet still has some constraints. The most significant limitations of MRI are the limited performance for lymph node staging and difficulties in differentiating between fibrosis and residual tumour after neoadjuvant treatment. These limitations restrict the use of MRI for response evaluation and thus to select patients for local excision or watchful waiting. These limitations are part of the reason that in recent years new imaging techniques and post processing methods have been developed and researched.

## NOVEL IMAGING AND IMAGE QUANTIFICATION METHODS

Since visual (qualitative) interpretation of MR images has some limitations, researchers are exploring more objective, quantitative methods of analysing rectal MRI in order to improve staging outcomes. A method that has been extensively investigated is the use of tumour size measurements. Several studies have shown that tumour size measurements and changes in tumours size after treatment may be valuable parameters to help assess tumour response<sup>8-12</sup>. However, based on current evidence there are no clear recommendations on how to perform such measurements, which are the most appropriate thresholds and how these may impact clinical practice.

Apart from these quantitative size measurements mainly focusing on tumour morphology on standard T2-weighted (T2W) sequences, another important focus of research has been the use of 'functional' imaging. Functional imaging refers to the assessment of biological tumour properties such as tumour cellularity or perfusion using imaging sequences designed specifically for this purpose. One of the most widely investigated functional imaging techniques is diffusion-weighted imaging (DWI). The image contrast in DWI is based on differences between normal and pathologic tissues due to the difference in the extracellular movement ('diffusion') of water protons between tissues<sup>13</sup>. In tissues with a normal cellularity water protons can diffuse freely, which causes a decrease of the MR signal. Conversely in tissues with increased cellularity (malignant tumour) the movement of water protons is restricted and the MR signal remains high. As a result DWI is a very sensitive to detect tumours, as it offers an excellent lesion-to-background contrast<sup>14</sup>. The addition of DWI to standard T2W sequences has shown to significantly improve the diagnostic performance of MRI<sup>15</sup>. For this purpose, the use of DWI is currently implemented in the restaging protocol as recommended in the expert consensus guidelines on rectal MR imaging from the European Society of Gastrointestinal and Abdominal Imaging<sup>16</sup>.

Other potential applications of DWI include primary tumour staging and assessment of lymph nodes. However, for these indications far less evidence is available and studies that have been published so far showed inconsistent results. Thus, further research is needed to establish the potential clinical value of DWI in these settings.

The general applicability of DWI needs to be studied and tested issue before it can be incorporated in the clinical imaging protocols. Many of the reports on DWI were based on studies performed in expert centres with readers who have many years of experience in reading abdominal or even specifically rectal MRI. This rai-

ses the question whether these results can also be translated to everyday clinical practice and reproduced by general radiologists.

Apart from visual interpretation of diffusion-weighted images, promising results have also been reported for quantitative volumetric assessment of rectal tumours on DWI<sup>17,18</sup> and for measuring the 'apparent diffusion coefficient' (ADC) of rectal tumours<sup>19,20</sup> and lymph nodes<sup>21,22</sup>. The ADC is the main quantitative measure of DWI and provides a measure of the tumour's cellular structure. An important issue with quantitative DWI evaluation (particularly ADC) is that the results that have been published so far are highly variable and need to be validated. Moreover, in order to acquire quantitative measures (volumes or ADC) from DWI, accurate tumour segmentations are needed which require detailed manual input from radiologists. This makes quantitative assessment of DWI a time-consuming process, which greatly hampers its potential implementation into clinical practice.

The overall goal of this thesis is to address the issues described above in order to help translate the use of novel imaging techniques and image quantification methods from research into daily practice with the ultimate aim to further improve the performance and clinical value of MRI for the staging and restaging of rectal cancer.

## SPECIFIC AIMS OF THIS THESIS

The first aim of this thesis is to investigate practical ways in which DWI may be used in daily clinical rectal cancer staging and how to teach non-expert (general) readers to interpret diffusion-weighted images. The second aim is to shed a light on the current evidence for quantitative tumour size measurements to assess treatment response in rectal cancer. Finally, we aim to search for ways to facilitate quantitative analysis of DWI in order to make it better applicable to daily practice.

## OUTLINE OF THIS THESIS

### **Part I - Diffusion-weighted imaging for daily rectal staging**

**Chapter 2** investigates the effect of using DWI as an alternative to standard T2-weighted MRI to determine rectal tumour height and length.

**Chapter 3** addresses the question whether visual interpretation of DWI can help to accurately diagnose patients in whom all metastatic lymph nodes have sterilised after chemoradiotherapy.

**Chapter 4** focuses on the effect of expert feedback and teaching using pitfalls to optimise the diagnostic performance of non-expert readers to assess the response of rectal tumours on restaging MRI after CRT.

## **Part II – Quantitative analysis of rectal cancer imaging– a practical approach**

**Chapter 5** reviews all available literature on tumour size measurements to assess treatment response in rectal cancer, including a prospective validation of previously reported thresholds in an independent multicentre cohort to establish their potential clinical value.

**Chapter 6** investigates the use of automated and semi-automated software methods to acquire volumetric tumour measurements on DWI as an alternative to manual tumour delineation in order to save time and make it more suitable for daily practice.

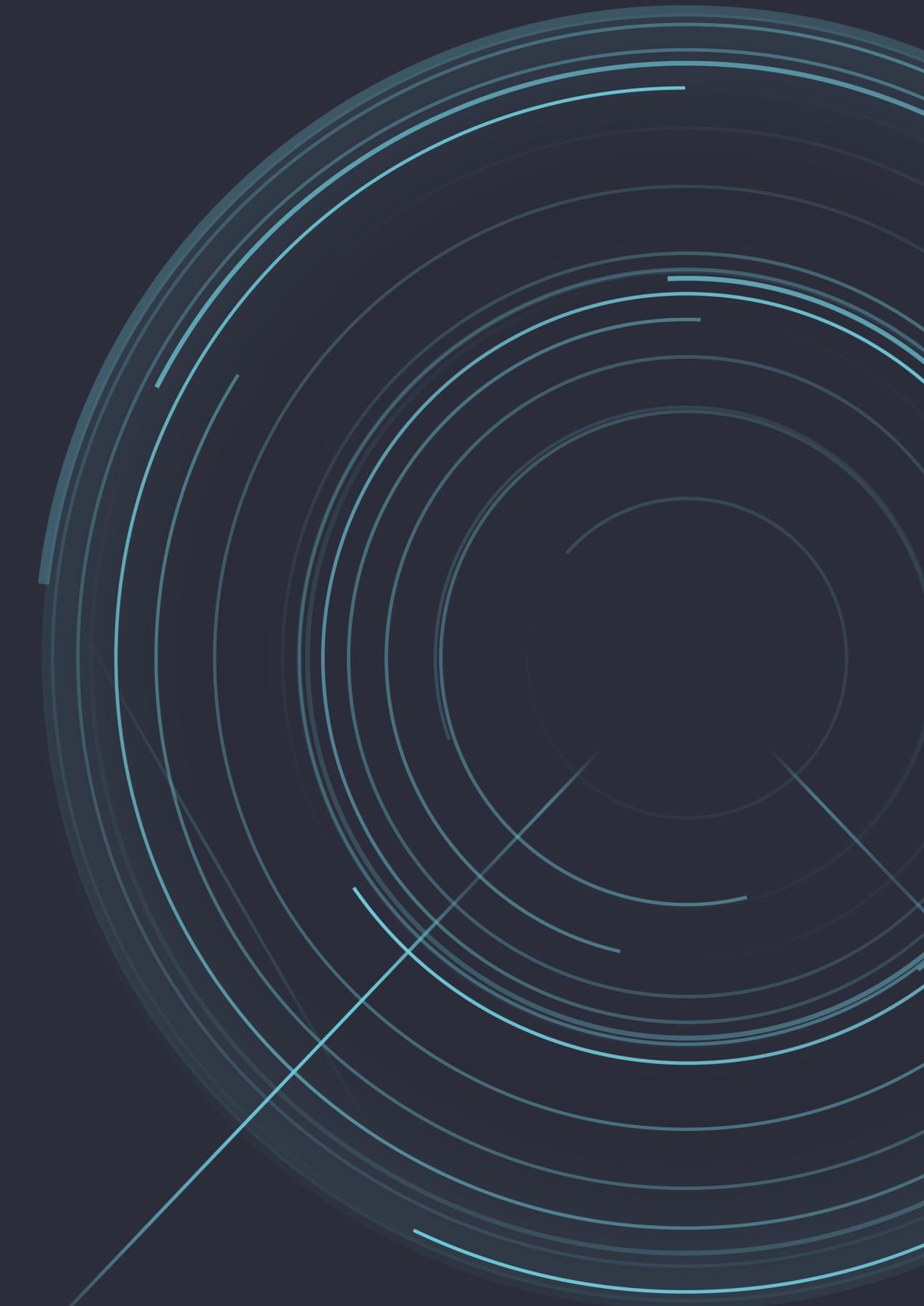
**Chapter 7** studies an alternative method to increase time efficacy in tumour delineation on DWI by combining non-precise tumour delineation with histogram post-processing to calculate tumour ADCs. Moreover, the prognostic value of ADC histogram analysis is investigated.



## REFERENCES

1. Landelijke Werkgroep Gastro-Intestinale tumoren. Landelijke richtlijn Colorectaal carcinoom (versie 3.0). [www.woncoline.nl](http://www.woncoline.nl). 2014.
2. Sauer R, Becker H, Hohenberger W, Rodel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med*. 2004;351(17):1731-40.
3. Tytherleigh MG, Warren BF, Mortensen NJ. Management of early rectal cancer. *Br J Surg*. 2008;95(4):409-23.
4. 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.
5. Verseveld M, de Graaf EJ, Verhoef C, van Meerten E, Punt CJ, de Hingh IH, et al. Chemoradiation therapy for rectal cancer in the distal rectum followed by organ-sparing transanal endoscopic microsurgery (CARTS study). *Br J Surg*. 2015;102(7):853-60.
6. Habr-Gama A, Perez RO, Nadalin W, Sabbaga J, Ribeiro U, Jr., Silva e Sousa AH, Jr., et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg*. 2004;240(4):711-7; discussion 7-8.
7. 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.
8. Kim YH, Kim DY, Kim TH, Jung KH, Chang HJ, Jeong SY, et al. Usefulness of magnetic resonance volumetric evaluation in predicting response to preoperative concurrent chemoradiotherapy in patients with resectable rectal cancer. *Int J Radiat Oncol Biol Phys*. 2005;62(3):761-8.
9. Dresen RC, Beets GL, Rutten HJ, Engelen SM, Lahaye MJ, Vliegen RF, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part I. Are we able to predict tumor confined to the rectal wall? *Radiology*. 2009;252(1):71-80.
10. Barbaro B, Fiorucci C, Tebala C, Valentini V, Gambacorta MA, Vecchio FM, et al. Locally advanced rectal cancer: MR imaging in prediction of response after preoperative chemotherapy and radiation therapy. *Radiology*. 2009;250(3):730-9.
11. Aiba T, Uehara K, Nihashi T, Tsuzuki T, Yatsuya H, Yoshioka Y, et al. MRI and FDG-PET for assessment of response to neoadjuvant chemotherapy in locally advanced rectal cancer. *Ann Surg Oncol*. 2014;21(6):1801-8.
12. Seierstad T, Hole KH, Groholt KK, Dueland S, Ree AH, Flatmark K, et al. MRI volumetry for prediction of tumour response to neoadjuvant chemotherapy followed by chemoradiotherapy in locally advanced rectal cancer. *Br J Radiol*. 2015;88(1051):20150097.
13. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*. 1986;161(2):401-7.

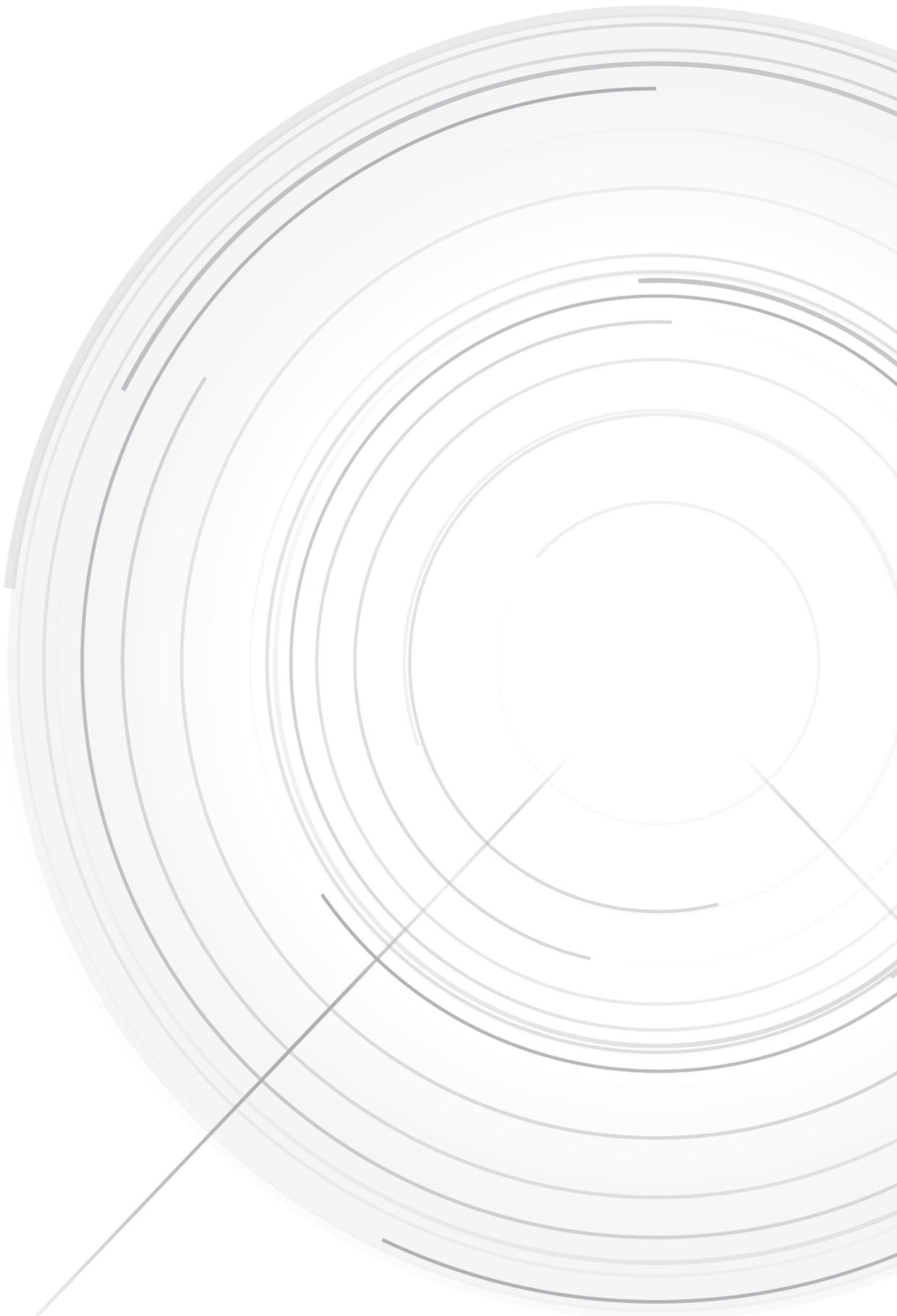
14. Nasu K, Kuroki Y, Kuroki S, Murakami K, Nawano S, Moriyama N. Diffusion-weighted single shot echo planar imaging of colorectal cancer using a sensitivity-encoding technique. *Jpn J Clin Oncol.* 2004;34(10):620-6.
15. van der Paardt MP, Zagers MB, Beets-Tan RG, Stoker J, Bipat S. Patients who undergo 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.
16. Beets-Tan RG, Lambregts DM, Maas M, Bipat S, Barbaro B, Caseiro-Alves F, et al. 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.* 2013;23(9):2522-31.
17. 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(12):3345-53.
18. Lambregts DM, Rao SX, Sassen S, Martens MH, Heijnen LA, Buijsen J, et al. MRI and Diffusion-weighted MRI Volumetry for Identification of Complete Tumor Responders After Preoperative Chemoradiotherapy in Patients With Rectal Cancer: A Bi-institutional Validation Study. *Ann Surg.* 2015;262(6):1034-9.
19. Curvo-Semedo L, Lambregts DM, Maas M, Beets GL, Caseiro-Alves F, Beets-Tan RG. Diffusion-weighted MRI in rectal cancer: apparent diffusion coefficient as a potential noninvasive marker of tumor aggressiveness. *J Magn Reson Imaging.* 2012;35(6):1365-71.
20. Ippolito D, Monguzzi L, Guerra L, Deponti E, Gardani G, Messa C, et al. Response to neoadjuvant therapy in locally advanced rectal cancer: assessment with diffusion-weighted MR imaging and 18FDG PET/CT. *Abdom Imaging.* 2012;37(6):1032-40.
21. Yasui O, Sato M, Kamada A. Diffusion-weighted imaging in the detection of lymph node metastasis in colorectal cancer. *Tohoku J Exp Med.* 2009;218(3):177-83.
22. Kim SH, Ryu KH, Yoon JH, Lee Y, Paik JH, Kim SJ, et al. Apparent diffusion coefficient for lymph node characterization after chemoradiation therapy for locally advanced rectal cancer. *Acta Radiol.* 2015;56(12):1446-53.





# PART I

Diffusion-weighted imaging  
for daily rectal staging





# Chapter 2

Measuring rectal tumor height and length on MRI - effect of using diffusion-weighted MRI instead of T2-weighted MRI

MM van Heeswijk, DMJ Lambregts, L den Ouden, MC Ageitos-Casais,  
M Maas, GL Beets, FCH Bakers, RGH Beets-Tan  
*Submitted for publication*

## ABSTRACT

### Aim

Endoscopy and MRI are the primary tools used to assess tumour height (distance from anorectal junction) and tumour length. Diffusion-weighted MRI is known to improve tumour conspicuity and may therefore influence assessment of tumour height and length on MRI. The aim was to compare diffusion-weighted MRI to standard (T2-weighted) MRI for assessment of rectal tumour height and length, both at primary staging and for restaging after chemoradiotherapy.

### Material & Methods

Seventy-two patients were included: 24 underwent only primary staging MRI, 48 underwent a second restaging-MRI after chemoradiotherapy. All MRIs (1.5T) included standard T2-weighted and diffusion-weighted (highest b-value b1000) sequences. Two readers (R1-experienced, R2-unexperienced) independently measured tumour length and height: [1] on T2-weighted MRI and [2] on b1000 diffusion-MRI. Height/length measurements were compared between T2-weighted MRI and diffusion-MRI and interobserver agreement (intraclass correlation coefficient) was calculated.

### Results

Median tumour length (R1/R2) was 44/44 mm on T2-weighted MRI versus 43/40 mm on diffusion-MRI at primary staging ( $P=0.17$  (R1) and  $P=0.001$  (R2)) and 30/25 mm on T2-weighted MRI versus 21/20 mm on diffusion-MRI for restaging post-chemoradiotherapy ( $P=0.01$  (R1) and  $P<0.001$  (R2)). Tumour height was 51/47 mm on T2-weighted MRI versus 52/57 mm on diffusion-MRI at primary staging ( $P=0.12$  (R1) and  $P=0.01$  (R2)) and 50/50 mm on T2-weighted MRI versus 51/55 mm on diffusion-MRI for restaging post-chemoradiotherapy ( $P=0.51$  (R1) and  $P<0.001$  (R2)). Interobserver agreement at primary staging was excellent both for height/length measurements (ICC 0.92/0.85 for T2-weighted and 0.93/0.87 for diffusion-MRI); for restaging after CRT, ICCs for height/length were 0.95/0.72 (T2-weighted MRI) and 0.86/0.48 (diffusion-MRI).

### Conclusion

Particularly when assessed by a reader with limited MR experience, rectal tumour length is significantly smaller and as a result tumour height is larger when measured on diffusion-MRI instead of standard T2-weighted MRI.

## INTRODUCTION

Magnetic Resonance Imaging (MRI) is part of the standard staging work-up of rectal cancer. As part of tumour staging, rectal tumour height and length are determined, which – apart from endoscopic assessment – is also increasingly done using MRI<sup>1</sup>. The tumour height, typically assessed on MRI as the distance from the anorectal junction to the lower pole of the tumour, is an important feature in rectal cancer staging for several reasons. The surgical approach of distal tumours depends on its distance from the anal sphincter complex, based on which the surgeon will determine whether the sphincter can be preserved and if it is possible to create an anastomosis<sup>2,3</sup>. The distance from the anorectal junction also (indirectly) influences the decision if neoadjuvant therapy is necessary, as lower tumours have a higher risk of mesorectal fascia (MRF) involvement due to the tapering of the mesorectum in the lower pelvis and the relatively small amount of mesorectal fat tissue around the distal rectum<sup>2,4</sup>. The tumour length itself does not have such a direct effect on treatment planning, but is mainly important as a measure to indicate the tumour size, which is known to be an indicator for prognostic outcome<sup>5</sup>. Larger tumours tend to show a poorer response to neoadjuvant treatment<sup>6-9</sup>. The one-dimensional measurement of tumour length is one of the easiest and fastest ways to get an impression of the tumour size and is typically performed on morphological T2-weighted (T2W) imaging sequences. However, there are some drawbacks to this approach. Buijsen et al compared one-dimensional tumour length measurements on T2-weighted MRI to the tumour length as established at pathology and found that MRI consistently overestimated the tumour length<sup>10</sup>. Moreover, a recent review paper by Martens et al. compared whole volume tumour measurements to one-dimensional (length) measurements to predict treatment response and found inferior results for the one-dimensional approach<sup>11</sup>. The overestimation of tumour on T2 may be due to partial volume effects or by the incorporation of the normal (edematous) rectal wall in the measurements<sup>12</sup>. Moreover, in small tumours it can be difficult to define the tumour boundaries, and fecal impaction may hamper image interpretation. Diffusion-weighted imaging (DWI) is increasingly incorporated in standard rectal MRI protocols because it offers an excellent lesion-to-background contrast and therefore is a very sensitive technique to detect malignant tumour<sup>12</sup>. The addition of DWI can increase tumour conspicuity and thereby help in determining the tumour boundaries.

Aim of this study was to test this hypothesis and compare DWI to standard T2-weighted MRI for the assessment of tumour height and tumour length in rectal cancer patients, both at primary staging and for restaging of tumours after CRT.

## MATERIAL & METHODS

### Patients

Ninety-seven patients diagnosed with rectal cancer between April 2009 and October 2011 at Maastricht University Medical Centre were considered for inclusion in this retrospective study. All patients underwent MRI as part of their diagnostic workup, with a DWI sequence being part of the routine imaging protocol. The study was approved by the local institutional review board. Due to the retrospective nature of the study informed consent was waived. Inclusion criteria were (a) biopsy proven rectal adenocarcinoma, and (b) availability of a primary staging MRI (and restaging MRI in case of neoadjuvant chemoradiotherapy) including a DWI sequence according to a standardized protocol. Twenty-five patients were excluded for the following reasons: (a) the tumour had already been excised with transanal endoscopic microsurgery (TEM) or polypectomy at the time of the MRI ( $n=19$ ), and (b) missing data due to referral by/to another hospital ( $n=6$ ). This left a total study population of 72 patients. Twenty-four patients underwent only a primary staging MRI followed by (a short course of 5x5 Gy radiation and) immediate surgery, the remaining 48 patients underwent a long course of neoadjuvant chemoradiotherapy (CRT) followed by a second restaging MRI (after 8-10 weeks), adding to a total of 72 primary staging examinations and 48 restaging examinations.

### MR Imaging

MR imaging was performed at 1.5T (Intera or Intera Achieva MR system; Philips Medical Systems, Best, The Netherlands) using a phased array coil. Patients did not receive any bowel preparation. In case of anticipated bowel movement artefacts on the planning scan, patients received 20 mg of scopolamine butylbromide (Buscopan, Boehringer Ingelheim, Germany) intravenously. The MR protocol included standard 2D T2-weighted (T2W) fast-spin echo (FSE) sequences in sagittal, transverse and coronal planes (TR/TE 8456/130 msec,  $90^\circ$  flip angle, 25 echotrain length, 6 NSA,  $0.78 \times 1.14 \times 3.00$  mm acquisition voxel size, 22 slices, 5.08 min acquisition time) and an axial diffusion-weighted sequence with background body signal suppression (b-values: 0, (100,) 500, 1000  $s/mm^2$ , TR/TE 4829/70 msec, echo-planar imaging factor 53, 4 number of signal averages,  $2.50 \times 3.11\text{-}3.18 \times 5.00$  mm acquisition voxel size, 42-50 slices, 10.37 minutes acquisition time). The transverse T2W and DWI sequences were angled in identical planes, perpendicular to the tumour axis as identified on the sagittal planning scan. The coronal images were angled parallel to the tumour axis (or parallel to the anal canal in case of very distal tumours).

## Image Evaluation

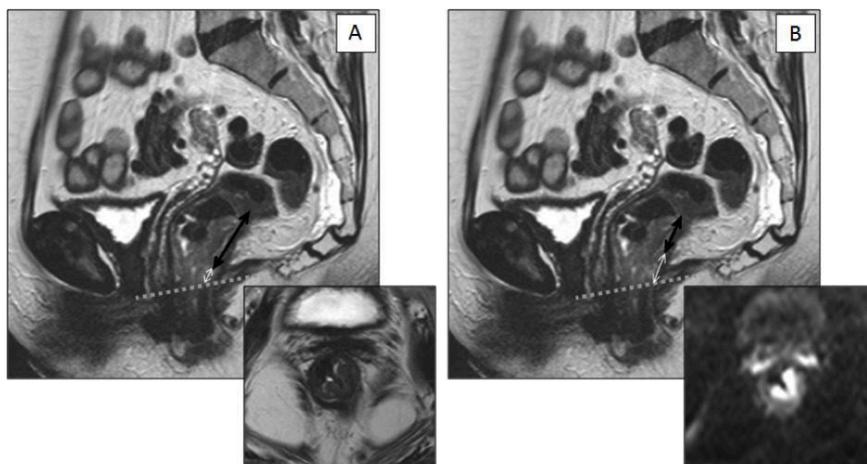
Two independent readers (R1 D-L, experienced in reading rectal MRI; and R2 LDO with no specific expertise in reading rectal MRI) evaluated the MRI scans to measure the tumour length and height for each patient, both on the primary staging MRIs as well as on the restaging MRIs when available. Examples illustrating the measurement methodology are given in **Figures 1 (primary staging) and 2 (restaging)**: tumour length was defined as the distance (in mm) between the distal and proximal margin of the tumour, tumour height was determined as the distance (in mm) between the anorectal junction and the distal tumour margin. Measurements were performed in twofold (with a two week interval between the two sessions to avoid recall bias);

(a) based on the T2-weighted MRI: the tumour boundaries were determined by scrolling through the transverse T2-weighted images to identify the first slice (distal tumour margin) and last slice (proximal tumour margin) containing tumour. These distal and proximal margins were projected as reference points on the sagittal T2-weighted planning scan (using reference lines) to measure the tumour length and height.

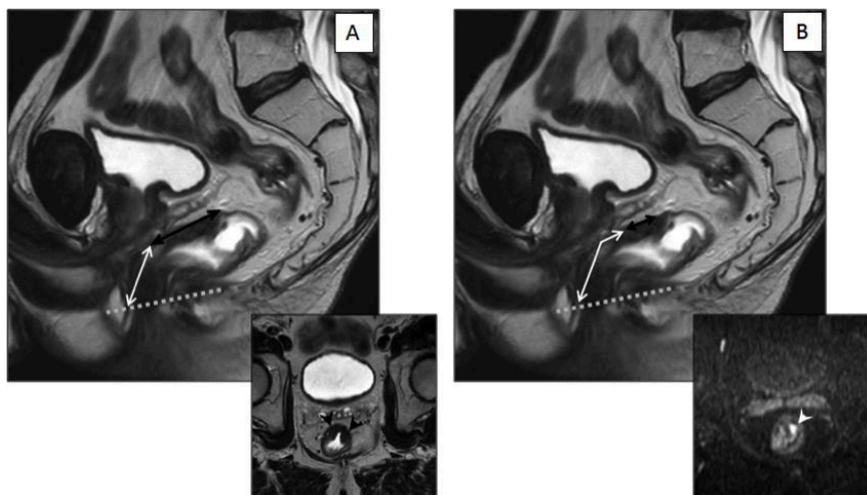
(b) based on DWI: the tumour boundaries were determined in a similar way by scrolling through the transverse high b-value (b1000) diffusion-weighted to identify the distal and proximal tumour margins. These were projected as reference points on the sagittal T2-weighted planning scan to measure tumour height and length.

In both sessions the transverse images were thus used as the main input to determine the tumour boundaries. The sagittal planning images were merely used as an anatomical reference to project the tumour boundaries and to visualize the anorectal junction as a landmark to determine the tumour height. When a reader could not identify any residual tumour on the restaging MRI (and diagnosed the patient as a complete responder), the tumour length was indicated as 0 mm and the patient was excluded from the tumour height measurements.

Statistical analysis Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS, version 23.0, Inc., Chicago, IL). The intraclass correlation coefficient (ICC) for single measures was calculated (using a 2-way mixed model with absolute agreement) to determine the interobserver agreement between the 2 readers for both measurement methods (T2-weighted and DWI) with 0 to 0.20 indicating poor agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 good agreement, and 0.81 to 1.0 excellent agreement. Data was tested for normality with the Shapiro-Wilk test for Normality. Measu-



**Figure 1** Example of height and length measurements based on the primary staging T2W-MRI (A) and high b-value (b1000) DWI (B) images of a 71-year-old female patient with a tumour in the distal rectum. The white arrows indicate the tumour height, measured as the distance from the anorectal junction (3 mm in A; 8 mm in B), the black arrows indicate the tumour length (41 mm in A; 22 mm in B). The dotted line represents the anorectal junction, which was the main landmark to determine the tumour height.



**Figure 2** Example of height and length measurements based on the restaging T2W-MRI (A) and high b-value (b1000) DWI (B) images of a 69-year-old male patient after chemoradiotherapy. On T2-weighted MRI, the fibrotic wall thickening in the anterior rectal wall (black arrowheads in A) was included in the length measurement as suspicious for residual tumour; on b1000 DWI only a small tumour remnant was visualised within the fibrosis (white arrowhead in B) and included in the measurement. The white arrows indicate the tumour height (11 mm in A; 16 mm B), the black arrows indicate the tumour length (14 mm in A; 7 mm in B). The dotted line grey represents the anorectal junction.

rements were compared between the two measurement methods with the Wilcoxon Signed Rank Test in case of non-normally distributed data and with a paired t-test in case of normally distributed data. All analyses were performed separately for the primary staging scans (n=72) and the restaging scans (n=48). P values <0.05 were considered statistically significant.

## RESULTS

### Patients

Baseline patient characteristics are given in **Table 1**. Forty-nine patients were male and 23 were female, median age was 69 years (range 45-86).

**Table 1.** Patient characteristics

Patient Characteristics	Number
<b>Sex</b>	
Male	49
Female	23
Median age at diagnosis (range)	69 (45-86)
<b>cT-stage at diagnosis</b>	
cT1-2	25
cT3	42
cT4	5
<b>cN-stage at diagnosis</b>	
cN0	20
cN1	27
cN2	25
<b>Neoadjuvant treatment</b>	
None	8
Short course (5x5Gy)	16
Long course (followed by restaging MRI)	48

### T2W-MRI versus DWI measurements

The height and length measurements for T2-weighted MRI and DWI are given in **Table 2** for the two readers. For Reader 2 the median tumour height was significantly larger when measured on DWI compared to T2W-MRI, both at primary staging (57 versus 47 mm, P=0.01) as well as for the restaging MRIs (55 versus 50

mm,  $P < 0.001$ ). Also, the tumour length was significantly smaller on DWI-MRI compared to T2W-MRI, both at primary staging (40 versus 44 mm,  $P = 0.001$ ) as well as for restaging after CRT (20 versus 25 mm,  $P < 0.001$ ). For Reader 1 there were no significant differences between T2W-MRI and DWI measurements, except for the tumour length on restaging MRI, which was significantly smaller on DWI compared to T2-weighted MRI (21 versus 30 mm,  $P = 0.01$ ).

**Table 2.** Median (range) tumour height and length measurements (in mm) for T2-weighted MRI and DWI per reader.

	Reader 1			Reader 2		
	T2W	DWI	P-value	T2W	DWI	P-value
<b>Primary staging (n=72)</b>						
Height	51 (0-105)	52 (0-116)	0.12	47 (0-143)	57 (0-165)	0.01
Length	44 (8-119)	43 (7-142)	0.17	44 (9-120)	40 (5-134)	0.001
<b>Restaging (N=48)</b>						
Height*	50 (0-133)	51 (0-122)	0.51	50 (0-138)	55 (7-155)	<0.001
Length	30 (5-84)	21 (0-88)	0.01	25 (0-84)	20 (0-73)	<0.001

\*Only patients in whom a residual tumour was recognized on both T2-weighted MRI and DWI were included (n=43 for reader 1 and n=37 for reader 2), patients with a suspected complete response on either T2W-MRI or DWI were excluded from the height measurements after CRT.

### Interobserver agreement

Intraclass correlations between the two readers for the T2W and DWI measurements, respectively, are given in **Table 3**. The interobserver agreement (IOA) on primary staging was excellent, both for assessment of tumour height (ICC 0.92 for T2W-MRI and 0.93 for DWI) and length (ICC 0.85 and 0.87). For the restaging MRIs, agreement for assessment of tumour height was excellent for both methods (ICC 0.95 for T2W-MRI and 0.86 for DWI). For assessment of tumour length on restaging MRI, agreement was good for T2W-MRI (ICC 0.72) and moderate for DWI (ICC 0.48).

**Table 3.** Interobserver agreement between the two readers expressed as the intraclass correlation coefficient (+ 95% confidence interval).

	T2W	DWI
<b>Primary staging</b>		
Height	0.92 (0.87-0.95)	0.93 (0.84-0.96)
Length	0.85 (0.77-0.90)	0.87 (0.80-0.92)
<b>Restaging*</b>		
Height	0.95 (0.91-0.97)	0.86 (0.72-0.93)
Length	0.72 (0.56-0.84)	0.48 (0.18-0.70)

\*ICCs for the restaging MRIs were only calculated for patients where both readers found a residual tumour, i.e. n=47 patients on T2W-MRI and n=35 patients on DWI. Patient in whom one or both readers diagnosed a complete tumour response were excluded.

## DISCUSSION

The aim of this study was to assess if the use of diffusion-weighted imaging as an alternative to standard morphological T2-weighted imaging affects the assessment of rectal tumour height and length on MRI. Our results show that – particularly in relatively unexperienced hands – the tumour length is consistently smaller and consequently the tumour height (distance from the anorectal junction) is larger when measured on DWI, both at primary staging as well as for restaging of tumour after CRT. For a reader with more experience in reading rectal MRI, differences were less obvious and only the tumour length after chemoradiotherapy was significantly smaller when measured on DWI instead of T2-weighted MRI. Apparently the effect of adding DWI is more pronounced for a reader with less experience in reading rectal MRI, for whom it may be more challenging to define the tumour boundaries on standard MRI, and moreover in the restaging setting when the tumour bed has undergone fibrotic changes as a result of radiation treatment. In these cases, it may be easier to determine the tumour boundaries on DWI since determining the extent of remaining high signal within the rectal wall is more straightforward than looking for residual tumour within fibrosis on morphological MRI, which is well known to be a difficult task for radiologists.

We acknowledge that the main limitation of our study is the lack of a true gold standard. Tumour length was not routinely reported at histopathology and could therefore not be used as a standard of reference. Moreover no true gold standard exists for assessing tumour height. Traditionally mainly endoscopy is adopted as a measurement method, but the results vary depending on the use of a flexible or rigid scope<sup>3</sup>. Nowadays, MRI is increasingly used because of the visualization of more consistent anatomical landmarks such as the anorectal junction that can be used as a reference point for height measurements. Due to the lack of a true uniform and valid reference standard we can only speculate about the implications of the difference we observed between the DWI and T2W-MRI based height and length measurements in our current report. Previous studies have suggested that DWI may be more accurate to detect tumour compared to T2-weighted MRI, particularly in the restaging setting after chemoradiotherapy where standard T2W-MRI is known to be of limited value to assess the degree of viable tumour within areas of post-radiation fibrosis<sup>13-15</sup>, typically leading to an overestimation of residual tumour on T2W-MRI<sup>14, 15</sup>. Moreover there are various previous reports that have compared DWI to T2-weighted MRI to assess tumour volume. These reports showed that the volume as measured on DWI is typically smaller than the volume as measured on T2W-MRI<sup>16, 17</sup>, which is in line with our finding that tumour length on DWI (particularly after CRT) is smaller when measured on DWI. These previous reports also showed that the DWI tumour volume after CRT correlates better with the actual amount of residual tumour at histopathology than the tumour volume derived from T2W-MRI<sup>16, 18, 19</sup>. These findings indirectly suggest that DWI (at least post-CRT) is likely superior in assessing the actual tumour volume and that T2W-MRI systematically overestimates the tumour load. This is also in line with previous findings by Buijsen et al. who compared T2W-MRI tumour length measurements to the tumour length at histopathology in patients who were immediately operated after primary staging. They found that tumour size was consistently overestimated on T2W-MRI<sup>10</sup>. Their study, however, did not include a comparison with DWI. Despite the lack of a gold standard in our current study, the above findings thus appear to support our hypothesis that measurements on DWI may be more reliable than those performed on T2W-MRI, although all together the differences between the two techniques observed in our study were small, ranging between 1-10 mm at primary staging and 1-9 mm for the restaging scans.

A remarkable finding is the relatively low interobserver agreement for the tumour length measurements on restaging DWI, especially compared to the good to excellent IOA for all other measurements. This is in contrast with the assumption

that the tumour boundaries would be easier to define on DWI than on T2W-MRI, which should have resulted in improved rather than decreased IOA. There are two possible explanations for this contradiction. First, the two readers had different experience levels and it has been reported that experience level influences the performance of DWI interpretation<sup>20</sup>. Second, although readers were instructed to primarily use the transverse T2W and DWI images to determine the tumour boundaries, the actual measurements were performed on the sagittal T2-weighted planning images. One can imagine that for such a 3-dimensional reading, it may have been more difficult to correlate the transverse diffusion-weighted images with the sagittal T2-weighted scan, given the difference in scan technique and the relative lack of anatomical reference on DWI. This methodology was chosen because DWI images were only acquired in transverse plane and therefore no sagittal DWI sequences were available to more accurately correlate the DWI findings.

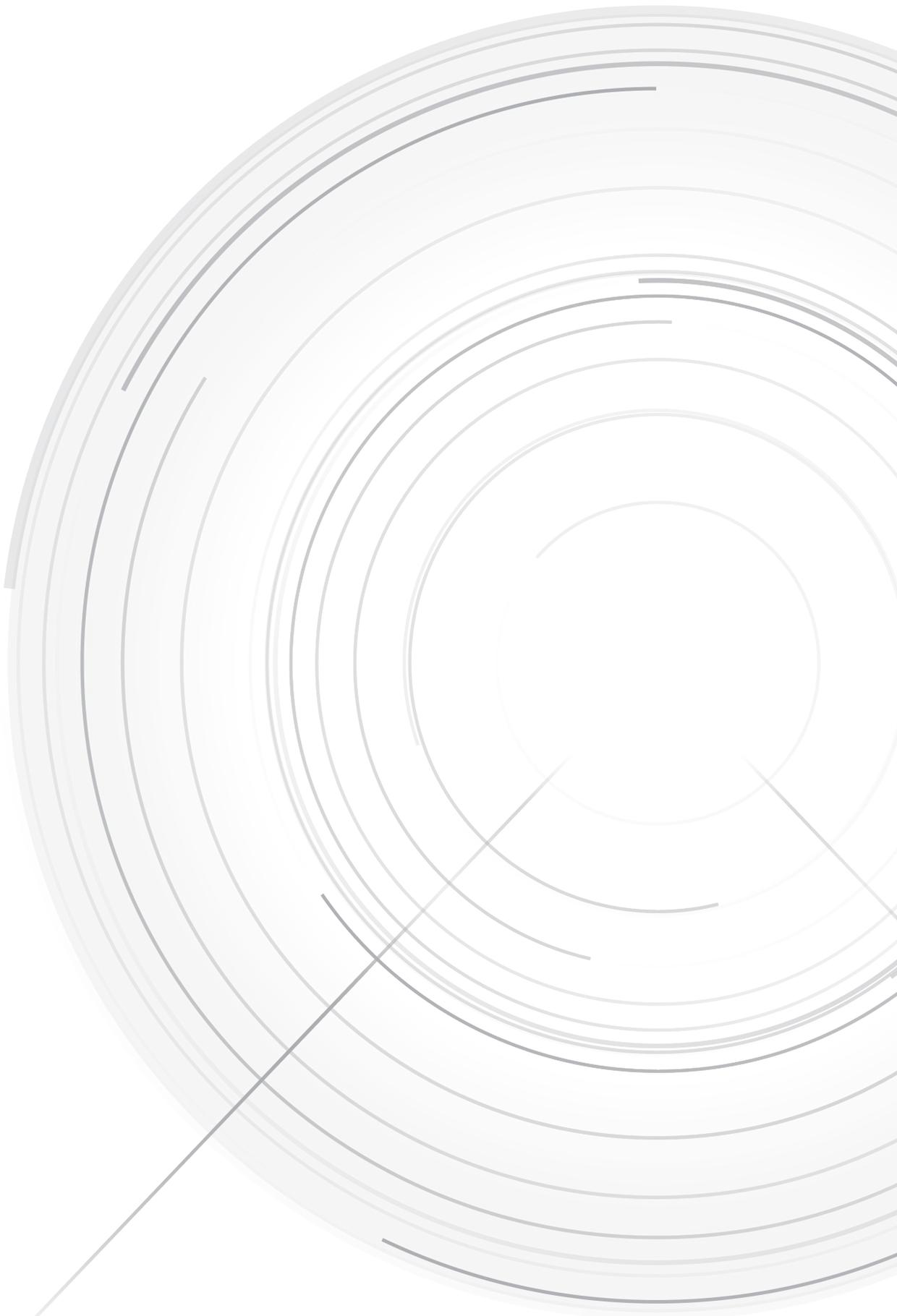
Apart from this issue, there were some limitations to our study design. The DWI sequence was acquired with a slice thickness of 5 mm, while the T2W images were acquired with 3 mm slices, which creates a margin of error when using the axial images to determine the tumour boundaries. Furthermore, the study was performed retrospectively and with a relatively small number of patients. Finally, the main limitation – as described above - was the lack of a valid standard of reference.

In conclusion, our results suggest that use of DWI may be helpful in determining rectal tumour boundaries on MRI. Particularly after chemoradiotherapy and in less experienced hands the use of diffusion-weighted MRI instead of standard T2-weighted MRI results in a reduction in measured tumour length and consequently in an increased distance from the anorectal junction.

## REFERENCES

1. Beets-Tan RG, Lambregts DM, Maas M, Bipat S, Barbaro B, Caseiro-Alves F, et al. 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.* 2013;23(9):2522-31.
2. Glimelius B, Pahlman L, Cervantes A, Group EGW. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2010;21 Suppl 5:v82-6.
3. van de Velde CJ, Boelens PG, Borrás JM, Coebergh JW, Cervantes A, Blomqvist L, et al. EURECCA colorectal: multidisciplinary management: European consensus conference colon & rectum. *Eur J Cancer.* 2014;50(1):1 e- e34.
4. Taylor FG, Quirke P, Heald RJ, Moran BJ, Blomqvist L, Swift IR, et al. Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the MERCURY study. *J Clin Oncol.* 2014;32(1):34-43.
5. Kim YH, Kim DY, Kim TH, Jung KH, Chang HJ, Jeong SY, et al. Usefulness of magnetic resonance volumetric evaluation in predicting response to preoperative concurrent chemoradiotherapy in patients with resectable rectal cancer. *Int J Radiat Oncol Biol Phys.* 2005;62(3):761-8.
6. Dresen RC, Beets GL, Rutten HJ, Engelen SM, Lahaye MJ, Vliegen RF, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part I. Are we able to predict tumour confined to the rectal wall? *Radiology.* 2009;252(1):71-80.
7. Kim NK, Baik SH, Min BS, Pyo HR, Choi YJ, Kim H, et al. A comparative study of volumetric analysis, histopathologic downstaging, and tumour regression grade in evaluating tumour response in locally advanced rectal cancer following preoperative chemoradiation. *Int J Radiat Oncol Biol Phys.* 2007;67(1):204-10.
8. Aiba T, Uehara K, Nishashi T, Tsuzuki T, Yatsuya H, Yoshioka Y, et al. MRI and FDG-PET for assessment of response to neoadjuvant chemotherapy in locally advanced rectal cancer. *Ann Surg Oncol.* 2014;21(6):1801-8.
9. Seierstad T, Hole KH, Groholt KK, Dueland S, Ree AH, Flatmark K, et al. MRI volumetry for prediction of tumour response to neoadjuvant chemotherapy followed by chemoradiotherapy in locally advanced rectal cancer. *Br J Radiol.* 2015;88(1051):20150097.
10. Buijsen J, van den Bogaard J, Janssen MH, Bakers FC, Engelsman S, Ollers M, et al. FDG-PET provides the best correlation with the tumour specimen compared to MRI and CT in rectal cancer. *Radiother Oncol.* 2011;98(2):270-6.
11. Martens MH, van Heeswijk MM, van den Broek JJ, Rao SX, Vandecaveye V, Vliegen RA, et al. Prospective, Multicenter Validation Study of Magnetic Resonance Volumetry for Response Assessment After Preoperative Chemoradiation in Rectal Cancer: Can the Results in the Literature be Reproduced? *Int J Radiat Oncol Biol Phys.* 2015;93(5):1005-14.
12. Nasu K, Kuroki Y, Kuroki S, Murakami K, Nawano S, Moriyama N. Diffusion-weighted single shot echo planar imaging of colorectal cancer using a sensitivity-encoding technique. *Jpn J Clin Oncol.* 2004;34(10):620-6.

13. 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.
14. 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 tumour response to neoadjuvant chemo- and radiation therapy. *Radiology.* 2009;253(1):116-25.
15. 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.
16. Lambregts DM, Rao SX, Sassen S, Martens MH, Heijnen LA, Buijsen J, 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.
17. Carbone SF, Pirtoli L, Ricci V, Venezia D, Carfagno T, Lazzi S, et al. Assessment of response to chemoradiation therapy in rectal cancer using MR volumetry based on diffusion-weighted data sets: a preliminary report. *Radiol Med.* 2012;117(7):1112-24.
18. 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(12):3345-53.
19. Curvo-Semedo L, Lambregts DM, Maas M, Thywissen T, Mehsen RT, Lammering G, 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.
20. van Heeswijk MM LD, Andrade L, Maas M, Beets GL, Beets-Tan RGH. Clinical assessment of diffusion-weighted MRI for restaging rectal cancer: is there a learning curve? [abstract]. *Insights into imaging* 2016;7(suppl 1):S412.





# Chapter 3

DWI for assessment of rectal cancer nodes  
after chemoradiotherapy:  
is the absence of nodes at  
DWI proof of a negative nodal status?

MM van Heeswijk, DMJ Lambregts, WM Palm, BMF Hendriks,  
M Maas, GL Beets, RGH Beets-Tan

*American Journal of Roentgenology 2017;208:W79-W84*

## ABSTRACT

### Objectives

When considering organ preservation in rectal cancer patients with good tumour response, assessment of a node negative status after chemoradiation (CRT) is important. Diffusion-weighted MRI (DWI) is a very sensitive technique to detect nodes. The study aim was to test the hypothesis that absence of nodes on DWI after CRT is concordant with a ypN0-status.

### Methods

Retrospective study of 90 rectal cancer patients treated with CRT followed by re-tagging MRI(1.5T) including DWI (highest b-value b1000). Two independent readers counted the number of nodes visible in the mesorectal compartment on post-CRT DWI. The number of nodes on DWI (0 vs  $\geq 1$ ) was compared with the number of metastatic nodes at histopathology/follow-up (yN0 vs yN+).

### Results

At histopathology/follow-up 71 patients were yN0, 19 were yN+. In 10 patients no nodes were observed on DWI, which was concordant with a yN0-status in 100%. In the other 61 yN0 patients the median number of nodes detected on DWI was 3 (range 1-17). To differentiate between yN0/yN+ sensitivity was 100%, specificity 14%, PPV 24% and NPV 100%.

### Conclusion

Although the absence of nodes on DWI is not a frequent finding, it appears to be a reliable predictor of yN0-status after CRT in rectal cancer. DWI may thus be a helpful adjunct in assessing response after chemoradiotherapy and help select patients for organ-saving treatment.

## INTRODUCTION

Magnetic Resonance Imaging (MRI) is part of the standard work-up for patients with rectal cancer. MRI is increasingly adopted for the restaging of rectal cancer patients after neoadjuvant chemoradiotherapy (CRT). Restaging MRI is valuable to evaluate the response of a locally advanced tumour to CRT and to determine whether the extent of the surgical resection can be decreased in case of significant tumour regression. Furthermore, in case of a very good or complete tumour response, patients may – although presently mainly in trial settings – be considered for organ-preserving treatment options, i.e. a local excision (transanal endoscopic microsurgery) or a “watchful waiting policy”<sup>1-4</sup>. If organ preservation is considered, it is crucial to accurately select the (near) complete responders. A combination of MRI including diffusion-weighted imaging together with endoscopy and biopsies offers the best results to determine whether or not a tumour remnant remains present within the bowel wall<sup>5</sup>. The assessment of the response of the lymph nodes mainly relies on imaging. With organ-preservation, one needs to ensure that there is no residual nodal disease as this induces a risk for a nodal recurrence<sup>6,7</sup>. It is known that standard morphological MRI is of limited use for assessing lymph nodes<sup>8,9</sup>, although more favourable results have been suggested for restaging of nodes after CRT compared with results for primary nodal staging<sup>8-11</sup>. Sensitivities reported for detection of nodal disease after CRT range between 42-89%<sup>12-15</sup>. Particularly in patients with a good tumour response (ypT0-2) that may be considered for organ-preservation low sensitivities of around 40% have been reported<sup>15</sup>, probably owing to the fact that in these cases residual disease typically occurs within very small nodes that are difficult to evaluate with imaging. This still limits the clinical utility of imaging for the prediction of the pathologic nodal status after CRT. In recent years, diffusion-weighted imaging (DWI) has proven its value and is now recommended as a routine addition to the MRI protocol for the assessment of rectal cancer, particularly in the restaging setting<sup>16-20</sup>. DWI highlights tissues with a dense cellular structure, making it a very suitable technique to detect malignant tumours, but also to detect lymph nodes, as lymphoid tissue has an intrinsically high cellular density<sup>21-25</sup>. With this in mind, DWI should also be well equipped to detect the typically small nodes that remain after CRT, as has also been suggested by a previous report that showed that in the restaging setting DWI detects 26% more nodes compared to standard T2-weighted MRI<sup>22</sup>. One could even argue that – given the high sensitivity of DWI to detect nodes – the absence of visible lymph nodes on DWI after CRT should indicate that no metastatic lymph nodes remain. If we can confirm this hypothesis, DWI could be valuable as a tool to help select

node-negative patients after CRT. The aim of this study therefore was to evaluate whether the absence of nodes on DWI can be used to establish a valid diagnosis of a yNO status in rectal cancer patients after neoadjuvant treatment.

## METHODS

### Patients

This study retrospectively assessed 130 consecutive patients diagnosed with locally advanced rectal cancer and treated with a long course of neoadjuvant CRT between November 2009 and May 2013 at Maastricht University Medical Centre. The study meets the guidelines of our country's responsible governmental agency; due to the retrospective nature of the study informed consent was not required. Inclusion criteria were (a) biopsy proven rectal adenocarcinoma, (b) long course of neoadjuvant treatment, (c) availability of a restaging MRI including a DWI sequence after neoadjuvant treatment, and (d) availability of histopathology after surgery or a sustained ycNO status and disease free survival of at least 2 years after organ preserving treatment as an endpoint for nodal response. The routine neoadjuvant treatment consisted of 50.4 Gy radiation combined with 2x825 mg/m<sup>2</sup>/day capecitabine. Four patients underwent 5x5 Gy of radiation combined with rapamycin and a long (8-10 week) waiting interval before restaging and surgery. Forty patients were excluded for the following reasons: a prolonged interval (>50 days) between restaging MRI and surgery (since histopathology in this setting will no longer be representative for the MRI findings) (n=24), missing data (e.g. missing histopathology reports; n=13), or DFS of <2 years after organ preserving treatment (n=3). This left a total study population of 90 patients who met all inclusion criteria. Some of the study patient were included in previous studies from our group on DWI for assessment of tumour response and on contrast-enhanced lymph node imaging. These previous studies did not focus on the use of DWI for lymph node evaluation<sup>26,27</sup>.

### MRI Imaging

All patients underwent a primary staging and restaging MRI at 1.5T MRI (Intera (Achieva) or Ingenia system, Philips Medical Systems, Best, The Netherlands). The standard protocol included two-dimensional T2-weighted (T2W) fast spin-echo sequences in 3 orthogonal directions (with the transverse images angled perpendicular and the coronal images angled parallel to the tumour axis as identified on the sagittal scan) and a transverse 3D T1W gradient echo (GRE) sequence with 1

mm isotropic voxels. This protocol was according to previous studies on lymph node evaluation<sup>27,28</sup>. The diffusion-weighted sequence was a transverse echo planar imaging (EPI) diffusion sequence with b1000 as the highest b-value, angled in the same plane as the transverse T2-weighted and T1-weighted sequences. All imaging sequences covered the full mesorectal compartment. Detailed sequence parameters are provided in **Table 1**.

### Image evaluation

MRIs were evaluated by two independent readers with 6 years and 3 years specific experience in reading pelvic MRI. The two readers evaluated the restaging high b-value (b1000) diffusion-weighted images and counted the number of visible lymph nodes within the mesorectal compartment (**Figure 1**). A lymph node on DWI was defined as a round or oval structure showing high signal intensity compared to the suppressed background signal of the mesorectal fat. The anatomical T2W and T1W sequences, as well as the primary staging MR examinations were at the reader's disposal for anatomical correlation and to confirm high signal structures as lymph nodes. The final nodal count on DWI was classified as 'no visible nodes' or '≥1 visible nodes'. The readers were blinded to all other patient data, including histopathological outcome, as well as to each other's results. In case of a discrepancy in lymph node count (i.e. no nodes versus ≥1 node seen on DWI) between the two observers, the readers re-evaluated the images in consensus to obtain a final count.

### Reference standard

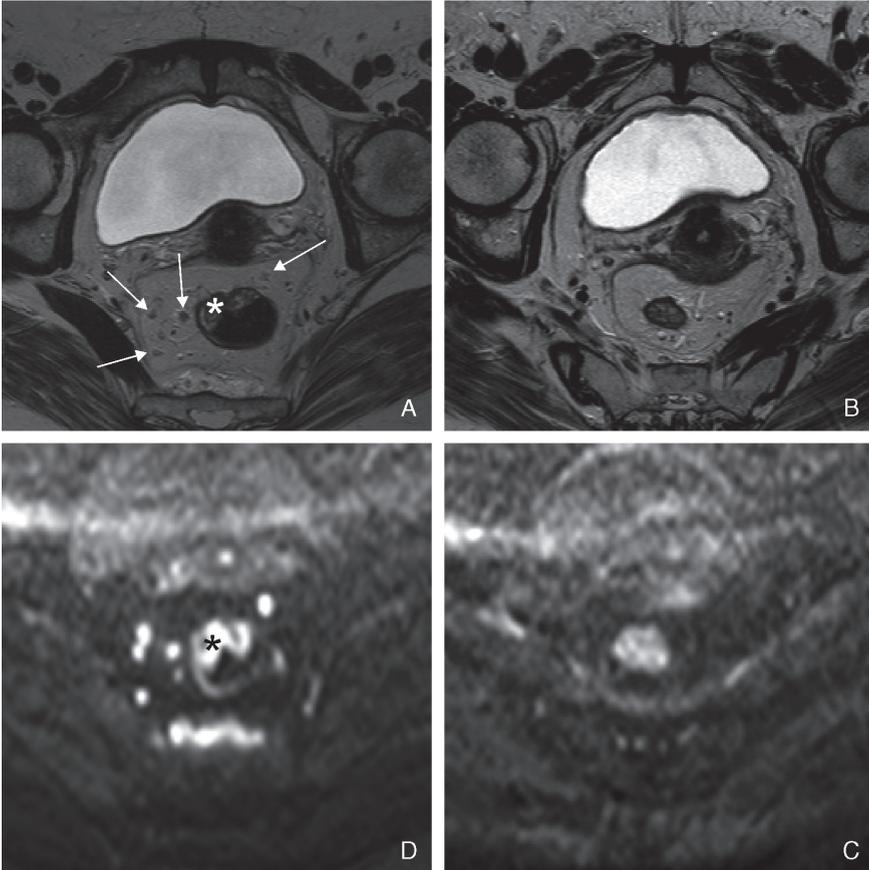
Sixty-eight patients underwent a total mesorectal excision. In these patients histological examination of the surgical resection specimens according to the Sixth American Joint Committee on Cancer TNM Staging System<sup>29</sup> (as advocated during the study period in our centre) served as the standard of reference. The total number of non-metastatic and metastatic lymph nodes at histopathology was recorded, as well as the overall nodal stage (ypN0, ypN1 or ypN2). The remaining twenty-two patients underwent a local excision (n=2) or a watchful waiting policy with intensive follow-up (n=20). For these patients a disease free survival (DFS) of ≥24 months with no evidence of growing/suspicious nodes on FU imaging was considered a surrogate endpoint for a yN0 status. Median DFS follow-up in these patients was 61 months (range 44-75 months).

**Table 1.** Sequence parameters for the T2-weighted and diffusion-weighted sequences.

Sequence Parameters	T2-Weighted Fast Spin-Echo	3D T1-Weighted Gradient-Recalled Echo	DWI Sequence 1 (n=56)	DWI-2 (n=17)	DWI-3 (n=22)
TR/TE	8456-9558/130-150	9.8/4.6	4808-4829/70	4971/70	4172-5241/68-70
No. Slices	22-33	200	50	24	20-24
Slice thickness (mm)	2	1	5	5	5
Slice gap (mm)	2	-	0.5	0.5	0.5
In-plane resolution (mm)	0.78 x 1.14	1.15 x 1.15	2.50 x 3.11-3.18	1.82 x 2.31	1.82 x 2.27
Echo-train length	25	1	1	1	1
No. Signal averages	2-6	1	4	5	5
b-Value (s/mm <sup>2</sup> )	N/A	N/A	0,(100)*,500,1000	0,500,1000	0,(25,50,100)*,500,1000
Fat-suppression technique	N/A	N/A	STIR	SPIR/fatsat	SPAIR
Echo planar imaging factor	N/A	N/A	53-55	55	61

Note - DWI sequences were used in the following time periods: DWI-1 Start study - November 2011; DWI-2 December 2011 - June 2012; DWI-3 July 2012 - End study.

\*Values in parentheses were only acquired in a subset of patients.



**Figure 1** 72-year-old woman with rectal tumour. T2-weighted MR image obtained before treatment (A) shows rectal tumour (asterisk, A) and several lymphnodes in mesorectal fat (arrows, A), some of which are difficult to discern. On corresponding DW image (C), more nodes are observed and nodes are more conspicuous because of their marked high signal (asterisk, C), compared with suppressed signal of mesorectal fat. No residual tumour is observed on T2-weighted MR image (B) or corresponding DW image (D) obtained after patient underwent chemoradiation therapy, and no residual nodes are observed. Patient had complete response (ypT0N0).

### Statistical analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, version 22.0, INC., Chicago, IL). To assess interobserver agreement for the number of lymph nodes detected on DWI, the intraclass correlation coefficient (ICC) was calculated with the two-way mixed method with absolute agreement for single measures. For all other analyses, the number of lymph nodes counted on

DWI was averaged between the two readers (for patients in whom an additional consensus reading was performed, the number reached after consensus reading was used). 2x2 contingency tables were constructed to correlate the number of nodes detected on post-CRT DWI (i.e. no nodes versus  $\geq 1$  nodes) with histopathology/follow-up. Sensitivity, specificity, positive and negative predictive values were calculated using a yN+ status as the positive outcome.

## RESULTS

### Patient and treatment characteristics.

Sixty patients were male and 30 were female. Seventy-one patients had a yN0 status (50 ypN0 after surgery and 21 ycN0 with a sustained ycN0 status and disease free survival of  $>24$  months). Nineteen patients had a ypN+ status after surgery (16 ypN1 and 3 ypN2). Detailed patient characteristics are shown in **Table 2**. Median interval between the restaging MRI and surgery was 16 days (range 5-50 days).

**Table 2.** Patient characteristics

Patient Characteristics	Number (%)	
<b>Sex</b>		
Male	60 (67%)	
Female	30 (33%)	
Median age (range)	65 (range 32-81)	
<b>yN-stage</b>		
N0	71 (79%)	
N+	19 (21%)	
N1	15 (17%)	
N2	4 (4%)	
<b>Number of patients per yT-stage</b>	<b>ypN0</b>	<b>ypN+</b>
T0	28 (31%)	0
T1	7 (8%)	0
T2	13 (14%)	3 (3%)
T3	20 (22%)	12 (13%)
T4	3 (3%)	4 (4%)
Median number of nodes harvested at histopathology in patients who underwent surgical resection (range)	12 (0-25)	11 (5-26)
Median number of malignant nodes (range)	-	2 (1-5)
Median size of malignant nodes at histopathology (range)	6 mm (1-21)	
Median size of benign nodes at histopathology (range)	3 mm (1-23)	

### Number of nodes on DWI to predict yN0 vs. yN+.

In 10/90 patients no nodes were observed on DWI, which was concordant with a yN0-status at histopathology/follow-up in all 10 (100%) cases. In the remaining 61 patients with a yN0 status the median number of nodes detected on DWI was 3 (range 1-17 nodes). In the 19 patients with a ypN+ status a median number of 4 nodes (range 1-7 nodes) was detected on DWI. In patients with a ypN1-status the median number of nodes on DWI was 4 (range 1-7) versus 4 (range 4-6) in patients with a ypN2-status. Diagnostic accuracy figures to discriminate yN0 versus yN+ (with yN+ being the positive outcome) are presented in **Table 3**: Sensitivity was 100%, specificity 14%, PPV 24% and NPV 100%.

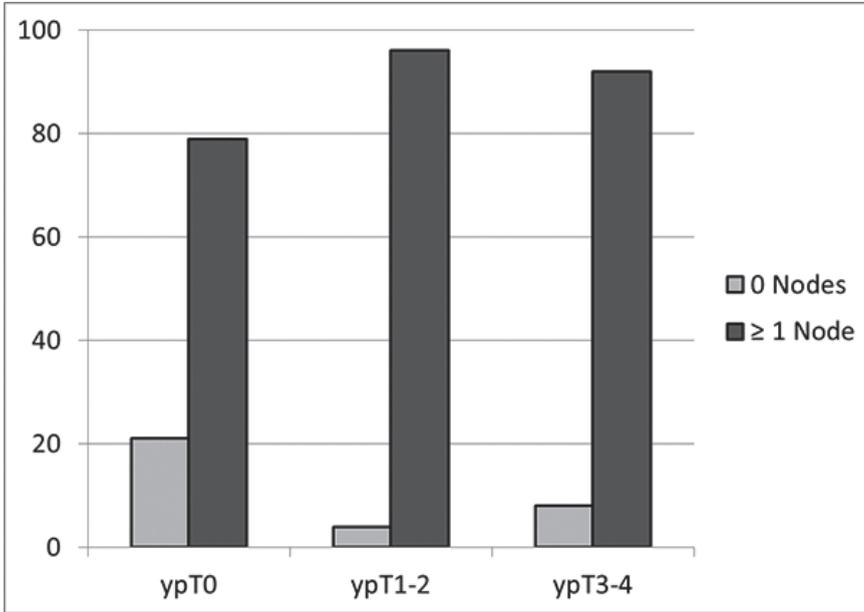
**Table 3.** Diagnostic performance to diagnose a ypN+ versus ypN0 status (with yN+ being the positive outcome) based on the presence/absence of visible nodes on post-CRT DWI.

Percentage (no. of nodes/total)	Sensitivity	Specificity	PPV	NPV
	100 (19/19)	14 (10/71)	24 (19/80)	100 (10/10)
95% CI	82-100	7-24	15-35	69-100

Note - in this analysis, yN-positive status is defined as the positive outcome.

### Results per yT-stage:

**Figure 2** shows the distribution of patients with 0 or  $\geq 1$  nodes on DWI among different tumour response groups. Of the patients with 0 nodes on DWI, 4 has residual tumour (1 ypT2, 2 ypT3 and 1 ypT4) and 6 were (clinical) complete tumour responders (yT0). The percentage of patients with no nodes on DWI was highest (21%) in the yT0 group, compared to 4% in the ypT1-2 group and 8% in the ypT3-4 group. The median number of nodes detected on DWI was 2 (range 0-17) in the yT0 group, 3 (range 0-12) in the ypT1-2 group and 3 (range 0-9) in the ypT3-4 group. In the subgroup of patients with a complete tumour response (yT0) 6 out of 28 patients had no visible nodes on DWI, which results in a specificity of 21%.



**Figure 2** Percentage of patients with zero nodes versus one or more nodes visible at DWI ( $b = 1000 \text{ s/mm}^2$ ) specified for patients with complete response (ypT0), good response (ypT1–2), and poor response (ypT3–4) of their primary rectal tumour, respectively.

*Interobserver agreement:*

Agreement between the two readers regarding the number of detected nodes on DWI was good with an ICC of 0.72 (95% CI: 0.46-0.84). In 7 patients, both readers initially observed no nodes, in 68 patients both readers observed  $\geq 1$  nodes. In the remaining 15 patients, there was a discrepancy in the number of nodes (0 versus  $\geq 1$  nodes) observed by the two readers and an additional consensus reading was performed (which resulted in the diagnosis of 0 nodes in 3 patients and  $\geq 1$  nodes in the other 12 patients).

**DISCUSSION**

The goal of this study was to evaluate whether the absence of visible lymph nodes on DWI after chemoradiotherapy is proof of a negative lymph node status. Although the absence of nodes on DWI was an infrequent finding only occurring in 10/90 patients, our results show that this was concordant with a yN0 status in all cases. This suggests that from a clinical point of view yN0 patients can safely be identified when these patients have no visible nodes on restaging DWI. However,

in the majority of patients with a yN0 status (61/71 yN0 patients), nodes were still visible on DWI. Hence, specificity was only 14%, which would imply a very stringent patient selection, where many of the complete nodal responders will be missed.

Interestingly, the number of patients with no visible nodes on DWI was higher in the patients with a complete tumour response compared to patients with residual tumour. In the yT0 subgroup, 21% of the patients had no visible nodes on DWI, compared to only 4-8% of the patients who still had residual tumour. Specificity to discriminate yN+ versus yN0 – albeit still far from optimal – was thus better for the complete tumour responders than for the whole patient group (21% versus 14%). This is beneficial as the yT0 group constitutes the most interesting subgroup with respect to the identification of potential candidates for organ-preserving treatments (watchful waiting). In these patients it is particularly important to identify residual nodal disease, since the nodes are left in situ and any missed residual nodal metastases harbour a potential risk for recurrence. The prevalence of a positive lymph node status after CRT is low in patients with a good tumour response (5-17%)<sup>4, 7, 30, 31</sup>. Moreover, as an effect of long course neoadjuvant therapy most lymph nodes decrease in size<sup>15, 31</sup>. As a result, residual nodal metastases mostly occur in very small nodes (< 3 mm). These small nodal sizes make it difficult to detect and therefore interpret the nodes after CRT, resulting in low reported sensitivities of 42% for detection of residual nodal metastases in patients with a good tumour response after CRT15.

Sensitivity to detect residual positive nodes in our study was 100%, which is better than reported in previous studies for routine T2-weighted MRI (42-89%)<sup>12-15</sup>. This means that when there is residual nodal disease, we will always see residual nodes on post-CRT DWI. However, it doesn't work the other way around: when we see residual nodes on DWI, this does not necessarily indicate that there is residual nodal disease. High signal intensity nodes were also still visualised after CRT in 61/71 of the yN0 patients, resulting in a low specificity of 14%, which is considerably lower than the specificities in the range of 46-95% previously reported for routine T2-weighted MRI<sup>12-15</sup>. This is in line with the findings of previous studies that showed that when considering all high signal nodes on DWI as positive for nodal disease this leads to false positive rates of up to 50%<sup>24, 33, 34</sup>. Hence, we will need to look for additional tools to further characterise the visible residual nodes after CRT. Inconsistent results have been reported for quantitatively measuring the apparent diffusion coefficient (ADC) of nodes on DWI to differentiate between positive and negative lymph nodes, both for the primary staging of lymph nodes as well as for restaging of nodes after CRT<sup>22, 23, 33-35</sup>. Moreover in small nodes it can be

very challenging to obtain reproducible ADC measurements<sup>22</sup>. A potentially promising alternative is the use of lymph node specific contrast-agents, with promising results reported for both ultrasmall superparamagnetic particles of iron oxide (USPIO) contrast as well as for the gadolinium-based contrast agent gadofosveset trisodium<sup>27, 28, 36, 37</sup>. However, these contrast-agents have not yet found their way into clinical practice.

Our study had some limitations. In 22 patients long-term follow-up was used as a surrogate reference standard instead of histopathological outcome. Patients had no signs of nodal recurrence for >2 years with a median follow-up of 61 months (minimum 44 months). Since previous studies have shown that both endoluminal and nodal recurrences mostly occur within the first two years, we believe this may be used as an alternative endpoint for a ypNO status<sup>3</sup>. Second, we only evaluated nodes within the mesorectal compartment since extramesorectal nodes are not routinely resected and can therefore not be correlated with histopathological outcome. Third, there were some variations in the DWI protocol (e.g. fat suppression techniques) employed throughout the course of the study. This, however, reflects daily practice where imaging protocols are subject to updates and optimisation. All of the images were, however, subjectively deemed to be of good diagnostic quality by both observers. Quality of fat suppression was similar and the signal to background ratio of the visualised lymph nodes was comparable in all studies. Fourth, the readers had a 3D T1-weighted sequence with 1 mm isotropic voxels at their disposal. This sequence was part of the protocol because of ongoing studies on lymph node imaging, but is not typically part of routine clinical rectal MRI protocols. Its use in the current study may thus potentially limit the generalisability of our study results. However, in practice lymph nodes were typically well recognised as nodes on the DWI sequence and the need to refer to the T1-weighted sequence was limited. Finally, we did not perform a within-patient comparison of our results for DWI to 'routine' assessment of the nodes on T2-weighted MRI, since the specific aim of our study was to test the hypothesis that the absence of nodes on DWI could be used to predict a ypNO status and various previous studies have already demonstrated the superiority of DWI compared to T2-weighted MRI for the detection of nodes. We did, however, compare our results to those previously reported in literature for T2-weighted MRI.

In conclusion, our results suggest that – although it is not a frequent finding – the absence of lymph nodes on a restaging DWI can be a reliable predictor of a node negative status after neoadjuvant CRT in patients with locally advanced rectal can-

cer and that the percentage of patients without visible nodes on DWI is highest in patients with a complete tumour response. DWI may therefore play a supporting role in the selection of patients for organ-preservation after CRT. When we see remaining nodes on DWI after CRT, we need to look for other tools to further characterise these nodes.



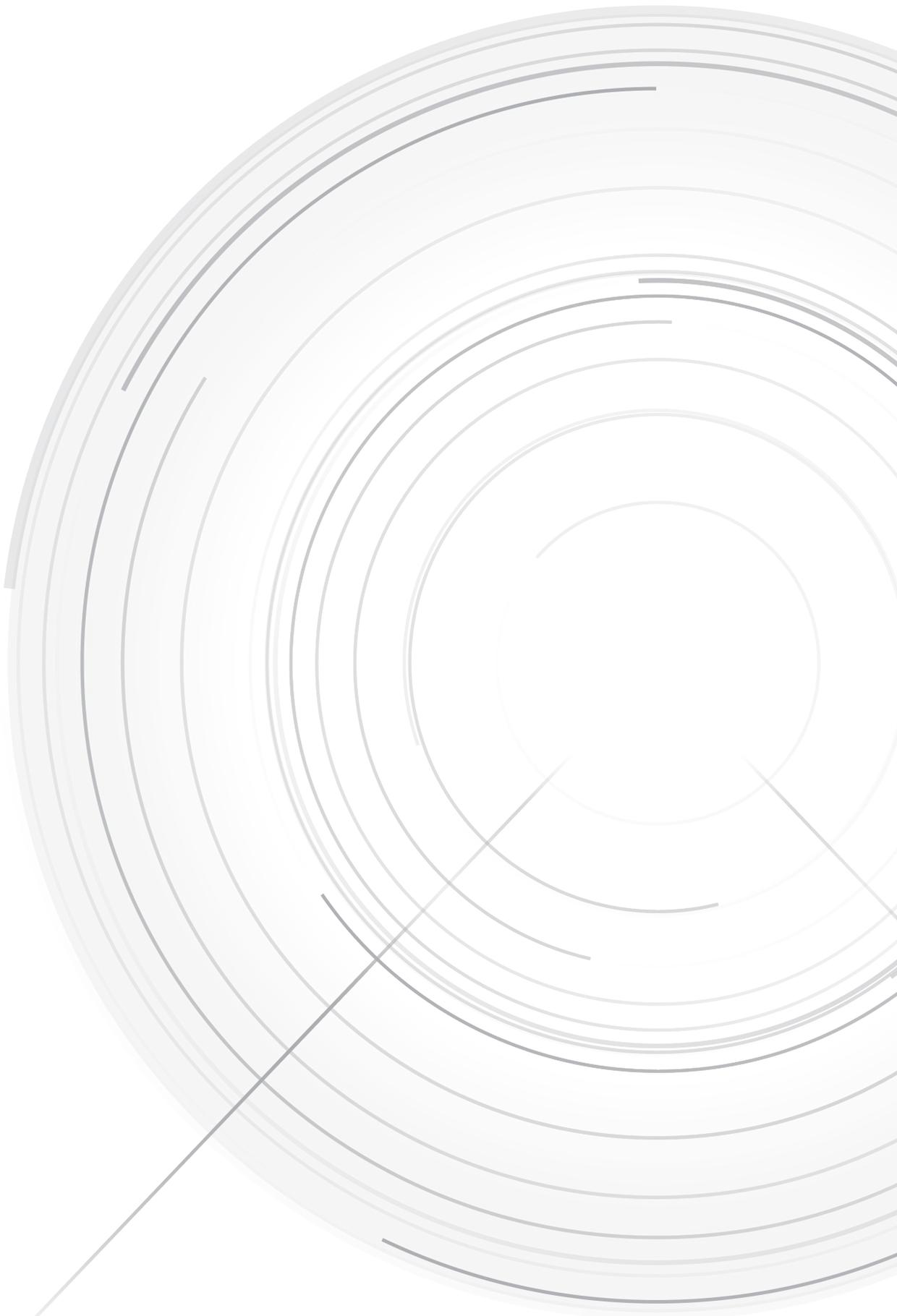
## REFERENCES

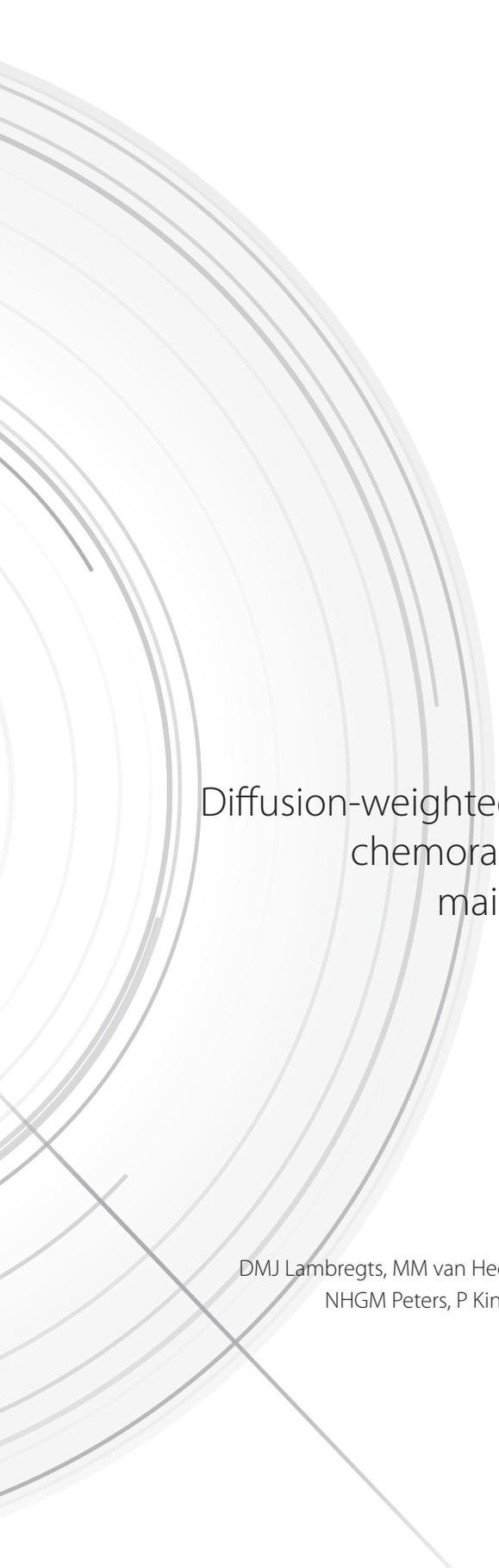
1. Lezoche G, Baldarelli M, Guerrieri M, et al. A prospective randomized study with a 5-year minimum follow-up evaluation of transanal endoscopic microsurgery versus laparoscopic total mesorectal excision after neoadjuvant therapy. *Surg Endosc*. 2008;22(2):352-8.
2. Verseveld M, de Graaf EJ, Verhoef C, et al. Chemoradiation therapy for rectal cancer in the distal rectum followed by organ-sparing transanal endoscopic microsurgery (CARTS study). *Br J Surg*. 2015;102(7):853-60.
3. Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg*. 2004;240(4):711-7; discussion 7-8.
4. Maas M, Nelemans PJ, Valentini V, 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.
5. Maas M, Lambregts DM, Nelemans PJ, et al. 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*. 2015;22(12):3873-80.
6. Gosens MJ, van Krieken JH, Marijnen CA, et al. Improvement of staging by combining tumour and treatment parameters: the value for prognostication in rectal cancer. *Clin Gastroenterol Hepatol*. 2007;5(8):997-1003.
7. Park JJ, You YN, Skibber JM, et al. Comparative analysis of lymph node metastases in patients with ypT0-2 rectal cancers after neoadjuvant chemoradiotherapy. *Dis Colon Rectum*. 2013;56(2):135-41.
8. Lahaye MJ, Engelen SM, Nelemans PJ, et al. Imaging for predicting the risk factors--the circumferential resection margin and nodal disease--of local recurrence in rectal cancer: a meta-analysis. *Semin Ultrasound CT MR*. 2005;26(4):259-68.
9. Bipat S, Glas AS, Slors FJ, Zwinderman AH, Bossuyt PM, Stoker J. Rectal cancer: local staging and assessment of lymph node involvement with endoluminal US, CT, and MR imaging--a meta-analysis. *Radiology*. 2004;232(3):773-83.
10. Engelen SM, Beets-Tan RG, Lahaye MJ, et al. MRI after chemoradiotherapy of rectal cancer: a useful tool to select patients for local excision. *Dis Colon Rectum*. 2010;53(7):979-86.
11. Lahaye MJ, Beets GL, Engelen SM, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part II. What are the criteria to predict involved lymph nodes? *Radiology*. 2009;252(1):81-91.
12. Lee JH, Jang HS, Kim JG, et al. Prediction of pathologic staging with magnetic resonance imaging after preoperative chemoradiotherapy in rectal cancer: pooled analysis of KROG 10-01 and 11-02. *Radiother Oncol*. 2014;113(1):18-23.
13. Koh DM, Chau I, Tait D, Wotherspoon A, Cunningham D, Brown G. Evaluating mesorectal lymph nodes in rectal cancer before and after neoadjuvant chemoradiation using thin-section T2-weighted magnetic resonance imaging. *Int J Radiat Oncol Biol Phys*. 2008;71(2):456-61.

14. Brown G, Radcliffe AG, Newcombe RG, Dallimore NS, Bourne MW, Williams GT. Preoperative assessment of prognostic factors in rectal cancer using high-resolution magnetic resonance imaging. *Br J Surg.* 2003;90(3):355-64.
15. Perez RO, Pereira DD, Proscurshim I, et al. Lymph node size in rectal cancer following neoadjuvant chemoradiation--can we rely on radiologic nodal staging after chemoradiation? *Dis Colon Rectum.* 2009;52(7):1278-84.
16. Kim SH, Lee JM, Hong SH, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumour response to neoadjuvant chemo- and radiation therapy. *Radiology.* 2009;253(1):116-25.
17. Park MJ, Kim SH, Lee SJ, Jang KM, Rhim H. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging for predicting tumour clearance of the mesorectal fascia after neoadjuvant chemotherapy and radiation therapy. *Radiology.* 2011;260(3):771-80.
18. Lambregts DM, Vandecaveye V, Barbaro B, 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.
19. 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.
20. Beets-Tan RG, Lambregts DM, Maas M, et al. 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.* 2013;23(9):2522-31.
21. Mir N, Sohaib SA, Collins D, Koh DM. Fusion of high b-value diffusion-weighted and T2-weighted MR images improves identification of lymph nodes in the pelvis. *J Med Imaging Radiat Oncol.* 2010;54(4):358-64.
22. Lambregts DM, Maas M, Riedl RG, et al. Value of ADC measurements for nodal staging after chemoradiation in locally advanced rectal cancer-a per lesion validation study. *Eur Radiol.* 2011;21(2):265-73.
23. Heijnen LA, Lambregts DM, Mondal D, et al. Diffusion-weighted MR imaging in primary rectal cancer staging demonstrates but does not characterise lymph nodes. *Eur Radiol.* 2013;23(12):3354-60.
24. Heusner TA, Kuemmel S, Koeninger A, et al. Diagnostic value of diffusion-weighted magnetic resonance imaging (DWI) compared to FDG PET/CT for whole-body breast cancer staging. *Eur J Nucl Med Mol Imaging.* 2010;37(6):1077-86.
25. Nakai G, Matsuki M, Inada Y, et al. Detection and evaluation of pelvic lymph nodes in patients with gynecologic malignancies using body diffusion-weighted magnetic resonance imaging. *J Comput Assist Tomogr.* 2008;32(5):764-8.
26. Lambregts DM, Rao SX, Sasses S, 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(2):1034-9.

27. Heijnen LA, Lambregts DM, Martens MH, et al. Performance of gadofosveset-enhanced MRI for staging rectal cancer nodes: can the initial promising results be reproduced? *Eur Radiol.* 2014;24(2):371-9.
28. Lambregts DM, Beets GL, Maas M, et al. Accuracy of gadofosveset-enhanced MRI for nodal staging and restaging in rectal cancer. *Ann Surg.* 2011;253(3):539-45.
29. Greene FL, Page DL, Fleming ID, eds. *AJCC cancer staging manual*, 6th ed. Chicago, IL: American Joint Committee on Cancer, 2002.
30. Bedrosian I, Rodriguez-Bigas MA, Feig B, et al. Predicting the node-negative mesorectum after preoperative chemoradiation for locally advanced rectal carcinoma. *J Gastrointest Surg.* 2004;8(1):56-62.
31. Yeo SG, Kim DY, Kim TH, et al. Pathologic complete response of primary tumour following preoperative chemoradiotherapy for locally advanced rectal cancer: long-term outcomes and prognostic significance of pathologic nodal status (KROG 09-01). *Ann Surg.* 2010;252(6):998-1004.
32. Heijnen LA, Maas M, Beets-Tan RG. Nodal restaging in rectal cancer: why is restaging after chemoradiation more accurate than primary nodal staging? *Int J Colorectal Dis* 2016;31(6):1157-62
33. Attenberger UI, Pilz LR, Morelli JN, et al. Multi-parametric MRI of rectal cancer - do quantitative functional MR measurements correlate with radiologic and pathologic tumour stages? *Eur J Radiol.* 2014;83(7):1036-43.
34. Yasui O, Sato M, Kamada A. Diffusion-weighted imaging in the detection of lymph node metastasis in colorectal cancer. *Tohoku J Exp Med.* 2009;218(3):177- 83.
35. Cho EY, Kim SH, Yoon JH, et al. Apparent diffusion coefficient for discriminating metastatic from non-metastatic lymph nodes in primary rectal cancer. *Eur J Radiol.* 2013;82(11):e662-8.
36. Koh DM, Brown G, Temple L, et al. Rectal cancer: mesorectal lymph nodes at MR imaging with USPIO versus histopathologic findings--initial observations. *Radiology.* 2004;231(1):91-9.
37. Lahaye MJ, Engelen SM, Kessels AG, et al. USPIO-enhanced MR imaging for nodal staging in patients with primary rectal cancer: predictive criteria. *Radiology.* 2008;246(3):804-11.







# Chapter 4

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

DMJ Lambregts, MM van Heeswijk, A Delli Pizzi, SCG van Elderen, LA Andrade, NHGM Peters, P Kint, M Osinga-de Jong, S Bipat, R Ooms, MJ Lahaye, M Maas, GL Beets, FCH Bakers, RGH Beets-Tan.

*European Radiology 2017; ePub ahead of print*

## 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 (b1000 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.



## 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<sup>1-3</sup>. 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<sup>4</sup>. 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<sup>5-7</sup>. 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<sup>8</sup>.

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<sup>5-7,9</sup>. 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<sup>10-12</sup>. Teaching is therefore an important issue.

It would be helpful to gain knowledge on the pitfalls and interpretation errors most commonly encountered by nonexpert 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 local institutional review board approved the study 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<sup>®</sup>; 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 b1000 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.

### 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<sup>13</sup>). The readers based their score on the high b-value (b1000) 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 b1000 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.



**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, or 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 effect (with corresponding high signal on ADC) or artefacts

## 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.

## 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<sup>5, 13-15</sup>. 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 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.



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

Variable	Value
<b>Sex:</b>	
Male	69
Female	31
Median age (range)	64 (31-82)
<b>Primary cTN stage (as assessed with MRI)</b>	
cT-stage	
cT1-2	17
cT3	74
cT4	9
cN-stage	
cN0	27
cN1	27
cN2	46
<b>Surgical treatment</b>	
Low Anterior Resection	42
Abdominoperianal Resection microsurgery	12
Transanal Endoscopic Microsurgery	4
Pelvic exenteration	4
None (watchful waiting)	38
Mean follow up in months ( $\pm$ SD)	32 ( $\pm$ 11)
<b>Final yTN stage</b>	
yT-stage	
yT0*	46
yT1	7
yT2	16
yT3	25
yT4	6
yN-stage	
yN0^	77
yN1	18
yN2	5
<b>Final tumour response (yT-stage)</b>	
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 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)	67% (19/25)	72% (21/29)	86% (25/29)
Specificity	80% (20/25)	68% (17/25)	90% (19/21)	76% (16/21)
Postive Predictive value	75% (15/20)	70% (19/20)	91% (21/23)	83% (25/23)
Negative predictive value	67% (20/30)	74% (17/30)	70% (19/27)	80% (16/27)
No. Of equivocal scores	11	4	6	0
Interobserver agreement (κ) (95% CI)	0.58 (0.41-0.74)		0.71 (0.57-0.85)	

CI Confidence interval, AUC Area under the ROC curve

### Interpretation pitfalls

The five most common interpretation difficulties and pitfalls encountered by the non-expert readers that were discussed 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 (**Figure 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 (**Figure 2**).

#### 3. T2 shine-through:

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 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; **Figure 3**).

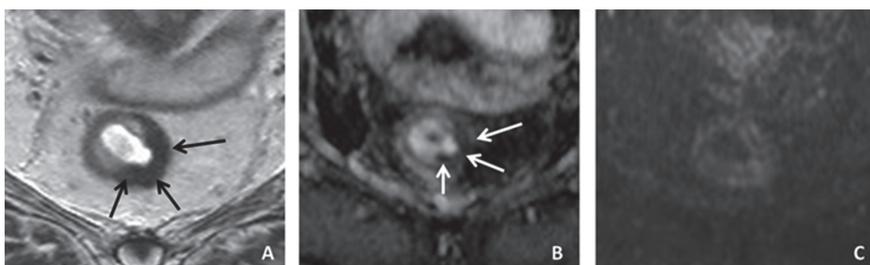


**Table 4.** Overview of main potential causes of error (pitfalls)

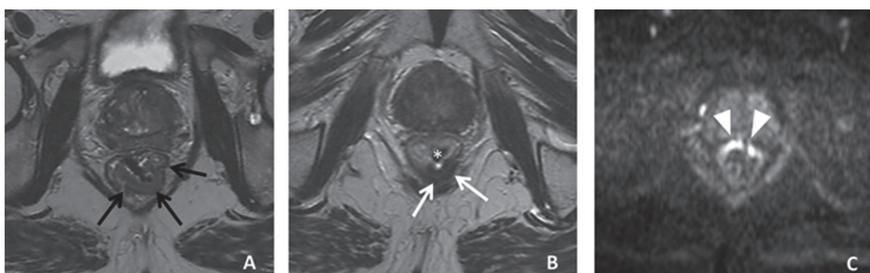
No	Potential pitfall	Illustrated in:	% of cases*
1	Misinterpretation of low signal on ADC map in case of fibrosis	Figure 1	41%
2	Susceptibility effects	Figure 2	33%
3	T2 shine through of fluid in the rectal lumen	Figure 3	12%
4	Suboptimal sequence angulation	Figure 4	6%
5	Collapsed rectal wall	Figure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

ADC = Apparent Diffusion Coefficient

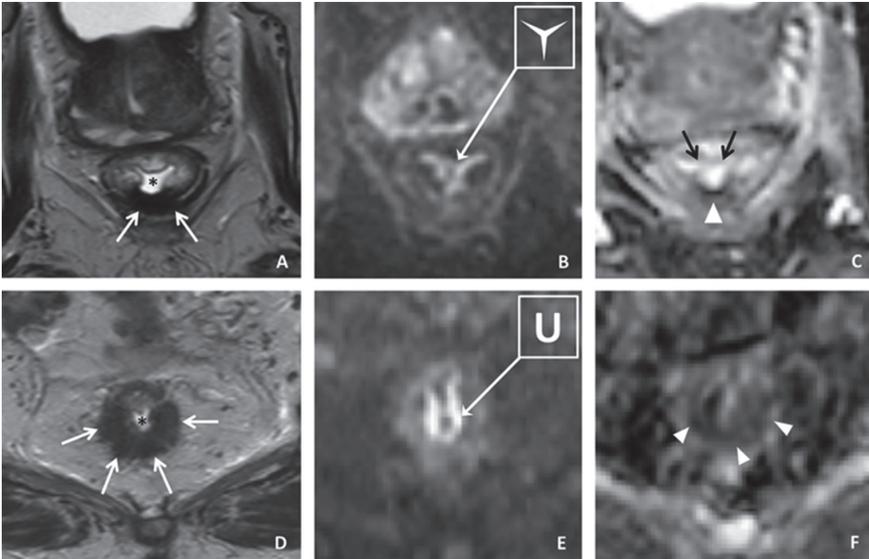
\* Represents the percentage of cases in which a potential pitfall was present *cq.* identified by the expert reader after discussing the case with the two non-expert readers. It does not necessarily indicate that the case was in fact misinterpreted by one or both readers.



**Figure 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 b1,000 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 b1,000 diffusion image.



**Figure 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 b1,000 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.



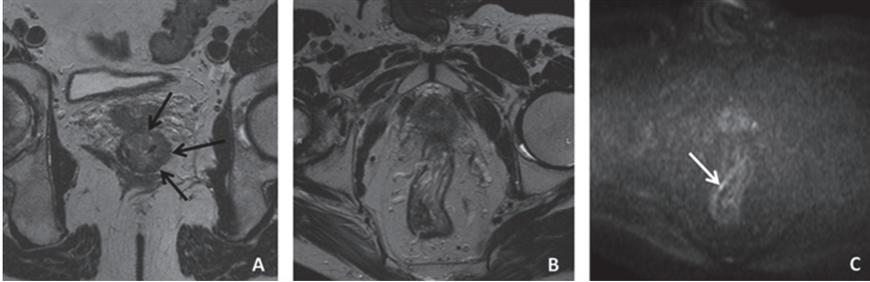
**Figure 3.** Two patients, both with fibrotic wall thickening on T2-weighted MR images on the dorsal side (a,d white arrows), after CRT. In both patients a high signal is seen on the corresponding b1,000 DW images (b, e). a-c First patient. On the DW image (b) the signal is star-shaped and correspond 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 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.

#### 4. Suboptimal sequence angulation:

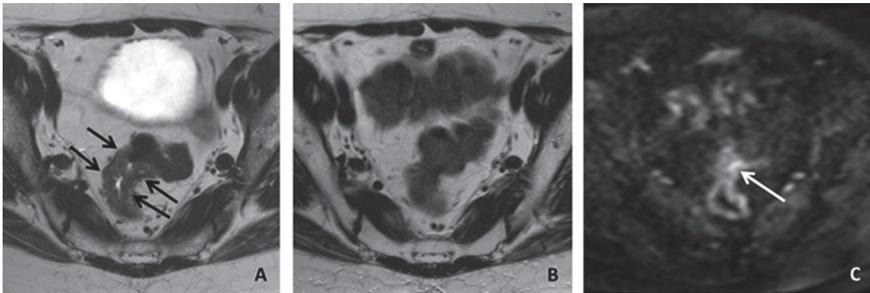
In proximal rectal tumours, angulation perpendicular to the tumour results in coronal-like imaging planes, which were found to be more difficult to interpret. Moreover when the images before and after CRT were not angled in identical planes, the readers found it difficult to compare the images and correctly interpret the diffusion images after CRT (**Figure 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 (**Figure 5**).



**Figure 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.



**Figure 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 rectum 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 b1,000 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.

## DISCUSSION

The main aim of this study was to document the most common potential interpretation pitfalls encountered by nonexpert 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<sup>16</sup>. 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<sup>17</sup>. 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 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 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 **Figure 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 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 tumoural 'region of interest' and consider it as non-suspicious, as illustrated in **Figure 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 (**Figure 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<sup>18,19</sup>. 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<sup>20</sup>, although the use of such sequences within the abdomen has scarcely been studied so far. Finally, it is important to ensure optimal sequence angulation by well-trained personnel. This can prevent interpretation difficulties caused by suboptimal 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 pre-treatment 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<sup>5-7</sup>. Moreover, the two readers showed good interobserver 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, the two readers assigned equivocal scores to six and zero cases, 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 keener 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 b1000 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%<sup>21</sup>. 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<sup>5, 13–15</sup>. 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<sup>1, 2, 22, 23</sup>.

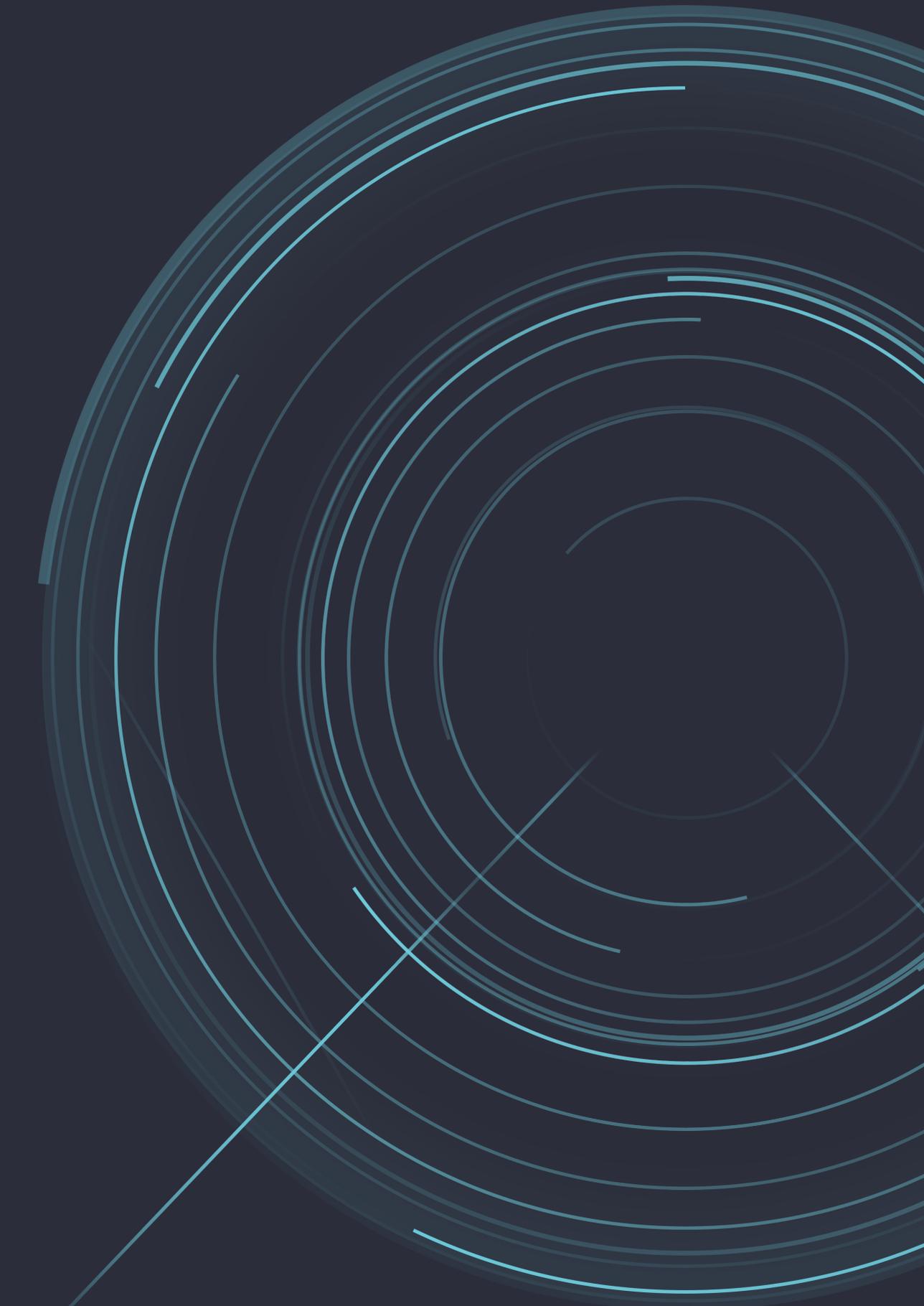
In conclusion, this study showed that there are five important potential DWI interpretation pitfalls, which were documented with imaging examples and may serve as a reference to teach future readers interested in the use of DWI for rectal tumour response evaluation. The study also showed that nonexpert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

## REFERENCES

1. Habr-Gama A, Gama-Rodrigues J, Sao Juliao GP 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;88:822–828
2. Maas M, Beets-Tan RG, Lambregts DM et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011;29:4633–4640
3. Lezoche E, Guerrieri M, Paganini AM 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;19:751–756
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;269:101–112
5. Lambregts DM, Vandecaveye V, Barbaro B et al. Diffusion weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. *Ann Surg Oncol* 2011;18:2224–2231
6. Kim SH, Lee JM, Hong SH 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;253:116–125
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;85:577–586
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;89:20150328
9. Sassen S, de Booiij M, Sosef M 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;23:3440–3449
10. Miglioretti DL, Gard CC, Carney PA et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology* 2009;253:632–640
11. Boellaard TN, Nio CY, Bossuyt PM, Bipat S, Stoker J. Can radiographers be trained to triage CT colonography for extracolonic findings? *Eur Radiol* 2012;22:2780–2789
12. Saba L, Guerriero S, Sulis R et al. Learning curve in the detection of ovarian and deep endometriosis by using magnetic resonance: comparison with surgical results. *Eur J Radiol* 2011;79:237–244
13. Maas M, Lambregts DM, Nelemans PJ et al. 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* 2015;22:3873–3880
14. Lambregts DM, Lahaye MJ, Heijnen LA et al. 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* 2016;26:2118–2125

15. Curvo-Semedo L, Lambregts DM, Maas M 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;260:734–743
16. Gatehouse PD, Bydder GM. Magnetic resonance imaging of short T2 components in tissue. *Clin Radiol* 2003;58:1–19
17. Weinreb JC, Barentsz JO, Choyke PL et al. PI-RADS Prostate Imaging – Reporting and Data System: 2015, Version 2. *Eur Urol* 2016;69:16–40
18. Slater A, Halligan S, Taylor SA et al. 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* 2006;61:65–70
19. Beets-Tan RG, Lambregts DM, Maas M et al. 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* 2013;23:2522–2531
20. Mikayama R, Yabuuchi H, Kobayashi K et al. 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* 2016. doi:10.1594/ecr2016/C-1045
21. Maas M, Nelemans PJ, Valentini V 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;11:835–844
22. Martens M, Maas M, Heijnen LA et al. Long term outcome of an organ preservation program after neoadjuvant treatment for rectal cancer. *J Natl Cancer Inst* 2016;108(12). doi:10.1093/jnci/djw171
23. Smith JD, Ruby JA, Goodman KA et al. Nonoperative management of rectal cancer with complete clinical response after neoadjuvant therapy. *Ann Surg* 2012;256:965–972

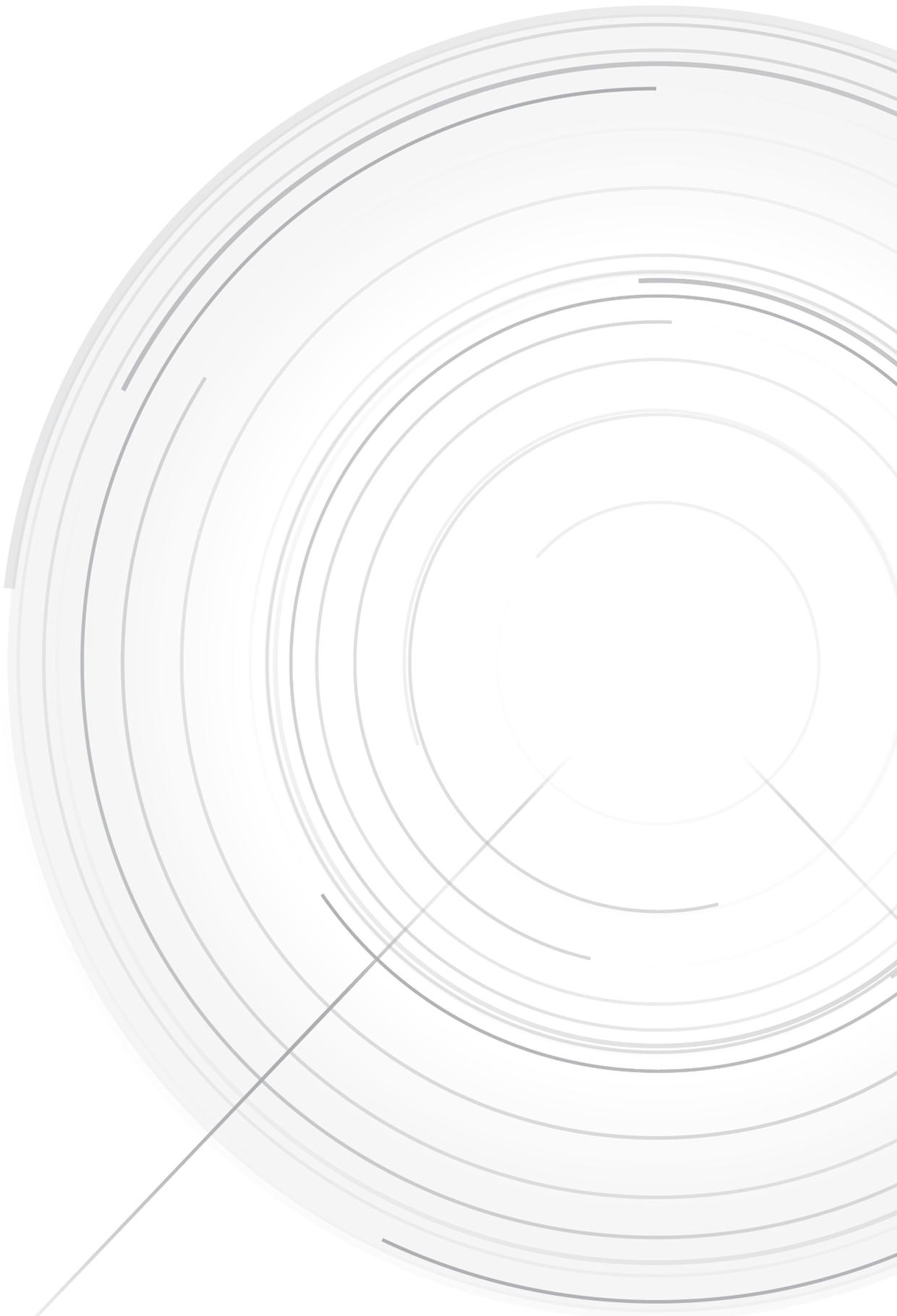


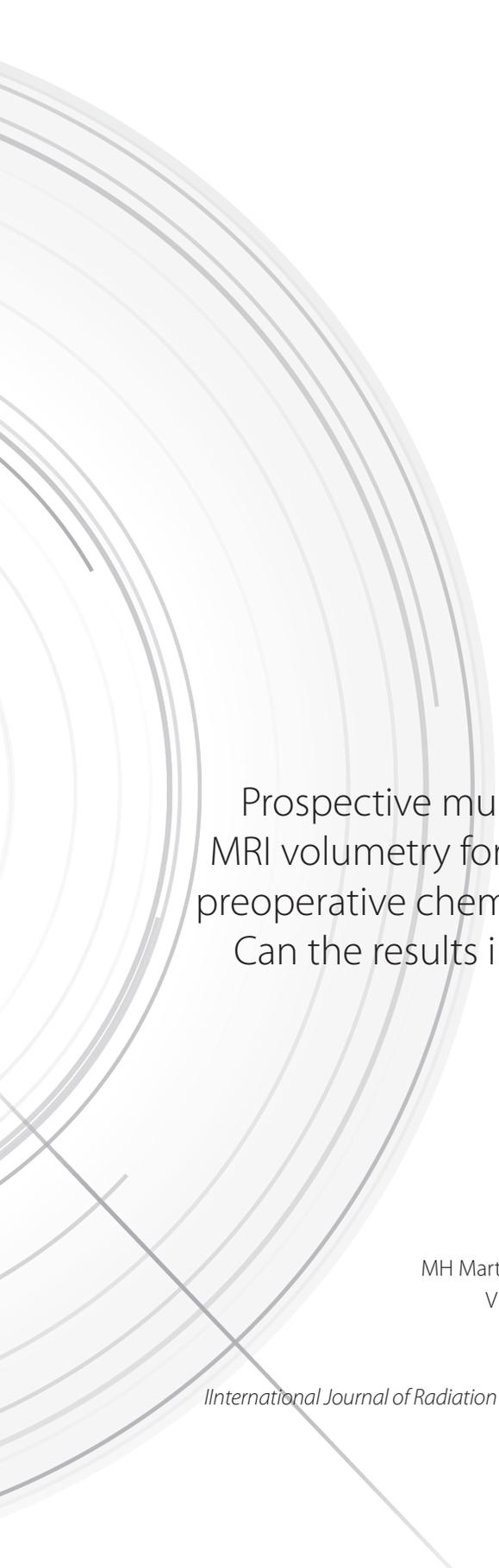




# PART II

Quantitative analysis of  
rectal cancer imaging -  
a practical approach





# Chapter 5

Prospective multicenter validation study of MRI volumetry for response assessment after preoperative chemoradiation in rectal cancer: Can the results in literature be reproduced?

MH Martens, MM van Heeswijk, JJ van den Broek, S-X Rao,  
V Vandecaveye, RA Vliegen, WH Schreurs, GL Beets,  
DMJ Lambregts, RGH Beets-Tan

*International Journal of Radiation Oncology, Biology, Physics. 2015 Dec 1;93(5):1005-14.*

## ABSTRACT

### Purpose

To review available literature on tumour size/volume measurements on MRI for response assessment after chemoradiotherapy, and validate these cut-offs in an independent multicenter patient cohort.

### Methods

The study included two parts: [1] Review of literature; papers were included that assessed the accuracy of tumour size/volume measurements on MRI for tumour response assessment. Size/volume cut-offs were extracted. [2] Multicenter validation; Extracted cut-offs from literature were tested in a multicenter cohort (n=146). Accuracies were calculated and compared to reported results from literature.

### Results

The review included fourteen papers, in which 3 different measurement methods were assessed: [1] tumour length, [2] 3-dimensional tumour size, and [3] whole volume. Study outcomes consisted of [1] complete response (ypT0) versus residual tumour, [2] tumour regression grade (TRG) 1-2 versus 3-5, and [3] T-downstaging (ypT<cT). In the multicenter cohort, best results were obtained for the validation of the whole-volume measurements, in particular for the outcome ypT0 (accuracy 44-80%), with the optimal cut-offs being 1.6 cm<sup>3</sup> (post-CRT) and a volume-reduction of  $\Delta$ 80-86.6%. Accuracies for whole-volume measurements to assess TRG1-2 were 52-61%, and for T-downstaging 51-57%. Overall accuracies for tumour length ranged between 48-53% and for 3D-size measurement between 52-56%.

### Conclusions

MR-volumetry using whole tumour volume measurements can be helpful in rectal cancer response assessment with selected cut-off values. Measurements of tumour length or three-dimensional tumour size are not helpful. MRI volumetry is mainly accurate to assess a complete tumour response (ypT0) after CRT (accuracies up to 80%).



## INTRODUCTION

Since the introduction of standardized total mesorectal excision (TME) and preoperative chemoradiation treatment (CRT), the outcome for rectal cancer patients has considerably improved<sup>1-3</sup>. Preoperative treatment often results in downsizing and downstaging of the tumour<sup>3,4</sup>, and in 15-20% of the patients even in a complete tumour regression<sup>5</sup>. In patients with a clinically complete or near complete response, organ-preservation (local excision or “watch-and-wait”) is now increasingly considered as a potentially good alternative to surgical resection<sup>6-8</sup>. The recommended tools to assess tumour response include digital rectal examination, endoscopy and imaging (mainly MRI). MRI, approximately 6-8 weeks after completion of chemoradiotherapy<sup>9, 10</sup>, performs well in assessing tumour regression from initially invaded structures and can be beneficial in determining the definitive surgical approach<sup>11</sup>. Although MRI has proven useful to assess changes in tumour volume and morphology<sup>12</sup>, evaluation of minimal residual tumour within post-radiation fibrosis remains difficult. As a result, the performance of MRI to assess response after chemoradiotherapy remains limited, as was shown in a recent meta-analysis with an overall sensitivity of 50.4% and specificity of 91.2% for MRI in restaging tumour after CRT<sup>10</sup>. Various studies have investigated the value of measuring changes in the tumour size and/or volume on MRI as a predictor of response. Reported results for these measurements vary considerably, with overall accuracies ranging between 42%-88%<sup>13-27</sup>. This could partly be explained by the variety in methods for the tumour size/volume measurements (e.g. one-dimensional versus 3-dimensional versus ‘whole-volume’ volumetry), by the variety in cut-off values, by the retrospective nature of the majority of often small-scale studies, and the different study endpoints used to define response. Furthermore, prospective validation of retrospectively determined cut-offs in an independent and large multicenter patient cohort is lacking. Aim of the current study was threefold: [1] to systematically review the available literature on the value of tumour size/volume measurements on MRI (using routine T2-weighted sequences) for tumour response prediction and assessment in patients with rectal cancer undergoing preoperative CRT, [2] to prospectively validate the cut-off values reported in literature in an independent multicenter patient cohort, and [3] to evaluate – based on the above two steps – whether MRI volumetry is reliable to predict and assess primary tumour response and which method and cut-off values would be most accurate.

## MATERIALS AND METHODS

The study consisted of two parts: [1] systematic review of the available literature and [2] multicenter validation of the data derived from the literature review.

### 1. Literature review

A literature search was performed to identify all papers focusing on assessment of response to neoadjuvant treatment in patients with locally advanced rectal cancer by means of tumour size or volumetric measurements using routine morphological (T2-weighted) MRI. We performed a search of three major databases: Embase, Medline, and Cochrane library using (a combination of) the following key words: “rectal cancer”, “rectal adenocarcinoma”, “rectal neoplasm”, “response evaluation”, “response assessment”, “response prediction”, “radiotherapy”, “neoadjuvant therapy”, “chemoradiation”, “chemoradiotherapy”, “tumour length”, “tumour volumetry”, “tumour size”, “tumour volume”, “tumour burden”, “tumour regression grade”, “tumour regression”, “downstaging”, and “pathological complete response”. Inclusion criteria were: (1) tumour size/length or volume was measured on T2-weighted MRI before and/or after neoadjuvant therapy, (2) the main study outcome was the response to neoadjuvant treatment with the standard of reference being pathological complete response, tumour-downstaging ( $cT > pT$ ), or favorable tumour regression grade (TRG), (3) cut-off values for tumour size/volume were used to determine the diagnostic performance for predicting response, and (4) the results were reported in 2x2 contingency tables or these could otherwise be derived from the data reported. Two reviewers checked the titles and abstracts of the identified studies to determine studies which potentially met the inclusion criteria. Additional studies were identified by checking the reference list of the selected studies. Full text copies were then studied to decide which articles met the final inclusion criteria.

### 2. Validation study

#### *Patients*

According to our country’s national law, institutional review board approval is not required for this retrospective study and patients consent was waived. For the validation study, 146 locally advanced rectal cancer patients were included between 2006-2013 in one of four study centres (2 academic referral centres and 2 non-academic centres). None of the patients were previously included in any of the previously reported papers included in the review. Inclusion criteria consisted of (1) histologically proven rectal cancer, (2) locally advanced rectal cancer ( $cT3-$

4 and/or cN+), (3) long course of neoadjuvant treatment followed by resection, and (4) availability of T2-weighted MRI before and after neoadjuvant treatment. The routine neoadjuvant treatment protocol consisted of either 50.4 Gy radiation combined with 2x825 mg/m<sup>2</sup>/day capecitabine or 45 Gy radiation combined with 225 mg/m<sup>2</sup>/day 5-fluorouracil (5-FU). Nineteen patients received an alternative scheme with additional oxaliplatin and/or bevacizumab.

### Magnetic Resonance Imaging

MRI was performed as part of the routine diagnostic workup for both primary staging and restaging after CRT in all centres. The standard interval between completion of CRT and the restaging MRI was 4-8 weeks in all centres. The standard MRI protocol included T2-weighted Fast Spin Echo (FSE) sequences in 3 orthogonal directions. Slice thickness was 3-3.5 mm. Detailed information on scanning systems, patient preparation, and sequence parameters are described in **Table 1**.

**Table 1** Overview of hardware, sequence parameters, and patient preparation

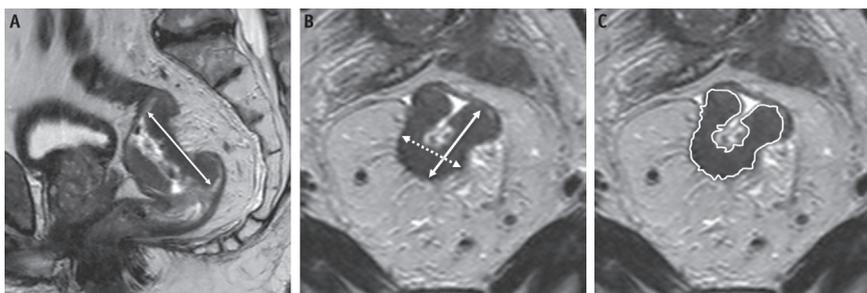
Parameter	Center 1	Center 2	Center 3	Center 4, protocol 1	Center 4, protocol 2
Manufacturer	Philips	Siemens	Siemens	Siemens	Philips
Field strength (T)	1.5	1.5	1.5	1.5	3.0
Type of coil (surface/endorectal)	Surface coil	Surface coil	Surface coil	Surface coil	Surface coil
Repetition time (ms)	8456-11,788	4000	3400-4160	8870	5000
Echo time (ms)	130	102	122	134	95
No. of slices	30	32	48	20	32
Slice thickness (mm)	3	3.5	3.5	3.5	3
Slice gap (mm)	2	0.7	0-0.35	0.7	0
In-plane resolution (mm × mm)	0.78 × 1.14	0.7 × 0.6	1.0 × 0.7	0.63 × 0.73	0.69 × 0.82
Echo train length	25	23	23	61	19
No. of signal averages	2	3	2	3	1
Bowel preparation	No	No	No	Endorectal filling	Endorectal filling

### Image evaluation

MR images were evaluated on an offline workstation using the closed polygon tool in OsiriX Medical Imaging Software<sup>28</sup> by one reader with 8 years of specific rectal MRI experience (S-XR). A second reader with 3 years of experience (MHM) re-evaluated the pre- and post-CRT images of a random sample of one third (n=48) of the study patients to evaluate interobserver reproducibility. The readers were blinded to each other's results and all clinical and histopathological data. Tumour size and volume measurements (including only the tumour in the rectal wall, excluding any lymph nodes or extraluminal tumour deposits) were performed on the primary staging MRI as well as on the restaging MRI. Measurements were manually performed according to 3 different methods (concordant with different methodologies described in literature): [1] One-dimensional size measurement of the maximum



tumour length on sagittal MRI<sup>15, 27</sup>, [2] 3-dimensional size measurements (i.e., tumour length (sagittal plane) x anterior-posterior (AP) diameter (axial plane) x left-right (LR) diameter (axial plane))<sup>26, 27</sup>, and [3] measurement of the whole tumour volume by manually tracing the tumour boundaries on each tumour-containing slice and multiplying these cross-sectional areas by the slice thickness (including the slice gap)<sup>13-25</sup>. For all 3 methods the percentage of volume/size reduction ( $V_{\Delta}$ ) was calculated by  $(V_{pre} - V_{post}) / V_{pre} \times 100$ , with  $V_{pre}$  and  $V_{post}$  representing the volumes/sizes pre- and post-CRT, respectively. The different measurement methods are illustrated in **Figure 1**.



**Figure 1.** (A) T2-weighted sagittal MRI, the arrow indicates the maximal tumour length (method 1). (B) T2-weighted axial MRI, the dashed arrow indicates the left-right (LR) diameter, the filled arrow represents the anterior-posterior (AP) diameter (method 2: length x LR x AP). (C) Delineation of the tumour boundaries for method 3 (tumour area x slice thickness).

### Reference standard

All patients underwent surgical resection. Histopathological examination was performed according to the method described by Quirke et al<sup>29</sup>. The pathological T-stage (ypT) was reported for each patient. Tumour regression grade was assessed as part of routine clinical care according to the method described by Mandard<sup>30</sup> or Dworak<sup>31</sup>. For study purposes, the TRGs measured according to the Dworak classification were converted to the method of Mandard, with TRG 1-2 (concordant with Dworak TRG 3-4) being a good response and TRG 3-5 (concordant with Dworak TRG 0-2) being a poor response<sup>32</sup>.

### Statistical Analyses

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, version 20.0, Inc., Chicago, Illinois, USA). For the measurements performed by both readers (n=48), interobserver agreement (IOA) was evaluated

by calculating the intraclass correlation coefficient (ICC) with ICC 0-0.20 indicating poor, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good and 0.81-1.0 indicating excellent agreement. For all other analyses, the measurements of reader 1 were used. Non-parametric Mann-Whitney U test was performed to compare the volumes between good/poor responders. Dichotomization between good/poor responders was performed according to different methods derived from the literature review: [1] complete response (ypT0) vs residual tumour, [2] Tumour-downstaging (cT-stage > ypT-stage) vs no Tumour-downstaging (cT-stage ≤ ypT-stage), and [3] TRG 1-2 vs TRG3-5, where a favorable response (i.e. ypT0, cT>pT, TRG1-2) was consistently defined as the positive outcome. Lymph nodes were not taken into account. P-values <0.05 were considered statistically significant. Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), and overall accuracies including 95% confidence intervals (CI) for identifying a good response were calculated using the predefined cut-off values derived from the results of the literature review.

## RESULTS

### 1.Literature Review

Fourteen papers met the final inclusion criteria. An overview of these studies is presented in **Table 2**. Two papers focused on the one-dimensional tumour size (method 1)<sup>15, 27</sup>. Two papers evaluated the 3-dimensional tumour size (method 2)<sup>26, 27</sup>. The majority of the papers (12/14) focused on whole tumour volume measurements (method 3). In 2 of the 14 studies, cut-off values used were predefined before onset of the study<sup>14, 27</sup>. In the other 12 studies, cut-off values were either retrospectively determined from the study dataset<sup>15-19, 21, 22, 26</sup> or it was not clearly reported how these were determined<sup>13, 20, 23, 25</sup>. The results (i.e. accuracy figures) of the different studies are given in **Table 3**.

**Table 2** Overview of literature investigating cut-off values for size or volume measurements to predict or assess a good response

Article (reference)	N	Inclusion criteria	Neoadjuvant therapy	Interval between CRT and surgery (wk)	Measurement	Cut-off	Outcome
Kim et al 2005 (13)*	112	cT3-4	50.4 Gy + 5FU + leucovorin	4-6	Volume	Pre-CRT 15 cm <sup>3</sup> Post-CRT 5 cm <sup>3</sup> Δ 70%	pCR, T-downstaging
Barbaro et al 2009 (14)†	53	cT3 >5 mm EMI, cT4, any cN, middle and low	50.4-55 Gy + oxaliplatin-based chemotherapy	6-8	Volume	Δ 70%	T-downstaging
Dresen et al 2009 (26)‡	67	MRF <2 mm, cT4	45-50.4 Gy + 5FU (±leucovorin)	8	3D size	Pre-CRT 50 cm <sup>3</sup> and Δ 70%	ypT0-2
Yeo et al 2010 (25)*	405	cT3-4, <9 cm of anal verge	50.4 Gy + 5FU + leucovorin OR capecitabine ± irinotecan	6	Volume	Δ 60% Δ 80%	pCR, TRG 1-2, ypT0-2
Kang et al 2010 (20)*	84	cT3-4 and/or cN+, <10 cm anal verge	50.4 Gy + 5FU + leucovorin OR capecitabine OR irinotecan	4-6	Volume	Post-CRT 5 mL Δ 75%	pCR
Curvo-Semedo et al 2011 (16)‡	50	cT3-4 and/or cN+	50.4 Gy + capecitabine	9-11	Volume	Pre-CRT 16.7 cm <sup>3</sup> Post-CRT 1.6 cm <sup>3</sup> Δ 88.6%	pCR
Nougaret et al 2012 (22)‡	58	cT3-4, <10 cm anal verge	45-63 Gy + capecitabine OR oxaliplatin + leucovorin + 5FU	4-6	Volume	Δ 70%	TRG 1-2
Lambrecht et al 2012 (21)‡	20	cT3-4, cN1-2	45-50.4 Gy + 5FU + oxaliplatin	6-9	Volume	Δ 77%	pCR
Patel et al 2012 (27)‡	78	MRF <2 mm, cT3-4 (<5 cm anorectal verge)	45 Gy + oxaliplatin + capecitabine	6-8	Length 3D size	Δ 30% Δ 60% Δ 80%	TRG 1-2, ypT0-3a
Genovesi et al 2013 (17)‡	28	cT3 and/or cN+	50.4 Gy + 5FU + cisplatin	8	Volume	Δ 72.1%	pCR
Ha et al 2013 (18)‡	100	cT3-4, any cN	50.4 Gy + 5FU OR capecitabine	7	Volume	Δ 84.3%	pCR
Suzuki et al 2013 (23)*	114	cT3-4Nx or cT2N+	40-45 Gy + uracil + tegafur OR S-1	6-8	Volume	Δ 70%	pCR, TRG 1-2, T-downstaging
Intven et al 2014 (19)‡	51	LARC	50 Gy + capecitabine	6-10	Volume	Post-CRT 14 cm <sup>3</sup> Δ 70%	pCR, TRG 1-2
Aiba et al 2014 (15)‡	40	CRM <1 mm, EMI >5 mm, cT4, or cN2	Oxaliplatin-based chemotherapy	3	Volume Length	Pre-CRT 29.3 cm <sup>3</sup> Post-CRT 10.2 cm <sup>3</sup> Δ 70% Pre-CRT 70 mm Post-CRT 46 mm	TRG 1-2

*Abbreviations:* 3D = 3-dimensional; 5FU = 5-fluorouracil; CRM = Circumferential resection margin; CRT = chemoradiation therapy; EMI = extramural invasion; LARC = Locally advanced rectal cancer; MRF = mesorectal fascia; pCR = pathologic complete response; TRG = tumor regression grade.  
 \* It was not reported how cut-offs were determined.  
 † Cut-offs were predefined.  
 ‡ Cut-offs were determined retrospectively using receive operating characteristic curves.

**Table 3** Diagnostic performance of all cut-off values as reported in the literature (left column) and as reported in the validation cohort (right column)

Cut-off (reference)	Reported results from the literature					Validation (n=148)				
	Sens	Spec	PPV	NPV	Accuracy	Sens	Spec	PPV	NPV	Accuracy
<b>Method 1: 1-dimensional measurements (tumor length)</b>										
Outcome = TRG 1-2 vs TRG 3-5										
Pre-CRT 70 mm (15)	94	46	53	92	65	92 (81-97)	13* (7-24)	47 (38-57)	64* (36-86)	49*
Post-CRT 46 mm (15)	100	46	55	100	68	93 (83-98)	10* (5-20)	47 (38-56)	64* (32-88)	48*
Δ 30% (27)	75	42	39	77	53	73 (60-83)	36 (25-49)	49 <sup>†</sup> (39-60)	61* (45-75)	53
<b>Method 2: 3-dimensional measurements (length × anterior–posterior diameter × left–right diameter)</b>										
Outcome = TRG 1-2 vs TRG3-5										
Δ 60% (27)	78	40	40	78	53	73 (60-83)	35 (24-47)	49 (38-60)	60* (43-75)	52
Δ 80% (27)	35	73	40	69	60	36 (24-49)	73 (60-82)	53 <sup>†</sup> (36-68)	57* (46-67)	56
Outcome = ypT0-2 vs ypT3-4										
Pre-CRT 50 cm <sup>3</sup> + Δ 70% (26)	100	75	42	100	79	31* (21-43)	76 (65-85)	58 <sup>†</sup> (41-73)	52* (42-62)	53*
<b>Method 3: Whole-volume (tumor area × slice thickness)</b>										
Outcome = complete response (ypT0) vs residual tumor										
Pre-CRT 15 cm <sup>3</sup> (13)	50	49	14	86	49	31* (16-51)	83 <sup>†</sup> (75-89)	31 <sup>†</sup> (16-49)	83 (75-89)	73 <sup>†</sup>
Pre-CRT 16.7 cm <sup>3</sup> (16)	43	83	50	79	72	38 (21-58)	79 (70-85)	31* (17-48)	84 (75-90)	71
Post-CRT 1.6 cm <sup>3</sup> (16)	64	94	82	87	86	14* (5-33)	97 (91-99)	50* (17-83)	82 (74-88)	80
Post-CRT 5 cm <sup>3</sup> (20)	44	74	28	85	68	48 (30-67)	76 (67-83)	33 (20-50)	86 (77-91)	71
Post-CRT 5 cm <sup>3</sup> (13)	63	41	15	87	44	48* (30-67)	76 <sup>†</sup> (67-83)	33 <sup>†</sup> (20-50)	86 (77-91)	71 <sup>†</sup>
Post-CRT 14 cm <sup>3</sup> (19)	100	69	30	100	73	100 (85-100)	30* (22-39)	26 (19-36)	100 (88-100)	44*
Δ 60% (25)	80	36	17	91	42	90 (72-97)	34 (26-44)	25 (17-35)	93 (80-98)	66 <sup>†</sup>
Δ 70% (13)	63	45	16	88	47	62.1 (42-79)	59 <sup>†</sup> (50-68)	27 <sup>†</sup> (17-40)	86 (76-93)	60 <sup>†</sup>
Δ 70% (23)	92	46	21	94	52	62* (42-79)	59 <sup>†</sup> (50-68)	27 (17-40)	86 (76-93)	60
Δ 70% (19)	83	78	33	97	78	62* (42-79)	59* (50-68)	27 (17-40)	86* (76-93)	60*
Δ 72.1% (17)	70	67	47	80	68	55* (36-73)	63 (54-72)	27* (17-41)	85 (75-92)	62
Δ 75% (20)	75	63	32	92	66	52* (33-70)	68 (58-76)	28 (17-43)	85 (76-91)	64
Δ 77% (21)	83	86	71	92	85	48* (30-67)	70* (62-79)	29* (17-44)	85 (76-91)	66*
Δ 80% (25)	43	76	23	89	72	45 (27-64)	80 (71-86)	35 <sup>†</sup> (21-53)	85 (77-91)	73
Δ 84.3% (18)	67	83	87	82	77	38* (21-58)	86 (78-92)	41* (23-61)	85 (77-91)	77
Δ 88.6% (16)	79	92	79	92	88	21* (9-40)	93 (87-97)	43* (19-70)	83 (75-88)	79
Outcome = T-downstaging (cT>ypT) vs no T-downstaging (cT≤ypT)										
Pre-CRT 15 cm <sup>3</sup> (13)	60	56	59	57	58	24* (16-35)	86 <sup>†</sup> (74-93)	69 (49-84)	46* (37-56)	51
Post-CRT 5 cm <sup>3</sup> (13)	72	52	61	64	63	33* (23-44)	76 <sup>†</sup> (64-86)	64 (48-78)	46* (36-56)	51*
Δ 70% (13)	67	55	60	61	61	52* (41-63)	64 (50-75)	65 (52-76)	50* (39-61)	57
Δ 70% (14)	78	88	94	64	81	52* (41-63)	64* (50-75)	65* (52-76)	50* (39-61)	57*
Δ 70% (23)	74	62	70	67	68	52* (41-63)	64 (50-75)	65 (52-76)	50* (39-61)	57*
Outcome = TRG 1-2 vs TRG 3-5										
Pre-CRT 29.3 cm <sup>3</sup> (15)	75	67	60	80	70	64* (51-76)	42* (30-55)	49* (37-60)	58* (43-72)	52*
Post-CRT 10.2 cm <sup>3</sup> (15)	75	88	80	84	73	76 (63-86)	44* (32-56)	54* (42-64)	68* (52-81)	59*
Post-CRT 14 cm <sup>3</sup> (19)	100	76	50	100	80	86* (75-94)	32* (22-44)	52 (42-62)	73* (54-87)	57*
Δ 60% (25)	76	38	34	79	49	81 (69-90)	35 (24-47)	52 <sup>†</sup> (41-62)	69* (51-83)	56
Δ 70% (22)	65	88	88	64	74	58 (44-70)	62* (50-74)	57* (43-69)	63 (51-74)	60*
Δ 70% (23)	80	56	53	81	65	58* (44-70)	62 (50-74)	57 (43-69)	63* (51-74)	60
Δ 70% (19)	78	81	47	94	82	58* (44-70)	62* (50-74)	57 (43-69)	63* (51-74)	60*
Δ 70% (15)	75	83	75	83	80	58* (44-70)	62* (50-74)	57* (43-69)	63* (51-74)	60*
Δ 80% (25)	35	77	39	74	65	36 (24-49)	83 (71-90)	64 <sup>†</sup> (45-79)	60* (49-70)	61

Abbreviations: CRT = chemoradiation therapy; NPV = negative predictive value; PPV = positive predictive value; Sens = sensitivity; Spec = specificity; TRG = tumor regression grade.

For the outcome parameter TRG1-2 versus TRG3-5 validation data were based on 128/146 study patients, since TRG scores were not available for 18/146 patients.

\* Validated performance is >10% lower than reported in the literature.

† Validated performance is >10% higher than reported in the literature.

## 2. Prospective validation study

### Patient and treatment characteristics

Of the 146 study patients 90 were male and 56 were female (mean age 64.6 years,  $\pm 10.8$ ). Mean interval between the completion of radiation and restaging MRI was 5.7 weeks ( $\pm 2.4$ ), and mean interval between the completion of radiation and surgery was 10 weeks ( $\pm 3.8$ ). Further patient and treatment details are given in **Table 4**.

Parameter	Total (n = 146)
Age (y)	64.6 $\pm$ 10.8
Sex	
Male	90 (61.6)
Female	56 (38.4)
Clinical T-stage	
cT2	2 (1.4)
cT3	122 (83.6)
cT4	22 (15.1)
MRF involvement	
No	61 (41.8)
Threatened (<2 mm)	38 (26)
Involvement	47 (32.2)
Distance from anorectal verge (cm)	
0-5	78 (53.4)
5-10	61 (41.8)
10-15	7 (4.8)
Clinical N-stage	
cN0	16 (11)
cN1	34 (23.3)
cN2	96 (65.8)
Neoadjuvant treatment	
45 Gy + 5FU	35 (24)
50.4 Gy + capecitabine	92 (63.0)
Other	19 (13.0)
Type of resection	
LAR	102 (69.9)
APR	38 (26.2)
Local excision	2 (1.4)
TEM	1 (0.7)
Extended surgery	2 (1.4)
Pathologic T-stage	
ypT0	29 (19.9)
ypT1	5 (3.4)
ypT2	41 (28.1)
ypT3	69 (47.3)
ypT4	2 (1.4)
Pathologic N-stage	
ypN0	99 (67.8)
ypN1	33 (22.6)
ypN2	14 (9.6)
Tumor regression grade according to Mandard	
1	30 (20.5)
2	29 (19.9)
3	47 (32.2)
4	18 (12.3)
5	4 (2.7)
Unknown	18 (12.3)
Good vs poor response	
pCR (ypT0)	29 (19.9)
No pCR	117 (80.1)
T-downstaging	83 (56.8)
No T-downstaging	63 (43.2)
TRG 1-2	59 (40.4)

(continued)

Parameter	Total (n = 146)
TRG 3-5	69 (47.3)
Unknown	18 (12.3)

Abbreviations: 5FU = 5-fluorouracil; APR = abdominoperineal resection; LAR = low anterior resection; MRF = mesorectal fascia; pCR = pathologic complete response; TEM = transanal endoscopic microsurgery; TRG = tumor regression grade.  
Values are number (percentage) or mean  $\pm$  SD.

### *Interobserver agreement*

Interobserver agreement for the one-dimensional measurements (method 1) was good pre-CRT (ICC 0.78) and post-CRT (ICC 0.66). For the 3-dimensional measurements (method 2) agreement was moderate pre-CRT (ICC 0.60) and good post-CRT (ICC 0.61). For the whole-volume measurements (method 3), agreement was excellent pre-CRT (ICC 0.94) and good post-CRT (ICC 0.65).

### *Tumour size and volume measurements.*

Mean tumour length (method 1) was 5.1 cm ( $\pm 1.5$ cm) before therapy and 3.1 cm ( $\pm 1.1$ cm) post-therapy, with a mean decrease ( $\Delta$ ) of -36.6% ( $\pm 20.2\%$ ). Mean volumes according to method 2 (length x AP x LR) were 57.9 cm<sup>3</sup> pre-CRT ( $\pm 50.6$  cm<sup>3</sup>), 17.0 cm<sup>3</sup> post-CRT ( $\pm 16.1$  cm<sup>3</sup>), and  $V_{\Delta}$  was -65.8% ( $\pm 23.9\%$ ). According to the whole volume measurements (method 3), mean volumes were 32.8 cm<sup>3</sup> pre-CRT ( $\pm 25.6$  cm<sup>3</sup>), 10.9 cm<sup>3</sup> post-CRT ( $\pm 11.3$  cm<sup>3</sup>), and  $V_{\Delta}$  was -64.6% ( $\pm 22.4\%$ ).

### *Diagnostic performance of predefined cut-off values and comparison with literature*

Diagnostic accuracy figures using the cut-off values derived from the literature review are provided in **Table 3**, together with those originally reported in literature.

[1] For the one-dimensional size measurements overall accuracy (vs. that originally reported in literature) to predict TRG1-2 ranged from 48-53% (vs. 53-68%), sensitivity was 73-93% (vs. 75-100%) and specificity 10-36% (vs. 42-46%); the outcomes ypT0 and T-downstaging were not reported/studied.

[2] For the 3-dimensional size measurements overall accuracy to predict TRG 1-2 was 52-56% (vs. 53-60%), sensitivity 36-73% (vs. 35-78%) and specificity 35-73% (vs. 40-73%); to predict ypT0-2 accuracy was 53% (vs. 79%), sensitivity 31% (vs. 100%) and specificity 76% (vs. 75%); the outcome ypT0 was not reported/studied.

[3] For the whole-volume measurements accuracy to predict a complete response (ypT0) was 44-80% (vs. 42-88%), sensitivity 14-100% (vs. 43-100%) and specificity 30-97% (vs. 36-94%); to predict T-downstaging accuracy was 51-57% (vs. 58-61%), sensitivity 24-52% (vs. 60-78%) and specificity 64-86% (vs. 52-88%); to predict TRG1-2 accuracy was 52-61% (vs. 49-82%), sensitivity 36-86% (vs. 35-100%) and specificity 32-83% (vs. 38-88%).

When comparing the results for the multicenter validation with those originally reported in literature, the prospectively calculated accuracy was >10% lower than originally reported in literature in 43% of the tested cut-offs, in 11% overall accuracy was >10% higher than originally reported and in the other 46% accuracy was similar (within a 10% range) to the originally reported data.

## DISCUSSION

Aim of this study was threefold. The first aim was to systematically review all available literature investigating the value of tumour size/volume measurements on MRI to predict and assess tumour response to CRT in patients with locally advanced rectal cancer. The second aim was to prospectively validate different cut-off values derived from the literature review in an independent multicenter cohort of rectal cancer patients treated with CRT. The third aim was to evaluate – based on the former – whether MRI volumetry is reliable to predict and assess the response of the primary tumour to CRT and which method and cut-off values would be most accurate.

As expected, the diagnostic performance of retrospectively determined cut-off values from literature decreased when tested ‘prospectively’ in an independent new cohort of patients. Of the fourteen papers included in the review, cut-off values had been determined retrospectively from the study data in the majority (8/14) of studies, while another four papers did not report how cut-off values had been defined. Only two papers had defined the cut-off values beforehand and already applied the cut-offs prospectively in their study. For 43% of the cut-offs that were tested, the overall accuracies in our prospective multicenter validation cohort were at least 10% lower compared to the results reported in the original studies. Similarly, the sensitivity and NPV were at least 10% lower than initially reported in 44% and 56%, respectively. The specificity and PPV on the other hand were typically similar or even >10% better than initially reported. As one would expect, a certain tradeoff between sensitivity and specificity was observed depending on the chosen cut-off values. A stricter cut-off will result in a higher specificity at the expense of a lower sensitivity and vice versa.

An important finding of our study is that both one-dimensional (i.e. tumour length) as well as 3-dimensional size measurements are insufficiently accurate to assess response. Overall accuracy for tumour length measurements in the review ranged between only 53%-68% and dropped to 48-53% in our validation cohort. For the three-dimensional size measurements (length x AP x LR), the overall accuracies showed a similar trend; moderate accuracies of 53-79% were reported in literature dropping to 52-56% when prospectively re-tested. In contrast, good results were reported for whole-tumour volume with overall accuracies of over 80% for both post-treatment and  $\Delta\%$  measurements, which could be reproduced in our validation cohort, with accuracies up to 79-80%. Moreover, interobserver agreement was considerably better for the whole-volume measurements (e.g. ICC was 0.94 pre-CRT versus 0.78 for tumour length and 0.60 for three-dimensional size

measurements). The whole-volume measurements offered the best results for assessment and prediction of complete response (ypT0) as the outcome variable, with overall accuracies ranging up to 80%. For the other outcome measures – tumour regression grade and T-downstaging – results were considerably poorer with accuracies of only 51-60%. Although we have to take into account that this discrepancy will also partly be influenced by the fact that a complete response is a less prevalent outcome measure than TRG1-2 or T-downstaging, it may also be related to intrinsic differences between the outcome measure ypT0 on the one hand, and TRG and T-downstaging on the other hand. The relation with the reduction of tumour volume is less straightforward in TRG and T-downstaging, than for a complete tumour response. A tumour can show a marked volume reduction but still be ypT3 because of a small remnant in the mesorectum. Likewise a tumour may shrink considerably but still be a non-responder on histological TRG assessment. Additionally, it is known that there is more interobserver variation in assessing the histological TRG, compared to assessing a complete regression or T-downstaging<sup>33</sup>. In our study we specifically focused on the response of the primary tumour itself and did not take into account the evolution in nodal status, which was beyond the scope of this study.

For whole-tumour volumetry, the most promising cut-off values to predict ypT0 – based both on literature and the validation results – were a volume of 1.6 cm<sup>3</sup> post-CRT and a volume reduction post-CRT of  $\Delta$ 80-86.6%. The accuracies reported in literature for these cut-offs ranged from 72%-88%, prospective validation resulted in comparable overall accuracies ranging from 73%-80%, with a corresponding sensitivity of 14-45% and specificity of 80%-97%. From a clinical point of view, it is particularly important to achieve a high specificity to ensure that no residual tumours are missed which would induce a risk for recurrence when considering patients for organ-preserving treatments. The high specificity of 80-97% in our validation cohort indicates that most patients with residual tumour will be correctly identified and patients have a <3-20% chance of being falsely classified as complete responders. However, corresponding sensitivities are relatively poor, indicating that many patients with a complete response will be missed and incorrectly diagnosed as having residual tumour. When selecting patients for organ-preservation after CRT using MR volumetry to support clinical selection tools would imply a very strict selection, since a substantial number of patients with a complete response might not be identified as such.

In daily radiological practice, a potentially limiting factor of whole-volume tumour volumetry is the fact that it is a time consuming procedure. Future work should therefore focus on the development of post-processing methods such as auto-segmentation techniques that may speed up the process of tumour delineation. In this regard, tumour volumetry using functional imaging techniques such as diffusion-weighted MRI (DWI) or FDG-PET have already been investigated with good sensitivities up to 70% for assessing a pCR with (semi-)automated segmentation<sup>34-36</sup>. The advantage is that both DWI and FDG-PET have superior strong contrast resolution, a prerequisite for automated tumour delineation.

Our study has some limitations. First, the validation part was performed using routinely acquired MRI data from 4 different hospitals. As such, there were variations in MRI hardware, scanning protocols and patient preparation. This approach however reflects daily clinical practice where protocols are subject to variations between different centres and therefore may render our results more generally applicable. Second, there was a subgroup of patients (n=19) that underwent an alternative neoadjuvant treatment protocol, potentially affecting the treatment response itself. Third, both in literature as well as in our validation cohort, there were variable intervals between imaging and surgery, potentially leading to the underestimation on imaging of the final histopathological response. Fourth, due to the time consuming nature of the study the second reader performed measurements on only a sample of patients to test for interobserver differences. Our interobserver agreement was, however, fully in line with various previously published results on IOA and we therefore believe that our results are representative<sup>16, 17, 22, 34</sup>. Finally, there were some variations in inclusion criteria in the different reports, from which the studied cut-off values were derived. Although overall the selection criteria used to establish our current study group corresponded well with those applied in literature, these slight variations may have affected the study outcome.

In conclusion, we have shown that MRI volumetry using whole tumour volume measurements may be helpful in rectal cancer response assessment when using selected cut-off values. Measurements of tumour length or three-dimensional tumour size are not helpful. MRI volumetry is mainly accurate to assess a complete tumour response (ypT0) after CRT, but less to assess T-downstaging or TRG1-2. Further research should focus on the development of automated segmentation techniques which may accelerate the process of tumour delineation and facilitate its implementation in clinical practice.

## REFERENCES

1. Heald RJ. A new approach to rectal cancer. *British journal of hospital medicine* 1979; 22(3): 277-81.
2. Kapiteijn E, Marijnen CA, Nagtegaal ID, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med* 2001; 345(9): 638-46.
3. Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004; 351(17): 1731-40.
4. Reerink O, Verschuuren RC, Szabo BG, Hospers GA, Mulder NH. A favourable pathological stage after neoadjuvant radiochemotherapy in patients with initially irresectable rectal cancer correlates with a favourable prognosis. *European journal of cancer* 2003; 39(2): 192-5.
5. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *The lancet oncology* 2010; 11(9): 835-44.
6. Lezoche G, Baldarelli M, Guerrieri M, et al. A prospective randomized study with a 5-year minimum follow-up evaluation of transanal endoscopic microsurgery versus laparoscopic total mesorectal excision after neoadjuvant therapy. *Surgical endoscopy* 2008; 22(2): 352-8.
7. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011; 29(35): 4633-40.
8. Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Annals of surgery* 2004; 240(4): 711-7; discussion 7-8.
9. Beets-Tan RG, Lambregts DM, Maas M, et al. 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* 2013; 23(9): 2522-31.
10. 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.
11. Vliegen RF, Beets GL, Lammering G, et al. Mesorectal fascia invasion after neoadjuvant chemotherapy and radiation therapy for locally advanced rectal cancer: accuracy of MR imaging for prediction. *Radiology* 2008; 246(2): 454-62.
12. Barbaro B, Vitale R, Leccisotti L, et al. Restaging locally advanced rectal cancer with MR imaging after chemoradiation therapy. *Radiographics : a review publication of the Radiological Society of North America, Inc* 2010; 30(3): 699-716.
13. Kim YH, Kim DY, Kim TH, et al. Usefulness of magnetic resonance volumetric evaluation in predicting response to preoperative concurrent chemoradiotherapy in patients with resectable rectal cancer. *International journal of radiation oncology, biology, physics* 2005; 62(3): 761-8.
14. Barbaro B, Fiorucci C, Tebala C, et al. Locally advanced rectal cancer: MR imaging in prediction of response after preoperative chemotherapy and radiation therapy. *Radiology* 2009; 250(3): 730-9.
15. Aiba T, Uehara K, Nihashi T, et al. MRI and FDG-PET for assessment of response to neoadjuvant chemotherapy in locally advanced rectal cancer. *Ann Surg Oncol* 2014; 21(6): 1801-8.



16. Curvo-Semedo L, Lambregts DM, Maas M, 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.
17. Genovesi D, Filippone A, Ausili Cefaro G, et al. Diffusion-weighted magnetic resonance for prediction of response after neoadjuvant chemoradiation therapy for locally advanced rectal cancer: preliminary results of a monoinstitutional prospective study. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 2013; 39(10): 1071-8.
18. 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(12): 3345-53.
19. Intven M, Reerink O, Philippens ME. Dynamic contrast enhanced MR imaging for rectal cancer response assessment after neo-adjuvant chemoradiation. *J Magn Reson Imaging* 2015; 41(6): 1646-53.
20. Kang JH, Kim YC, Kim H, et al. Tumour volume changes assessed by three-dimensional magnetic resonance volumetry in rectal cancer patients after preoperative chemoradiation: the impact of the volume reduction ratio on the prediction of pathologic complete response. *International journal of radiation oncology, biology, physics* 2010; 76(4): 1018-25.
21. Lambrecht M, Vandecaveye V, De Keyzer F, et al. Value of diffusion-weighted magnetic resonance imaging for prediction and early assessment of response to neoadjuvant radiochemotherapy in rectal cancer: preliminary results. *International journal of radiation oncology, biology, physics* 2012; 82(2): 863-70.
22. Nougaret S, Rouanet P, Molinari N, et al. MR volumetric measurement of low rectal cancer helps predict tumour response and outcome after combined chemotherapy and radiation therapy. *Radiology* 2012; 263(2): 409-18.
23. Suzuki T, Sadahiro S, Tanaka A, et al. Relationship between histologic response and the degree of tumour shrinkage after chemoradiotherapy in patients with locally advanced rectal cancer. *Journal of surgical oncology* 2014; 109(7): 659-64.
24. van Stiphout RG, Valentini V, Buijsen J, et al. Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: A multicentric prospective study with external validation. *Radiother Oncol* 2014; 113(2): 215-22.
25. Yeo SG, Kim DY, Kim TH, et al. Tumour volume reduction rate measured by magnetic resonance volumetry correlated with pathologic tumour response of preoperative chemoradiotherapy for rectal cancer. *International journal of radiation oncology, biology, physics* 2010; 78(1): 164-71.
26. Dresen RC, Beets GL, Rutten HJ, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part I. Are we able to predict tumour confined to the rectal wall? *Radiology* 2009; 252(1): 71-80.
27. Patel UB, Brown G, Rutten H, et al. Comparison of magnetic resonance imaging and histopathological response to chemoradiotherapy in locally advanced rectal cancer. *Ann Surg Oncol* 2012; 19(9): 2842-52.

28. Rosset A, Spadola L, Ratib O. OsiriX: an open-source software for navigating in multidimensional DICOM images. *J Digit Imaging* 2004; 17(3): 205-16.
29. Quirke P, Durdey P, Dixon MF, Williams NS. Local recurrence of rectal adenocarcinoma due to inadequate surgical resection. Histopathological study of lateral tumour spread and surgical excision. *Lancet* 1986; 2(8514): 996-9.
30. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumour regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994; 73(11): 2680-6.
31. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis* 1997; 12(1): 19-23.
32. Santos MD, Silva C, Rocha A, Matos E, Nogueira C, Lopes C. Prognostic value of mandard and dworak tumour regression grading in rectal cancer: study of a single tertiary center. *ISRN surgery* 2014; 2014: 310542.
33. Chetty R, Gill P, Govender D, et al. International study group on rectal cancer regression grading: interobserver variability with commonly used regression grading systems. *Human pathology* 2012; 43(11): 1917-23.
34. Carbone SF, Pirtoli L, Ricci V, et al. Assessment of response to chemoradiation therapy in rectal cancer using MR volumetry based on diffusion-weighted data sets: a preliminary report. *La Radiologia medica* 2012; 117(7): 1112-24.
35. Chen SW, Chen WT, Wu YC, et al. Which FDG/PET parameters of the primary tumours in colon or sigmoid cancer provide the best correlation with the pathological findings? *European journal of radiology* 2013; 82(9): e405-10.
36. Lambregts DM, Rao SX, Sassen S, 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. *Annals of surgery* 2014; Epub ahead of print.





# Chapter 6

Automated and semi-automated  
segmentation of rectal tumor volumes on  
diffusion-weighted MRI:  
can it replace manual volumetry?

MM van Heeswijk, DMJ Lambregts, JJM van Griethuysen,  
SA Oei, S-X Rao, CAM de Graaff, RFA Vliegen, GL Beets,  
N Papanikolaou, RGH Beets-Tan

*International Journal of Radiation Oncology, Biology, Physics. 2016 Mar 15;94(4):824-31.*

## ABSTRACT

### Purpose

Diffusion-weighted imaging (DWI) tumour volumetry is promising for rectal cancer response assessment, but an important drawback is that manual per-slice tumour delineation can be highly time consuming. Aim was to investigate whether manual DWI-volumetry can be reproduced using a (semi-)automated segmentation approach.

### Methods and materials

Seventy-nine patients underwent MRI including DWI (highest b-value b1000/b1100) pre- and post-chemoradiotherapy (CRT). Tumour volumes were assessed on b1000/b1100 DWI pre-/post-CRT by means of: [1] automated segmentation (by 2 inexperienced readers), [2] semi-automated segmentation (manual adjustment of the volumes obtained by method 1 by two radiologists), and [3] manual segmentation (by 2 radiologists), the latter which served as the reference standard. Intraclass correlation coefficients (ICC) and dice similarity indices (DSI) were calculated to evaluate agreement between different methods and observers. Measurement times (from a radiologists' perspective) were recorded for each method.

### Results

Tumour volumes were not significantly different between the three methods, neither pre- nor post-CRT (P-values 0.08-0.92). ICCs compared to manual segmentation were 0.80-0.91 / 0.53-0.66 pre-/post-CRT for the automated segmentation and 0.91-0.97 / 0.61-0.75 for the semi-automated method. Interobserver agreement (ICC) pre-/post-CRT was 0.82/0.59 for automated segmentation, 0.91/0.73 for semi-automated segmentation and 0.91/0.75 for manual segmentation. Mean DSI between the automated and semi-automated method was 0.83/0.58 pre-CRT/post-CRT; DSI between the automated and manual segmentation was 0.68/0.42 and between the semi-automated and manual 0.70/0.41. Median measurement time for the radiologists was 0 sec (pre- and post-CRT) for the automated method, 41-69 sec (pre-CRT) and 60-67 sec (post-CRT) for the semi-automated method, and 180-296 sec (pre-CRT) and 84-91 sec (post-CRT) for the manual method.

### Conclusions

DWI-volumetry using a semi-automated segmentation approach is promising and a potentially time saving alternative to manual tumour delineation, particularly for primary tumour volumetry. Once further optimized it could be a helpful tool for tumour response assessment in rectal cancer.

## INTRODUCTION:

Response assessment to neoadjuvant chemoradiotherapy (CRT) is an increasingly discussed issue in the treatment of rectal cancer. Previously, long course CRT (for patients with locally advanced rectal cancer) was routinely followed by total mesorectal excision, regardless of the treatment response. Organ-preserving alternatives such as a local excision or a “watch-and-wait” policy are, however, gaining ground for selected patients who respond very well to CRT, albeit so far mainly performed in trial settings<sup>1,2</sup>. Apart from digital rectal examination and endoscopy, imaging – in particular MRI – can play an important role in the evaluation of response. In a recent meta-analysis it was shown that for re-assessment of T-stage after CRT, morphological MRI has a high specificity of 91% but limited sensitivity of 50%. In the same meta-analysis, use of diffusion-weighted MRI (DWI) increased the sensitivity of MRI to 84%<sup>3</sup>. The routine use of DWI for tumour response assessment is now even recommended in a recent consensus guideline from the European Society of Gastrointestinal and Abdominal Radiology (ESGAR)<sup>4</sup>. There are several ways in which DWI can be used to assess response. The first and most basic approach is to visually assess high b-value images for the presence of a remaining high signal (indicating residual tumour) after CRT, which can significantly improve the ability of MRI to differentiate residual tumour within post-radiation fibrosis<sup>5</sup>. Second, one can quantitatively measure tumour “apparent diffusion coefficient” (ADC) values, although varying success rates have been reported for this approach<sup>6-9</sup>. A third approach is to measure the volume of the high signal areas (i.e. tumour) on DWI, which has shown great promise to identify patients with a complete tumour response<sup>6, 10-12</sup>. However, an important drawback of DWI-volumetry is that measurements are typically performed manually, i.e. by tracing the tumour boundaries on DWI on every single slice, which can add up to well over 30 slices in large tumours. Although this method has been found to be precise and well reproducible<sup>6, 10</sup>, it is time consuming and therefore not likely to find its way into daily practice. As such we need to investigate ways to facilitate the assessment of tumour volumes on DWI and speed up the measurement process if we truly want to be able to use DWI volume measurements as an imaging biomarker for treatment response. Aim of this study was to investigate whether manual DWI volume measurements can be accurately reproduced using a (semi-)automated segmentation software tool.

## MATERIALS & METHODS:

### Patients

This study included a random sample of 79 patients from a previously reported bi-institutional study cohort<sup>10</sup>. All patients had biopsy proven rectal adenocarcinoma, and were treated with a long course of neoadjuvant CRT. The routine neoadjuvant treatment scheme consisted of 50.4 Gy radiation combined with 2x825 mg/m<sup>2</sup>/d capecitabine. Patients underwent a primary staging MRI before CRT and a second restaging MRI after CRT, both including a DWI sequence. The restaging MRI was routinely performed 6-8 weeks after completion of CRT.

### MR imaging

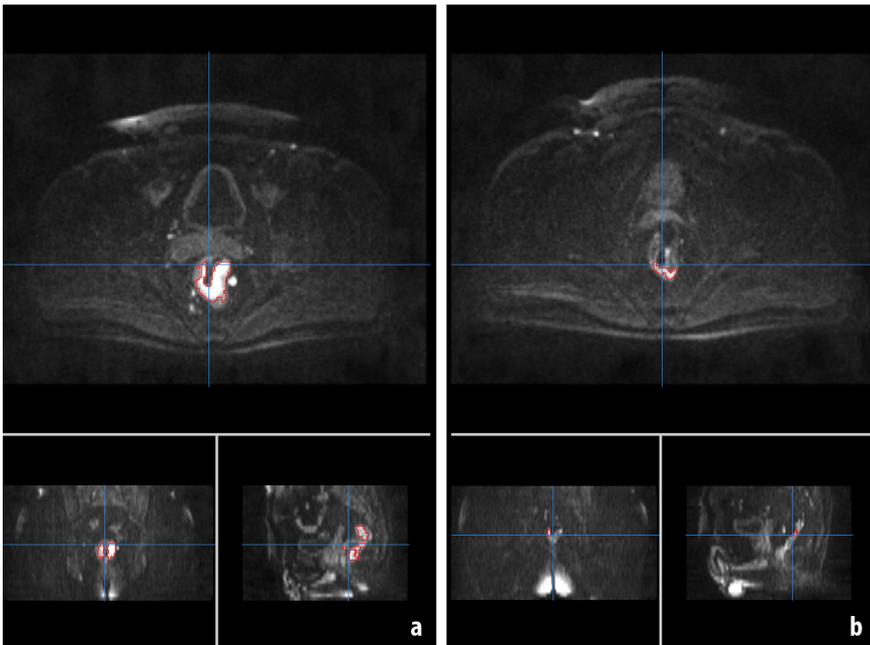
MR imaging was performed at 1.5T (using either an Intera (Achieva) or Ingenia system, Philips Medical Systems, Best, The Netherlands (centre 1) or a Siemens Magnetom Avanto system, Siemens Healthcare, Forchheim, Germany (centre 2)) with a phased array surface coil. The DWI sequence was angled axially, perpendicular to the tumour axis as defined on a sagittal planning scan. The highest b-value was 1000-1100. Detailed DWI sequence parameters are given in Supplementary material 1<sup>10</sup>. The further MRI protocol consisted of standard 2-dimensional T2-weighted (T2W) fast spin-echo sequences in 3 orthogonal directions, of which the axial T2W sequence was angled identical to the DWI. Patients did not receive bowel preparation or rectal distension.

### Image evaluation / volumetric measurements

MR images were transferred in DICOM format and evaluated on an offline workstation. Measurements were performed using a freely distributed software tool (MANGO; Multi-image Analysis GUI, version 2.6, Research Imaging Institute, University of Texas Health Science Centre, San Antonio, USA). For each patient, tumour volumes were measured on the b1000-11000 DWI images on the primary staging MRI and on the restaging MRI. The corresponding T2-weighted images were at the readers' disposal for anatomical reference. Window-level settings could be adjusted at the discretion of the readers. Measurements were performed in threefold, by means of: [1] 'automated' tumour segmentation, [2] 'semi-automated' tumour segmentation, and [3] manual tumour segmentation, the latter which served as the standard of reference. Time required to finalize the measurements (in seconds) was recorded for each method. For each reading session, the respective observers were blinded to each other's results, clinical patient data and histopathological outcome.

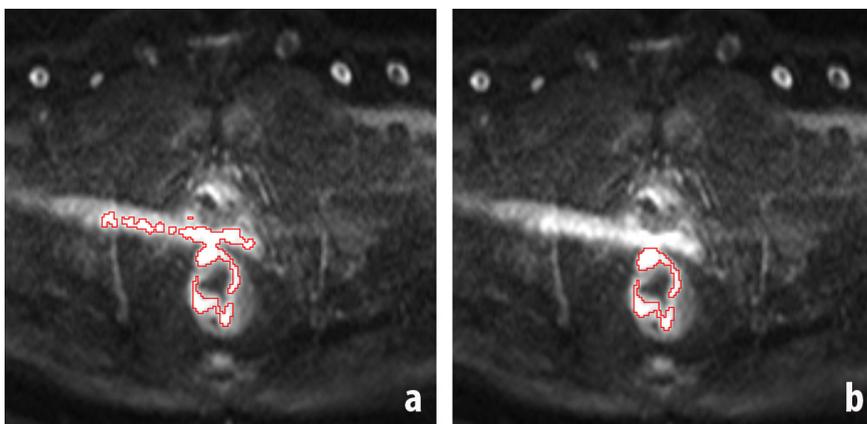
## 1. Automated tumour segmentation

Two independent readers without any specific experience in reading pelvic MRI performed the ‘automated’ segmentation. Both readers received a short training before the start of the study on rectal tumour assessment on DWI and specifically on how to recognize tumoural signal intensities, other organs and structures (e.g. prostate, lymph nodes) and artefacts such as susceptibility effects. Using a magic wand tool, the readers ‘scanned’ the range of signal intensities present within the suspected tumour area to identify the minimum and maximum signal intensity values. The lower boundary of the signal intensity range to be included in the automated segmentation was set at the minimum identified signal intensity with an additional margin of 5% (i.e. in case of a minimum signal intensity of 100, the lower boundary was set at 95); the upper boundary was set at “infinitely high”. All voxels within the defined signal intensity range were then automatically segmented by means of region-growing in 3D on all slices (**Figure 1**). Total volumes (cm<sup>3</sup>) of the acquired volumes of interest (VOI) were automatically calculated by the software.



**Figure 1.** Example of automatically generated 3D tumour segmentation before (a) and after chemoradiotherapy (b) in a male patient with a tumour in the mid-rectum.

When obvious artefacts (i.e. susceptibility artefacts) or other organs outside the rectum (such as the prostate) were included in the VOIs, these were manually erased by the readers (**Figure 2**). VOIs were then saved and served as the basis for the semi-automated measurements.



**Figure 2.** Example of erroneous inclusion of a fat suppression artefact in the volume of interest by the automated segmentation method (a). The artefact was manually excluded from the VOI (b).

## 2. Semi-automated tumour segmentation

Two experienced radiologists independently performed the ‘semi-automated’ measurements. The readers uploaded the VOIs acquired by method 1, after which they could manually adjust the segmented areas per slice as they saw fit. After manual adjustment of the VOIs, tumour volumes ( $\text{cm}^3$ ) were again automatically calculated by the software.

## 3. Manual tumour segmentation

For the manual segmentation two radiologists manually delineated the high signal tumour areas on each axial slice containing tumour to obtain the cross sectional tumour areas, which were multiplied by slice thickness (including the interslice gap) to obtain the total tumour volume ( $\text{cm}^3$ ).

## Statistical Analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, version 21.0 Inc, Chicago, IL). Volumes obtained with the different

measurement methods were compared using the Wilcoxon Signed Rank test (since data was not normally distributed). P-values <0.05 were considered statistically significant. The intraclass correlation coefficient (ICC) for single measures was calculated (using a two-way mixed model with absolute agreement) to determine agreement between the two readers for all three methods, as well as the agreement of the automated and semi-automated segmented volumes, respectively, compared with the manually segmented volumes as the standard of reference. ICC is a measure representing agreement between different observers or evaluation methods on the same set of continuous data (i.e. volume in cm<sup>3</sup> for the current study), with 0-0.20 indicating poor, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good, and 0.81-1.0 excellent agreement. Finally, Dice Similarity Indices (DSIs) were calculated to compare the spatial overlap of the VOIs generated by the different individual observers for the different segmentation methods<sup>13</sup>. DSIs were calculated using 3D-slicer, version 4.4.0<sup>14</sup> using the module 'Dice Computation'<sup>15</sup>. The DSI was defined as  $DSI = 2(A \cap B) / (A + B)$  in which A and B represent the different methods or observers, respectively. The DSI ranges from 0 (no spatial overlap between VOIs) and 1 (complete overlap between VOIs).

## RESULTS

### Patients

56 patients were male, 23 were female. Median age was 65 years (range 43-86). The initial tumour stage was cT1-2 (n=7), cT3 (n=62) and cT4 (n=10); nodal stage was N0 (n=4), N1 (n=25) and N2 (n=50); the mesorectal fascia was involved ( $\leq 1$  mm) in 49 and non-involved in 30 patients. 75 patients underwent total mesorectal excision (n=15 ypT0, n=5 ypT1, n=16 ypT2, n=39 ypT3, and n=4 ypT4). The remaining 4 patients had clinical evidence of a complete response after CRT and were followed according to a wait-and-see policy with a median disease free follow-up period of 41 months (range 27-50). These patients were considered yT0.

### Volumetry results

**Table 1** shows the median tumour volumes for the three methods. There were no significant differences in volumes between the three different approaches (P 0.08-0.92). **Table 2** shows the ICCs for agreement of the automated/semi-automated segmentation compared with the manual segmentation and agreement between observers. Agreement of the automated segmentation compared to manual segmentation was good-excellent pre-CRT (ICC 0.80-0.91) and moderate-good post-CRT (ICC 0.53-0.66). For the semi-automated segmentation, agreement with

**Table 1.** Tumour volumes obtained by the 3 different measurement methods.

	<b>Pre-CRT volume (range) (cm<sup>3</sup>)</b>	<b>Post-CRT volume (range) (cm<sup>3</sup>)</b>
Automated segmentation	19.7 (2.0 - 198.2)	1.9 (0.0 – 15.9)
Semi-automated segmentation	18.0 (2.4 - 97.4)	1.9 (0.0 – 27.8)
Manual segmentation	17.4 (0.4 – 185.7)	1.5 (0.0 – 16.4)

Numbers are medians (for the value averaged between the 2 observers). Comparison of the 3 measurement methods using the Wilcoxon signed rank test revealed no significant differences among the 3 measurement methods (P = .08-.92 pre-CRT; P = .80-.92 post-CRT)

**Table 2.** ICCs for assessment of agreement between the automated and semi-automated segmentation with manual segmentation as the reference standard, as well as agreement between observers.

<b>Parameters</b>	<b>Pre-CRT</b>	<b>Post-CRT</b>
Correlation with manual segmentation*		
Automated segmentation	0.80-0.91	0.53-0.66
Semi-automated segmentation	0.91-0.97	0.61-0.75
Correlation between observers		
Automated segmentation	0.82 (0.73-0.88)	0.59 (0.42-0.71)
Semi-automated segmentation	0.91 (0.86-0.94)	0.73 (0.60-0.82)
Manual segmentation	0.91 (0.86-0.94)	0.75 (0.63-0.83)

ICC = Intraclass Correlation Coefficient. ICCs represent results for single measurements. 95% confidence intervals are given in parentheses.

\* = range of ICCs between the two methods for the different observers.

manual segmentation was excellent pre-CRT (ICC 0.91-0.97) and good post-CRT (ICC 0.61-0.75). Interobserver agreement was excellent for all three methods (ICCs 0.82-0.91) pre-CRT. Post-CRT, interobserver agreement was good for the manual and semi-automated segmentation (ICC 0.75 and 0.73) and moderate for the automated segmentation (ICC 0.59).

### Dice similarity indices

**Table 3** shows the DSIs between the different segmentation methods and between observers. DSI between the automated/semi-automated approaches was excellent pre-CRT (DSI 0.83) and moderate post-CRT (DSI 0.58); between the automated/manual approaches DSI was good pre-CRT (DSI 0.68) and moderate post-CRT (DSI 0.42); between the semi-automated/manual approaches DSI was good pre-CRT (DSI 0.70) and moderate post-CRT (DSI 0.41). DSIs between observers were

0.76/0.50 pre-/post-CRT for the automated segmentation, 0.85/0.57 for the semi-automated segmentation, and 0.68/0.53 for the manual segmentation. In supplementary material 2, DSIs are given separately for tumours in the low, mid, and high rectum.

**Table 3.** Dice similarity indices ( $\pm$  SD) for the automated and semi-automated segmentations

DSI	Pre-CRT	Post-CRT
Automated vs. semi-automated (all observers)	0.83 $\pm$ 0.19	0.58 $\pm$ 0.33
Automated vs. manual (all observers)	0.68 $\pm$ 0.15	0.42 $\pm$ 0.23
Semi-automated vs. manual (all observers)	0.70 $\pm$ 0.14	0.41 $\pm$ 0.22
Observer 1 vs. observer 2 (automated)	0.76 $\pm$ 0.20	0.50 $\pm$ 0.28
Observer 1 vs. observer 2 (semi-automated)	0.85 $\pm$ 0.19	0.57 $\pm$ 0.30
Observer 1 vs. observer 2 (manual)	0.68 $\pm$ 0.11	0.53 $\pm$ 0.11

Numbers are means, standard deviations (SD) are given in parentheses.

### Effects on timing

Measurement times for the different segmentation methods (from a radiologist perspective) are given in **Table 4**. For manual segmentation, median measurement time was 180-296 sec pre-CRT and 84-91 sec post-CRT. For the semi-automated segmentation, median measurement time was 41-69 sec pre-CRT (median decrease of 96-236 sec,  $P < 0.001$ ) and 60-67 sec post-CRT (median decrease of 20-21 sec,  $P = 0.001-0.03$ ). Since the automated measurements were performed by non-radiological readers, the time required for the radiologists themselves was 0 seconds.



**Table 4.** Median time in seconds required from the radiologists' perspective to finalize the segmentations

	Pre-CRT		Post-CRT	
	Obs 1	Obs 2	Obs 1	Obs 2
Automated segmentation	0	0	0	0
Semi-automated segmentation	69 (17-251)	41 (19-300)	67 (0-220)	60 (15-370)
Manual segmentation	180 (60-438)	296 (61-1118)	91 (20-191)	84 (11-383)
<b>Decrease in time required compared to manual segmentation</b>				
Automated segmentation	180 (60-438)*	296 (61-1118)*	91 (20-191)*	84 (11-383)*
Semi-automated segmentation	96 (-104-290)*	236 (-125-1085)*	21 (-110-130)*	20 (-172-346)*

Numbers are medians (seconds); ranges are given in parentheses. Time measurements for the automated and semi-automated segmentations were compared to the manual segmentation using the Wilcoxon Signed Rank test.

\* indicates a significant difference compared to the manual segmentation method ( $P < 0.05$ ).

## DISCUSSION

Aim was to investigate whether (semi-)automated segmentation tools can replace manual tumour segmentation to measure rectal tumour volumes on diffusion-weighted MRI. Results show that particularly a semi-automated approach – where automatically generated volumes are manually adjusted by a radiologist – can accurately reproduce manual measurements with good-excellent interobserver agreement, whilst resulting in a median decrease in median measurement time of up to 4minutes per patient pre-CRT.

Our results are in line with those of Carbone and colleagues who used a semi-automated approach to delineate rectal tumour volumes on DWI in a small group of 14 patients with excellent interobserver agreement (ICC 0.96)<sup>12</sup>. Encouraging results have also been suggested for auto-segmentation on DWI in metastatic bone lesions<sup>16</sup>, brain tumours<sup>17</sup> and prostate cancer<sup>18</sup>. An important aim of (semi-) automated tumour segmentation is to reduce measurement times. Manual segmentation requires delineation of tumour per slice, which is time consuming, particularly in large and irregularly shaped tumours. In our cohort, measurement time ranged from 60 to 1118 seconds ( $\pm 18.5$  minutes) for the manual delineation of primary tumours. Post-CRT, measurement time was considerably shorter (20-383 seconds), which is logical since tumour remnants after CRT are considerably smaller

than tumours at primary staging. The 'automated' segmentation was performed by two inexperienced readers, who were merely asked to select a high signal area within the bowel wall/tumour after which the software automatically segmented and calculated the total tumour volume. The only additional action from their part was the manual exclusion of obvious incorrectly included structures such as other organs, lymph nodes or artefact areas (**Figure 2**). Readers without any experience in MR pelvis were deliberately chosen because – if such an approach works – this would imply that we can allow non-radiological readers (i.e. MRI technicians) to perform segmentations, which would free the time of the radiologist. For the semi-automated approach – where the radiologists had to manually adjust the automatically generated VOIs – the median measurement time decreased by 96-236 seconds for the two readers pre-CRT. After CRT median measurement time reduced by only 20-21 seconds. Albeit statistically significant, this is not a relevant reduction in measurement time from a clinical perspective. Although overall the post-CRT segmentation required less time, results were considerably less reproducible compared to the pre-CRT measurements. Pre-CRT, agreement of both the automated and semi-automated approach compared to manual segmentation was excellent (ICC 0.80-0.91 and 0.91-0.97) and the DSIs between the different methods were good to excellent (DSI 0.68-0.83). After CRT, ICCs of the semi-automated segmentation compared to manual segmentation were considerably better than for the automated method (ICC 0.61-0.75 versus 0.53-0.66). DSIs between the different methods were only 0.41-0.58. The lower DSIs after CRT can partly be explained by the fact that DSIs will always be lower when delineating smaller volumes. In addition, ICCs and DSIs post-CRT are lower because after CRT it is more difficult to determine which areas still represent tumour. Remaining high signal areas are often much less conspicuous and sometimes scattered throughout the fibrosis. This also explains our lower interobserver agreement after CRT, a finding concordant with previous reports<sup>6, 10</sup>. Results, including time efficacy, may be further enhanced in the future by optimization of autosegmentation algorithms, improvement of DWI image quality (e.g. reduce artefacts) and integration of the measurement tools into clinically used PACS viewing systems. An important clinically relevant application of diffusion-weighted tumour volumetry is to predict which patients will respond well to chemoradiotherapy as this knowledge would allow for further tailoring of treatment with one of the options being organ preserving therapy. Initial retrospective reports in patient cohorts of 14 to up to 100 patients consistently showed good results for DWI-volumetry in analysing response to CRT<sup>6, 11, 12</sup>. Carbone et al. showed a very good cor-

relation between post-CRT DWI-volumetry and histopathological tumour regression grade (TRG) of  $\rho=0.847$ , which was significantly better compared to the correlation between volumetry performed using standard T2-weighted images ( $\rho=0.597$ )<sup>12</sup>. The groups of Curvo-Semedo<sup>6</sup> and Ha<sup>11</sup> both reported very good results for DWI-volumetry in selecting complete responding rectal tumours. Particular good results were obtained for the DWI-volume post-CRT and the relative decrease in DWI volume ( $\Delta$ volume) with areas under the Receiver Operator Characteristics (ROC) curve of up to 0.91 and 0.93, results superior to those of both tumour volumetry using standard T2-weighted MRI as well as to quantitative ADC measurements. In a recent report, DWI volume thresholds retrospectively determined in the study by Curvo-Semedo et al. were prospectively tested in a new cohort of 112 patients, which confirmed the good results of DWI-volumetry with a prospectively calculated sensitivity of 70% and specificity of 98% to identify the complete responders<sup>10</sup>. Another potential application of DWI-volumetry is image guided radiotherapy planning. Regini et al. used diffusion-weighted images for delineation of the gross target volume (GTV), after which a software tool automatically calculated 3D tumour volumes for radiotherapy planning. They concluded that use of DWI (compared to T2W-MRI) resulted in significantly smaller GTVs, which may be attributed to better lesion conspicuity on DWI<sup>19</sup>. In a preliminary communication from van Brussel et al. diffusion images were successfully used for dose painting in a small group of 5 rectal cancer patients<sup>20</sup>. This is a potentially interesting application of DWI that should be further explored.

There are some limitations to our study design. First, the segmentations were performed on an offline workstation using a software tool that was not linked to the hospital's PACS system. Should DWI-volumetry find its way into clinical practice, (auto-)segmentation should preferably be performed using a tool that can be integrated into the clinical PACS system. Also, the auto-segmentation algorithms used in our study should be compared with other segmentation methods. Second, there were some variations in acquisition parameters of the DWI protocols used throughout the course of the study, which may have introduced some variations in scan quality and signal to noise ratios which may have affected the study results. Moreover, window-level settings were not standardised. On the other hand, such variations reflect daily practice and our favourable results despite standardisation of such factors suggest that our results may be generalized to daily practice. Third, our time measurements may not be fully representative for daily practice, since readers who measure many consecutive patients in a study setting will tend to be faster than readers in clinical practice who only perform these measurements

'every once in a while'. Fourth, despite good ICCs of 0.91 pre-CRT and 0.75 post-CRT, the DSIs for the manual segmentation (the standard of reference) were relatively low (0.68 pre-CRT and 0.53 post-CRT). This effect has been previously reported and makes sense, since manual segmentations are obtained by subjective interpretation as opposed to the more standardised computer algorithms used for the automated segmentations<sup>21</sup>. Finally, segmentation was performed using high b-value (b1000-1100) DWI. We chose this approach, as the high signal to background ratio makes it the most suitable method for automated segmentation. There are, however, potential drawbacks that need to be taken in to account: high signal caused by T2 shine-through effects, surrounding structures such as the prostate or lymph nodes and artefacts (e.g. susceptibility effects) may erroneously be included in the automated VOIs, as illustrated in **figure 2**. These effects will need to be corrected.

In conclusion, we have shown that (semi-)automated segmentation tools can accurately reproduce rectal tumour volumes as measured on DWI by manual delineation. At primary staging, the best results are obtained with semi-automated segmentation (in which automatically generated volumes are manually adjusted by a radiologist): agreement compared to manual segmentation is excellent whilst resulting in a median decrease in measurement time of up to 4 minutes per patients. Acceptable results can also be achieved at primary staging with an automated approach, which results in a 100% decrease in measurement time for a radiologist, since measurements can be performed by non-radiological readers. After chemoradiotherapy, a semi-automated approach offers the best results and more accurately reproduces manually delineated tumour volumes compared to an automated approach, although it does not result in a substantial reduction in measurement time ( $\pm 20$  sec). Overall, semi-automated tumour volume measurements could be a good alternative to manual DWI tumour volumetry, the latter of which has been proven to be an accurate imaging tool to assess a complete response to CRT, yet not practical because it is time consuming. Future studies should focus on further optimising DWI image acquisition and segmentation algorithms as well as on integrating these tools into clinical viewing systems to allow implementation in daily practice and further improve time efficacy.



## REFERENCES

1. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011;29:4633-4640.
2. Habr-Gama A, Gama-Rodrigues J, Sao Juliao GP, 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;88:822-828.
3. 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;269:101-112.
4. Beets-Tan RG, Lambregts DM, Maas M, et al. 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* 2013;23:2522-2531.
5. Kim SH, Lee JM, Hong SH, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumour response to neoadjuvant chemo- and radiation therapy. *Radiology* 2009;253:116-125.
6. Curvo-Semedo L, Lambregts DM, Maas M, 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;260:734-743.
7. Sun YS, Zhang XP, Tang L, et al. Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MRI for early detection of tumour histopathologic downstaging. *Radiology* 2010;254:170-178
8. Lambrecht M, Vandecaveye V, De Keyzer F, et al. Value of diffusion-weighted magnetic resonance imaging for prediction and early assessment of response to neoadjuvant radiochemotherapy in rectal cancer: preliminary results. *Int J Radiat Oncol Biol Phys* 2012;82:863-870
9. Kim SH, Lee JY, Lee JM, Han JK, Choi BI. Apparent diffusion coefficient for evaluating tumour response to neoadjuvant chemoradiation therapy for locally advanced rectal cancer. *Eur Radiol* 2011;21:987-995
10. Lambregts DM, Rao SX, Sassen S, 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* 2014, Sep 10 [ePub ahead of print], DOI 10.1097/SLA.0000000000000909
11. 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;23:3345-3353.
12. Carbone SF, Pirtoli L, Ricci V, et al. Assessment of response to chemoradiation therapy in rectal cancer using MR volumetry based on diffusion-weighted data sets: a preliminary report. *Radiol Med* 2012;117:1112-1124.
13. Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. *Strahlenther Onkol* 2012;188:160–167



14. Fedorov A., Beichel R., Kalpathy-Cramer J., et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012 ;30:1323-1341.
15. Bharatha A, Hirose M, Hata N, et al. Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. *Med Phys* 2001;28:2551-2560.
16. Blackledge MD, Collins DJ, Tunariu N, et al. Assessment of treatment response by total tumour volume and global apparent diffusion coefficient using diffusion-weighted MRI in patients with metastatic bone disease: a feasibility study. *PLoS One* 2014;9:e91779.
17. Simon D, Fritzsche KH, Thieke C, et al. Diffusion-weighted imaging-based probabilistic segmentation of high- and low-proliferative areas in high-grade gliomas. *Cancer Imaging* 2012;12:89-99.
18. Ozer S, Langer DL, Liu X, et al. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Med Phys* 2010;37:1873-1883.
19. Regini F, Gourtsoyianni S, Cardoso De Melo R, et al. Rectal tumour volume (GTV) delineation using T2-weighted and diffusion-weighted MRI: Implications for radiotherapy planning. *Eur J Radiol* 2014;83:768-772.
20. van Brussel S, Deckers F, van Laere S, et al. IMRT for rectal cancer patients based on diffusion-weighted MRI (DWMRI). *Int J Radiat Oncol Biol Phys* 2008;72 (Suppl):S251.
21. Velazquez ER, Parmar C, Jermoumi M, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep* 2013;18:3529.

# SUPPLEMENTARY MATERIAL 1

**Table 1.** Sequence parameters for the T2-weighted and diffusion-weighted sequences.

Sequence Parameters	T2-Weighted Fast Spin-Echo	3D T1-Weighted Gradient-Recalled Echo	DWI Sequence 1 (n=56)	DWI-2 (n=17)	DWI-3 (n=22)
TR/TE	8456-9558/130-150	9.8/4.6	4808-4829/70	4971/70	4172-5241/68-70
No. Slices	22-33	200	50	24	20-24
Slice thickness (mm)	2	1	5	5	5
Slice gap (mm)	2	-	0.5	0.5	0.5
In-plane resolution (mm)	0.78 x 1.14	1.15 x 1.15	2.50 x 3.11-3.18	1.82 x 2.31	1.82 x 2.27
Echo-train length	25	1	1	1	1
No. Signal averages	2-6	1	4	5	5
b-Value (s/mm <sup>2</sup> )	N/A	N/A	0,(100)*,500,1000	0,500,1000	0,(25,50,100)*,500,1000
Fat-suppression technique	N/A	N/A	STIR	SPIR/fatsat	SPAIR
Echo planar imaging factor	N/A	N/A	53-55	55	61

Note - DWI sequences were used in the following time periods: DWI-1 Start study - November 2011; DWI-2 December 2011 - June 2012; DWI-3 July 2012 - End study.

\*Values in parentheses were only acquired in a subset of patients.



## SUPPLEMENTARY MATERIAL 2

**Table 5.** Dice similarity indices ( $\pm$  SD) specified for low, mid and high rectal tumours

DSI	Height	Pre-CRT	Post-CRT
Automated vs. semi-automated (all observers)	Low	0.85 $\pm$ 0.15	0.56 $\pm$ 0.29
	Mid	0.93 $\pm$ 0.09	0.53 $\pm$ 0.31
	High	0.66 $\pm$ 0.27	0.57 $\pm$ 0.32
Automated vs. manual (all observers)	Low	0.71 $\pm$ 0.13	0.53 $\pm$ 0.13
	Mid	0.70 $\pm$ 0.10	0.52 $\pm$ 0.10
	High	0.55 $\pm$ 0.16	0.52 $\pm$ 0.08
Semi-automated vs. manual (all observer)	Low	0.73 $\pm$ 0.09	0.43 $\pm$ 0.22
	Mid	0.73 $\pm$ 0.11	0.46 $\pm$ 0.18
	High	0.51 $\pm$ 0.24	0.37 $\pm$ 0.24
Observer 1 vs. observer 2 (automated)	Low	0.77 $\pm$ 0.19	0.41 $\pm$ 0.26
	Mid	0.82 $\pm$ 0.14	0.62 $\pm$ 0.29
	High	0.57 $\pm$ 0.26	0.49 $\pm$ 0.24
Observer 1 vs. observer 2 (semi-automated)	Low	0.86 $\pm$ 0.13	0.59 $\pm$ 0.28
	Mid	0.92 $\pm$ 0.09	0.57 $\pm$ 0.32
	High	0.60 $\pm$ 0.30	0.49 $\pm$ 0.37
Observer 1 vs. observer 2 (manual)	Low	0.68 $\pm$ 0.12	0.53 $\pm$ 0.13
	Mid	0.69 $\pm$ 0.10	0.52 $\pm$ 0.10
	High	0.64 $\pm$ 0.10	0.52 $\pm$ 0.08

Numbers are means, standard deviations are given in parentheses.





# Chapter 7

Measuring the Apparent Diffusion Coefficient in primary rectal tumours – is there a benefit in performing histogram analyses?

MM van Heeswijk, DMJ Lambregts, M Maas, MJ Lahaye, Z Ayas,  
JMGM Slenter, GL Beets, FCH Bakers, RGH Beets-Tan

*Abdominal Radiology (2017)42:1627-1636.*

## ABSTRACT:

### Purpose

The Apparent Diffusion Coefficient (ADC) is a potential prognostic imaging-marker in rectal cancer. Typically mean ADC values are used, derived from precise manual whole-volume tumour delineations by experts. Aim was to explore whether non-precise circular delineation combined with histogram analysis can be a less cumbersome alternative to acquire similar ADC-measurements and second to explore whether histogram analyses provide additional prognostic information.

### Methods

Thirty-seven patients who underwent a primary staging MRI including diffusion-weighted imaging (DWI; b0,25,50,100,500,1000; 1.5T) were included. Volumes-of-interest (VOIs) were drawn on b1000-DWI: [1] precise delineation, manually tracing tumour boundaries (2 expert readers); [2] non-precise delineation, drawing circular VOIs with a wide margin around the tumour (2 non-experts). Mean ADC and histogram metrics (mean,min,max,median,SD,skewness,kurtosis,5<sup>th</sup>-95<sup>th</sup> percentiles) were derived from the VOIs and delineation time was recorded. Measurements were compared between the two methods and correlated with prognostic outcome parameters.

### Results

Median delineation time reduced from 47-165 sec (precise) to 25-43 sec (non-precise). The 45<sup>th</sup> percentile of the non-precise delineation showed the best correlation with the mean ADC from the precise delineation as the reference standard (ICC 0.71-0.75). None of the mean ADC or histogram parameters showed significant prognostic value; only the total tumour (VOI) volume was significantly larger in patients with positive clinical N-stage and mesorectal fascia involvement.

### Conclusion

When performing non-precise tumour delineation, histogram analysis (in specific 45<sup>th</sup> ADC percentile) may be used as an alternative to obtain similar ADC-values as with precise whole-tumour delineation. Histogram analyses are not beneficial to obtain additional prognostic information.



## INTRODUCTION

In recent years, diffusion-weighted imaging (DWI) has increasingly found its way to the clinical practice of Magnetic Resonance (MR) imaging in rectal cancer. Its value has particularly been demonstrated in the restaging setting to assess the response of the primary tumour to neoadjuvant chemoradiotherapy (CRT) and determine whether or not a residual tumour mass is still present within the post-radiation fibrosis<sup>1-3</sup>. In this setting the addition of DWI has been shown to improve the sensitivity for tumour restaging after CRT by >30% in a recent meta-analysis<sup>4</sup>.

Furthermore, in research settings, several studies have shown that quantifying the diffusion of rectal tumours by measuring the Apparent Diffusion Coefficient (ADC) may be used as an imaging biomarker. This could be beneficial in clinics to predict prognostic factors such as nodal stage, mesorectal fascia (MRF) involvement and histological differentiation grade<sup>5-8</sup>. Moreover, ADC may have value to predict therapeutic response<sup>9-13</sup>, which (in the future) could impact treatment stratification. For example neoadjuvant treatment may be further tailored depending on the anticipated treatment response. Moreover, accurate response evaluation can benefit the selection of good responding patient who can be candidates for minimally invasive follow-up treatments such as local excision or watchful waiting.

ADC values are most often expressed as the mean ADC, which can be acquired either from a single tumour slice<sup>14, 15</sup>, from tumour sample measurements<sup>3, 5</sup> or by manually delineating the whole tumour volume on the diffusion-weighted images<sup>1, 12, 16, 17</sup>. The latter approach is most commonly advocated and has been shown to provide the most reproducible results<sup>18, 19</sup>. As such, it is now considered more or less the standard reference method of choice to acquire tumour ADC measurements. It is, however, a labor intensive and time-consuming method, which is one of the factors that hampers translation of the use of quantitative ADC measures from research settings to clinical practice.

Hypothetically, an alternative approach to obtain whole-volume ADC measurements could be to perform a non-precise tumour delineation, e.g. by roughly placing a circular volume of interest (VOI) with a wide margin around the tumour. This saves time, but non-tumoural tissues such as mesorectal fat, normal bowel wall and lumen and surrounding organs and muscle structures will be included in the VOIs, which will affect the ADC measurements. In theory, this effect may be compensated for by adding histogram analyses as a post-processing step. With histogram analysis we analyze the spectrum of ADC values obtained from all voxels within the VOI. By doing so we can not only extract the mean ADC values, but also for example calculate minimum and maximum values and different percentile

ranges. In principle, this information may be used to specifically focus on those ADC values within the VOI representing tumour (which will typically be the lower ADC values within the spectrum), for example by extracting only lower percentile values thereby filtering out 'noise' from other included tissues. This could allow radiologist to save time with the delineation process whilst in principle obtaining the same ADC information. In addition, adding histogram analysis provides information about the distribution of ADC values within the tumour, which can potentially offer valuable additional insights into tumour structure and heterogeneity. This information could be of added benefit to predict response and prognosis, as has been suggested by previous authors<sup>9,10</sup>.

Therefore the aim of this study was two-fold. The primary aim was to test – on primary staging (pre-treatment) DWI-MRI - if histogram ADC analysis can be used to compensate non-precise tumour delineation and may be used as an alternative method to acquired similar ADC values as would normally be derived from precise whole-volume tumour delineation as the current 'standard of reference'. The second aim was to evaluate if histogram analysis provides valuable additional information to predict treatment outcome and prognosis.

## MATERIALS AND METHODS

### Patients

Forty-four patients diagnosed with and treated for rectal cancer at Maastricht University Medical Centre between October 2012 and June 2014 were considered for inclusion in this retrospective study. The study was approved by the local ethical institutional review board. Due to the retrospective nature of the study, informed consent was waived. Inclusion criteria were (a) biopsy proven non mucinous type rectal adenocarcinoma, (b) availability of a primary staging MRI including DWI (with a standardized acquisition protocol at 1.5T), (c) availability of follow-up data on treatment and outcome. Seven patients were excluded for the following reasons: severe artifacts on DWI, e.g. susceptibility artifacts due to air or metal prostheses (n=5), multiple tumour sites in the rectum (n=1), and prior pelvic radiation (n=1). This left a final study population of 37 patients.

### MR Imaging

All patients underwent a primary staging MRI at 1.5T MRI (Ingenia system, Philips Medical Systems, Best, The Netherlands), using a phased-array 16-channel body coil. Patients did not receive any bowel preparation. An intravenous bolus injection of 20 mg of butylscopolamine (Buscopan®, Boehringer Ingelheim bv, Ingelheim,



Germany) was administered intravenously to reduce peristaltic movement. The standard imaging protocol included standard two-dimensional T2-weighted (T2W) fast spin-echo sequences in 3 orthogonal directions (with the transverse images angled perpendicular and the coronal images angled parallel to the tumour axis as identified on the sagittal scan), and an axial echo planar imaging (EPI) DWI sequence angled in the same plane as the T2W transverse images. The DWI sequence was performed with spectral attenuated inversion recovery (SPAIR) fat-suppression (b-values 0,25,50,100,500,1000 s/mm<sup>2</sup>; TR/TE 4147/66 msec; EPI factor 77; 5 number of signal acquired; 1.82 x 2.26 x 5.00 mm acquisition voxel size, 20 slices, slice gap 0.5 mm; 6:44 minutes acquisition time). Apparent diffusion coefficient maps were automatically generated by the operating system, using a mono-exponential decay model including all six b-values.

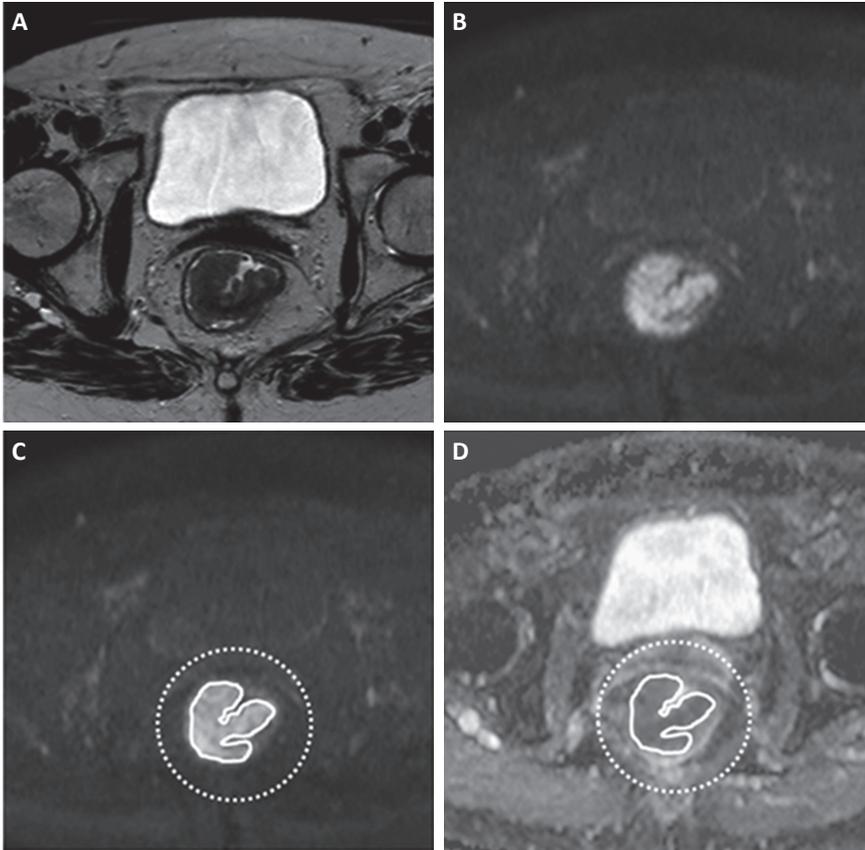
### **Precise and non-precise tumour delineation**

All primary staging MR images were transferred to an offline workstation for tumour delineation, which was performed using the freely available program Medview (Github clmedview, Maastricht, the Netherlands). Volumes-of-interest (VOIs) were drawn on the high b-value (b1000) diffusion images in two ways: [1] precise delineation and [2] non-precise delineation. For the precise delineation two experienced radiologists (R1 and R2; both with 8 years of experience in reading rectal MRI) manually closely traced the tumour boundaries on each consecutive slice to include the whole tumour volume. For the non-precise delineation two additional non-expert resident level readers (R3 and R4; both with no specific previous experience in reading rectal MRI) drew a circular/oval VOI with a margin around the tumour on each slice. The T2W-images were at the disposal of all four readers for anatomical reference. An example illustrating the two delineation methods is given in **Figure 1**. Time required to perform the delineations was recorded in a representative sample of n=18 study patients (for 1 expert and 1 non-expert reader) in order to quantify the potential reduction in measurement time.

### **Histogram metrics**

VOIs were transferred from the b1000 diffusion images to the ADC map to calculate mean ADCs and histogram metrics. Histogram plots were generated using a dedicated script written in MATLAB (The MathWorks Inc., Natick, MA, 2000) by one of the authors [BLINDED]. The following histogram parameters were calculated: minimum, maximum, mean, median (50<sup>th</sup> percentile), standard deviation (SD), skewness, kurtosis and every fifth percentile (5<sup>th</sup>- 95<sup>th</sup>). The total volume (cm<sup>3</sup>) of each VOI was also recorded.





**Figure 1** Example of the MR exam of a 82-year old female patient. A: Axial T2W image shows a semi-circular tumour, B: the b-1000 DWI shows high signal in the tumour area, C: example of the precise delineation (*solid line*; performed by reader 1) and the non-precise delineation (*dotted line*; performed by reader 3), the latter including both tumour and surrounding tissues, D: both delineations transferred to the ADC map.

### Outcome parameters

Various prognostic and therapeutic outcome parameters were collected. [1] From the primary staging MRI reports the cT-stage, cN-stage, and mesorectal; fascia (MRF) involvement. [2] From the clinical patient database the presence of distant metastases, neoadjuvant and surgical procedures. [3] From the pathology reports (of the biopsy and surgical specimens) the tumour differentiation grade, and [4] in patients undergoing a long course of neoadjuvant treatment the final treatment response defined as the tumour regression grade (TRG; method of Mandard) assessed at histopathology after surgery, where TRG1-2 was considered a good response and TRG3-5 a poor response<sup>20</sup>.

## Statistical analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, version 23.0, INC., Chicago, IL). The intraclass correlation coefficient (ICC; using a two-way mixed method with absolute agreement for single measures) was used to calculate agreement between the different readers and delineation methods. ICCs were also used to compare the various ADC histogram metrics of the non-precise delineation to the mean ADC of the precise delineation (being the most commonly used parameter in previous literature and therefore serving as the 'standard of reference'). Agreement was additionally assessed using Bland-Altman statistics. Independent sample T-tests (or Mann-Whitney U/Wilcoxon Rank test in case of non-normally distributed data) were used to compare mean ADC values and various histogram metrics between (1) cN0 vs. cN+ patients, (2) cMRF- vs. cMRF+ patients, (3) patients with vs. without metastases, (4) tumours with good-moderate vs. poor differentiation grade at histopathology and (5) good (TRG 1-2) vs. poor (TRG 3-5) responders. The Holm-Bonferroni correction method was applied to correct for multiple testing<sup>21</sup>. A Wilcoxon Rank test was performed to compare the delineation time between the two methods. P-values  $\leq 0.05$  were considered statistically significant.

## RESULTS

### Patient and treatment characteristics

Of the 37 study patients, 28 were male, nine female. Median age was 72 (range 29-86). Initial tumour stage on MRI was cT1-2 in 7 patients, cT3 in 27 patients and cT4 in 3 patients. Eleven patients were cN0 and 26 cN+. Eight patients had distant metastasis. Fifteen patients with non-locally advanced tumours underwent surgery without neoadjuvant treatment or immediately after a short course of 5x5 Gy. Seventeen locally advanced patients underwent a long course of CRT (28x1.8Gy radiotherapy with 2 x 825 mg/m<sup>2</sup>/d capecitabine) or 5x5 Gy with a prolonged waiting interval before surgery. Five patients received palliative care.

### Precise versus non-precise delineation method

VOI volumes, mean ADCs and histogram metrics derived from the primary staging MRIs are provided in **Table 1**. The mean volume of the VOIs used to calculate the ADCs was 3.22 cm<sup>3</sup> (R1) and 2.86cm<sup>3</sup> (R2) for the precise delineation method versus 8.65 cm<sup>3</sup> (R3) and 9.63 cm<sup>3</sup> (R4) for the non-precise delineation method (P<0.001). Mean ADC (\*10<sup>-3</sup> mm<sup>2</sup>/s) was 1.44 (R1) and 1.43 (R2) for the precise

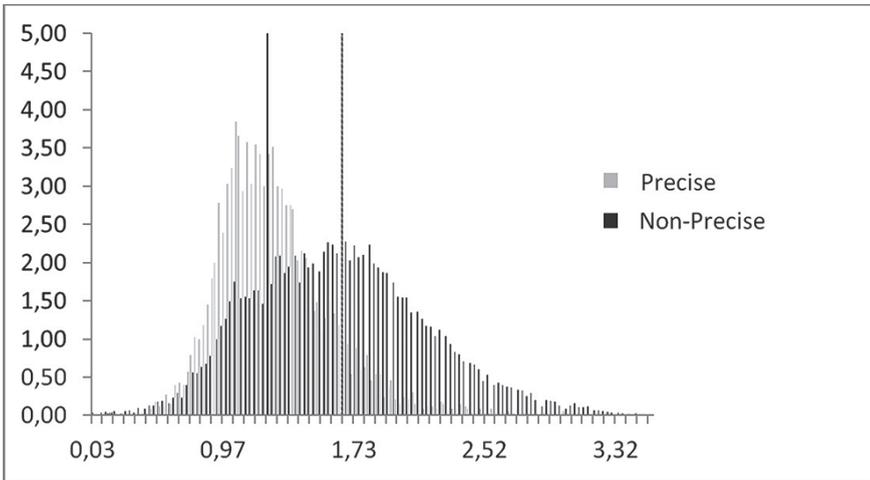


**Table 1.** Mean volumes, ADCs and histogram metrics including interobserver agreement between the readers and between the delineation methods.

	Precise delineation				Non-precise delineation				Precise vs. non-precise			
	Reader 1	Reader 2	ICC	Reader 3	Reader 4	ICC	R1 + R2*	R3 + R4*	ICC			
Volume	3218	2858	0.91	8653	9629	0.88	3038	9141	0.49			
Mean	1.44	1.43	0.98	1.51	1.51	0.96	1.43	1.51	0.64			
Minimum	0.31	0.32	0.78	0.01	0.01	0.80	0.31	0.01	0.00			
Maximum	3.18	3.18	0.82	3.58	3.61	0.95	3.18	3.60	0.64			
SD	0.40	0.40	0.92	0.57	0.58	0.93	0.40	0.57	0.35			
Skewness	0.66	0.67	0.87	0.15	0.17	0.83	0.67	0.16	0.12			
Kurtosis	4.28	4.27	0.80	3.42	3.36	0.73	4.28	3.39	0.31			
5 <sup>th</sup>	0.89	0.88	0.93	0.58	0.57	0.90	0.88	0.57	0.20			
10 <sup>th</sup>	1.00	0.98	0.94	0.79	0.78	0.91	0.99	0.78	0.26			
15 <sup>th</sup>	1.07	1.05	0.96	0.94	0.92	0.90	1.06	0.93	0.35			
20 <sup>th</sup>	1.13	1.11	0.97	1.05	1.03	0.89	1.12	1.04	0.46			
25 <sup>th</sup>	1.18	1.16	0.98	1.14	1.12	0.90	1.17	1.13	0.57			
30 <sup>th</sup>	1.23	1.21	0.98	1.22	1.21	0.92	1.22	1.22	0.64			
35 <sup>th</sup>	1.28	1.25	0.98	1.30	1.29	0.93	1.27	1.29	0.67			
40 <sup>th</sup>	1.32	1.30	0.98	1.37	1.36	0.94	1.35	1.36	0.67			
45 <sup>th</sup>	1.37	1.34	0.98	1.44	1.43	0.94	1.36	1.43	0.67			
50% (median)	1.40	1.38	0.98	1.50	1.50	0.95	1.39	1.50	0.65			
55 <sup>th</sup>	1.46	1.44	0.98	1.57	1.57	0.95	1.45	1.57	0.64			
60 <sup>th</sup>	1.51	1.49	0.98	1.64	1.64	0.96	1.50	1.64	0.63			
65 <sup>th</sup>	1.56	1.54	0.98	1.71	1.71	0.96	1.55	1.71	0.62			
70 <sup>th</sup>	1.62	1.60	0.98	1.79	1.79	0.96	1.61	1.79	0.62			
75 <sup>th</sup>	1.69	1.67	0.98	1.87	1.88	0.96	1.68	1.88	0.60			
80 <sup>th</sup>	1.76	1.75	0.98	1.97	1.98	0.97	1.75	1.98	0.58			
85 <sup>th</sup>	1.85	1.84	0.97	2.09	2.10	0.97	1.84	2.09	0.56			
90 <sup>th</sup>	1.96	1.96	0.96	2.24	2.25	0.97	1.96	2.25	0.53			
95 <sup>th</sup>	2.15	2.15	0.95	2.46	2.50	0.94	2.15	2.48	0.52			

\* Results are averaged between R1+R2 and R3+R4 respectively.

delineation versus 1.51 (R3) and 1.51 (R4) for the non-precise delineation ( $P=0.01-0.06$ ). An example comparing the histograms of the precise and non-precise delineation method is shown in **Figure 2**. **Table 2** shows the delineation times for the two measurement methods. Compared to precise delineation, non-precise delineation significantly reduced the delineation time for small, intermediate sized as well as for the large tumours with a median reduction in measurement time of 28 up to 132 seconds per patient/tumour.



**Figure2** Example of the normalized histograms for the precise and non-precise delineation in the same patient. The vertical lines in bold represent the mean ADC per method (solid line indicating a mean ADC of  $1.13 \cdot 10^{-3}$  mm<sup>2</sup>/s for the precise delineation and the dotted line indicating a mean ADC of  $1.64 \cdot 10^{-3}$  mm<sup>2</sup>/s for the non-precise delineation). These normalized histograms shows that the ADC values of the non-precise delineation are much more spread out due to the inclusion of other tissues, resulting in a higher mean ADC.

**Table 2.** Delineation times for the non-precise and precise delineation method

	Tumour volume (cm <sup>3</sup> )	Delineation time (sec)			
		Precise	Non-precise	Difference	P
Small (n=6)	0.57 (0.16-1.05)	47 (29-65)	21 (15-25)	28 (14-40)	0.03
Intermediate (n=6)	2.28 (1.87-2.49)	101(77-145)	31 (30-37)	71 (40-109)	0.03
Large (n=6)	6.24 (4.99-18.32)	165 (107-292)	43 (34-64)	123 (68-228)	0.03

NB, numbers are medians with ranges provided in parentheses

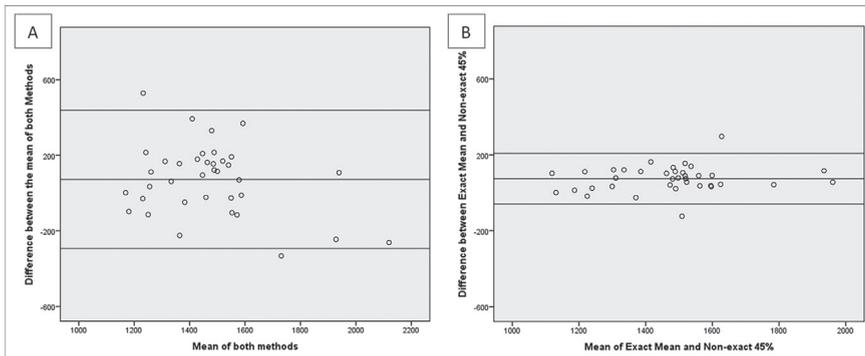
Delineation time was measured for reader 1 (precise) and reader 3 (non-precise) in a representative sample of  $n = 18$  tumours that were categorized into 'small', 'intermediate' and 'large' tumours by sorting all  $n = 37$  tumour ascendingly according to their volume (derived from Table 1), dividing the group into 3 equal subsets based on tumour volume and randomly selecting a sample of  $n = 6$  from each group.

## Interobserver and intermethod agreement

ICCs between the different readers and delineation methods are given in **Table 1**. Interobserver agreement was excellent, both for the precise method (R1 vs. R2) and the non-precise method (R3 vs. R4) with ICCs ranging between 0.80-0.98. ICCs comparing the precise and non-precise method were poor to good (ICC 0.00-0.67). **Table 3** shows the correlations of the various histogram metrics of the non-precise method to the mean ADC of the precise method (as the standard of reference): best correlation was found for the 45<sup>th</sup> percentile ADC of the non-precise method (ICC of 0.71-0.75). Results for the mean ADC and 45<sup>th</sup> percentile measurements are illustrated using Bland-Altman plots in **Figure 3**.

## ADC histogram analysis vs. prognostic factors

In **Table 4**, VOI volumes, mean ADCs and various histogram metrics are compared between different prognostic and response subgroups for both delineation methods. VOI volumes were significantly larger in the cN+ patients (both for the precise and non-precise delineation;  $P=0.04$ ) and in patients with MRF involvement on MRI (for the precise delineation;  $P=0.04$ ). Mean ADCs and the various histogram metrics were not significantly different between these subgroups. For the other outcome parameters (good vs. poor histopathological differentiation grade, good vs. poor response, metastasized versus non-metastasized patients) none of the volume, ADC or histogram metrics resulted in any significant differences between subgroups.



**Figure 3** Bland-Altman plots showing the reproducibility for tumour ADC measurements plotted as the mean ADC of the two methods (x-axis) against the difference in ADC between the two methods (y-axis). The middle line represent the mean absolute difference (bias) between the two methods while the outer lines represent the 95% confidence intervals (limits of agreement). Plot A compares the mean ADC derived from the precise delineation and non-precise delineation. Plot B compares the mean of the precise method and the 45<sup>th</sup> percentile of the non-precise method. Measurements were averaged for the two readers for both methods.

**Table 3.** ICCs constructed to explore which histogram parameter derived from non-precise delineation correlates best with the mean ADC of the precise delineation (as the standard of reference).

Precise (reference standard)	Mean ADC (R1):		Mean ADC (R2):	
	R3	R4	R3	R4
Non-precise				
Mean	0.63	0.63	0.63	0.62
Min	0.00	0.00	0.01	0.00
5 <sup>th</sup>	0.02	0.01	0.02	0.01
10 <sup>th</sup>	0.05	0.04	0.05	0.05
15 <sup>th</sup>	0.09	0.09	0.09	0.09
20 <sup>th</sup>	0.16	0.16	0.17	0.16
25 <sup>th</sup>	0.27	0.26	0.28	0.26
30 <sup>th</sup>	0.41	0.40	0.42	0.41
35 <sup>th</sup>	0.55	0.55	0.56	0.55
40 <sup>th</sup>	0.67	0.67	0.66	0.67
<b>45<sup>th</sup></b>	<b>0.73</b>	<b>0.75</b>	<b>0.71</b>	<b>0.73</b>
50 <sup>th</sup> (median)	0.72	0.74	0.71	0.72
55 <sup>th</sup>	0.65	0.67	0.62	0.65
60 <sup>th</sup>	0.55	0.58	0.53	0.55
65 <sup>th</sup>	0.45	0.48	0.44	0.46
70 <sup>th</sup>	0.36	0.38	0.36	0.37
75 <sup>th</sup>	0.28	0.29	0.29	0.29
80 <sup>th</sup>	0.22	0.22	0.22	0.23
85 <sup>th</sup>	0.16	0.16	0.17	0.17
90 <sup>th</sup>	0.11	0.11	0.12	0.12
95 <sup>th</sup>	0.08	0.07	0.08	0.08
Max	0.02	0.02	0.02	0.02

NB Best results are printed in bold.



**Table 4.** Correlation of histogram metrics with prognostic and therapeutic outcome parameters for both methods.

	cMRF involvement			cN-stage		
	cMRF- n=23	cMRF+ n=14	P	cN- n=11	cN+ n=26	P
<b>Method 1 (precise)</b>						
Volume	1995±1686	4751±5086	<b>0.04</b>	1388±1195	3737±4040	<b>0.04</b>
Minimum	0.33±0.30	0.28±0.23	1	0.46±0.32	0.25±0.23	0.64
Maximum	2.96±0.59	3.53±0.46	0.05	2.83±0.64	3.32±0.54	0.20
Mean	1.37±0.18	1.54±0.32	1	1.50±0.27	1.40±0.25	0.96
Median	1.32±0.19	1.50±0.34	1	1.48±0.27	1.35±0.26	0.63
SD	0.36±0.08	0.45±0.12	0.22	0.35±0.09	0.42±0.10	0.22
Skewness	0.64±0.46	0.71±0.63	1	0.33±0.38	0.81±0.51	0.12
Kurtosis	4.08±0.96	4.60±1.92	1	3.59±0.84	4.57±1.50	0.45
5th	0.87±0.16	0.90±0.25	1	0.98±0.18	0.84±0.19	0.54
30th	1.18±0.20	1.29±0.29	1	1.31±0.22	1.18±0.24	0.56
45th	1.30±0.21	1.45±0.33	1	1.44±0.25	1.32±0.27	0.55
70th	1.54±0.24	1.73±0.40	1	1.67±0.32	1.56±0.32	0.46
95th	2.04±0.30	2.32±0.44	0.3	2.17±0.38	2.17±0.38	0.70
<b>Method 2 (non-precise)</b>						
Volume	7018±3820	12627±10419	0.36	5371±3750	10735±8107	<b>0.04</b>
Minimum	0.01±0.02	0.02±0.02	0.48	0.01±0.01	0.01±0.02	1
Maximum	3.42±0.55	3.89±0.30	0.13	3.22±0.57	3.75±0.41	0.12
Mean	1.45±0.18	1.59±0.21	0.44	1.48±0.19	1.52±0.21	1
Median	1.45±0.15	1.58±0.23	0.50	1.50±0.17	1.49±0.21	1
SD	0.56±0.12	0.59±0.13	0.86	0.54±0.11	0.59±0.13	1
Skewness	0.04±0.29	0.36±0.53	0.54	0.14±0.37	0.17±0.44	1
Kurtosis	3.24±0.61	3.63±1.03	1	3.16±0.46	3.48±0.90	1
5th	0.52±0.20	0.65±0.27	0.54	0.52±0.21	0.59±0.24	1
30th	1.17±0.16	1.28±0.18	0.48	1.23±0.17	1.21±0.18	1
45th	1.39±0.16	1.51±0.22	0.42	1.44±0.17	1.43±0.20	1
70th	1.74±0.19	1.87±0.29	1	1.78±.21	1.80±0.25	1
95th	2.41±0.40	2.60±0.32	0.55	2.34±0.33	2.55±0.39	0.88

NB. Values for each method are averaged for the two readers;  
 # 'Good' = good, good-moderate and moderate differentiation, 'Poor' = poor or poor-moderate differentiation;  
 †'Good' = tumour regression grade 1-2, 'Poor' = TRG 3-5. To correct for multiple testing Holm-Bonferroni correction was performed (and mainly the most commonly reported percentile ranges were tested to reduce the number of testing variables). Significant results are printed in bold.

Differentiation grade <sup>a</sup>			Response after CRT <sup>†</sup>			Distant Metastases		
Good N=19	Poor N=3	P	Good N=7	Poor N=10	P	M- N=29	M+ N=8	P
3903±4761	1417±1029	1	1698±745	4080±5151	1	2717±3370	4202±4360	1
0.30±0.27	0.49±0.27	1	0.22±0.23	0.30±0.23	1	0.29±0.26	0.40±0.26	1
3.17±0.63	2.86±0.74	1	2.88±0.43	3.21±0.55	1	3.13±0.61	3.33±0.59	1
1.45±0.25	1.44±0.18	1	1.50±0.19	1.45±0.26	1	1.44±0.26	1.43±0.23	1
1.40±0.26	1.41±0.15	1	1.23±0.19	1.41±0.27	1	1.39±0.28	1.38±0.22	0.96
0.39±0.08	0.33±0.05	1	0.38±0.09	0.39±0.08	1	0.39±0.10	0.41±0.10	1
0.64±0.39	0.34±0.93	1	0.57±0.48	0.62±0.38	1	0.61±0.49	0.88±0.59	1
3.90±1.03	3.91±1.09	1	4.03±1.28	3.93±0.90	0.89	4.11±0.97	4.86±2.42	1
0.91±0.16	0.93±0.20	1	0.74±0.30	0.91±0.16	1	0.88±0.16	0.88±0.30	1
1.22±0.22	1.27±0.18	0.98	1.14±0.30	1.23±0.23	1	1.23±0.25	1.19±0.22	1
1.36±0.25	1.37±0.14	1	1.27±0.29	1.37±0.26	1	1.36±0.28	1.33±0.21	1
1.32±0.30	1.57±0.19	1	1.51±0.30	1.61±0.31	1	1.62±0.34	1.59±0.26	1
2.16±0.37	2.02±0.32	1	1.99±0.31	2.15±0.34	1	2.15±0.40	2.15±0.32	1
3275±3327	11109±9973	1	7199±3551	12398±11173	1	8742±7525	10586±7552	1
0.02±0.02	0.01±0.02	1	0.01±0.01	0.01±0.02	1	0.01±0.01	0.02±0.02	1
3.07±0.54	3.61±0.57	0.88	3.37±0.31	3.55±0.56	1	3.56±0.51	3.72±0.58	1
1.35±0.20	1.52±0.22	1	1.48±0.09	1.54±0.19	1	1.51±0.20	1.52±0.22	1
1.39±0.18	1.50±0.20	1	1.47±0.10	1.54±0.18	1	1.50±0.19	1.49±0.22	1
0.48±0.02	0.58±0.13	0.84	0.56±0.10	0.55±0.08	1	0.58±0.13	0.53±0.07	1
0.19±0.43	0.06±0.28	1	0.07±0.42	0.34±0.49	1	0.13±0.43	0.29±0.36	1
3.03±0.55	3.28±0.07	0.72	3.16±0.52	3.08±0.29	1	3.22±0.65	3.96±1.09	1
0.49±0.20	0.58±0.27	1	0.54±0.22	0.61±0.23	1	0.54±0.21	0.68±0.29	1
1.12±0.22	1.22±0.19	1	1.21±0.17	1.26±0.15	1	1.21±0.17	1.24±0.22	1
1.33±0.18	1.44±0.20	1	1.42±0.13	1.47±0.17	1	1.43±0.18	1.43±0.22	1
1.62±0.17	1.80±0.24	1	1.72±0.07	1.83±0.22	1	1.80±0.24	1.75±0.24	1
2.08±0.25	2.52±0.42	0.65	2.28±0.22	2.46±0.31	1	2.50±0.41	2.42±0.27	1



## DISCUSSION

The primary aim of this study was to assess the feasibility of calculating ADC values of rectal tumours at the time of primary staging using non-precise rectal tumour delineation combined with histogram analysis as an alternative to precise manual tumour delineation, aiming to simplify and speed up the delineation process. Precise volumetric delineation (typically performed by manual tracing of the tumour boundaries by expert readers) is the most commonly used method in current literature to calculate mean tumour ADCs and therefore in a way considered the current 'standard of reference' method. The benefit of a non-precise delineation (e.g. simply placing a circular ROI with a margin around the tumour area) is that it is faster and can be performed by non-experienced readers. The main drawback, however, is that tissues other than tumour such as the normal rectal wall, perirectal fat and adjacent organs will be included in the delineation, which will affect the mean ADC. Our hypothesis was that this effect may be overcome by adding histogram analysis to filter out these effects and specifically focus on ADC values of the tumour within the histogram in order to acquire similar ADCs as would have normally been derived by calculating the mean ADC from a precise delineation.

Our results show that when using histogram post-processing in such a way, the 45<sup>th</sup> percentile ADC from the non-precise delineation showed the best correlation with the mean ADC from the precise delineation as the standard of reference (ICC 0.71-0.75). Results without the addition of histogram post-processing were considerably poorer with an ICC of only 0.64 between the mean ADCs of the precise and non-precise method. The main reason for this poorer correlation is that the non-precise delineation resulted in remarkably higher overall ADC values, which can be explained by the fact that voxels with relatively high ADCs for example from the bladder, seminal vesicles, prostate and normal rectal wall were often included in the VOIs (see Figures 1 and 2).

Although use of the 45<sup>th</sup> percentile instead of mean ADC from the non-precise delineation thus improved the results, an agreement with a maximum ICC of 0.75 was still suboptimal, especially when comparing it for example to the ICC of 0.98 between the two readers for the precise delineation method. A previous study explored the use of (semi-) automated tumour segmentation (using computer algorithms) as an alternative method to overcome the problem of time consuming and labor-intensive manual tumour segmentation in rectal cancer. Similar to the current study the results of manual delineation were used as the reference standard. Although the main outcome was the volume of the VOI itself (and not the ADC as in our study), high ICCs of 0.91-0.97 for semi-automated tumour segmentation

were reported<sup>22</sup>. It would seem logical to assume that such an approach (given the excellent agreement when comparing the VOI volumes) would also result in a good agreement in ADC measurements if these were to be derived from these (semi-) automatically generated VOIs. It was reported that median delineation time decreased from 180-296 seconds for precise manual delineation to 41-69 seconds for semi-automated segmentation, which entails a considerable decrease in time and input required from readers<sup>22</sup>. In our study delineation time also significantly reduced with the non-precise method to a median measurement time ranging between 21 and 43 seconds per tumour/patient depending on the tumour volume. This would make it a similarly or even more effective solution with regard to time efficacy with the added benefit that the non-precise delineations can be performed by non-experienced readers and does not necessarily require expert input. The clinically relevant question is, however, if and how different delineation methods affect the utility of the acquired ADC measurements as a prognostic imaging biomarker.

Therefore, the second aim of this study was to explore the prognostic value of ADC measurements derived from the two delineation methods. In addition, we aimed to evaluate whether addition of histogram analysis provides valuable extra information. Previous studies have shown that mean ADC at primary staging may differentiate prognostically unfavorable tumour subtypes (e.g. tumours with MRF involvement, clinical N+ stage, tumour deposits and poor differentiation grade)<sup>5, 6</sup>. In our study we could unfortunately not reproduce these findings. Mean ADC values were not useful to differentiate between cMRF- and cMRF+ tumours, cN- and cN+ tumours, well and poorly differentiated histological tumour subtypes, or between patients with/without metastasized disease. Moreover, addition of histogram parameters did not lead to improved results. For the prediction of treatment response in the subgroup of patients undergoing chemoradiotherapy, neither mean ADC nor any of the histogram parameters showed significant results to differentiate between the poor and good responders using the tumour regression grade at histopathology as the outcome. Results in literature regarding this issue have also been conflicting. Some groups reported significantly lower mean ADC values in patients who showed a good response to treatment, suggesting that pre-treatment ADC may have potential to predict response, which could be of potential clinical benefit to tailor (neoadjuvant) treatment strategies depending on the anticipated response<sup>11-13, 23-25</sup>. Conversely, other groups found – similar to the current study – no significant differences in pre-treatment ADC between



responders and non-responders<sup>25-29</sup>. To date, only a few studies have investigated the potential benefit of adding histogram post-processing to predict rectal tumour response. Nougaret et al. reported that histogram metrics did not add to median ADC values for assessment of rectal tumour response after CRT<sup>19</sup>. Choi et al. found some promising results after CRT, with significant differences between poor and good responders after CRT for several histogram parameters (minimum ADC, 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> percentiles)<sup>10</sup>. This was confirmed by another study by Cho et al. who reported significantly different 10<sup>th</sup> and 25<sup>th</sup> percentile values between responders and non-responders with better diagnostic performance compared to mean ADC [9]. However, similar to the findings of our current study, Choi et al. find no benefit in any of the ADC or histogram metrics for pre-treatment prediction of response<sup>10</sup>.

Interestingly, the VOI volume, was the only parameter that resulted in significant differences between some of the favorable and unfavorable (cN+, cMRF+) subgroups. Previous studies also showed superior results for DWI tumour-volumetry compared to ADC measurements, albeit that these studies focused on assessment of tumour response after chemoradiotherapy rather than for predicting prognostic factors at primary staging. Pre-treatment DWI tumour volumes in those previous reports did not show any significant correlations with the final treatment outcome<sup>1,30</sup>, which is in line with our current findings.

Our study had some limitations, the first of which being its retrospective nature and the relatively small number of patients. Second, for some of the study patients not all of the outcome variables were available (for example because patients underwent palliative treatment and did not proceed to surgery). Moreover, the prognostic outcome factors (such as N-stage and MRF involvement) were primarily based on the MRI staging result rather than histopathology. This method – also previously used by other authors<sup>5,6</sup> – was chosen, since part of the study patients underwent neoadjuvant treatment before surgery and in these patients the final histopathology will no longer reflect the primary tumour stage. We, however, acknowledge that the clinical staging is a subjective measure that will be influenced by known limitations of MRI in assessing these factors, as well as radiologists' experience. Third, mean ADC derived from a precise expert delineation was arbitrarily defined as a standard of reference, because it is the measure that is currently most widely used in published reports. We acknowledge, however, that this is a subjective standard of reference that will vary for example depending on the DWI image protocol (e.g. acquisition parameters, patient preparation etc.) as well as the experience of the readers. Finally we performed our analyses based on the



assumption that the ADC values from both the precise and non-precise methods are normally distributed within patients (as illustrated in **Figure 2**), which would make it acceptable to identify the non-precisely measured 45<sup>th</sup> percentile as the best surrogate measurement for the mean ADC from the precise method. This was, however, not tested for each individual patient.

In conclusion, the 45<sup>th</sup> percentile ADC of the histogram derived from non-precise delineation correlates well with the mean ADC of the precise method and may thus be used as an alternative measure. In our study we could not confirm the previously reported potential value of ADC measurements to predict the prognostic tumour profile or response to treatment. Moreover, histogram ADC analysis did not appear to provide any additional prognostic information. Tumour volume was the only parameter found to correlate with prognostic features (N+ and MRF+ status).



## REFERENCES

1. Curvo-Semedo L, Lambregts DM, Maas M, Thywissen T, Mehsen RT, Lammering G, 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.
2. 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.
3. 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 tumour response to neoadjuvant chemo- and radiation therapy. *Radiology*. 2009;253(1):116-25.
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. Curvo-Semedo L, Lambregts DM, Maas M, Beets GL, Caseiro-Alves F, Beets-Tan RG. Diffusion-weighted MRI in rectal cancer: apparent diffusion coefficient as a potential noninvasive marker of tumour aggressiveness. *J Magn Reson Imaging*. 2012;35(6):1365-71.
6. Sun Y, Tong T, Cai S, Bi R, Xin C, Gu Y. Apparent Diffusion Coefficient (ADC) value: a potential imaging biomarker that reflects the biological features of rectal cancer. *PLoS One*. 2014;9(10):e109371.
7. Akashi M, Nakahusa Y, Yakabe T, Egashira Y, Koga Y, Sumi K, et al. Assessment of aggressiveness of rectal cancer using 3-T MRI: correlation between the apparent diffusion coefficient as a potential imaging biomarker and histologic prognostic factors. *Acta Radiol*. 2014;55(5):524-31.
8. Tong T, Yao Z, Xu L, Cai S, Bi R, Xin C, et al. Extramural depth of tumour invasion at thin-section MR in rectal cancer: associating with prognostic factors and ADC value. *J Magn Reson Imaging*. 2014;40(3):738-44.
9. Cho SH, Kim GC, Jang YJ, Ryeom H, Kim HJ, Shin KM, et al. Locally advanced rectal cancer: post-chemoradiotherapy ADC histogram analysis for predicting a complete response. *Acta Radiol*. 2015;56(9):1042-50.
10. Choi MH, Oh SN, Rha SE, Choi JI, Lee SH, Jang HS, et al. Diffusion-weighted imaging: Apparent diffusion coefficient histogram analysis for detecting pathologic complete response to chemoradiotherapy in locally advanced rectal cancer. *J Magn Reson Imaging*. 2016;44(1):212-20.
11. Barbaro B, Vitale R, Valentini V, Illuminati S, Vecchio FM, Rizzo G, et al. Diffusion-weighted magnetic resonance imaging in monitoring rectal cancer response to neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys*. 2012;83(2):594-9.
12. Sun YS, Zhang XP, Tang L, Ji JF, Gu J, Cai Y, et al. Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging for early detection of tumour histopathologic downstaging. *Radiology*. 2010;254(1):170-8.
13. Intven M, Reerink O, Philippens ME. Diffusion-weighted MRI in locally advanced rectal cancer: pathological response prediction after neo-adjuvant radiochemotherapy. *Strahlenther Onkol*. 2013;189(2):117-22.

14. Dzik-Jurasz A, Domenig C, George M, Wolber J, Padhani A, Brown G, et al. Diffusion MRI for prediction of response of rectal cancer to chemoradiation. *Lancet*. 2002;360(9329):307-8.
15. Kremser C, Judmaier W, Hein P, Griebel J, Lukas P, de Vries A. Preliminary results on the influence of chemoradiation on apparent diffusion coefficients of primary rectal carcinoma measured by magnetic resonance imaging. *Strahlenther Onkol*. 2003;179(9):641-9.
16. Kim SH, Lee JY, Lee JM, Han JK, Choi BI. Apparent diffusion coefficient for evaluating tumour response to neoadjuvant chemoradiation therapy for locally advanced rectal cancer. *Eur Radiol*. 2011;21(5):987-95.
17. Roth Y, Tichler T, Kostenich G, Ruiz-Cabello J, Maier SE, Cohen JS, et al. High-b-value diffusion-weighted MR imaging for pretreatment prediction and early monitoring of tumour response to therapy in mice. *Radiology*. 2004;232(3):685-92.
18. Lambregts DM, Beets GL, Maas M, Curvo-Semedo L, Kessels AG, Thywissen T, et al. Tumour ADC measurements in rectal cancer: effect of ROI methods on ADC values and interobserver variability. *Eur Radiol*. 2011;21(12):2567-74.
19. Nougaret S, Vargas HA, Lakhman Y, Sudre R, Do RK, Bibeau F, et al. Intravoxel Incoherent Motion-derived Histogram Metrics for Assessment of Response after Combined Chemotherapy and Radiation Therapy in Rectal Cancer: Initial Experience and Comparison between Single-Section and Volumetric Analyses. *Radiology*. 2016;280(2):446-454.
20. Mandard AM, Dalibard F, Mandard JC, Marnay J, Henry-Amar M, Petiot JF, et al. Pathologic assessment of tumour regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer*. 1994;73(11):2680-6.
21. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand J Statist*. 1979;6:65-70.
22. van Heeswijk MM, Lambregts DM, van Griethuysen JJ, Oei S, Rao SX, de Graaff CA, 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.
23. Jung SH, Heo SH, Kim JW, Jeong YY, Shin SS, Soung MG, et al. Predicting response to neoadjuvant chemoradiation therapy in locally advanced rectal cancer: diffusion-weighted 3 Tesla MR imaging. *J Magn Reson Imaging*. 2012;35(1):110-6.
24. Lambrecht M, Deroose C, Roels S, Vandecaveye V, Penninckx F, Sagaert X, et al. The use of FDG-PET/CT and diffusion-weighted magnetic resonance imaging for response prediction before, during and after preoperative chemoradiotherapy for rectal cancer. *Acta Oncol*. 2010;49(7):956-63.
25. Lambrecht M, Vandecaveye V, De Keyzer F, Roels S, Penninckx F, Van Cutsem E, et al. Value of diffusion-weighted magnetic resonance imaging for prediction and early assessment of response to neoadjuvant radiochemotherapy in rectal cancer: preliminary results. *Int J Radiat Oncol Biol Phys*. 2012;82(2):863-70.
26. DeVries AF, Kremser C, Hein PA, Griebel J, Krezcy A, Ofner D, et al. Tumour microcirculation and diffusion predict therapy outcome for primary rectal carcinoma. *Int J Radiat Oncol Biol Phys*. 2003;56(4):958-65.



27. Ippolito D, Monguzzi L, Guerra L, Deponti E, Gardani G, Messa C, et al. Response to neoadjuvant therapy in locally advanced rectal cancer: assessment with diffusion-weighted MR imaging and 18FDG PET/CT. *Abdom Imaging*. 2012;37(6):1032-40.
28. Kim YC, Lim JS, Keum KC, Kim KA, Myoung S, Shin SJ, et al. Comparison of diffusion-weighted MRI and MR volumetry in the evaluation of early treatment outcomes after preoperative chemoradiotherapy for locally advanced rectal cancer. *J Magn Reson Imaging*. 2011;34(3):570-6.
29. Monguzzi L, Ippolito D, Bernasconi DP, Trattenero C, Galimberti S, Sironi S. Locally advanced rectal cancer: value of ADC mapping in prediction of tumour response to radiochemotherapy. *Eur J Radiol*. 2013;82(2):234-40.
30. Lambregts DM, Rao SX, Sassen S, Martens MH, Heijnen LA, Buijsen J, 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.









# Chapter 8

General Discussion

In recent years, novel MR techniques and image quantification methods have been developed and studied to investigate their potential to address the diagnostic challenges we are facing in rectal cancer staging and restaging. Radiologists need to keep adapting to changes in order to remain indispensable partners in the multidisciplinary team involved in the treatment of rectal cancer. This is especially to the current trend toward more patient tailored treatments, based on the specific needs of the individual patient. Medical research is the beating heart of this evolution. Unfortunately, many promising research discoveries never make it past the early research phase, because they are not sufficiently validated, unpractical for daily use or because they lack a clear clinical incentive. A good example of a research technique that has proven to be successful is diffusion-weighted imaging (DWI). Since its introduction as a potential oncologic imaging technique, over 5000 scientific reports have been published on the use of DWI in oncology, with over 200 specific publication on the topic of DWI for rectal cancer imaging.

With this thesis we aimed to help overcome this problem by taking previous research (including work from our own group) one step further, and to investigate new ways that help translate the use of novel imaging and image quantification techniques from research to daily clinical practice.

## **PART I - DIFFUSION-WEIGHTED IMAGING FOR RECTAL STAGING AND RESTAGING IN DAILY PRACTICE**

### **DWI for response evaluation after chemoradiotherapy**

Diffusion-weighted imaging was originally mainly used for imaging of the brain as a highly sensitive technique to detect brain ischemia<sup>1</sup>. Over the last few years, DWI has slowly also found its way to body imaging. DWI is a functional imaging technique that measures the extracellular movement ('diffusion') of water protons between tissues. As this diffusion varies in different tissues, DWI can distinguish normal from pathological tissues<sup>2</sup>. The first studies on the use of DWI in rectal cancer originate from 2002 and focused on response evaluation after chemoradiotherapy (CRT)<sup>3, 4</sup>. In 2009, Kim et al. were the first to convincingly show that the addition of DWI increases the diagnostic accuracy in the visual diagnosis of a complete tumour response after neoadjuvant treatment compared to the use of only the standard morphological T2-Weighted (T2W) imaging sequences<sup>5</sup>. After this publication, several reports, including a recent meta-analysis from 2013<sup>6</sup> have supported this finding and DWI is nowadays implemented in many centres as part of the standard rectal MRI protocol, particularly for restaging after CRT. However, a critical note is that in almost every study on DWI for response evaluation, the

images were read by abdominal radiologists or even specific rectal MR experts, often with many years of previous MR reading experience. In liver MR imaging, it has been demonstrated that the level of experience is a significant factor that can contribute to DWI reading performance<sup>7</sup>. In rectal cancer DWI imaging, this factor has not been specifically addressed before. This was one of the reasons why our research group initiated a nationwide multicentre study in 2012, funded by the Dutch Cancer Society, in which eight centres in the Netherlands participated. In this study, DWI was routinely added to the rectal MRI protocol for primary staging as well as for response evaluation after neoadjuvant CRT. DW images were read by local radiologists and then re-assessed by expert readers, who also provided the local radiologists with feedback. Inclusion of patients has recently been completed and the results are currently being analysed. As an adjunct to this multicentre study, we initiated a project (**chapter 4**) in which we asked two radiology residents with no specific previous rectal MRI experience to read 100 consecutive restaging MRIs including DWI to evaluate response to CRT. The readers received regular expert feedback. The aim was to explore the most common interpretation pitfalls encountered by non-expert readers when assessing DWI scans of the rectum and to study the effects of structured expert teaching. At the end of the study, the two readers were able to differentiate between patients with a complete response and patients with residual tumour after CRT with AUCs of up to 0.88, which is similar to that previously reported for experts. These results suggest that, after some training, DWI can be a useful addition to standard MRI and other clinical tools, such as the digital rectal examination and endoscopy. Together these tools can be used in daily clinical practice to assess response after CRT, and may help select clinical complete responders, who may be eligible for a watchful waiting approach. In this setting, the combination of DWI-MRI together with clinical evaluation including endoscopy has been shown highly accurate to select complete responders after CRT with a post-test probability of 98%<sup>8</sup>.

### **DWI for primary tumour staging**

While the role of DWI for response evaluation after CRT is already relatively well established, evidence for its use in the primary staging setting is still sparse. Some studies have shown that DWI may help in the primary detection of rectal tumours on MRI. Although from a clinical point of view the role of MRI is not so much to detect but rather to stage tumours, the high conspicuity of tumours on DWI may still be of added value in this staging process. In **chapter 2** we evaluated the potential value of DWI to measure tumour height and tumour length in the primary staging setting. The incentive to perform this study was that previous studies have sugge-

sted that use of standard (T2-weighted) sequences may result in an overestimation of the extent of rectal tumours on MRI<sup>9</sup>. Our results in **chapter 2** appear to support this finding: particularly when measurements are performed by a relatively inexperienced reader, use of DWI instead of T2-weighted MRI results in a significantly smaller assessment of tumour length and consequently a larger tumour height, suggesting that addition of DWI may help reducing the overestimation of tumour size.

### **DWI for Lymph node staging**

The role of DWI for lymph node imaging is a topic that has been the subject of ongoing debate in recent literature. There is a general consensus that DWI is a very powerful tool to detect lymph nodes, due to the fact that lymph node tissue has a very dense cellular structure, which results in high signal and thus excellent visibility on DWI<sup>10-14</sup>. Several studies, in rectal cancer as well in other malignancies, have shown that DWI detects a much higher number of nodes compared to standard T2W imaging. For example Lambregts et al. reported that on restaging MRI after CRT DWI detected 26% more nodes than T2-weighted MRI<sup>13</sup>. Despite this high detection rate so far DWI has not shown a convincing clinical benefit for nodal staging. This is because it lacks the ability to further characterize the nodes once detected. Various studies have looked into the signal intensities of nodes on DWI to differentiate between malignant and benign lymph nodes, but did not find any difference between the two groups<sup>14, 15</sup>. In addition, studies looked into quantitative DWI analysis of nodes using Apparent Diffusion Coefficient (ADC) measurements, again with disappointing results. Different studies could not produce a significant difference in ADC values between benign and malignant nodes<sup>14</sup>, or did find a significant difference but with significant overlap in ADC values between the malignant and benign lymph nodes<sup>13</sup>. This limits the clinical utility of measuring ADC, which was also reported by several groups who found no clear added value in terms of diagnostic accuracy when adding ADC measurements to routine visual (size and morphology based) assessment of lymph nodes<sup>13, 16, 17</sup>. Therefore we set out in **chapter 3** to use DWI in an alternative, more practical way, focussing on the main strength of DWI: its high sensitivity to detect even very small lymph nodes. After CRT, lymph nodes are known to reduce in size, which makes them harder to detect and assess<sup>18, 19</sup>. We hypothesized that if any positive nodes remain after CRT, these should be visible on DWI. In other words, if there are no nodes visible on DWI after CRT, this should indicate that all nodes have been sterilised and considered proof of an yNO status. Selection of these node-negative patients after CRT

is important, particularly when considering patients for organ saving treatments such as watchful waiting. We indeed found that - although the total absence of nodes on DWI was not a frequent finding – it was a 100% reliable predictor for a yNO-status. When used in this way, DWI may thus still be of practical added benefit for lymph node evaluation after CRT, in addition to its already established role for response assessment of the primary tumour described above.

## PART II – QUANTIFICATION – A PRACTICAL APPROACH

There is much more information that can be derived from imaging than what can be observed with visual assessment alone. Therefore, in recent years a lot of research has focussed on quantitative imaging assessment. One of the most often studied quantitative measurements in rectal cancer is the tumour size. The common notion is that with increasing tumour size, outcome will be worse: larger tumours have shown to have a more aggressive tumour profile, and to have a poorer response to neoadjuvant therapy<sup>20-23</sup>. Moreover, when a tumour shows little reduction in size as a result of treatment, this is typically associated with a poor response. Although over 15 studies have been published on the topic of rectal tumour size/volume, it is difficult to draw firm conclusions about the best way to perform measurements (one-dimensional, 3-dimensional or whole-volume), the reproducibility of the variable threshold values reported, their clinical implications and how these measurements may be used in daily practice. This is the reason why we performed the study described in **chapter 5**. In this study, we conducted a review of all existing literature on tumour size/volume measurements for response assessment in rectal cancer and validated the findings of these studies in an independent multicentre patient cohort of n=146 patients. We found that one dimensional (length) or 3-dimensional size measurements are of limited use to assess response. Whole-volume measurements on the other hand, can be helpful in rectal cancer response assessment when using selected cut-off values, particularly to assess a complete tumour response (ypT0) after CRT with accuracies – after prospective validation – of up to 80%. However, these results for whole tumour volumetry on T2-weighted MRI are still inferior to those reported for visual assessment on T2W images combined with DWI (with reported accuracies up to 94%)<sup>23-25</sup>. Tumour volumes on MRI are typically acquired by the exact manual delineation of the tumour boundaries on each slice by experienced readers. These precise delineations are also required to calculate tumour ADC values, as whole-volume delineations have been shown to provide the best reproducible tumour ADC values with superior results compared to sample or single slice measurements<sup>24, 25</sup>. Howe-

ver, the obvious downside of these whole-volume delineations is that they require extensive input from radiologists and are highly time-consuming. In order to make tumour quantification more practically applicable, there is thus an obvious need to study methods that can facilitate tumour delineation. In **chapter 6** we investigated the option of automated tumour segmentation on DWI. The practical advantage of DWI is that it has a very high tumour-to-background contrast. This makes DWI potentially suitable for automated detection and segmentation of tumours using software algorithms. When using manual tumour segmentation as the reference standard, our first results showed that similar outcomes can be achieved when the radiologists manually adjust the inaccuracies of automatically generated segmentations. This approach saved up to 18 minutes delineation time per patient when compared to full manual delineation. The input required from radiologists may be reduced even more when these auto-segmentation software models are further developed, for example using machine-learning algorithms<sup>26</sup>.

In **chapter 7** we investigated an alternative method to reduce the delineation input and time that is required from radiologists to calculate tumour ADC values. We investigated whether non-precise tumour delineations performed by non-expert readers could be used as an alternative if we compensate for the delineation inaccuracies by adding histogram analysis as a post-processing step. Histograms are used to assess the distribution of individual voxels values within a certain region of interest. As such histograms may theoretically be used to focus on specific ADC values (e.g. tumoural values) within the spectrum, thereby filtering out 'noise' created by inaccurate delineations. We indeed found that histograms may be used in such a way and that the time saved with the delineation procedure was almost similar to that for the auto-delineation software algorithms used in **chapter 6**. As an added benefit, no expert input was required, as delineations could be performed entirely by non-expert (resident level) readers. On the other hand, reproducibility compared to whole-volume manual delineations was considerably poorer than the results for our software algorithm approach in **chapter 6**. As a secondary goal, we tested the prognostic value of the ADC values obtained in our study and compared these to results from previous literature. ADC is the most commonly used quantitative diffusion measure and (indirectly) provides information about the tumoural cell and tissue architecture. Some studies have shown that ADC can be used as a prognostic biomarker to predict tumours with a more or less aggressive tumour tissue profile<sup>27-30</sup>. Moreover, some studies have demonstrated significantly lower ADC values in tumours that show a better response to treatment<sup>31, 32</sup>. The hypothesis is that tumours with lower ADC values have less necrotic changes and



are therefore better susceptible to neoadjuvant CRT. Despite these promising initial results, various others reports could not confirm these findings. We were also not able to demonstrate the benefit of measuring ADC to predict prognostic tumour factors or treatment outcome in our study. Furthermore, we could not find a clear additional prognostic value for calculating histogram metrics. Therefore, the role of ADC as a clinical biomarker to date remains questionable.

## FUTURE PERSPECTIVES

The development of novel imaging and post-processing methods is a hot topic in medical imaging research. The model of development, research and implementation of DWI described in this thesis could serve as an example to stimulate research on other promising functional imaging techniques, such as dynamic contrast enhancement (DCE) MRI, hypoxia imaging, fibrosis imaging (magnetization transfer MRI), hybrid PET-MR imaging, etcetera. In the future we should attempt to define the most clinically meaningful (functional) imaging parameters and combine these with other clinical, histopathological and immunohistochemical markers into comprehensive multiparametric models to help further optimize and tailor treatments to a patient's individual tumour profile. The ultimate aim is to offer each patient the best possible treatment, chance of cure and quality of life.

## RECOMMENDATIONS FOR CLINICAL PRACTICE

With this thesis, we aimed to explore ways to facilitate the general use of DWI in the daily management of rectal cancer patients in a practical manner, and to see what steps should be taken to help translate the use of more novel and quantitative DWI applications such as DWI tumour volumetry and ADC measurements from research to daily clinical practice.

As shown by previous research, DWI is a useful adjunct to the clinical and endoscopic selection of rectal cancer patients with a complete response after CRT. This thesis further proves that DWI can aid in the selection of patients for organ-preservation by accurately identifying patients with a node negative status after CRT. Absence of nodes on post CRT DWI is 100% predictive of a  $\gamma$ N0 status. Also, it has been shown that a general radiologist needs to be trained in 40-60 cases to reach diagnostic performance levels similar to that of expert readers.

With regard to tumour quantification, the results in this thesis do not support the routine use of ADC as an imaging biomarker. Tumour volumetry on standard T2-weighted MRI is inferior to that on DWI, and T2W volumetry for the assessment of response therefore is of limited use for daily practice. DWI volumetry

could work but it is a time consuming process and therefore unpractical for daily use. In this thesis we have shown that mainly (semi-) automated software algorithms are promising to reduce the time required for tumour segmentations on DWI. However the currently available algorithms still requires manual input from radiologists. Therefore, future research should focus on further optimizing these algorithms in order to make DWI tumour volumetry a practically applicable tool for daily use.

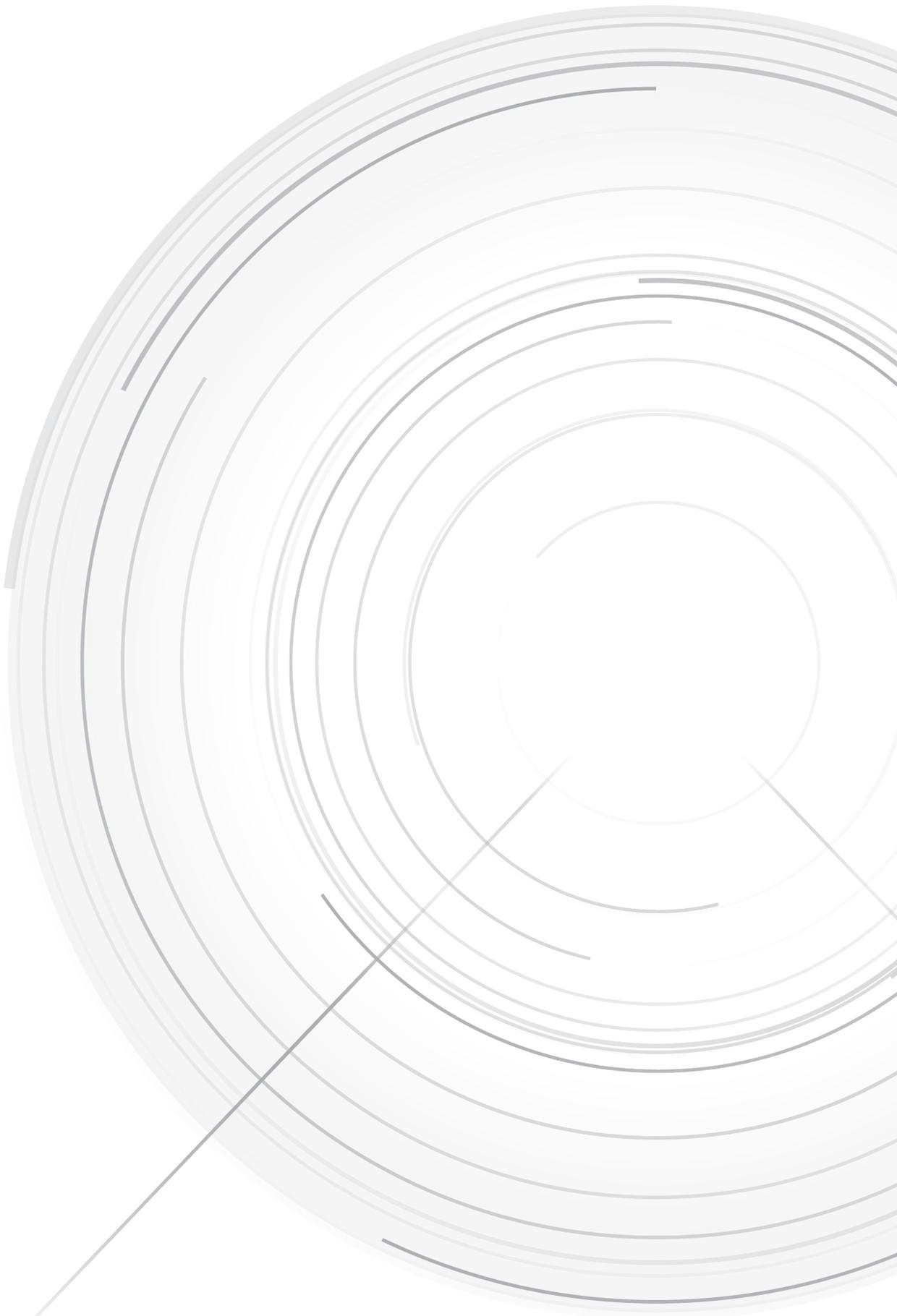


## REFERENCES

1. Le Bihan D. Diffusion/perfusion MR imaging of the brain: from structure to function. *Radiology*. 1990;177(2):328-9.
2. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*. 1986;161(2):401-7.
3. Hein PA, Kremser C, Judmaier W, Griebel J, Pfeiffer KP, Kreczy A, et al. Diffusion-weighted magnetic resonance imaging for monitoring diffusion changes in rectal carcinoma during combined, preoperative chemoradiation: preliminary results of a prospective study. *Eur J Radiol*. 2003;45(3):214-22.
4. Dzik-Jurasz A, Domenig C, George M, Wolber J, Padhani A, Brown G, et al. Diffusion MRI for prediction of response of rectal cancer to chemoradiation. *Lancet*. 2002;360(9329):307-8.
5. 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 tumour response to neoadjuvant chemo- and radiation therapy. *Radiology*. 2009;253(1):116-25.
6. van der Paardt MP, Zegers 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.
7. Fukumoto W, Nakamura Y, Higaki T, Tatsugami F, Iida M, Awai K. Additional Value of Diffusion-weighted MRI to Gd-EOB-DTPA-enhanced Hepatic MRI for the Detection of Liver Metastasis: the Difference Depending on the Experience of the Radiologists. *Hiroshima J Med Sci*. 2015;64(1-2):15-21.
8. Maas M, Lambregts DM, Nelemans PJ, Heijnen LA, Martens MH, Leijtens JW, et al. 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*. 2015;22(12):3873-80.
9. Buijssen J, van den Bogaard J, Janssen MH, Bakers FC, Engelsman S, Ollers M, et al. FDG-PET provides the best correlation with the tumour specimen compared to MRI and CT in rectal cancer. *Radiother Oncol*. 2011;98(2):270-6.
10. Nakai G, Matsuki M, Inada Y, Tatsugami F, Tanikake M, Narabayashi I, et al. Detection and evaluation of pelvic lymph nodes in patients with gynecologic malignancies using body diffusion-weighted magnetic resonance imaging. *J Comput Assist Tomogr*. 2008;32(5):764-8.
11. Mir N, Sohaib SA, Collins D, Koh DM. Fusion of high b-value diffusion-weighted and T2-weighted MR images improves identification of lymph nodes in the pelvis. *J Med Imaging Radiat Oncol*. 2010;54(4):358-64.
12. Heusner TA, Kuemmel S, Koeninger A, Hamami ME, Hahn S, Quinsten A, et al. Diagnostic value of diffusion-weighted magnetic resonance imaging (DWI) compared to FDG PET/CT for whole-body breast cancer staging. *Eur J Nucl Med Mol Imaging*. 2010;37(6):1077-86.
13. Lambregts DM, Maas M, Riedl RG, Bakers FC, Verwoerd JL, Kessels AG, et al. Value of ADC measurements for nodal staging after chemoradiation in locally advanced rectal cancer—a per lesion validation study. *Eur Radiol*. 2011;21(2):265-73.

14. Heijnen LA, Lambregts DM, Mondal D, Martens MH, Riedl RG, Beets GL, et al. Diffusion-weighted MR imaging in primary rectal cancer staging demonstrates but does not characterise lymph nodes. *Eur Radiol.* 2013;23(12):3354-60.
15. Mizukami Y, Ueda S, Mizumoto A, Sasada T, Okumura R, Kohno S, et al. Diffusion-weighted magnetic resonance imaging for detecting lymph node metastasis of rectal cancer. *World J Surg.* 2011;35(4):895-9.
16. Cho EY, Kim SH, Yoon JH, Lee Y, Lim YJ, Kim SJ, et al. Apparent diffusion coefficient for discriminating metastatic from non-metastatic lymph nodes in primary rectal cancer. *Eur J Radiol.* 2013;82(11):e662-8.
17. Kim SH, Ryu KH, Yoon JH, Lee Y, Paik JH, Kim SJ, et al. Apparent diffusion coefficient for lymph node characterization after chemoradiation therapy for locally advanced rectal cancer. *Acta Radiol.* 2015;56(12):1446-53.
18. Perez RO, Pereira DD, Proscuschim I, Gama-Rodrigues J, Rawet V, Sao Juliao GP, et al. Lymph node size in rectal cancer following neoadjuvant chemoradiation--can we rely on radiologic nodal staging after chemoradiation? *Dis Colon Rectum.* 2009;52(7):1278-84.
19. Heijnen LA, Maas M, Beets-Tan RG, Berkhof M, Lambregts DM, Nelemans PJ, et al. Nodal staging in rectal cancer: why is restaging after chemoradiation more accurate than primary nodal staging? *Int J Colorectal Dis.* 2016;31(6):1157-62.
20. Kim NK, Baik SH, Min BS, Pyo HR, Choi YJ, Kim H, et al. A comparative study of volumetric analysis, histopathologic downstaging, and tumour regression grade in evaluating tumour response in locally advanced rectal cancer following preoperative chemoradiation. *Int J Radiat Oncol Biol Phys.* 2007;67(1):204-10.
21. Dresen RC, Beets GL, Rutten HJ, Engelen SM, Lahaye MJ, Vliegen RF, et al. Locally advanced rectal cancer: MR imaging for restaging after neoadjuvant radiation therapy with concomitant chemotherapy. Part I. Are we able to predict tumour confined to the rectal wall? *Radiology.* 2009;252(1):71-80.
22. Aiba T, Uehara K, Nihashi T, Tsuzuki T, Yatsuya H, Yoshioka Y, et al. MRI and FDG-PET for assessment of response to neoadjuvant chemotherapy in locally advanced rectal cancer. *Ann Surg Oncol.* 2014;21(6):1801-8.
23. Seierstad T, Hole KH, Groholt KK, Dueland S, Ree AH, Flatmark K, et al. MRI volumetry for prediction of tumour response to neoadjuvant chemotherapy followed by chemoradiotherapy in locally advanced rectal cancer. *Br J Radiol.* 2015;88(1051):20150097.
24. Lambregts DM, Beets GL, Maas M, Curvo-Semedo L, Kessels AG, Thywissen T, et al. Tumour ADC measurements in rectal cancer: effect of ROI methods on ADC values and interobserver variability. *Eur Radiol.* 2011;21(12):2567-74.
25. Nougaret S, Vargas HA, Lakhman Y, Sudre R, Do RK, Bibeau F, et al. Intravoxel Incoherent Motion-derived Histogram Metrics for Assessment of Response after Combined Chemotherapy and Radiation Therapy in Rectal Cancer: Initial Experience and Comparison between Single-Section and Volumetric Analyses. *Radiology.* 2016:150702.

26. Trebeshi S vGJ, Aerts HJWL, Lambregts DMJ, Lahaye MJ, Bakers FCH, Peters NHGM, Verheij M, Beets-Tan RGH. Fully automated segmentation of rectal carcinomas using supervised learning techniques with expert-reader input [abstract]. *Insights into imaging* 2017.
27. Curvo-Semedo L, Lambregts DM, Maas M, Beets GL, Caseiro-Alves F, Beets-Tan RG. Diffusion-weighted MRI in rectal cancer: apparent diffusion coefficient as a potential noninvasive marker of tumour aggressiveness. *J Magn Reson Imaging*. 2012;35(6):1365-71.
28. Tong T, Yao Z, Xu L, Cai S, Bi R, Xin C, et al. Extramural depth of tumour invasion at thin-section MR in rectal cancer: associating with prognostic factors and ADC value. *J Magn Reson Imaging*. 2014;40(3):738-44.
29. Akashi M, Nakahusa Y, Yakabe T, Egashira Y, Koga Y, Sumi K, et al. Assessment of aggressiveness of rectal cancer using 3-T MRI: correlation between the apparent diffusion coefficient as a potential imaging biomarker and histologic prognostic factors. *Acta Radiol*. 2014;55(5):524-31.
30. Sun Y, Tong T, Cai S, Bi R, Xin C, Gu Y. Apparent Diffusion Coefficient (ADC) value: a potential imaging biomarker that reflects the biological features of rectal cancer. *PLoS One*. 2014;9(10):e109371.
31. Sun YS, Zhang XP, Tang L, Ji JF, Gu J, Cai Y, et al. Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging for early detection of tumour histopathologic downstaging. *Radiology*. 2010;254(1):170-8.
32. Intven M, Reerink O, Philippens ME. Diffusion-weighted MRI in locally advanced rectal cancer : pathological response prediction after neo-adjuvant radiochemotherapy. *Strahlenther Onkol*. 2013;189(2):117-22.





# Chapter 9

Valorisation

In the last century, significant steps have been made in the treatment of cancer. Although there is no therapy to cure every type of cancer, in many types of cancer treatment strategies have evolved to make it a chronic illness that can be managed and controlled for a long period of time. Despite these developments, cancer is still in the top 10 of causes of death in the Western World. Colorectal cancer in specific is the 3rd and 2nd most common type of cancer in men and women, respectively and worldwide the 3rd most common cause of cancer death. Rectal cancer accounts for approximately one third of all colorectal cancers, with 500 new cancer diagnoses in the Netherlands every year. Patients with rectal cancer traditionally had a very poor prognosis, mainly due to the high risk of a local recurrence of the tumour after surgery. The introduction of standardised surgical techniques (total mesorectal excision), as well as neoadjuvant treatment (chemotherapy and radiation therapy) for patients with more advanced tumours have lowered the number of local recurrences drastically. These discoveries further stimulated the urge to personalize treatment for each individual patient. To decide which patient benefits from which treatment it is important to divide the patients in different risk categories based on their primary tumour status. There are three main risk categories: low risk tumours that can be treated with direct surgery without neoadjuvant treatment, intermediate risk tumours who receive a short course of radiotherapy before surgery, and the high risk group with locally advanced tumours who receive a long course of neoadjuvant chemoradiotherapy (CRT) to downstage the tumour before surgery. MRI plays a key role to stage rectal tumours in order to determine the tumour risk profile and thus the appropriate treatment for each individual patient. Recently there have been new developments to further personalize treatment in rectal cancer, the most important one being the introduction of ‘organ saving’ treatment strategies. Up until recently all patients with advanced rectal tumours treated with neoadjuvant chemoradiotherapy proceeded to surgery, regardless of the response to the chemoradiotherapy. In a small percentage ( $\pm 15\%$ ) of these patients, the response to chemoradiotherapy is so substantial that no residual viable tumour is found at surgical resection. For these patients it may now be considered to defer from radical surgery and instead monitor the patients (‘wait-and-see’) or perform a small local excision, thereby saving the rectum and preventing a stoma and other peri- and postoperative morbidity. In the light of these organ-saving treatments, it becomes more and more important to accurately evaluate the response to chemoradiotherapy with imaging or even predict before the start of treatment which patients are likely to show a good response to chemoradiotherapy so that the neoadjuvant treatment may be further optimized. Standard morphological MRI will no longer be sufficient to answer these more advanced

imaging questions and new techniques will need to be developed and studied to meet the new clinical needs.

One of the most important new techniques that has been developed and studied for rectal cancer imaging in recent years is diffusion-weighted imaging (DWI). DWI is a type of MR sequence that can easily be added to any routine rectal MRI protocol. It can provide radiologists with unique information on the cellularity of tissues and has shown great benefit for the differentiation between high-cellular (tumour) and non-cellular (non-tumour) tissues. At the start of this thesis, the use of DWI had mainly been studied in purely research settings and studies focussing on implementation of DWI in clinical settings were lacking. Moreover, previously studied methods of DWI analysis so far often included quantitative measurement methods that are impractical and time-consuming and therefore difficult to implement in daily practice. This triggered us to study practical ways to use DWI in a clinical setting, and to make it available and applicable for every radiologist, to avoid that this potentially promising and potentially clinically beneficial technique may get stuck in research settings. Therefore, the main aim of this thesis was to look for strategies and tools that can benefit the clinical use of DWI in different clinically relevant scenarios within the spectrum of rectal cancer treatment.

## RELEVANCE OF SCIENTIFIC RESULTS IN THIS THESIS

The results of the thesis offer several practical points of advice on how DWI may be used for rectal cancer management. First, it shows that after a certain training every radiologist can use DWI for tumour response assessment after neoadjuvant treatment, with results similar to those previously reported for expert readers in research settings. Moreover, we also discovered several important pitfalls and learning points, that can serve as a teaching reference for future radiologists. Second, this thesis shows that DWI can very practically be put to use to help select patients with a complete response of their lymph node metastases after chemoradiotherapy by simply counting the number of lymph nodes on DWI, a method that is easily applicable for daily use. Finally this thesis shows different ways in which quantitative analysis of DWI can be performed in a more practical manner. The use of DWI tumour volumetry has previously been shown to be promising. However due to the time-consuming nature of manual delineation of tumour volumes it has not yet found its way to clinical practice. This thesis shows that there are other ways to measure tumour volumes on DWI that are easier, quicker and offer similar results. Based on the results of this thesis, the use of more advanced quantification methods at this point has no clear added value for clinical use.

## TARGET POPULATION

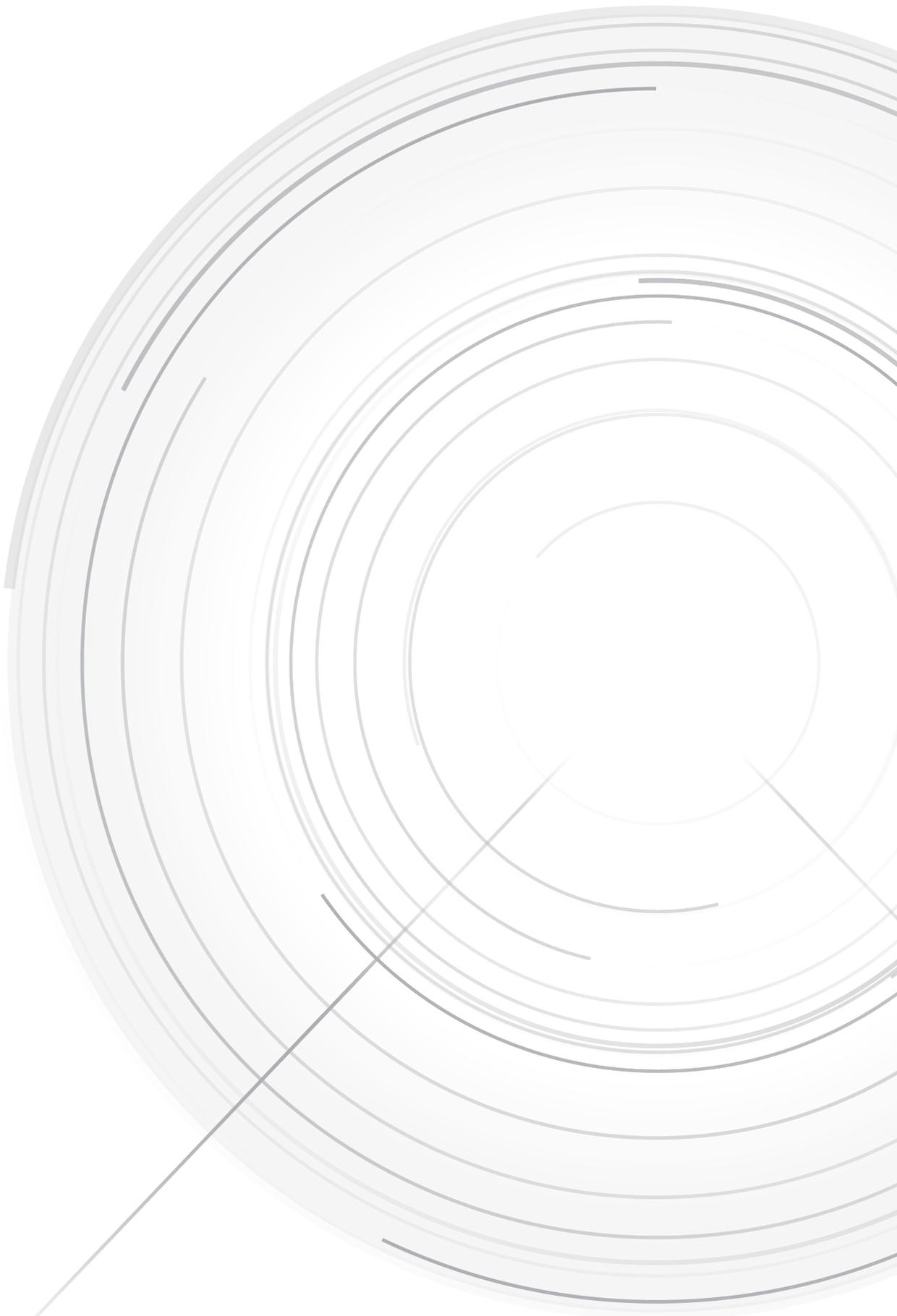
All patients with rectal tumours routinely undergo MRI as part of the clinical diagnostic path, according to the national guidelines. Patients with intermediate or high risk tumours then receive neoadjuvant treatment, followed by surgery (or in some cases by organ-saving treatments, as discussed above). These patients undergo a second restaging MRI to assess the tumour response. The target population includes all these patients undergoing MRI. In addition, the radiologists who assess the MR Images are part of the target population of this thesis, as the teaching of radiologists is a very important feature that will benefit the actual use of novel MR techniques and its implementation in clinical protocols.

## INNOVATION AND FUTURE

The role of imaging in the treatment of cancer has grown considerably with ongoing developments in cancer treatment. Introduction of novel imaging strategies such as the use of diffusion-weighted MR imaging in rectal cancer can facilitate the selection of patients for organ-saving treatments. However, there is no clear consensus on the best strategy to put DWI to use in daily practice. Although the results of these thesis already contribute to solving this issues, further research is still required, particularly with respect to quantitative DWI evaluation. An important next step is multicenter validation of (semi-)automated DWI volumetry methods

Moreover, the same path to translate research into daily practice should be explored for other novel imaging techniques apart from DWI (which was the main focus of the current thesis), for example perfusion and fibrosis imaging. Ultimately, the future most probably lies in finding a way to smartly combine these different techniques. Such multiparametric imaging evaluations most likely will offer the most comprehensive insights into tumour biology and can be expected to contribute a better understanding of cancer and treatment processes and thereby ultimately help to further personalize and optimize treatment individual cancer patients.







# Chapter 10

Summary

The aims of this thesis were to investigate practical ways to enhance the role of diffusion-weighted imaging (DWI) in daily rectal cancer staging and restaging and to find ways to facilitate the role of quantitative MRI and DWI image analysis in rectal cancer.

## **Part I Diffusion-weighted imaging for daily rectal staging**

**Chapter 2** explores the measuring of the tumour length and height with DWI as an alternative to the commonly used standard T2-Weighted imaging (T2W) sequences. In previous studies, it was shown that T2W-MRI tends to overestimate tumour length. DWI may help to overcome this problem. Indeed we found that the tumour length was significantly smaller when measured on DWI instead of standard T2W, both at primary staging as well as for restaging after chemoradiotherapy. This consequently translated into a larger tumour height when measured on DWI, particularly at restaging. However, differences were small and probably will have limited impact on treatment decision-making.

In **chapter 3** we tested whether use of DWI could aid in differentiating between patients with and without residual nodal metastases in the restaging setting, after completion of chemoradiotherapy. We hypothesized that – given the known excellent sensitivity of DWI to detect lymph nodes – the absence of nodes on DWI after CRT would be proof that all metastatic nodes have vanished. We indeed found in a group of 90 patients that the 10 patients (14%) who did not had any visible nodes on the b1000 DWI scans all had a confirmed yN0 status (resulting in a sensitivity of 100% for DWI). Unfortunately, this does not work as well the other way around: most patients with an yN0 status still had visible lymph nodes on DWI resulting in a low specificity of only 14%. We thus concluded that – although it is not a frequent finding – the absence of lymph nodes on a restaging DWI can play a supporting role in the selection of node-negative patients which is an important factor when considering patients for organ-preservation after CRT.

In **chapter 4**, we explored how expert teaching can be used to improve the diagnostic performance of general or resident radiologists when assessing DWI to evaluate the response of rectal tumours after neoadjuvant treatment. Results showed that after training by an expert readers with approximately 50 cases non-expert readers are able to achieve a diagnostic accuracy of around 85% to differentiate between patients with a complete response and residual tumour, similar to those previously reported for experts. Furthermore, we documented five main interpretation pitfalls that may serve as a reference to teach future readers interested in the use of DWI for rectal tumour response evaluation.



## Part II – Quantitative analysis of rectal cancer imaging – a practical approach

In **chapter 5** we reviewed all available literature on tumour size and volume measurements on standard T2-weighted MR for rectal tumour response assessment after CRT. From these results we extracted reported size and volume threshold, which we then tested and validated in a new prospective multicenter cohort to determine their potential clinical value to assess and predict response. We tested 3 different reported measurement methods: one-dimensional tumour length, 3 dimensional measurements (length x height x width) and whole tumour volumetry. Results showed that whole tumour volume measurements provided the best results and might be helpful in rectal cancer response assessment when using selected cut-off values, with accuracies up to 80%. MRI whole tumour volumetry was mainly accurate to assess a complete tumour response (ypT0) after CRT, but less helpful to assess T-downstaging or TRG1-2. Measurements of tumour length or three-dimensional tumour size are not useful with accuracies ranging only between 48% and 56%.

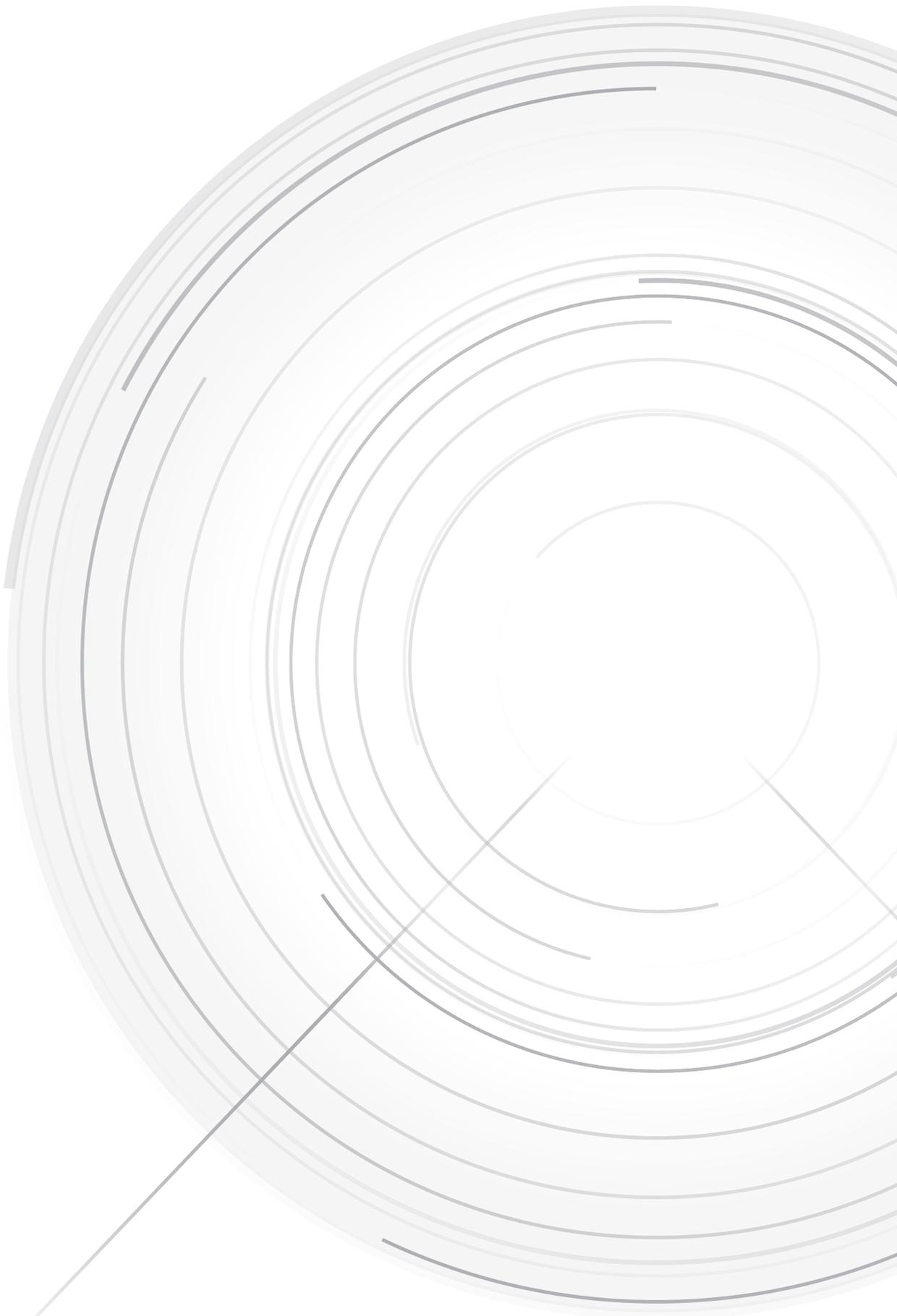
In **chapter 6** we subsequently focussed on whole tumour volumetry using DWI instead of T2-weighted MRI. Previous studies have shown that DWI-volumetry is more useful than standard T2W-volumetry to assess response, but the typically used manual segmentation methods required to calculate tumour volumes are far too time consuming to be useful in daily practice. For this project we studied the use of automated segmentation techniques (that can be performed by non-radiologists) and a semi-automated method (where automatically generated volumes are manually adjusted by expert radiologists) and compared this to manual tumour segmentation. We found that there is excellent agreement between the manual delineation method and the (semi) automated segmentation tools, with Intraclass coefficients (ICCs) ranging from 0.80-0.91 at primary staging and 0.53-0.66 after CRT for the automated segmentation, and 0.91-0.97 (staging) and 0.61-0.75 (after CRT) for the semi-automated segmentation. Moreover, both methods can save the radiologist a considerable amount of time compared to manual delineation (especially for the primary staging MRI) with a decrease in measurement time of up to 18 minutes per patient for both the semi-automated and automated approach. Given the superior reproducibility (compared to manual delineation) of the semi-automated method, this strategy could have real potential to provide a good alternative to manual DWI tumour volumetry for rectal tumour response assessment. Future studies should focus on further optimising DWI image acquisition and segmentation algorithms as well as on integrating these tools into clinical viewing systems to allow implementation in daily practice and further improve time efficacy.



Manual whole-volume tumour delineations are also typically used to obtain tumour Apparent Diffusion Coefficient (ADC) measurements, the main quantitative measure of DWI reflecting tumour cellularity. In **chapter 7** we investigated an alternative way to overcome the fact that these delineations are time consuming and cumbersome. We investigated whether non-precise delineation combined with histogram analysis (to compensate for delineation inaccuracies) could provide the same ADC information as when derived from precise delineations. Moreover, we investigated whether the addition of histogram analysis would provide valuable extra information to predict prognostic tumour profiles and response. We found that the 45th percentile of the ADC histogram derived from non-precise delineation correlated well with the mean ADC derived from exact manual tumour delineation, and may thus be used as an alternative measure whilst considerably saving time in the delineation process. We could unfortunately not confirm results previously reported that ADC measurements have potential as a clinical prognostic biomarker. Moreover, addition of histogram ADC analysis did not offer any additional prognostic information. The only parameter found to correlate with important prognostic features (N+ and MRF+ status), was the DWI tumour volume.









# Chapter 10

Samenvatting

De doelstellingen in deze thesis waren om te onderzoeken of er een praktische manier is om de rol van diffusie gewogen MRI (DWI) te vergroten in de dagelijkse praktijk bij zowel primaire stadiëring als bij de respons evaluatie van het rectumcarcinoom, en daarnaast om te kijken of het mogelijk is om het gebruik van kwantitatieve analyse van MRI en specifiek DWI bij het rectumcarcinoom te faciliteren.

## **Deel I Diffusie gewogen MRI in de dagelijkse praktijk voor de stadiëring van rectumcarcinomen**

In **hoofdstuk 2** wordt onderzocht of het meten van de tumorlengte en –hoogte op DWI een goed alternatief is voor het meten hiervan op de standard gebruikte T2-gewogen (T2W) sequentie. Eerdere studies hebben laten zien dat op T2W MRI de lengte van de tumor kan worden overschat. DWI kan dit probleem wellicht helpen. We vonden inderdaad dat de tumorlengte significant kleiner was wanneer deze werd gemeten op DWI in plaats van standaard T2W MRI beelden. Dit was zowel het geval bij de primaire stadiëring als bij respons evaluatie na behandeling met chemoradiotherapie (CRT). Dit had als gevolg dat de tumor hoogte op DWI beelden groter werd gemeten, met name bij respons evaluatie. Echter, de gemeten verschillen waren klein en waarschijnlijk is de invloed beperkt voor de besluitvorming rondom behandeling.

In **hoofdstuk 3** hebben we getest of het gebruik van DWI zou kunnen helpen bij het differentiëren of patiënten wel of geen lymfeklier metastasen hebben op het moment van de respons evaluatie na de neoadjuvante behandeling met CRT. Onze hypothese was, gezien het feit dat DWI heel sensitief is voor het detecteren van lymfeklieren, dat de afwezigheid van klieren op DWI bewijs zou kunnen zijn voor het feit dat alle lymfeklier metastasen zouden zijn verdwenen. We vonden inderdaad in een groep van 90 patiënten dat de 10 patiënten (14%) bij wie geen klieren meer werden gezien op de b1000 DWI scans allemaal een yN0 status hadden (wat maakt dat de sensitiviteit 100% was voor DWI). Jammer genoeg werkt dit niet andersom: bij de meeste patiënten met een yN0 status waren nog lymfeklieren zichtbaar op de DWI, waardoor de specificiteit van deze bevinding slechts 14% is. Derhalve hebben we geconcludeerd dat, ook al komt het niet vaak voor, de afwezigheid van lymfeklieren op de DWI scan voor respons evaluatie een ondersteunende rol kan hebben voor het selecteren van patiënten met een yN0 status. Dit is een belangrijke factor voor het bepalen of een patiënt in aanmerking komt voor orgaan-preservatie na CRT.

In **hoofdstuk 4** hebben we onderzocht of scholing door experts kan worden gebruikt om de diagnostische vaardigheden bij het beoordelen van DWI voor respons evaluatie van het rectumcarcinoom van algemene radiologen of artsen in

opleiding tot radioloog te verbeteren. De resultaten lieten zien dat na een scholing door experts van ongeveer 50 casussen ervoor zorgde dat een diagnostische accuraatheid van 85% werd gehaald voor het onderscheiden van patiënten met of zonder complete tumorrespons door de niet-experts. Dit is vergelijkbaar voor de accuratesse van experts die in eerdere studies werd beschreven. Bovendien hebben we de vijf belangrijkste valkuilen gedocumenteerd van de interpretatie van DWI, die als referentie kunnen worden gebruikt om in de toekomst radiologen te scholen die zijn geïnteresseerd in het gebruik van DWI voor tumorrespons evaluatie bij rectumtumoren.

## **Deel II – Kwantitatieve analyse van MRI bij rectumcarcinoom – een praktische benadering.**

In **hoofdstuk 5** wordt een overzicht gegeven van de beschikbare literatuur over het meten van tumor grootte en volume op standaard T2W MRI voor de respons evaluatie na CRT. Uit deze resultaten werden verschillende grootte en volume afkappunten gehaald, die vervolgens prospectief werden getest en gevalideerd in een nieuwe multicenter cohort. Hier werd bij 3 verschillende meetmethoden (lengte, 3D-grootte en volume) bepaald wat de klinische waarde was van deze afkappunten om de respons evaluatie te bepalen en voorspellen. Er werd gevonden dat het meten van het tumorvolume de beste resultaten liet zien en dat dit bruikbaar kan zijn wanneer de juiste afkapwaarden worden gebruikt, met een accuratesse tot 80% voor het bepalen van de tumorrespons. Hierbij lijkt het meten van het volume van de tumor vooral bruikbaar om te bepalen of de patiënt een complete tumorrespons heeft na CRT (ypT0). De accuratesse van het meten van tumorlengte en 3D volume op T2W liet lage accuratesse zien (tussen 48% en 56%) .

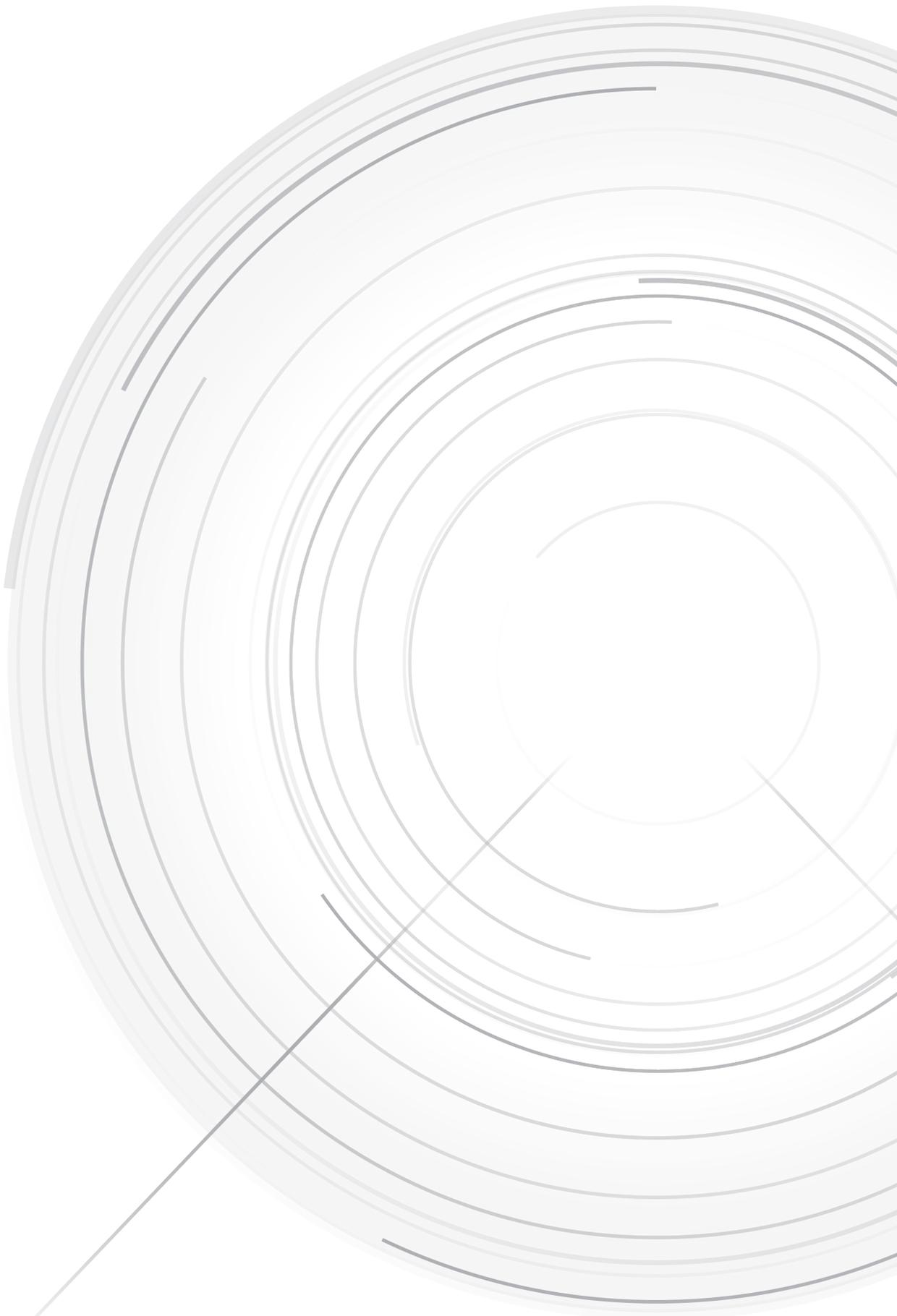
In **hoofdstuk 6** lag de focus vervolgens op het bepalen van tumor volume op DWI beelden in plaats van de T2W beelden. Eerdere studies lieten al zien dat volumemetingen op DWI geschikter zijn voor het bepalen van respons na CRT, maar doorgaans wordt het volume gemeten door het handmatig intekenen van de tumorgrenzen op de beelden. Deze methode kost te veel tijd om in de dagelijkse praktijk te gebruiken. Derhalve hebben we in dit project onderzoek gedaan naar het gebruik van geautomatiseerde segmentatie technieken en semi-automatische methoden (waarbij de radioloog het automatisch gegenereerde volume handmatig aanpast). We vonden hierbij dat er sprake is van een uitstekende overeenstemming tussen de handmatige en de (semi-)automatische methoden, waarbij de Intraclass coefficients (ICCs) tussen 0.80-0.91 lagen bij primare tumoren en tussen 0.53-0.66 na CRT voor de automatische methode en voor de semi-automatische methode tussen 0.91-0.97 (primair) en 0.61-0.75 (na CRT). Bovendien,

beide methodes besparen de radioloog een aanzienlijke hoeveelheid tijd vergeleken met de handmatige methode. Dit geldt vooral voor de volume metingen bij primaire tumoren, waar tot wel 18 minuten per patient werd bespaard met zowel de automatische als semi-automatische methode. Omdat de semi-automatische een betere reproduceerbaarheid had (vergeleken met de handmatige methode), heeft deze meetmethode potentie om een goed alternatief te worden voor de handmatige tumor volume metingen voor respons evaluatie.

Bij het bepalen van handmatige tumor volumes op DWI, kan ook de Apparent Diffusion Coefficient (ADC) worden bepaald, wat een kwantitatieve meting is voor de cellulariteit van de tumor. In **hoofdstuk 7** is onderzocht of er een alternatieve manier is voor deze handmatige, tijdrovende methode om de tumor ADC te bepalen. Hierbij is onderzocht of ruime omlijning van de tumor gecombineerd met Histogram analyse (om te compenseren voor ruis) vergelijkbare ADC informatie kan verschaffen als de exacte tumor delineatie methode. Daarbij werd ook onderzocht of het toevoegen van histogram analyse wellicht meer informatie kan verschaffen over het prognostische tumorprofiel dan de standaard gebruikte gemiddelde ADC waarde. Hierbij vonden we dat het 45e percentile van de ruime delineatie het beste overeenkomst met het gemiddelde ADC van de exacte delineatie method. Dit betekent dat de ruime delineatie de 45e percentile kan worden gebruikt als tijdbesparend alternatief. We konden helaas het eerder gerapporteerde potentieel van ADC als klinische prognostische biomarker bevestigen in deze studie. Ook bleek dat het toevoegen van ADC histogram analyse geen toegevoegde waarde had. De enige parameter die correleerde met belangrijke prognostische factoren (N+ en MRF+ status), was het volume van de tumor op DWI.









# Chapter 10

List of publications



DMJ Lambregts, **MM van Heeswijk**, A Delli Pizzi, S van Elderen, LA Andrade, HGM Peters, P Kint, M Osinga-de Jong, S Bipat, H Ooms, MJ Lahaye, M Maas, GL Beets, FCH Bakers, RGH Beets-Tan

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

*Eur Rad (2017) Epub ahead of print.*

**MM van Heeswijk**, DMJ Lambregts, M Maas, MJ Lahaye, Ayas S, J Slenter, GL Beets, FCH Bakers, Beets-Tan RGH. Measuring the Apparent Diffusion Coefficient in primary rectal tumours – is there a benefit in performing histogram analyses? *Abdom Radiol (NY)*. 2017 Jun;42(6):1627-1636.

**MM van Heeswijk**, DMJ Lambregts, WM Palm, BMF Hendriks, MM, GL Beets, RGH Beets-Tan

DWI for assessment of rectal cancer nodes after chemoradiotherapy: is the absence of nodes on DWI proof of a negative nodal status? *AJR 2017; 208:W79–W84.*

**MM van Heeswijk**, DM Lambregts, JJ van Griethuysen, SA Oei, S-X Rao, CA de Graaff, RF Vliegen, GL Beets, N Papanikolaou, RGH Beets-Tan. Automated and semi-automated segmentation of rectal tumour volumes on diffusion-weighted MRI: can it replace manual volumetry?

*Int J Radiat Oncol Biol Phys; 2016 Mar 15;94(4):824-31.*

MH Martens, **MM van Heeswijk**, JJ van den Broek, S-X Rao, V Vandecaveye, RA Vliegen, WH Schreurs, GL Beets, DMJ Lambregts, RGH Beets-Tan. Prospective multicenter validation study of MRI volumetry for response assessment after preoperative chemoradiation in rectal cancer: can the results in literature be reproduced?

*Int J Radiat Oncol Biol Phys; 2015 Dec 1;93(5):1005-14.*

**MM van Heeswijk**, DMJ Lambregts, L den Ouden, MC Ageitos-Casais, M Maas, GL Beets, FCH Bakers, RGH Beets-Tan. Measuring rectal tumour height and length on MRI - effect of using diffusion-weighted MRI instead of T2-weighted MRI. *Submitted for publication.*







# Chapter 10

Dankwoord

Dan het dankwoord. Natuurlijk hét moment om veel mensen te bedanken, maar ook om te laten zien hoeveel fijne mensen ik de afgelopen jaren heb ontmoet. Het spijt me oprecht als ik iemand ben vergeten, ik had dit proefschrift nog gemakkelijk kunnen verdubbelen in dikte door lofzangen en dankbetuigingen.

Prof. Dr. Regina Beets-Tan, beste Regina, heel erg bedankt dat je me hebt overtuigd om onderzoek te gaan doen! Wie had dat gedacht een aantal jaar geleden. Ook al zaten we de laatste maanden wat verder uit elkaar, je vertrouwen en enthousiasme betekenen veel voor mij. Hartelijk dank voor de kansen die je me hebt gegeven. Ik ben trots om jou als promotor te hebben en van het Recteam deel uit te mogen maken.

Prof. Dr. Geerard Beets, beste Geerard, ook jou wil ik van harte bedanken voor de afgelopen jaren. Jouw nuchtere kijk en relaxte instelling maken je een bijzonder fijn persoon om mee te werken. Je houdt altijd de patiënt in het oog. Ik weet zeker dat het watch-and-wait project zich verder zal verspreiden door het land en internationaal! Hartelijk dank en succes verder in Amsterdam!

Dr. Doenja Lambregts, beste Doenja, jouw onuitputbaarheid en inzicht zijn jaloersmakend. Een betere begeleider had ik me niet kunnen wensen. Je enthousiasmeert, bent rechtvaardig en bracht me regelmatig het overzicht terug. Ik vond het fijn om met je te sparren, maar zeker ook om een borrel te drinken. Succes in de toekomst, die zeker gouden bergen zal hebben. En ik beloof dat ik nooit meer statistiek met een rekenmachine zal doen.

Prof. dr. Roy Kruitwagen, het is een eer dat u voorzitter van mijn beoordelingscommissie wilde worden. Ik wil u en de andere leden van de beoordelingscommissie, Prof. dr. Simon Robben, Prof. dr. Laurent Stassen, Prof. Dr. Ulke van der Heiden en Dr. Elleke Dresen, alsmede de leden van de oppositiecommissie, hartelijk danken voor de kritische beoordeling van mijn proefschrift.

Lieve Rianne en Britt, heel veel dank voor alles wat jullie de afgelopen jaren voor mij hebben betekend. Jullie hebben nogal wat met me te stellen gehad, maar ik heb jullie hopelijk ook veel gebracht (en niet alleen in gebakken vorm). Rianne, jouw 'doe-maar-normaal/rustig' instelling was precies wat ik af en toe nodig had. Die instelling is ook de reden dat ik hier niet te veel veren in je derrière zal steken. Je weet namelijk zelf precies hoe blij ik met jou als collega ben. Britt, we zaten soms angstaanjagend veel op 1 lijn, of in ieder geval vaak in herkenbare situaties. Ik wil je bedanken voor je gezelligheid, je lunch als je die weer niet op kreeg, en tot slot natuurlijk een quote: "If I don't input those numbers... it doesn't make much of a difference."



Het Recteam. Wat ben ik trots om met jullie een clubje met zo'n vreemde obsessie voor keutels te vormen! Mo: dank voor alle keren dat je me advies gaf. Niet alleen over Radiologie, statistiek en onderzoek, maar ook over pensioen, make-up en carrière. Max, jouw kijk op het leven bevalt me wel, dank voor de keren dat ik je om advies vroeg! Milou, ik hoop dat je niet onderschat hoeveel ik van jou heb geleerd. Ik was maar wat trots dat ik van je student later je collega werd. Hetzelfde geldt voor jou Luc! Joost: jouw kennis en vaardigheden zijn snel onmisbaar geworden in het team, maar ik was vooral blij dat er nog een koffie liefhebber op onze kamer kwam! Ook Marit, Rebecca, Jasenko en Elias, hartelijk dank voor alle gezelligheid!

Frans Bakers, je enthousiasme voor abdominale radiologie is bijzonder inspirerend. Bedankt voor alle keren dat je de tijd nam om me iets uit te leggen en het vertrouwen dat je met hebt gegeven. Ik hoop dat onze paden snel weer kruisen!

Natuurlijk de rest van de afdeling Radiologie. Ik vond het ontzettend leuk om alles bij elkaar 3.5 jaar op de afdeling in Maastricht te mogen rondlopen! De sfeer tussen assistenten, staf en alle rest van het personeel is bijzonder prettig en heeft mijn voorliefde voor radiologie alleen maar versterkt. Een speciale dank voor alle collega's met wie ik op de afdelingsfeesten en congressen een dansje kon wagen. Janneke Houwers, Miguel Palm en Stanley Oei, bedankt voor de fijne samenwerking. En ook zeker een speciaal woord van dank voor Prof. Robben, beste Simon, bedankt voor alle inzichten en advies.

De dames van het secretariaat Radiologie: Monique, Elfie, Christianne en Monique. Bedankt voor alle hulp voor mij als coassistent, WESP-student, GEZP-student EN als onderzoeker. Ik denk dat ik niemand zoveel heb lastig gevallen als jullie. Bedankt voor de gelpennen, engelse drop, maar vooral voor de gezellige praatjes afgelopen jaren.

Jos Slenter, zoals ik je vaker heb gezegd: je bent een ontzettend fijne vent. Dank dat ik je altijd domme vragen mocht stellen. Natuurlijk ook bedankt Marc voor alle technische ondersteuning!

Dank ook alle WESP-studenten en student-assistenten die ons hebben geholpen de afgelopen jaren in de kamer met de keutel op de deur. Het was ontzettend leuk om jullie inhoudelijk iets over de behandeling van rectumcarcinomen te leren. Maar daarnaast vond ik het vooral leuk dat we jullie een beeld hebben geven over de gang van zaken in een ziekenhuis en in de wetenschap. Babs, van je begeleider naar collega, maar vooral ook naar vrienden. Ik hoop dat we ook later als collega's kunnen werken! Geertje, ik vond het ontzettend leuk zo'n goede vriendin kennis te



laten maken met onderzoek. Bedankt ook collega-onderzoekers bij Radiologie en Heelkunde, met name Paul, Yvonne, Martine, Nienke, Koos, en natuurlijk Jeroen van 2 deuren verderop.

Hartelijk dank aan Roland Kersemakers en alle MRI laboranten voor jullie uitleg, geduld met ons als we weer helemaal vanuit UNS40 moesten komen, het meedenken en de gezelligheid in de afgelopen jaren!

De dames en heren van de balie: Anita, Anja, Annie, Audrey, Bea, Daisy, Guus, Imane, Irene, Jan, Jeanette, Lieke, Maria, Rosita, Sylvia en Vivian. Ik wil jullie ontzettend bedanken voor alle gezellige uren die ik met jullie heb doorgebracht en vooral voor jullie onvermoeibare inzet voor het plannen van patiënten die van ver kwamen! Zonder jullie was dit boekje en het wait-and-see beleid niet wat het nu is! Ook hartelijk dank voor al je hulp met alle beelden van alle verwijzingen, Frans Kusters.

De afdeling Chirurgie, prof. dr. Stassen, Drs Keymeulen, Dr. Boerma, Dr. Engelen (als onze voorganger) en in het bijzonder Jarno Melenhorst en Stephanie Breukink. Jarno, dank voor al het vertrouwen dat je in ons had. Stephanie, bedankt dat je ons in het oog hield. Christel Gielen, ik heb ontzettend respect voor jouw inzet en omgang met patiënten. Zoals jij ken ik er maar één.

Alle radiologen, chirurgen, MDL-artsen, verpleegkundigen en epidemiologen die aan de MRI2-studie en de wait-and-see studie hebben meegewerkt, bedankt! Vooral ook Jeroen Leijtens, Jeroen Heemskerk, Nicky Peters, Roy Vliegen en Peter Kint hartelijk dank voor de fijne samenwerking. Nikos Papnikolau, thank you for your help and collaboration.

Mijn collega arts-assistenten en de radiologen in het UMC Utrecht, dank voor het warme welkom in een nieuwe stad en nieuw ziekenhuis!

Lieve Richard, bedankt voor het prachtige omslag ontwerp! Roy Sanders, bedankt voor de mooie opmaak van dit proefschrift.

Dan nu een stuk over mijn vrienden, wat waarschijnlijk voornamelijk interessant is voor de desbetreffende personen. Ik prijs mijzelf ontzettend gelukkig met een grote groep fijne mensen om mij heen, niet alleen in Maastricht en Utrecht, maar door het hele land. Allereerst lieve musketiers: Merel, Aimée en mijn homie Kim, jullie hebben mij tijdens mijn studententijd ontzettend veel plezier gebracht. Wat fijn dat we allemaal zo onze plek hebben gevonden. Lieve dames: Yvonne, Maartje, Jeske en Jolijn, vanaf onderwijsgroep 1 een vast klikje en een stabiele



basis. We kunnen altijd terugvallen op elkaar en schromen niet advies te geven. Ik ben dankbaar dat jullie mijn vriendinnen zijn. Lieve INKOM helden, met vooral Chantie, Baaten, Bart, Jolien, Geertje, Lorân, mijn nieuwe huisgenootje Maron en mijn collega Paul (maar zeker nog veel meer!): een van de beste beslissingen die ik ooit heb gemaakt is solliciteren voor de Centrale Post. Jullie zijn stuk voor stuk energiek, houden van aanpakken en ik voel me helemaal thuis bij jullie. Ik vind het fijn nu dichterbij de meesten van jullie te wonen! Lieve Tomas, wij hebben maar 1 woord nodig: knuffel! Lieve Lucas, mijn oudste vriend. Al meer dan meer dan 25 jaar vriendschap, je bent onmisbaar! Daarnaast zijn er nog een paar namen: Dul, Paulus, Nadine, Gijs, Daphne, Steffie, Yvette, want ook jullie ben ik niet vergeten!

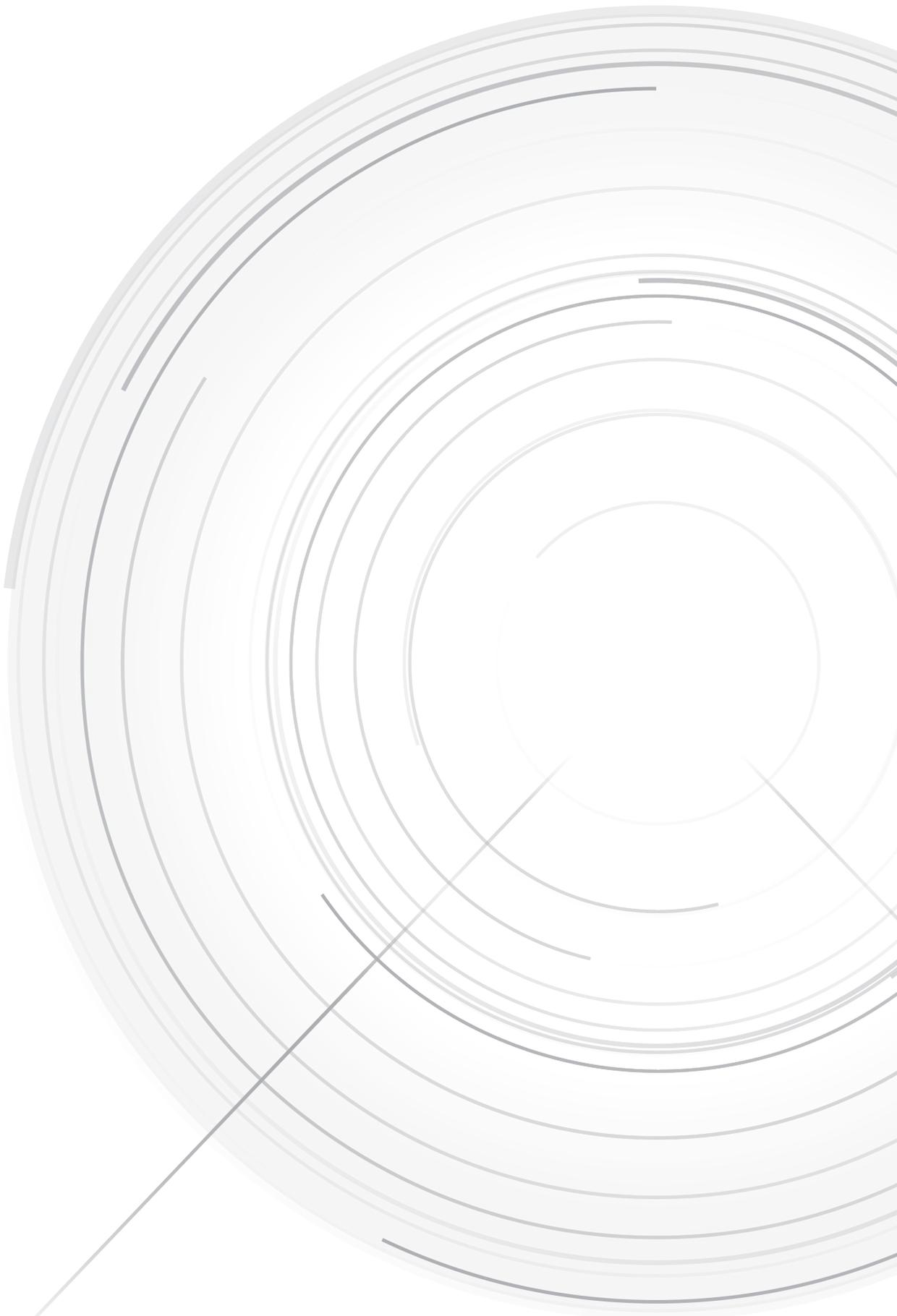
Ooh Juul... Lieve Julia, Guliorupio, mijn paranimf/wingman. Ondanks dat we elkaar meteen niet mochten bij de eerste ontmoeting, ben je in rap tempo uitgegroeid tot een van de beste vriendinnen die een mens zich kan wensen. Je kan schaamteloos met me lachen om al die bizarre situaties, maar hebt ook altijd wijze raad en haalt me uit mijn chaos als dat weer eens nodig is. Je bent attent, gezellig en, ook al laat je me soms mijn eigen grenzen testen, heel lief. Ik ben blij dat we, na al die avonden bellen op de fiets na het werk in Maastricht en Leusden/Naarden, nu eindelijk weer in dezelfde stad wonen.

Lieve Merel, mini(-Hommen), muis, ik weet nog goed de eerste keer dat wij elkaar ontmoetten in café de Uni tijdens de INKOM 2007. Het begin van een hechte en vooral hele fijne vriendschap. Je bent een steun en toeverlaat, eerlijk als nodig en je bent regelmatig mijn geweten. En daarnaast vind ik je vooral heel lief! Je vriendschap betekent veel voor me, natuurlijk moet jij mijn paranimf zijn. Omdat we nog maar veel zullen culinairen.

Lieve familie en iedereen die ik als familie beschouw (Van Heeswijk, Beerens, Driessen, Kloek), hartelijk dank voor jullie vertrouwen in mij en voor al jullie gezelligheid! Cees, dank voor alle bemoedigende woorden. Ria, dank voor al je support en kaartjes de afgelopen jaren. Lieve Oma, dit boekje is voor u. Ik hoop dat ik u trots heb kunnen maken. En luister naar uw dokter-kleindochter: u mag best ook af en toe twee paracetamolletjes nemen.

Lieve Esau, van grote zus naar mijn gelijke. Ik ben ontzettend trots op je doorzettingsvermogen en gun jou alles wat je hartje begeert. Lieve papa en mama, kleine mimi's worden groot. Bedankt voor alles wat jullie mij hebben geleerd. En daarbij vooral ook dank voor jullie warme opvoeding, geduld en onuitputbare interesse. Zonder de steun van jullie drieën waren dit allemaal lege bladzijden geweest.







# Chapter 10

Curriculum vitae





Miriam Maria van Heeswijk was born on May 28th, 1989 in Waalwijk, the Netherlands. After obtaining her Gymnasium degree at the Dr. Mollercollege in Waalwijk in 2007, she started her medical study at the Faculty of Health care, Medicine and Life sciences at Maastricht University.

Between her 5th and 6th year of her medical study she also obtained a Master of Science-degree at Maastricht University in Healthcare, Policy, Innovation

and Management. She was a graduate Intern at Medtronic Hospital Solutions where she wrote her Master Thesis '*Carelink Express Lane: Redesigning the patient process of pacemaker follow-up with CareLink® and Mediconnect*', on patient logistics and innovation implementation in the Reinier de Graaf Gasthuis, Delft, NL.

In her final year of her medical study she did her clinical internship at the department of Radiology and a research internship in the research team of Prof. Dr. Regina Beets-Tan in Maastricht, where she co-wrote her first medical publication. In September 2014, immediately after graduation, she continued this research as a PhD student at Maastricht University (GROW) under the supervision of Prof. Dr. Regina Beets-Tan and Prof. Dr. Geerard Beets. During this period she was one of the coordinators of the Dutch Colorectal Cancer Group Multicenter trial: the Maastricht Rectal Imaging (MRI<sub>2</sub>) study. She was nominated as an ESGAR top 20 presenter in 2015. For the yearly national Radiology congress in the Netherlands she wrote a top 5 abstract in 2015 and gave a key-note presentation in 2017.

In Januari 2017, she started working as a Radiology resident at the University Medical Centre in Utrecht under the supervision of dr. R.J. Nijvelstein, with a special focus on medical education.



