

Multiparametric image modelling

Citation for published version (APA):

Schurink, N. W. (2022). *Multiparametric image modelling: predicting treatment response in rectal cancer*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20220607ns>

Document status and date:

Published: 01/01/2022

DOI:

[10.26481/dis.20220607ns](https://doi.org/10.26481/dis.20220607ns)

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

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Multiparametric image modelling

predicting treatment response
in rectal cancer

Niels Schurink



Multiparametric image modelling

predicting treatment response
in rectal cancer

Niels Schurink

ISBN: 978-94-6421-701-8

Cover design & lay-out: Esther Beekman (www.estherontwerpt.nl)

Cover riso printed by: De Kijm & zonen, Den Haag

Printed by: Ipskamp printing, Enschede

The publication of this thesis was financially supported by the Netherlands Cancer Institute, Maastricht University and the Dutch Cancer Society (project number 10138).

©2022 N.W. Schurink

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

Multiparametric image modelling: predicting treatment response in rectal cancer

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. Dr. Pamela Habibović
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op dinsdag 7 juni 2022 om 16:00 uur

door
Nicolaas Wilhelm Schurink

Promotor:

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

Copromotores:

Dr. D.M.J. Lambregts, Antoni van Leeuwenhoek Ziekenhuis

Dr. S.R. van Kranen, Antoni van Leeuwenhoek Ziekenhuis

Beoordelingscommissie:

Prof. dr. J.E. Wildberger (voorzitter)

Dr. S.O. Breukink

Prof. dr. C.A.M. Marijnen, Technische Universiteit Delft

Prof. dr. F.M. Mottaghy

Prof. dr. W.J. Niessen, Technische Universiteit Delft



Contents

Chapter 1	General introduction, aim and outline of the thesis	9
Chapter 2	Diffusion-weighted imaging in rectal cancer: current applications and future perspectives	21
Chapter 3	Value of combined multiparametric MRI and FDG-PET/CT to identify well-responding rectal cancer patients before the start of neoadjuvant chemoradiation.	57
Chapter 4	Studying local tumour heterogeneity on MRI and FDG-PET/CT to predict response to neoadjuvant chemoradiotherapy in rectal cancer.	79
Chapter 5	Sources of variation in multicentre rectal MRI data and their effect on Radiomics feature reproducibility.	97
Chapter 6	Development and multicentre validation of a multiparametric imaging model to predict treatment response in rectal cancer.	117
Chapter 7	General discussion	139
Appendices	Summary	152
	Samenvatting	154
	Scientific impact	156
	List of publications	160
	Dankwoord	162
	Curriculum Vitae	166



General introduction,
aim and outline of
the thesis

INTRODUCTION

The standard treatment for locally advanced rectal cancer (LARC) is neoadjuvant long-course chemoradiotherapy (CRT) followed by complete surgical resection of the rectum and its surrounding mesorectal compartment (total mesorectal excision or 'TME') [1–4]. The main aim of CRT is to downstage the tumor prior to surgery, which leads to a complete pathological tumor response (i.e. no residual tumor cells are found in the surgical specimen) in up to 15-27% of the cases [5, 6]. This has sparked debate on the justification of invasive surgery in these patients [7], as TME has a high morbidity, significant mortality risk (2-8%, rising up to 30% in elderly patients [8, 9]), and often results in a temporary or permanent colostomy.

Recent studies have shown that patients that have no detectable tumor on post-treatment imaging (clinical complete response) may also be managed with active surveillance ("Watch-and-Wait" (W&W)) instead of surgery. W&W has been shown to result in similar overall survival as radical surgery [10–16], but leads to better functional outcome and quality of life [7, 16–19]. Achieving a complete response is therefore nowadays an increasingly important goal for patients.

If we could predict at the start how well patients will respond to chemoradiotherapy, this could create opportunities to further personalize treatment and improve patient outcomes. In patients likely to respond well, neoadjuvant treatment may be intensified to increase their chance of a complete response. Likewise, potentially harmful CRT may be avoided in patients that are not expected to benefit from it. Patients with smaller (cT1-2 or early stage cT3) tumors typically show higher complete response rates (up to 30-50%) [20–22]. These early tumors are currently treated with immediate surgery without CRT with good prognostic outcomes. If treatment response can be reliably predicted, CRT may become a viable alternative for these patients in the future. To date, offering CRT to small tumors to achieve W&W is still experimental and only performed in trial settings [23, 24]. Nevertheless, it is a development that further stresses the need for accurate response prediction. The ultimate goal here is to offer more patients a chance on organ-preservation.

The role of imaging for response prediction

Anatomical imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT) play a key role in the staging and treatment stratification of oncologic patients by offering detailed information on the location, size, and

spread of tumor lesions. The introduction of 'functional' molecular imaging has further transformed the field of oncological imaging. Functional MRI sequences such as diffusion-weighted imaging (DWI) and dynamic contrast enhanced (DCE or 'perfusion') imaging allow visualization and quantification of biological tumor properties such as tumor cell structure [25, 26] and angiogenesis [26, 27]. Metabolic tumor activity can be studied using 18F-fluorodeoxyglucose positron emission tomography (FDG-PET) [25, 28]. Hybrid PET/CT and more recently introduced PET/MRI systems integrate the benefits of functional imaging with high-resolution anatomical imaging within a single acquisition [28, 29]. Another important evolution has been the clinical introduction of modern post-processing and artificial intelligence (AI) techniques such as 'Radiomics'. With Radiomics, large numbers of quantitative imaging 'features' are extracted from routinely acquired imaging studies [30, 31]. These features can be used to describe a radiological phenotype of a disease and provide new insights into underlying tumor characteristics such as for example tumor heterogeneity, a feature that has often been associated with tumor aggressiveness and response outcomes [32–35]. Parameters derived from anatomical imaging, functional imaging and Radiomics have increasingly been studied in recent years as 'biomarkers' with the potential to assess response to neoadjuvant treatment in rectal cancer. Promising results have for example been reported for tumor volume measurements [36–38], metabolic tumor parameters derived from FDG-PET/CT [25, 39, 40], heterogeneity features from Radiomics [41], and for the 'apparent diffusion coefficient' (ADC; the main quantitative DWI measure) [42–51] as predictors of response.

Multiparametric imaging

'Multiparametric' imaging refers to a combined approach where the information from different anatomical, functional and/or post-processing techniques is integrated aiming to achieve a more comprehensive understanding of underlying tumor processes than what can be achieved using any single technique on its own. Evidence on the use of multiparametric imaging for treatment response prediction in rectal cancer is sparse. Most available image biomarker studies in rectal cancer concern relatively small scale and single-center studies with a focus on a single imaging technique. Results between individual studies are conflicting and little is known about the complimentary value of combining different techniques. Another important gap is that results of previous single-center works have scarcely been validated in external cohorts or in multicenter study settings. Whether previously obtained results are reproducible and generalizable towards other centers and populations therefore remains largely unknown. Finally, imaging features should be compared and combined with other clinical information

and clinical radiological staging parameters to gain a better insight into the added value of quantitative and multiparametric imaging in a clinical setting.

AIM OF THIS THESIS

The overall aim of this thesis is to study the value of multiparametric imaging to predict response to neoadjuvant treatment in rectal cancer and determine the optimal combination of clinical predictors, functional imaging ‘biomarkers’ and modern post-processing methods.

The main study questions are:

1. What is the value of imaging features derived from anatomical MRI, diffusion-weighted MRI and FDG-PET/CT to predict neoadjuvant treatment response?
2. What is the complementary value of these parameters when combined with other basic clinical patient information and staging outcomes?
3. Can we develop a generalizable image-based prediction model using multicenter imaging data from everyday clinical practice?

OUTLINE OF THIS THESIS

Chapter 2 summarizes the current evidence on the clinical applications and utility of DWI for rectal cancer, including the current status of DWI for the purpose of response prediction.

Chapter 3 combines quantitative imaging features derived from PET/CT, anatomical MRI and diffusion-weighted MRI to investigate their potential complementary value for predicting treatment response to neoadjuvant chemoradiotherapy.

Chapter 4 focuses on local tumor heterogeneity and assesses whether advanced local image texture descriptors are beneficial to study response on PET/CT, MRI and DWI.

Chapter 5 evaluates the effects of variations in acquisition, hardware and inter-reader differences on the reproducibility of quantitative imaging features derived from anatomical and diffusion-weighted MRI in a large multicenter MRI rectal cancer cohort.

In *Chapter 6*, findings from previous chapters were integrated and used to develop and externally validate a multiparametric response prediction model combining both clinical and quantitative imaging variables in a multicenter study setting.

1

REFERENCES

1. Bujko K, Nowacki MP, Nasierowska-Guttmejer A, et al (2006) Long-term results of a randomized trial comparing preoperative short-course radiotherapy with preoperative conventionally fractionated chemoradiation for rectal cancer. *Br J Surg* 93:1215–1223
2. Ngan SY, Burmeister B, Fisher RJ, et al (2012) Randomized Trial of Short-Course Radiotherapy Versus Long-Course Chemoradiation Comparing Rates of Local Recurrence in Patients With T3 Rectal Cancer: Trans-Tasman Radiation Oncology Group Trial 01.04. *J Clin Oncol* 30:3827–3833
3. Latkauskas T, Pauzas H, Kairevice L, et al (2016) Preoperative conventional chemoradiotherapy versus short-course radiotherapy with delayed surgery for rectal cancer: results of a randomized controlled trial. *BMC Cancer* 16:927
4. Ali F, Keshinro A, Weiser MR (2021) Advances in the treatment of locally advanced rectal cancer. *Ann Gastroenterol Surg* 5:32–38
5. Francois Y, Nemoz CJ, Baulieux J, et al (1999) Influence of the Interval Between Preoperative Radiation Therapy and Surgery on Downstaging and on the Rate of Sphincter-Sparing Surgery for Rectal Cancer: The Lyon R90-01 Randomized Trial. *J Clin Oncol* 17:2396–2396
6. Park IJ, You YN, Agarwal A, et al (2012) Neoadjuvant Treatment Response As an Early Response Indicator for Patients With Rectal Cancer. *J Clin Oncol* 30:1770–1776
7. Habr-Gama A, Perez RO, Nadalin W, et al (2004) Operative Versus Nonoperative Treatment for Stage 0 Distal Rectal Cancer Following Chemoradiation Therapy. *Ann Surg* 240:711–718
8. de Neree tot Babberich MPM, van Groningen JT, Dekker E, et al (2018) Laparoscopic conversion in colorectal cancer surgery; is there any improvement over time at a population level? *Surg Endosc* 32:3234–3246
9. Paun BC, Cassie S, MacLean AR, et al (2010) Postoperative Complications Following Surgery for Rectal Cancer. *Ann Surg* 251:807–818
10. van der Valk MJM, Hilling DE, Bastiaannet E, et al (2018) Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. *Lancet* 391:2537–2545

11. Habr-Gama A, Gama-Rodrigues J, São Julião GP, et al (2014) Local Recurrence After Complete Clinical Response and Watch and Wait in Rectal Cancer After Neoadjuvant Chemoradiation: Impact of Salvage Therapy on Local Disease Control. *Int J Radiat Oncol* 88:822–828
12. Habr-Gama A, Perez R, Proscurshim I, et al (2006) Patterns of Failure and Survival for Nonoperative Treatment of Stage c0 Distal Rectal Cancer Following Neoadjuvant Chemoradiation Therapy. *J Gastrointest Surg* 10:1319–1329
13. Maas M, Beets-Tan RGH, Lambregts DMJ, et al (2011) Wait-and-See Policy for Clinical Complete Responders After Chemoradiation for Rectal Cancer. *J Clin Oncol* 29:4633–4640
14. Lambregts DMJ, Maas M, Bakers FCH, et al (2011) Long-term Follow-up Features on Rectal MRI During a Wait-and-See Approach After a Clinical Complete Response in Patients With Rectal Cancer Treated With Chemoradiotherapy. *Dis Colon Rectum* 54:1521–1528
15. Li J, Li L, Yang L, et al (2016) Wait-and-see treatment strategies for rectal cancer patients with clinical complete response after neoadjuvant chemoradiotherapy: a systematic review and meta-analysis. *Oncotarget* 7:44857–44870
16. Dossa F, Chesney TR, Acuna SA, Baxter NN (2017) A watch-and-wait approach for locally advanced rectal cancer after a clinical complete response following neoadjuvant chemoradiation: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2:501–513
17. Habr-Gama A (2006) Assessment and management of the complete clinical response of rectal cancer to chemoradiotherapy. *Color Dis* 8:21–24
18. Habr-Gama A, Perez R, Nadalin W, et al (2005) Long-term results of preoperative chemoradiation for distal rectal cancer correlation between final stage and survival. *J Gastrointest Surg* 9:90–101
19. Hupkens BJP, Martens MH, Stoot JH, et al (2017) Quality of Life in Rectal Cancer Patients After Chemoradiation: Watch-and-Wait Policy Versus Standard Resection – A Matched-Controlled Study. *Dis Colon Rectum* 60:1032–1040
20. Maas M, Nelemans PJ, Valentini V, et al (2010) Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: A pooled analysis of individual patient data. *Lancet Oncol* 11:835–844
21. Bujko K, Richter P, Smith FM, et al (2013) Preoperative radiotherapy and local excision of rectal cancer with immediate radical re-operation for poor responders: A prospective multicentre study. *Radiother Oncol* 106:198–205

22. Garcia-Aguilar J, Shi Q, Thomas CR, et al (2012) A Phase II Trial of Neoadjuvant Chemoradiation and Local Excision for T2N0 Rectal Cancer: Preliminary Results of the ACOSOG Z6041 Trial. *Ann Surg Oncol* 19:384–391
23. Rombouts AJM, Al-Najami I, Abbott NL, et al (2017) Can we Save the rectum by watchful waiting or TransAnal microsurgery following (chemo) Radiotherapy versus Total mesorectal excision for early REctal Cancer (STAR-TREC study)? protocol for a multicentre, randomised feasibility study. *BMJ Open* 7:e019474
24. Bach SP, Gilbert A, Brock K, et al (2021) Radical surgery versus organ preservation via short-course radiotherapy followed by transanal endoscopic microsurgery for early-stage rectal cancer (TREC): a randomised, open-label feasibility study. *Lancet Gastroenterol Hepatol* 6:92–105
25. Joye I, Deroose CM, Vandecaveye V, Haustermans K (2014) The role of diffusion-weighted MRI and 18F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: A systematic review. *Radiother Oncol* 113:158–165
26. Pham TT, Liney GP, Wong K, Barton MB (2017) Functional MRI for quantitative treatment response prediction in locally advanced rectal cancer. *Br J Radiol* 90:20151078
27. Dijkhoff RAP, Beets-Tan RGH, Lambregts DMJ, et al (2017) Value of DCE-MRI for staging and response evaluation in rectal cancer: A systematic review. *Eur J Radiol* 95:155–168
28. Histed SN, Lindenberg ML, Mena E, et al (2012) Review of functional/anatomical imaging in oncology. *Nucl Med Commun* 33:349–361
29. Dhingra VK, Mahajan A, Basu S (2015) Emerging clinical applications of PET based molecular imaging in oncology: the promising future potential for evolving personalized cancer care. *Indian J Radiol Imaging* 25:332–341
30. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5:
31. Lambin P, Rios-Velazquez E, Leijenaar R, et al (2012) Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48:441–446
32. O'Connor JPB, Rose CJ, Waterton JC, et al (2015) Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. *Clin Cancer Res* 21:249–257
33. Ramón y Cajal S, Sesé M, Capdevila C, et al (2020) Clinical implications of intratumor heterogeneity: challenges and opportunities. *J Mol Med* 98:161–177

34. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C (2015) Translational Implications of Tumor Heterogeneity. *Clin Cancer Res* 21:1258–1266
35. Greenbaum A, Martin DR, Bocklage T, et al (2019) Tumor Heterogeneity as a Predictor of Response to Neoadjuvant Chemotherapy in Locally Advanced Rectal Cancer. *Clin Colorectal Cancer* 18:102–109
36. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2011) Rectal Cancer: Assessment of Complete Response to Preoperative Combined Radiation Therapy with Chemotherapy—Conventional MR Volumetry versus Diffusion-weighted MR Imaging. *Radiology* 260:734–743
37. Lambregts DMJ, Rao S-X, Sassen S, et al (2015) MRI and Diffusion-weighted MRI Volumetry for Identification of Complete Tumor Responders After Preoperative Chemoradiotherapy in Patients With Rectal Cancer. *Ann Surg* 262:1034–1039
38. Ha HI, Kim AY, Yu CS, et al (2013) Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after preoperative chemoradiation therapy. *Eur Radiol* 23:3345–3353
39. Ryan JE, Warriar SK, Lynch AC, Heriot AG (2015) Assessing pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. *Color Dis* 17:849–861
40. Maffione AM, Chondrogiannis S, Capirci C, et al (2014) Early prediction of response by 18F-FDG PET/CT during preoperative therapy in locally advanced rectal cancer: A systematic review. *Eur J Surg Oncol* 40:1186–1194
41. Staal FCR, van der Reijdt DJ, Taghavi M, et al (2021) Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review. *Clin Colorectal Cancer* 20:52–71
42. Intven M, Reerink O, Philippens MEP (2013) Diffusion-weighted MRI in locally advanced rectal cancer : pathological response prediction after neo-adjuvant radiochemotherapy. *Strahlenther Onkol* 189:117–22
43. Lambrecht M, Deroose C, Roels S, et al (2010) 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 (Madr)* 49:956–963
44. Jung SH, Heo SH, Kim JW, et al (2012) Predicting response to neoadjuvant chemoradiation therapy in locally advanced rectal cancer: Diffusion-weighted 3 tesla MR imaging. *J Magn Reson Imaging* 35:110–116

45. Sun Y-S, Zhang X-P, Tang L, et al (2010) Locally Advanced Rectal Carcinoma Treated with Preoperative Chemotherapy and Radiation Therapy: Preliminary Analysis of Diffusion-weighted MR Imaging for Early Detection of Tumor Histopathologic Downstaging. *Radiology* 254:170–178
46. Barbaro B, Vitale R, Valentini V, et al (2012) Diffusion-Weighted Magnetic Resonance Imaging in Monitoring Rectal Cancer Response to Neoadjuvant Chemoradiotherapy. *Int J Radiat Oncol* 83:594–599
47. Birlik B, Obuz F, Elibol FD, et al (2015) Diffusion-weighted MRI and MR- volumetry - in the evaluation of tumor response after preoperative chemoradiotherapy in patients with locally advanced rectal cancer. *Magn Reson Imaging* 33:201–212
48. Doi H, Beppu N, Kato T, et al (2015) Diffusion-weighted magnetic resonance imaging for prediction of tumor response to neoadjuvant chemoradiotherapy using irinotecan plus S-1 for rectal cancer. *Mol Clin Oncol* 3:1129–1134
49. Ippolito D, Monguzzi L, Guerra L, et al (2012) Response to neoadjuvant therapy in locally advanced rectal cancer: Assessment with diffusion-weighted MR imaging and 18FDG PET/CT. *Abdom Imaging* 37:1032–1040
50. Jacobs L, Intven M, van Lelyveld N, et al (2016) Diffusion-weighted MRI for Early Prediction of Treatment Response on Preoperative Chemoradiotherapy for Patients With Locally Advanced Rectal Cancer. *Ann Surg* 263:522–528
51. Lambrecht M, Vandecaveye V, De Keyzer F, et al (2012) 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* 82:863–870



© 2007 The McGraw-Hill Companies
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage or retrieval system, without the prior written permission of The McGraw-Hill Companies, Inc.

ISBN 0-07-089671-0
9 780070 896710 >

Diffusion-weighted imaging in rectal cancer: current applications and future perspectives

Niels W. Schurink, Doenja M.J. Lambregts, Regina G.H. Beets-Tan

Br J Radiol. 2019; 92(1096):20180655. doi: 10.1259/bjr.20180655

ABSTRACT

This review summarizes current applications and clinical utility of diffusion-weighted imaging (DWI) for rectal cancer and in addition provides a brief overview of more recent developments (including intravoxel incoherent motion imaging, diffusion kurtosis imaging, and novel postprocessing tools) that are still in more early stages of research.

More than 140 papers have been published in the last decade, during which period the use of DWI have slowly moved from mainly qualitative (visual) image interpretation to increasingly advanced methods of quantitative analysis. So far, the largest body of evidence exists for assessment of tumour response to neoadjuvant treatment. In this setting, particularly the benefit of DWI for visual assessment of residual tumour in post-radiation fibrosis has been established and is now increasingly adopted in clinics. Quantitative DWI analysis (mainly the apparent diffusion coefficient) has potential, both for response prediction as well as for tumour prognostication, but protocols require standardization and results need to be prospectively confirmed on larger scale. The role of DWI for further clinical tumour and nodal staging is less well-defined, although there could be a benefit for DWI to help detect lymph nodes. Novel methods of DWI analysis and post-processing are still being developed and optimized; the clinical potential of these tools remains to be established in the upcoming years.

INTRODUCTION

Over the last decade, more than 140 papers have been published on diffusion-weighted imaging (DWI) for rectal cancer varying from small, purely technical and pre-clinical studies to multicentre clinical patient studies in cohorts of up to 128 patients. Increasing evidence shows that DWI provides added benefit compared to conventional morphological sequences, in particular for the assessment of treatment response. The routine use of DWI for rectal cancer restaging was recently also recommended in the expert consensus guidelines of the European Society of Gastrointestinal Abdominal Radiology [1] and DWI is increasingly incorporated in clinical rectal MRI exams worldwide. **Figure 1** illustrates how the research on rectal DWI has evolved over the years and what have been the main topics under investigation.

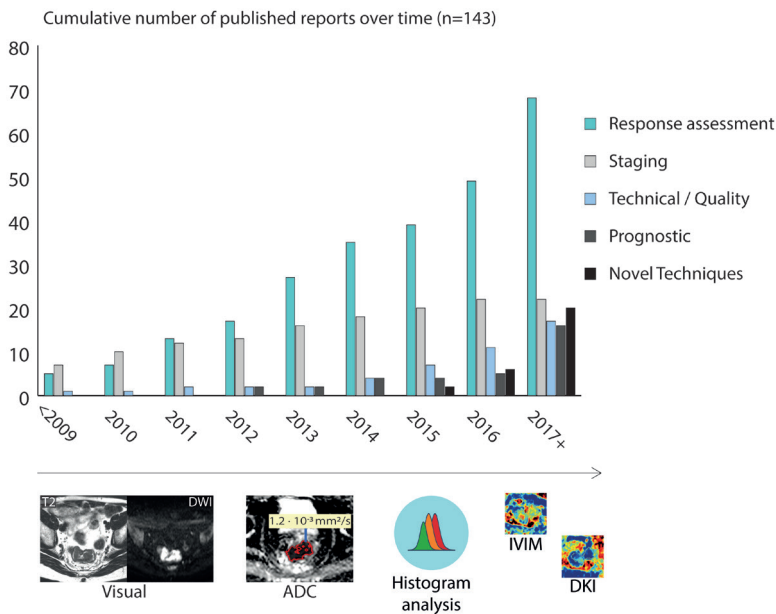


Figure 1. Overview of the cumulative number of studies published on DWI and rectal cancer in the last decade. The majority focused on response assessment to CRT, initially followed by studies on DWI for staging though now overtaken by studies focusing on new techniques. Over time the focus of research has shifted from simple qualitative evaluation to increasingly advanced quantitative methods, which is also reflected by the increased proportion of studies focusing on the development of novel DWI models such as intravoxel incoherent motion (IVIM) and diffusion kurtosis imaging (DKI). Technical/Quality papers indicate papers that focusing on image quality or protocol development.

Initially, the main focus of research was the role of DWI for qualitative (visual) assessment of rectal cancer for either staging or response assessment. This focus has slowly shifted towards more quantitative methods of DWI assessment, including a large number of studies on the use of the apparent diffusion coefficient (ADC), the main quantitative measure of DWI. More recently, several papers have been published on more advanced DWI models and post-processing methods such as histogram analysis, intravoxel incoherent motion (IVIM), diffusion kurtosis imaging (DKI), and automated DWI tumour segmentation.

This paper aims to give an overview of the various clinical applications of DWI and discuss their potential role for rectal cancer imaging.

PRIMARY RECTAL CANCER STAGING

Rectal tumour detection

The main goal of MRI for rectal cancer is staging rather than tumour detection, since typically the presence of tumour has already been established by endoscopy or CT-colonography [2]. It is probably therefore that only a few studies have focused on DWI for the primary detection of rectal cancer [3–10]. Nevertheless, published reports have shown consistently good results for DWI to detect rectal tumours. In a recent meta-analysis, albeit focusing on colorectal tumours in general and not specifically on rectal cancer, pooled sensitivity, specificity and area under the curve (AUC) were 95%, 93% and 0.98, respectively [11]. The few studies that did specifically focus on rectal cancer detection found similar high sensitivities of 86-100%, specificities of 84-100%, and AUCs of 0.96-0.99 [3–6]. These results indicate that overall, the detection level of DWI is high, with a low risk for false positive findings, although these may occur, e.g. because DWI can also result in high signal in non-malignant colorectal polyps [6, 7]. In practice, DWI may mainly be helpful in some specific more difficult cases, to help direct the eye of the radiologist, e.g. in the case of small tumours (see **Figure 2**) or when tumours are obscured by large amounts of faeces. DWI is less useful for the detection of mucinous rectal tumours as due to their high mucin content, these tumours show less restricted diffusion and assessment is limited by T2 effects [12]. They typically show a relatively low signal on high b-value DWI with corresponding high signal on the ADC-map [12]. Mucinous type rectal tumours are generally better appreciated on routine T2W-MRI because their high mucin content results in markedly high T2 signal intensity [13, 14].

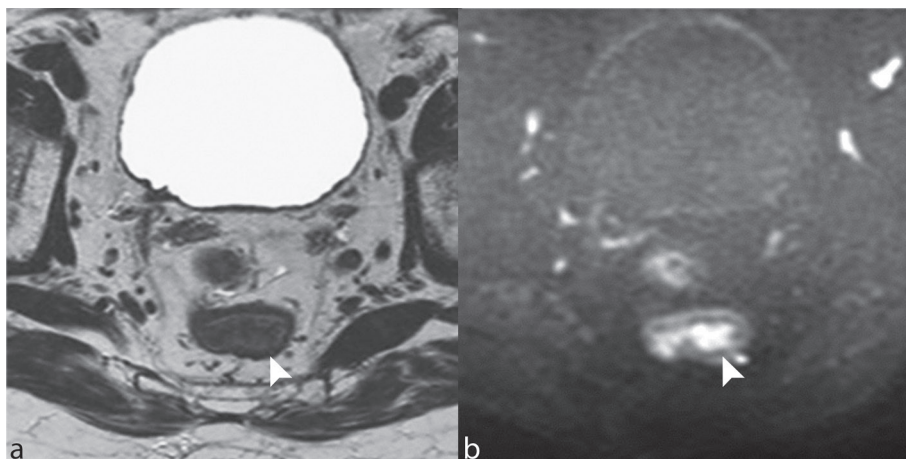


Figure 2. T2-weighted MRI (a) and b1000 s/mm² DWI images (b) of a male patient with a small tumour (white arrowhead) that is hard to detect and initially missed on T2-weighted MRI, but is clearly visible on DWI.

Tumour staging

DWI appears to have only a minor role in the primary staging of rectal tumours. Two groups studied the added benefit of DWI for T-staging of rectal cancer, compared to routine staging using T2W-MRI [15, 16]. They found no clear benefit: for T1-2 tumours sensitivity was 64-90% for DWI vs. 60-80% for T2W-MRI and specificity was 83-100% for DWI vs. 78-92% for T2W-MRI; for T3-4 tumours sensitivity was 50-100% for DWI vs. 50-100% for T2W-MRI and specificity was 83-100% for DWI vs. 77-100% for T2W-MRI. Differences in staging performance all lacked statistical significance. To the best of our knowledge, no papers have specifically focused on using DWI for other primary staging outcomes such as mesorectal fascia (MRF) involvement and extramural venous invasion (EMVI).

Lymph node staging

Nodal staging, remains one of the most challenging tasks for radiologists [17]. Traditionally, nodal staging relied heavily on nodal size as the main criterion. Additional morphological criteria such as nodal border, shape and signal intensity have been shown to be helpful and are now commonly employed, although these criteria may be difficult to evaluate in very small nodes [18–20]. Two meta-analyses reported suboptimal sensitivities and specificities in the range of 55-78% for nodal staging with standard (T2-weighted) MRI [21, 22]. The use of DWI for lymph node staging is appealing, since owing to the high cellular density of lymphoid tissue, nodes should typically be well

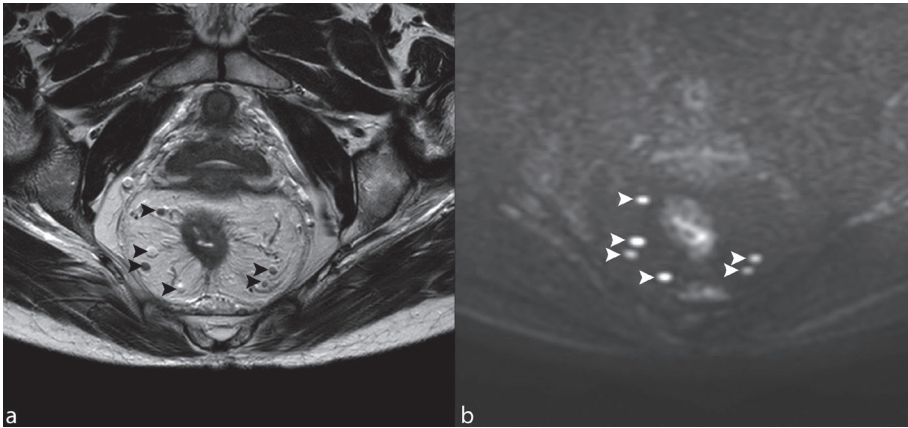


Figure 3. Pre-treatment, primary staging T2-weighted (a) and DWI b1000 s/mm² (b) images of a female patient with a spiculated tumour in the mid-rectum. Note how the various mesorectal lymph nodes (arrowheads) are very easily detectable on DWI.

detectable on DWI (see **Figure 3**). Indeed, a 10-83% increase in the overall number of detected lymph nodes has been reported when using DWI compared to T2W-MRI to detect pelvic lymph nodes [23–26]. The value of visual DWI evaluation for nodal characterization is less apparent. Two studies each reported a positive-predictive value of only 52% when using a high signal on DWI as a criterion for malignancy, indicating that use of DWI entails a risk for overstaging [20, 23]. One study looked at the morphology of lymph nodes on DWI and found that a more heterogeneous signal on DWI was associated with malignancy [27], an observation that has also previously been reported for nodes on T2W-MRI [18, 19] and may be less attributable to the use of DWI itself. This single-centre result has so far not been validated by other groups.

The majority of studies on DWI for characterizing rectal nodes focused on quantitatively measuring nodal ADC values [23, 28–31]. An overview of these studies is presented in **Table 1**. Most studies reported significantly higher ADC values (indicating a lower cellular density) for benign nodes than for malignant nodes [28–31]. Sensitivities and specificities to characterize nodes based on the ADC (using retrospectively determined threshold values) ranged from 67-88% and 60-97% respectively, which is only slightly higher than previously reported values [21, 22, 24, 32]. Moreover, reported ADC values vary across studies (using different MR vendors and protocols) and show considerable overlap between malignant and benign nodes. Also, feasibility and reproducibility of nodal ADC measurements has been reported as a potential drawback, owing to the

Table 1. Overview of studies that compared the mean ADC values of benign and malignant nodes in rectal cancer, both in the primary staging setting, as well as for restaging of nodes after CRT.

Author (year)	(ref)	N (pt)	N (nodes)	ADC benign nodes	ADC malignant nodes	p=	Cut-off	AUC	Sens	Spec
Primary staging										
Yasui (2009)	[28]	46	163	1.85 ± 0.53	1.36 ± 0.42	0.001	1.44	0.79	75	74
Cho (2013)	[29]	34	114	1.10 ± 0.22	0.90 ± 0.15	< 0.0001	1.00	0.73	78	67
Zhao (2014)	[30]	72	454	0.91 ± 0.19	0.77 ± 0.12	<0.01#	-	-	88	97
Cerny (2016)	[31]	24	44	1.38 ± 0.32	1.10 ± 0.19	0.0012	-	0.76	-	-
Heijnen (2013)	[23]	21	102	1.15 ± 0.24	1.04 ± 0.22	0.1	1.07	0.64	67	60
Restaging after CRT										
Kim (2015)	[32]	53	115	1.13 ± 0.23	1.36 ± 0.27	<0.0001	1.25	0.74	66	74
Lambregts (2011)	[24]	30	115	1.19 ± 0.27	1.43 ± 0.38	<0.001	1.25	0.66	53	82

NB. Statistically significant (P<0.05) results are printed in bold.

For Zhao (2014), results are presented on a per patient (N+ v.s. N0 stage) basis, all other concern node-by-node analyses.

2

typical small size of rectal nodes combined with the suboptimal resolution on ADC-maps, which can make it hard to delineate the nodes to measure their ADC [24, 33].

Two studies specifically reported that ADC could not be measured in a subset (21-27%) of the nodes identified on DWI because they were either too small or due to local image distortions [23, 24]. Along the same line, two groups omitted measurements on lymph nodes smaller than 2 mm in diameter, as delineations were technically too challenging in these nodes [29, 32].

TUMOUR RESPONSE ASSESSMENT

As illustrated in **Figure 1**, the major focus of research in rectal DWI has been the assessment of response to neoadjuvant treatment. This specific focus follows a recent paradigm shift in the treatment of rectal cancer, based on current evidence that in patients that show a (near)-complete response to chemoradiotherapy (CRT) organ-preserving treatments such as 'watch-and-wait' may be considered as safe alternatives

to major surgery [34]. This shift in treatment management increases the demand for an accurate radiological response evaluation.

Different methods of response evaluation have been studied varying from visual DWI assessment to quantitative volumetric or ADC measurements, some with the addition of post-processing steps such as histogram analysis. An imaging example illustrating these various methods is provided in **Figure 4**.

Visual (qualitative) response assessment

Routine MRI has well-known difficulties to discern viable tumour within post-treatment fibrosis, which is reflected by its poor performance (sensitivity of only 19% in a recent meta-analysis [35]) to differentiate between patients with a complete response (i.e. sterilized fibrosis) and patients with residual tumour. Increasing evidence suggests that

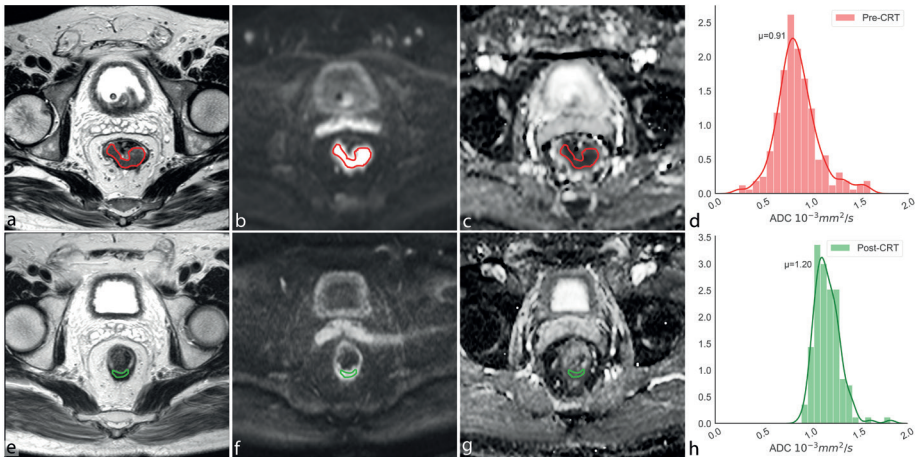


Figure 4. Pre- (upper row) and post-CRT (bottom row) T2-weighted (a,e), b1000 s/mm² DWI (b,f) and ADC images (c,g) of a patient with a midrectal tumour that responded well to CRT (Histopathology after surgery indicated a very good response with predominant fibrosis and only rare residual tumour cells; Mandard tumour regression grade of 2). The images illustrate the different ways DWI can be used to assess response: on pre-CRT a clear high signal mass can be appreciated on DWI (b), after CRT only a small high signal remnant is visible within the fibrosis on DWI indicating a small residual tumour (f). The tumour volume on DWI decreased from 13.2 cm³ to 0.26 cm³, while the ADC value increased from 0.91·10⁻³ mm²/s, to 1.20·10⁻³ mm²/s. Concordantly the histograms show that the distribution of ADC values within the tumour has shifted towards more high ADC values, indicating a good response.

DWI is better equipped to make this differentiation (overview of studies provided in **Table 2**). Fibrotic tissue consists of loose connective fibres resulting in relatively minor diffusion restriction. Moreover, the high collagen content in fibrosis typically has a very short T2 relaxation time, leading to low signal on the ADC-map as well as the DW images [43]. In contrast, areas containing persistent viable tumour typically show restricted diffusion resulting in a high signal on high b-value DWI. Reported AUCs for DWI range between 0.70 and 0.96, compared to 0.67-0.85 for standard MRI, with a statistically significant effect in the majority of reports [36–42]. One recent study combined morphologic patterns of response on T2W-MRI with distinct signal patterns on DWI (specific locations of focal diffusion restriction within different patterns of fibrosis).

With this approach, the authors reached a sensitivity of 94% and specificity of 77% to differentiate patients with residual tumour from patients with a complete response [44], results that remain to be validated by other groups.

Table 2. Overview of studies that compared the performance of DWI and T2W imaging to visually assess complete response to CRT

Author (year)	(ref)	N (pt)	T2W					DWI				
			AUC	Sens	Spec	PPV	NPV	AUC	Sens	Spec	PPV	NPV
Kim (2009)*	[36]	40	0.67	50	78	46	80	0.85	87	83	66	94
Lambrechts (2011)*[37]		120	0.67	11	94	18	82	0.79	57	93	71	89
Park (2011)*	[38]	45	0.81	0.44	84	88	38	0.94	96	79	96	87
Song (2012)*	[39]	50	-	71	67	94	25	-	88	42	92	55
Sassen (2013)*	[40]	70	0.76	25	94	64	88	0.80	55	96	72	93
Marouf (2015)	[41]	19	-	60	33	-	-	-	79	80	-	-
Foti (2016)	[42]	31	-	20	100	100	88	-	80	100	100	97

NB. Presented results are the average of two readers except for [37], who used 3 readers, and [41] who did not mention the number of readers.

*In studies [36–40] a statistically significant improvement in results was observed for at least one of the readers after adding DWI images compared to T2W alone for assessing complete response to CRT.

Tumour restaging

Most published reports on DWI for tumour response assessment focused on the general differentiation between a favourable response (i.e. complete or good response) vs poor response. In a meta-analysis by van der Paardt et al. pooled sensitivity for predicting response (defined as either ypT0, ypT0-2 or T-downstaging compared to primary staging) was significantly higher for studies that included DWI in the MR protocol compared to studies that did not (83.6% vs 50.4%) [35]. A second meta-analysis by Wu et al. included studies focusing both on visual DWI analysis as well as studies focusing on ADC (with no subanalyses between these two groups) making it difficult to draw conclusion about the one or the other [45]. To the best of our knowledge, so far no studies exist on visual use of DWI for further yT substaging. One study recently proposed a new 3-point MR based tumour regression grade (mrTRG) incorporating both T2W-MRI and DWI for evaluating response after CRT. Both the accuracy for assessing response and interreader agreement improved significantly compared to the more well-known 5-point mrTRG score which uses T2W-MRI only [46]. A single report by Park and colleagues evaluated the use of DWI in addition to T2W-MRI to predict tumour clearance of the MRF after neoadjuvant CRT [38]. The authors reported a significantly improved performance after the addition of DWI (AUC 0.92-0.96) compared to use of only T2-weighted MRI (AUC 0.77-0.85).

Nodal restaging

A small number of papers reported on the diagnostic value of visual lymph node assessment using DWI in the restaging setting. Lambregts et al performed a node-by-node analysis of 157 nodes detected on DWI post-CRT and found that nodal signal intensity on DWI did not differ between yN- and yN+ nodes (AUC 0.52 and 0.64 for two readers) [24]. Two groups assessed the use of DWI to predict lymph node eradication (i.e. yN0 stage) after CRT on a patient basis [47, 48]. Van Heeswijk et al reported that the visual absence of nodes on DWI after CRT was a highly reliable predictor of a negative nodal status (sensitivity 100%), but the presence of nodes on post-CRT DWI was an unspecific finding, which could indicate either the presence of benign or malignant nodes, resulting in a low specificity of only 14%, again illustrating the limited capacity of DWI to visually characterize lymph nodes [47]. Ryu et al used a confidence level score to predict lymph node eradication after CRT with and without DWI and found no improvement in diagnostic performance for DWI compared to T2W-MRI with AUCs in the same range of 0.77-0.80 [48].

Table 3. Overview of studies that have compared DWI and T2W tumour volumetry to predict a complete response to CRT.

Author (year)	(ref)	N=	T2W AUC	DWI AUC
Pre-CRT volume				
Curvo-Semedo (2011)	[49]	50	0.57	0.63
Lambregts (2014)	[50]	112	0.73	0.77
Post-CRT volume				
Curvo-Semedo (2011)	[49]	50	0.70	0.93*
Lambregts (2014)	[50]	112	0.82	0.92*
Sathyakumar (2016)	[51]	64	-	0.88
ΔVolume				
Curvo-Semedo (2011)	[49]	50	0.84	0.92
Lambregts (2014)	[50]	112	0.78	0.86
Sathyakumar (2016)	[51]	64	-	0.84
Ha (2013)	[52]	100	0.79	0.91*

* = DWI volumetry performed significantly ($P < 0.05$) better than T2W volumetry for predicting complete response to CRT.

NB. For all studies, manual tumour segmentations were performed slice-by-slice on T2W and DWI by experienced readers.

Quantitative response assessment

DWI tumour volumetry

Table 3 summarizes the findings of four studies that assessed the value of measuring tumour volumes on high b-value DW images to diagnose a complete response. Similar good results were found for the DWI tumour volume after CRT and the relative change in DWI tumour volume after CRT (Δ volume) with AUCs of 0.84-0.93 [49–52]. In the three studies that compared DWI to T2W-MRI, DWI volumetry significantly outperformed T2W-volumetry [49–51]. Pre-treatment volumes showed only moderate performance with AUCs of 0.57-0.77 for both T2W-MRI and DWI, indicating that volumetry is of limited value for pre-treatment response prediction.

Tumour ADC

ADC has been extensively studied as an imaging biomarker to assess and predict response. A summary of these studies, in particular the studies focusing on pre-CRT ADC, post-CRT ADC and Δ ADC, is presented in **Table 4**.

Table 4. Overview of studies evaluating mean tumour ADC to predict response to chemoradiotherapy.

Author (year)	(ref)	N=	Standard of reference	Pre-CRT ADC			Post-CRT ADC			ΔADC% (post versus pre-CRT)		
				good	poor	P	good	poor	P	good	poor	P
Outcome 1 – Prediction of good versus poor response												
Jung (2012)	[53]	35	cT > ypT	0.93	1.03	0.034	1.29	1.18	0.009	+0.36#	+0.14#	<0.0005
Birik (2015)	[54]	43	cT > ypT	0.63	0.73	<0.05	1.26	0.93	0.001	+103%	+30%	<0.001
Sun (2010)	[55]	37	cT > ypT	1.07	1.19	0.013	1.30	1.28	0.560	+23%	+10%	<0.001
Elmi (2013)	[56]	49	cT > ypT	0.97	0.84	0.035	-	-	-	-	-	-
Hu (2015)	[57]	56	cT > ypT	0.85	0.85	0.944	1.25	1.10	0.001	0.52*	0.32*	0.015
Iannicelli (2016)	[58]	34	cT > ypT	0.92	0.90	0.268	1.34	1.15	0.010	0.41	0.30	0.168
Kim (2011)	[59]	34	cT > ypT	0.87	0.91	0.610	-	-	-	+21%	+18%	0.430
Nougaret (2016)	[60]	31	TRG3-4 vs TRG0-2	1.10	0.90	0.460	1.40	1.10	0.002	+40%	+6%	0.004
Lu (2017)	[61]	42	TRG3-4 vs TRG0-2	1.21	1.25	0.503	1.93	1.82	0.282	+58%	+37%	0.181
Quaia (2016)	[62]	45	RCRG1-2 vs RCRG3-4	0.94	0.91	0.830	1.42	1.23	0.160	+51%	+35%	0.250
Hu (2015)	[57]	56	TRG0-1 vs TRG2-3	0.83	0.86	0.524	1.27	1.10	<0.001	0.55*	0.32*	-0.006
Jacobs (2016)	[63]	22	TRG1-2 vs TRG3-5	0.94	1.11	0.040	1.45	1.35	0.010	+46%	+16%	<0.001
Barbaro (2012)	[64]	62	TRG1-2 vs TRG3-5	1.50	1.20	0.007	-	-	-	>+23%	<+23%	0.011
Intven (2013)	[65]	59	TRG1-2 vs TRG3-5	0.95	1.12	0.001	1.44	1.36	NS	+50%	+23%	<0.001
Cai (2013)	[66]	15	TRG1-2 vs TRG3-5	0.66	0.89	0.021	-	-	-	-	-	-
Blazic (2015)	[67]	58	TRG1-2 vs TRG3-5	0.88	0.87	0.409	1.36	1.12	<0.001	+55%	+30%	<0.001
Ippolito (2015)	[68]	31	TRG1-2 vs TRG3-5	0.88	0.78	0.076	1.47	1.19	0.009	+73%	+56%	0.008
Ippolito (2012)	[69]	30	TRG1-2 vs TRG3-5	0.88	0.78	0.331	1.48	1.19	0.007	+71%	+52%	0.113
Iannicelli (2016)	[58]	34	TRG1-2 vs TRG3-5	0.94	0.87	0.151	1.43	1.16	0.001	+0.49#	+0.29#	0.01
Monguzzi (2013)	[70]	31	TRG1-2 vs TRG3-5	0.83	0.82	0.273	1.43	1.25	0.004	+63%	+60%	0.124
Intven (2015)	[71]	55	TRG1-2 vs TRG3-5	-	-	-	-	-	-	+48%	+26%	<0.001
Kim (2011)	[59]	34	TRG1-2 vs TRG3-5	0.89	0.91	0.530	-	-	-	+18%	+20%	0.460
Bakke (2017)	[72]	27	TRG1-2 vs TRG3-5	0.74	0.61	>0.1	0.65	0.69	>0.1	-16%	+15%	<0.01
Foti (2016)	[42]	31	ypCR + ypPR vs ypSD	0.83	0.91	<0.05	1.19	1.01	<0.05	+0.36#	+0.11#	<0.020
Krenser (2003)	[73]	8	ypT0-2 vs ypT3	0.80	0.70	<0.02	-	-	-	-	-	-
Hein (2003)	[74]	16	ypT0-2 vs ypT3	0.48	0.70	0.012	-	-	-	-	-	-
Lambrecht (2012)	[75]	20	ypT0-2 vs ypT3	1.06	1.19	0.270	-	-	-	+46%	+17%	0.080
deVries (2003)	[76]	34	ypT0-2 vs ypT3	0.65	0.66	0.800	-	-	-	-	-	-
Hein (2003)	[77]	9	ypT0-2 vs ypT3	-	-	NS	-	-	-	-	-	-
Lambrecht (2010)	[78]	22	ypT0-2 vs ypT3	-	-	-	-	-	-	+55%	+32%	0.320

DIFFUSION-WEIGHTED IMAGING IN RECTAL CANCER: CURRENT APPLICATIONS AND FUTURE PERSPECTIVES

Outcome 2– Prediction of complete versus incomplete response

				complete	incomplete	P	complete	incomplete	P	complete	incomplete	P
Chen (2016)	[79]	100	pCR vs non-pCR	0.86	0.90	<0.001	1.44	1.33	<0.001	+68%	+48%	<0.001
Intven (2013)	[65]	59	"	0.97	1.09	0.010	1.46	1.35	0.047	+50%	+25%	<0.001
Lambrech (2012)	[75]	20	"	0.94	1.19	0.003	-	-	-	+88%	+26%	0.0011
Lambrech (2010)	[78]	22	"	0.94	1.20	0.002	-	-	-	-	-	-
Blazic (2016)	[80]	62	"	0.85	0.88	0.157	1.36	1.16	<0.001	+61%	+33%	<0.001
Genovesi (2013)	[81]	28	"	1.01	1.29	0.330	1.79	1.37	0.003	+77%	+36%	0.05
Hu (2015)	[57]	56	"	0.82	0.86	0.332	1.31	1.12	<0.001	0.64*	0.33*	<0.001
Kim (2011)	[82]	76	"	0.85	0.88	0.410	1.43	1.14	<0.0001	+70%	+30%	<0.0001
Intvent (2015)	[71]	55	"	-	-	-	-	-	-	+48%	+26%	0.012
Lu (2017)	[61]	42	"	1.20	1.25	0.406	1.94	1.83	0.420	+63%	+48%	0.042
Bassaneze (2017)	[83]	33	"	-	-	-	1.53	1.16	<0.010	-	-	-
Choi (2016)	[84]	86	"	-	-	-	1.60	1.42	<0.004	-	-	-
Ha (2013)	[52]	100	"	0.59	0.49	0.484	1.33	1.13	0.001	-	-	-
Kim (2009)	[36]	40	"	-	-	-	1.62	1.04	<0.0001	-	-	-
Cho (2015)	[85]	50	"	-	-	-	1.60	1.41	0.019	-	-	-
Cai (2014)	[86]	80	"	-	-	-	1.65	1.52	0.024	-	-	-
Song (2012)	[39]	50	"	-	-	-	1.55	0.93	<0.0001	-	-	-
Barbaro (2012)	[64]	57	"	-	-	>0.05	-	-	-	-	-	-
De Cecco (2016)	[87]	12	"	0.93	0.85	0.818	-	-	-	-	-	-
De Felice (2017)	[88]	37	"	0.81	1.05	>0.05	1.18	1.50	0.050	-	-	-
Curvo-Semedo (2011)[49]	50		"	1.07	1.10	0.610	1.39	1.45	0.480	+35%	+36%	0.960
Engin (2012)	[89]	30	"	0.88	0.83	0.066	1.29	1.11	0.071	-	-	-
Foti (2016)	[42]	31	"	0.78	0.87	-	1.28	1.10	-	+0.51#	+0.23#	-
Sathyakumar (2016)[51]	64		"	0.98	1.01	NS	1.46	1.41	>0.05	+54%	+45%	NS
Cai (2014)	[86]	65	"	-	-	-	1.56	1.44	0.152	-	-	-

= absolute difference in ADC between pre- and post-treatment scan; * = reported as ADCratio = (ADCpost – ADCpre)/ADCpre; NS = not significant; " = same as above
 Statistically significant results (P<0.05) are printed in bold
 Note: cT = clinical T-stage, ypT = histopathological T-stage, TRG = tumour regression grade, RCRG = rectal cancer regression grade, ypCR/pCR= pathological complete response, ypPR = pathological partial response, ypSD = pathological stable disease
 Studies [60, 61] used Dworak's TRG; [62] used the Wheeler's RCRG. [57] used Ryan's TRG. [58, 59, 72, 63–72] used Mandard's TRG



Regardless of the definition of response used in these reports (i.e. good or complete response), all studies reported an increase in mean tumour ADC after CRT [36, 39, 42, 49, 51–58, 60–90], which is thought to be due to radiation-induced cellular damage and necrosis [91, 92]. The disruption of cell membranes reduces the diffusion restriction and therefore increases the ADC. Both the final post-CRT ADC and the relative increase in ADC ($\Delta\text{ADC}\%$) were typically higher in the favourable response groups, with statistically significant results in the majority of studies [36, 42, 52–55, 57, 58, 60–65, 67–72, 79–81, 83, 84]. In addition, several studies found significantly higher pre-CRT ADC values in the unfavourable response groups [42, 53–56, 63–66, 73–75, 78, 79], although a similar number of studies did not find a significant difference in pre-CRT ADC between response groups [51, 52, 58, 59, 61, 62, 64, 67–70, 72, 75–77, 80–82, 84, 87–89]. A high ADC is believed to be associated with tissue necrosis, which in turn leads to decreased tissue perfusion and hypoxia, making tumours less susceptible to CRT effects [93, 94].

In addition, six groups investigated the prognostic value of measuring changes in ADC early during CRT. The groups of Jacobs et al and Cai et al reported significant differences between good and poor responders in mean tumour ADC in week 3 and weeks 3-5, respectively [63, 66]. Published results for the first 2 weeks of CRT have so far been inconsistent: some authors already found significant differences in ADC at these very early timepoints [55, 64], while others could not reproduce this [59, 66, 75]. Altogether, the majority of studies that investigated ADC as a biomarker to assess or predict response to treatment found significant results at one or more time points, although a subgroup (21%) of studies could not produce any statistically significant results [49, 51, 76, 77, 86–90]. As a critical note, most study cohorts presented so far are small and single centre, and reported ADC and cut-off values show large variation and overlap between studies and have never been validated in prospective study cohorts. This stresses the need for standardization and multicentre validation studies. With this in mind, a meta-analysis by Joye et al concluded that based on current evidence the results look promising but need work with pooled sensitivities and specificities to predict complete response of 69 and 68% for pre-CRT ADC, 78 and 72% for post-CRT ADC, and 80% and 78% for $\Delta\text{ADC}\%$ [95]. Because of the limited number of studies and the small study sizes, no conclusions can be drawn yet with respect to the added benefit of performing ADC measurements (early) during CRT treatment.

Histogram analysis

While most studies investigated only mean tumour ADC values, some evaluated the added benefit of performing histogram analysis. With histogram analysis, the whole spectrum of ADC values within the tumour is analyzed, allowing extraction of not only mean (or median) values but also additional parameters such as the minimum and maximum, standard deviation and different percentile ranges. Based on the limited evidence available so far, these parameters do not seem to offer a clear additional benefit. Of the papers that have reported an association between ADC histogram metrics (in particular 10th-25th percentile ranges) and response [60, 84, 85], the majority also reported that histogram parameters did not significantly outperform median or mean ADC values [60, 85]. Two other reports by van Heeswijk et al and Chidambaram et al failed to produce any significant correlation between histogram ADC measurements and the final treatment, although these reports did not find any significant results for mean ADC values either [96, 97].

Lymph node ADC

Two of the studies included in **Table 1** compared the mean ADC values of benign and malignant nodes in the restaging setting [24, 32]. Both reported a significantly higher ADC for malignant nodes, and interestingly an identical optimal cut-off value of $1.25 \cdot 10^{-3}$ mm²/s to differentiate between benign and malignant nodes. However, both groups also reported that adding ADC measurements to size-based assessment on routine T2W-MRI yielded no statistically significant diagnostic gain, suggesting that from a clinical point of view the benefit of measuring nodal ADCs may be limited.

DWI FOR FOLLOW-UP AFTER TREATMENT

There is limited evidence that DWI may help diagnose locally recurrent disease during follow-up after primary treatment. Two groups assessed the value of DWI to detect pelvic recurrences post-surgery. High AUCs of 0.87-0.99 were found for MRI + DWI, though results were not significantly different compared to using only standard MRI [98, 99], except for less experienced, resident readers in one of the two reports [99]. In addition, it was reported that DWI may aid in very specific cases with multiple local recurrent sites, or for the detection of small and/or anastomotic tumours. Two other reports compared the use of MRI with and without DWI for follow-up of rectal cancer

patients treated with (local excision or wait-and-see) [100, 101]. Although both found no overall improvement in diagnostic performance to detect local tumour regrowths in terms of AUC, adding DWI did offer some potential benefits. In one report, adding DWI improved the sensitivity of MRI and lowered the rate of inconclusive MRI outcomes [100]. Both studies also suggested that DWI may aid in detecting recurrences earlier during follow-up [100, 101].

DWI AS A PROGNOSTIC MARKER

Over the last few years there has been a growing interest for the use of quantitative DWI parameters as prognostic imaging biomarkers to predict various outcomes ranging from clinical TNM-stage, to histopathological or immunohistochemical markers, and measures of long-term outcome such as disease-free survival. Although a comprehensive discussion of the results of these studies with this wide range of outcomes is beyond the scope of this paper, a brief overview is provided in **Table 5** and discussed below.

The majority of reports published so far focused on the correlation of DWI with relatively simple clinical prognostic markers (such as TN-stage) and histopathological markers such as the tumour differentiation grade, with the aim to differentiate tumours with a more or less favourable overall prognostic profile. Around half of these reports found significant correlations between DWI-derived parameters and clinical or histopathology outcomes [96, 102–111]. Of those studies that found significant results, the majority reported low ADC values for the unfavourable outcome groups (e.g. higher TN-stage, lower differentiation grade, MRF+ stage and extranodal tumour deposits), and high ADC values for the more favourable outcome groups, suggesting that tumours with a more dense cellular structure (low ADC) tend to show a more aggressive growth pattern. As discussed in the previous section (on “Quantitative response assessment”), these low ADC tumours have also been associated with a more favourable outcome in terms of response to treatment by some groups [42, 53–56, 63–66, 73–75, 78, 79]. This might suggest that the same factors that give rise to a generally more aggressive tumour profile may also render tumours more susceptible to anticancer treatment. However, given the ambiguous results published so far (with approximately 50% of studies lacking statistically significant findings), this hypothesis remains to be further tested before any definite conclusions can be drawn.

DIFFUSION-WEIGHTED IMAGING IN RECTAL CANCER: CURRENT APPLICATIONS AND FUTURE PERSPECTIVES

Table 5. Overview of studies that investigated the relationship between DWI and prognostic outcomes.

	Total number of studies [total N° of patients]		N° studies with positive outcome* (refs)		N° studies with negative outcome* (refs)	
Clinical outcomes						
AJCC-stage	1	[n= 52]	-		1	[102]
T-stage#	12	[n= 650]	5	[102–106]	7	[96, 105, 107–111]
N-stage#	11	[n= 609]	5	[102, 103, 106, 109, 111]	6	[96, 104, 105, 107, 108, 110]
Mesorectal fascia involvement	6	[n= 307]	2	[103, 109]	4	[96, 102, 104, 108]
Extramural Venous Invasion	1	[n= 52]	1	[102]	4	-
M-stage	3	[n= 124]	1	[105]	2	[96, 107]
Histopathological outcomes						
Differentiation grade	8	[n= 421]	5	[102, 108–111]	3	[96, 104, 107]
Extranodal tumour deposits	1	[n= 49]	1	[104]	-	
Lymphovascular invasion	5	[n= 264]	2	[103, 107]	3	[104, 108, 109]
Neural invasion	2	[n= 95]	-		2	[104, 107]
Laboratory and immunohistochemical outcomes						
P21	1	[n= 49]	-		1	[104]
P53	1	[n= 49]	-		1	[104]
Her2/neu	1	[n= 49]	-		1	[104]
CD44	1	[n= 49]	-		1	[104]
Ki-67	4	[n= 314]	4	[104, 110, 112, 113]	-	
AgNOR	1	[n= 49]	1	[104]	-	
Hif-1 α	1	[n= 91]	1	[110]	-	
VEGF	1	[n= 91]	1	[110]	-	
Cell count	1	[n= 17]	-		1	[113]
Total nucleic area	1	[n= 17]	-		1	[113]
Average nucleic area	1	[n= 17]	-		1	[113]
Microvessel density	1	[n= 17]	1	[113]	-	
KRAS status	1	[n= 51]	1	[114]	-	
CEA	4	[n= 252]	1	[110]	3	[102, 103, 109]
CA19-9	2	[n= 101]	1	[104]	1	[102]
Long term outcomes						
Disease free survival	1	[n= 61]	1	[115]	-	
3 year local recurrence rate	1	[n= 128]	1	[116]	-	
3 year distant relapse-free survival	1	[n= 128]	1	[116]	-	
Local or distant recurrence	2	[n= 101]	1	[115]	1	[108]

* Positive outcome indicates that ≥ 1 of the DWI parameters under investigation (e.g. mean ADC, ADC histogram parameters or parameters derived from, IVIM, DKI, or DWI texture) had a significant correlation with the studied outcome.

All authors used pathological T- and N-stage as the outcome except for ref [109], that used mrT- and mrN-stage.

Note: All presented studies included mean ADC as an input variable. References [96, 102, 103, 105–107, 111, 113, 114] additionally included more advanced parameters related to DKI [103, 105, 111], IVIM [102, 107, 113, 114], texture [106] or histogram parameters [96, 103, 106]

Some studies looked at more advanced DWI parameters derived from IVIM imaging (discussed in more detail in section on “Recent advances” below) [102, 107]. Higher IVIM perfusion related parameters were associated with poorer TN-stage, differentiation grade, lymphovascular invasion and extramural venous invasion.

In an attempt to better understand the relation between DWI-parameters and underlying tumour biology, several investigators have studied the relationship between DWI-parameters and immunochemical marker expressions related to cell proliferation/apoptosis (p21, p53, Ki-67, AgNOR), vascularization (VEGF), cell adhesion (CD44, CEA) and hypoxia (Hif1- α). So far, evidence mainly comes from single centre studies. Approximately half of these studies found significant correlations between tumour ADC and the studied marker, which mainly consisted of either proliferation related biomarkers (Ki-67, AgNOR) or biomarkers related to perfusion (VEGF, microvessel density) and hypoxia (Hif1- α) [104, 110, 112–114].

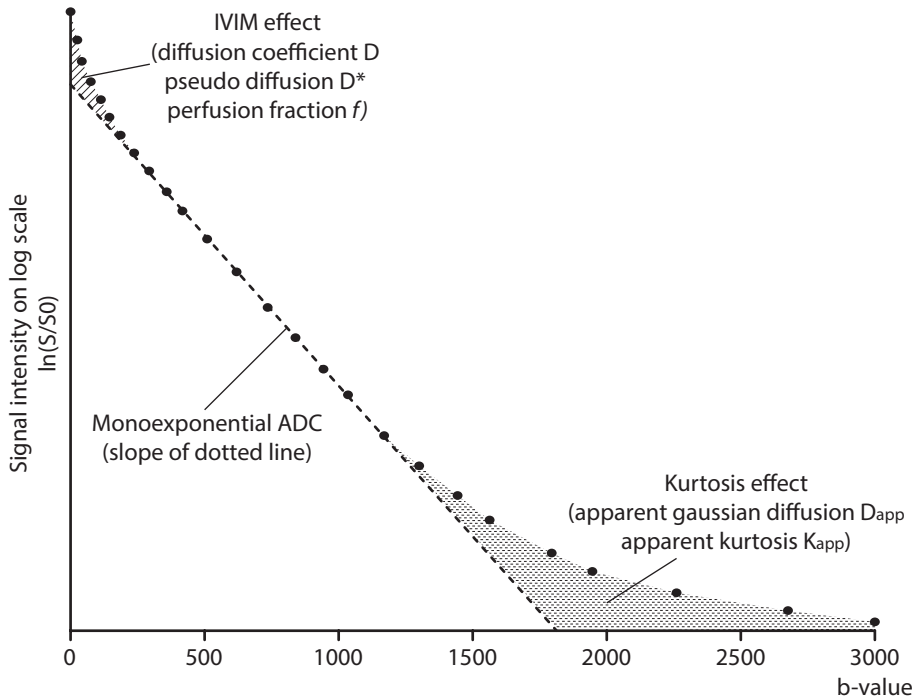
Finally, a small subset of studies focused on the correlation between DWI-derived parameters and long-term outcome, suggesting that lower ADC values are associated with shorter disease-free survival and higher recurrence rates [115, 116]. Future research is needed to confirm these findings.

RECENT ADVANCES

Methods of DWI analysis have rapidly evolved over the years. Whereas traditionally DWI analysis was limited to visual image interpretation or “simple” ADC measurements, more complex methods of DWI acquisition, multiexponential diffusion quantification models and novel DWI post-processing tools have recently been introduced. To provide a comprehensive overview of these new methods is beyond the scope of the current paper, but a brief overview is provided below and illustrated in **Figure 5**.

Intravoxel incoherent motion (IVIM)

Traditionally, ADC is calculated using a monoexponential fit of ≥ 2 b-value DWI images (typical in the range of $b = 0$ to $b = 1000$). Instead of using a monoexponential fit, the IVIM model uses a biexponential fit that separates effects of (micro-)perfusion (measurable in the low b-value range) and true diffusion effects (measurable at higher b-values). This effect is schematically illustrated in **Figure 5** [117]. The IVIM model separates the true diffusion coefficient D from two perfusion related parameters



2

Figure 5. Traditional DWI models use a mono-exponential fit of two or more b-value images between $b=0$ and $b=1000$ to calculate the ADC value as the slope of a straight line between these points. At low b-values ($b < 200$) the signal decay will, however, deviate from this line as it is not only affected by tissue diffusion, but also by microperfusion effects (the IVIM or 'intravoxel incoherent motion' effect). Another phenomena is the deviation of the signal curve when applying very high b-values ($b > 1000-1500$). This effect is caused by non-Gaussian diffusion as a result of complex structures (such as cell membranes, organelles etc) that hinder diffusion. The degree of non-Gaussian behaviour is referred to as the kurtosis effect. Formula's to calculate the various parameters described in the Figure are as follows: Monoexponential ADC: $S/S_0 = \exp(-b \cdot \text{ADC})$; IVIM: $S/S_0 = f \cdot \exp(-b(D+D^*)) + (1-f) \exp(-b \cdot D)$; Kurtosis: $S/S_0 = \exp(-b \cdot D_{\text{app}} + b^2 \cdot D_{\text{app}}^2 \cdot K_{\text{app}}^2 / 6)$; where S = signal intensity with (S) and without (S_0) diffusion-weighting; b = b-value (s/mm^2) used; ADC = apparent diffusion coefficient (mm^2/s ; observed diffusion); D = diffusion coefficient (mm^2/s ; true diffusion in the tissue; depends on cell density); D^* = pseudo diffusion coefficient (mm^2/s ; depends on mean capillary segment length and average blood velocity in a voxel); f = the perfusion fraction (indicates the fractional volume (%) of capillary blood flowing within a voxel); D_{app} = apparent gaussian diffusion coefficient (mm^2/s ; diffusion coefficient under a Gaussian assumption); K_{app} = apparent kurtosis (describes how much the measured diffusion departs from the assumed gaussian distribution; a measure for heterogeneity).

called the pseudodiffusion D^* and perfusion fraction f [118]. The potential benefit of IVIM is that it can provide parameters related to tissue microcirculation and perfusion in addition to cellularity, without the need for exogenous contrast agents (such as required for dynamic contrast-enhanced perfusion imaging).

Some encouraging first results have been shown for IVIM in rectal cancer to predict response [60, 61, 72], for differentiating between metastatic (N+) and non-metastatic (N-) lymph nodes [119], and to predict prognostic markers such as TNM-stage [102, 113], tumour differentiation grade [102], lymphovascular invasion [107], micro-vessel density [113] and KRAS status [114]. Potential drawbacks of the IVIM method are its test-retest reproducibility [120] and that measurements may be significantly influenced by scan parameters such as the echo time [121]. There is currently no consensus on how IVIM analysis should best be performed, as is also illustrated by the different imaging protocols used in the current literature [60, 61, 72, 102, 107, 113, 114, 119]. Moreover, results as to whether IVIM parameters provide added benefit compared to simple mean ADC measurements have so far been conflicting.

Diffusion kurtosis imaging

In addition to the perfusion effects that can be captured by the IVIM model, DKI takes into account effects of non-Gaussian diffusion (see **Figure 5**). In a free medium, diffusion is assumed to follow a Gaussian distribution. Since tissue contains barriers like cell membranes and vessels that influence the diffusivity, this assumption does not hold true for tissues. Especially for high b -values ($>b1000$) non-Gaussian diffusion effects can be observed. This non-Gaussian behaviour can be expressed in terms of kurtosis, which can be seen as a measure of a tissue's degree of heterogeneity [122]. The DKI model separates the signal into the apparent diffusion coefficient D_{app} (assuming a Gaussian distribution) and an apparent diffusional kurtosis K_{app} which expresses how much the measured signal departs from the assumed Gaussian distribution. A potential downside is that, similar to IVIM, DKI uses multiple model-based parameters and is therefore relatively susceptible to measurement inaccuracies [123, 124].

Evidence for DKI so far is limited. Two studies investigated diffusion kurtosis for response prediction in rectal cancer. Although these reports were in agreement in the sense that both found significant differences in $\Delta D_{app}\%$ after CRT between good and poor responders, results with respect to other studied parameters (e.g. pre-CRT DKI measures) were contradictory and the authors could not produce a statistically significant benefit for DKI parameters compared to routine DWI parameters (i.e.

ADC) to predict response [57, 125]. In other (preliminary) reports, the diffusion kurtosis coefficient has shown promise as a prognostic marker to predict metastases [105], tumour differentiation grade [103, 111], T-stage [103], N-stage [103, 111], lymphovascular invasion and involvement of the mesorectal fascia [103].

Automated DWI post-processing methods

Tumour segmentation is an important aspect of the workflow to be able to extract quantitative tumour parameters. Unfortunately, manual segmentation of rectal tumours is labour intensive, time consuming and often requires a relatively high level of experience. Given the high lesion-to-background ratio of tumours on DWI, it is a potentially suitable technique for automated (or semi-automated) segmentation methods. One study investigated the accuracy and time needed for tumour segmentation on DWI, using a semi-automated region growing algorithm with and without manual adjustments. The semi-automated method (with some manual adjustments) had excellent agreement with full manual segmentation and resulted in a significant reduction in delineation time for the radiologist [126]. Another group investigated a deep learning segmentation approach, incorporating information from both T2-weighted MRI and DWI to train a convolutional neural network to perform fully automated segmentation. The algorithm resulted in segmentations that were very comparable to those performed manually by expert readers with a good dice similarity index (DSI: a measure indicating the spatial overlap of voxels within the segmentations on a scale from 0 to 1) of 0.70 [127]. Although these automated segmentation methods will need to be further optimized and validated, they appear promising and will likely be helpful to reduce the workload of radiologists in future research and clinics.

2

CONCLUSIONS AND CLINICAL RECOMMENDATIONS

DWI in rectal cancer is an emerging topic of research and is now also increasingly finding its way to clinical practice. Over the last decade, use of DWI has evolved from qualitative visual image interpretation to increasingly advanced methods of quantitative analysis. So far the largest body of evidence exists for assessment of tumour response to neoadjuvant treatment. In this setting, particularly the benefit of DWI for visual assessment of residual tumour in post-radiation fibrosis has been established and is now increasingly adopted and highly recommended for clinical use. Promising

results have also been reported for quantitative DWI analysis (mainly ADC), both for response prediction as well as for overall tumour prognostication, but protocols require standardization and results will need to be prospectively confirmed on larger scale. Until then, clinical evaluation of DWI should be limited to visual (qualitative) assessment with no role for quantification in current daily practice. The role of DWI for further clinical tumour and nodal staging is less well-defined but appears to be limited, although there could be a benefit for DWI to help detect lymph nodes. Novel methods of analysis as well as new post-processing tools are still being developed; the role of these tools remains to be established in the upcoming years.

REFERENCES

1. Beets-Tan RGH, Lambregts DMJ, Maas M, et al (2018) Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 28:1465–1475. <https://doi.org/10.1007/s00330-017-5026-2>
2. Vega P, Valentín F, Cubiella J (2015) Colorectal cancer diagnosis: Pitfalls and opportunities. *World J Gastrointest Oncol* 7:422. <https://doi.org/10.4251/wjgo.v7.i12.422>
3. Soyer P, Lagadec M, Sirol M, et al (2010) Free-breathing diffusion-weighted single-shot echo-planar MR imaging using parallel imaging (GRAPPA 2) and high b value for the detection of primary rectal adenocarcinoma. *Cancer Imaging* 10:32–39. <https://doi.org/10.1102/1470-7330.2010.0011>
4. Ichikawa T, Erturk SM, Motosugi U, et al (2006) High-B-Value Diffusion-Weighted MRI in Colorectal Cancer. *Am J Roentgenol* 187:181–184. <https://doi.org/10.2214/AJR.05.1005>
5. Rao S-X, Zeng M-S, Chen C-Z, et al (2008) The value of diffusion-weighted imaging in combination with T2-weighted imaging for rectal cancer detection. *Eur J Radiol* 65:299–303. <https://doi.org/10.1016/j.ejrad.2007.04.001>
6. Leufkens AM, Kwee TC, van den Bosch MAAJ, et al (2013) Diffusion-weighted MRI for the detection of colorectal polyps: Feasibility study. *Magn Reson Imaging* 31:28–35. <https://doi.org/10.1016/j.mri.2012.06.029>
7. Tomizawa M, Shinozaki F, Uchida Y, et al (2015) Diffusion-weighted whole-body imaging with background body signal suppression/T2 image fusion and positron emission tomography/computed tomography of upper gastrointestinal cancers. *Abdom Imaging* 40:3012–3019. <https://doi.org/10.1007/s00261-015-0545-2>
8. Hosonuma T, Tozaki M, Ichiba N, et al (2006) Clinical usefulness of diffusion-weighted imaging using low and high b-values to detect rectal cancer. *Magn Reson Med Sci* 5:173–7
9. Ono K, Ochiai R, Yoshida T, et al (2009) Comparison of diffusion-weighted MRI and 2-[fluorine-18]-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for detecting primary colorectal cancer and regional lymph node metastases. *J Magn Reson Imaging* 29:336–340. <https://doi.org/10.1002/jmri.21638>
10. Shinya S, Sasaki T, Nakagawa Y, et al (2009) The efficacy of diffusion-weighted imaging for the detection of colorectal cancer. *Hepatogastroenterology* 56:128–132

11. Jia H, Ma X, Zhao Y, et al (2015) Meta-analysis of diffusion-weighted magnetic resonance imaging in identification of colorectal cancer. *Int J Clin Exp Med* 8:17333–17342
12. Nasu K, Kuroki Y, Minami M (2012) Diffusion-weighted imaging findings of mucinous carcinoma arising in the ano-rectal region: Comparison of apparent diffusion coefficient with that of tubular adenocarcinoma. *Jpn J Radiol* 30:120–127. <https://doi.org/10.1007/s11604-011-0023-x>
13. Kim M-J, Park JS, Park S II, et al (2003) Accuracy in Differentiation of Mucinous and Nonmucinous Rectal Carcinoma on MR Imaging. *J Comput Assist Tomogr* 27:48–55. <https://doi.org/10.1097/00004728-200301000-00010>
14. Hussain SM, Outwater EK, Siegelman ES (1999) Mucinous versus Nonmucinous Rectal Carcinomas: Differentiation with MR Imaging. *Radiology* 213:79–85. <https://doi.org/10.1148/radiology.213.1.r99se3879>
15. Lu Z-H, Hu C-H, Qian W-X, Cao W-H (2016) Preoperative diffusion-weighted imaging value of rectal cancer: Preoperative T staging and correlations with histological T stage. *Clin Imaging* 40:563–568. <https://doi.org/10.1016/j.clinimag.2015.12.006>
16. Feng Q, Yan YQ, Zhu J, XU JR (2014) T staging of rectal cancer: Accuracy of diffusion-weighted imaging compared with T2-weighted imaging on 3.0 tesla MRI. *J Dig Dis* 15:188–194. <https://doi.org/10.1111/1751-2980.12124>
17. Glynne-Jones R, Wyrwicz L, Tiret E, et al (2017) Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 28:iv22–iv40. <https://doi.org/10.1093/annonc/mdx224>
18. Brown G, Richards CJ, Bourne MW, et al (2003) Morphologic Predictors of Lymph Node Status in Rectal Cancer with Use of High-Spatial-Resolution MR Imaging with Histopathologic Comparison. *Radiology* 227:371–377. <https://doi.org/10.1148/radiol.2272011747>
19. Kim JH, Beets GL, Kim M-J, et al (2004) High-resolution MR imaging for nodal staging in rectal cancer: are there any criteria in addition to the size? *Eur J Radiol* 52:78–83. <https://doi.org/10.1016/j.ejrad.2003.12.005>
20. Mizukami Y, Ueda S, Mizumoto A, et al (2011) Diffusion-weighted Magnetic Resonance Imaging for Detecting Lymph Node Metastasis of Rectal Cancer. *World J Surg* 35:895–899. <https://doi.org/10.1007/s00268-011-0986-x>
21. Bipat S, Glas AS, Slors FJM, et al (2004) Rectal Cancer: Local Staging and Assessment of Lymph Node Involvement with Endoluminal US, CT, and MR Imaging—A Meta-Analysis. *Radiology* 232:773–783. <https://doi.org/10.1148/radiol.2323031368>

22. Lahaye MJ, Engelen SME, Nelemans PJ, et al (2005) 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 MRI* 26:259–268. <https://doi.org/10.1053/j.sult.2005.04.005>
23. Heijnen LA, Lambregts DMJ, Mondal D, et al (2013) Diffusion-weighted MR imaging in primary rectal cancer staging demonstrates but does not characterise lymph nodes. *Eur Radiol* 23:3354–3360. <https://doi.org/10.1007/s00330-013-2952-5>
24. Lambregts DMJ, Maas M, Riedl RG, et al (2011) Value of ADC measurements for nodal staging after chemoradiation in locally advanced rectal cancer—a per lesion validation study. *Eur Radiol* 21:265–273. <https://doi.org/10.1007/s00330-010-1937-x>
25. Nakai G, Matsuki M, Inada Y, et al (2008) Detection and Evaluation of Pelvic Lymph Nodes in Patients With Gynecologic Malignancies Using Body Diffusion-Weighted Magnetic Resonance Imaging. *J Comput Assist Tomogr* 32:764–768. <https://doi.org/10.1097/RCT.0b013e318153fd43>
26. Mir N, Sohaib SA, Collins D, Koh DM (2010) 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* 54:358–364. <https://doi.org/10.1111/j.1754-9485.2010.02182.x>
27. Kim SH, Yoon J-H, Lee Y (2015) Added value of morphologic characteristics on diffusion-weighted images for characterizing lymph nodes in primary rectal cancer. *Clin Imaging* 39:1046–1051. <https://doi.org/10.1016/j.clinimag.2015.07.022>
28. Yasui O, Sato M, Kamada A (2009) Diffusion-Weighted Imaging in the Detection of Lymph Node Metastasis in Colorectal Cancer. *Tohoku J Exp Med* 218:177–183. <https://doi.org/10.1620/tjem.218.177>
29. Cho EY, Kim SH, Yoon JH, et al (2013) Apparent diffusion coefficient for discriminating metastatic from non-metastatic lymph nodes in primary rectal cancer. *Eur J Radiol* 82:e662–e668. <https://doi.org/10.1016/j.ejrad.2013.08.007>
30. Zhao Q, Liu L, Wang Q, et al (2014) Preoperative diagnosis and staging of rectal cancer using diffusion-weighted and water imaging combined with dynamic contrast-enhanced scanning. *Oncol Lett* 8:2734–2740. <https://doi.org/10.3892/ol.2014.2590>
31. Cerny M, Dunet V, Prior JO, et al (2016) Initial Staging of Locally Advanced Rectal Cancer and Regional Lymph Nodes. *Clin Nucl Med* 41:289–295. <https://doi.org/10.1097/RLU.0000000000001172>
32. Kim SH, Ryu KH, Yoon J-H, et al (2015) Apparent diffusion coefficient for lymph node characterization after chemoradiation therapy for locally advanced rectal cancer. *Acta Radiol* 56:1446–1453. <https://doi.org/10.1177/0284185114560936>

33. Kwee TC, Takahara T, Luijten PR, Nieuvelstein RAJ (2010) ADC measurements of lymph nodes: Inter- and intra-observer reproducibility study and an overview of the literature. *Eur J Radiol* 75:215–220. <https://doi.org/10.1016/j.ejrad.2009.03.026>
34. van der Valk MJM, Hilling DE, Bastiaannet E, et al (2018) Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. *Lancet* 391:2537–2545. [https://doi.org/10.1016/S0140-6736\(18\)31078-X](https://doi.org/10.1016/S0140-6736(18)31078-X)
35. van der Paardt MP, Zagers MB, Beets-Tan RGH, et al (2013) Patients Who Undergo Preoperative Chemoradiotherapy for Locally Advanced Rectal Cancer Restaged by Using Diagnostic MR Imaging: A Systematic Review and Meta-Analysis. *Radiology* 269:101–112. <https://doi.org/10.1148/radiol.13122833>
36. Kim SH, Lee JM, Hong SH, et al (2009) Locally Advanced Rectal Cancer: Added Value of Diffusion-weighted MR Imaging in the Evaluation of Tumor Response to Neoadjuvant Chemo- and Radiation Therapy. *Radiology* 253:116–125. <https://doi.org/10.1148/radiol.2532090027>
37. Lambregts DMJ, Vandecaveye V, Barbaro B, et al (2011) Diffusion-Weighted MRI for Selection of Complete Responders After Chemoradiation for Locally Advanced Rectal Cancer: A Multicenter Study. *Ann Surg Oncol* 18:2224–2231. <https://doi.org/10.1245/s10434-011-1607-5>
38. Park MJ, Kim SH, Lee SJ, et al (2011) Locally Advanced Rectal Cancer: Added Value of Diffusion-weighted MR Imaging for Predicting Tumor Clearance of the Mesorectal Fascia after Neoadjuvant Chemotherapy and Radiation Therapy. *Radiology* 260:771–780. <https://doi.org/10.1148/radiol.11102135>
39. Song I, Kim SH, Lee SJ, et al (2012) Value of diffusion-weighted imaging in the detection of viable tumour after neoadjuvant chemoradiation therapy in patients with locally advanced rectal cancer: comparison with T 2 weighted and PET/CT imaging. *Br J Radiol* 85:577–586. <https://doi.org/10.1259/bjr/68424021>
40. Sassen S, de Booi M, Sosef M, et al (2013) Locally advanced rectal cancer: is diffusion weighted MRI helpful for the identification of complete responders (ypT0N0) after neoadjuvant chemoradiation therapy? *Eur Radiol* 23:3440–3449. <https://doi.org/10.1007/s00330-013-2956-1>
41. Marouf RA, Tadros MY, Ahmed TY (2015) Value of diffusion-weighted MR imaging in assessing response of neoadjuvant chemo and radiation therapy in locally advanced rectal cancer. *Egypt J Radiol Nucl Med* 46:553–561. <https://doi.org/10.1016/j.ejrn.2015.03.005>

42. Foti PV, Privitera G, Piana S, et al (2016) Locally advanced rectal cancer: Qualitative and quantitative evaluation of diffusion-weighted MR imaging in the response assessment after neoadjuvant chemo-radiotherapy. *Eur J Radiol Open* 3:145–152. <https://doi.org/10.1016/j.ejro.2016.06.003>
43. Lambregts DMJ, van Heeswijk MM, Delli Pizzi A, et al (2017) Diffusion-weighted MRI to assess response to chemoradiotherapy in rectal cancer: main interpretation pitfalls and their use for teaching. *Eur Radiol* 27:4445–4454. <https://doi.org/10.1007/s00330-017-4830-z>
44. Lambregts DMJ, Delli Pizzi A, Lahaye MJ, et al (2018) A Pattern-Based Approach Combining Tumor Morphology on MRI With Distinct Signal Patterns on Diffusion-Weighted Imaging to Assess Response of Rectal Tumors After Chemoradiotherapy. *Dis Colon Rectum* 61:1. <https://doi.org/10.1097/DCR.0000000000000915>
45. Wu L-M, Zhu J, Hu J, et al (2013) Is there a benefit in using magnetic resonance imaging in the prediction of preoperative neoadjuvant therapy response in locally advanced rectal cancer? *Int J Colorectal Dis* 28:1225–1238. <https://doi.org/10.1007/s00384-013-1676-y>
46. Lee MA, Cho SH, Seo AN, et al (2017) Modified 3-Point MRI-Based Tumor Regression Grade Incorporating DWI for Locally Advanced Rectal Cancer. *AJR Am J Roentgenol* 209:1247–1255. <https://doi.org/10.2214/AJR.16.17242>
47. van Heeswijk MM, Lambregts DMJ, Palm WM, et al (2017) DWI for Assessment of Rectal Cancer Nodes After Chemoradiotherapy: Is the Absence of Nodes at DWI Proof of a Negative Nodal Status? *Am J Roentgenol* 208:W79–W84. <https://doi.org/10.2214/AJR.16.17117>
48. Ryu KH, Kim SH, Yoon J-H, et al (2016) Diffusion-weighted imaging for evaluating lymph node eradication after neoadjuvant chemoradiation therapy in locally advanced rectal cancer. *Acta radiol* 57:133–141. <https://doi.org/10.1177/0284185114568908>
49. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2011) Rectal Cancer: Assessment of Complete Response to Preoperative Combined Radiation Therapy with Chemotherapy—Conventional MR Volumetry versus Diffusion-weighted MR Imaging. *Radiology* 260:734–743. <https://doi.org/10.1148/radiol.11102467>
50. Lambregts DMJ, Rao S-X, Sassen S, et al (2015) MRI and Diffusion-weighted MRI Volumetry for Identification of Complete Tumor Responders After Preoperative Chemoradiotherapy in Patients With Rectal Cancer. *Ann Surg* 262:1034–1039. <https://doi.org/10.1097/SLA.0000000000000909>

51. Sathyakumar K, Chandramohan A, Masih D, et al (2016) Best MRI predictors of complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. *Br J Radiol* 89:20150328. <https://doi.org/10.1259/bjr.20150328>
52. Ha HI, Kim AY, Yu CS, et al (2013) Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after preoperative chemoradiation therapy. *Eur Radiol* 23:3345–3353. <https://doi.org/10.1007/s00330-013-2936-5>
53. Jung SH, Heo SH, Kim JW, et al (2012) Predicting response to neoadjuvant chemoradiation therapy in locally advanced rectal cancer: Diffusion-weighted 3 tesla MR imaging. *J Magn Reson Imaging* 35:110–116. <https://doi.org/10.1002/jmri.22749>
54. Birlik B, Obuz F, Elibol FD, et al (2015) Diffusion-weighted MRI and MR- volumetry - in the evaluation of tumor response after preoperative chemoradiotherapy in patients with locally advanced rectal cancer. *Magn Reson Imaging* 33:201–212. <https://doi.org/10.1016/j.mri.2014.08.041>
55. Sun Y-S, Zhang X-P, Tang L, et al (2010) Locally Advanced Rectal Carcinoma Treated with Preoperative Chemotherapy and Radiation Therapy: Preliminary Analysis of Diffusion-weighted MR Imaging for Early Detection of Tumor Histopathologic Downstaging. *Radiology* 254:170–178. <https://doi.org/10.1148/radiol.2541082230>
56. Elmi A, Hedgire SS, Covarrubias D, et al (2013) Apparent diffusion coefficient as a non-invasive predictor of treatment response and recurrence in locally advanced rectal cancer. *Clin Radiol* 68:e524–e531. <https://doi.org/10.1016/j.crad.2013.05.094>
57. Hu F, Tang W, Sun Y, et al (2015) The value of diffusion kurtosis imaging in assessing pathological complete response to neoadjuvant chemoradiation therapy in rectal cancer: a comparison with conventional diffusion-weighted imaging. *Oncotarget* 8:75597–75606. <https://doi.org/10.18632/oncotarget.17491>
58. Iannicelli E, Di Pietropaolo M, Pilozi E, et al (2016) Value of diffusion-weighted MRI and apparent diffusion coefficient measurements for predicting the response of locally advanced rectal cancer to neoadjuvant chemoradiotherapy. *Abdom Radiol* 41:1906–1917. <https://doi.org/10.1007/s00261-016-0805-9>
59. Kim YC, Lim JS, Keum KC, et al (2011) 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* 34:570–576. <https://doi.org/10.1002/jmri.22696>

60. Nougaret S, Vargas HA, Lakhman Y, et al (2016) 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* 280:446–454. <https://doi.org/10.1148/radiol.2016150702>
61. Lu W, Jing H, Ju-Mei Z, et al (2017) Intravoxel incoherent motion diffusion-weighted imaging for discriminating the pathological response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Sci Rep* 7:8496. <https://doi.org/10.1038/s41598-017-09227-9>
62. Quaiia E, Gennari AG, Ricciardi MC, et al (2016) Value of percent change in tumoral volume measured at T2-weighted and diffusion-weighted MRI to identify responders after neoadjuvant chemoradiation therapy in patients with locally advanced rectal carcinoma. *J Magn Reson Imaging* 44:1415–1424. <https://doi.org/10.1002/jmri.25310>
63. Jacobs L, Intven M, van Lelyveld N, et al (2016) Diffusion-weighted MRI for Early Prediction of Treatment Response on Preoperative Chemoradiotherapy for Patients With Locally Advanced Rectal Cancer. *Ann Surg* 263:522–528. <https://doi.org/10.1097/SLA.0000000000001311>
64. Barbaro B, Vitale R, Valentini V, et al (2012) Diffusion-Weighted Magnetic Resonance Imaging in Monitoring Rectal Cancer Response to Neoadjuvant Chemoradiotherapy. *Int J Radiat Oncol* 83:594–599. <https://doi.org/10.1016/j.ijrobp.2011.07.017>
65. Intven M, Reerink O, Philippens MEP (2013) Diffusion-weighted MRI in locally advanced rectal cancer: pathological response prediction after neo-adjuvant radiochemotherapy. *Strahlenther Onkol* 189:117–22. <https://doi.org/10.1007/s00066-012-0270-5>
66. Cai G, Xu Y, Zhu J, et al (2013) Diffusion-weighted magnetic resonance imaging for predicting the response of rectal cancer to neoadjuvant concurrent chemoradiation. *World J Gastroenterol* 19:5520. <https://doi.org/10.3748/wjg.v19.i33.5520>
67. Blažić I, Maksimović R, Gajić M, Šaranović Đ (2015) Apparent diffusion coefficient measurement covering complete tumor area better predicts rectal cancer response to neoadjuvant chemoradiotherapy. *Croat Med J* 56:460–469. <https://doi.org/10.3325/cmj.2015.56.460>
68. Ippolito D, Fior D, Trattenero C, et al (2015) Combined value of apparent diffusion coefficient-standardized uptake value max in evaluation of post-treated locally advanced rectal cancer. *World J Radiol* 7:509. <https://doi.org/10.4329/wjr.v7.i12.509>

69. Ippolito D, Monguzzi L, Guerra L, et al (2012) Response to neoadjuvant therapy in locally advanced rectal cancer: Assessment with diffusion-weighted MR imaging and 18FDG PET/CT. *Abdom Imaging* 37:1032–1040. <https://doi.org/10.1007/s00261-011-9839-1>
70. Monguzzi L, Ippolito D, Bernasconi DP, et al (2013) Locally advanced rectal cancer: Value of ADC mapping in prediction of tumor response to radiochemotherapy. *Eur J Radiol* 82:234–240. <https://doi.org/10.1016/j.ejrad.2012.09.027>
71. Intven M, Monninkhof EM, Reerink O, Philippens MEP (2015) Combined T2w volumetry, DW-MRI and DCE-MRI for response assessment after neo-adjuvant chemoradiation in locally advanced rectal cancer. *Acta Oncol (Madr)* 54:1729–1736. <https://doi.org/10.3109/0284186X.2015.1037010>
72. Bakke KM, Hole KH, Dueland S, et al (2017) Diffusion-weighted magnetic resonance imaging of rectal cancer: tumour volume and perfusion fraction predict chemoradiotherapy response and survival. *Acta Oncol (Madr)* 56:813–818. <https://doi.org/10.1080/0284186X.2017.1287951>
73. Kremser C, Judmaier W, Hein P, et al (2003) Preliminary Results on the Influence of Chemoradiation on Apparent Diffusion Coefficients of Primary Rectal Carcinoma Measured by Magnetic Resonance Imaging. *Strahlentherapie und Onkol* 179:641–649. <https://doi.org/10.1007/s00066-003-1045-9>
74. Hein PA, Kremser C, Judmaier W, et al (2003) Diffusionsgewichtete magnetresonanztomographie - Ein neuer prognoseparameter beim fortgeschrittenen rektumkarzinom? *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgeb Verfahren* 175:381–386. <https://doi.org/10.1055/s-2003-37836>
75. Lambrecht M, Vandecaveye V, De Keyzer F, et al (2012) 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* 82:863–870. <https://doi.org/10.1016/j.ijrobp.2010.12.063>
76. DeVries AF, Kremser C, Hein PA, et al (2003) Tumor microcirculation and diffusion predict therapy outcome for primary rectal carcinoma. *Int J Radiat Oncol* 56:958–965. [https://doi.org/10.1016/S0360-3016\(03\)00208-6](https://doi.org/10.1016/S0360-3016(03)00208-6)
77. Hein PA, Kremser C, Judmaier W, et al (2003) 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* 45:214–222. [https://doi.org/10.1016/S0720-048X\(02\)00231-0](https://doi.org/10.1016/S0720-048X(02)00231-0)

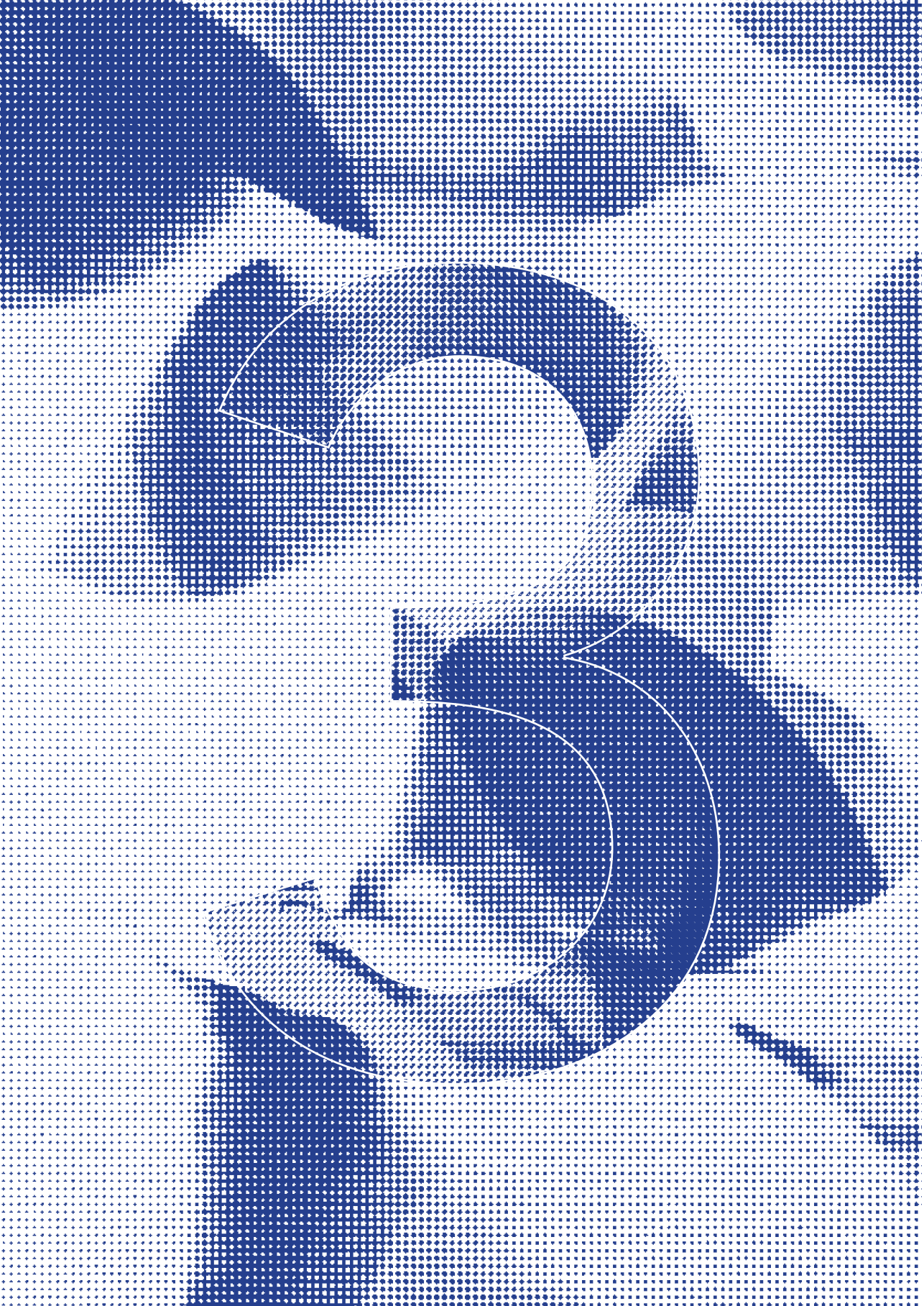
78. Lambrecht M, Deroose C, Roels S, et al (2010) The use of FDG-PET/CT and diffusion-weighted magnetic resonance imaging for response prediction before, during and after preoperative chemoradiotherapy for rectal cancer. *Acta Oncol (Madr)* 49:956–963. <https://doi.org/10.3109/0284186X.2010.498439>
79. Chen Y-G, Chen M-Q, Guo Y-Y, et al (2016) Apparent Diffusion Coefficient Predicts Pathology Complete Response of Rectal Cancer Treated with Neoadjuvant Chemoradiotherapy. *PLoS One* 11:e0153944. <https://doi.org/10.1371/journal.pone.0153944>
80. Blazic IM, Lilic GB, Gajic MM (2017) Quantitative Assessment of Rectal Cancer Response to Neoadjuvant Combined Chemotherapy and Radiation Therapy: Comparison of Three Methods of Positioning Region of Interest for ADC Measurements at Diffusion-weighted MR Imaging. *Radiology* 282:418–428. <https://doi.org/10.1148/radiol.2016151908>
81. Genovesi D, Filippone A, Ausili Cèfaro G, et al (2013) Diffusion-weighted magnetic resonance for prediction of response after neoadjuvant chemoradiation therapy for locally advanced rectal cancer: Preliminary results of a monoinstitutional prospective study. *Eur J Surg Oncol* 39:1071–1078. <https://doi.org/10.1016/j.ejso.2013.07.090>
82. Kim SH, Lee JY, Lee JM, et al (2011) Apparent diffusion coefficient for evaluating tumour response to neoadjuvant chemoradiation therapy for locally advanced rectal cancer. *Eur Radiol* 21:987–995. <https://doi.org/10.1007/s00330-010-1989-y>
83. Bassaneze T, Gonçalves JE, Faria JF, et al (2017) Quantitative aspects of diffusion-weighted magnetic resonance imaging in rectal cancer response to neoadjuvant therapy. *Radiol Oncol* 51:270–276. <https://doi.org/10.1515/raon-2017-0025>
84. Choi MH, Oh SN, Rha SE, et al (2016) Diffusion-weighted imaging: Apparent diffusion coefficient histogram analysis for detecting pathologic complete response to chemoradiotherapy in locally advanced rectal cancer. *J Magn Reson Imaging* 44:212–220. <https://doi.org/10.1002/jmri.25117>
85. Cho SH, Kim GC, Jang Y-J, et al (2015) Locally advanced rectal cancer: post-chemoradiotherapy ADC histogram analysis for predicting a complete response. *Acta radiol* 56:1042–1050. <https://doi.org/10.1177/0284185114550193>
86. Cai P-Q, Wu Y-P, An X, et al (2014) Simple measurements on diffusion-weighted MR imaging for assessment of complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 24:2962–70. <https://doi.org/10.1007/s00330-014-3251-5>

87. De Cecco CN, Ciolina M, Caruso D, et al (2016) Performance of diffusion-weighted imaging, perfusion imaging, and texture analysis in predicting tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3T MR: initial experience. *Abdom Radiol* 41:1728–1735. <https://doi.org/10.1007/s00261-016-0733-8>
88. De Felice F, Magnante AL, Musio D, et al (2017) Diffusion-weighted magnetic resonance imaging in locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. *Eur J Surg Oncol* 43:1324–1329. <https://doi.org/10.1016/j.ejso.2017.03.010>
89. Engin G, Sharifov R, Gural Z, et al (2012) Can diffusion-weighted MRI determine complete responders after neoadjuvant chemoradiation for locally advanced rectal cancer? *Diagnostic Interv Radiol* 18:574–81. <https://doi.org/10.4261/1305-3825.DIR.5755-12.1>
90. Kim SH, Lee JM, Moon SK, et al (2012) Evaluation of lymph node metastases: Comparison of gadofluorine M-enhanced MRI and diffusion-weighted MRI in a rabbit VX2 rectal cancer model. *J Magn Reson Imaging* 35:1179–1186. <https://doi.org/10.1002/jmri.23513>
91. Fusco R, Sansone M, Petrillo A (2017) A comparison of fitting algorithms for diffusion-weighted MRI data analysis using an intravoxel incoherent motion model. *Magn Reson Mater Physics, Biol Med* 30:113–120. <https://doi.org/10.1007/s10334-016-0591-y>
92. Herneth AM, Guccione S, Bednarski M (2003) Apparent Diffusion Coefficient: a quantitative parameter for in vivo tumor characterization. *Eur J Radiol* 45:208–213. [https://doi.org/10.1016/S0720-048X\(02\)00310-8](https://doi.org/10.1016/S0720-048X(02)00310-8)
93. Gray LH, Conger AD, Ebert M, et al (1953) The Concentration of Oxygen Dissolved in Tissues at the Time of Irradiation as a Factor in Radiotherapy. *Br J Radiol* 26:638–648. <https://doi.org/10.1259/0007-1285-26-312-638>
94. Jordan BF, Sonveaux P (2012) Targeting Tumor Perfusion and Oxygenation to Improve the Outcome of Anticancer Therapy1. *Front Pharmacol* 3:1–15. <https://doi.org/10.3389/fphar.2012.00094>
95. Joye I, Deroose CM, Vandecaveye V, Haustermans K (2014) The role of diffusion-weighted MRI and 18F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: A systematic review. *Radiother Oncol* 113:158–165. <https://doi.org/10.1016/j.radonc.2014.11.026>

96. van Heeswijk MM, Lambregts DMJ, Maas M, et al (2017) Measuring the apparent diffusion coefficient in primary rectal tumors: is there a benefit in performing histogram analyses? *Abdom Radiol* 42:1627–1636. <https://doi.org/10.1007/s00261-017-1062-2>
97. Chidambaram V, Brierley JD, Cummings B, et al (2017) Investigation of volumetric apparent diffusion coefficient histogram analysis for assessing complete response and clinical outcomes following pre-operative chemoradiation treatment for rectal carcinoma. *Abdom Radiol* 42:1310–1318. <https://doi.org/10.1007/s00261-016-1010-6>
98. Lambregts DMJ, Cappendijk VC, Maas M, et al (2011) Value of MRI and diffusion-weighted MRI for the diagnosis of locally recurrent rectal cancer. *Eur Radiol* 21:1250–1258. <https://doi.org/10.1007/s00330-010-2052-8>
99. Colosio A, Soyer P, Rousset P, et al (2014) Value of diffusion-weighted and gadolinium-enhanced MRI for the diagnosis of pelvic recurrence from colorectal cancer. *J Magn Reson Imaging* 40:306–313. <https://doi.org/10.1002/jmri.24366>
100. Lambregts DMJ, Lahaye MJ, Heijnen LA, et al (2016) MRI and diffusion-weighted MRI to diagnose a local tumour regrowth during long-term follow-up of rectal cancer patients treated with organ preservation after chemoradiotherapy. *Eur Radiol* 26:2118–2125. <https://doi.org/10.1007/s00330-015-4062-z>
101. Hupkens BJP, Maas M, Martens MH, et al (2017) MRI surveillance for the detection of local recurrence in rectal cancer after transanal endoscopic microsurgery. *Eur Radiol* 4960–4969. <https://doi.org/10.1007/s00330-017-4853-5>
102. Sun H, Xu Y, Song A, et al (2018) Intravoxel Incoherent Motion MRI of Rectal Cancer: Correlation of Diffusion and Perfusion Characteristics With Prognostic Tumor Markers. *Am J Roentgenol* 210:W139–W147. <https://doi.org/10.2214/AJR.17.18342>
103. Cui Y, Yang X, Du X, et al (2018) Whole-tumour diffusion kurtosis MR imaging histogram analysis of rectal adenocarcinoma: Correlation with clinical pathologic prognostic factors. *Eur Radiol* 28:1485–1494. <https://doi.org/10.1007/s00330-017-5094-3>
104. Sun Y, Tong T, Cai S, et al (2014) Apparent Diffusion Coefficient (ADC) Value: A Potential Imaging Biomarker That Reflects the Biological Features of Rectal Cancer. *PLoS One* 9:e109371. <https://doi.org/10.1371/journal.pone.0109371>
105. Yu J, Huang D-YY, Li Y, et al (2016) Correlation of standard diffusion-weighted imaging and diffusion kurtosis imaging with distant metastases of rectal carcinoma. *J Magn Reson Imaging* 44:221–229. <https://doi.org/10.1002/jmri.25137>

106. Liu L, Liu Y, Xu L, et al (2017) Application of texture analysis based on apparent diffusion coefficient maps in discriminating different stages of rectal cancer. *J Magn Reson Imaging* 45:1798–1808. <https://doi.org/10.1002/jmri.25460>
107. Yan C, Pan X, Chen G, et al (2017) A pilot study on correlations between preoperative intravoxel incoherent motion MR imaging and postoperative histopathological features of rectal cancers. *Transl Cancer Res* 6:1050–1060. <https://doi.org/10.21037/tcr.2017.08.23>
108. Akashi M, Nakahusa Y, Yakabe T, et al (2014) 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* 55:524–531. <https://doi.org/10.1177/0284185113503154>
109. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2012) Diffusion-weighted MRI in rectal cancer: Apparent diffusion coefficient as a potential noninvasive marker of tumor aggressiveness. *J Magn Reson Imaging* 35:1365–1371. <https://doi.org/10.1002/jmri.23589>
110. Meng X, Li H, Kong L, et al (2016) MRI In rectal cancer: Correlations between MRI features and molecular markers Ki-67, HIF-1 α , and VEGF. *J Magn Reson Imaging* 44:594–600. <https://doi.org/10.1002/jmri.25195>
111. Zhu L, Pan Z, Ma Q, et al (2017) Diffusion Kurtosis Imaging Study of Rectal Adenocarcinoma Associated with Histopathologic Prognostic Factors: Preliminary Findings. *Radiology* 284:66–76. <https://doi.org/10.1148/radiol.2016160094>
112. Surov A, Meyer HJ, Wienke A (2017) Associations between apparent diffusion coefficient (ADC) and Ki 67 in different tumors: a meta-analysis. Part 1: ADCmean. *Oncotarget* 8:75434–75444. <https://doi.org/10.18632/oncotarget.20406>
113. Surov A, Meyer HJ, Höhn A, et al (2017) Correlations between intravoxel incoherent motion (IVIM) parameters and histological findings in rectal cancer: preliminary results. *Oncotarget* 8:21974–21983. <https://doi.org/10.18632/oncotarget.15753>
114. Xu Y, Xu Q, Sun H, et al (2018) Could IVIM and ADC help in predicting the KRAS status in patients with rectal cancer? *Eur Radiol*. <https://doi.org/10.1007/s00330-018-5329-y>
115. Noda Y, Goshima S, Kajita K, et al (2018) Prognostic Value of Diffusion MR Imaging and Clinical-Pathologic Factors in Patients with Rectal Cancer. *Iran J Radiol In Press*: <https://doi.org/10.5812/iranjradiol.57080>
116. Moon SJ, Cho SH, Kim GC, et al (2016) Complementary value of pre-treatment apparent diffusion coefficient in rectal cancer for predicting tumor recurrence. *Abdom Radiol* 41:1237–1244. <https://doi.org/10.1007/s00261-016-0648-4>

117. Le Bihan D, Breton E, Lallemand D, et al (1988) Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology* 168:497–505. <https://doi.org/10.1148/radiology.168.2.3393671>
118. Bihan D Le, Turner R (1992) The capillary network: a link between ivim and classical perfusion. *Magn Reson Med* 27:171–178. <https://doi.org/10.1002/mrm.1910270116>
119. Yu X-P, Wen L, Hou J, et al (2016) Discrimination between metastatic and nonmetastatic mesorectal lymph nodes in rectal cancer using intravoxel incoherent motion diffusion-weighted magnetic resonance imaging. *Acad Radiol* 23:479–485. <https://doi.org/10.1016/j.acra.2015.12.013>
120. Sun H, Xu Y, Xu Q, et al (2017) Rectal cancer: Short-term reproducibility of intravoxel incoherent motion parameters in 3.0T magnetic resonance imaging. *Medicine (Baltimore)* 96:e6866. <https://doi.org/10.1097/MD.0000000000006866>
121. Lemke A, Laun FB, Simon D, et al (2010) An in vivo verification of the intravoxel incoherent motion effect in diffusion-weighted imaging of the abdomen. *Magn Reson Med* 64:1580–1585. <https://doi.org/10.1002/mrm.22565>
122. Jensen JH, Helpert JA, Ramani A, et al (2005) Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med* 53:1432–1440. <https://doi.org/10.1002/mrm.20508>
123. Jensen JH, Helpert JA (2010) MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed* 23:698–710. <https://doi.org/10.1002/nbm.1518>
124. Hutchinson EB, Avram A V, Irfanoglu MO, et al (2017) Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models. *Magn Reson Med* 78:1767–1780. <https://doi.org/10.1002/mrm.26575>
125. Yu J, Xu Q, Song J-C, et al (2017) The value of diffusion kurtosis magnetic resonance imaging for assessing treatment response of neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 27:1848–1857. <https://doi.org/10.1007/s00330-016-4529-6>
126. Van Heeswijk MM, Lambregts DMJ, van Griethuysen JJM, et al (2016) Automated and semiautomated segmentation of rectal tumor volumes on diffusion-weighted MRI: Can it replace manual volumetry? *Int J Radiat Oncol Biol Phys* 94:824–831. <https://doi.org/10.1016/j.ijrobp.2015.12.017>
127. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, et al (2017) Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR. *Sci Rep* 7:5301. <https://doi.org/10.1038/s41598-017-05728-9>



Value of combined multiparametric MRI and FDG-PET/CT to identify well-responding rectal cancer patients before the start of neoadjuvant chemoradiation

Niels W. Schurink*, Lisa A. Min*, Maaïke Berbee, Wouter van Elmpt, Joost J.M. van Griethuysen, Frans C.H. Bakers, Sander Roberti, Simon R. Van Kranen, Max J. Lahaye, Monique Maas, Geerard L. Beets, Regina G.H. Beets-Tan, Doenja M.J. Lambregts

Eur Radiol. 2020; 30(5):2945-2954. doi: 10.1007/s00330-019-06638-2

* Shared first authorship

ABSTRACT

OBJECTIVES

To explore the value of multiparametric MRI combined with FDG-PET/CT to identify well-responding rectal cancer patients before the start of neoadjuvant chemoradiation.

METHODS

Sixty-one locally-advanced rectal cancer patients who underwent a baseline FDG-PET/CT and MRI (T2W+DWI) and received long course neoadjuvant chemoradiotherapy were retrospectively analyzed. Tumours were delineated on MRI and PET/CT from which the following quantitative parameters were calculated: T2W volume and entropy, ADC mean and entropy, CT density (mean-HU), SUV maximum and mean, metabolic tumor volume (MTV42%) and total lesion glycolysis (TLG). These features, together with sex, age, mrTN-stage ("baseline parameters") and the CRT-surgery interval were analyzed using multivariable stepwise logistic regression. Outcome was a good (TRG 1-2) versus poor histopathological response. Performance (AUC) to predict response was compared for different combinations of baseline +/- quantitative imaging parameters and performance in an 'independent' dataset was estimated using bootstrapped leave-one-out cross-validation (LOOCV).

RESULTS

The optimal multivariable prediction model consisted of a combination of baseline + quantitative imaging parameters and included mrT-stage (OR 0.004, $P < 0.001$), T2W-signal entropy (OR 7.81, $P = 0.0079$) and T2W volume (OR 1.028, $P = 0.0389$) as the selected predictors. AUC in the study dataset was 0.88 and 0.83 after LOOCV. No PET/CT-features were selected as predictors.

CONCLUSIONS

A multivariable model incorporating mrT-stage and quantitative parameters from baseline MRI can aid in identifying well-responding patients before the start of treatment. Addition of FDG-PET/CT is not beneficial.

INTRODUCTION

Current standard treatment for locally advanced rectal cancer (LARC) consists of long-course neo-adjuvant chemoradiotherapy (CRT) followed by surgery. In 15-25% of these patients, no residual tumour is found in the resection specimen [1, 2]. This has raised the question whether for this group surgery may be avoided [3, 4]. Organ-preserving treatments like the 'watch-and-wait' approach (W&W) are nowadays increasingly considered as an alternative to surgery, with good reported functional outcome, disease-free and overall survival [5-9].

At this point there is no pre-therapy classification method to predict how patients will respond to CRT. Although this information would currently not likely impact treatment, predicting response before the start of therapy could have a clinical impact in the future: in patients likely to respond well, neoadjuvant treatment may be further intensified to increase the chance of organ preservation, while in predicted non-responders futile CRT may be avoided. Pre-treatment response prediction may furthermore help create opportunities to select small and low-risk tumours (now typically managed with surgery without neoadjuvant treatment) to undergo CRT in case of a predicted good response, with the specific aim to achieve organ preservation [10]. These developments urge the need for accurate predictive biomarkers.

There is a growing interest in the value of imaging as a potential source for these biomarkers, with numerous reports exploring the potential of metabolic imaging (FDG-PET/CT) [11-14] and MRI with the addition of functional imaging sequences such as diffusion-weighted imaging (DWI) [15-19]. Most studies so far have focused on single-modality imaging and included only one or a few imaging markers. Linking multiparametric data from PET and MRI may be beneficial to provide a more comprehensive insight into underlying tumour biology. The few reports that have investigated such a multimodality PET/CT + MRI assessment in rectal cancer, suggested its potential, in particular when applying sequential imaging (pre- and post-CRT) and for higher-order (Radiomics) imaging variables [15, 23].

This study aims to further explore the value of combining baseline FDG-PET/CT and multiparametric MRI to identify before onset of treatment those patients that will respond well to neoadjuvant chemoradiation.

Methods

This study was approved by the local institutional review board. Informed consent was not required due to the retrospective nature of this study.

Patients

From 2008-2015 a cohort of 104 locally advanced ($\geq T3$ and/or N+) rectal cancer patients was identified from the local institutional database of the department of Radiation Oncology of Maastricht University Medical Center (Maastro Clinic), that underwent both routine MRI for primary tumour staging and an additional FDG-PET/CT at baseline (prior to any treatment), either as part of a previous study protocol (trial number NCT00969657) or for standard of care radiotherapy planning. From this cohort, 61 patients were selected based on the following inclusion criteria: (1) treatment consisting of long-course CRT followed by surgery or W&W, and (2) sufficient information to establish the treatment response outcome (histopathology or ≥ 2 years of clinical follow-up in case of W&W-surveillance). The standard CRT protocol consisted of 50.4 Gy with concurrent capecitabine-based chemotherapy. Patients who received a non-standardized treatment, had insufficient quality imaging or mucinous tumour histology were excluded (see **Figure 1**).

Baseline (pre-treatment) imaging

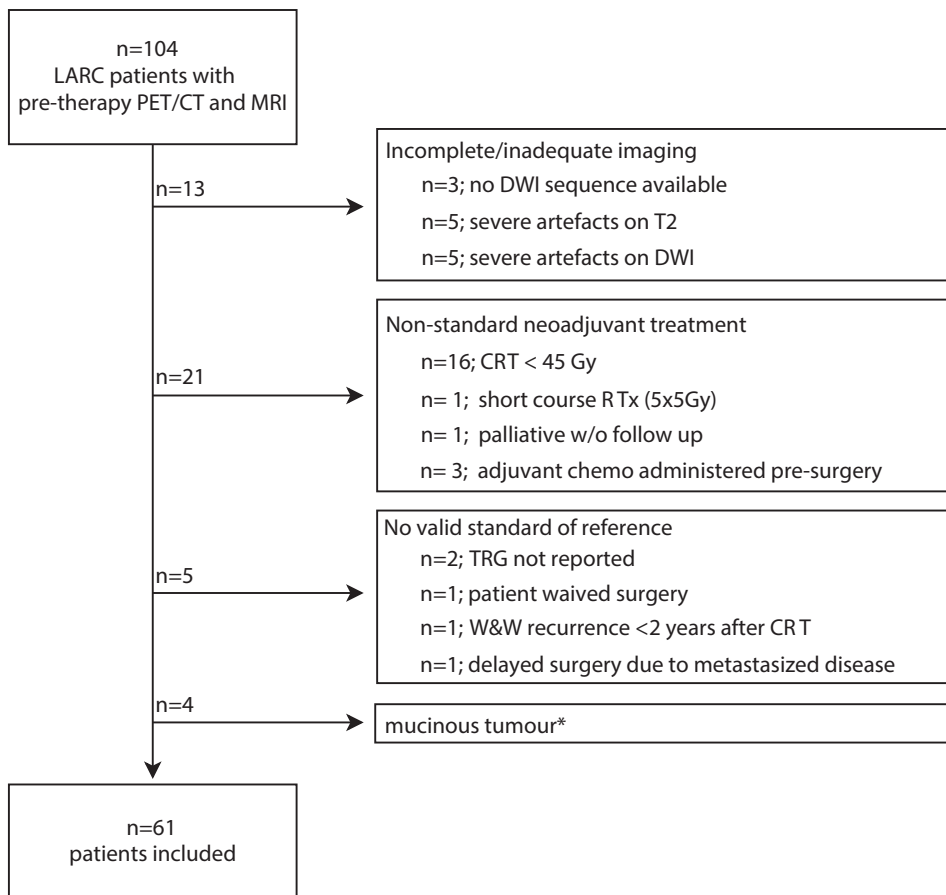
MRI

MRIs were performed at 1.5 Tesla (Intera (Achieva) $n=43$ or Ingenia $n=18$, Philips Healthcare) and included a T2W-sequence in 3 orthogonal directions, and an axial DWI-sequence including b-values $b=0$ and $b=1000$ s/mm². Apparent diffusion coefficient (ADC) maps were calculated by fitting a mono-exponential decay function to the $b=0$ and $b=1000$ s/mm² images. The axial T2W-MRI and DWI were angled in identical planes, perpendicular to the tumour axis. Further protocol details are given in **Table 1**. Patients received no spasmolytic or bowel preparation/filling.

FDG-PET/CT

¹⁸F-FDG PET/CT was performed on a Siemens Biograph 40 TruePoint PET/CT scanner (SIEMENS medical). A bolus of 2-deoxy-2-[¹⁸F]fluoro-D-glucose (¹⁸F FDG, from here on: FDG) of 2.5MBq/kg ($n=52$) or 4.0MBq/kg ($n=9$) was administered intravenously, after a 6-hour fast (blood glucose level <10 mmol/L). Scanning started after an incubation time of 60 (± 5) minutes, with 5 minutes per bed position, and ran from the skull base to upper-thighs (reconstructed to 3mm slice thickness, 4.07mm in-plane resolution). A non-enhanced CT scan (120 KVp, 113-297 mAs with automatic dose modulation) was acquired for attenuation correction, anatomical correlation and radiotherapy planning (reconstructed to 3mm slice thickness, 0.98mm in-plane resolution).

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF NEOADJUVANT CHEMORADIATION



3

Figure 1. Patient in- and exclusion flowchart. CRT: chemoradiotherapy; LARC: locally-advanced rectal cancer ($\geq T3$ and/or N+) RTx: radiotherapy; TRG: tumour regression grade (Mandard's); W&W: watch-and-wait. * Predominantly mucinous tumours were excluded because these typically exhibit distinctly different characteristics on PET and MRI and show a different response to CRT.

Quantitative MRI and PET/CT parameters

The image analysis workflow is illustrated in **Figure 2**. PET/CT and MR images were transferred to an offline workstation for tumour segmentation, performed using dedicated software (3D Slicer, version 4.8.1). Feature extraction was performed using the open-source software PyRadiomics (version 2.1.2) [24].

Table 1. MRI protocol

	T2-weighted	Diffusion-weighted
Echo time (ms)	130-150	65.74-84.88
Repetition time (ms)	3427-16738	2480-5545
Echo train length	25-28	53-87
Slice thickness (mm)	3-5 ^a	5
Slice gap (mm)	3.3-7.03	4-6.02
In-plane resolution (mm)	0.78125	1.25-1.71875
Number of averages	2-6	3-10
b-values (s/mm ²)	-	0, 1000 ^b
Fat-suppression	-	STIR (n=32), SPIR (n=7), SPAIR (n=22)

STIR: Short-TI Inversion Recovery; SPIR: Spectral Presaturation with Inversion Recovery; SPAIR: Spectral Attenuated Inversion Recovery.

^a n = 23 patients were scanned with 5 mm and n=38 with 3 mm axial slice thickness

^b Protocols included 3-7 b-values ranging from b0 to b2000 s/mm², but for the purpose of this study only the b=0 and b=1000 s/mm² series were used for analyses and to calculate the ADC map.

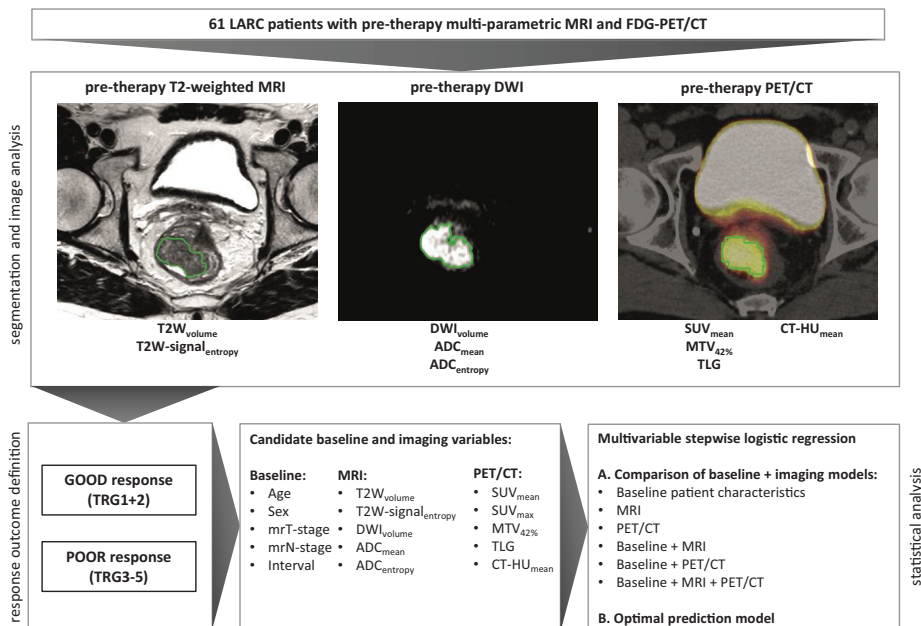


Figure 2. Schematic study outline

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF NEOADJUVANT CHEMORADIATION

A board-certified radiologist (DL, >9 years of rectal MRI experience) manually delineated whole-tumour volumes on the axial T2W-MRI and b1000-DWI, respectively, to calculate the following features: volume on T2W ($T2W_{\text{volume}}$, mesh-volume in PyRadiomics), entropy of the T2W signal intensity histogram ($T2W\text{-signal}_{\text{entropy}}$), volume on DWI (DWI_{volume} , mesh-volume in PyRadiomics), mean ADC (ADC_{mean}), and entropy of the ADC intensity histogram (ADC_{entropy}).

Metabolic tumour volumes ($MTV_{42\%}$) on PET/CT were semi-automatically segmented by one of the researchers experienced in PET segmentation (NS) by placing a volume of interest (VOI) over the tumour while taking care to avoid inclusion of physiologic uptake in the bladder. From this VOI the metabolic tumour volume was calculated using a threshold of 42% of the maximum standardized uptake value (SUV_{max}), according to methods previously described [25–27]. The $MTV_{42\%}$ was used to calculate the mean standardized uptake value (SUV_{mean}) and total lesion glycolysis (TLG; defined as $SUV_{\text{mean}} \times MTV_{42\%}$). The $MTV_{42\%}$ segmentation was transferred to CT to calculate the mean Hounsfield unit (HU) ($CT\text{-}HU_{\text{mean}}$).

The specific MRI and PET features described above were chosen as they represent relatively straightforward (1st order) variables reflecting tumour size, heterogeneity, cellularity and metabolism, which have all shown potential in previous reports and which are relatively simple to reproduce [16, 28–32].

Baseline patient characteristics

The following clinical baseline patient characteristics were documented: sex, age, and T- and N-stage derived from routine clinical staging with MRI (further referred to as mrT-stage and mrN-stage). The latter were dichotomized as mrT3c-4 vs. mrT1-3b and mrN+ vs. mrN0, respectively.

Response to chemoradiotherapy (standard of reference)

The primary outcome was the histopathological tumour regression grade (TRG) by Mandard [33]. Patients were classified as good responders (TRG1-2) or poor responders (TRG3-5). For W&W-patients, a recurrence-free follow-up of ≥ 2 years was used as a surrogate endpoint of a complete response. For the purpose of this study, these patients were considered complete responders (TRG1) and classified in the good responders group.

3

Statistical analysis

Statistical analysis was performed using R software (version 3.4.3; R Foundation for Statistical Computing, 2017).

The value of the quantitative MRI and PET/CT features and baseline patient characteristics to predict a good response was analyzed by multivariable logistic regression, consisting of a forward stepwise feature selection method based on the Akaike Information Criterion (AIC). The AIC describes model quality as a tradeoff between model fit and model complexity (i.e., the number of variables). A lower AIC indicates a better model, and is achieved by a better goodness of fit or fewer variables [34, 35]. The analysis workflow is summarized as follows:

- As described above, only a limited number of parameters ($T2W_{\text{volume}}$, $T2W\text{-signal}_{\text{entropy}}$, DWI_{volume} , ADC_{mean} , ADC_{entropy} , $MTV_{42\%}$, SUV_{max} , SUV_{mean} , TLG, CT-HU, mrT-stage, mrN-stage, age, sex) were assessed to limit overfitting. These parameters were defined before the onset of the study based on previous literature showing their potential promise as predictors of response [16, 28–31]. The interval between the last radiotherapy fraction and the final response evaluation (dichotomized as ≤ 10 vs. > 10 weeks) was added as an additional variable, as longer intervals have been reported to result in higher response rates and could thus act as a potential confounder [36].
- When two features showed a strong correlation (Pearson's correlation coefficient $\rho \geq 0.8$), only one was entered in the feature selection process to reduce effects of multicollinearity.
- The multivariable modelling process was repeated separately for different subsets and combinations of baseline and/or imaging variables (baseline only, MRI only, PET/CT only, baseline+MRI, baseline+PET/CT, baseline+PET/CT+MRI). To limit effects of overfitting, the number of variables selected for each model was set to a maximum of 1 feature per 10 patients in the smallest outcome group (3 features in total).
- Predictive performance of each model was assessed by calculating the area under the receiver-operating curve (AUC). Since our cohort size did not allow splitting of the data in a test and validation set, performance in an 'independent' dataset was estimated by performing leave-one-out cross-validation (LOOCV) with 500 bootstrap samples (to calculate confidence intervals). LOOCV involves building a model using the original dataset multiple times, while excluding one different

patient each time to predict the outcome. The cross-validated AUC is determined on the collective of these different predictions, and approximates the AUC in independent data.

To provide a complete overview of all investigated features, additional univariable logistic regression analysis was performed for each baseline and quantitative imaging variable. This was done independent of the multivariable analysis. P-values <0.05 were considered statistically significant.

RESULTS

3

Patient characteristics

Baseline patient characteristics are reported in **Table 2**. In total, 54/61 patients underwent surgery: 6 (10%) had a TRG1, 18 (30%) TRG2, 19 (31%) TRG3, 11 (18%) TRG4 and 0 (0%) TRG5. The remaining seven patients (12%) were monitored with W&W and had a sustained clinical complete response (median follow-up of 59 months, range 26-89). This resulted in 31 good responders (51%, TRG 1-2) and 30 poor responders (49%, TRG 3-5).

Comparison of different baseline and imaging models and their combinations

Results of the stepwise feature selection process including the different combinations of baseline patient characteristics, MRI and PET/CT variables are shown in **Table 3A**. The best fitting model (based on the smallest AIC) was the baseline + MRI model. The model PET/CT-only model had the poorest fit and addition of PET/CT features to the 'baseline only' or 'baseline + MRI' model was not beneficial. AUCs were 0.81 (baseline only), 0.70 (MRI only), 0.50 (PET/CT only), 0.88 (baseline + MRI), 0.81 (baseline + PET/CT) and 0.88 (baseline + MRI + PET/CT), respectively.

Optimized multivariable model

The optimized baseline + MRI model is summarized in **Table 3B**, and included mrT-stage (OR 0.004; 95%CI: 0.00 – 0.09 for cT3c-4 vs. cT1-3b), T2W-signal_{entropy} (OR per IQR 4.33; 95%CI: 1.47 – 12.77) and T2W_{volume} (OR 1.028 per cm³; 95%CI: 1.00 – 1.05). The model had an AUC of 0.88 to predict good responders within our dataset, with a sensitivity of 0.68 (95%CI: 0.49 – 0.83) when the ROC threshold was set at a specificity of 0.90. With leave-one-out cross-validation the found AUC was 0.83 (95%CI 0.70-0.96) with a sensitivity of 0.61 (95%CI: 0.42 – 0.78) at a specificity of 0.90.

Table 2. Baseline characteristics of study population

Baseline + staging	Male / Female	47 (77%) / 14 (23%)
	Age mean (sd)	68 (9)
	MRI-based T-stage (mrT-stage)	
	Early stage (mrT1-3b)	
	mrT1-2	5 (8%)
	mrT3a	0 (0%)
	mrT3b	34 (56%)
	Advanced stage (mrT3c-4b)	
	mrT3c	15 (25%)
	mrT3d	1 (2%)
	mrT4a	2 (3%)
	mrT4b	4 (7%)
	MRI-based N-stage (mrN-stage)	
	mrN0	16 (26%)
	mrN1	30 (49%)
	mrN2	15 (25%)
	Treatment post-CRT	
	Surgery	54 (88%)
	W&W	7 (12%)
Outcome	TRG (Mandard)	
	1 ^a	13 (21%)
	2	18 (30%)
	3	19 (31%)
	4	11 (18%)
	5	0 (0%)
	Good response (=TRG1-2) / Poor response (=TRG3-5)	31 (51%) / 30 (49%)
Treatment intervals (median N° days and interquartile range)	RT treatment duration	37 (36 – 51)
	Time from MRI to start of CRT	27 (9)
	Time from PET to start of CRT	7 (2)
	Time between PET and MRI	20 (9)
	Time from last RT fraction to sur- gery (n=54 patients)	71 (8)
	Time from last RT fraction to W&W inclusion (n=7 patients)	56 (4)

CRT: chemoradiotherapy; W&W: watch-and-wait follow-up; TRG: tumour regression grade;
RT: radiotherapy

^a7/13 patients were followed up according to a watch-and-wait program and had a sustained clinical complete response for at least 2 years (median follow-up 59 months, range 26-89). This was used as a surrogate endpoint for a pathological complete response (TRG1).

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY
WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF
NEOADJUVANT CHEMORADIATION

Table 3. Multivariable stepwise logistic regression analysis

A. Comparison of baseline + imaging models			
Candidate variable subset	AIC	AUC (training dataset)	Selected variables
I. Baseline patient characteristics	67.9	0.81	mrT-stage (mrT1-3b vs. mrT3c-4d), Time to surgery (≤ 10 vs. > 10 weeks)
II. MRI	83.7	0.70	T2W-signal _{entropy} (per unit), ADC _{entropy} (per unit)
III. PET/CT	86.5	0.50	– ^a
IV. Baseline + MRI	58.0	0.88	mrT-stage (mrT1-3b vs. mrT3c-4d) T2W-signal _{entropy} (per unit), T2W _{volume} (per cm ³)
V. Baseline + PET/CT	67.9	0.81	mrT-stage (mrT1-3b vs. mrT3c-4d), Time to surgery (≤ 10 vs. > 10 weeks)
VI. Baseline+ MRI + PET/CT	58.0	0.88	mrT-stage (mrT1-3b vs. mrT3c-4d), T2W-signal _{entropy} (per unit), T2W _{volume} (per cm ³)
B. Optimal prediction model (baseline + MRI model)			
Modality	Selected Variable	Odds Ratio (95% CI)	P-value
Baseline	mrT-stage (mrT1-3b vs. mrT3c-4d)	0.004 (0.00017 - 0.092)	<0.001
MRI	T2W-signal _{entropy} (per unit)	7.810 (1.713 - 35.612)	0.0079
	T2W _{volume} (per cm ³)	1.028 (1.001 - 1.054)	0.0389
AUC (training dataset)	0.88		
AUC (LOOCV)	0.83 (bootstrap 95%CI: 0.70 - 0.96)		

AIC = Akaike Information Criterion, which reflects the relative efficiency of a statistical model compared to other models, with a lower value indicating a more efficient model. AUC: area under the receiver operating characteristic curve. LOOCV = Leave-one-out cross-validation. CI: confidence interval.

^aNo variables were selected as predictors when only PET/CT variables were offered to the model.

Supplementary Table 1 illustrates the results of the univariable analysis (which was performed independently of the multivariable feature selection process) and correlation analysis for all baseline and imaging variables. Since there was a strong correlation between DWI_{volume} and T2W_{volume} ($\rho = 0.96$), SUV_{max} and SUV_{mean} ($\rho = 0.99$) and MTV_{42%} and TLG ($\rho = 0.80$), only T2W_{volume}, SUV_{mean} and TLG were entered in the multivariable selection process described above.

DISCUSSION

This study explores the value of combining quantitative imaging features from baseline, pre-treatment FDG-PET/CT and MRI with common baseline patient characteristics to predict response to neoadjuvant CRT in rectal cancer. Our findings demonstrate that a multivariable model incorporating mrT-stage, combined with (semi-) quantitative MRI features (T2W-signal_{entropy} and tumour volume) can aid in identifying good responders before the start of treatment, with an estimated predictive performance of AUC 0.83. Addition of FDG-PET/CT variables was not beneficial.

Our results indicated mrT-stage as the strongest baseline predictor of response, with a higher mrT-stage resulting in a lower probability of achieving a good response. This is in line with previous studies, including a pooled analysis of >3000 patients that showed that higher T-stage is negatively associated with complete response rates after CRT [1]. More recent large retrospective cohort studies by Joye *et al.* and Al-Sukhni *et al.* confirmed T-stage to be amongst the main baseline predictors of response [37, 38]. In these two previous works, contradictory results were reported for the predictive value of N-stage: while Joye *et al.* reported higher N-stage to be associated with a favourable response, Al-Sukhni reported the opposite. mrN-stage was not identified as a significant predictor in our study. These conflicting findings are likely related to the known inaccuracies of imaging for lymph node staging [39, 40]. Al-Sukhni *et al.* also found a longer interval between CRT and surgery to be associated with a higher probability of response, which is consistent with several other reports [36, 41–45]. For this reason we chose to include time to surgery as a potential confounder in our analyses (although it can clearly not be used as a pre-therapy predictor). While it was indeed associated with response, it was not amongst the strongest parameters ultimately included in the optimal predictive model.

In addition to mrT-stage, only the MRI-based quantitative features significantly contributed to the optimal prediction model. A positive predictive effect was observed for T2W-signal_{entropy}, indicating that tumours with a higher entropy (i.e. a more heterogeneous texture) have a higher probability of achieving a good response. Similarly, a recent prospective study by Shu *et al.* found entropy on pre-CRT T2W MRI to be higher in patients who achieved a complete response after CRT [28]. In contrast, Meng *et al.* found lower pre-treatment T2W entropy to be associated with complete response [46], while a third report by De Cecco *et al.* found no significant differences at all in baseline tumour entropy between response groups [47]. Although

in literature tumour heterogeneity is generally regarded as a factor associated with tumour aggressiveness, the precise relation between heterogeneity (as assessed on imaging) and response to treatment is not well understood. In addition, variations in methodology concerning patient selection, image processing, outcome definition and statistics may have contributed to inconsistent findings between reports. The baseline tumour volume ($T2W_{\text{volume}}$) was the third independent predictor included in the model, though its effect was relatively small. This is in line with data from previous studies that reported suboptimal performance for pre-therapy tumour volumetry to predict response [48–56].

Interestingly, our study showed limited predictive value for baseline PET and DWI variables. This confirms previous evidence showing disappointing or conflicting results for pre-treatment response prediction based on DWI (using mainly ADC) and PET (SUV_{mean} and SUV_{max}) [16, 17]. In a systematic review by Joye *et al.*, sub-optimal pooled predictive performance was reported in the pre-treatment setting for both PET (SUV_{max} pooled sensitivity 0.78; pooled specificity 0.35) and DWI (ADC_{mean} pooled sensitivity 0.69; pooled specificity 0.68) [17]. More positive results for PET or DWI were mainly reported when (sequential) imaging data acquired during and/or after completion of CRT, rather than at baseline was used [16, 17, 19]. To our knowledge, only two other groups have performed a multivariable analysis combining pre-treatment PET/CT and MRI to predict rectal tumour response. Joye *et al.* combined PET/CT and DWI features measured before, during and after CRT, together with volume on T2W-MRI. Their multivariable model reached an AUC of 0.83 to predict a good response (ypT0-1N0). However, only features dependent on post-treatment measurements (post-CRT and ΔCRT) were selected as predictors and no pre-treatment features were included, again indicating the limited value of PET- and DWI in the pre-therapy setting [15]. The second study, by Giannini *et al.* specifically focused on image texture and combined first-order and second-order texture features derived from pre-treatment PET, DWI and T2W-MRI together with PET volume. Their multivariable model reached an AUC of 0.86 in which notably 5 out of 6 selected variables were based on PET. However, this good result was achieved in a test dataset without further (cross-)validation [23]. Validation is required to estimate the performance of a model in actual clinical practice (unseen data), since the accuracy as established in a test dataset will likely be an overestimation. Unfortunately, our current dataset was too small to allow splitting of the data into a test and validation set. Therefore, we chose to simulate validation on an ‘independent’ dataset by performing leave-one-out cross-validation (LOOCV), which resulted in an AUC of 0.83. Apart from the relatively small size, our study is limited

by its retrospective nature. As a consequence, variations in scanning protocols (in particular MRI) and hardware used over time may have introduced heterogeneity not related to the treatment outcome. The study further used a single-reader design for image segmentation, which does not account for inter-observer variations, particularly for the manual (MRI) delineations. These effects are expected to be limited, however, based on previously reported excellent inter-reader reproducibility [48, 49, 51]. Along the same line, some of the baseline characteristics included in the analyses were based on radiological staging (mrT-stage and mrN-stage) which are also known to be subject to interobserver variations. An in-depth analysis of such effects, however, was beyond the scope of the current study. Histopathologic response evaluation was not available for all patients due to the inclusion of W&W patients, for which the surrogate endpoint to establish the treatment outcome was a recurrence-free follow-up of at least 2-years (median 59 months). Since locoregional regrowths indicating incomplete response occur almost exclusively within these first two years, we believe this can be considered an acceptable endpoint in these cases [5]. Future validation and replication of this work may be limited by the fact that PET/CT is typically not routinely performed as a first-line staging modality. Finally, for this study we deliberately chose to explore the predictive value of only a selective number of relatively well-known and reproducible variables (reported to be of potential value in previous literature), to prevent overfitting of a large number of features to a small sample size. As a result, alternative useful predictors may have been neglected. This would be an interesting area for further research in larger datasets (using Radiomics or deep learning approaches) and should also include a more comprehensive integration of imaging features with other clinical, immunological, histological and genetic variables.

CONCLUSION AND CLINICAL OUTLOOK

Prediction of response to neoadjuvant treatment is an increasingly relevant issue in rectal cancer, especially given the growing interest in organ-preserving treatment programs. Our findings demonstrate that a model incorporating (semi-)quantitative imaging features from routine staging MRI combined with mrT-stage can aid in identifying patients likely to show a good response to neoadjuvant chemoradiation. Addition of PET/CT variables was not beneficial, indicating that pre-treatment PET/CT (which is currently not typically used as a first-line modality for rectal cancer staging)

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY
WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF
NEOADJUVANT CHEMORADIATION

probably has a limited added value for pre-therapy response prediction. These results are an encouragement for further development of clinical response prediction models incorporating routine pre-therapy MR imaging in rectal cancer, which will need to be further studied and validated in large prospective patient cohorts.

3

REFERENCES

1. Maas M, Nelemans PJ, Valentini V, et al (2010) Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: A pooled analysis of individual patient data. *Lancet Oncol* 11:835–844.
2. Habr-Gama A, Perez RO, Wynn G, et al (2010) Complete Clinical Response After Neoadjuvant Chemoradiation Therapy for Distal Rectal Cancer: Characterization of Clinical and Endoscopic Findings for Standardization. *Dis Colon Rectum* 53:1692–1698.
3. Paun BC, Cassie S, MacLean AR, et al (2010) Postoperative complications following surgery for rectal cancer. *Ann Surg* 251:807–818.
4. Hendren SK, O'Connor BI, Liu M, et al (2005) Prevalence of male and female sexual dysfunction is high following surgery for rectal cancer. *Ann Surg* 242:212–223.
5. van der Valk MJM, Hilling DE, Bastiaannet E, et al (2018) Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. *Lancet* 391:2537–2545.
6. Maas M, Beets-Tan RGH, Lambregts DMJ, et al (2011) Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 29:4633–4640.
7. Habr-Gama A, Gama-Rodrigues J, São Julião GP, et al (2014) Local Recurrence After Complete Clinical Response and Watch and Wait in Rectal Cancer After Neoadjuvant Chemoradiation: Impact of Salvage Therapy on Local Disease Control. *Int J Radiat Oncol* 88:822–828.
8. Smith JD, Ruby JA, Goodman KA, et al (2012) Nonoperative management of rectal cancer with complete clinical response after neoadjuvant therapy. *Ann Surg* 256:965–972.
9. Appelt AL, Pløen J, Harling H, et al (2015) High-dose chemoradiotherapy and watchful waiting for distal rectal cancer: a prospective observational study. *Lancet Oncol* 16:919–927.
10. Rombouts AJM, Al-Najami I, Abbott NL, et al (2017) Can we Save the rectum by watchful waiting or TransAnal microsurgery following (chemo) Radiotherapy versus Total mesorectal excision for early REctal Cancer (STAR-TREC study)?: protocol for a multicentre, randomised feasibility study. *BMJ Open* 7:e019474.
11. Van Stiphout RGPM, Valentini V, Buijsen J, et al (2014) Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: A multicentric prospective study with external validation. *Radiother Oncol* 113:215–222.

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY
WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF
NEOADJUVANT CHEMORADIATION

12. Janssen MHM, Öllers MC, Van Stiphout RGPM, et al (2012) PET-based treatment response evaluation in rectal cancer: Prediction and validation. *Int J Radiat Oncol Biol Phys* 82:871–876.
13. Maffione AM, Marzola MC, Capirci C, et al (2015) Value of 18 F-FDG PET for Predicting Response to Neoadjuvant Therapy in Rectal Cancer: Systematic Review and Meta-Analysis. *Am J Roentgenol* 204:1261–1268.
14. Cliffe H, Patel C, Prestwich R, Scarsbrook A (2017) Radiotherapy response evaluation using FDG PET-CT—established and emerging applications. *Br J Radiol* 90:20160764.
15. Joye I, Debuquoy A, Deroose CM, et al (2017) Quantitative imaging outperforms molecular markers when predicting response to chemoradiotherapy for rectal cancer. *Radiother Oncol* 124:104–109.
16. Schurink NW, Lambregts DM, Beets-Tan RG (2019) Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. *Br J Radiol* 92:20180655.
17. Joye I, Deroose CM, Vandecaveye V, Haustermans K (2014) The role of diffusion-weighted MRI and 18F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: A systematic review. *Radiother Oncol* 113:158–165.
18. Mahadevan LS, Zhong J, Venkatesulu BP, et al (2018) Imaging predictors of treatment outcomes in rectal cancer: An overview. *Crit. Rev. Oncol. Hematol.* 129:153–162.
19. Meng X, Huang Z, Wang R, Yu J (2014) Prediction of response to preoperative chemoradiotherapy in patients with locally advanced rectal cancer. *Biosci Trends* 8:11–23.
20. Meng X, Li H, Kong L, et al (2016) MRI In rectal cancer: Correlations between MRI features and molecular markers Ki-67, HIF-1 α , and VEGF. *J Magn Reson Imaging* 44:594–600.
21. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2012) Diffusion-weighted MRI in rectal cancer: Apparent diffusion coefficient as a potential noninvasive marker of tumor aggressiveness. *J Magn Reson Imaging* 35:1365–1371.
22. Surov A, Meyer HJ, Wienke A (2017) Associations between apparent diffusion coefficient (ADC) and Ki 67 in different tumors: a meta-analysis. Part 1: ADC_{mean}. *Oncotarget* 8:75434–75444.
23. Giannini V, Mazzetti S, Bertotto I, et al (2019) Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18 F-FDG PET and MRI radiomics features. *Eur J Nucl Med Mol Imaging* 46:878–888.
24. Van Griethuysen JJM, Fedorov A, Parmar C, et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107.

25. Erdi YE, Mawlawi O, Larson SM, et al (1997) Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* 80:2505–2509.
26. Miccò M, Vargas HA, Burger IA, et al (2014) Combined pre-treatment MRI and 18F-FDG PET/CT parameters as prognostic biomarkers in patients with cervical cancer. *Eur J Radiol* 83:1169–1176.
27. Ueno Y, Lisbona R, Tamada T, et al (2017) Comparison of FDG PET metabolic tumour volume versus ADC histogram: Prognostic value of tumour treatment response and survival in patients with locally advanced uterine cervical cancer. *Br J Radiol* 90:20170035.
28. Shu Z, Fang S, Ye Q, et al (2019) Prediction of efficacy of neoadjuvant chemoradiotherapy for rectal cancer: the value of texture analysis of magnetic resonance images. *Abdom Radiol* 21:1051–1058.
29. Greenbaum A, Martin DR, Bocklage T, et al (2019) Tumor Heterogeneity as a Predictor of Response to Neoadjuvant Chemotherapy in Locally Advanced Rectal Cancer. *Clin Colorectal Cancer* 18:102–109.
30. Bozkaya Y, Özdemir NY, Erdem GU, et al (2018) Clinical predictive factors associated with pathologic complete response in locally advanced rectal cancer. *J Oncol Sci* 4:5–10.
31. Deantonio L, Caroli A, Puta E, et al (2018) Does baseline [18F] FDG-PET/CT correlate with tumor staging, response after neoadjuvant chemoradiotherapy, and prognosis in patients with rectal cancer? *Radiat Oncol* 13:211.
32. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol* 102:1143–1158.
33. Mandard A-M, Dalibard F, Mandard J-C, et al (1994) Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 73:2680–2686.
34. Akaike H (1974) A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr* 19:716–723.
35. Burnham KP, Anderson DR (2004) Multimodel Inference. *Sociol Methods Res* 33:261–304.
36. Akgun E, Caliskan C, Bozbiyik O, et al (2018) Randomized clinical trial of short or long interval between neoadjuvant chemoradiotherapy and surgery for rectal cancer. *Br J Surg* 105:1417–1425.
37. Joye I, Debucquoy A, Fieuids S, et al (2016) Can clinical factors be used as a selection tool for an organ-preserving strategy in rectal cancer? *Acta Oncol (Madr)* 55:1047–1052.

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY
WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF
NEOADJUVANT CHEMORADIATION

38. Al-Sukhni E, Attwood K, Mattson DM, et al (2016) Predictors of Pathologic Complete Response Following Neoadjuvant Chemoradiotherapy for Rectal Cancer. *Ann Surg Oncol* 23:1177–1186.
39. Lahaye MJ, Engelen SME, Nelemans PJ, et al (2005) 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 MRI* 26:259–268.
40. Gröne J, Loch FN, Taupitz M, et al (2018) Accuracy of Various Lymph Node Staging Criteria in Rectal Cancer with Magnetic Resonance Imaging. *J Gastrointest Surg* 22:146–153.
41. Francois Y, Nemoz CJ, Baulieux J, et al (1999) Influence of the Interval Between Preoperative Radiation Therapy and Surgery on Downstaging and on the Rate of Sphincter-Sparing Surgery for Rectal Cancer: The Lyon R90-01 Randomized Trial. *J Clin Oncol* 17:2396–2396.
42. Kalady MF, de Campos-Lobato LF, Stocchi L, et al (2009) Predictive Factors of Pathologic Complete Response After Neoadjuvant Chemoradiation for Rectal Cancer. *Trans . Meet Am Surg Assoc* 127:213–220.
43. Foster JD, Jones EL, Falk S, et al (2013) Timing of Surgery After Long-Course Neoadjuvant Chemoradiotherapy for Rectal Cancer. *Dis Colon Rectum* 56:921–930.
44. Probst CP, Becerra AZ, Aquina CT, et al (2015) Extended Intervals after Neoadjuvant Therapy in Locally Advanced Rectal Cancer: The Key to Improved Tumor Response and Potential Organ Preservation. *J Am Coll Surg* 221:430–440.
45. Petrelli F, Sgroi G, Sarti E, Barni S (2016) Increasing the Interval Between Neoadjuvant Chemoradiotherapy and Surgery in Rectal Cancer. *Ann Surg* 263:458–464.
46. Meng Y, Zhang C, Zou S, et al (2018) MRI texture analysis in predicting treatment response to neoadjuvant chemoradiotherapy in rectal cancer. *Oncotarget* 9:11999-12008.
47. Cecco CN De, Rengo M, Meinel FG, et al (2015) Texture Analysis as Imaging Biomarker of Tumoral Response to Neoadjuvant Chemoradiotherapy in Rectal Cancer Patients Studied with 3-T Magnetic Resonance. *Invest Radiol* 50:239–245.
48. Martens MH, Van Heeswijk MM, Van Den Broek JJ, et al (2015) 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* 93:1005–1014.
49. Lambregts DMJ, Rao S-X, Sassen S, et al (2015) MRI and Diffusion-weighted MRI Volumetry for Identification of Complete Tumor Responders After Preoperative Chemoradiotherapy in Patients With Rectal Cancer. *Ann Surg* 262:1034–1039.

50. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2011) Rectal Cancer: Assessment of Complete Response to Preoperative Combined Radiation Therapy with Chemotherapy—Conventional MR Volumetry versus Diffusion-weighted MR Imaging. *Radiology* 260:734–743.
51. Quaia E, Gennari AG, Ricciardi MC, et al (2016) Value of percent change in tumoral volume measured at T_2 -weighted and diffusion-weighted MRI to identify responders after neoadjuvant chemoradiation therapy in patients with locally advanced rectal carcinoma. *J Magn Reson Imaging* 44:1415–1424.
52. Ha HI, Kim AY, Yu CS, et al (2013) Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after preoperative chemoradiation therapy. *Eur Radiol* 23:3345–3353.
53. Young HK, Dae YK, Tae HK, et al (2005) 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* 62:761–768.
54. Okuno T, Kawai K, Koyama K, et al (2018) Value of FDG–PET/CT Volumetry After Chemoradiotherapy in Rectal Cancer. *Dis Colon Rectum* 61:320–327.
55. Dos Anjos DA, Perez RO, Habr-Gama A, et al (2016) Semiquantitative Volumetry by Sequential PET/CT May Improve Prediction of Complete Response to Neoadjuvant Chemoradiation in Patients with Distal Rectal Cancer. *Dis Colon Rectum* 59:805–812.
56. Park J, Chang KJ, Seo YS, et al (2014) Tumor SUVmax Normalized to Liver Uptake on 18 F-FDG PET/CT Predicts the Pathologic Complete Response After Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Nucl Med Mol Imaging* (2010) 48:295–302.

VALUE OF COMBINED MULTIPARAMETRIC MRI AND FDG-PET/CT TO IDENTIFY
WELL-RESPONDING RECTAL CANCER PATIENTS BEFORE THE START OF
NEOADJUVANT CHEMORADIATION

3



Studying local tumour heterogeneity on MRI and FDG-PET/CT to predict response to neoadjuvant chemoradiotherapy in rectal cancer

Niels W. Schurink, Simon R. van Kranen, Maaïke Berbee, Wouter van Elmpt, Frans C.H. Bakers, Sander Roberti, Joost J.M. van Griethuysen, Lisa A. Min, Max J. Lahaye, Monique Maas, Geerard L. Beets, Regina G.H. Beets-Tan, Doenja M.J. Lambregts

Eur Radiol. 2021; 31(9):7031-7038. doi: 10.1007/s00330-021-07724-0

ABSTRACT

OBJECTIVE

To investigate whether quantifying local tumour heterogeneity has added benefit compared to global tumour features to predict response to chemoradiotherapy using pre-treatment multiparametric PET and MRI data

METHODS

Sixty-one locally advanced rectal cancer patients treated with chemoradiotherapy and staged at baseline with MRI and FDG-PET/CT were retrospectively analyzed. Whole-tumour volumes were segmented on the MRI and PET/CT scans from which global tumour features ($T2W_{\text{volume}}/T2W_{\text{entropy}}/ADC_{\text{mean}}/SUV_{\text{mean}}/TLG/CT_{\text{mean-HU}}$) and local texture features (histogram features derived from local entropy/mean/standard deviation maps) were calculated. These respective feature sets were combined with clinical baseline parameters (e.g. age/gender/TN-stage) to build multivariable prediction models to predict a good (Mandard TRG1-2) versus poor (Mandard TRG3-5) response to chemoradiotherapy. Leave-one-out-cross validation (LOOCV) with bootstrapping was performed to estimate performance in an 'independent' dataset.

RESULTS

When using only imaging features, local texture features showed an AUC=0.81 versus AUC=0.74 for global tumour features. After internal cross-validation (LOOCV), AUC to predict a good response was highest for the combination of clinical baseline variables + global tumour features (AUC=0.83), compared to AUC=0.79 for baseline + local texture and AUC=0.76 for all combined (baseline + global + local texture).

CONCLUSION

In imaging-based prediction models, local texture analysis has potential added value compared to global tumour features to predict response. However, when combined with clinical baseline parameters such as cTN-stage, the added value of local texture analysis appears to be limited. Overall performance to predict response when combining baseline variables with quantitative imaging parameters is promising and warrants further research.

INTRODUCTION

In the era of organ preservation, the assessment of response of rectal cancer to neoadjuvant chemoradiotherapy (CRT) has become an increasingly important issue. Early- and pre-treatment prediction of response is gaining interest as this could allow further optimization of treatment based on the anticipated treatment response. The ultimate aim is to further increase response rates and avoid ineffective and potential harmful treatments in those who are unlikely to benefit.

Several studies have investigated the potential of imaging to predict response, with promising – albeit somewhat inconsistent – results for quantitative imaging biomarkers derived directly from either MRI (including functional sequences like diffusion and perfusion MRI) [1–5] or 2-[¹⁸F]-fluoro-2-deoxy-D-glucose-PET (FDG-PET) [1, 4, 6]. So far, there are only few reports that have combined different imaging modalities to build multiparametric response prediction models. These studies have varying success rates (AUCs ranging from 0.51 – 0.94), which is likely related to the large methodological differences between studies and variations in outcome parameters used [7–11].

In addition to these reports, several groups have focused on more advanced parameters derived from image post processing, in specific parameters related to image texture as a measure of tumour heterogeneity [12–21]. This makes sense from a histopathological perspective, as tumours are typically not homogeneous but show spatial variations in cellular microarchitecture, necrosis, vascularization and gene expression [22]. Increased intra-tumoural heterogeneity has often been suggested as a potential prognostic factor in oncology as it has been related to the emergence of resistant subpopulations of cells that drive resistance to treatment [12, 23, 24]. A similar relation has been reported for rectal cancer, as several pathology reports have shown an association between tumour heterogeneity and response to neoadjuvant treatment [25, 26].

Image texture analysis offers opportunities to non-invasively study tumour heterogeneity, although quantifying heterogeneity at a microstructural level using medical imaging is challenging due to the large difference in scale between MRI and histology. Hence, texture analysis in medical imaging is most often applied to the tumour as a whole, rendering only global measures of heterogeneity such as mean tumour entropy and uniformity. These measures do not specifically take into account the local variations within the tumour (local heterogeneity), as a result of which potentially important predictive information may be overlooked or averaged out [12].

4

The goal of this study is to investigate whether quantifying local tumour heterogeneity has added benefit compared to global tumour features to predict response to CRT using pre-treatment multiparametric PET and MRI data.

MATERIALS AND METHODS

This study was approved by the local institutional review board and informed consent was waived due to the retrospective nature of this study.

Patients and outcome definition

The study cohort consisted of 61 locally advanced (\geq T3 and/or N+) rectal cancer patients, derived from a previously published single institute study cohort [13]. All patients were treated with neoadjuvant chemoradiotherapy (50.4Gy with concurrent capecitabine-based chemotherapy), underwent pre-treatment MRI+DWI and FDG-PET/CT, and sufficient information was available to establish the final treatment response (histopathology after surgery or >2 years clinical follow-up to confirm a sustained clinical complete response in patients undergoing watch-and-wait (W&W)). Final response was documented according to the 5-point histopathological tumour regression grade (TRG) by Mandard [27]. Patients with TRG1-2 were categorized as 'good responders'. Patients with TRG 3-5 were categorized as 'poor responders'. Watch-and-wait patients ($n=7$) with a sustained clinical complete response for >2 years were considered TRG1 for the purpose of this study.

Imaging and image segmentation

MRI was acquired at 1.5T (Intera Achieva $n=43$ or Ingenia $n=18$, Philips Healthcare). Imaging and image segmentation were performed according to previously published protocols [13] provided in detail in **Table 1**. In short, the protocol consisted of T2-weighted sequences with 3-5 mm slice thickness in 3 planes, and an axial echo planar imaging (EPI) DWI sequence with b-values 0-1000 s/mm^2 and slice thickness 5 mm, angled in the same plane as the axial T2W-MRI. No bowel preparation or spasmolytics were given. Apparent diffusion coefficient (ADC) maps were derived from the DWI images by fitting the data to a mono-exponential decay function. 18F-FDG-PET/CTs (3 mm slice thickness) were acquired on a Siemens Biograph 40 TruePoint PET/CT scanner (SIEMENS Healthineers AG) 60 minutes after an intravenous bolus of 18F-FDG with an activity of either 2.5 MBq/kg ($n=52$) or 4.0 MBq/kg ($n=9$) after 6 hours fasting and with a full bladder. Attenuation correction was performed using a non-enhanced

STUDYING LOCAL TUMOUR HETEROGENEITY ON MRI AND FDG-PET/CT TO
 PREDICT RESPONSE TO NEOADJUVANT CHEMORADIOTHERAPY
 IN RECTAL CANCER.

CT scan with automatic dose modulation. Standardized uptake values (SUV) were calculated according to the recommendations of the Quantitative Imaging Biomarker Alliance [28].

For each patient volumes of interest (VOI) covering the whole tumour volume were segmented by a board-certified expert radiologist on the T2W and b1000-DWI MRI. PET images were segmented using semi-automated segmentation (using a threshold of 42% of the maximum SUV) [13, 29–31]. These segmentations were then transferred to the corresponding ADC-maps and unenhanced CT images.

Table 1: MRI protocol

	T2-weighted	Diffusion-weighted
Echo time (ms)	130-150	65.74-84.88
Repetition time (ms)	3427-16738	2480-5545
Echo train length	25-28	53-87
Slice thickness (mm)	3-5 ^a	5
Slice gap (mm)	3.3-7.03	4-6.02
In-plane resolution (mm)	0.78125	1.25-1.71875
Number of averages	2-6	3-10
b-values (s/mm ²)	-	0, 1000 ^b
Fat-suppression	-	STIR (n=32), SPIR (n=7), SPAIR (n=22)

STIR: Short-TI Inversion Recovery; SPIR: Spectral Presaturation with Inversion Recovery; SPAIR: Spectral Attenuated Inversion Recovery.

^a n= 23 patients were scanned with 5 mm and n=38 with 3 mm axial slice thickness

^b Protocols included 3-7 b-values ranging from b0 to b2000 s/mm², but for the purpose of this study only the b=0 and b=1000 s/mm² series were used for analyses and to calculate the ADC map.

Reprint of Table 1 [13]

4

Quantitative modelling of local texture

Local texture features were extracted from the segmented tumour VOIs of each imaging modality/sequence using the following steps:

1. Normalization: because T2W-MRI and b1000-DWI images are expressed in arbitrary units that can differ between study visits, protocol settings and subjects, these images were normalized (mean pixel intensity=300 and standard deviation=100). ADC, SUV and CT were processed using their original units.
2. Local texture maps: For each imaging type 3 local texture maps (entropy, mean and standard deviation) were derived from the VOIs using PyRadiomics [32] with a neighborhood of 5x5x5 pixels. Bin widths of 20, 20, 0.05, 0.5 and 50 were used for T2W-MRI, b1000-DWI, ADC, SUV and CT respectively. A visualization of the extracted local entropy is depicted in **Figure 1**.
3. Histogram metrics: From each local texture map histogram features (10th, 25th, 50th, 75th, 90th percentile, mean and standard deviation) were derived to describe the distribution of local texture features within the tumour.

This resulted in 21 features per image type/sequence, i.e. 63 MRI features (T2W-MRI, b1000-DWI and ADC) and 42 PET/CT features (SUV and CT). In addition, clinical baseline parameters (mrT-stage, mrN-stage, age, gender and interval between last

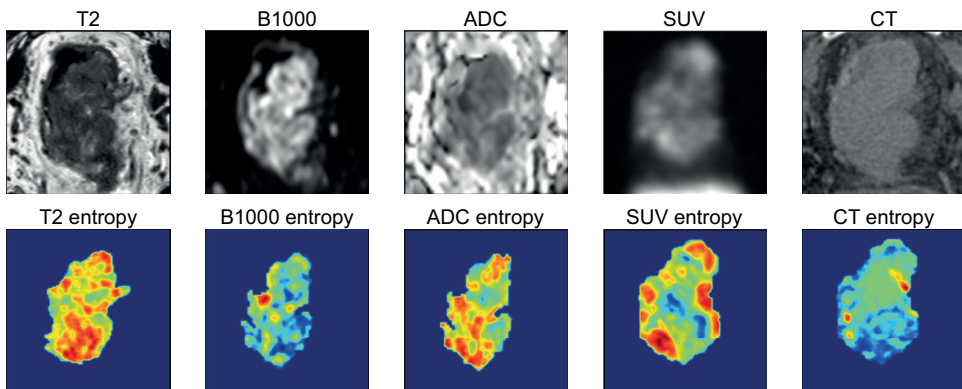


Figure 1. Example of local entropy heatmaps in a 66 year old male patient with a large (clinically T4aN1) rectal tumour. Red areas correspond to a high local entropy whereas blue areas represent areas of low local entropy.

radiotherapy fraction and final response evaluation), and global tumour measures (tumour volume, mean entropy, mean tumour ADC, mean SUV, total lesion glycolysis (TLG) and mean Hounsfield units) were collected, that were available from a previously published study in the same patient cohort [13].

Statistical analysis:

The analysis workflow comprised 4 main steps. First, when two or more features showed a strong correlation (Pearson's $\rho > 0.8$) only the feature with the lowest mean absolute correlation was retained for further analysis to reduce effects of multicollinearity. Second, multivariable logistic regression models were trained with the remaining feature set using forward stepwise feature selection. The modelling process was repeated separately for (1) global imaging features only (including separate sub-analyses using only MRI or PET/CT data respectively), (2) local texture features only (including MRI-only and PET/CT only sub-analyses), (3) global imaging features + clinical baseline parameters, (4) local texture features + baseline parameters, and (5) global + local texture features + clinical baseline parameters combined. Third, to prevent effects of overfitting, the maximum numbers of selected features for each model was set to 3 (i.e. 1 per 10 patients in the smallest outcome group). Finally, the performance of the different models was compared using Akaike's Information Criterion (AIC), as a measure of the 'goodness-of-fit' of the model taking into account model complexity [33, 34], and by calculating the area under the receiver operator curve (AUC) to predict a good (TRG1-2) versus poor (TRG3-5) response. Confidence intervals for the training AUC were calculated using the DeLong method [35]. For the models combining clinical baseline parameters with imaging features an estimate of the model's performance on unseen data was obtained by using bootstrapped leave-one-out cross validation (LOOCV AUC; 5000 bootstrap samples to calculate LOOCV AUC confidence intervals), in line with methods previously reported [13].

4

RESULTS

Patient characteristics

Of the 61 study patients, 47 were male, median age was 69 years (range 46-88). In total 13 patients were TRG1 (6 after surgery (10%) and 7 undergoing W&W (11%)), 18 TRG2 (30%), 19 TRG3 (31%), 11 TRG4 (18%) and 0 TRG5 (0%), resulting in a total of 31 good responders (TRG1-2) and 30 poor responders (TRG3-5).

Table 2. Multivariable logistic regression analysis

Imaging only - Local texture versus global tumour models			
Model	AIC	AUC (training)	Selected features
Local texture	68	0.81 (0.70 – 0.91)	CT_mean_75perc, ADC_entropy_10perc, T2_standarddeviation_mean
Local texture MRI-only	76	0.73 (0.60 – 0.85)	ADC_entropy_10perc
Local texture PET/CT-only	74	0.78 (0.66 – 0.90)	CT_mean_75perc, CT_standarddeviation_10perc
Global tumour	82	0.74 (0.61 – 0.87)	T2_entropy, ADC_entropy, SUV_mean
Global tumour MRI-only	83	0.70 (0.57 – 0.84)	T2entropy, ADC_entropy
Global tumour PET/CT-only	87	0.50 (0.50 – 0.50)	-
Combined models including clinical baseline variables			
Model	AIC	AUC (training)	AUC (LOOCV)
Clinical baseline + local texture	60	0.87 (0.78 – 0.96)	0.79 (0.46 – 0.90)
Clinical baseline + global tumour	58	0.88 (0.80 – 0.97)	0.83 (0.70 – 0.96)
Clinical baseline + global tumour + local texture	56	0.89 (0.81 – 0.98)	0.76 (0.57 – 0.90)

AIC Akaike Information Criterion, a measure for 'goodness of fit' with lower values indicating a better fit. AUC area under the receiver operator curve. LOOCV leave-one-out cross-validation.

NB: 95% confidence intervals (CI) are given in parentheses. CIs for the training AUC were determined using the DeLong method. The 95% CIs of the LOOCV AUC were determined using 500 bootstrap samples.

Quantitative modelling of local texture

A correlation matrix for the total feature set and exclusion of features with Pearson's $\rho > 0.8$ is provided in **Supplementary Table 1**. Results of the multivariable regression models are summarized in **Table 2**.

When looking at imaging features only, local texture features performed better than global tumour features based on the AIC (AIC=68 vs. AIC=82), resulting in a training AUC of 0.81 (95%CI 0.70 – 0.91) for the local texture features and AUC of 0.74 (95%CI 0.61 – 0.87) for the global tumour features. Similarly, in the MRI-only or PET/CT only analyses, local texture features performed better (AUC 0.73 for MRI and 0.78 for

PET/CT) than global features (AUC 0.70 and 0.50 respectively), though for MRI the difference was small. When imaging features were combined with baseline clinical parameters (e.g. T- and N-stage), performance for local versus global texture features was similar (AIC=60 vs AIC=58; AUC=0.87 vs AUC=0.88). Adding local texture features to the combination of baseline + global tumour features did not result in improved performance (AIC=56 and AUC=0.89). After leave one out cross-validation, the model incorporating clinical baseline + global tumour features achieved the highest AUC of 0.83 (95% CI 0.70-0.96) to predict a good response, though differences with the other models were small (Table 2).

DISCUSSION

The aim of this study was to investigate whether there is added value in quantifying local tumour texture (heterogeneity) compared to global tumour features on baseline MRI and 18-F-FDG-PET/CT to predict response to neoadjuvant treatment in rectal cancer. Results indicate that, when including only imaging-parameters, local texture features show a better performance compared to global tumour features to predict a good response with AUCs of up to 0.81 (versus 0.74 for global features) when tested within our own study dataset. While global tumour PET/CT features did not show any predictive performance (AUC 0.50, no significant features), local texture PET/CT features did perform substantially better with an AUC of 0.78. However, the selected local PET/CT features were all derived from CT imaging indicating a limited role for PET. When local texture features were combined with clinical baseline parameters (such as the cT-and cN-stage) addition of local texture features did not improve the predictive performance of the model compared to addition of global tumour features, neither on the training data nor after internal cross-validation, indicating that from a clinical point of view local texture features may be of limited added benefit. Performance for the combination of clinical baseline parameters and global tumour features was encouraging, with an AUC of 0.83 to predict a good response after internal cross-validation.

In rectal cancer, there are multiple previous reports correlating image texture features with response to neoadjuvant treatment [7, 13–19, 21, 36–48]. Some of these studies focused on relatively simple first and second order features derived from histogram analysis and gray-level co-occurrence matrices (GLCM) [7, 13, 41, 42, 46], while others assessed advanced features by applying more sophisticated Radiomics modelling

[14–19, 21, 38–40, 43–45, 47–50]. Overall, studies incorporating higher order features into their analysis tended to achieve a higher performance to predict response (AUCs of 0.69-0.97) versus those using only simpler first and second order features (AUCs 0.51-0.89). Although this difference in performance can be attributed to many factors, it might be a hint supporting the hypothesis of our current study that assessing spatial (local) variations within the tumour may be beneficial, as higher order features tend to take the local texture throughout the tumour into account (for example, often used radiomics features such as gray level size zone matrix features summarize the size and number of local homogeneous patches throughout the tumour). In our current image-based analysis local texture features indeed showed a better predictive performance compared to global tumour measures (AUC 0.81 vs. 0.74).

Our analysis suggests that among the local texture features, those derived from the ADC-map, CT and T2W-MRI provide the best predictive value as these were amongst the selected features in the forward selection process. With respect to CT, there have only been a limited number of reports focusing on rectal tumour response prediction using CT texture analysis [43, 45, 48, 50], which is probably related to the fact that CT is not routinely used for the local staging of rectal cancer. Bibault et al. derived 1683 texture features from planning CTs of which 28 features (first order histogram, and 2D and 3D second order gray-level co-occurrence matrix features) were significantly correlated with pathological complete response and were combined with clinical T-stage to train a deep neural network. The performance of the resulting network (AUC 0.72) was compared to classical methods such as linear regression analysis (AUC 0.59) and support vector machines (AUC 0.62) [48]. However, since the selected features were not specified by the authors it is not possible to directly compare these findings to our current results. In a second study Hamerla et al. tried to reproduce the findings of Bibault, but with limited success (balanced accuracy score of 0.50) [43]. In another study by Chee et al. good responding patients showed significantly lower entropy, higher uniformity and lower standard deviation on baseline CT images [50]. As a critical note, our results with respect to CT should perhaps be interpreted with some caution, because of the low-dose quality of our CT scans (that were merely acquired for PET attenuation correction), which are inherently more heterogeneous due to increased noise levels. Therefore, these findings should still be confirmed in future studies using diagnostic quality CTs. Based on our findings, the contributing role of PET appears to be limited. From a clinical point of view, the impact of this negative result is probably small, as PET is generally not routinely performed as part of the diagnostic workup for the local staging of rectal cancer.

Similar to our current findings, promising results have been previously reported for DWI texture features, in particular those derived from ADC maps, to predict response to CRT in rectal cancer. Reported AUCs when using these features in addition to other MRI features range between 0.69-0.97 [15, 17, 19, 36]. Most of these studies combined features derived from DWI/ADC with other modalities such as T1-weighted, T2-weighted and/or DCE-MRI and showed better performance when incorporating these features in a multiparametric model compared to models using features derived from only a single modality [17, 19, 36]. In contrast, there are also some reports in which ADC texture features were not included amongst the selected predictors when entered into a multivariate model together with other imaging features [7, 13, 18, 21]. For T2-weighted features (either alone or in combination with clinical variables) AUCs ranging between 0.71-0.93 have previously been reported to predict response, though methods of image analysis, image post-processing and statistical analysis differ substantially between reports [14, 16, 39, 46, 47, 49]. In addition to rectal cancer, there have been numerous reports focusing on image-based prediction models of tumour response and outcome in other target organs such as head and neck [51], breast [52], liver [53], lungs [54], and bladder and prostate [55]. Reported model performance and image modalities used vary widely between reports and there are major methodological differences between the various studies, making it difficult to draw any general conclusions yet on the value of such models in oncologic settings.

There are some limitations to our study design. Due to the retrospective design and small study cohort, only a limited number of features (max 3) were used in the multivariable analysis to prevent overfitting, which limited our analyses to simpler models. We assessed local texture for each modality individually (using histogram metrics), which does not take into account the local spatial correlation between different imaging modalities. Such analysis would require that cross-modality images are accurately registered on a voxel-basis, which is a challenge on its own as image registration can also introduce new uncertainties [56]. Finally, although all images were resampled to a common pixel spacing, T2W images were in part acquired with 5mm instead of the standard 3mm slice thickness, which may have had an influence on the extracted feature values. Since the majority of features included in our final prediction models were derived from other modalities than T2W MRI, we believe this potential effect will likely be small.

In our current study we correlated local texture patterns on imaging to one "global" measure of the final tumour response (TRG). For future research it would be interesting

to gain more in depth insights into how local therapeutic effects as assessed on imaging correlate to underlying response effects on corresponding histopathology. In addition, recent technical advances such as the introduction of integrated MR-linear accelerator systems into the clinical setting create new opportunities for day to day monitoring of global and local tumour changes during treatment [57]. Such developments can significantly increase our understanding of how local texture features can be related to local treatment outcome which might provide a basis for further personalized treatment (e.g. targeted boosting strategies) in the future.

In conclusion, this study has shown that it is possible to quantify local tumour heterogeneity using local texture maps derived from baseline MRI and FDG-PET/CT imaging. When entered in multivariate prediction models using only imaging data, quantification of local texture features offered a better performance to predict response compared to “simple” global tumour measures, with features derived from MRI and CT being the main predictors and no clear role for PET. When combined with clinical baseline parameters (such as T- and N-stage) the added value of local texture versus global tumour analysis was limited, indicating that from a clinical perspective the added benefit of such more advanced image analysis will likely be small. Predictive performance of our optimal model – combining clinical baseline variables with global imaging features – was encouraging, warranting further research in this direction on a larger scale.

REFERENCES

1. Joye I, Deroose CM, Vandecaveye V, Haustermans K (2014) The role of diffusion-weighted MRI and 18F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: A systematic review. *Radiother Oncol* 113:158–165
2. Pham TT, Liney GP, Wong K, Barton MB (2017) Functional MRI for quantitative treatment response prediction in locally advanced rectal cancer. *Br J Radiol* 90:20151078
3. Hötker AM, Garcia-Aguilar J, Gollub MJ (2014) Multiparametric MRI of Rectal Cancer in the Assessment of Response to Therapy. *Dis Colon Rectum* 57:790–799
4. Ryan JE, Warrier SK, Lynch AC, Heriot AG (2015) Assessing pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. *Color Dis* 17:849–861
5. Schurink NW, Lambregts DMJ, Beets-Tan RGH (2019) Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. *Br J Radiol* 92:20180655
6. Maffione AM, Chondrogiannis S, Capirci C, et al (2014) Early prediction of response by 18F-FDG PET/CT during preoperative therapy in locally advanced rectal cancer: A systematic review. *Eur J Surg Oncol* 40:1186–1194
7. Nie K, Shi L, Chen Q, et al (2016) Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. *Clin Cancer Res* 22:5256–5264
8. Lambrecht M, Deroose C, Roels S, et al (2010) 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 (Madr)* 49:956–963
9. Joye I, Debucquoy A, Deroose CM, et al (2017) Quantitative imaging outperforms molecular markers when predicting response to chemoradiotherapy for rectal cancer. *Radiother Oncol* 124:104–109
10. Ippolito D, Fior D, Trattenero C, et al (2015) Combined value of apparent diffusion coefficient-standardized uptake value max in evaluation of post-treated locally advanced rectal cancer. *World J Radiol* 7:509
11. Intven M, Monninkhof EM, Reerink O, Philippens MEP (2015) Combined T2w volumetry, DW-MRI and DCE-MRI for response assessment after neo-adjuvant chemoradiation in locally advanced rectal cancer. *Acta Oncol (Madr)* 54:1729–1736

12. O'Connor JPB, Rose CJ, Waterton JC, et al (2015) Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. *Clin Cancer Res* 21:249–257
13. Schurink NW, Min LA, Berbee M, et al (2020) Value of combined multiparametric MRI and FDG-PET/CT to identify well-responding rectal cancer patients before the start of neoadjuvant chemoradiation. *Eur Radiol* 30:2945–2954
14. Shayesteh SP, Alikhassi A, Fard Esfahani A, et al (2019) Neo-adjuvant chemoradiotherapy response prediction using MRI based ensemble learning method in rectal cancer patients. *Phys Medica* 62:111–119
15. van Griethuysen JJM, Lambregts DMJ, Trebeschi S, et al (2020) Radiomics performs comparable to morphologic assessment by expert radiologists for prediction of response to neoadjuvant chemoradiotherapy on baseline staging MRI in rectal cancer. *Abdom Radiol* 45:632–643
16. Cusumano D, Dinapoli N, Boldrini L, et al (2018) Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. *Radiol Med* 123:286–295
17. Zhou X, Yi Y, Liu Z, et al (2019) Radiomics-Based Pretherapeutic Prediction of Non-response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer. *Ann Surg Oncol* 26:1676–1684
18. Giannini V, Mazzetti S, Bertotto I, et al (2019) Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18F-FDG PET and MRI radiomics features. *Eur J Nucl Med Mol Imaging* 46:878–888
19. Cui Y, Yang X, Shi Z, et al (2019) Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 29:1211–1220
20. Nie K, Shi L, Chen Q, et al (2016) Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. *Clin Cancer Res* 22:5256–5264
21. Metser U, Jhaveri KS, Murphy G, et al (2015) Multiparametric PET-MR Assessment of Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer: PET, MR, PET-MR and Tumor Texture Analysis: A Pilot Study. *Adv Mol Imaging* 05:49–60
22. Molinari C, Marisi G, Passardi A, et al (2018) Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *Int J Mol Sci* 19:3733
23. Ramón y Cajal S, Sesé M, Capdevila C, et al (2020) Clinical implications of intratumor heterogeneity: challenges and opportunities. *J Mol Med* 98:161–177

STUDYING LOCAL TUMOUR HETEROGENEITY ON MRI AND FDG-PET/CT TO
PREDICT RESPONSE TO NEOADJUVANT CHEMORADIOTHERAPY
IN RECTAL CANCER.

24. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C (2015) Translational Implications of Tumor Heterogeneity. *Clin Cancer Res* 21:1258–1266
25. Greenbaum A, Martin DR, Bocklage T, et al (2019) Tumor Heterogeneity as a Predictor of Response to Neoadjuvant Chemotherapy in Locally Advanced Rectal Cancer. *Clin Colorectal Cancer* 18:102–109
26. Diaz Jr LA, Williams RT, Wu J, et al (2012) The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486:537–540
27. Mandard A-M, Dalibard F, Mandard J-C, et al (1994) Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 73:2680–2686
28. QIBA (2018) SUV vendorneutral pseudocode 20180626 DAC. Quantitative Image Biomarker Initiative, Oak Brook. Available via https://qibawiki.rsna.org/images/8/86/SUV_vendorneutral_pseudocode_20180626_DAC.pdf. Accessed 11 May 2020
29. Erdi YE, Mawlawi O, Larson SM, et al (1997) Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. *Cancer* 80:2505–9
30. Miccò M, Vargas HA, Burger IA, et al (2014) Combined pre-treatment MRI and 18F-FDG PET/CT parameters as prognostic biomarkers in patients with cervical cancer. *Eur J Radiol* 83:1169–1176
31. Ueno Y, Lisbona R, Tamada T, et al (2017) Comparison of FDG PET metabolic tumour volume versus ADC histogram: prognostic value of tumour treatment response and survival in patients with locally advanced uterine cervical cancer. *Br J Radiol* 90:20170035
32. Van Griethuysen JJM, Fedorov A, Parmar C, et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107
33. Burnham KP, Anderson DR (2004) Multimodel Inference. *Sociol Methods Res* 33:261–304
34. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723
35. DeLong ER, Carolina N (1988) Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves : A Nonparametric Approach. *Biometrics* 44:837–845
36. Shi L, Zhang Y, Nie K, et al (2019) Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI. *Magn Reson Imaging* 61:33–40

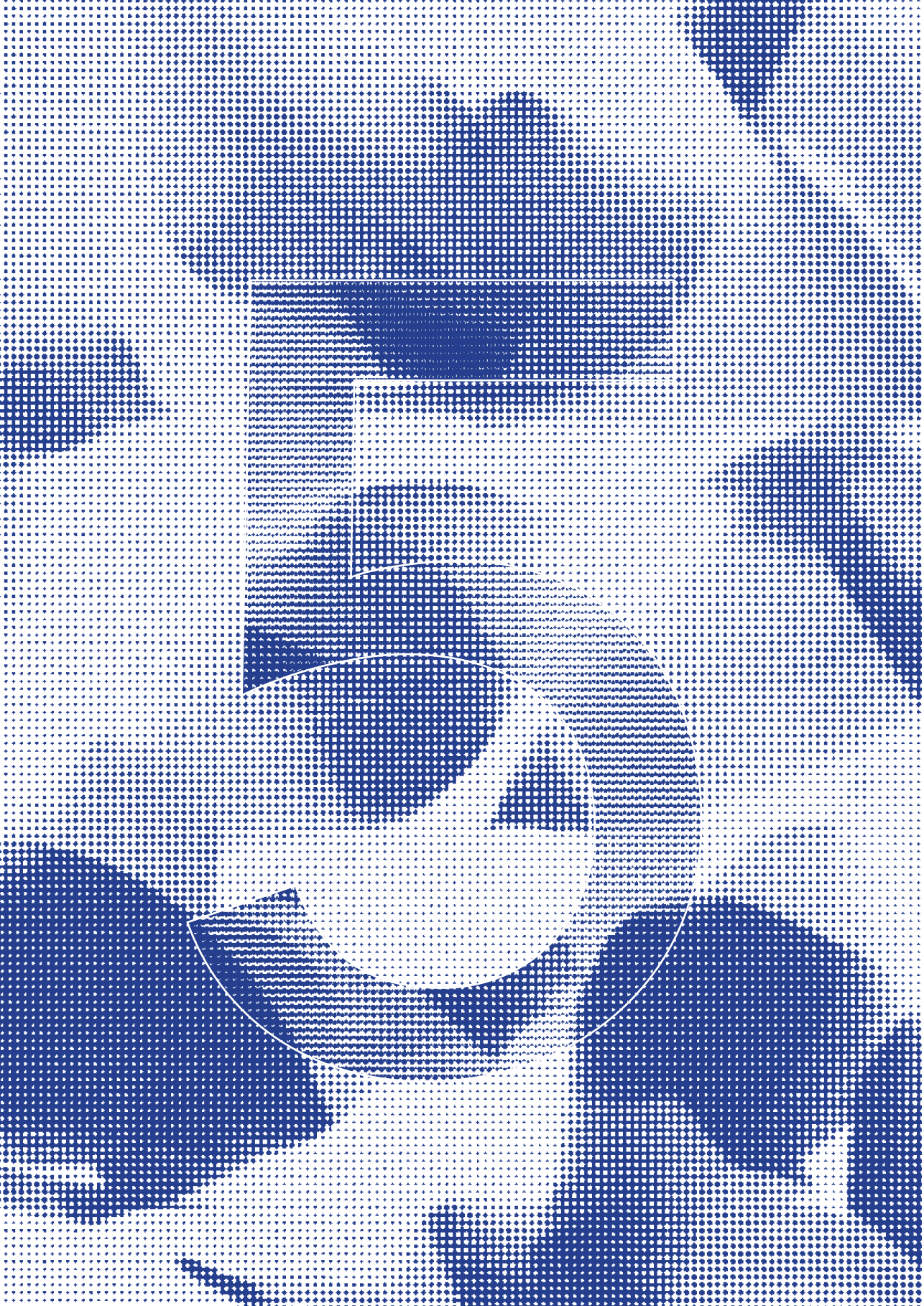
4

37. Antunes JT, Ofshteyn A, Bera K, et al (2020) Radiomic Features of Primary Rectal Cancers on Baseline T2-Weighted MRI Are Associated With Pathologic Complete Response to Neoadjuvant Chemoradiation: A Multisite Study. *J Magn Reson Imaging* 52:1531–1541
38. Shayesteh SP, Alikhassi A, Farhan F, et al (2020) Prediction of Response to Neoadjuvant Chemoradiotherapy by MRI-Based Machine Learning Texture Analysis in Rectal Cancer Patients. *J Gastrointest Cancer* 51:601–609
39. Petkovska I, Tixier F, Ortiz EJ, et al (2020) Clinical utility of radiomics at baseline rectal MRI to predict complete response of rectal cancer after chemoradiation therapy. *Abdom Radiol*. <https://doi.org/10.1007/s00261-020-02502-w>
40. Shen W-C, Chen S-W, Wu K-C, et al (2020) Predicting pathological complete response in rectal cancer after chemoradiotherapy with a random forest using 18F-fluorodeoxyglucose positron emission tomography and computed tomography radiomics. *Ann Transl Med* 8:207–207
41. Zou H, Yu J, Wei Y, et al (2019) Response to neoadjuvant chemoradiotherapy for locally advanced rectum cancer: Texture analysis of dynamic contrast-enhanced MRI. *J Magn Reson Imaging* 49:885–893
42. Liu S, Wen L, Hou J, et al (2019) Predicting the pathological response to chemoradiotherapy of non-mucinous rectal cancer using pretreatment texture features based on intravoxel incoherent motion diffusion-weighted imaging. *Abdom Radiol* 44:2689–2698
43. Hamerla G, Meyer H-J, Hamsch P, et al (2019) Radiomics Model Based on Non-Contrast CT Shows No Predictive Power for Complete Pathological Response in Locally Advanced Rectal Cancer. *Cancers (Basel)* 11:1680
44. Fu J, Zhong X, Li N, et al (2020) Deep learning-based radiomic features for improving neoadjuvant chemoradiation response prediction in locally advanced rectal cancer. *Phys Med Biol* 65:075001
45. Lovinfosse P, Polus M, Van Daele D, et al (2018) FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. *Eur J Nucl Med Mol Imaging* 45:365–375
46. Shu Z, Fang S, Ye Q, et al (2019) Prediction of efficacy of neoadjuvant chemoradiotherapy for rectal cancer: the value of texture analysis of magnetic resonance images. *Abdom Radiol* 21:1051–1058
47. Yi X, Pei Q, Zhang Y, et al (2019) MRI-Based Radiomics Predicts Tumor Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Front Oncol* 9:1–10

STUDYING LOCAL TUMOUR HETEROGENEITY ON MRI AND FDG-PET/CT TO
PREDICT RESPONSE TO NEOADJUVANT CHEMORADIOTHERAPY
IN RECTAL CANCER.

48. Bibault J-E, Giraud P, Housset M, et al (2018) Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep* 8:12611
49. Dinapoli N, Barbaro B, Gatta R, et al (2018) Magnetic Resonance, Vendor-independent, Intensity Histogram Analysis Predicting Pathologic Complete Response After Radiochemotherapy of Rectal Cancer. *Int J Radiat Oncol* 102:765–774
50. Chee CG, Kim YH, Lee KH, et al (2017) CT texture analysis in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy: A potential imaging biomarker for treatment response and prognosis. *PLoS One* 12:e0182883
51. Guha A, Connor S, Anjari M, et al (2020) Radiomic analysis for response assessment in advanced head and neck cancers, a distant dream or an inevitable reality? A systematic review of the current level of evidence. *Br J Radiol* 93:20190496
52. Granzier RWY, van Nijnatten TJA, Woodruff HC, et al (2019) Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: A systematic review. *Eur J Radiol* 121:108736
53. Fiz F, Viganò L, Gennaro N, et al (2020) Radiomics of Liver Metastases: A Systematic Review. *Cancers (Basel)* 12:2881
54. Lee G, Park H, Bak SH, Lee HY (2020) Radiomics in Lung Cancer from Basic to Advanced: Current Status and Future Directions. *Korean J Radiol* 21:159
55. Liu L, Yi X, Lu C, et al (2020) Applications of radiomics in genitourinary tumors. *Am J Cancer Res* 10:2293–2308
56. Oliveira FPM, Tavares JMRS (2014) Medical image registration: a review. *Comput Methods Biomech Biomed Engin* 17:73–93
57. Raaymakers BW, Lagendijk JJW, Overweg J, et al (2009) Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol* 54:N229–N237

4



Sources of variation in multicenter rectal MRI data and their effect on Radiomics feature reproducibility

Niels W. Schurink, Simon R van Kranen, Sander Roberti, Joost J.M. van Griethuysen, Nino Bogveradze, Francesca Castagnoli, Najim El Khababi, Frans C.H. Bakers, Shira H. de Bie, Gerlof P.T. Bosma, Vincent C. Cappendijk, Remy W.F. Geenen, Peter A. Neijenhuis, Gerald M. Peterson, Cornelis J. Veeken, Roy F.A. Vliegen, Regina G.H. Beets-Tan, Doenja M.J. Lambregts

Eur Radiol. 2021; doi: [10.1007/s00330-021-08251-8](https://doi.org/10.1007/s00330-021-08251-8)

ABSTRACT

OBJECTIVES

To investigate sources of variation in a multicenter rectal cancer MRI dataset focusing on hardware and image acquisition, segmentation methodology and Radiomics feature extraction software.

METHODS

T2W and DWI/ADC MRIs from 649 rectal cancer patients were retrospectively acquired in 9 centers. Fifty-two imaging features (14 first-order/6 shape/32 higher-order) were extracted from each scan using whole-volume (expert/non-expert) and single-slice segmentations using two different software packages (PyRadiomics/CapTk). Influence of hardware, acquisition and patient-intrinsic factors (age/gender/cTN-stage) on ADC was assessed using linear regression. Feature reproducibility was assessed between segmentation methods and software packages using the intraclass correlation coefficient.

RESULTS

Image features differed significantly ($p < 0.001$) between centers with more substantial variations in ADC compared to T2W-MRI. 64.3% of the variation in mean ADC was explained by differences in hardware and acquisition, compared to 0.4% by patient-intrinsic factors. Feature reproducibility between expert and non-expert segmentations was good to excellent (median ICC 0.89-0.90). Reproducibility for single-slice versus whole-volume segmentations was substantially poorer (median ICC 0.40-0.58). Between software packages, reproducibility was good to excellent (median ICC 0.99) for most features (first-order/shape/GLCM/GLRLM) but poor for higher-order (GLSZM/NGTDM) features (median ICC 0.00-0.41).

CONCLUSIONS

Significant variations are present in multicenter MRI data, particularly related to differences in hardware and acquisition, which will likely negatively influence subsequent analysis if not corrected for. Segmentation variations had minor impact when using whole volume-segmentations. Between software packages, higher-order features were less reproducible and caution is warranted when implementing these in prediction models.

INTRODUCTION

Over the past decade more than 100 papers have been published on the use of MR imaging biomarkers to predict various clinical outcomes in rectal cancer such as treatment response and survival [1–3]. Imaging biomarkers range from relatively simple measures (tumor size and volume) [1, 2] to ‘functional’ measures derived from imaging sequences such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced MRI [4]. More recently, the focus of research has shifted towards more advanced post-processing techniques such as Radiomics used to extract large numbers of quantitative features to construct a radiological phenotype of the studied lesion [2, 5]. Common Radiomics features include “first-order” histogram features (e.g. mean, standard deviation), shape features (e.g. volume, sphericity), and more complex higher-order texture features (e.g. gray-level co-occurrence matrix features) that describe patterns within the image.

While imaging biomarker studies have shown promising results to predict oncologic outcomes, several authors have voiced concern about the poor reproducibility and repeatability of these studies [6–8], related to small/underpowered single-center study designs, lack of independent model validation, and poor reproducibility of imaging features [6–8]. Important factors affecting reproducibility are data variations introduced by differences in acquisition, post-processing, and statistical analysis [9]. This is especially relevant for multicenter studies where data is generated using different hardware, software, and acquisition protocols, and where data is often evaluated by different readers. These variations are often referred to as “center effects” [10] and defined as “non-biological systematic differences between measurements of different batches of experiments” [11] that can negatively affect the performance of multicenter models [12].

Studies investigating sources of variation in imaging data have so far mainly focused on CT and PET and only one of 35 studies in a systematic review on Radiomic feature reproducibility focused on MRI [9]. Some recent studies have explored variations in quantitative MRI analysis, though mainly in phantoms [13–16] or small (<48 patients) single-center [13, 17, 18] or bi-institutional [19] patient cohorts. The current study aimed to add to these previous data by analyzing a large representative sample of rectal MRIs acquired at multiple institutions in the Netherlands to gain insight into how variations in “real life” clinical MRI data can affect Radiomics studies. In specific, the goal was to investigate sources of variation focusing on hardware, image acquisition, and effects of post-processing related to segmentation methodology and feature extraction software.

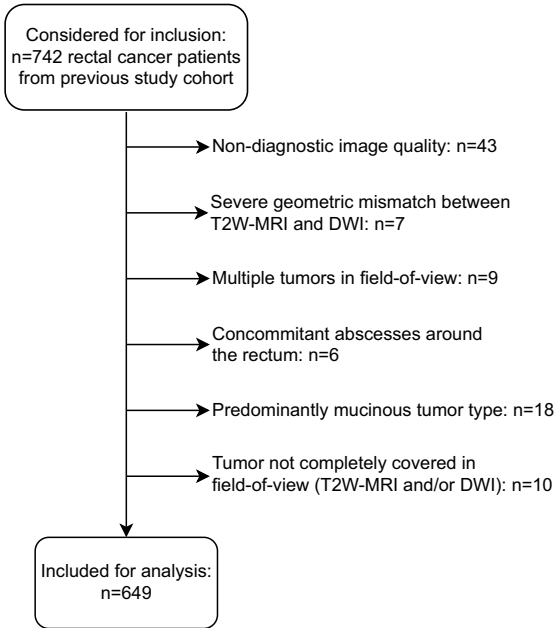


Figure 1. In- and exclusion flowchart

MATERIALS AND METHODS

Patients

For this retrospective study we analyzed a dataset of rectal MRI scans (scanned between 2012-2017) previously collected as part of an ongoing IRB-approved retrospective multicenter study on prediction of response to neoadjuvant treatment, including patients from nine different centers in the Netherlands (1 tertiary oncologic referral center, 1 academic and 7 non-academic centers). Inclusion criteria for this previous study were: (1) biopsy-proven rectal adenocarcinoma, (2) neoadjuvant treatment (chemoradiotherapy or 5x5Gy radiotherapy with a long waiting interval) followed by surgery or watch-and-wait (W&W), (3) availability of baseline staging MRI (including T2W-MRI and DWI) and (4) availability of clinical outcome to establish response. From this initial cohort of 742 patients, 93 were excluded for reasons detailed in the in-/exclusion flowchart in **Figure 1**, leaving a total study population of 649 patients. The overall study methodology is illustrated in **Figure 2**.

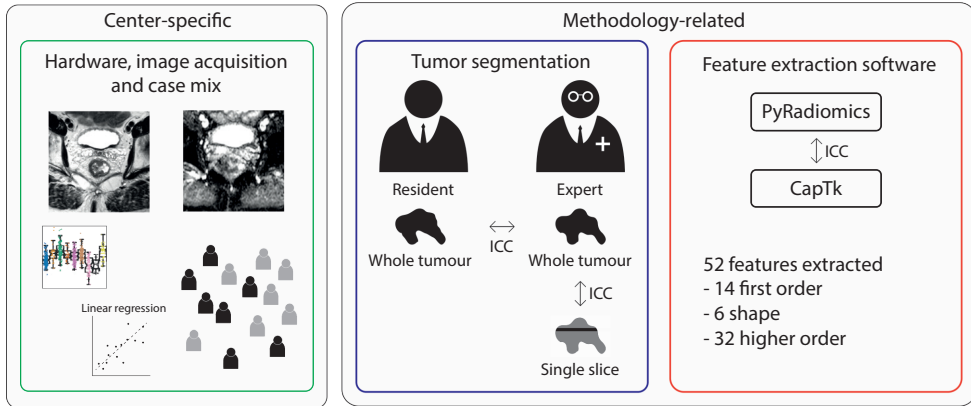


Figure 2. Study overview Two types of data variation between centers were analyzed: center-specific variations (related to hardware and image acquisition protocols, and case-mix) and methodology-related sources of variation (related to segmentation and feature extraction methodology). ICC=intra-class correlation coefficient.

Imaging and image processing steps

All images were acquired according to routine practice in the participating centers using various vendors and acquisition protocols. The transverse T2W-MRI and apparent diffusion coefficient (ADC) maps were selected for analysis, as these were most commonly reported in previous rectal cancer image biomarker studies [2]. ADC-maps were calculated from the DWI-series with a mono-exponential fit of the signal intensity using all available b-values (varying from 2-7 b-values per sequence; b-values ranging between b0-b2000). Negative ADC values (<0) or ADC values larger than 3 standard deviations from the mean (>mean+3SD) were marked invalid. As T2W pixel values are represented on an arbitrary scale, these images were normalized to mean=0 and standard deviation=100. All images were then resampled to a common pixel spacing of 2x2x2mm.

To explore effects of segmentation methodology, three types of tumor segmentations were generated using 3D-slicer (version-4.10.2). Segmentations were generated on high b-value DWI, using the T2W-MRI as an anatomical reference, and then copied to the T2W-MRI and ADC-maps. First, a non-expert reader (JvG or NS; resident-level with no specific expertise in reading rectal MRI) segmented the rectal tumors by applying the “level-tracing” algorithm on DWI and manually adjusting it to exclude obvious artefacts or non-tumor tissues (e.g. adjacent organs or lymph nodes). Second, a board-certified radiologist (DL; with >10 years of experience in rectal MRI) manually revised these segmentations, taking care to precisely delineate the tumor boundaries slice-

by-slice. Third, a single-slice segmentation was derived from this expert-segmentation including the axial slice with the largest tumor surface area. These three segmentations will be further referred to as: (1) non-expert, (2) expert, and (3) single-slice segmentation. Imaging features were extracted using PyRadiomics (version-v3.0). To explore effects of feature extraction methodology, features (for the whole-volume expert-segmentations) were additionally extracted with similar software settings using a different open-source software package, CapTk (version-1.8.1). Only features defined in both software packages were extracted, including 52 features in total: 14 first-order, 6 shape, and 32 higher-order (7 gray-level co-occurrence matrix (GLCM), 16 gray-level size zone matrix (GLSZM), 4 gray-level run-length matrix (GLRLM), and 5 neighboring gray-tone difference matrix (NGTDM)).

Analysis of sources of variation:

1. Center variations (case-mix, hardware, and image acquisition)

To investigate potential effects of “case-mix” differences, baseline patient characteristics were compared between centers using the Kruskal-Wallis test for age, T-stage and N-stage, and Chi-squared test for sex.

As a first exploratory step, we derived 6 basic imaging features (minimum, maximum, mean, standard deviation, entropy, and tumor volume) for each patient for both the T2W-MRI and ADC maps. The distribution of these features within our cohort was then visualized for each center separately using notched boxplots. To test whether the medians of the derived features were significantly different between patients from different centers we used Kruskal-Wallis; to identify which specific centers have different feature distributions, a post-hoc pairwise Mann-Whitney U-test was performed with Bonferroni correction to account for multiple testing. **Supplementary Materials 1** describes a sub-analysis exploring whether differences between centers can be harmonized retrospectively by adjusting the b-values for ADC-calculation or by performing data normalization using reference organs or z-transformation.

Using multivariable linear regression, we further explored effects of variations in hardware (vendor/scanner model, field strength) and acquisition parameters (slice thickness, acquired in-plane resolution, repetition time, echo time, number of signal averages, maximum b-value, number of b-values, signal-to-noise ratio) on ADC, and compared these to various patient-intrinsic (baseline and clinical outcome) parameters previously reported to be correlated with ADC (including sex, age, cT-stage, cN-stage, response to chemoradiotherapy and tumor volume [1, 2, 20]). “Center” (i.e. hospital) was investigated as a final parameter to account for unknown variations not covered by the other variables (e.g. patient preparation and undocumented acquisition parameters).

Table 1. Baseline patient characteristics and variations between centers

	Total	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6	Center 7	Center 8	Center 9	P-Value
TOTAL, n (%)	n=649 (100%)	n=133 (20.5%)	n=27 (4.2%)	n=99 (15.3%)	n=21 (3.2%)	n=137 (21.1%)	n=97 (14.9%)	n=81 (12.5%)	n=11 (1.7%)	n=43 (6.6%)	
sex, n (%)											
Female	232 (35.7)	45 (33.8)	12 (44.4)	37 (37.4)	6 (28.6)	52 (38.0)	35 (36.1)	28 (34.6)	6 (54.5)	11 (25.6)	0.686*
Male	417 (64.3)	88 (66.2)	15 (55.6)	62 (62.6)	15 (71.4)	85 (62.0)	62 (63.9)	53 (65.4)	5 (45.5)	32 (74.4)	
age, median (range)	65 (26-88)	67 (31-87)	69 (41-88)	67 (26-86)	65 (46-77)	65 (39-86)	64 (35-83)	65 (44-82)	55 (32-79)	62 (44-80)	0.112**
cT, n (%)											
2	40 (6.2)	7 (5.3)	1 (3.7)	7 (7.1)	1 (4.8)	11 (8.0)	4 (4.1)	6 (7.4)	0 (0.0)	3 (7.0)	
3	523 (80.6)	101 (75.9)	21 (77.8)	75 (75.8)	16 (76.2)	112 (81.8)	83 (85.6)	65 (80.2)	11 (100.0)	39 (90.7)	0.228**
4	86 (13.3)	25 (18.8)	5 (18.5)	17 (17.2)	4 (19.0)	14 (10.2)	10 (10.3)	10 (12.3)	0 (0.0)	1 (2.3)	
0	100 (15.4)	6 (4.5)	9 (33.3)	20 (20.2)	2 (9.5)	31 (22.6)	13 (13.4)	10 (12.3)	2 (18.2)	7 (16.3)	
cN, n (%)											
1	170 (26.2)	31 (23.3)	7 (25.9)	26 (26.3)	6 (28.6)	38 (27.7)	23 (23.7)	15 (18.5)	7 (63.6)	17 (39.5)	<0.001**.#
2	379 (58.4)	96 (72.2)	11 (40.7)	53 (53.5)	13 (61.9)	68 (49.6)	61 (62.9)	56 (69.1)	2 (18.2)	19 (44.2)	

*Calculated using Chi-squared test

** Calculated using Kruskal-Wallis test

Post-hoc analysis using pair-wise Mann-Whitney U test indicated that Center 1 significantly differed from the other centers with respect to cN-stage ($p < 0.001$); no significant differences were detected between the remaining 8 centers.

Analyses were performed using R version-3.6.1, and p-values <0.05 were considered statistically significant. Further details on the regression analysis are provided in **Supplementary Materials 2**.

2. Image segmentation

Imaging features were compared between the expert, non-expert, and single-slice segmentations using the two-way absolute agreement intra-class correlation coefficient (ICC), with $ICC < 0.50$ indicating poor agreement, $0.50 \leq ICC < 0.75$ moderate agreement, $0.75 \leq ICC < 0.90$ good agreement, and $ICC > 0.90$ excellent agreement [21].

3. Feature extraction software

Imaging features derived with PyRadiomics (using expert segmentations) were compared to those derived using CapTk using the two-way absolute agreement ICC and the same cut-offs for agreement detailed above [21].

RESULTS

Sources of variation

1. Center variations (patient-mix, hardware, and image acquisition)

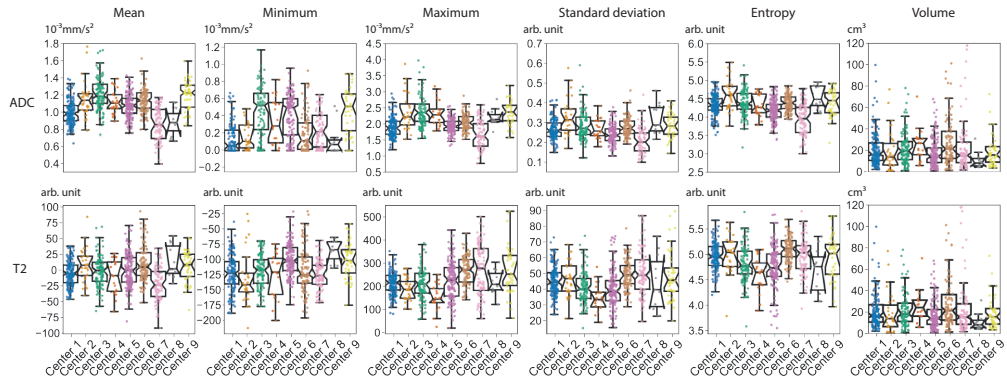
Baseline characteristics of the 649 study patients (417 male, median age 65 years) are provided in **Table 1**. There were no significant differences in cT-stage, age, and sex distribution between the nine centers ($p=0.11-0.69$), except for cN-stage that was significantly higher in one center ($p < 0.001$). An overview of the main variations in hardware and acquisition protocols is provided in **Table 2**. The distribution of basic first-order feature values per center and post-hoc analyses illustrating differences between individual centers are depicted in **Figure 3**. All tested features differed significantly between centers (Kruskal-Wallis $p < 0.001$) on both T2W-MRI and ADC. Pairwise comparisons between individual centers revealed that mainly ADC mean, minimum and maximum showed significant differences between the majority of the centers, while for T2W-MRI features, and ADC standard deviation and entropy, differences were limited to 2-4 individual centers. Data variations between centers did not improve after b-value harmonization and remained significant after applying different retrospective normalization methods, though normalization using inguinal lymph nodes as a reference organ did have a positive effect in reducing data variations as outlined in **Supplementary Materials 1**. Tumor volumes were mostly comparable between centers and only differed significantly between a minority of individual centers.

Table 2. Overview of main variations in hardware and acquisition protocols between the 9 participating centers

Hardware		
Total number of scanners		n=26
Total number of scanner models		n=13
Vendor type		
Philips Healthcare (used in 6 centers)		n=11 (incl. 4 different scanner models)
Siemens Healthineers (used in 5 center)		n=12 (incl. 7 different scanner models)
GE Healthcare (used in 2 centers)		n=3 (incl. 2 different scanner models)
Field strength		
1.5 T		n=19
3.0 T		n=7
Acquisition protocol		
Parameter	T2W-MRI median (range)	DWI median (range)
TR (ms)	4235 (866-16738)	5475 (948-11000)
TE (ms)	108 (60-250)	80 (37-117)
Flip angle (°)	150 (90-173)	90 (70-180)
NSA	2 (1-6)	5 (1-15)
Slice thickness (mm)	3 (3-5)	5 (2.7-8)
Pixel spacing (mm)	0.63 (0.29-1.48)	1.63 (0.63-3.52)
Field of view (mm)	200 (150-400)	320 (160-520)
Total number of b-values	N/A	3 (2-7)
Lowest b-value	N/A	0 (0-50)
Highest b-value	N/A	1000 (600-2000)

NSA: Number of signal averages, T: Tesla, TE: Echo Time, TR: repetition time

A. Overall difference between centers



B. Pairwise comparison between individual centers

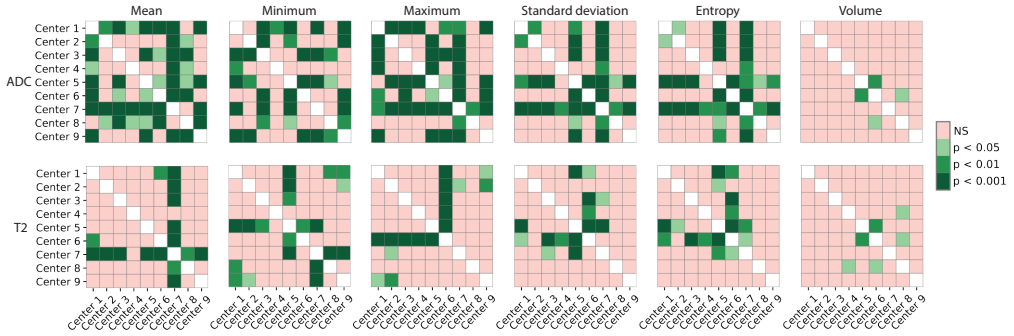


Figure 3. Center variations

A. Visualization of the distribution of 6 basic (first-order + volume) imaging features within our study cohort, grouped by center. The imaging features were extracted from the rectal tumors on the ADC map (upper row) and T2W-MRI (bottom row), respectively. The boxplots show the distribution of the feature values for all patients within each center, with the notches in each box plot representing the 95% confidence intervals of the median feature value within a center. Kruskal-Wallis tests showed that for all features these median values were significantly different between the centers ($p < 0.001$). B. Additional post-hoc pairwise significance tests to explore which specific centers had significantly different feature values, with pink indicating no significant differences between centers and green indicating a significant difference (darker green corresponding to higher level of significance). Bonferroni correction was used to account for multiple testing.

Table 3. Factors attributing to mean tumor ADC	
Factors:	Proportion of variance in ADC predicted by these factors (LOOCV R²)
A. Hardware and acquisition parameters:	64.3%
<ul style="list-style-type: none"> - Repetition time (TR)* - Echo time (TE)* - Flip angle - Pixel Bandwidth - In plane resolution* - Slice thickness - Number of signal averages (NSA)* - Maximum b-value* - Number of b-values* - Signal to noise ratio (SNR)* - Scanner model* - Magnetic field strength 	
B. Patient-intrinsic parameters	0.4%
<ul style="list-style-type: none"> - Age - Sex - cT-stage (assessed on baseline MRI) - cN-stage (assessed on baseline MRI) - Response to chemoradiotherapy (complete versus incomplete response) - Tumor volume 	
C. Center	32.5%
Umbrella variable to account for any additional unknown variations between centers (e.g. patient preparation protocols, types of coils used, fat suppression techniques, etc.) Significant parameters: center	
All (A + B + C) combined	63.5%
Significant parameters: center, age, TR, TE, in plane resolution, slice thickness, NSA, maximum b-values, number of b-values, scanner model, SNR	

Further details of the regression analysis can be found in **Supplementary Materials 2**. The LOOCV R² value is a leave-one-out cross-validated goodness-of-fit measure indicating the proportion of the variance in the dependent (i.e. ADC) variable that can be explained by the independent variables using a linear regression model.

* Indicate the variables that were significant with a p-value < 0.05 based on a likelihood ratio test. For continuous variables, all polynomial terms were tested jointly.

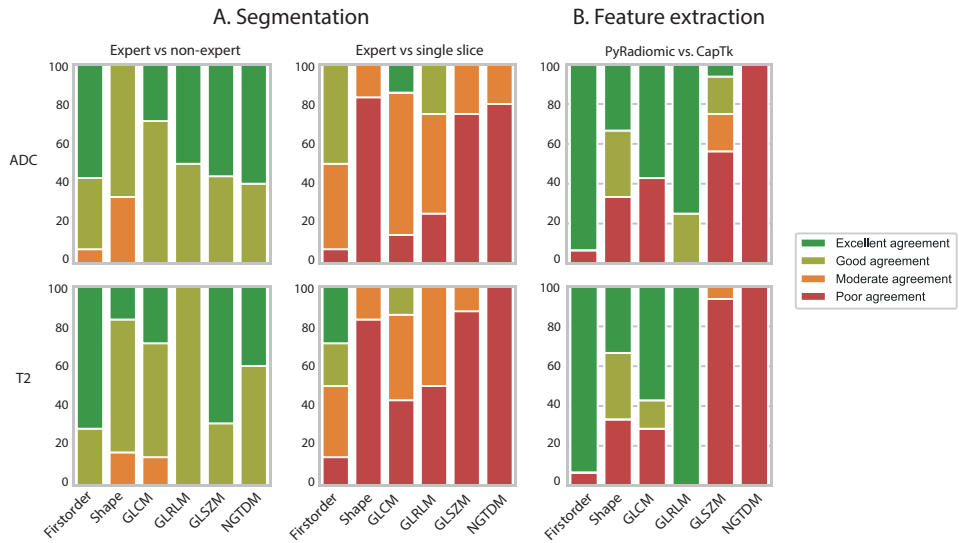


Figure 4. Effects of image segmentation and feature extraction methodology. Feature reproducibility for ADC (upper row) and T2W-MRI (bottom row) using different segmentation methods (A) and feature extraction packages (B). Each column corresponds to the percentage of features showing excellent (dark green, $ICC > 0.90$), good (green, $0.90 > ICC > 0.75$), moderate (orange, $0.75 > ICC > 0.5$) or poor (red, $ICC < 0.5$) agreement. In total, 52 features were analyzed, including 14 first-order, 6 shape, 7 Gray-Level Co-occurrence Matrix (GLCM), 4 Gray-Level Run-Length Matrix (GLRLM), 16 Gray-Level Size Zone Matrix (GLSZM) and 5 Neighboring Gray-Tone Difference Matrix (NGTDM) features.

The results of the regression models developed to investigate the influence of hardware, acquisition, and patient-intrinsic (baseline and clinical outcome) parameters on mean tumor ADC are reported in **Table 3**. Acquisition parameters had the strongest association with ADC and on their own were able to explain 64.3% of the variation in ADC present in the data. Patient-intrinsic parameters (e.g. age, gender, TN-stage, treatment response) had a negligible effect on ADC and were able to explain only 0.4% of the variation in ADC. The umbrella variable “Center” was able to explain 32.5% of the variation in ADC. When combining all factors in one model, the model explained 63.5% of the data variation, with acquisition and hardware parameters as the main predictors.

2. Image segmentation

The results of the reproducibility analysis using different segmentation strategies are depicted in **Figure 4A**. Reproducibility between expert and non-expert segmentations was good-excellent for the majority of features (first-order, shape, and higher-order) with ICC values ranging between 0.72-0.99 (median 0.90) for T2W-MRI and 0.53-0.99 (median 0.89) for ADC. Compared to the expert whole-volume segmentations, the extracted single-slice segmentations resulted in considerably lower reproducibility with an ICC of 0.00-0.94 (median 0.40) for T2W-MRI and ICC of 0.00-0.97 (median 0.58) for ADC, with poor results for shape, GLSZM, and NGTDM features.

3. Feature extraction software

The influence of feature extraction software is depicted in **Figure 4B**. The majority of first-order, shape, GLCM and GLRLM features showed good-excellent reproducibility with similar results for features derived from T2W-MRI or ADC with ICCs ranging between 0.00-1.00 (median 0.99) for both modalities. In contrast, the majority of GLSZM and NGTDM features were poorly reproducible with ICCs of 0.00-0.56 (median 0.00) for T2W-MRI and ICCs 0.01-0.99 (median 0.41) for ADC.

DISCUSSION

The results of this study show that variations in quantitative imaging (Radiomics) in a large clinical multicenter dataset of rectal MRIs were more substantial for DWI/ADC than for T2W-MRI and mainly related to hardware and acquisition protocols (i.e. "center effects"). Effects of segmentation methodology and feature extraction software on feature variation were less significant, particularly for the more basic first-order and shape-related features that showed overall good reproducibility

An exploratory analysis on the feature distribution of 6 basic imaging features (first-order + volume) showed significant variation between patient populations coming from different centers, with more significant differences for ADC than for T2W-MRI. Tumor volume was the most robust feature with most comparable results between centers. This is in line with a previous report on the repeatability of MRI features in a small cohort (n=48) of patients with brain glioblastoma, which showed that shape features (including volume) resulted in higher repeatability than features derived from T2W-MRI pixel intensities [18]. Similarly, shape features were found to have the highest repeatability and reproducibility on T2W-MRI of cervical cancer [22].

Since ADC data showed the largest variations, we developed a linear regression model to investigate in-depth which factors influence mean tumor ADC. The majority (>60%) of the ADC variation could be predicted using only hardware and acquisition-related parameters while patient-intrinsic (clinical and tumor) features alone predicted only 0.4% of the variation in ADC. This suggests that—when building multicenter prediction models—any potential relation between clinical outcomes and ADC will likely be obscured when no correction is performed to account for acquisition and hardware differences. This is a very relevant issue when incorporating ADC in retrospective multicenter studies without protocol standardization. Even in controlled prospective study designs with harmonized acquisition protocols, variation in ADC can still be a limiting factor as coefficients of variation range as high as 4-37% depending on the measured organ [23, 24]. Differences in acquisition settings have previously also been shown to substantially reduce inter- and intra-scanner reproducibility of T2W-MRI radiomics features, with particularly poor results for higher-order features [13, 16].

Several methods have been suggested to correct for data variations between centers. A first option is to discard features that are poorly reproducible across centers, with the advantage of creating simpler models (though with the drawback of losing potentially valuable information) [12]. Two alternative options are normalization in the image domain (e.g. z-transform or within patient normalization using reference tissues) or feature domain (e.g. ComBat harmonization). These techniques have been shown to significantly improve T2W-MRI feature reproducibility [18], and to successfully correct for “batch-effects” (similar to “center-effects”) in genomic studies [12]. The latter approach has recently been adopted for Radiomics with promising results [25]. In our exploratory analysis described in **Supplementary Materials 1**, data normalization using lymphoid tissue (benign inguinal lymph nodes) as a reference organ had a positive effect to reduce ADC data variations between centers, though differences remained statistically significant and the benefits of this approach will need to be further investigated in studies where features are tested against a clinical outcome. The fourth option is to use statistical models that specifically take center effects into account (e.g. random/mixed-effect models [10]). These various options all have their strengths and weaknesses and evidence-based guidelines on the preferred (combination of) methods to handle center effects in multicenter Radiomics research are so far lacking and urgently needed.

Regarding image segmentation methodology, we found poor reproducibility for higher-order features (e.g. GLSZM and GLRLM) but overall good reproducibility for

simpler features (e.g. first-order, GLCM) similar to previous reports [9, 26, 27]. The features derived from single-slice segmentations showed the poorest reproducibility, which is in line with previous single-center reports [28, 29], indicating that—though less cumbersome—single-slice methods are not recommendable. Interestingly, feature reproducibility was good-excellent between expert and non-expert readers, indicating that input from expert-radiologists is not necessarily required. This is in line with a previous report where a Radiomics model was trained to predict response to chemoradiotherapy in rectal cancer and achieved similar performance regardless of whether segmentations were performed by expert (AUC 0.67-0.83) or non-expert readers (AUC 0.69-0.79) [30]. This is reassuring given the tremendous workload associated with image segmentation, especially when analyzing large volumes of imaging data. Another potential solution to reduce this workload could be to use computer algorithms to (semi-)automatically generate tumor segmentations [31]. There have been some promising reports showing that computer algorithms may generate segmentations similar to manual tumor delineations [31, 32], provided that image quality is good [33].

When comparing feature reproducibility using different software packages, we found that the majority of first-order, shape, GLCM and GLRLM features showed good to excellent reproducibility whereas the majority of higher-order features (GLSZM and NGTDM) were poorly reproducible. This is in line with earlier findings that higher-order features are generally less reproducible than first-order and shape features [9] which can probably be partly attributed to technical differences in the implementation of features and/or image processing by different software and computational algorithms. This underlines the importance of accurately reporting software versions, and preferably using packages with standardized feature definitions such as those defined by the image biomarker standardization initiative (IBSI). Considering the poor reproducibility of the higher-order features in our dataset, caution should be taken when incorporating more advanced features into clinical prediction models.

The main novelty of the current study lies in its multicenter aspect. Although previous studies have already identified acquisition parameters [13, 15, 16], segmentation [19] and post-processing methods [15, 18] as factors affecting feature reproducibility, these studies have so far mainly been performed on non-MRI data or in small patient cohorts or phantoms. The extent to which these effects influence feature reproducibility and may obscure correlations with common clinical outcomes in a representative “real life” clinical cohort of MRI data acquired at various institutions has not been previously

reported. There were, however, some limitations to our study design in addition to its retrospective nature. As the data was acquired and anonymized to comply with privacy regulations, only basic acquisition information could be extracted from the DICOM headers. Other potential sources of variation, such as coil use, fat suppression, MRI software version, and patient preparation, therefore, remain underexposed. In addition, all segmentations were done on the high b-value DWI (and then copied to the other modalities). Although care was taken to take the anatomical information of T2W-MRI into account during the segmentations, ideally a separate segmentation would have been performed on T2W-MRI. This, however, was not feasible to accomplish within an acceptable timeframe. Finally, several data-processing choices (e.g. resampling voxel-size, bin-width, gray-level discretization, and T2W-MRI normalization) were made which may have influenced the extracted features [34–36]. Although some of these steps may have introduced some bias in our analyses a more detailed analysis on the impact of these choices was beyond the scope of this paper.

In conclusion, this study has shown that significant variations between centers are present in multicenter rectal MRI data with more substantial variations in DWI/ADC compared to T2W-MRI, which are mainly related to hardware and image acquisition protocols (i.e. "center effects"). These effects need to be accounted for when analyzing multicenter MRI datasets to avoid that potential correlations with the clinical outcome under investigation are overlooked. Image segmentation has relatively minor effects on image quantification provided that whole-volume segmentations are performed. Expert segmentation input is not necessarily required to acquire stable features, which could shift the daunting task of image segmentation from expert-radiologists to less experienced readers or even (semi-)automatic software algorithms. Higher-order features were less reproducible between software packages and caution is therefore warranted when implementing these into clinical prediction models.

REFERENCES

1. Schurink NW, Lambregts DMJ, Beets-Tan RGH (2019) Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. *Br J Radiol* 92:20180655
2. Di Re AM, Sun Y, Sundaresan P, et al (2021) MRI radiomics in the prediction of therapeutic response to neoadjuvant therapy for locoregionally advanced rectal cancer: a systematic review. *Expert Rev Anticancer Ther* 21:425–449
3. Staal FCR, van der Reijdt DJ, Taghavi M, et al (2021) Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review. *Clin Colorectal Cancer* 20:52–71
4. Pham TT, Liney GP, Wong K, Barton MB (2017) Functional MRI for quantitative treatment response prediction in locally advanced rectal cancer. *Br J Radiol* 90:20151078
5. Lambin P, Rios-Velazquez E, Leijenaar R, et al (2012) Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48:441–446
6. Wright BD, Vo N, Nolan J, et al (2020) An analysis of key indicators of reproducibility in radiology. *Insights Imaging* 11:65
7. Song J, Yin Y, Wang H, et al (2020) A review of original articles published in the emerging field of radiomics. *Eur J Radiol* 127:108991
8. Aerts HJWL (2018) Data Science in Radiology: A Path Forward. *Clin Cancer Res* 24:532–534
9. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol* 102:1143–1158
10. Kahan BC (2014) Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? *BMC Med Res Methodol* 14:20
11. Luo J, Schumacher M, Scherer A, et al (2010) A comparison of batch effect removal methods for enhancement of prediction performance using MAQC-II microarray gene expression data. *Pharmacogenomics J* 10:278–291
12. Da-Ano R, Visvikis D, Hatt M (2020) Harmonization strategies for multicenter radiomics investigations. *Phys Med Biol* 65:24TR02
13. Mi H, Yuan M, Suo S, et al (2020) Impact of different scanners and acquisition parameters on robustness of MR radiomics features based on women’s cervix. *Sci Rep* 10:20407
14. Baeßler B, Weiss K, Pinto dos Santos D (2019) Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging. *Invest Radiol* 54:221–228

15. Ammari S, Pitre-Champagnat S, Dercle L, et al (2021) Influence of Magnetic Field Strength on Magnetic Resonance Imaging Radiomics Features in Brain Imaging, an In Vitro and In Vivo Study. *Front Oncol* 10:1–11
16. Yuan J, Xue C, Lo G, et al (2021) Quantitative assessment of acquisition imaging parameters on MRI radiomics features: a prospective anthropomorphic phantom study using a 3D-T2W-TSE sequence for MR-guided-radiotherapy. *Quant Imaging Med Surg* 11:1870–1887
17. Gourtsoyianni S, Doumou G, Prezzi D, et al (2017) Primary Rectal Cancer: Repeatability of Global and Local-Regional MR Imaging Texture Features. *Radiology* 284:552–561
18. Hoebel K V., Patel JB, Beers AL, et al (2021) Radiomics Repeatability Pitfalls in a Scan-Rescan MRI Study of Glioblastoma. *Radiol Artif Intell* 3:e190199
19. Traverso A, Kazmierski M, Shi Z, et al (2019) Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing. *Phys Medica* 61:44–51
20. Attenberger UI, Pilz LR, Morelli JN, et al (2014) Multi-parametric MRI of rectal cancer - Do quantitative functional MR measurements correlate with radiologic and pathologic tumor stages? *Eur J Radiol* 83:1036–1043
21. Perinetti G (2018) StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. *South Eur J Orthod Dentofac Res* 5:3–5
22. Fiset S, Welch ML, Weiss J, et al (2019) Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol* 135:107–114
23. Michoux NF, Ceranka JW, Vandemeulebroucke J, et al (2021) Repeatability and reproducibility of ADC measurements: a prospective multicenter whole-body-MRI study. *Eur Radiol* 31:4514–4527
24. Donati OF, Chong D, Nanz D, et al (2014) Diffusion-weighted MR Imaging of Upper Abdominal Organs: Field Strength and Intervendor Variability of Apparent Diffusion Coefficients. *Radiology* 270:454–463
25. Orlhac F, Lecler A, Savatovski J, et al (2021) How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol* 31:2272–2280
26. Haarbuerger C, Schock J, Truhn D, et al (2019) Radiomic Feature Stability Analysis based on Probabilistic Segmentations
27. Lee J, Steinmann A, Ding Y, et al (2021) Radiomics feature robustness as measured using an MRI phantom. *Sci Rep* 11:3973

28. Lambregts DMJ, Beets GL, Maas M, et al (2011) Tumour ADC measurements in rectal cancer: effect of ROI methods on ADC values and interobserver variability. *Eur Radiol* 21:2567–2574
29. Nougaret S, Vargas HA, Lakhman Y, et al (2016) 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* 280:446–454
30. van Griethuysen JJM, Lambregts DMJ, Trebeschi S, et al (2020) Radiomics performs comparable to morphologic assessment by expert radiologists for prediction of response to neoadjuvant chemoradiotherapy on baseline staging MRI in rectal cancer. *Abdom Radiol* 45:632–643
31. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, et al (2017) Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR. *Sci Rep* 7:5301
32. Van Heeswijk MM, Lambregts DMJ, van Griethuysen JJM, et al (2016) Automated and semiautomated segmentation of rectal tumor volumes on diffusion-weighted MRI: Can it replace manual volumetry? *Int J Radiat Oncol Biol Phys* 94:824–831
33. Van Griethuysen J, Schurink N, Lahaye MJ, et al (2020) Deep learning for fully automated segmentation of rectal tumours on MRI in a multicentre setting. In: *ESGAR 2020 Book of Abstracts*. Insights Imaging, 11:64
34. Li Q, Bai H, Chen Y, et al (2017) A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme. *Sci Rep* 7:14331
35. Duron L, Balvay D, Vande Perre S, et al (2019) Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS One* 14:e0213459
36. Isaksson LJ, Raimondi S, Botta F, et al (2020) Effects of MRI image normalization techniques in prostate cancer radiomics. *Phys Medica* 71:7–13



Development and multicenter validation of a multiparametric imaging model to predict treatment response in rectal cancer

Niels W. Schurink, Simon van Kranen, Joost J.M. van Griethuysen, Sander Roberti, Petur Snaebjornsson, Frans C.H. Bakers, Shira H. De Bie, Gerlof P.T. Bosma, Vincent C. Cappendijk, Remy W.F. Geenen, Peter A. Neijenhuis, Gerald M. Peterson, Cornelis J. Veeken, Roy F.A. Vliegen, Femke P. Peters, Nino Bogveradze, Najim el Khababi, Max J. Lahaye, Monique Maas, Geerard L. Beets, Regina G.H. Beets-Tan, Doenja M.J. Lambregts

Submitted for publication

ABSTRACT

AIM

To develop and validate a multiparametric imaging model to predict response to neoadjuvant treatment in rectal cancer using a clinically representative heterogeneous multicenter dataset of baseline staging MRIs.

METHODS

Primary staging MRIs (T2W-MRI and DWI/ADC) of 509 rectal cancer patients treated with neoadjuvant chemoradiotherapy (CRT) were collected from nine centers. Response outcomes were defined as (1) complete versus incomplete response, or (2) good (Mandard tumor regression grade, TRG1-2) versus poor response (TRG3-5). Models to predict these outcomes were developed based on combinations of the following variable groups:

1. non-imaging: age/sex/tumor-location/tumor-morphology/CRT-surgery interval
2. basic imaging staging: cT-stage/cN-stage/mesorectal fascia involvement, derived from either (2a) the original staging reports, or (2b) re-evaluation staging by a dedicated expert
3. advanced imaging staging: variables from 2b combined with cTN-substaging/depth of invasion/extramural vascular invasion/tumor length
4. quantitative imaging: volume + first-order histogram features derived from T2W-MRI and DWI/ADC using whole-tumor delineations.

Prediction models were developed with data from 6 centers (n=412) using logistic regression with LASSO feature selection, and internally validated using repeated (n=100) random hold-out validation. The best performing model was tested on an external validation cohort (3 centers; n=97).

RESULTS

After external validation, the best performing model (including non-imaging and advanced staging variables) achieved an area under the receiver-operating characteristic curve (AUC) of 0.60 (95%CI 0.48-0.72) to predict complete response and AUC 0.65 (95%CI 0.53-0.76) to predict a good response. Quantitative imaging variables did not contribute to improved model performance. Basic staging variables consistently achieved lower performance compared to advanced image staging variables.

DISCUSSION

Overall model performance to predict response to chemoradiotherapy was moderate. Best results were obtained with advanced image-based staging variables, highlighting the importance of accurate staging according to current guidelines. No added value was found for quantitative imaging features in this heterogeneous multicenter dataset.

INTRODUCTION

Locally advanced rectal tumors are typically treated with neoadjuvant chemoradiotherapy (CRT) aiming to downstage the tumor prior to surgery to achieve a more effective oncological resection, thereby reducing the risk for a local recurrence [1]. In up to 15-27% of the cases a complete tumor remission may be achieved as a result of CRT [2]. This has contributed to the recent paradigm shift in rectal cancer treatment towards organ preservation (e.g. “watch-and-wait” or local treatment of small tumor remnants) for selected patients with clinical evidence of a very good or complete tumor response after CRT. For these organ-preservation approaches, the morbidity and mortality risks associated with major surgery are avoided, with good reported clinical outcomes regarding the quality of life and overall survival [3, 4]. Predicting the response to CRT and thus the chance of achieving organ-preservation before the start of treatment, i.e. at baseline, may open up new possibilities to further personalize neoadjuvant treatment strategies depending on the anticipated treatment benefit, particularly for smaller tumors that do not necessarily require CRT for oncological reasons.

In recent years, an increasing volume of research indicates a possible role for imaging in this setting [5–9]. Promising results have been reported for clinical staging variables (MRI-based TN-stage) [6, 7], tumor volume [10–12], and functional imaging parameters (e.g. from diffusion-weighted imaging (DWI) [8, 9] or dynamic contrast enhanced MRI (DCE) [13]) to predict rectal tumor response on baseline MRI, and more recently also for more advanced quantitative variables derived using modern post-processing tools such as radiomics [5]. However, available evidence so far mainly comes from single-center studies focusing on a limited number of imaging variables. Multicenter studies incorporating clinical, functional as well as advanced quantitative imaging data are scarce [14, 15]. Moreover, the effects of multicenter data variations related to image acquisition and diagnostic staging differences between observers so far remain largely uninvestigated. Prediction studies on larger multicenter patient cohorts with imaging data acquired and analyzed as part of everyday clinical routine are therefore urgently

6

needed to develop a more realistic view of the potential role of image-based treatment prediction models in general clinical practice.

In this retrospective multicenter study, we set out to develop and validate a model to predict response to neoadjuvant treatment in rectal cancer using rectal MRIs acquired for baseline staging in 9 different centers in the Netherlands, intended to be a representative sample of rectal imaging performed in everyday clinical practice.

METHODS

Patients

As part of an institutional review board approved multicenter study project, the clinical and imaging data of 670 patients undergoing standard of care neoadjuvant chemoradiotherapy for newly diagnosed locally advanced rectal cancer between February 2008 and March 2018 were retrospectively collected from 9 participating study centers (1 university hospital, 7 large teaching hospitals and 1 comprehensive cancer center). Patients were identified based on the following inclusion criteria: (a) biopsy-proven rectal adenocarcinoma, (b) non-metastasized disease, (c) availability of a pre-treatment MRI (including at least T2-weighted (T2W) sequences in multiple planes and an axial DWI sequence) with corresponding radiological staging report (d) routine long-course neoadjuvant treatment consisting of radiotherapy (total dose 50.0-50.4 Gray) with concurrent capecitabine-based chemotherapy, (e) final treatment consisting of surgery or watch-and-wait with >2 years clinical follow-up to establish a reliable final response to CRT. From this initial cohort, 161 patients were excluded for reasons detailed in the inclusion/exclusion flowchart in **Figure 1**, leaving a total study population of 509 patients. Due to the retrospective nature of this study, informed consent was waived.

Imaging and image pre-processing

MRIs were acquired according to routine practice in the participating centers with substantial variations in scan protocols and corresponding image quality between (and within) centers, as illustrated in **Figure 2**. During the study inclusion period, images were acquired using 25 different scanners (19 1.5T; 6 3.0T) and a total of 112 unique T2W and 94 unique DWI protocols. Further parameters are summarized in **Supplementary Materials A**. From the source DW images we calculated the Apparent Diffusion Coefficient (ADC) maps using all available b-values (varying from 2-7 b-values

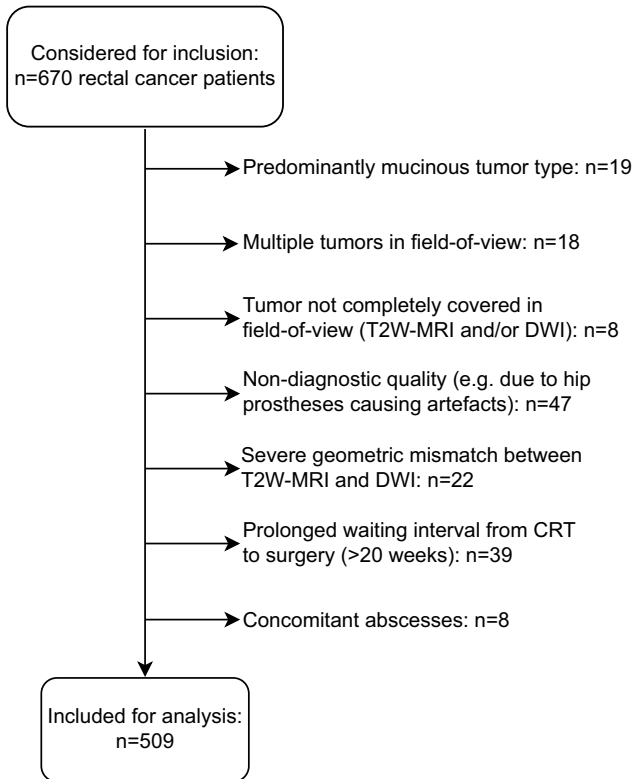


Figure 1. In- and exclusion flowchart. Note, mucinous tumors were excluded because these are known to exhibit distinctly different signal characteristics on both T2W-MRI and DWI.

per sequence; b-values ranging between b0-b2000) by applying a mono-exponential fit of the signal intensity. Negative ADC values (<0) and ADC values larger than 3 standard deviations from the tumor mean (>mean+3SD) were marked as invalid. Since T2W pixel values are represented on an arbitrary scale, these images were normalized to mean=0 and standard deviation=100 (in line with PyRadiomics documentation [16]). All images were then resampled to a common isotropic pixel spacing of 2x2x2mm.

Image evaluation

In addition to the original staging reports, all MRIs were re-evaluated by a dedicated radiologist (DMJL, pelvic MR expert with >10 years' experience in reading rectal MRI)

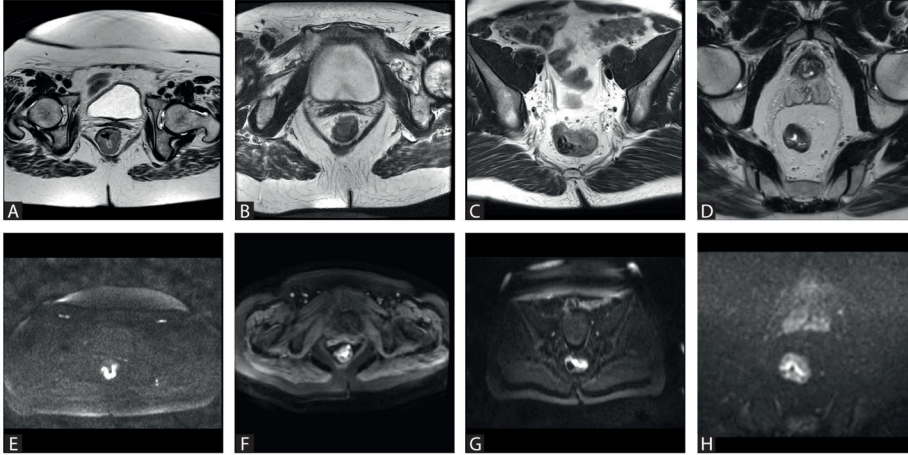


Figure 2. Examples illustrating differences in image quality and acquisition for T2W-MRI (A-D) and DWI (E-H) between centers, related to for example field-of-view, tissue contrast (e.g. TR/TE settings), image resolution, and noise. For the DWI scans, the highest acquired b-values shown in these examples were b1000 (E), b600 (F), b800 (F) and b1000 (H).

who staged all cases in line with the modern staging guidelines as outlined in the structured report template published by the European Society of Gastrointestinal and Abdominal Radiology (ESGAR) in 2018 [17]. For quantitative image analysis, all rectal tumors were segmented using 3D Slicer (version-4.10.2), using previously published methods [18]. In summary, segmentation was performed semi-automatically using a level-tracing algorithm applied to the high b-value DWI. Segmentations were then manually adjusted by an expert radiologist (DMJL, the same reader who also staged the cases) on a slice-by-slice level taking into account the anatomical information from the corresponding T2W-MRI. Care was taken to include only tumor tissue on the DWI and T2W images, excluding the rectal lumen and any non-tumor perirectal tissues. Segmentations were then copied from the DWI to the ADC-map and T2W-MRI, after which tumor volume and other quantitative features (first-order features such as mean signal intensity, standard deviation, entropy) were extracted from these images with PyRadiomics (version-3.0) using a bin-width of 5 (T2W-MRI) and 5×10^{-5} (ADC). This bin-width was chosen such that the number of histogram bins was between 30 and 130, as proposed by PyRadiomics documentation [16]. Quantitative features were deliberately limited to simpler volume, and first-order features as these have previously been reported to be most reproducible [19–23] and least dependent on acquisition differences between centers [18].

Variable definitions

Five distinct variable categories were defined:

1. *Non-imaging variables*; these included baseline patient characteristics (age and sex), basic tumor descriptors available from clinical examination and endoscopy (tumor location and basic tumor morphology, e.g. polyp/circular) and the time interval between completion of neoadjuvant CRT and surgery.
2. *Basic image-based staging variables*:
 - (2a) *derived from the original reports*; these included the main clinical staging descriptors that could routinely be derived from the original radiological staging reports: cT-stage (cT1-2, cT3, cT4), cN-stage (cN0, cN1, cN2), and involvement of the mesorectal fascia (MRF).
 - (2b) *derived from expert re-evaluation*; the same descriptors given in (2a) were derived from the re-evaluations performed by the dedicated expert reader.
3. *Advanced image-based staging variables*; these included, in addition to the descriptors given in (2b), advanced staging descriptors that were not routinely available from the original staging reports but were derived from the expert re-evaluations in line with the ESGAR reporting guidelines: tumor length, cT-substage (cT1-2; cT3a,b,c,d; cT4a,b), depth of extramural invasion, and extramural vascular invasion (EMVI).
4. *Quantitative imaging features*; these included tumor volume, which was extracted directly from the whole-tumor segmentations, and the following first-order features that were derived from the pixel values of the rectal tumor for both the T2W-MRI and the ADC maps: mean, median, minimum, maximum, variance, mean absolute deviation, range, robust mean absolute deviation, root mean squared, 10th percentile, 90th percentile, energy, entropy, interquartile range, kurtosis, skewness, total energy and uniformity.

These variable categories were combined into eight variable sets for the statistical analysis as detailed in **Table 1**.

Response outcome

The final treatment response outcome was defined in twofold [8, 24, 25]:

- *Complete response (CR) versus incomplete response*: CR was defined as either a pathological complete response after surgery (pCR; ypT0N0) or a sustained clinical complete response (cCR) with no evidence of a tumor on repeated follow-up MRI and endoscopy for a follow-up period of >2 years in patients undergoing watch-

and-wait. Patients with ypT1-4 disease after surgery were classified as incomplete response.

- *Good response (GR) versus poor response*: the good response group included all patients with a histopathological Mandard's tumor regression grade (TRG) of 1-2 (total and subtotal regression); patients with TRG of 3-5 (moderate, limited and no regression) were classified as poor responders. For the purpose of this study, patients with a sustained cCR for >2 years were considered TRG1. If the pathology report did not explicitly mention a TRG score, the complete pathology reports were reviewed with a dedicated gastrointestinal pathologist (PS with >8 years of experience) to assign a TRG score retrospectively.

Table 1. Variable category definition and variable sets

Variable categories	Features
1. Non-imaging	Age, sex, time between CRT and surgery, tumor morphology (polyp, semicircular, or circular) and tumor height (distal-mid versus proximal-rectosigmoid)
2a. Basic imaging staging (original reports)	cT-stage (cT12, cT3, cT4), cN-stage (cN0, cN1, cN2), involvement of the mesorectal fascia (MRF-, MRF+)
2b. Basic imaging staging (expert re-evaluation)	cT-stage (cT12, cT3, cT4), cN-stage (cN0, cN1, cN2), involvement of the mesorectal fascia (MRF-, MRF+)
3. Advanced imaging staging (expert re-evaluation)	All variables included in 2b (basic imaging staging – expert re-evaluation) + cT-substage (cT12, cT3abcd, cT4ab), extramural invasion depth, EMVI, tumor length
4. Quantitative imaging (derived from T2W-MRI and ADC)	Tumor volume*, mean, median, minimum, maximum, variance, mean absolute deviation, range, robust mean absolute deviation, root mean squared, 10 th percentile, 90 th percentile, energy, entropy, interquartile range, kurtosis, skewness, total energy, uniformity

Variable sets

1. Non-imaging only
2. Non-imaging + basic imaging staging (original reports)
3. Non-imaging + basic imaging staging (expert re-evaluation)
4. Non-imaging + advanced imaging staging (expert re-evaluation)
5. Non-imaging + quantitative imaging
6. Non-imaging + basic imaging staging (original reports) + quantitative imaging
7. Non-imaging + basic imaging staging (expert re-evaluation) + quantitative imaging
8. Non-imaging + advanced imaging staging (expert re-evaluation) + quantitative imaging

* Tumor volume was derived directly from the whole-tumor segmentations.

DEVELOPMENT AND MULTICENTER VALIDATION OF A MULTIPARAMETRIC IMAGING MODEL TO PREDICT TREATMENT RESPONSE IN RECTAL CANCER

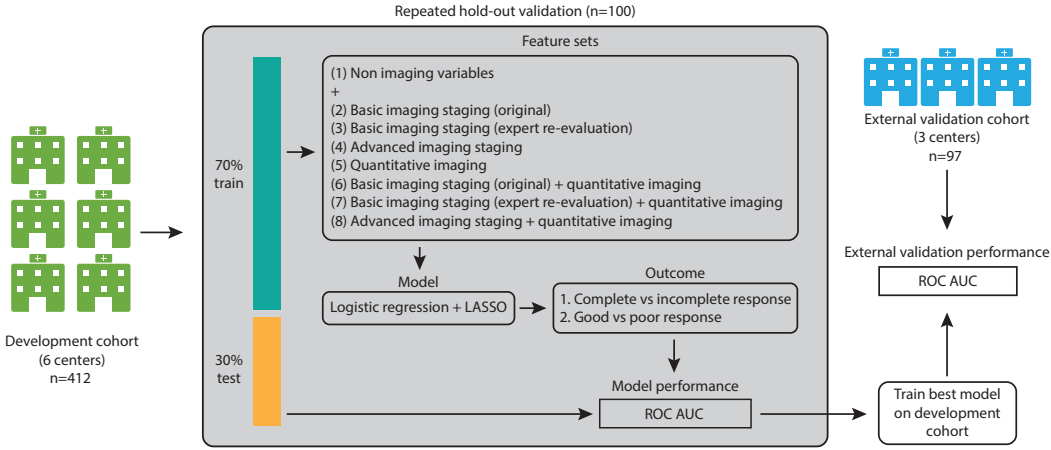


Figure 3. Schematic overview of the study workflow and statistical analysis. From a total cohort of 509 patients from 9 centers, 412 patients (from 6 centers) were used to develop a prediction model to predict two respective outcomes (complete response, good response) using repeated hold-out validation. For both outcomes, the best performing model was tested on an external and independent validation cohort consisting of 97 patients (from 3 different centers)

Statistical analysis

The 9 centers were divided into a development and (external) validation set, including 6 centers (n=412) and 3 centers (n=97), respectively. Differences between development and validation sets were assessed using Chi-squared tests for categorical (sex and response) and Kruskal-Wallis tests for continuous/ordinal variables (age, cT and cN). The model development and validation process is summarized in **Figure 3**. Predictive performance for the eight variable sets (see **Table 1**) to predict the two respective response outcomes (complete vs incomplete response; good vs poor response) was assessed in the development cohort by calculating the average area under the receiver operator characteristic curve (AUC) after repeated (n=100) random hold-out validation. During each iteration, the development cohort was randomly split into a 70% training / 30% test dataset.

All training variables were then scaled (mean = 0, standard deviation = 1), with the same scaling (i.e. using the mean and standard deviation derived from the training set) applied to the test set. When two or more features in a variable set were correlated (with Pearson's $\rho > 0.8$ in the training data), only the feature with the lowest mean absolute correlation was retained for further analysis. The remaining variables were used to train a logistic regression model with Least Absolute Shrinkage and



Selector Operator (LASSO) regularization [26]. The LASSO regularization parameter (λ) was tuned to select only the most relevant variables by minimizing the negative binomial log-likelihood loss using internal repeated ($n=100$) 10-fold cross-validation. Each model's performance was measured on the test dataset, and the model achieving the best average test AUC was trained on the whole development cohort. As a final step, the performance of this best performing model (N.B. one model to predict a complete response and one model to predict a good response) was tested on the external validation cohort. 95% confidence intervals for averaged AUCs in the development data were estimated through bootstrapping (200 samples). Confidence intervals for the validation cohort were obtained using DeLong's method [27].

Supplementary Materials B describes two additional analyses: (1) testing the effects of 3 different previously described methods for multicenter data normalization (using a reference organ [28], statistical correction of imaging features using the ComBat algorithm [29], and statistical correction using mixed-effects models [30]), and (2) comparing model performance in the multicenter dataset to a single-center data subset from the cohort acquired with a harmonized MRI acquisition protocol. The latter was done to mimic comparison of our results with a single center study design.

RESULTS

Patients

Baseline information of the 509 included patients (332 (65%) male; median age 65 (range 25-87) years) is presented in **Table 2**. For the outcome complete (versus incomplete) response, 141 patients (28%) were classified as complete responders. For the outcome good (versus poor) response, 225 patients (44%) were classified as good responders. The development and validation cohort showed no significant differences in sex, age, cT-stage, cN-stage and tumor response ($p=0.37-0.98$).

Model performance and predictive variables

Results for model development and performance are detailed in **Table 3**. The best performing model included non-imaging and advanced imaging staging variables and achieved an average AUC of 0.60 (95%CI 0.48-0.72) to predict a complete response and AUC 0.65 (95% CI 0.53-0.76) to predict a good response in the external validation cohort, results very similar to those obtained during testing in the development cohort. The addition of quantitative imaging features did not lead to improved predictive

Table 2. Baseline patient characteristics and variations between centers

		Total	Development cohort	Validation cohort	P-value
TOTAL, n (%)		n=509 (100%)	n=412 (81%)	n=97 (19%)	
Sex, n (%)	Female	177 (35%)	139 (34%)	38 (39%)	0.37*
	Male	332 (65%)	273 (66%)	59 (61%)	
Age, median (range)		65 (25-87)	66 (25-87)	65 (33-81)	0.37**
cT, n (%)	1-2	35 (7%)	30 (7%)	5 (5%)	0.57**
	3	441 (81%)	334 (81%)	80 (83%)	
	4	60 (12%)	48 (12%)	12 (12%)	
cN, n (%)	0	68 (13%)	52 (13%)	16 (17%)	0.98**
	1	122 (24%)	103 (25%)	19 (20%)	
	2	319 (63%)	257 (62%)	62 (64%)	
Complete response, n (%)	CR	141(28%)	111 (27%)	30 (31%)	0.51*
	Not-CR	368 (72%)	301 (73%)	67 (69%)	
Good response, n (%)	Good	225 (44%)	184 (45%)	41 (42%)	0.75*
	Poor	284 (56%)	228 (55%)	56 (58%)	

*Calculated using Chi-squared test

** Calculated using Kruskal-Wallis test

performance in any of the model combinations. Basic staging variables consistently achieved lower predictive performance compared to the advanced staging variables, especially (though 95% confidence intervals showed some overlap) when the basic staging variables were derived from the original reports. Based on the model coefficients, a more proximal tumor location, shorter tumor length, longer waiting interval after CRT, lower cT-substage and cN-stage, negative MRF, lower extramural invasion depth and negative EMVI status were associated with a favorable response outcome (full model coefficients are provided in **Supplementary Materials C**).

The results of our supplementary analyses (**Supplementary Materials B**) show that none of the normalization methods applied to retrospectively harmonize the data improved the predictive performance of our models. When mimicking a single-center study design (i.e. when performing the same analysis on a single-center subset within our cohort with homogeneous imaging protocols), results were highly variable. However, there was a trend towards better single-center model performance for most variable subsets to predict a complete response.

The best performing single-center model (including non-imaging and advanced staging variables) achieved an AUC of 0.79, compared to an AUC of 0.69 achieved in the total multicenter development cohort.

Table 3. Model performance

Average AUC on the development cohort		
Variable groups and combinations	Outcome	
	CR (95% CI)	GR (95% CI)
Non-imaging	0.58 (0.49 – 0.66)	0.53 (0.42 – 0.58)
Non-imaging + basic imaging staging (original reports)	0.63 (0.55 – 0.70)	0.52 (0.39 – 0.54)
Non-imaging + basic imaging staging (expert re-evaluation)	0.66 (0.58 – 0.70)	0.62 (0.56 – 0.68)
Non-imaging + advanced imaging staging (expert re-evaluation)	0.69 (0.62 – 0.74)	0.67 (0.62 – 0.73)
Non-imaging + quantitative imaging	0.59 (0.46 – 0.61)	0.58 (0.47 – 0.61)
Non-imaging + basic imaging staging (original reports) + quantitative imaging	0.59 (0.44 – 0.60)	0.57 (0.44 – 0.59)
Non-imaging + basic imaging staging (expert re-evaluation) + quantitative imaging	0.63 (0.51 – 0.68)	0.62 (0.53 – 0.68)
Non-imaging + advanced imaging staging (expert re-evaluation) + quantitative imaging	0.68 (0.59 – 0.71)	0.67 (0.61 – 0.72)
Performance of best performing model on external validation cohort		
Non imaging + advanced imaging staging (expert re-evaluation)	0.60 (0.48 – 0.72)	0.65 (0.53 – 0.76)
Features selected in CR model: [Intercept], tumor height, weeks to surgery, cTsub-stage, cN-stage, invasion depth (mm), tumor length (mm)		
Features selected in GR model: Tumor height, weeks to surgery, cTsub-stage, MRF-status, invasion depth (mm), EMVI status		
95% CI: 95% confidence interval, CR: Complete response (pCR and cCR), GR: Good response (TRG1-2)		

NB: Confidence intervals on the development cohort AUC are based on the non-studentized pivotal bootstrap method [31] using 200 bootstrap samples. For the external validation cohort DeLong's method [27] was used.

DISCUSSION

This multicenter study shows that when combining clinical baseline variables with various image-based staging variables and quantitative imaging features, overall model performance to predict neoadjuvant treatment response in rectal cancer is disappointing, with AUCs ranging between 0.60 and 0.65 when tested in an external validation cohort to predict either a complete response (ypT0) or a good response (TRG1-2). The best model performance was achieved when combining clinical baseline information and image-based staging variables. Quantitative imaging variables derived from T2W-MRI and ADC maps had no added value. Notably, model performance was considerably better when performing state-of-the-art staging, including advanced staging parameters such as cT-substage, extramural invasion depth and EMVI, compared to more traditional staging including only simplified cTN-stage and MRF involvement. Moreover, model performance seemed to be affected by staging variations between observers with better performance when staging was performed by a dedicated reader compared to the original staging reports acquired by a multitude of readers.

The latter is an interesting finding as previous studies included staging variables such as cTN-stage as part of the “baseline patient variables”, which suggest that these are used as ‘objective’ variables with little variation between readers [32, 33]. While measurement variations are commonly considered when analyzing quantitative imaging data, our current results demonstrate that the more basic staging variables are also subject to inter-observer variations, considering the differences in model performance when using the same variables derived from either the original staging reports or expert re-evaluation of the same images. The improved model performance when including also modern staging variables such as cT-substage and EMVI in these expert re-evaluations further highlight the importance of high-quality diagnostic staging according to up-to-date guidelines such as the ESGAR rectal staging template [17]. The clinical impact of state-of-the-art staging was also demonstrated by Bogveradze et al., who showed in a retrospective analysis of 712 patients that compared to “traditional” staging methods, advanced staging according to recent guideline updates would have led to a change in risk classification (and therefore potentially in treatment stratification) in up 18% of patients [34].

The fact that our cohort dates back as far as 2008 and covers a 10-year inclusion period explains why many of these advanced staging variables could mostly not be derived

from the original reports. The use of older data will likely also have impacted the quality of the images and thus the quantitative imaging features derived from the data. Following developments in acquisition guidelines, software and hardware updates, image quality will have evolved over time. This is also reflected by the large number of different imaging protocols (including 112 unique T2W and 94 unique DWI protocols) derived from the 9 participating centers over time. The question, therefore, remains if and how model performance would have improved using only current-day state-of-the-art and/or more harmonized (prospectively acquired) MRI data.

In our current dataset quantitative imaging features showed no added benefit for the purpose of pre-treatment response prediction. This is contrary to different previous single center and smaller bi- and tri-institutional studies that achieved more encouraging AUCs ranging from 0.63 to as high as 0.97 using quantitative imaging features derived from MRI [5, 14, 15]. These previous results are likely at least in part an overestimation of how such models would perform in everyday practice, as especially earlier pilot studies are hampered by limitations in methodological design (e.g., small patient cohorts, re-using of training data for testing, multiple testing, etc.) as also outlined in several review papers reporting on the quality and/or reproducibility of image biomarker studies [5, 19, 35–39].

The fact that most previous studies have been single-center reports will likely have also played an important role. Though reflective of data acquired in everyday practice, our results confirm the known difficulty of building generally applicable prediction models using heterogeneous retrospectively collected multicenter data. While some data variations are necessary to identify robust features to vendor and acquisition differences, too much variation will negatively impact model generalizability. Attempting to directly compare and investigate the effects of multicenter (heterogeneous) versus single-center (homogeneous) modelling using our own data, we mimicked a single center comparison by repeating our study analyses on a homogeneous single-center subset within our cohort. Though results have to be interpreted with caution considering the wider confidence intervals and lack of external validation in the single-center arm, this comparison suggests that the best performing model indeed appeared to be better for the homogeneous single-center subset (AUC 0.79) than for the multicenter (AUC 0.69) cohort. Though full data harmonization will likely never be achieved in daily clinical practice, these findings do support a need for further protocol guidelines and standardization to benefit future multicenter research.

There are some limitations to our study design. As mentioned above, data was acquired over the time span of a decade including scans acquired using outdated protocols dating back as far as 2008. All segmentations were performed on high b-value DWI and then copied to T2W-MRI and ADC maps. Although care was taken to include anatomical information of the T2W-MRI during segmentation, ideally a separate segmentation would have been performed. Finally, comparison between the original basic staging reports and the advanced staging performed as part of this study was influenced by the fact that all re-evaluations were done by a single reader. In contrast, original staging reports were performed by a multitude of readers with varying levels of expertise.

In conclusion, this multicenter study combining clinical information and MRIs acquired as part of everyday clinical practice over the time span of a decade rendered disappointing performance to predict response to neoadjuvant treatment in rectal cancer. Best results were obtained when combining clinical baseline information with state-of-the-art image-based staging variables, highlighting the importance of good quality staging according to current guidelines and staging templates. No added value was found for quantitative imaging features in this multicenter retrospective study setting. This is likely at least in part the result of acquisition variations, which is a major problem for feature reproducibility and thus model generalizability. To benefit from quantitative imaging features—assuming a predictive potential—further optimization and harmonization of acquisition protocols will be essential to reduce feature variation across centers. For future research it would also be interesting to see how model performance may improve when combining the information that can be derived from imaging with other clinical biomarkers such as molecular markers (e.g. DNA mutations, gene expression, microRNA) [40, 41], blood biomarkers (e.g. CEA, circulating tumor DNA) [40, 42], metabolomics (e.g. metabolites, hormones and other signaling molecules) [43], organoids [44] and immune profiling [45].



REFERENCES

1. Aklilu M, Eng C (2011) The current landscape of locally advanced rectal cancer. *Nat Rev Clin Oncol* 8:649–659. <https://doi.org/10.1038/nrclinonc.2011.118>
2. Maas M, Nelemans PJ, Valentini V, et al (2010) Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: A pooled analysis of individual patient data. *Lancet Oncol* 11:835–844. [https://doi.org/10.1016/S1470-2045\(10\)70172-8](https://doi.org/10.1016/S1470-2045(10)70172-8)
3. López-Campos F, Martín-Martín M, Fornell-Pérez R, et al (2020) Watch and wait approach in rectal cancer: Current controversies and future directions. *World J Gastroenterol* 26:4218–4239. <https://doi.org/10.3748/wjg.v26.i29.4218>
4. van der Valk MJM, Hilling DE, Bastiaannet E, et al (2018) Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multicentre registry study. *Lancet* 391:2537–2545. [https://doi.org/10.1016/S0140-6736\(18\)31078-X](https://doi.org/10.1016/S0140-6736(18)31078-X)
5. Staal FCR, van der Reijdt DJ, Taghavi M, et al (2021) Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer: A Systematic Review. *Clin Colorectal Cancer* 20:52–71. <https://doi.org/10.1016/j.clcc.2020.11.001>
6. Huang Y, Lee D, Young C (2020) Predictors for complete pathological response for stage II and III rectal cancer following neoadjuvant therapy - A systematic review and meta-analysis. *Am J Surg* 220:300–308. <https://doi.org/10.1016/j.amjsurg.2020.01.001>
7. Fischer J, Eglinton TW, Richards SJG, Frizelle FA (2021) Predicting pathological response to chemoradiotherapy for rectal cancer: a systematic review. *Expert Rev Anticancer Ther* 21:489–500. <https://doi.org/10.1080/14737140.2021.1868992>
8. Schurink NW, Lambregts DMJ, Beets-Tan RGH (2019) Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. *Br J Radiol* 92:20180655. <https://doi.org/10.1259/bjr.20180655>
9. Joye I, Deroose CM, Vandecaveye V, Haustermans K (2014) The role of diffusion-weighted MRI and 18F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: A systematic review. *Radiother Oncol* 113:158–165. <https://doi.org/10.1016/j.radonc.2014.11.026>
10. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2011) Rectal Cancer: Assessment of Complete Response to Preoperative Combined Radiation Therapy with Chemotherapy—Conventional MR Volumetry versus Diffusion-weighted MR Imaging. *Radiology* 260:734–743. <https://doi.org/10.1148/radiol.11102467>

11. Lambregts DMJ, Rao S-X, Sassen S, et al (2015) MRI and Diffusion-weighted MRI Volumetry for Identification of Complete Tumor Responders After Preoperative Chemoradiotherapy in Patients With Rectal Cancer. *Ann Surg* 262:1034–1039. <https://doi.org/10.1097/SLA.0000000000000909>
12. Ha HI, Kim AY, Yu CS, et al (2013) Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after preoperative chemoradiation therapy. *Eur Radiol* 23:3345–3353. <https://doi.org/10.1007/s00330-013-2936-5>
13. Dijkhoff RAP, Beets-Tan RGH, Lambregts DMJ, et al (2017) Value of DCE-MRI for staging and response evaluation in rectal cancer: A systematic review. *Eur J Radiol* 95:155–168. <https://doi.org/10.1016/j.ejrad.2017.08.009>
14. van Griethuysen JJM, Lambregts DMJ, Trebeschi S, et al (2020) Radiomics performs comparable to morphologic assessment by expert radiologists for prediction of response to neoadjuvant chemoradiotherapy on baseline staging MRI in rectal cancer. *Abdom Radiol* 45:632–643. <https://doi.org/10.1007/s00261-019-02321-8>
15. Antunes JT, Ofshteyn A, Bera K, et al (2020) Radiomic Features of Primary Rectal Cancers on Baseline T2-Weighted MRI Are Associated With Pathologic Complete Response to Neoadjuvant Chemoradiation: A Multisite Study. *J Magn Reson Imaging* 52:1531–1541. <https://doi.org/10.1002/jmri.27140>
16. Van Griethuysen JJM, Fedorov A, Parmar C, et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
17. Beets-Tan RGH, Lambregts DMJ, Maas M, et al (2018) Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 28:1465–1475. <https://doi.org/10.1007/s00330-017-5026-2>
18. Schurink NW, van Kranen S, Roberti S, et al (2021) Sources of variation in multicenter rectal MRI data and their effect on radiomics feature reproducibility. *Eur Radiol*. <https://doi.org/https://doi.org/10.1007/s00330-021-08251-8>
19. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol* 102:1143–1158. <https://doi.org/10.1016/j.ijrobp.2018.05.053>
20. Traverso A, Kazmierski M, Shi Z, et al (2019) Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing. *Phys Medica* 61:44–51. <https://doi.org/10.1016/j.ejmp.2019.04.009>

21. Fiset S, Welch ML, Weiss J, et al (2019) Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol* 135:107–114. <https://doi.org/10.1016/j.radonc.2019.03.001>
22. Yuan J, Xue C, Lo G, et al (2021) Quantitative assessment of acquisition imaging parameters on MRI radiomics features: a prospective anthropomorphic phantom study using a 3D-T2W-TSE sequence for MR-guided-radiotherapy. *Quant Imaging Med Surg* 11:1870–1887. <https://doi.org/10.21037/qims-20-865>
23. Mi H, Yuan M, Suo S, et al (2020) Impact of different scanners and acquisition parameters on robustness of MR radiomics features based on women's cervix. *Sci Rep* 10:20407. <https://doi.org/10.1038/s41598-020-76989-0>
24. Xie H, Sun T, Chen M, et al (2015) Effectiveness of the Apparent Diffusion Coefficient for Predicting the Response to Chemoradiation Therapy in Locally Advanced Rectal Cancer. *Medicine (Baltimore)* 94:e517. <https://doi.org/10.1097/MD.0000000000000517>
25. Maffione AM, Marzola MC, Capirci C, et al (2015) Value of 18 F-FDG PET for Predicting Response to Neoadjuvant Therapy in Rectal Cancer: Systematic Review and Meta-Analysis. *Am J Roentgenol* 204:1261–1268. <https://doi.org/10.2214/AJR.14.13210>
26. Tibshirani R (1994) Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 58:267–288
27. DeLong ER, Carolina N (1988) Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves : A Nonparametric Approach. *Biometrics* 44:837–845
28. Koc Z, Erbay G, Karadeli E (2017) Internal comparison standard for abdominal diffusion-weighted imaging. *Acta radiol* 58:1029–1036. <https://doi.org/10.1177/0284185116681040>
29. Orhac F, Lecler A, Savatovski J, et al (2021) How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol* 31:2272–2280. <https://doi.org/10.1007/s00330-020-07284-9>
30. Kahan BC (2014) Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? *BMC Med Res Methodol* 14:20. <https://doi.org/10.1186/1471-2288-14-20>
31. Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 19:1141–1164. [https://doi.org/10.1002/\(SICI\)1097-0258\(20000515\)19:9<1141::AID-SIM479>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F)

32. Yang C, Jiang Z-K, Liu L-H, Zeng M-S (2020) Pre-treatment ADC image-based random forest classifier for identifying resistant rectal adenocarcinoma to neoadjuvant chemoradiotherapy. *Int J Colorectal Dis* 35:101–107. <https://doi.org/10.1007/s00384-019-03455-3>
33. Yi X, Pei Q, Zhang Y, et al (2019) MRI-Based Radiomics Predicts Tumor Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Front Oncol* 9:1–10. <https://doi.org/10.3389/fonc.2019.00552>
34. Bogveradze N, el Khababi N, Schurink NW, et al (2021) Evolutions in rectal cancer MRI staging and risk stratification in The Netherlands. *Abdom Radiol*. <https://doi.org/https://doi.org/10.1007/s00261-021-03281-8>
35. Chalkidou A, O'Doherty MJ, Marsden PK (2015) False discovery rates in PET and CT studies with texture features: A systematic review. *PLoS One* 10:1–18. <https://doi.org/10.1371/journal.pone.0124165>
36. Zwanenburg A (2019) Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging* 46:2638–2655. <https://doi.org/10.1007/s00259-019-04391-8>
37. Sanduleanu S, Woodruff HC, de Jong EEC, et al (2018) Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiother Oncol* 127:349–360. <https://doi.org/10.1016/j.radonc.2018.03.033>
38. Keenan KE, Delfino JG, Jordanova K V., et al (2021) Challenges in ensuring the generalizability of image quantitation methods for MRI. *Med Phys* mp.15195. <https://doi.org/10.1002/mp.15195>
39. Hagiwara A, Fujita S, Ohno Y, Aoki S (2020) Variability and Standardization of Quantitative Imaging. *Invest Radiol* 55:601–616. <https://doi.org/10.1097/RLI.0000000000000666>
40. Dayde D, Tanaka I, Jain R, et al (2017) Predictive and Prognostic Molecular Biomarkers for Response to Neoadjuvant Chemoradiation in Rectal Cancer. *Int J Mol Sci* 18:573. <https://doi.org/10.3390/ijms18030573>
41. El Sissy C, Kirilovsky A, Van den Eynde M, et al (2020) A Diagnostic Biopsy-Adapted Immunoscore Predicts Response to Neoadjuvant Treatment and Selects Patients with Rectal Cancer Eligible for a Watch-and-Wait Strategy. *Clin Cancer Res* 26:5198–5207. <https://doi.org/10.1158/1078-0432.CCR-20-0337>
42. Massihnia D, Pizzutilo EG, Amatu A, et al (2019) Liquid biopsy for rectal cancer: A systematic review. *Cancer Treat Rev* 79:101893. <https://doi.org/10.1016/j.ctrv.2019.101893>

43. Jia H, Shen X, Guan Y, et al (2018) Predicting the pathological response to neoadjuvant chemoradiation using untargeted metabolomics in locally advanced rectal cancer. *Radiother Oncol* 128:548–556. <https://doi.org/10.1016/j.radonc.2018.06.022>
44. Flood M, Narasimhan V, Wilson K, et al (2021) Organoids as a Robust Preclinical Model for Precision Medicine in Colorectal Cancer: A Systematic Review. *Ann Surg Oncol*. <https://doi.org/10.1245/s10434-021-10829-x>
45. Yuan Z, Frazer M, Ahmed KA, et al (2020) Modeling precision genomic-based radiation dose response in rectal cancer. *Futur Oncol* 16:2411–2420. <https://doi.org/10.2217/fo-2020-0060>

DEVELOPMENT AND MULTICENTER VALIDATION OF A MULTIPARAMETRIC IMAGING
MODEL TO PREDICT TREATMENT RESPONSE IN RECTAL CANCER

6



General discussion

GENERAL DISCUSSION

The aim of this thesis was to study the clinical role of multiparametric imaging to help predict neoadjuvant treatment response in rectal cancer and to determine the optimal combination of clinical predictors, functional imaging ‘biomarkers’ and modern post-processing methods.

Importance of good quality radiological imaging and staging

Chapters 3, 4 and 6 consistently show that clinical-radiological staging parameters were the most significant predictors for response, outperforming the advanced quantitative imaging features studied in this thesis. Routine clinical staging parameters such as tumour stage (cT-stage), nodal stage (cN-stage), invasion depth and mesorectal fascia involvement are thus not only crucial for risk- and treatment stratification, but also have value for prediction of response. However, our results from Chapter 6 also show that the predictive performance of these staging variables can vary depending on visual staging quality, with better model performance when staging was performed by dedicated radiologists using current state-of-the-art reporting guidelines. This is a factor that has commonly been neglected in previous works, where variables such as cT- and cN-stage have been included as “baseline” patient variables not taking into account inter-reader variations in staging. In line with our findings, a recent report by Bogveradze et al. showed that dedicated expert staging using up to date guideline criteria resulted in significant downstaging of cT-stage and cN-stage compared to original radiological reports performed by multiple readers and using older guidelines. This resulted in significant risk migration which would retrospectively have altered treatment stratification in up to 18% of the patients [1]. These results clearly highlight the importance of having dedicated radiologists involved in rectal cancer management, both for treatment stratification and prognostic modeling. Moreover, it underscores the importance of dedicated radiological training and investment in guideline implementation.

Image quality was another important factor that affected the predictive performance of our models. In our multicenter cohort described in Chapters 5 and 6, there were substantial differences in acquisition protocols between—but also within—centers. Differences in for example in-plane resolution, slice-thickness, signal-to-noise ratio, number of signal averages, and DWI b-values resulted in significant variations in quantitative imaging features but also in overall (visual) scan quality between centers. Our results from Chapter 6 suggest that these differences also indirectly affected the

clinical-radiological staging variables. When the predictive modelling was repeated on a single-center subset of patients with consistent imaging settings and good quality images, staging variables achieved a substantially higher predictive performance compared to the full heterogeneous multicenter dataset. This emphasizes the importance of using optimized imaging protocols. It also suggests that current MR guidelines (that mainly include basic recommendations on slice thickness and sequence angulation [2, 3]) should be expanded with more specific recommendations on other scan parameters affecting image quality.

Quantitative MRI variables

Quantitative ('Radiomics') MR imaging parameters were of limited predictive value in the various predictive models investigated in this thesis. In Chapter 4, some predictive value was found for DWI texture variables, but only when included in an "imaging-only" prediction model. When combined with other clinical-radiological variables (cT-stage, cN-stage, age, sex and interval between CRT and final response evaluation), the predictive benefit of quantitative DWI analysis was negligible. Our review on the role of DWI for the management of rectal cancer (Chapter 2) also showed conflicting results. While several studies reported a potential value for pre-treatment DWI analysis (in particular ADC) to predict neoadjuvant treatment response [4–13], others were not able to reproduce these findings [14–20]. Moreover, there were large variations in reported ADC values and response threshold values in the different studies. We found similar large variations in ADC values in our own multicenter dataset. Our analysis in Chapter 5 showed that 64% of these ADC variations were related to acquisition and hardware differences between centers, whereas only 0.4% of the variations in ADC could be attributed to a correlation with clinical outcomes such as treatment response. This demonstrates that the influence of imaging variations is substantially larger than any relation between ADC and treatment outcome, which also explains why ADC had limited predictive value in our multicenter analysis (Chapter 6). In the current clinical setting — with inherent variations in data between centers — ADC will therefore likely have a limited role as a biomarker for response.

The same conclusion can be drawn for the other MRI features studied in this thesis. Even though we deliberately chose to include mainly basic features that are generally regarded as the most reproducible (e.g. mean, min, max, standard deviation, entropy and volume) [21–23], these features still showed significant variations between centers (Chapter 5). Several normalization techniques were used attempting to correct for these variations. Nevertheless, their predictive value in our multicenter models remained

disappointing with AUCs not exceeding 0.69. Acquisition variations are thus a serious problem when building image-based prediction models, but are also the current clinical reality. Not only are there inherent differences between images acquired using different vendors and field strengths, acquisition protocols will likely also continue to vary because centers continuously update and optimize their own protocols. While efforts such as specialized phantoms for quality assessment of Radiomics features [24], best-practice guidelines on how to perform quantitative imaging research [25–29], and ongoing developments in statistical post-processing methods [30] may reduce these variations to some extent, it remains questionable whether these solutions will ultimately be sufficient to implement MR imaging features as biomarkers for response prediction in everyday clinical practice. It could very well be that prediction models developed in one center (with specific scanner and acquisition settings) will prove to be center-specific and cannot be generalized. When designing prospective clinical trials integrating the use of imaging biomarkers in the decision process, standardization of scan protocols, as well as defining specific requirements with respect to field strength and further hardware and software will be of the utmost importance.

Combination of PET/CT and MRI

In chapters 3 and 4 we studied the potential complementary value of combining FDG-PET/CT and MRI parameters to predict response. Results indicated no added value for PET/CT variables when combined with clinical-radiological staging and quantitative MRI variables. Again, clinical staging variables (in specific cT-stage) showed the best predictive performance. Chapter 4 did indicate a potential benefit for performing a more in depth analysis of local texture features on both PET and MRI to assess response, but only when analyzed in image-only prediction modes. In a clinical setting, i.e. when combined in a multiparametric model together with other staging variables, these positive effects were negligible.

Our disappointing findings for combining PET and MRI to predict response are in line with a recent review by Min et al. who reported that with current generation stand-alone MRI and PET techniques there seems to be limited value in combining quantitative PET and MR parameters in prediction models for abdominal tumors. Of the 14 studies that were included in this review (including 7 studying rectal cancer), only 6/14 showed a potential value for combining these two techniques [31]. Given that PET/CT appears to have little predictive value for prediction of response, there is currently no hard argument to include PET/CT in the primary work up of rectal cancer.

Future directions for research

Even though our single-center results were more encouraging, the moderate predictive performance of our multicenter model was too low to base actual clinical decision making on. This is likely at least in part explained by the large variations in the imaging data. Therefore, future multicenter studies should be conducted prospectively and with harmonized imaging protocols. Whether this leads to improved prediction models remains an open question, as it is also possible that the studied imaging features have no predictive value at all.

The search for biomarkers that can be used in decision-support tools to personalize neoadjuvant treatment in rectal cancer thus continues and should also be expanded to other fields. Promising options include circulating biomarkers (e.g. CEA, microRNA, circulating tumor DNA) [32, 33], gene expression profiles [34] and gut microbiome components [35]. Integration of these techniques with imaging biomarkers to improve predictive models is a field that should be further explored. In our studies, imaging techniques were limited to methods that are already applied in clinics (i.e. T2/DWI MRI and PET/CT), but exciting advancements such as MRI fingerprinting, susceptibility imaging, magnetic transfer, metabolic spectroscopy and photon counting CT may also provide new ways to characterize tissue [36].

In this thesis we have only looked at baseline imaging. Given that tumor response is a dynamic process, it would also be interesting to investigate the evolution of imaging features early after the onset of treatment (e.g. in the first or second week). Such longitudinal response monitoring has the added benefit that an 'internal' normalization is performed (by comparing imaging to baseline). These "relative" measurements comparing two time points (often referred to as Δ -measurements) typically show less variations between centers than absolute measurements taken at a single time point. Literature suggests that PET/CT and parametric MRI performed during treatment may have potential to monitor response early after onset of CRT [7, 8, 12, 37]. Novel technologies such as the MR Linear Accelerator (MR-Linac) that integrate MRI with a radiotherapy machine too enables day-to-day imaging during treatment, this is a subject of research with potential clinical impact.

Finally, artificial intelligence solutions such as deep learning may offer new ways to handle imaging differences [38]. The potential of deep learning was for example demonstrated in a recent study that developed a method called "DeepHarmony" which can harmonize the contrast of MR images regardless of the MRI protocol that was used [39].

CONCLUSION

The role of imaging in the management of rectal cancer has evolved substantially over the past decade. Aiming to offer more patients the chance of achieving organ-preservation there is an increased need for accurate prediction of response to help guide neoadjuvant treatment choices. This thesis investigated whether combining imaging with clinical variables into multiparametric prediction models has value. Routine radiological staging variables were found to be most important with the highest predictive performance in all of our studies. Staging quality influenced predictive performance indicating the importance of reader experience and adherence to state-of-the-art guidelines. The predictive value of quantitative imaging variables is limited in the multicenter setting and negatively affected by large variations in acquisition and hardware. Though heterogeneous data will always remain part of the clinical reality, efforts should therefore be taken towards further protocol harmonization and standardization. The complimentary role of PET/CT when combined with clinical staging and MRI for response prediction at baseline appears to be limited.

REFERENCES

1. Bogveradze N, el Khababi N, Schurink NW, et al (2021) Evolutions in rectal cancer MRI staging and risk stratification in The Netherlands. *Abdom Radiol*. <https://doi.org/https://doi.org/10.1007/s00261-021-03281-8>
2. Beets-Tan RGH, Lambregts DMJ, Maas M, et al (2018) Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 28:1465–1475
3. Gollub MJ, Arya S, Beets-Tan RG, et al (2018) Use of magnetic resonance imaging in rectal cancer patients: Society of Abdominal Radiology (SAR) rectal cancer disease-focused panel (DFP) recommendations 2017. *Abdom Radiol* 43:2893–2902
4. Intven M, Reerink O, Philippens MEP (2013) Diffusion-weighted MRI in locally advanced rectal cancer : pathological response prediction after neo-adjuvant radiochemotherapy. *Strahlenther Onkol* 189:117–22
5. Lambrecht M, Deroose C, Roels S, et al (2010) 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 (Madr)* 49:956–963
6. Jung SH, Heo SH, Kim JW, et al (2012) Predicting response to neoadjuvant chemoradiation therapy in locally advanced rectal cancer: Diffusion-weighted 3 tesla MR imaging. *J Magn Reson Imaging* 35:110–116
7. Sun Y-S, Zhang X-P, Tang L, et al (2010) Locally Advanced Rectal Carcinoma Treated with Preoperative Chemotherapy and Radiation Therapy: Preliminary Analysis of Diffusion-weighted MR Imaging for Early Detection of Tumor Histopathologic Downstaging. *Radiology* 254:170–178
8. Barbaro B, Vitale R, Valentini V, et al (2012) Diffusion-Weighted Magnetic Resonance Imaging in Monitoring Rectal Cancer Response to Neoadjuvant Chemoradiotherapy. *Int J Radiat Oncol* 83:594–599
9. Birlik B, Obuz F, Elibol FD, et al (2015) Diffusion-weighted MRI and MR- volumetry - in the evaluation of tumor response after preoperative chemoradiotherapy in patients with locally advanced rectal cancer. *Magn Reson Imaging* 33:201–212
10. Doi H, Beppu N, Kato T, et al (2015) Diffusion-weighted magnetic resonance imaging for prediction of tumor response to neoadjuvant chemoradiotherapy using irinotecan plus S-1 for rectal cancer. *Mol Clin Oncol* 3:1129–1134

11. Ippolito D, Monguzzi L, Guerra L, et al (2012) Response to neoadjuvant therapy in locally advanced rectal cancer: Assessment with diffusion-weighted MR imaging and 18FDG PET/CT. *Abdom Imaging* 37:1032–1040
12. Jacobs L, Intven M, van Lelyveld N, et al (2016) Diffusion-weighted MRI for Early Prediction of Treatment Response on Preoperative Chemoradiotherapy for Patients With Locally Advanced Rectal Cancer. *Ann Surg* 263:522–528
13. Lambrecht M, Vandecaveye V, De Keyzer F, et al (2012) 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* 82:863–870
14. Curvo-Semedo L, Lambregts DMJ, Maas M, et al (2011) Rectal Cancer: Assessment of Complete Response to Preoperative Combined Radiation Therapy with Chemotherapy—Conventional MR Volumetry versus Diffusion-weighted MR Imaging. *Radiology* 260:734–743
15. Monguzzi L, Ippolito D, Bernasconi DP, et al (2013) Locally advanced rectal cancer: Value of ADC mapping in prediction of tumor response to radiochemotherapy. *Eur J Radiol* 82:234–240
16. Blažić I, Maksimović R, Gajić M, Šaranović Đ (2015) Apparent diffusion coefficient measurement covering complete tumor area better predicts rectal cancer response to neoadjuvant chemoradiotherapy. *Croat Med J* 56:460–469
17. DeVries AF, Kremser C, Hein PA, et al (2003) Tumor microcirculation and diffusion predict therapy outcome for primary rectal carcinoma. *Int J Radiat Oncol* 56:958–965
18. Genovesi D, Filippone A, Ausili Cèfaro G, et al (2013) Diffusion-weighted magnetic resonance for prediction of response after neoadjuvant chemoradiation therapy for locally advanced rectal cancer: Preliminary results of a monoinstitutional prospective study. *Eur J Surg Oncol* 39:1071–1078
19. Kim SH, Lee JY, Lee JM, et al (2011) Apparent diffusion coefficient for evaluating tumour response to neoadjuvant chemoradiation therapy for locally advanced rectal cancer. *Eur Radiol* 21:987–995
20. Kim YC, Lim JS, Keum KC, et al (2011) 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* 34:570–576
21. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol* 102:1143–1158

22. Lee J, Steinmann A, Ding Y, et al (2021) Radiomics feature robustness as measured using an MRI phantom. *Sci Rep* 11:3973
23. Haarbuerger C, Schock J, Truhn D, et al (2019) Radiomic Feature Stability Analysis based on Probabilistic Segmentations
24. Bianchini L, Botta F, Origgi D, et al (2020) PETER PHAN: An MRI phantom for the optimisation of radiomic studies of the female pelvis. *Phys Medica* 71:71–81
25. Shukla-Dave A, Obuchowski NA, Chenevert TL, et al (2019) Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging* 49:e101–e121
26. Zwanenburg A, Vallières M, Abdalah MA, et al (2020) The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* 295:328–338
27. Current challenges, future prospects, and the proposal of a new framework. *Methods* 188:20–29
28. Lambin P, Leijenaar RTH, Deist TM, et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* In press:
29. Hagiwara A, Fujita S, Ohno Y, Aoki S (2020) Variability and Standardization of Quantitative Imaging. *Invest Radiol* 55:601–616
30. Mali SA, Ibrahim A, Woodruff HC, et al (2021) Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. *J Pers Med* 11:842
31. Min LA, Castagnoli F, Vogel W V., et al (2021) A decade of multi-modality PET and MR imaging in abdominal oncology. *Br J Radiol* 94:20201351
32. Bedin C, Crotti S, D'Angelo E, et al (2020) Circulating Biomarkers for Response Prediction of Rectal Cancer to Neoadjuvant Chemoradiotherapy. *Curr Med Chem* 27:4274–4294
33. Dayde D, Tanaka I, Jain R, et al (2017) Predictive and Prognostic Molecular Biomarkers for Response to Neoadjuvant Chemoradiation in Rectal Cancer. *Int J Mol Sci* 18:573
34. Yuan Z, Frazer M, Ahmed KA, et al (2020) Modeling precision genomic-based radiation dose response in rectal cancer. *Futur Oncol* 16:2411–2420
35. Yi Y, Shen L, Shi W, et al (2021) Gut Microbiome Components Predict Response to Neoadjuvant Chemoradiotherapy in Patients with Locally Advanced Rectal Cancer: A Prospective, Longitudinal Study. *Clin Cancer Res* 27:1329–1340

36. van Zijl P, Knutsson L (2019) In vivo magnetic resonance imaging and spectroscopy. Technological advances and opportunities for applications continue to abound. *J Magn Reson* 306:55–65
37. Joye I, Deroose CM, Vandecaveye V, Haustermans K (2014) The role of diffusion-weighted MRI and 18F-FDG PET/CT in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: A systematic review. *Radiother Oncol* 113:158–165
38. Chan H-P, Samala RK, Hadjiiski LM, Zhou C (2020) Deep Learning in Medical Image Analysis. In: *Advances in Experimental Medicine and Biology*. pp 3–21
39. Dewey BE, Zhao C, Reinhold JC, et al (2019) DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. *Magn Reson Imaging* 64:160–170

The image features a blue and white halftone pattern. A white rectangular box is centered horizontally and contains the word "Appendices" in a blue, sans-serif font.

Appendices

Summary / samenvatting

Scientific impact

List of publications

Dankwoord

Curriculum Vitae

SUMMARY

The overall aim of this thesis is to determine which clinical parameters, functional imaging techniques and modern post-processing methods are appropriate for predicting response to chemoradiation therapy in rectal cancer.

Chapter 2 provides a literature review of the applications and clinical benefit of diffusion weighted MRI (DWI) in rectal cancer. In the past 10 years there has been a clear trend from simple qualitative (visual) evaluation of DWI towards increasingly advanced methods of quantitative image analysis. Most of the studies reviewed focus on DWI for determining rectal tumour response after neoadjuvant chemoradiotherapy. In this setting, DWI has particular value for detecting residual tumour in post-radiation fibrosis. The role of DWI for further tumour and nodal staging is less well-defined. Quantitative DWI analysis (particularly measurement of the "Apparent Diffusion Coefficient") may have added value for predicting tumour response and other prognostic outcomes, but protocol standardization is an issue and results will need to be confirmed prospectively and on a larger scale.

In **Chapter 3**, we examine the individual and combined value of commonly used quantitative MRI and FDG-PET/CT variables in combination with baseline clinical staging parameters to build models for predicting rectal tumour response. Based on our analysis, we can conclude that a model incorporating quantitative variables derived from T2-weighted MRI in combination with clinical tumour stage can help identify patients who will respond well to chemoradiotherapy. Adding quantitative variables derived from PET/CT or DWI does not improve the predictive value of the model.

In **Chapter 4**, we focus specifically on texture analysis and investigate whether quantitative variables describing local tumour texture on MRI and FDG-PET/CT can be of added value in predicting tumour response, compared to more global tumour features (such as those examined in Chapter 3). Our analysis shows that in prediction models based solely on imaging, local texture variables (especially those derived from ADC, CT and T2-weighted MRI) appear to have additional predictive value compared to global tumour features. However, when local texture variables are combined with clinical variables (such as tumour stage), local texture features have similar predictive value compared to more traditional global tumour features. This shows that from a clinical perspective these local texture variables have no clear added value when building models to predict response.

Chapter 5 examines in a multicentre MRI dataset the influence of different sources of variation, such as hardware and acquisition settings, segmentation methodology and image quantification software, on the reproducibility of quantitative imaging features. This analysis shows that there are large variations in imaging features between centres. When looking at ADC (one of the most commonly used quantitative imaging parameters) we find that more than 60% of the variation in mean tumour ADC can be linked to variations in hardware and acquisition settings. Differences in segmentation between readers has little influence on feature reproducibility, provided that whole-volume segmentations are performed. Software variations does not affect simpler features, but results in poor reproducibility for more advanced ('higher order') quantitative features. Overall, simpler 'first order' and shape features shows good reproducibility, with better reproducibility for T2-weighted MRI compared to DWI features.

In **Chapter 6**, we integrate findings from previous chapters to develop and validate a multicentre model to predict tumour response based on clinical and quantitative MR image analysis. Our analysis shows that after external validation, the best performing model achieves only a moderate performance to predict tumour response with areas under the curve in the range of 0.60-0.65. We find that mainly clinical baseline and staging variables are predictive of treatment outcome. More advanced staging variables (acquired by expert readers using state-of-the-art staging guidelines) show the best performance, highlighting the need of good quality staging using up to date guidelines. Quantitative imaging variables has no added value, which could probably at least in part be explained by the large variations in these variables between centers as also discussed in chapter 5.

SAMENVATTING

Het doel van deze thesis is om te bepalen welke klinische parameters, functionele imaging technieken en moderne post-processing methodes geschikt zijn voor het voorspellen van respons op chemoradiatietherapie bij endeldarm kanker.

Hoofdstuk 2 geeft een literatuuroverzicht van de toepassingen en klinische meerwaarde van diffusie gewogen MRI (DWI) bij rectumkanker. In de afgelopen 10 jaar is er een duidelijke trend geweest van eenvoudige kwalitatieve (visuele) evaluatie van DWI naar steeds geavanceerdere methoden van kwantitatieve beeldanalyse. De meeste van de studies die we hebben onderzocht richten zich op DWI voor het bepalen van de respons van de endeldarm tumor na neoadjuvante chemoradiotherapie. In deze setting heeft DWI met name waarde voor het opsporen van resttumor in post-bestralingsfibrose. De rol van DWI voor verdere tumor- en nodale stadiëring is minder goed gedefinieerd. Kwantitatieve DWI analyse (in het bijzonder meting van de "Apparent Diffusion Coefficient") kan toegevoegde waarde hebben voor het voorspellen van tumorrespons en andere prognostische uitkomsten, echter is protocol standaardisatie een probleem en zullen de resultaten prospectief en op grotere schaal bevestigd moeten worden.

In **Hoofdstuk 3** onderzoeken we de individuele en gecombineerde waarde van veelgebruikte kwantitatieve MRI en FDG-PET/CT variabelen in combinatie met baseline klinische stageringsparameters om modellen te bouwen voor het voorspellen van endeldarm tumor respons. Op basis van onze analyse kunnen we concluderen dat een model met kwantitatieve variabelen afgeleid van T2-gewogen MRI in combinatie met klinisch tumorstadium kan helpen bij het identificeren van patiënten die goed zullen reageren op chemoradiotherapie. Toevoeging van kwantitatieve variabelen afgeleid van PET/CT of DWI verbetert de voorspellende waarde van het model niet.

In **Hoofdstuk 4** richten we ons specifiek op textuur analyse en onderzoeken we of kwantitatieve variabelen die de lokale tumortextuur op MRI en FDG-PET/CT beschrijven van toegevoegde waarde kunnen zijn bij het voorspellen van tumorrespons in vergelijking met meer globale tumorkenmerken (zoals die zijn onderzocht in hoofdstuk 3). Uit onze analyse blijkt dat in modellen die alleen op beeldvorming zijn gebaseerd, lokale textuurvariabelen (vooral die zijn afgeleid van ADC, CT en T2-gewogen MRI) extra voorspellende waarde lijken te hebben in vergelijking met globale tumorkenmerken. Echter, wanneer lokale textuurvariabelen worden gecombineerd

met klinische variabelen (zoals tumorstadium), hebben lokale textuurkenmerken een vergelijkbare voorspellende waarde in vergelijking met de meer traditionele globale tumorkenmerken. Dit toont aan dat vanuit een klinisch perspectief deze lokale textuur variabelen geen duidelijke toegevoegde waarde hebben bij het bouwen van modellen om respons te voorspellen.

Hoofdstuk 5 onderzoekt in een multicenter MRI dataset de invloed van verschillende bronnen van variatie, zoals hardware en acquisitie instellingen, segmentatie methodologie en beeldkwantificatie software, op de reproduceerbaarheid van kwantitatieve beeldkenmerken. Uit deze analyse blijkt dat er grote variaties in beeldvormingskenmerken bestaan tussen centra. Wanneer we kijken naar ADC (een van de meest gebruikte kwantitatieve beeldkenmerken), zien we dat meer dan 60% van de variatie in de gemiddelde ADC van de tumor in verband kan worden gebracht met variaties in hardware- en acquisitie-instellingen. Verschillen in segmentatie tussen beoordelaars hebben weinig invloed op de reproduceerbaarheid van deze beeldkenmerken, op voorwaarde dat segmentaties van het hele volume worden uitgevoerd. Softwarevariaties hebben geen invloed op eenvoudigere beeldkenmerken, maar leiden tot een slechte reproduceerbaarheid voor meer geavanceerde ("hogere orde") kwantitatieve kenmerken. Over het algemeen laten eenvoudigere 'eerste orde' en vormkenmerken een goede reproduceerbaarheid zien, met een betere reproduceerbaarheid voor T2-gewogen MRI vergeleken met DWI.

In **Hoofdstuk 6** integreren we bevindingen uit voorgaande hoofdstukken om een multicentermodel te ontwikkelen en te valideren waarmee tumorrespons kan worden voorspeld op basis van klinische en kwantitatieve MR-beeldanalyse. Onze analyse laat zien dat na externe validatie het best presterende model slechts matig presteert in het voorspellen van tumorrespons met oppervlakten onder de curve in het bereik van 0,60-0,65. Vooral klinische basislijn- en stadiëringsvariabelen blijken voorspellend te zijn voor het behandelingsresultaat. Meer geavanceerde stadiëringsvariabelen (verkregen door deskundige beoordelaars gebruikmakend van state-of-the-art stadiëringsrichtlijnen) laten de beste prestatie zien, hetgeen de noodzaak onderstreept van een goede kwaliteit van stadiëring gebruikmakend van up to date richtlijnen. Kwantitatieve beeldvormingskenmerken hebben geen toegevoegde waarde, wat waarschijnlijk tenminste gedeeltelijk verklaard kan worden door de grote variaties in deze variabelen tussen centra, zoals ook besproken in hoofdstuk 5.

MAIN AIMS AND OUTCOMES

The standard treatment for patients with locally advanced rectal cancer is chemoradiotherapy (CRT) followed by surgical resection. In a minority of patients ($\pm 20\%$) the tumour completely disappears as a result of CRT. In these “complete responders” surgery may be avoided. This “watch-and-wait” approach has been shown to result in similar survival outcomes as surgery, but with considerably better functional outcome and quality of life. It is impossible to know upfront which patients will achieve a complete response after CRT. If we could predict how well patients will respond to CRT beforehand, this may create opportunities to further personalize and optimize neoadjuvant treatment, for example by intensifying CRT in patients likely to respond, ultimately aiming to increase complete response rates and offer more patients the chance of organ-preservation.

This thesis focuses on investigating the role of different medical imaging techniques and image analysis tools in predicting neoadjuvant treatment response. In addition to ‘anatomical’ imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI), there are to date several ‘functional’ or ‘molecular’ imaging techniques available that can help us visualize biological tumor properties such as cell structure (diffusion weighted imaging (DWI)), tumor perfusion/angiogenesis (dynamic contrast enhanced imaging (DCE)), or metabolic tumor activity (positron emission tomography (PET)). In addition, novel image post-processing tools such as Radiomics can help us to study the detailed imaging phenotype of a tumour lesion on a pixel-per-pixel level, providing us with new insights into for example underlying tumor heterogeneity (i.e. texture analysis). Quantitative parameters derived from functional imaging and image post-processing can be used as imaging biomarkers of disease which in turn may be used to predict clinical outcomes such as treatment response. Although promising results have been reported for various imaging biomarkers in previous studies, evidence mainly comes from small single-center studies focusing on a single imaging or post-processing technique at a time. Results are conflicting and little is known about the complementary value of combining parameters acquired using different imaging techniques, an approach commonly referred to as “multiparametric imaging”. With this thesis we set out to investigate the value of multiparametric imaging to predict response to chemoradiotherapy in rectal cancer, to identify the best predictive imaging biomarkers and evaluate how these should best be combined with other clinical information such as the radiological tumour stage. In addition we have explored whether it is feasible to build prediction models that are reproducible when using heterogeneous multicenter data from everyday clinical practice.

Chapter 2 summarized the role of diffusion-weighted imaging – one of the most commonly used functional imaging techniques in oncology – for the clinical management of rectal cancer, including its role for prediction of response. In Chapters 3 and 4 the information from anatomical (MRI, CT), functional (DWI, PET) and image post-processing techniques (e.g. texture analysis) was combined into a ‘multiparametric’ imaging model to predict response. Best performance was achieved when combining imaging features derived from MRI with clinical staging variables; the added value of PET and CT was limited. In Chapter 5, we investigated how imaging parameters derived from multicenter data – i.e. from MRI scans acquired in different institutions – are affected by variations in scanner hardware, scan protocols, radiologists and image analysis software. We found that functional DWI parameters and more advanced (complex) imaging features are more affected by variations between centers. For this reason we chose to include mainly simpler and better reproducible parameters to develop our clinical response prediction model described in Chapter 6. In this model (similar to the models, described in Chapters 3 and 4) clinical radiological staging variables such as the tumour (T-) and nodal (N-) stage had the highest predictive value, particularly when staging was performed by experienced radiologists using current state-of-the-art clinical guideline criteria. Quantitative imaging parameters contributed little in the multicenter setting, despite efforts to account for data variations between centers.

RELEVANCE

The results presented in this thesis are relevant for future studies that are aimed at developing image-based clinical prediction models. Lessons learned from our multicenter data analyses in Chapters 5 and 6 can offer guidance on how to handle data variations that are inherently present when dealing with everyday clinical imaging data derived from multiple institutions. Though heterogeneous data will always remain part of the clinical reality, efforts should be taken towards further protocol harmonization and standardization.

The results of this thesis are also relevant for the clinical management of rectal cancer patients. Our literature review presented in Chapter 2 can serve as a quick reference for clinicians on the pros and cons of using DWI to help guide clinical treatment management. From Chapters 3 and 4 we have learned that PET/CT does not seem to have any added value when combined with clinical and MRI variables in the setting of pre-treatment response prediction, indicating that there is currently no

incentive to include PET/CT in the primary staging work up of rectal cancer. Finally, the results from Chapters 3, 4 and 6 consistently demonstrate the importance of clinical radiological staging variables as predictors of response. Chapter 6, however, also showed that the predictive performance of these staging variables was dependent on the experience level of the radiologist performing the staging and on whether staging was performed using state-of-the-art staging guidelines. These results highlight the importance of having dedicated radiologists involved in rectal cancer management, and also underscore the importance of dedicated radiological teaching and training, and investment in clinical guideline implementation.

TARGET POPULATION

The results of this thesis are relevant to several groups. Although this thesis focused specifically on rectal cancer, the impact of multicenter imaging variations on multiparametric imaging modelling and lessons learned on how to handle these can also be translated to other organs and outcomes. These findings are thus important for all researchers focused on multiparametric image model development.

Our findings are also relevant for radiologists performing rectal cancer staging, as well as other clinicians dealing with the management of rectal cancer. Our results describe the current role and limitations of imaging to help predict neoadjuvant treatment response. Though the predictive performance of our multicenter model in Chapter 6 was too low to base actual clinical decision making on, the search for biomarkers continues and should be expanded to other fields. If imaging can be integrated with other clinical biomarkers to build stronger clinical prediction models of response, these models could be used as decision support tools to help further personalize treatment in rectal cancer and optimize treatment outcomes.

ACTIVITIES

The results of this thesis have been presented to a wide audience at both national and international conferences, and have been published in peer-reviewed journals. The paper publication of Chapter 3 was awarded the European Radiology ESGAR award 2020. Lessons learned from this thesis are incorporated into follow-up projects on multicenter data modelling to predict other clinical outcomes and investigating other target organs. Moreover, the large multicenter dataset that was acquired as

part of this thesis has been further expanded and is currently being used in several ongoing collaborative international projects focusing on optimizing the diagnostic performance of imaging for the staging and response evaluation of rectal cancer patients. The current thesis and results of these future studies may serve as a basis for future guideline updates.

LIST OF PUBLICATIONS

Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. **Schurink NW**, Lambregts DMJ, Beets-Tan RGH. *Br J Radiol.* 2019; 92(1096):20180655. doi: 10.1259/bjr.20180655

Value of combined multiparametric MRI and FDG-PET/CT to identify well-responding rectal cancer patients before the start of neoadjuvant chemoradiation. **Schurink NW**, Min LA, Berbee M, van Elmpt W, van Griethuysen JJM, Bakers FCH, Roberti S, van Kranen SR, Lahaye MJ, Maas M, Beets GL, Beets-Tan RGH, Lambregts DMJ. *Eur Radiol.* 2020; 30(5):2945-2954. doi: 10.1007/s00330-019-06638-2

Studying local tumour heterogeneity on MRI and FDG-PET/CT to predict response to neoadjuvant chemoradiotherapy in rectal cancer. **Schurink NW**, van Kranen SR, Berbee M, van Elmpt W, Bakers FCH, Roberti S, van Griethuysen JJM, Min LA, Lahaye MJ, Maas M, Beets GL, Beets-Tan RGH, Lambregts DMJ. *Eur Radiol.* 2021; 31(9):7031-7038. doi: 10.1007/s00330-021-07724-0

Sources of variation in multicentre rectal MRI data and their effect on Radiomics feature reproducibility. **Schurink NW**, van Kranen SR, Roberti S, van Griethuysen JJM, Bogveradze N, Castagnoli F, el Khababi N, Bakers FCH, de Bie SH, Bosma GPT, Cappendijk VC, Geenen RWF, Neijenhuis PA, Peterson GM, Veeken CJ, Vliegen RFA, Beets-Tan RGH, Lambregts DMJ. *Eur Radiol.* 2021; doi: 10.1007/s00330-021-08251-8

Development and multicentre validation of a multiparametric imaging model to predict treatment response in rectal cancer. **Schurink NW**, van Kranen S, van Griethuysen JJM, Roberti S, Snaebjornsson P, Bakers FCH, de Bie SH, Bosma GPT, Cappendijk VC, Geenen RWF, Neijenhuis PA, Peterson GM, Veeken CJ, Vliegen RGA, Peters FP, Bogveradze N, el Khababi N, Lahaye MJ, Maas M, Beets GL, Beets-Tan RGH, Lambregts DMJ. *Submitted for publication*

Multiparametric Imaging for the Locoregional Follow-up of Rectal Cancer. Lambregts DMJ, Min LA, **Schurink N**, Beets-Tan RGH. *Colorectal Cancer Rep.* 2020; 16:19-28. doi: 10.1007/s11888-020-00450-7

Evolutions in rectal cancer MRI staging and risk stratification in The Netherlands. Bogveradze N, el Khababi N, **Schurink NW**, van Griethuysen JJM, de Bie S, Bosma G, Cappendijk VC, Geenen RWF, Neijenhuis P, Peterson G, Veeken CJ, Vliegen RFA, Maas M, Lahaye MJ, Beets GL, Beets-Tan RGH, Lambregts DMJ. *Abdom Radiol* (2021). doi: 10.1007/s00261-021-03281-8

DANKWOORD

Vol trots ligt hij er dan, een proefschrift waar 4 jaar hard aan is gewerkt. Maar laten we wel duidelijk zijn, een prachtig team effort, want zonder de hulp van vele collega's en de steun van vrienden en vriendinnen was dit niet mogelijk was geweest. Daarom een poging om jullie hier allen te bedanken voor jullie geweldige hulp!

Allereerst wil ik graag mijn begeleiders **Regina Beets-Tan, Doenja Lambregts en Simon van Kranen** bedanken voor de cruciale rol die zij hebben gespeeld.

Regina, 4 jaar geleden ben ik als onderzoeker bij je begonnen. Een prachtige kans waar ik met veel plezier op terug kijk. Je enthousiasme voor onderzoek is een inspiratie, dag in en dag uit leef jij voor de research, en praat jij vol trots over de PhDers die je begeleid wat echt voor je spreekt. Bedankt voor je hulp en begeleiding.

Doenja, dank je wel dat je m'n co-promotor bent. De artikelen zijn zo veel beter geworden toen ik ze, vaak volledig rood maar wel altijd voorzien van een positieve begeleidende tekst, van je terug kreeg. Ik heb ontzettend veel van je geleerd en vond het erg fijn om met je samen te werken. Hopelijk ben jij na deze 4 jaar het beoordelen van MRIs nog niet zat bent geworden ;), want oef wat heb je veel scans beoordeeld voor m'n studies. Gelukkig heb je daarvoor je nuchtere humor om je er door heen te slaan, die me—net als je heerlijke directe manier van dingen zeggen en de gezelligheid tijdens congressen of team-uitjes— zeker gaan bijblijven. Ontzettend bedankt voor de goede tijd!

Simon, jij kwam na het 1^e jaar als co-promotor bij het team om me op technisch gebied te begeleiden. Eigenlijk voelde het al heel snel alsof je er vanaf het begin bij was. Wekelijks pakten we er een goede kop koffie bij om de voortgang te bespreken, wat we tijdens corona hebben doorgezet. Vaak met een relaxte sfeer, over hoe het ging, maar daarnaast ook altijd met kritische vragen over het onderzoek, of... de planning ;). Die vragen hebben me scherp gehouden en hebben enorm geholpen bij de opzet en uitwerkingen van de verschillende onderzoeken, en dankzij het plannen is alles wel mooi afgekomen. Bedankt voor je inzichten, maar ook zeker voor de goede momenten. Ik heb erg genoten van onze samenwerking!

Leden van de beoordelingscommissie: prof. dr. J.E. Wildberger, dr. S.O. Breukink, prof. dr. W.J. Niessen, prof. dr. C.A.M. Marijnen en prof. dr. F.M. Mottaghy, ook jullie wil ik graag bedank voor jullie tijd en het beoordelen van mijn proefschrift.

Leden van de OOA commissie commissie: prof. dr. M.P.M. Stokkel, dr. K.F.D. Kuhlman en dr. P.J. van Houdt, bedankt voor jullie betrokkenheid bij mijn promotietraject.

Bij elk artikel in deze thesis zijn **vele co-auteurs** betrokken geweest die in vele vormen hebben bijgedragen aan het onderzoek. Zonder jullie waren de verschillende studies absoluut niet mogelijk geweest. Ik wil jullie allen bedanken voor de goede samenwerking, jullie hulp en voor de waardevolle input die jullie hebben geleverd. **Sander**, jou wil ik nog in het bijzonder bedanken voor de tijd die je nam om al m'n lastige statistische vragen te beantwoorden, je uitgebreide hulp en de tijd en energie die je in de analyses hebt gestoken hebben enorm bijgedragen. **Lisa**, PET buddy! Jij ook enorm bedankt. Wat hebben we hard gewerkt aan ons gezamenlijke paper. Daarnaast mogen we ook nog eens op dezelfde dag ons praatje houden, het wordt ongetwijfeld een top dag!

Collega's van het tuinhuis, met heel veel plezier heb ik de afgelopen jaren met jullie samengewerkt. Ik herinner me vele mooie team uitjes met jullie, de OOA retraite op Renesse, wedstrijden bedrijfs volleybal en hockey, maar ook mooie congressen samen en gezellige vrijdag middag borrels. Een ontzettend fijne omgeving waar altijd tijd is voor een praatje, een gezellige lunch, maar ook vooral om een goede kop koffie te halen ;) Want, zoals iedereen weet, is koffie (en thee) de brandstof voor het tuinhuis. Ik hoop jullie zeker nog in de toekomst weer tegen te komen!

Kamergenootjes, ik weet nog goed dat ik begon in 'de Geekroom', en ondanks dat we achter in het tuinhuis zaten, waarschijnlijk by far de meestbezochte kamer. Niet alleen voor het dagelijkse ICT advies (hi Joost!) maar ook vanwege de rijkelijk gevulde snoeppot en gezelligheid. Het was ontzettend fijn om met jullie samen te werken. **Joost**, je bent opgeleid als dokter, maar stiekem geboren als programmeur, vol enthousiasme kan jij vertellen over je code. Het is tekenend voor je hoeveel passie je hier voor hebt en hoe je hiermee anderen weet te inspireren. **Paula**, blij ei, wat hebben we veel gesprekken gehad. Over de carnaval, je PhD, of de je foto's en filmpjes die je aan het sorteren was. Het was altijd gezellig als je er was, je vrolijkheid is aanstekelijk! **Marjaneh**, I really enjoyed your funny jokes and kindness. I hope you don't have too much of a trauma of your bouncing computer screen though (NIEELSSS..). Sorry for

DANKWOORD

rocking my foot so much ;) Thank you for the good times and also teaching me some basic Farsi. To mahshari! **Kay**, in het begin was je nog wat terughoudend, maar al gauw een grote prater. Altijd bereid om te helpen en mee te denken. Bedankt voor de goede gesprekken! **Najim**, door COVID hebben we helaas maar korte periodes samengewerkt, maar altijd prettig! Kamergenootjes, jullie allemaal, enorm bedankt voor de mooie tijd samen!

Joost, met heel veel plezier kijk ik terug op onze tijd samen in het AVL. We hebben vaak gelachen over droge programeer humor, met soms (jouw intentie ;)) een lange XKCD internetspiraal tot gevolg. Bijna dagelijks gingen we koffie halen en vele weken zijn we op pad geweest samen om, zoals je dat zo mooi kon zeggen, lekker te data-slaven. Maar het was zeker niet alleen op werk gezellig, ook daar buiten hebben we mooie borrels, feestjes en spelletjes avonden gehad. Met jouw promotie heb je alvast het goede voorbeeld gegeven. En zoals ze zeggen, goed voorbeeld doet volgen, ik ben blij dat je mijn paranimf wilt zijn!

Jelle, we kennen we elkaar al meer dan 15 jaar en hebben heel wat beleefd met de BCH. Met je enthousiasme ben je altijd een echte sfeermaker, want van slap ouwehoeren tot diepe gesprekken er is altijd wel wat te beleven. Je bent betrokken met anderen en weet over de meest uiteenlopende onderwerpen wat, dus mocht ik het even niet meer weten tijdens m'n promotie dan ben ik er zeker van dat jij er dan vast een mooi verhaal van weet te breiden ;). Ik ben blij dat je m'n paranimf wilt zijn, met jou erbij weet ik zeker dat het een geslaagde promotie gaat worden!

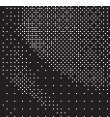
Lieve vrienden en vriendinnen uit Deventer, Enschede en Utrecht, bedankt voor alle welkome afleiding, en de goede momenten samen! Van de vele festivals en feestjes, vakanties, spelletjes avonden en bbq smoke sessies tot kerstdiners, wiskey borrels, lunch wandelingen en sportieve activiteiten, maar ook goede gesprekken afgewisseld met slap geouwehoer. Ik waardeer jullie allemaal enorm en ik hoop dat er nog veel mooie momenten volgen!

Mijn lieve familie, bedankt voor jullie steun door alle jaren heen ik kan me geen betere familie wensen. **Mam**, bedankt voor de basis die je me hebt gegeven, je staat altijd voor me klaar en hebt me, ook als je het ergens niet mee helemaal eens bent, altijd gesteund in m'n keuzes. **Lineke**, mijn grote kleine zusje, ik ben super trots op je en zal dat altijd blijven. Ondanks dat we niet bij elkaar om de hoek wonen ga ik altijd met veel plezier met jou en met **Elmer**, lunchen, wandelen of boulderen. **Oma Els**, leeftijd

zegt niks, je bent immers zo jong als je je voelt. Het is prachtig om te zien hoe energiek en vol plezier u van het leven geniet. Ik hoop dat ik op uw leeftijd nog net zo mooi in het leven mag staan. En **Pap**, je bent er helaas niet meer om mijn verdediging bij te wonen maar ik weet dat je trots op me bent. Je bent en blijft altijd een inspiratie voor me.

Ook mijn (schoon)familie wil ik graag bedanken: **Dick en Anja, Tianne, Ton, Minke en Jouk en Teska**, bedankt voor alle gezelligheid samen, de goede gesprekken en de leuke spelletjes. Het voelt altijd als een warm onthaal als ik bij jullie ben.

Lieke, jij hebt bijna alles van m'n proefschrift meegemaakt, van de mooie momenten tot de worstelingen. Altijd betrokken, lief, zorgzaam en positief, je bent er altijd voor me. Het is ongelooflijk fijn om samen met jou te zijn. We hebben al veel moois meegemaakt samen, maar ik kijk uit naar alle avonturen die we nog gaan beleven. Laten we samen van het leven blijven genieten en er een feestje van maken!



CURRICULUM VITAE

Nicolaas (Niels) Wilhelm Schurink was born on January 29th 1992 in Deventer, the Netherlands. He graduated from secondary school (Ety Hillesum Lyceum) in 2010 after which he started studying Technical Medicine at the University of Twente in Enschede. Besides his study he was involved in multiple extracurricular committees at his student association which included organizing symposia and a Study Tour to India—including 21 business visits with 30 students—where he was the Treasurer (budget 100k€). As part of his master program he did 5 clinical research internships in hospitals throughout the Netherlands. His graduation internship was focused on machine learning and hyperspectral imaging for identifying tumor tissue under the supervision of prof. dr. Theo Ruers in the surgery department of the Netherlands Cancer Institute (NKI).



After receiving his masters' degree in 2017 and inspired by the research in the NKI, Niels started as a PhD-candidate at the radiology department of the NKI under the supervision of prof. dr. Beets-Tan, dr. Lambregts and dr. van Kranen. The focus of his thesis was the prediction of treatment response outcome in rectal cancer patients using modern image analysis techniques and machine learning. Niels has presented his research results in both national (Radiologendagen, Technical innovations in Medicine (TiiM)) and international radiology congresses (European Congress of Radiology (ECR), European Society of Gastrointestinal and Abdominal Radiology (ESGAR), European Multidisciplinary Colorectal Cancer Congress (EMCCC), European Society of Oncologic Imaging (ESOI)). For his research he has been awarded the ECR 2019 best presenter award (Chapter 4 of this thesis) and the ESGAR 2020 best paper award (Chapter 3 of this thesis).

Since December 2021, Niels finished his PhD contract and started as scientific collaboration manager at Siemens Healthineers, where he will be involved in setting up research collaborations to continue to improve healthcare and medical technology.

