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

European Radiology (2020) 30:2945–2954
https://doi.org/10.1007/s00330-019-06638-2

GASTROINTESTINAL

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[1,2] · Lisa A. Min[1,2] · Maaike Berbee[2,3] · Wouter van Elmpt[2,3] · Joost J. M. van Griethuysen[1,2] · Frans C. H. Bakers[4] · Sander Roberti[5] · Simon R. van Kranen[6] · Max J. Lahaye[1] · Monique Maas[1] · Geerard L. Beets[2,7] · Regina G. H. Beets-Tan[1,2] · Doenja M. J. Lambregts[1]

Received: 7 November 2019 / Revised: 5 December 2019 / Accepted: 17 December 2019 / Published online: 7 February 2020

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 analysed. 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 tumour volume ($MTV_{42\%}$) and total lesion glycolysis (TLG). These features, together with sex, age, mrTN-stage ("baseline parameters") and the CRT-surgery interval were analysed 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.

Key Points

- *A multivariable model incorporating the mrT-stage and quantitative features derived from baseline MRI can aid in identifying well-responding patients before the start of neoadjuvant chemoradiotherapy.*
- *mrT-stage was the strongest predictor in the model and was complemented by the tumour volume and signal entropy calculated from T2W-MRI.*
- *Adding quantitative features derived from pre-treatment PET/CT or DWI did not contribute to the model's predictive performance.*

Niels W. Schurink and Lisa A. Min contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06638-2) contains supplementary material, which is available to authorized users.

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

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

[2] GROW School for Oncology and Developmental Biology, University of Maastricht, Maastricht, The Netherlands

[3] Department of Radiation Oncology (Maastro Clinic), Maastricht University Medical Centre, Maastricht, The Netherlands

[4] Department of Radiology, Maastricht University Medical Center, Maastricht, The Netherlands

[5] Department of Epidemiology and Biostatistics, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[6] Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[7] Department of Surgery, The Netherlands Cancer Institute, Amsterdam, The Netherlands

Keywords Rectal neoplasms · Neoadjuvant therapy · Magnetic resonance imaging · Positron emission tomography computed tomography · Logistic models

Abbreviations

ADC	Apparent diffusion coefficient
AIC	Akaike Information Criterion
AUC	Area under the receiver operating characteristic (ROC) curve
CRT	Chemoradiotherapy
CT	Computed tomography
DWI	Diffusion-weighted imaging
FDG	Fluorodeoxyglucose; 2-deoxy-2-[18F]fluoro-D-glucose; 18F-FDG
Gy	Gray
HU	Hounsfield unit
LARC	Locally advanced rectal cancer
LOOCV	Leave-one-out cross-validation
MRI	Magnetic resonance imaging
MTV	Metabolic tumour volume
OR	Odds ratio
PET/CT	Positron-emission tomography/computed tomography
SUV	Standardised uptake value
T2W	T2-weighted
TLG	Total lesion glycolysis
TRG	Tumour regression grade
W&W	Watch-and-wait

Introduction

Current standard treatment for locally advanced rectal cancer (LARC) consists of long-course neoadjuvant 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) [15–20]. 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, 20].

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 to 2015, a cohort of 104 locally advanced (≥ 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-standardised treatment, had insufficient quality imaging or mucinous tumour histology were excluded (see Fig. 1).

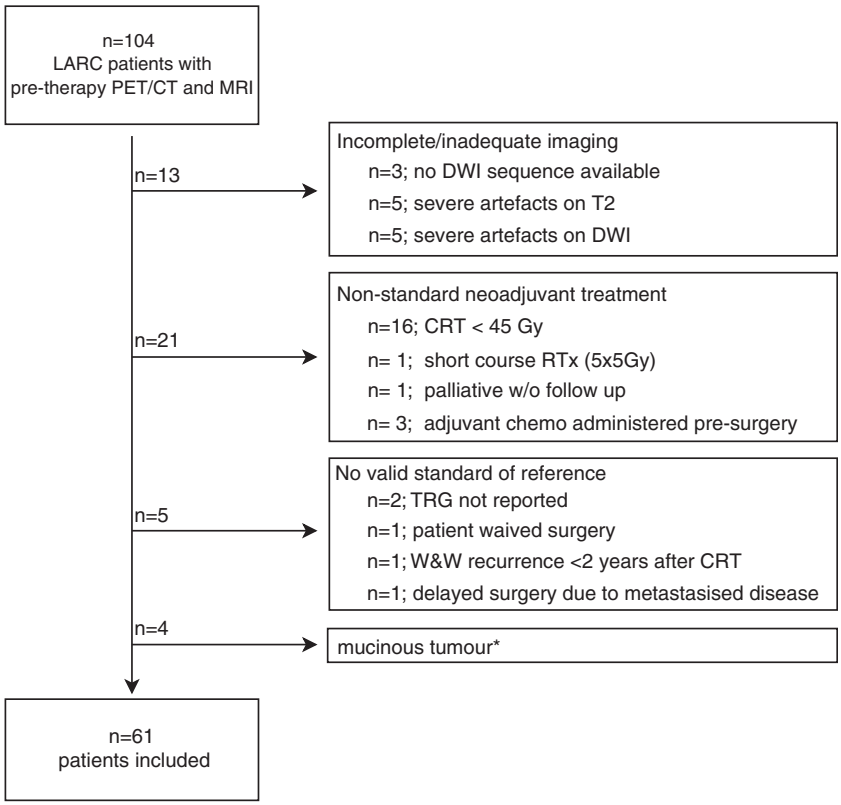

Fig. 1 Patient in- and exclusion flowchart. *CRT* chemoradiotherapy, *LARC* locally advanced rectal cancer (≥ T3 and/or N+), *RTx* radiotherapy, *TRG* tumour regression grade (Mandard's), *W&W* watch-and-wait. * predominantly mucinous tumours were excluded because these typically exhibit distinctly different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

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/mm^2. Apparent diffusion coefficient (ADC) maps were calculated by fitting a mono-exponential decay function to the $b = 0$ and $b = 1000$ s/mm^2 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

^{18}F-FDG-PET/CT was performed on a Siemens Biograph 40 TruePoint PET/CT scanner (SIEMENS medical). A bolus of 2-deoxy-2-[^{18}F]fluoro-D-glucose (^{18}F FDG, from here on: FDG) of 2.5 MBq/kg ($n = 52$) or 4.0 MBq/kg ($n = 9$) was administered intravenously, after a 6-h fast (blood glucose level < 10 mmol/L). Scanning started after an incubation time of 60 (± 5) min, with 5 min per bed position, and ran from the skull base to upper-thighs (reconstructed to 3 mm slice thickness, 4.07 mm 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 3 mm slice thickness, 0.98 mm in-plane resolution).

Quantitative MRI and PET/CT parameters

The image analysis workflow is illustrated in Fig. 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) [21].

A board-certified radiologist (D.L., > 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 ($T2W_{volume}$, mesh-volume in PyRadiomics), entropy of the T2W signal intensity histogram ($T2W\text{-}signal_{entropy}$), 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

Table 1 MRI protocol

	T2-weighted	Diffusion-weighted
Echo time (ms)	130–150	65.74–84.88
Repetition time (ms)	3427–16,738	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/mm^2, but for the purpose of this study only the $b = 0$ and $b = 1000$ s/mm^2 series were used for analyses and to calculate the ADC map

threshold of 42% of the maximum standardised uptake value (SUV_{max}), according to methods previously described [22–24]. The $MTV_{42\%}$ was used to calculate the mean standardised uptake value (SUV_{mean}) and total lesion glycolysis (TLG; defined as $SUV_{mean} \times MTV_{42\%}$). The $MTV_{42\%}$ segmentation was transferred to CT to calculate the mean Hounsfield unit (HU) ($CT\text{-}HU_{mean}$).

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, 25–29].

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 dichotomised as mrT3c-4 vs. mrT1-3b and mrN+ vs. mrN0, respectively.

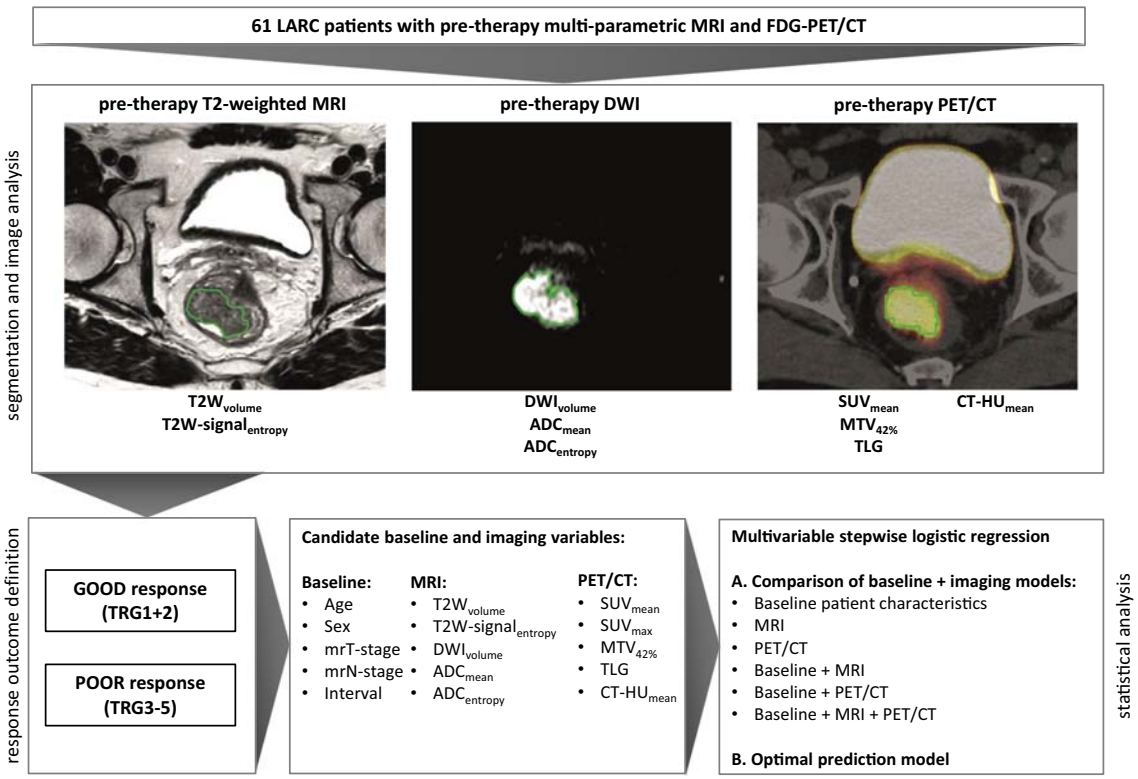

Fig. 2 Schematic study outline

Response to chemoradiotherapy (standard of reference)

The primary outcome was the histopathological tumour regression grade (TRG) by Mandard [30]. 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.

Statistical analysis

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

The value of the quantitative MRI and PET/CT features and baseline patient characteristics to predict a good response was analysed 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 [31, 32]. The analysis workflow is summarised as follows:

- As described above, only a limited number of parameters ($T2W_{volume}$, $T2W\text{-}signal_{entropy}$, 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, 25–28]. The interval between the last radiotherapy fraction and the final response evaluation (dichotomised 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 [33].
- 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

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 3, A. The best fitting model (based on the smallest AIC) was the baseline + MRI model. The 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.

Optimised multivariable model

The optimised baseline + MRI model is summarised in Table 3, B, and included mrT-stage (OR 0.004; 95% CI 0.00–0.09 for cT3c-4 vs. cT1-3b), $T2W\text{-}signal_{entropy}$ (OR per IQR 4.33; 95% CI 1.47–12.77) and $T2W_{volume}$ (OR 1.028 per cm^3; 95% CI 1.00–1.05). The model had an AUC of 0.88 to predict good responders within our dataset, with a sensitivity

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	RT treatment duration	37 (36–51)
(Median N° days and interquartile range)	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 surgery ($n = 54$ 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

[a] 7/13 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)

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.

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 ($T2W\text{-}signal_{entropy}$ 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.

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

[a] No variables were selected as predictors when only PET/CT variables were offered to the model

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 [34, 35]. 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 [36, 37]. 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 [33, 38–42]. 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 [25]. In contrast, Meng et al found lower pre-treatment T2W entropy to be associated with complete response [43], while a third report by De Cecco et al found no significant differences at all in baseline tumour entropy between response groups [44]. 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$_{volume}$) 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 [45–53].

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, suboptimal 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 [20]. 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 [45, 46, 48]. 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 inter-observer 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 2 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) 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.

Funding information This study has received funding by the Dutch Cancer Society (project number 10138).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Doenja MJ Lambregts.

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

Statistics and biometry One of the authors, Mr. Sander Roberti, has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Boards (retrospective analysis of prospectively obtained observational data).

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap Some study subjects or cohorts have been previously reported in:

Cusomano (2018) Radiol Med
Van Stiphout (2014) Radiother Oncol
Janssen (2012) Int J Rad Oncol Biol Phys
Van Stiphout (2011) Radiother Oncol
Janssen (2010) Int J Rad Oncol Biol Phys
Janssen (2010) Radiother Oncol
Lambregts (2011) Ann Surg Oncol
Lambregts (2015) Ann Surg
Martens (2015) Int J Radiat Oncol Biol Phys
Lambregts (2018) Dis Colon rectum

Methodology

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

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, Marks J, Kessler H, Gama-Rodrigues J (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, Dixon E, Buie WD (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 Biol Phys 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
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, Colletti PM, Rubello D (2015) Value of 18 F-FDG PET for predicting response to neoadjuvant therapy in rectal cancer: systematic review and meta-analysis. AJR 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, 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
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. 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
21. Van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77:e104–e107
22. 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
23. 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
24. Ueno Y, Lisbona R, Tamada T, Alaref A, Sugimura K, Reinhold C (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
25. 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 (NY) 21:1051–1058
26. Greenbaum A, Martin DR, Bocklage T, Lee JH, Ness SA, Rajput A (2019) Tumor heterogeneity as a predictor of response to neoadjuvant chemotherapy in locally advanced rectal cancer. Clin Colorectal Cancer 18:102–109
27. 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
28. 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
29. Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys 102:1143–1158
30. 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
31. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723
32. Burnham KP, Anderson DR (2004) Multimodel Inference. Sociol Methods Res 33:261–304
33. 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
34. Joye I, Debucquoy A, Fieuws S et al (2016) Can clinical factors be used as a selection tool for an organ-preserving strategy in rectal cancer? Acta Oncol 55:1047–1052
35. Al-Sukhni E, Attwood K, Mattson DM, Gabriel E, Nurkin SJ (2016) Predictors of pathologic complete response following neoadjuvant chemoradiotherapy for rectal cancer. Ann Surg Oncol 23:1177–1186
36. 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 MR 26:259–268
37. Gröne J, Loch FN, Taupitz M, Schmidt C, Kreis ME (2018) Accuracy of various lymph node staging criteria in rectal cancer with magnetic resonance imaging. J Gastrointest Surg 22:146–153
38. 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
39. 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
40. Foster JD, Jones EL, Falk S, Cooper EJ, Francis NK (2013) Timing of surgery after long-course neoadjuvant chemoradiotherapy for rectal cancer: a systematic review of the literature. Dis Colon Rectum 56:921–930
41. 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
42. 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
43. 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
44. De Cecco CN, 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
45. 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
46. 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
47. 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
48. 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
49. Ha HI, Kim AY, Yu CS, Park SH, Ha HK (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
50. 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
51. 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
52. 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
53. 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 48:295–302