Received: 20 October 2021 | Revised: 31 January 2022 | Accepted: 1 February 2022

DOI: 10.1002/ijc.33975

CANCER THERAPY AND PREVENTION

IJC INTERNATIONAL JOURNAL of CANCER

Modest agreement between magnetic resonance and pathological tumor regression after neoadjuvant therapy for rectal cancer in the real world

Pietro Achilli[1,2] | Carmelo Magistro[2] | Mohamed A. Abd El Aziz[1] | Giacomo Calini[1] | Camillo L. Bertoglio[2] | Giovanni Ferrari[2] | Giulio Mari[3] | Dario Maggioni[3] | Georgios Peros[4] | Sara Tamburello[4] | Elisabetta Coppola[4] | Antonino Spinelli[5,4] | Fabian Grass[6] | David Martin[6] | Dieter Hahnloser[6] | Andrea Salvatori[7] | Silvia De Simoni[8] | Shannon P. Sheedy[9] | Joel G. Fletcher[9] | David W. Larson[1]

[1]Division of Colon and Rectal Surgery, Mayo Clinic, Rochester, Minnesota, USA

[2]Department of Mini-invasive Surgery, Grande Ospedale Metropolitano Niguarda, Milan, Italy

[3]Department of Surgery, Desio Hospital, Desio, Italy

[4]Department of Biomedical Sciences, Humanitas University, Milan, Italy

[5]IRCCS Humanitas Research Hospital, Milan, Italy

[6]Department of Visceral Surgery, Lausanne University Hospital CHUV, Lausanne, Switzerland

[7]Branch of Medical Statistics, Biometry, and Epidemiology "G. A. Maccacaro," Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy

[8]Department of Radiology, Grande Ospedale Metropolitano Niguarda, Milan, Italy

[9]Mayo Clinic Department of Radiology, Rochester, Minnesota, USA

Correspondence
Pietro Achilli, Mayo Clinic Rochester MN, 200 First Street SW, Rochester, MN 55905, USA.
Email: achilli.pietro@mayo.edu; pietro.achilli@89gmail.com

Abstract

Magnetic resonance imaging (MRI) is routinely used for preoperative tumor staging and to assess response to therapy in rectal cancer patients. The aim of our study was to evaluate the accuracy of MRI based restaging after neoadjuvant chemoradiotherapy (CRT) in predicting pathologic response. This multicenter cohort study included adult patients with histologically confirmed locally advanced rectal adenocarcinoma treated with neoadjuvant CRT followed by curative intent elective surgery between January 2014 and December 2019 at four academic high-volume institutions. Magnetic resonance tumor regression grade (mrTRG) and pathologic tumor regression grade (pTRG) were reviewed and compared for all the patients. The agreement between radiologist and pathologist was assessed with the weighted *k* test. Risk factors for poor agreement were investigated using logistic regression. A total of 309 patients were included. Modest agreement was found between mrTRG and pTRG when regression was classified according to standard five-tier systems ($k = 0.386$). When only two categories were considered for each regression system, (pTRG 0-3 vs pTRG 4; mrTRG 2-5 vs mrTRG 1) an accuracy of 78% (95% confidence interval [CI] 0.73-0.83) was found between radiologic and pathologic assessment with a k value of 0.185. The logistic regression model revealed that "T3 greater than 5 mm extent" was the only variable significantly impacting on disagreement (OR 0.33, 95% CI 0.15-0.68, $P = .0034$). Modest agreement exists between mrTRG and pTRG. The chances of appropriate assessment of the regression grade after neoadjuvant CRT appear to be higher in case of a T3 tumor with at least 5 mm extension in the mesorectal fat at the pretreatment MRI.

Abbreviations: ARJ, anorectal junction; CR, complete response; CRT, chemoradiotherapy; EMVI, extramural venous invasion; LASSO, least absolute shrinkage and selection operator; MRI, magnetic resonance imaging; NCCN, National Comprehensive Cancer Network; TRG, tumor regression grade.

KEYWORDS
agreement analysis, magnetic resonance, neoadjuvant therapy, rectal cancer, tumor regression grade

What's new?

The watch-and-wait approach is a promising organ-preserving strategy for rectal cancer patients with complete or near-complete response to neoadjuvant chemotherapy. The early clinical success of this approach warrants reassessment of the current magnetic resonance imaging (MRI)-based tumor regression grade (mrTRG) scoring system for detecting pathological response. Here, the authors analyzed agreement between the mrTRG system and the classical pathological tumor regression grade (pTRG) system. Analyses reveal modest agreement between the two when regression is classified according to standard five-tier systems. The findings indicate that mrTRG is not a reliable surrogate marker of complete pathologic response in locally advanced rectal cancer.

1 | INTRODUCTION

Neoadjuvant chemoradiotherapy (CRT) followed by radical surgical resection represents the standard of care for locally advanced rectal cancer.[1] Organ-preserving treatments such as the watch-and-wait strategy are increasingly considered as an alternative to surgery in patients showing a complete clinical response (cCR) or near-complete response (nCR) to CRT.[2] The most extensive study of 880 patients published by the International Watch and Wait Data registry reported a 5-year overall survival of 85% and 5-year disease-specific survival of 95%.[3] The encouraging oncological results of the watch and wait approach[2-4] increase the need for an accurate clinical response evaluation tool to reliably identify those patients with cCR or nCR.[5]

Magnetic resonance imaging (MRI) is routinely used for preoperative tumor staging and to assess response to therapy in rectal cancer patients.[6] The size of the tumor and qualitative change in signal intensity are the main criteria used for response evaluation through an imaging-based tumor regression scoring system (mrTRG)[6] developed based on the principles of the pathological tumor regression grading (pTRG).[7] Nevertheless, few studies have analyzed the accuracy of mrTRG scoring in predicting response to neoadjuvant therapy.[8,9] Hence, there is a need to investigate the agreement between radiologic and pathological tumor regression to assess the safety of pursuing a watch and wait strategy in a real-world practice setting.

The aim of our study was to evaluate the accuracy of MRI based restaging after neoadjuvant CRT in predicting pathologic response in four high-volume institutions with longstanding experience in rectal cancer treatment.

2 | METHODS

2.1 | Patients

This observational study included adult (age > 18 years old) patients with histologically confirmed locally advanced rectal adenocarcinoma (clinical T3-T4N0M0 or T[any]N + M0) treated with neoadjuvant CRT followed by curative intent elective surgery between January 2014 and December 2019 at four academic high-volume institutions: the Division of Colon and Rectal Surgery, Mayo Clinic, Rochester, MN; the Department of Visceral Surgery, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland; the Division of Minimally-Invasive Surgery, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy and the Division of Colon and Rectal Surgery, IRCCS Humanitas Research Hospital, Rozzano, Milan. Exclusion criteria comprised stage IV disease, inflammatory bowel disease-related cancer, urgent surgery, short-course radiotherapy and lack of MRI synoptic report data. Patients with high rectal tumors who were not treated with neoadjuvant radiotherapy were not included in the present study.

All included patients had a pelvic MRI at baseline and a second pelvic MRI after neoadjuvant CRT. Interpretation and reporting of MRI for baseline staging was performed through a structured report adopted by each center involved in the present study.[10,11] All MRI reports were presented according to international standards and validated by board-certified specialized lower gastrointestinal radiologists. Threatened radiological margins were defined as the shortest distance of ≤1 mm between the mesorectal fascia and the tumor margin. For T3 tumors, the presence of extramural extension equal to or beyond 5 mm was recorded. Extramural venous invasion (EMVI) was defined as presence of tumor cells in the vasculature beyond the muscularis propria. Tumor height was defined on MRI as the shortest distance in a straight line from the lower edge of the tumor to the top of the anorectal junction, and divided into less than 1 cm, 1 to 5 cm and greater than 5 cm. Time intervals between the end of neoadjuvant CRT and the second MRI and between the second MRI and the date of surgery were reported.

According to the National Comprehensive Cancer Network (NCCN) guidelines, neoadjuvant treatment was administered at each center with different long-course CRT regimens.[1] To simplify both data collection and analysis, all the regimens were classified into two different categories: standard CRT (capecitabine/long-course RT or

infusional 5-FU/long-course RT or bolus 5-FU/leucovorin/long-course RT) or a combined regimen of long course chemotherapy (FOLFOX or CAPEOX or 5-FU/leucovorin or capecitabine) followed by standard CRT.[1]

Data were collected separately in each center, and all the relevant information was collectively stored in a dedicated anonymized database (REDCAP platform) to facilitate data management and analysis.[12]

2.2 | Tumor regression grade (mrTRG and pTRG)

The assessment of response to neoadjuvant CRT was reported according to the five tiers classification of the mrTRG. Initially developed by the MERCURY study group, the mrTRG score derived from pTRG was defined by the proportion of presumed residual tumor and fibrotic change on T2-weighted images[6] (Table 1). On postneoadjuvant CRT T2-weighted imaging, the fibrotic portion shows dark signal intensity, while the amount of residual tumor demonstrating intermediate signal intensity, similar to that of the initial tumor. Diffusion-weighted imaging (DWI) and tumor volumetry were included in the rectal MRI protocols.[13] DWI analyses how water molecules are free to move in a certain tissue. In highly cellular tissues, such as tumor tissue, the diffusion capacity of water is restricted and the signal is retained. Conversely, free diffusion can be seen in hypocellular tissues, such as fibrotic scar. Therefore, the addition DWI to standard T2-weighted improves the ability MRI to differentiate between patients with a complete tumor response and those with residual tumor.[14]

As already reported in the literature, treatment response in the primary tumor and the lymph nodes may show discrepancies.[15] In case of remnant suspicious lymph nodes after CRT at the restaging MRI, even in case of complete or near complete response of the primary tumor (cT0N+), patients were considered to be moderate responders (mrTRG3). In general, any lymph node with an irregular border, mixed signal intensity and/or size >8 mm in the short axis was reported as suspicious.[16]

Rectal specimens were routinely sampled and analyzed by board-certified pathologists and Dworak regression scale was used to estimate the response to neoadjuvant chemotherapy[7] (Table 1). All the histological specimens were analyzed with similar methods across the four institutions. Tumor or fibrotic area was identified and described macroscopically. Five to ten blocks were obtained from formalin fixed areas of macroscopic residual tumor or fibrotic tissue (in case of no macroscopic residual tumor). Five micrometer thick slides from each block were examined. If no tumor was found on the first slide, three levels were cut on all blocks from the tumor site (or fibrotic area) following a step section technique. Histological typing and grading were performed according to the WHO classification of tumors[17] and staging to the UICC 7th edition.[18]

2.3 | Study objectives

The study's primary aim was to evaluate the accuracy of the radiologic presurgery assessment by comparing the mrTRG to the pTRG score (defined as gold standard), to detect pathological response after neoadjuvant CRT. The secondary aim was to define factors (patient-related or tumor-related) that can impact the ability of the mrTRG score to predict pathologic response.

2.4 | Statistical analysis

The calculated sample size required for our study, considering 90% power and a 5% two-sided level of statistical significance, was

TABLE 1 Magnetic resonance and pathological tumor regression grade

Magnetic resonance tumor regression grade		Pathological tumor regression grade	
mrTRG 5 (no response)	No regression (intermediate signal intensity, same appearances as original tumor)	pTRG 0 (no response)	No regression
mrTRG 4 (slight response)	Slight regression (little areas of low signal intensity fibrosis or mucin but mostly tumor)	pTRG 1 (minimal response)	Dominant tumor mass with obvious fibrosis and/or vasculopathy
mrTRG 3 (moderate response)	Moderate regression (low signal intensity fibrosis predominates but there are obvious areas of intermediate signal intensity)	pTRG 2 (partial response)	Dominantly fibrotic changes with few tumor cells or groups (easy to find)
mrTRG 2 (near complete response)	Good regression (predominant low signal intensity fibrosis with no obvious residual tumor signal)	pTRG 3 (near complete response)	Very few (difficult to find microscopically) tumor cells in fibrotic tissue with or without mucous substance
mrTRG 1 (complete response)	Complete regression (absence of tumor signal and barely visible treatment related scar)	pTRG 4 (complete response)	No tumor cells, only fibrotic mass (total regression or response)

Abbreviations: mr, magnetic resonance; p, pathologic; TRG, tumor regression grade.

292 patients. This value was calculated for a weighted Cohen's *k* with $k_0 = 0.7$ and $k_1 = 0.8$. In order to correctly estimate the sample size, a preliminary analysis of 88 cases was conducted. Sample size calculation was performed with R (package "irr"). Continuous variables are shown as mean ± SD and medians with range, while categorical variables as frequency (percentage). Interrater reliability between mrTRG and pTRG was assessed using the weighted *k* test. This analysis was conducted both with a five-tier regression scale and a two-tier regression scale (ie, pTRG 0-3 vs pTRG 4; mrTRG 2-5 vs mrTRG 1 and pTRG 0-3 vs pTRG 4; mrTRG 3-5 vs mrTRG 1-2). Cohen's kappa statistic, *k*, is a measure of agreement between categorical variables *X* and *Y*. For example, Kappa statistic can be used to assess the agreement between alternative methods of categorical assessment when new techniques are under study. Kappa statistic can also be used to compare the ability of different raters to classify subjects into one of several groups, as it was used in our study.[19] Perfect agreement is evident when Cohen's kappa equals 1; a value of Cohen's kappa equal to zero suggests that the agreement is no better than that which would be obtained by chance alone. Although there is no formal scale, the level of agreement is considered moderate when *k* is greater than 0.4.[20]

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated using a dichotomous classification to assess the ability of MRI to identify complete pathological responders (ie, pTRG 0-3 vs pTRG 4; mrTRG 2-5 vs mrTRG 1 and pTRG 0-3 vs pTRG 4; mrTRG 3-5 vs mrTRG 1-2). Predictive factors associated with a worse agreement between radiological and pathological response to CRT using a two-tier regression scale ((ie, pTRG 0-3 vs pTRG 4; mrTRG 3-5 vs mrTRG 1 and 2) were investigated through multivariate logistic regression analysis. Factors related to poor agreement between mTRG and pTRG for the multivariate regression model were determined using the least absolute shrinkage and selection operator (LASSO) with bootstrapping technique. A *P*-value <.05 was considered to be statistically significant. All tests were two-sided. Statistical analysis was performed using R (version 4.1.0).

3 | RESULTS

A total of 309 patients represented the study population. One hundred and eight patients (34.9%) had a complete or near complete radiologic response (mrTRG 1 or 2), 143 patients (46.4%) had a moderate response (mrTRG 3) while in 58 patients (18.7%) a poor regression (mrTRG 4 or 5) was observed. Additional patient and tumor characteristics are summarized in Table 2. The median time from the end of CRT to the preoperative MRI scan was 48 days (IQR: 35-64). The median time from the preoperative MRI scan to surgery was 29.5 days (IQR: 18-44).

Abdominoperineal resection was performed in 56 patients (18.1%), while 253 subjects were submitted to low anterior resection and anastomosis. Pathologic regression analysis on resected specimens revealed 112 patients (36.2%) with complete or good regression (pTRG 4 or 3), 139 patients (45%) with partial regression, while 58 patients (18.8%) showed poor or no response to treatment (Table 3).

3.1 | Interrater agreement analysis (five-tier regression scale)

The agreement plot with the five-tier regression scale is displayed in Figure 1. In 127 out of 309 cases (41.1%) radiologic and pathological quantitated tumor regression using specular regression categories and weighted *k* analysis showed a *k* value of 0.386. Percentage of underestimation and overestimation of the pathological regression by the MRI for each grade (pTRG 4-0) are displayed in Figure 2. Among 57 patients who were found to be complete responders to neoadjuvant treatment at pathological report (pTRG 4), 24 (42.1%) were considered to have less pronounced regression according to MRI (mrTRG 3-4).

TABLE 2 Patients and tumor characteristics

Variable	Total n = 309 (%)
Age (years), mean (±SD)	66.1 (±13.7)
Gender (male)	182 (58.9%)
Body mass index (kg/m^2), mean (±SD)	27.9 (±3.1)
BMI >30	61 (19.7%)
Height from ARJ in cm	
<1 cm	77 (24.9%)
≥1 cm	232 (75.1%)
Clinical T-stage	
cT1-2	48 (15.5%)
cT3	237 (76.7%)
cT4	24 (7.8%)
Clinical N-stage	
N0	54 (17.4%)
N+	255 (82.6%)
Clinical extramesorectal lymph nodes	71 (22.9%)
Threatened/involved mesorectal fascia	121 (39.1%)
Anterior location	159 (51.4%)
EMVI positive on baseline MRI	31 (10.3%)
T3 greater than 5 mm extent	128 (41.4%)
Longitudinal extent >40 mm	152 (49.3%)
mrTRG	
1	40 (12.9%)
2	68 (22.0%)
3	143 (46.3%)
4	44 (14.2%)
5	14 (4.5%)

Abbreviations: ARJ, anorectal junction; EMVI, extramural venous invasion; mr, magnetic resonance; MRI, magnetic resonance imaging; TRG, tumor regression grade.

Understaging occurred in 22 patients (40%) who were graded as mrTRG 1 at the second MRI and turn out to be only near complete responders (pTRG 3) at the pathologic examination.

3.2 | Interrater agreement analysis (two-tier regression scale)

Table 4 summarizes the accuracy of mrTRG 1 and mrTRG 1 to 2 for the diagnosis of pathologic CR (pTRG 4) within the entire cohort. When only two categories were considered for each regression system, (pTRG 0-3 vs pTRG 4; mrTRG 2-5 vs mrTRG 1) an accuracy of 78% (95% confidence interval [CI] 0.73-0.83) was found between radiologic and pathologic assessment with a *k* value of 0.185 (95% CI 0.05-0.31). The sensitivity and specificity of mrTRG 1 for the diagnosis of pTRG 4 were 38% (95% CI 23-54%) and 84% (95% CI 0.79-0.88), respectively.

When considering both mrTRG 1 and 2 for the diagnosis of pTRG 4 the weight *k* analysis showed a *k* value of 0.209 (95% CI 0.10-0.31), while the accuracy, sensitivity and specificity were 0.67 (95% CI 0.62-0.73), 0.57 (95% CI 0.44-0.70) and 0.70 (95% CI 0.64-0.75), respectively.

TABLE 3 Operative details and pathological outcomes

Variable	Total n = 309 (%)
Procedure	
APR	56 (18.1%)
LAR	253 (81.9%)
pTRG	
0	12 (3.9%)
1	46 (14.9%)
2	139 (45%)
3	55 (17.8%)
4	57 (18.4%)
Tumor stage (p)	
pT0-1	85 (27.5%)
pT2	83 (26.9%)
pT3	130 (42.1%)
pT4	11 (3.5%)
N-stage	
N0	197 (63.7%)
N1	87 (28.2%)
N2-3	25 (8.1%)

Abbreviations: APR, abdominoperineal resection; LAR, low anterior resection; p, pathologic; TRG, tumor regression grade.

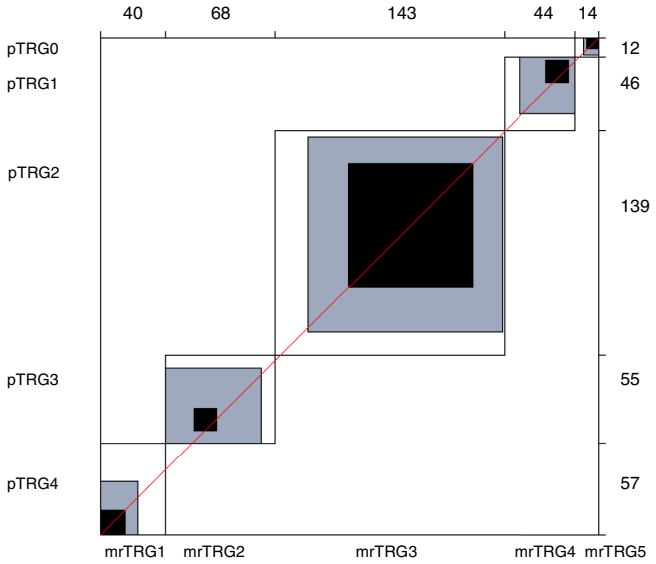

FIGURE 1 Agreement chart: mrTRG and pTRG (five-tier regression scale). Percent of agreement = 41.1%; *k* = 0.386; 95% CI = 0.28 to 0.49. In case of perfect agreement, the rectangles determined by the marginal totals are all perfect squares and the shaded squares determined by the diagonal cell entries are exactly equal to the rectangles. Lesser agreement is visualized by comparing the area of the blackened squares to the area of the rectangles, while observer bias is visualized by examining how rectangles deviate from the 45° diagonal line

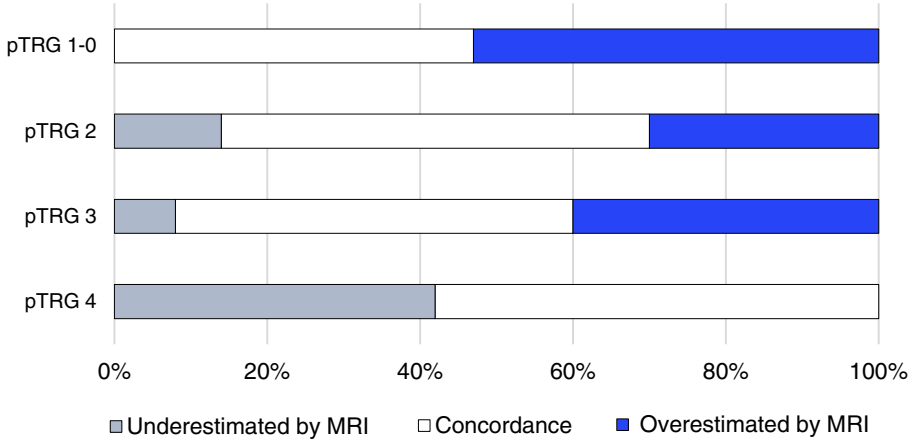

FIGURE 2 Accuracy of mrTRG for the assessment of histopathologic regression (five-tier regression scale)

A series of variables, both patient-and tumor-related, were identified within the dataset in order to investigate which factors could impact the ability of the mrTRG 1 to 2 to predict pathologic CR ("Clinical T2," "Clinical T3," "preoperative MRI scan to surgery," "end of CRT to surgery," "Clinical extramesorectal lymph nodes," "anterior location," "threatened/involved mesorectal fascia," "Clinical N+," "standard CRT," "LCCT + standard CRT," "height from ARJ <1 cm," "end of CRT to preoperative MRI," "EMVI positive on baseline MR" and "T3 greater than 5 mm extent"). A Lasso regression analysis was then performed in order to enhance the resulting statistical model, retaining eight variables with non-null coefficients ("pre-operative MRI scan to surgery (days)," "anterior location," "threatened/involved mesorectal fascia," "Clinical N+," "height from ARJ < 1 cm," "end of CRT to preoperative MRI (days)," "EMVI positive on baseline MR" and "T3 greater than 5 mm extent"). The logistic regression model on those variables is shown in Table 5, with the only variable significantly impacting on disagreement being "T3 greater than 5 mm extent" (OR 0.31, 95% CI 0.14-0.65, $P = .0024$).

TABLE 4 Diagnostic accuracy of mrTRG 1 and mrTRG1-2 for the diagnosis of pathological complete response (pTRG4)

	Accuracy of mrTRG 1	Accuracy of mrTRG 1-2
Number of patients	309	309
True positive	15	33
False positive	42	75
False negative	25	24
True negative	227	177
Diagnostic accuracy	0.78 (0.73-0.83)	0.67 (0.62-0.73)
Sensitivity (95% CI)	0.38 (0.23-0.54)	0.57 (0.44-0.70)
Specificity (95% CI)	0.84 (0.79-0.88)	0.70 (0.64-0.75)
Positive predictive value (95% CI)	0.26 (0.15-0.39)	0.30 (0.22-0.40)
Negative predictive value (95% CI)	0.90 (0.86-0.93)	0.88 (0.82-0.92)

Abbreviations: mr, magnetic resonance; p, pathologic; TRG, tumor regression grade.

4 | DISCUSSION

Radiological presurgery assessment, using the mrTRG score, for detecting pathological CR after neoadjuvant CRT in a real-world scenario involving four high-volume institutions with longstanding experience in rectal cancer treatment is reported in this series. The agreement between an MRI-based tumor regression grading system and the classical pTRG system proposed by Dworak was analyzed. The results suggest only modest agreement between radiologic and pathologic assessment. Moreover, a T3 extent of greater than 5 mm extent was the only factor impacting on interrater reliability between mrTRG and pTRG in our study.

Monitoring treatment efficacy has a pivotal role in the management of cancer patients. Assessing the effects of treatment is not only essential to define treatment duration, but it also has the potential to provide valuable prognostic information and guide subsequent therapeutic and follow-up strategies. When it comes to rectal cancer, predicting the response to neoadjuvant treatment becomes crucial when choosing an organ-preserving strategy. Indeed, nonoperative management strategies are increasingly considered as an alternative to surgery in patients showing complete response to CRT.[2] Thus, there is a need for a reliable noninvasive imaging tool to identify those patients who are more likely to achieve a complete response to treatment and, therefore, may be spared from surgery.[21] In this scenario, mTRG has recently emerged as a dynamic, noninvasive surrogate method for assessing tumor regression after neoadjuvant treatment prior to surgical resection.[22] The evidence, however, is conflicting. While many studies have shown its ability to predict pathological findings on surgical specimens as well as long-term prognosis,[23,8] some others described rather poor overall accuracy. In a recent meta-analysis published by Memon et al, the MRI treatment response was significantly predictive of mesorectal fascia involvement but showed poor accuracy for T stage prediction (52%).[24] In a large cohort by Maretto et al, complete responders could be detected only partially on restaging MRI.[25] Records from two phase II trials (EXPERT and EXPERT-C) were used to compare mrTRG and pTRG in rectal cancer patients by Sclafani et al, showing fair agreement between the two grading modalities (33.5%) when regression was classified according to standard five-tier systems ($k = 0.24$).[9] The authors concluded that

Variable	OR (95% CI)	(95% CI)	*P* value
Preoperative MRI scan to surgery (days)	1.01	(1.00-1.02)	.1133
Anterior location	0.6	(0.21-1.24)	.3951
Threatened/involved mesorectal fascia	1.7	(0.84-3.92)	.1508
Height from ARJ <1 cm	0.97	(0.41-2.04)	.9368
Clinical N+	0.92	(0.50-1.69)	.9837
End of CRT to preoperative MRI (days)	1	(1.00-1.01)	.1121
Extramural venous invasion	0.635	(0.19-1.78)	.6512
T3 >5 mm extent	0.31	(0.14-0.65)	.0024

TABLE 5 Preoperative factors influencing the disagreement between mrTRG 1 to 2 and pTRG 4

Abbreviations: ARJ, anorectal junction; CRT, chemoradiotherapy; mr, magnetic resonance; MRI, magnetic resonance imaging; p, pathologic; TRG, tumor regression grade.

the agreement between mrTRG and pTRG was low, questioning the use of mrTRG as a surrogate of pTRG. Our findings are in line to the aforementioned study, showing 41.1% agreement with a *k* value of 0.386.

The limits of mrTRG as a surrogate of pTRG become more evident when it comes to complete or near complete radiologic response evaluation. In our cohort, nearly 50% of the patients who were considered to be good responders to neoadjuvant CRT (mrTRG 1 or 2) turned out to have only partial response to treatment at pathology review (pTRG 3-0; Figure 1). The difficulty of post-CRT assessment of rectal cancer might be explained by CRT-induced changes and their overlapping appearances on MRI. Edema, inflammation, necrosis and particularly fibrosis can be hard to distinguish from residual tumor. Small islands of tumor cells within a large amount of reactive fibrotic tissue might not be detected. On the other hand, spiculated fibrosis may erroneously mimic a tumor. Some radiologists consider areas of very low signal intensity as sterile fibrosis,[6] whereas others consider these to be areas of residual tumor,[26] leading to either underestimation or overestimation of tumor response (Figure 2). Even the routinely use of DWI in the rectal MRI protocols does not seem to overcome the problem of differentiating fibrotic tissue from residual vital tumor. This might be related to the shortcomings of DWI, such as image distortion due to artifacts or interobserver variation. Moreover, fluids or mucinous cells can mimic restricted diffusion signal, further hampering the correct interpretation of diffusion images.[14] It is likely that the problem of differentiating between residual microscopic disease and sterile fibrosis will continue to limit the clinical value of MRI response assessment and the predictive role of mTRG in the assessment of complete response to CRT.[27]

Several regression scales are currently used by pathologists to assess response to neoadjuvant CRT, revealing a lack of standardization and potentially leading to confusion in interpreting data among different institutions worldwide.[23] Although the original reports utilized a five-tier system, many authors have found no loss in accuracy when the system is collapsed to a three-tier system.[28] Moreover, the 7th edition of the AJCC Cancer Staging Manual proposed a scale based on four tiers, which was found to be more accurate in terms of prognostication.[29]

In the present study, even if many other scales were described and available in the pathological reports, the mrTRG was compared to the Dworak regression system.[7] The Dworak system was preferred over other scales because it classified regression into five grades from 0 (no regression) to 4 (total regression), reflecting the five points of the mrTRG system. It can be argued that the choice of a particular scale could impact the reliability of the agreement analysis. However, since all the available pTRG scoring systems resulted to be comparable both in terms of classification criteria and prognostication,[30] it is unlikely that classifying pathological tumor regression according to a different method would change the outcome of the agreement analysis.

The secondary aim of our study was to identify factors (patient-related or tumor-related) affecting the ability of the mrTRG score to predict pathologic regression. Interestingly, the only factor positively impacting the agreement between mTRG and pTRG on multivariate analysis was the extramural invasion >5 mm of the tumor in the mesorectal tissue (T3 extent >5 mm). It appears that the more tumor invades the perirectal fat beyond the muscularis propria, the more likely it is that radiological and pathological regression scores will be concordant. In case of tumors with deeper extension into the mesorectal fat, the radiation-induced inflammation and regression after neoadjuvant CRT might be easier to appreciate, reducing the misinterpretation of fibrotic reactive tissue as tumor remnant.

Our study has limitations related to but not limited to its retrospective design. First, the international, multicenter setting was chosen to provide more representative and generalizable results. However, this approach bears an inherent risk of heterogeneity of data assessment and interpretation. Second, data collection was limited by the availability of standardized synoptic reporting, leading to a risk of bias due to patient selection. Third, the assessment of mrTRG and pTRG might be affected by subjective interpretation and depend on experience and quality of imaging, despite standardized protocols throughout the centers. Therefore, the results of our study should be interpreted considering these inherent limitations.

5 | CONCLUSION

In conclusion, the present study revealed modest agreement between mrTRG and pTRG. Hence, mrTRG cannot be considered as reliable surrogate marker of complete pathologic response assessment in locally advanced rectal cancer. The chances of appropriate assessment of the regression grade after neoadjuvant CRT appear to be higher in case of a T3 tumor with at least 5 mm extension in the mesorectal fat at the pretreatment MRI.

CONFLICT OF INTEREST

Antonino Spinelli received speaker fees from Johnson&Johnson and Takeda. No other disclosures or conflict of interest related to this project to declare.

DATA AVAILABILITY STATEMENT

Access to the de-identified data on the REDCAP platform can be made available to qualified researchers upon reasonable request to the author (David W. Larson).

ETHICS STATEMENT

All patient information in the REDCAP database is de-identified. Our study was approved by the Institutional Review Board (IRB) and informed consent was obtained by patients involved in this research.

ORCID

Pietro Achilli https://orcid.org/0000-0001-9021-9521
Fabian Grass https://orcid.org/0000-0001-9884-2879

REFERENCES

1. Benson AB, Venook AP, Al-Hawary MM, et al. Rectal cancer, version 2.2018 clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. 2018;16:874-901.

2. Martens MH, Maas M, Heijnen LA, et al. Long-term outcome of an organ preservation program after neoadjuvant treatment for rectal cancer. *J Natl Cancer Inst*. 2016;108:1-10.
3. van der Valk MJM, Hilling DE, Bastiaannet E, et al. 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*. 2018; 391:2537-2545.
4. Smith JJ, Strombom P, Chow OS, et al. Assessment of a watch-and-wait strategy for rectal cancer in patients with a complete response after neoadjuvant therapy. *JAMA Oncol*. 2019;5:e185896.
5. Sathyakumar K, Chandramohan A, Masih D, Jesudasan MR, Pulimood A, Eapen A. Best MRI predictors of complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. *Br J Radiol*. 2016;89:20150328.
6. Patel UB, Blomqvist LK, Taylor F, et al. MRI after treatment of locally advanced rectal cancer: how to report tumor response—the MERCURY experience. *Am J Roentgenol*. 2012;199:W486-W495.
7. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *Int J Colorectal Dis*. 1997;12:19-23.
8. Siddiqui MRS, Gormly KL, Bhoday J, et al. Interobserver agreement of radiologists assessing the response of rectal cancers to preoperative chemoradiation using the MRI tumour regression grading (mrTRG). *Clin Radiol*. 2016;71:854-862.
9. Sclafani F, Brown G, Cunningham D, et al. Comparison between MRI and pathology in the assessment of tumour regression grade in rectal cancer. *Br J Cancer*. 2017;117:1478-1485.
10. Beets-Tan RGH, Lambregts DMJ, Maas M, et al. Correction to: 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*. 2018;1.
11. Gollub MJ, Arya S, Beets-Tan RG, et al. Use of magnetic resonance imaging in rectal cancer patients: Society of Abdominal Radiology (SAR) rectal cancer disease-focused panel (DFP) recommendations 2017. *Abdom Radiol*. 2018;43:2893-2902.
12. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
13. Seo N, Kim H, Cho MS, Lim JS. Response assessment with MRI after chemoradiotherapy in rectal cancer: current evidences. *Korean J Radiol*. 2019;20:1003-1018.
14. Lambregts DMJ, Boellaard TN, Beets-Tan RGH. Response evaluation after neoadjuvant treatment for rectal cancer using modern MR imaging: a pictorial review. *Insights Imaging*. 2019;10:15. doi: 10.1186/s13244-019-0706-x
15. Loftas P, Sturludóttir M, Hallböök O, Almlöv K, Arbman G, Blomqvist L. Assessment of remaining tumour involved lymph nodes with MRI in patients with complete luminal response after neoadjuvant treatment of rectal cancer. *Br J Radiol*. 2018;91:20170938.
16. Gollub MJ. *User's Guide for the Synoptic MRI Report*. Society of Abdominal Radiology, Rectal and Anal Cancer DFP: SAR DFP; 2018.
17. Hu B, El HN, Sittler S, Lammert N, Barnes R, Meloni-Ehrig A. Gastric cancer: Classification, histology and application of molecular pathology. *J Gastrointest Oncol*. 2012;3(3):251-261.
18. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti AE. *AJCC Cancer Staging Manual*. 7th ed. New York: Springer; 2010.
19. Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology*. 2010;73:1167-1179. doi:10.1016/j.theriogenology.2010.01.003
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data published by: international biometric society stable. *Biometrics*. 1977;33:159.
21. Habr-Gama A, Perez RO, Nadalin W, et al. Operative vs nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg*. 2004;240:711-718.
22. Patel UB, Taylor F, Blomqvist L, et al. Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol*. 2011;29:3753-3760.
23. Siddiqui MRS, Bhoday J, Battersby NJ, et al. Defining response to radiotherapy in rectal cancer using magnetic resonance imaging and histopathological scales. *World J Gastroenterol*. 2016;22:8414-8434.
24. Memon S, Lynch AC, Bressel M, Wise AG, Heriot AG. Systematic review and meta-analysis of the accuracy of MRI and endorectal ultrasound in the restaging and response assessment of rectal cancer following neoadjuvant therapy. *Color Dis*. 2015;17:748-761.
25. Maretto I, Pomerri F, Pucciarelli S, et al. The potential of restaging in the prediction of pathologic response after preoperative chemoradiotherapy for rectal cancer. *Ann Surg Oncol*. 2007;14:455-461.
26. Lambregts DMJ, Rao SX, Sassen S, et al. MRI and diffusion-weighted MRI volumetry for identification of complete tumor responders after preoperative chemoradiotherapy in patients with rectal cancer: A bi-institutional validation study. *Ann Surg*. 2015;262:1034-1039.
27. Jang JK, Choi SH, Park SH, et al. MR tumor regression grade for pathological complete response in rectal cancer post neoadjuvant chemoradiotherapy: a systematic review and meta-analysis for accuracy. *Eur Radiol*. 2020;30:2312-2323.
28. Shia J, Guillem JG, Moore HG, Tickoo SK, Qin J. Patterns of morphologic alteration in residual rectal carcinoma following preoperative chemoradiation and their association with long-term outcome. *Am J Surg Pathol*. 2004;28:215-223.
29. Ryan R, Gibbons D, Hyland JMP, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology*. 2005;47:141-146.
30. Trakarnsanga A, Gönen M, Shia J, et al. Comparison of tumor regression grade systems for locally advanced rectal cancer after multimodality treatment. *J Natl Cancer Inst*. 2014;106:1-6.

How to cite this article: Achilli P, Magistro C, Abd El Aziz MA, et al. Modest agreement between magnetic resonance and pathological tumor regression after neoadjuvant therapy for rectal cancer in the real world. *Int. J. Cancer*. 2022;1-8. doi:10.1002/ijc.33975