HHS Public Access

Author manuscript

Eur Radiol. Author manuscript; available in PMC 2024 August 01.

Published in final edited form as:

Eur Radiol. 2023 August ; 33(8): 5761–5768. doi:10.1007/s00330-023-09480-9.

MRI assessment of rectal cancer response to neoadjuvant therapy: a multireader study

Jonathan B. Yuval[1], **Sujata Patil**[2], **Natalie Gangai**[3], **Dana M. Omer**[1], **Dmitriy G. Akselrod**[4], **Alice Fung**[5], **Carla B. Harmath**[6], **Rony Kampalath**[7], **Kyle Krehbiel**[8], **Sonia Lee**[7], **Peter S. Liu**[9], **John D. Millet**[10], **Ryan B. O'Malley**[11], **Andrei S. Purysko**[9], **Joseph C. Veniero**[9], **Ashish P. Wasnik**[10], **Julio Garcia-Aguilar**[1], **Marc J. Gollub**[3,☖]

[1]Colorectal Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, USA

[2]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, USA

[3]Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, USA

[4]Department of Radiology, University of Vermont, Burlington, USA

[5]Department of Radiology, Oregon Health & Sciences University, Portland, USA

[6]Department of Radiology, University of Chicago, Chicago, USA

[7]Department of Radiology, University of California Irvine, Irvine, USA

[8]Department of Radiology, Creighton University Medical Center – Bergan Mercy, Omaha, USA

[9]Department of Radiology, Cleveland Clinic, Cleveland, USA

[10]Department of Radiology, University of Michigan, Ann Arbor, USA

[11]Department of Radiology, University of Washington, Seattle, USA

Abstract

Objectives: A watch and wait strategy with the goal of organ preservation is an emerging treatment paradigm for rectal cancer following neoadjuvant treatment. However, the selection of appropriate patients remains a challenge. Most previous efforts to measure the accuracy of MRI in assessing rectal cancer response used a small number of radiologists and did not report variability among them.

Methods: Twelve radiologists from 8 institutions assessed baseline and restaging MRI scans of 39 patients. The participating radiologists were asked to assess MRI features and to categorize the overall response as complete or incomplete. The reference standard was pathological complete response or a sustained clinical response for >2 years.

Results: We measured the accuracy and described the interobserver variability of interpretation of rectal cancer response between radiologists at different medical centers. Overall accuracy was 64%, with a sensitivity of 65% for detecting complete response and specificity of 63%

☖ Marc Gollub, gollubm@mskcc.org, 1-212-639-2183.

for detecting residual tumor. Interpretation of the overall response was more accurate than the interpretation of any individual feature. Variability of interpretation was dependent on the patient and imaging feature investigated. In general, variability and accuracy were inversely correlated.

Conclusions: MRI-based evaluation of response at restaging is insufficiently accurate and has substantial variability of interpretation. Although some patients' response to neoadjuvant treatment on MRI may be easily recognizable, as seen by high accuracy and low variability, that is not the case for most patients.

Keywords

Rectal Cancer; MRI; Organ Preservation; Neoadjuvant Treatment; Watch and Wait

Introduction

In about 15–25% of patients with locally advanced rectal cancer treated with chemoradiotherapy (CRT), no residual tumor is found in the resected specimen, and this subset of patients has excellent oncological outcomes [1]. With a full course of preoperative chemotherapy in addition to CRT (total neoadjuvant therapy [TNT]), the rate of complete response can approach 40% [2; 3]. Since the combination of CRT and radical resection is associated with significant and long-term functional impairments affecting the quality of life [4], many have raised the question whether some patients are overtreated with total mesorectal excision (TME).

A watch and wait (WW) strategy with the goal of organ preservation is a new treatment paradigm for patients with a clinical complete response (cCR) following neoadjuvant treatment (NAT). Selection of appropriate patients for WW remains a challenge. The assessment of tumor response to NAT is based on digital rectal examination, endoscopy, and pelvic MRI with DWI [5–11]. While acceptance for WW is gaining traction among physicians and patients there is an increasing need to assess accuracy of the endoscopic and imaging methods to assess rectal cancer response to NAT across different practice settings.

Imaging plays a fundamental role in the treatment of rectal cancer. Baseline rectal MRI has been routinely used for staging, prognostication, and determining the initial treatment strategy and whether the patient may benefit from NAT. Increasingly, MRI is also being used at completion of NAT to evaluate response, restage the disease, and help guide further management. While endoscopy is considered the most reliable modality for evaluating mucosal response, restaging MRI can complement endoscopy with its ability to detect submucosal, mesenteric, and nodal evidence of resistance to treatment [12].

Most previous efforts to measure the accuracy of MRI in assessing rectal cancer response to NAT used a small number of radiologists and some did not report variability among them [13–16]. Our aims in this study were to (i) measure the accuracy of interpretation of T2-weighted (T2W) and DW MRI response to NAT in comparison with reference standards (histopathology following surgery for patients who underwent TME and presence of a sustained cCR in patients who underwent WW) and (ii) describe interobserver variability between radiologists at different medical centers.

Materials and Methods

Patient selection, treatment, and follow-up

With approval from the institutional review board, including a waiver of informed consent, an institutional database was searched for patients treated by the same surgeon who underwent baseline staging MRI between January 2013 and June 2017. The criteria for inclusion in the study were (i) diagnosis of locally advanced rectal cancer, (ii) treatment with NAT, (iii) availability of good-quality baseline and restaging MRI data including DWI sequences (judged by MJG), and (iv) availability of a reference standard in the form of either histopathology following TME or a sustained (>2-year) cCR.

Baseline demographics, clinicopathological parameters, treatment characteristics, and clinical follow-up data were manually retrieved from electronic medical records. NAT consisted of 4500 and 5400 cGy in 180- to −200-cGy fractions with concurrent intravenous fluorouracil or oral capecitabine. FOLFOX (leucovorin, fluorouracil, oxaliplatin) or Cap-OX (capecitabine, oxaliplatin) was administered either as induction (before CRT) of consolidation (after CRT) neoadjuvant therapy. All patients received TNT and upon completion of therapy, patients proceeded to either WW or TME based on the surgeon's assessment of response. Patients in WW were monitored according to the same protocol at the same center by the same surgeon. The protocol included digital rectal examination and endoscopy every 3 months and MRI every 6 months for the first 2 years following NAT. In addition, CT of the chest, abdomen, and pelvis was performed yearly. Patients who underwent TME completed follow-up according to the National Comprehensive Cancer Network guidelines.

MRI

MR images were acquired with GE MRI units at a field strength of 1.5 (28 scans) or 3 T (50 scans). A phased array coil was used for signal reception. Multiplanar T2W and DW images were acquired using a standard institutional rectum protocol (see Electronic Supplementary Material). All but 4 scans were performed at our institutional imaging sites.

Baseline and restaging MR images were anonymized, and DICOM (Digital Imaging and Communications in Medicine) files were uploaded into open-source software (XNAT version 1.7.4.1; Washington University School of Medicine) for interpretation. An electronic questionnaire (see Electronic Supplementary Material) was created and disseminated using a browser-based software application (REDCap version 9.1.20; Vanderbilt University). Requests to participate in the study were sent to radiologists at 17 sites collaborating in the Organ Preservation in Rectal Adenocarcinoma (OPRA) phase 2 randomized control trial (NCT02008656) [17; 18]. All cases were read by all radiologists. For each patient, the participating radiologists (blinded to the patients' clinical course) assessed the primary tumor response to NAT according to T2W imaging morphology, restricted diffusion on DWI, and the combination of T2W and DW imaging features. The radiologists also assessed the overall lymph node response and the response of the primary tumor together with the lymph nodes on both T2W and DW images. The radiologists categorized the response for each of these items as either complete or incomplete. (Table 1). The assessment of response

of the primary tumor together with the lymph nodes according to both T2W and DW images was considered the overall response assessment. The questionnaire stated that lymph nodes with a short-axis diameter of <5 mm on restaging MRI should be considered negative [19]. Explanations regarding the split scar sign[20] and DW artifact from luminal contents were given in the questionnaire, however no strict definition of primary tumor response was provided, and the radiologists were asked to use their best judgment.

Reference standards

The reference standard was pathological complete response in the resected TME specimen or a cCR sustained for >2 years following NAT for WW patients. Incomplete response was defined for TME patients as cancer in the resected specimen and for WW patients as persistent tumor or regrowth in the first 2 years following NAT. Complete response was defined as a positive study outcome for all features; sensitivity was defined as the proportion of patients with complete response correctly identified by the radiologists, and specificity was defined as the proportion of patients with incomplete response correctly identified. The reference standard for uninvolved lymph nodes was uninvolved nodes in the TME specimen or a cCR sustained for >2 years following NAT.

Statistical analysis

Categorical variables are presented as frequency (percent) and continuous variables are presented as mean with standard deviations. Analyses are presented in a descriptive and graphical manner. The percent of radiologists that chose a specific survey response category for a given patient's scan was plotted in a heatmap to provide the reader an indication of radiologist variability in providing a diagnosis using a specific metric. Average sensitivity, specificity, and accuracy were calculated for each radiologist, for each patient, and across all radiologists. For questions with more than two possible response categories, responses were categorized as incomplete or complete based on consensus of the study team (see Electronic Supplementary Material). All statistical analyses were performed in R version 3.6.0.

Results

Patient characteristics and response to treatment

Thirty-nine patients met the inclusion criteria and were included in the study (Table 2). Mean age at diagnosis was 57 ± 10 years. Mean tumor distance from the anal verge was 5.4 ± 2.5 cm. Twenty-five tumors (64%) were located in the distal third of the rectum, 13 (33%) were located in the middle third, and one (3%) was located in the proximal third. Thirty-five patients (90%) had clinical stage III disease at baseline, and the remaining four patients had stage II disease. Mean radiotherapy dose was 4800 ± 302 cGy (range 4500–5400 cGy). FOLFOX was administered to 35 patients (90%), CapOX was administered to 2 patients (5%), and 2 patients switched between these regimens. The mean time from the end of NAT to restaging MRI was 44.5 ± 24.9 days, and 30 patients (77%) underwent restaging MRI within 8 weeks of NAT completion.

Of the 25 patients (64%) for whom the post-NAT recommendation was WW, 21 had a sustained cCR. The mean time from restaging MRI to the decision to pursue WW for

these 25 patients was 9.7 ± 19.3 days. Of the 14 patients (36%) for whom the post-NAT recommendation was upfront TME, 13 patients had residual tumor and 1 patient had a pathological complete response. The mean time from restaging MRI to surgery for these 14 patients was 52.5 ± 38.9 days, and 12 patients (86%) underwent surgery within 8 weeks of completing the restaging MRI. The patient with a pathological complete response and the 21 patients with a sustained cCR constituted the complete-response group (n = 22; 56%). The remaining 17 patients (44%) constituted the incomplete-response group. These included the 13 patients treated with upfront TME with residual tumor, 3 patients managed by WW with luminal regrowth, and 1 patient managed by WW for a near-complete response (minimally persistent tumor) that progressed both locoregionally and with external Iliac lymph node metastasis.

Radiologist Characteristics:

We contacted 17 sites that contributed to the OPRA trial for radiologists to participate in this study. Twelve radiologists from eight institutions agreed to participate (Table 3). The group's mean number of rectal MRI scans interpreted per year was 48.9 ± 25.1 (range 20–100).

Accuracy Parameters of Overall Interpretation:

One of the questionnaire items asked respondents to categorize overall response in the T2W and DW images as either complete or incomplete, considering both the primary tumor and the lymph nodes. The radiologists' accuracy in categorizing overall response correctly for the 39 patients was 64% (range 46–72%), with 65% sensitivity (range 32–91%) and 63% specificity (range 35–76%) (Fig. 1, Table 4). Average accuracies of radiologists in practice >10 and ≤10 years were 63 ± 9% and 65 ± 5%, respectively (P = 0.64).

Accuracy Parameters of MRI features:

Accuracy parameters of the radiologists' assessments of the primary tumor response by T2W images, restricted diffusion of DW images and the combination of T2W and DW images are shown in Fig. 2. Accuracy of interpretation of these three features was 63% (range 54–72%), 58% (range 46–74%) and 61% (range 49–69%), respectively. All three features were interpreted with higher sensitivity than specificity. The difference was largest for the presence of restricted diffusion in DW imaging: sensitivity 72% (range 55–100%) vs. specificity 41% (range 12–65%) (Fig. 2. and Table 4).

Accuracy parameters for the interpretation of lymph node involvement were calculated by a different reference standard (see methods) and are shown in Table 4. According to the reference standard, 6 patients had positive nodes and 31 had negative nodes. Two patients had unknown lymph node status since they did not experience a sustained cCR in the rectal mucosa but did not undergo TME. The accuracy of interpretation of the lymph nodes for the remaining 37 patients was high (88%, range 76%−95%), and was interpreted with higher sensitivity (94%, range 77–100%) than specificity (54%, range 17–83%).

Variability of Interpretation of MRI Features:

Differences in the radiologists' interpretations of the overall response and the three key features—tumor response on T2W images, presence of restricted diffusion on DW imaging,

and tumor response on both T2W and DW images— for each of the 39 patients as well as the patients' response in the reference standard are shown in Fig. 3. A detailed heatmap showing the proportion of readers selecting each response for each patient is provided in the supplementary material.

Accuracy of the overall response assessment, and of the interpretation of the T2W and DW images, varies by patient. While for some patients all radiologists (or the overwhelming majority) were able to identify complete response or incomplete response with high accuracy (dark red and dark blue in both extremes of the heatmap in figure 3), for other patients the interpretations were highly variable. The pattern in Fig. 3 suggests an inverse relationship between variability and accuracy, with the highest variability and lowest accuracy for patients in the middle. The patients with the least variability in interpretation (agreement of either 11 or all 12 radiologists), were the five patients on the far left and the five patients on the far right in Fig. 3. The accuracy of interpretation for these patients was 95%. Representative axial T2W and DW images at baseline and restaging from three patients with different variabilities of interpretation are included in the supplementary material.

Fleiss Kappa values were 0.108 to 0.213 (slight to fare agreement) for T2W features, 0.116 to 0.195 (slight agreement) for DW features, 0.428 (moderate agreement) for lymph nodes, 0.202 (fair agreement) for primary tumor response on both T2W and DW images, and 0.247 (fair agreement) for the overall response assessment (see supplementary material).

Discussion

In this study we found a relatively low accuracy of MRI imaging in the identification of patients' complete or incomplete response to NAT. We found a significant variability between radiologists in the accuracy of identifying complete response, and that the variability was not related to experience. Finally, we found that the variability in the interpretation was patient dependent, while in some patients the interpretation was almost uniform, in others the interpretation was almost random.

The accuracy of 64% appears low compared to the 75% accuracy found in a meta-analysis of 16 studies with 790 patients following NAT [21]. The sensitivity reported was 95% (95% CI, 87%–98%), but specificity was only 31% (95% CI, 14%–56%); PPV, 83% (95% CI, 77%– 87%); NPV, 47% (95% CI, 32%–62%) suggesting that MRI tends to both overestimate and underestimate tumor response [21]. Fair agreement (K, 0.24) between an MRI-based regression scale and a pathological regression scale was found for NAT patients from two prospective clinical trials [14]; however, only 4 of 26 patients with pathologic complete response were correctly identified by MRI. Another study reported 88% accuracy for a pattern-based approach combining tumor morphology on T2-weighted images with DWI sequences [13]. However, the reference standard included WW patients followed for only 1 year, and although MRIs from 75 patients were assessed by two readers, for 147 other patients MRIs were assessed only by a single reader, limiting the generalizability of their results.

Our results are consistent with the low accuracy and high interobserver variability in a retrospective analysis that found low correlation between MRI assessments and pathological tumor regression scores [22]. Observers in that study agreed on tumor regression grade 1–2 vs. 3–4 for 60% to 67% of the patients. Haak et al. [23] found that radiologists were relatively accurate in identifying poor response but much less accurate in identifying complete response: only 44%−67% of patients classified as good responders had a complete response in the reference standard, consistent with the accuracy level in our study. In a study by Siddiqui et al. [24], median agreement between 35 radiologists and a highly experienced radiologist who served as the reference standard was 0.57 (with 1 corresponding to complete agreement and 0 corresponding to random chance).

Fleiss Kappa values in our study are lower than expected, because of the large number of readers and multiple response options for most questions. The complexity in the variability of interpretation is evident in the heatmaps in Fig. 3 and supplementary material. In general, agreement was higher for T2W features than for DW features. Agreement was highest for lymph node assessment and second highest for the overall response assessment. The latter finding suggests that readers may arrive at the same overall conclusion through contrasting assessments of individual features.

Although the overall accuracy in our study may seem low, it is important to note that tumor response in WW patients can evolve over time [25]. A longer interval from NAT completion to assessment of response can potentially increase the accuracy of image interpretation. Recent data from our institution indicate that 22% of patients with seemingly false-positive diffusion restriction at a given time point show evidence of tumor on MRI and endoscopy 3 months later [26]. Patients whose response cannot be assessed with confidence at the initial restaging following NAT may benefit from additional frequent reassessments, as suggested for patients with a near-complete response in the OPRA trial [17; 27; 28]

In our study, accuracy increased when all features—the primary tumor and lymph nodes, using both T2W and DWI sequences—were incorporated in the analysis, indicating that data aggregation improves accuracy and reduces interobserver variability. While the primary tumor response was interpreted with similar accuracies on T2W images and DW images, sensitivity was higher and specificity was lower for DW images compared to with T2W images.

MRI interpretation involves an element of judgment susceptible to two types of errors: bias and noise. The former is inaccuracy skewed in a particular direction; the latter corresponds to unwanted random variability [29]. The similar sensitivity and specificity in the overall assessment of response indicate that errors in judgment were made in both overestimation and underestimation of response, excluding bias as the primary source of errors. Although the presence of restricted diffusion and lymph node involvement had an element of bias (high sensitivity and low specificity), the large differences in accuracy between patients and the inverse correlation between overall accuracy and interobserver variability suggest that most interpretation errors are probably related to noise attributable to the intrinsic limitation of MRI in distinguishing between residual tumor and scar. The lack of correlation between accuracy and radiologist experience is consistent with this interpretation. Incorporation

of artificial intelligence into image interpretation may help improve the identification of patients with complete response who are most likely to benefit from an organ-preserving treatment approach.

The strengths of our study include the relatively high number of radiologists, the use of long-term WW and TME reference standards, and interpretation of both baseline and restaging rectal MRI. The high rate of complete response (56%) is consistent with the OPRA trial's organ preservation rates of 43–58% at 3 years [18; 27]. Our study also has potential limitations. The questionnaire used has not been formally validated and may not have included all relevant features for assessment of response. Patient information was collected in a retrospective manner, and the cohort was subject to potential selection bias. The patients were treated by a single surgeon at a high-volume cancer center, and their management may not be generalizable to all practice settings. Finally, MRIs were interpreted without clinical context. Maas et al. [12] found that when endoscopy, T2W, and DW images all indicate complete response, the likelihood of a true complete response is 98%, and when all three modalities indicate incomplete response, the likelihood of a true incomplete response is 85% [12]. These data suggest that multimodal surveillance is crucial for effective and safe WW management.

In summary, our findings indicate that MRI-based evaluation of rectal cancer response at restaging following NAT is associated with low accuracy and substantial variability of interpretation. A few patients have elements of response to NAT that are easily recognizable and are interpreted with high accuracy and low variability; however, that is not the case for most patients. For correct interpretation of response, assessment by more than one modality and at more than one time point is needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

Cap-OX	Capecitabine-oxaliplatin
cCR	Clinical complete response
CRT	Chemoradiotherapy
FOLFOX	Leucovorin-fluorouracil-oxaliplatin
NAT	Neoadjuvant treatment
TME	Total mesorectal excision
TNT	Total neoadjuvant therapy
T2W	T2 weighted
WW	Watch and wait

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 [PubMed: 20692872]
2. Garcia-Aguilar J, Chow OS, Smith DD et al. (2015) Effect of adding mFOLFOX6 after neoadjuvant chemoradiation in locally advanced rectal cancer: a multicentre, phase 2 trial. Lancet Oncol 16:957–966 [PubMed: 26187751]
3. Cercek A, Roxburgh CSD, Strombom P et al. (2018) Adoption of Total Neoadjuvant Therapy for Locally Advanced Rectal Cancer. JAMA Oncol 4:e180071 [PubMed: 29566109]
4. Chen TY, Wiltink LM, Nout RA et al. (2015) Bowel function 14 years after preoperative short-course radiotherapy and total mesorectal excision for rectal cancer: report of a multicenter randomized trial. Clin Colorectal Cancer 14:106–114 [PubMed: 25677122]
5. Maas M, Beets-Tan RG, Lambregts DM et al. (2011) Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. J Clin Oncol 29:4633–4640 [PubMed: 22067400]
6. Araujo RO, Valadao M, Borges D et al. (2015) Nonoperative management of rectal cancer after chemoradiation opposed to resection after complete clinical response. A comparative study. Eur J Surg Oncol 41:1456–1463 [PubMed: 26362228]
7. Martens MH, Maas M, Heijnen LA et al. (2016) Long-term Outcome of an Organ Preservation Program After Neoadjuvant Treatment for Rectal Cancer. J Natl Cancer Inst 108
8. Appelt AL, Ploen 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 [PubMed: 26156652]
9. 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 [PubMed: 23154394]
10. Renehan AG, Malcomson L, Emsley R et al. (2016) Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. Lancet Oncol 17:174–183 [PubMed: 26705854]
11. Smith JJ, Strombom P, Chow OS et al. (2019) Assessment of a Watch-and-Wait Strategy for Rectal Cancer in Patients With a Complete Response After Neoadjuvant Therapy. JAMA Oncol 5:e185896 [PubMed: 30629084]
12. Maas M, Lambregts DM, Nelemans PJ et al. (2015) Assessment of Clinical Complete Response After Chemoradiation for Rectal Cancer with Digital Rectal Examination, Endoscopy, and MRI: Selection for Organ-Saving Treatment. Ann Surg Oncol 22:3873–3880 [PubMed: 26198074]
13. 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 61:328–337 [PubMed: 29369900]
14. Sclafani F, Brown G, Cunningham D et al. (2017) Comparison between MRI and pathology in the assessment of tumour regression grade in rectal cancer. Br J Cancer 117:1478–1485 [PubMed: 28934761]
15. Lambregts DM, 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 [PubMed: 21347783]
16. Aker M, Boone D, Chandramohan A, Sizer B, Motson R, Arulampalam T (2018) Diagnostic accuracy of MRI in assessing tumor regression and identifying complete response in patients with locally advanced rectal cancer after neoadjuvant treatment. Abdom Radiol (NY) 43:3213–3219 [PubMed: 29767284]
17. Smith JJ, Chow OS, Gollub MJ et al. (2015) Organ Preservation in Rectal Adenocarcinoma: a phase II randomized controlled trial evaluating 3-year disease-free survival in patients with locally advanced rectal cancer treated with chemoradiation plus induction or consolidation chemotherapy, and total mesorectal excision or nonoperative management. BMC Cancer 15:767 [PubMed: 26497495]

18. Garcia-Aguilar J, Patil S, Gollub MJ et al. (2022) Organ Preservation in Patients With Rectal Adenocarcinoma Treated With Total Neoadjuvant Therapy. J Clin Oncol. 10.1200/JCO.22.00032:JCO2200032

19. 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 [PubMed: 29043428]

20. Santiago I, Barata M, Figueiredo N et al. (2020) The split scar sign as an indicator of sustained complete response after neoadjuvant therapy in rectal cancer. Eur Radiol 30:224–238 [PubMed: 31350587]

21. de Jong EA, ten Berge JC, Dwarkasing RS, Rijkers AP, van Eijck CH (2016) The accuracy of MRI, endorectal ultrasonography, and computed tomography in predicting the response of locally advanced rectal cancer after preoperative therapy: A metaanalysis. Surgery 159:688–699 [PubMed: 26619929]

22. van den Broek JJ, van der Wolf FS, Lahaye MJ et al. (2017) Accuracy of MRI in Restaging Locally Advanced Rectal Cancer After Preoperative Chemoradiation. Dis Colon Rectum 60:274–283 [PubMed: 28177989]

23. Haak HE, Maas M, Lahaye MJ et al. (2020) Selection of Patients for Organ Preservation After Chemoradiotherapy: MRI Identifies Poor Responders Who Can Go Straight to Surgery. Ann Surg Oncol 27:2732–2739 [PubMed: 32172333]

24. Siddiqui MR, Gormly KL, Bhoday J et al. (2016) Interobserver agreement of radiologists assessing the response of rectal cancers to preoperative chemoradiation using the MRI tumour regression grading (mrTRG). Clin Radiol 71:854–862 [PubMed: 27381221]

25. Yuval JB, Thompson HM, Garcia-Aguilar J (2020) Organ Preservation in Rectal Cancer. J Gastrointest Surg 24:1880–1888 [PubMed: 32314234]

26. Gollub MJ, Das JP, Bates DDB et al. (2021) Rectal cancer with complete endoscopic response after neoadjuvant therapy: what is the meaning of a positive MRI? Eur Radiol. 10.1007/s00330-020-07657-0

27. Garcia-Aguilar J, Patil S, Kim JK et al. (2020) Preliminary results of the organ preservation of rectal adenocarcinoma (OPRA) trial. Journal of Clinical Oncology 38:4008–4008

28. Thompson H, Kim JK, Yuval JB et al. (2021) Survival and organ preservation according to clinical response after total neoadjuvant therapy in locally advanced rectal cancer patients: A secondary analysis from the organ preservation in rectal adenocarcinoma (OPRA) trial. Journal of Clinical Oncology 39:3509–3509

29. Kahneman D, Rosenfield A, Gandhi L, Blaser T (2016) Noise. Harvard Bus Rev:38–46

Key Points

- The overall accuracy of MRI based response assessment is low and radiologists differed in their interpretation of key imaging features.
- Some patients' scans were interpreted with high accuracy and low variability, suggesting that these patients' pattern of response is easier to interpret.
- The most accurate assessments were those of the overall response, which took into consideration both T2W and DWI sequences and the assessment of both the primary tumor and the lymph nodes.

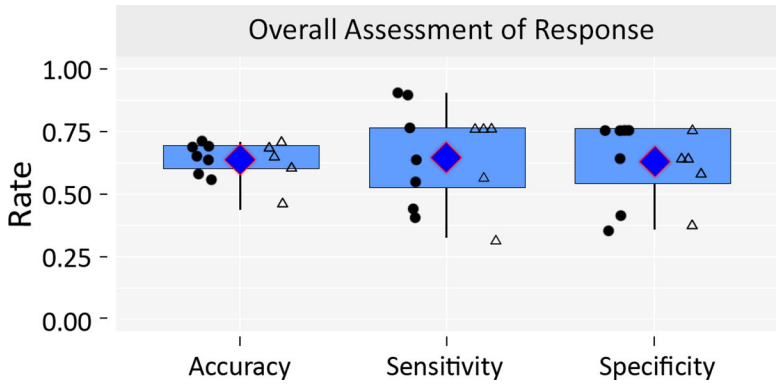

Fig. 1.

Accuracy, sensitivity, and specificity of radiologists' categorization of the overall response to NAT as either complete or incomplete. The dark blue diamonds indicate the accuracy parameter for the group of radiologists as a whole. Light blue boxes indicate the interquartile range. Circles and triangles indicate the accuracy parameters of individual radiologists. Circles indicate more than 10 years attending experience, and triangles indicate 10 or less years of attending experience.

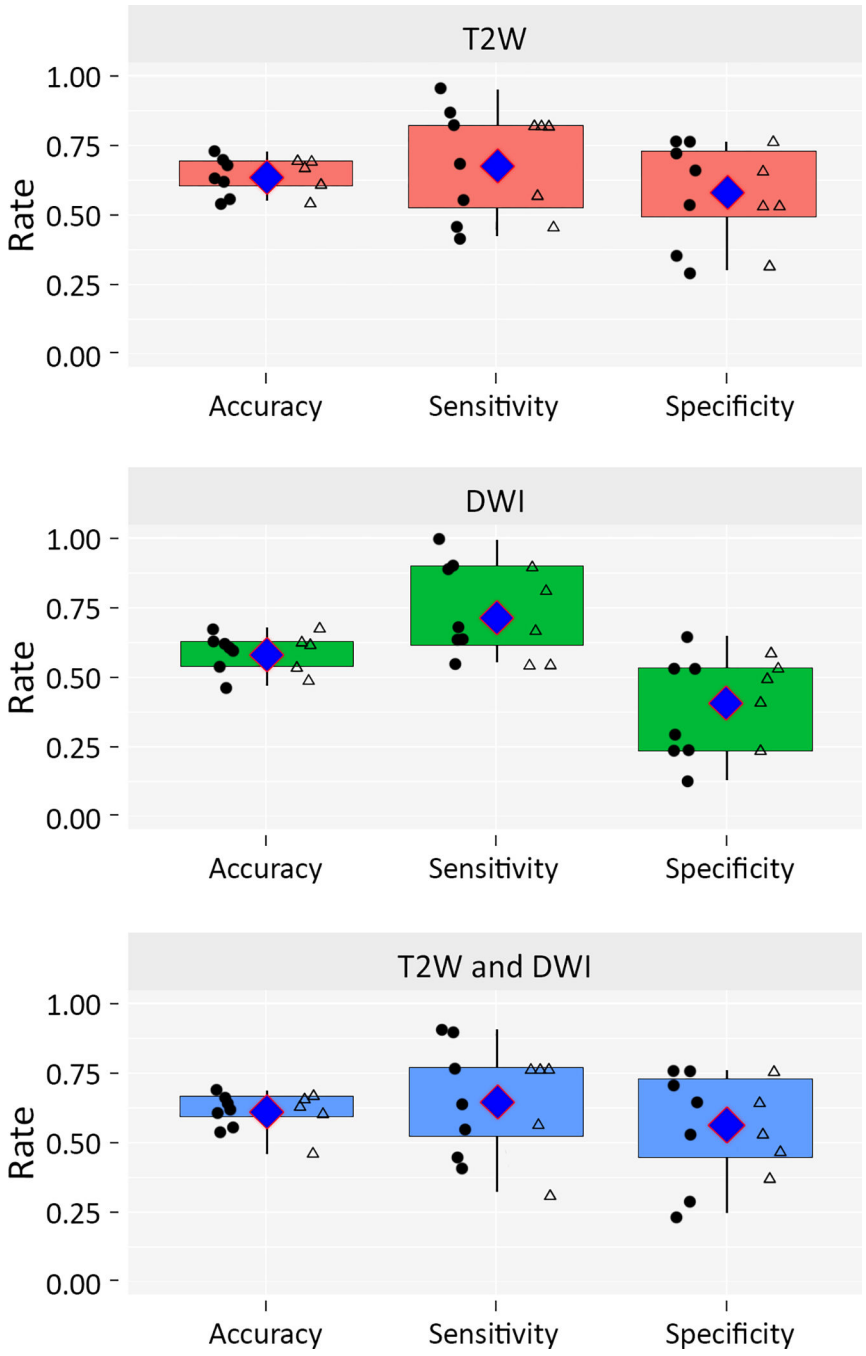

Fig. 2.

Accuracy, sensitivity, and specificity of radiologists' interpretations of the primary tumor response according to T2W images, DW images, and the combination of features of T2W images and features of DW images. The dark blue diamonds indicate the accuracy parameter for the group of radiologists as a whole. Colored boxes indicate the interquartile range. Circles and triangles indicate the accuracy parameters of individual radiologists. Circles indicate more than 10 years attending experience, and triangles indicate 10 or less years of attending experience.

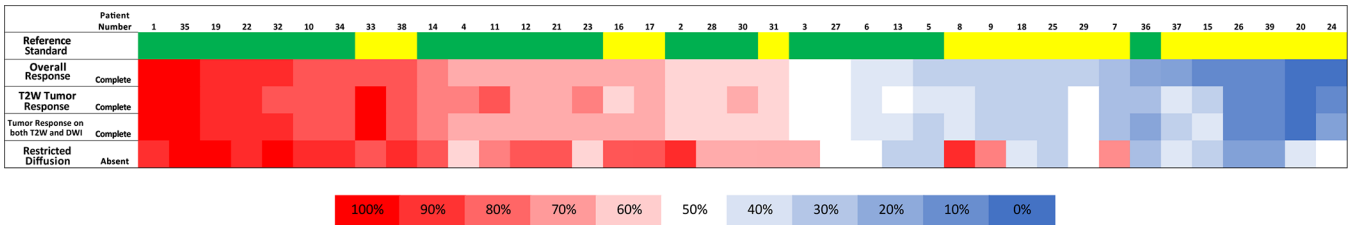

Fig. 3.

Variability among the 12 radiologists in interpreting three key features and in categorizing the overall response to NAT. Patients are listed from left to right in decreasing order of selection of *complete* response in the overall assessment. The top row includes the reference standard (complete response in green; incomplete response in yellow). The color intensity corresponds to the proportion of observers (dark red 100% and dark blue 0%) who selected complete response for the overall assessment and for each one of the features (tumor on T2W sequences alone, tumor on both T2W and DWI sequences, and tumor on DWI sequences alone).

Table 1

MRI features assessed

Feature	Description
T2W images	
Tumor response	General impression of morphologic primary tumor response
DW images	
Presence of restricted diffusion	Presence of restricted diffusion that is not attributed to artifact or T2 shine-through
Both T2W and DW images	
Tumor response	Assessment of primary tumor response
Lymph node involvement	Assessment of lymph node involvement
Overall response	Combination of tumor response and lymph node involvement

Table 2

Patient, tumor, and treatment characteristics

Characteristic	No. (%) of patients (N = 39)
Age[a]	57 ± 10 yr
Female	17 (43.6)
BMI[a]	23.7 ± 4.3
Distance from anal verge[a]	5.4 ± 2. 5 cm
cT category	
1/2	3 (7.7)
3	32 (82.1)
4	4 (10.3)
cN category	
0	5 (12.8)
1/2	34 (87.2)
Initial approach	
WW	25 (64.1)
TME	14 (35.9)
Chemotherapy timing	
Induction	30 (76.9)
Consolidation	9 (23.1)
Time from NAT to restaging MRI[a]	44.5 ± 24.9 days
pCR or sustained cCR	22 (56.4)

BMI, body mass index; *WW*, watch and wait; *TME*, total mesorectal excision; *NAT*, neoadjuvant therapy; *pCR*, pathological complete response; *cCR*, clinical complete response

[a]Mean ± standard deviation.

Table 3

Radiologist characteristics

Characteristic	No. of radiologists (N = 12)
Female	3
Years in practice	
≤10	5
>10	7
MRI rectum reads per week	
≤5	11
6–10	0
≥11	1
Subspecialty	
Gastrointestinal	10
Other	2
Track	
Academic	6
Clinical	1
Equally academic and clinical	5

Table 4

Accuracy parameters of the overall assessment of response and individual MRI features

	Accuracy	Sensitivity	Specificity	PPV	NPV
Overall Assessment	64%	65%	63%	69%	58%
Tumor-T2W and DWI	61%	65%	56%	65%	55%
Tumor-T2W only	63%	67%	57%	70%	57%
Tumor-DWI only	58%	72%	41%	61%	53%
Lymph Nodes	88%	94%	54%	91%	65%