Original Article | Gastrointestinal Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2023.1213
Korean J Radiol 2024;25(4):351-362

Korean Journal of Radiology
KJR

Interpretation of Complete Tumor Response on MRI Following Chemoradiotherapy of Rectal Cancer: Inter-Reader Agreement and Associated Factors in Multi-Center Clinical Practice

Hae Young Kim[1]*, Seung Hyun Cho[2], Jong Keon Jang[3], Bohyun Kim[4], Chul-min Lee[5], Joon Seok Lim[6], Sung Kyoung Moon[7], Soon Nam Oh[4], Nieun Seo[6], Seong Ho Park[3]

[1]Department of Radiology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[2]Department of Radiology, Kyungpook National University Chilgok Hospital, School of Medicine, Kyungpook National University, Daegu, Republic of Korea
[3]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea
[4]Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[5]Department of Radiology, Hanyang University College of Medicine, Seoul, Republic of Korea
[6]Department of Radiology, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea
[7]Department of Radiology, Kyung Hee University Hospital, Seoul, Republic of Korea

Objective: To measure inter-reader agreement and identify associated factors in interpreting complete response (CR) on magnetic resonance imaging (MRI) following chemoradiotherapy (CRT) for rectal cancer.

Materials and Methods: This retrospective study involved 10 readers from seven hospitals with experience of 80–10210 cases, and 149 patients who underwent surgery after CRT for rectal cancer. Using MRI-based tumor regression grading (mrTRG) and methods employed in daily practice, the readers independently assessed mrTRG, CR on T2-weighted images (T2WI) denoted as $mrCR_{T2W}$, and CR on all images including diffusion-weighted images (DWI) denoted as $mrCR_{overall}$. The readers described their interpretation patterns and how they utilized DWI. Inter-reader agreement was measured using multi-rater kappa, and associated factors were analyzed using multivariable regression. Correlation between sensitivity and specificity of each reader was analyzed using Spearman coefficient.

Results: The $mrCR_{T2W}$ and $mrCR_{overall}$ rates varied widely among the readers, ranging 18.8%–40.3% and 18.1%–34.9%, respectively. Nine readers used DWI as a supplement sequence, which modified interpretations on T2WI in 2.7% of cases (36/1341 [149 patients x 9 readers]) and mostly (33/36) changed $mrCR_{T2W}$ to non-$mrCR_{overall}$. The kappa values for mrTRG, $mrCR_{T2W}$, and $mrCR_{overall}$ were 0.56 (95% confidence interval: 0.49, 0.62), 0.55 (0.52, 0.57), and 0.54 (0.51, 0.57), respectively. No use of rectal gel, larger initial tumor size, and higher initial cT stage exhibited significant association with a higher inter-reader agreement for assessing $mrCR_{overall}$ ($P \leq 0.042$). Strong negative correlations were observed between the sensitivity and specificity of individual readers (coefficient, -0.718 to -0.963; $P \leq 0.019$).

Conclusion: Inter-reader agreement was moderate for assessing CR on post-CRT MRI. Readers' varying standards on MRI interpretation (i.e., threshold effect), along with the use of rectal gel, initial tumor size, and initial cT stage, were significant factors associated with inter-reader agreement.

Keywords: Magnetic resonance imaging; Rectal neoplasms; Chemoradiotherapy; Reader; Rater; Observer; Reliability; Reproducibility; Agreement; Variability; Variation

Received: December 6, 2023 **Revised:** January 29, 2024 **Accepted:** February 3, 2024

*Current affiliation: Department of Radiology, Asan Medical Center, Seoul, Republic of Korea
Corresponding author: Seong Ho Park, MD, PhD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Republic of Korea
• E-mail: parksh.radiology@gmail.com

INTRODUCTION

While the standard treatment for locally advanced rectal cancer remains neoadjuvant chemoradiotherapy (CRT) followed by surgical resection, watch-and-wait strategy or minimally invasive treatment are increasingly being considered for patients anticipated to attain pathological complete response (pCR) [1], reported to occur in 10%–25% of patients undergoing CRT [2]. Such organ-preserving strategies, when selectively applied to patients, may preclude perioperative mortality or morbidities such as urinary or sexual dysfunction without compromising survival [1,3].

With the paradigm shift in the treatment strategy, the role of magnetic resonance imaging (MRI) in tumor response assessment and identification of pCR following CRT has become clinically more relevant [4]. However, despite the importance of identifying pCR, there is substantial evidence that the accuracy of MRI in identifying cases of pCR is moderate at best [5]. Microscopic islets of residual tumor cells are frequently dispersed amidst post-CRT fibrosis [6], posing a challenge in their identification even with high-resolution MRI. Nonetheless, MRI remains the preferred diagnostic modality, and is increasingly utilized in real-world clinical practice as a tool for predicting pCR [7,8].

For a diagnostic test to be widely adopted in daily clinical practice beyond research settings, consistency among readers in interpretation is critical. Inter-reader variability can be a concern, even for seemingly well-established interpretative imaging tasks, when applied to daily practice outside the research environment [9-11]. This is particularly true for diagnostic tests with limited accuracy, since difficulties associated with their use may exacerbate inter-observer variability. Thus, it is important to determine the consistency among radiologists in determining complete responders following CRT of rectal cancer on rectal MRI denoted as mrCR. However, the current body of knowledge on this matter is sparse. Apart from two recent studies [12,13], previous studies on inter-reader agreement in evaluating mrCR were generally limited by a small number of readers (three or fewer), strictly controlled research setting, or single-center design [14-21]. Moreover, these studies lacked dedicated analyses on potential factors influencing inter-reader agreement.

In this study, we aimed to measure the agreement among a large number of readers in interpreting mrCR on MRI after CRT for rectal cancer, mirroring real-world clinical practice and explore factors associated with inter-reader agreement. Moreover, we sought to evaluate the accuracy of MRI in predicting pCR within the clinical practice setting.

MATERIALS AND METHODS

Study Setting

This multicenter, retrospective study involved 10 readers from seven general hospitals (Asan Medical Center, Hanyang University Seoul Hospital, Kyung Hee University Hospital, Kyungpook National University Chilgok Hospital, Seoul National University Bundang Hospital, Seoul St. Mary's Hospital, and Severance Hospital) and patients recruited from three of those seven centers (Asan Medical Center, Kyungpook National University Chilgok Hospital, and Severance Hospital). The Institutional Review Boards of the patient-recruiting centers approved the study, and waived the requirement for written informed consent. The study period, based on the timing of post-CRT MRI, spanned from June 2017 to December 2021. This report adhered to relevant reporting guideline [22].

Patients

The study moderators from the three centers each selected 50 eligible patients meeting the following criteria: 1) adults aged > 18 years, 2) non-mucinous rectal cancer with its lower margin located within 10 cm of the anal verge, 3) pre-CRT MRI and post-CRT MRI obtained > 4 weeks after completing CRT, 4) absence of remarkable image artifacts (such as due to endoscopic clip and hip prosthesis) that hindered interpretation, and 5) interval between post-CRT MRI and surgery < 8 weeks. We employed a case-control design in our study. This entailed recruiting an equal number of good (pCR or near pCR [23]) and poor responders, instead of recruiting all patients in a specified period. We reasoned that recruiting consecutive patients, the majority of whom will be poor responders, may result in inflation of inter-reader agreement since agreement was expected to be generally higher among poor responders than among good responders. Therefore, the study moderators at each center initially selected 25 good responders consecutively. Subsequently, they randomly selected 25 poor responders from the remaining eligible patients within the same period. A total of 149 patients were finally included (Table 1), following the exclusion of one patient who was later found to be inadvertently not fulfilling the eligibility criteria (Fig. 1).

https://doi.org/10.3348/kjr.2023.1213

Korean Journal of Radiology
KJR

Table 1. Patient characteristics

Characteristics (n = 149)	All patients
Baseline characteristics before treatment	
Age, yr	63 (57–71)
< 40	6 (4.0)
40–49	11 (7.4)
50–59	45 (30.2)
60–69	43 (28.9)
70–79	35 (23.5)
≥ 80	9 (6.0)
Sex	
Women	60 (40.3)
Men	89 (59.7)
Body mass index, kg/m^2	23.8 (21.8–26.1)
< 18.5 (underweight)	8 (5.4)
18.5–24.9 (normal)	91 (61.1)
25.0–29.9 (overweight)	45 (30.2)
≥ 30.0 (obese)	5 (3.4)
Serum carcinoembryonic antigen, mg/dL	2.4 (1.5–6.1)
< 5.0	103 (69.1)
≥ 5.0	44 (29.5)
Unavailable	2 (1.3)
Tumor distance from the anal verge, cm	5.3 (4.0–7.0)
Tumor size, cm	3.9 (3.0–5.0)
cT stage	
T1 or 2	19 (12.8)
T3ab	65 (43.6)
T3cd	37 (24.8)
T4	28 (18.8)
Treatment-related characteristics	
Radiation dose, Gy	50 (50–50)
Interval from CRT to post-CRT MRI, day	43 (39–50)
Interval from post-CRT MRI to surgery, day	16 (8–32)
Surgery	
Total mesorectal excision	126 (84.6)
Abdominoperineal resection	12 (8.1)
Transanal excision	11 (7.4)
Pathological tumor response grade*	
pCR	32 (21.5)
Near pCR	43 (28.9)
Moderate regression	64 (43.0)
Minimal regression	10 (6.7)
Pathological staging (for non-pCR)	
ypTis or T1	9 (6.0)
ypT2	42 (28.2)
ypT3	63 (42.3)
ypT4	3 (2.0)

Data are median (interquartile range) or number (%). All cancer staging followed the 8th American Joint Committee on Cancer TNM staging system.
*Based on categorization in [23].
CRT = chemoradiotherapy, pCR = pathologic complete response

MRI Protocol

Rectal MRI examinations were performed on either 1.5T or 3T MRI systems. The set of images from one institution were acquired with rectal filling using ultrasound gel, while those from the other two institutions were obtained without rectal filling. High-resolution fast spin-echo T2-weighted images (T2WI) with in-plane resolution of 0.4–0.6 mm pixel size and 3 mm thickness, conforming to expert recommendations [24-28], were used. Diffusion-weighted images (DWI) with b-factor of 1000 s/mm^2 and corresponding apparent diffusion coefficient map were also obtained. The details of the MRI techniques are summarized in Table 2.

Image Review

We used a Clinical Trial Imaging Management System (CTIMS: https://aim-aicro.com/system/AiCRO) [29] for image review. Any identifier of the patient or institution were thoroughly anonymized, and the examinations were randomly reordered in terms of the institution and the timing of image acquisition before being uploaded to the central server of the CTIMS system.

Ten board-certified radiologists with sub-special expertise in abdominal imaging independently reviewed the pre- and post-CRT MR images together, blinded to clinical information and pathology. The readers had varying level of experience with rectal MRI (3–17 years [median, 9 years], average number of monthly examinations ranging 2–77 [median, 9]) (Supplementary Table 1). Under instruction to solely evaluate the primary tumor and disregard other findings such as lymph node status, the readers encoded their assessment on electronic case report form as follows: 1) prediction of pCR based on T2WI alone without DWI ($mrCR_{T2W}$ vs. not) according to their daily practice criteria, 2) assessment of MRI-based tumor regression grading (mrTRG) (grade 1 to 5) [30], and 3) prediction of pCR based on all MR images ($mrCR_{overall}$ vs. not) according to their daily practice. Regarding cases with recently reported "split scar sign [31]," which is not specifically accounted for in the original mrTRG descriptions, the readers were instructed to consider the presence of nodular or mass-like intermediate signal intensity suggestive of residual tumor as mrTRG 3–5 [32]. The readers were not provided with any pre-specified case examples; however, all readers were familiar with relevant literature and illustrative cases in it [6,30].

We designed the image review to closely resemble daily clinical reading practices. Therefore, although the readers

https://doi.org/10.3348/kjr.2023.1213

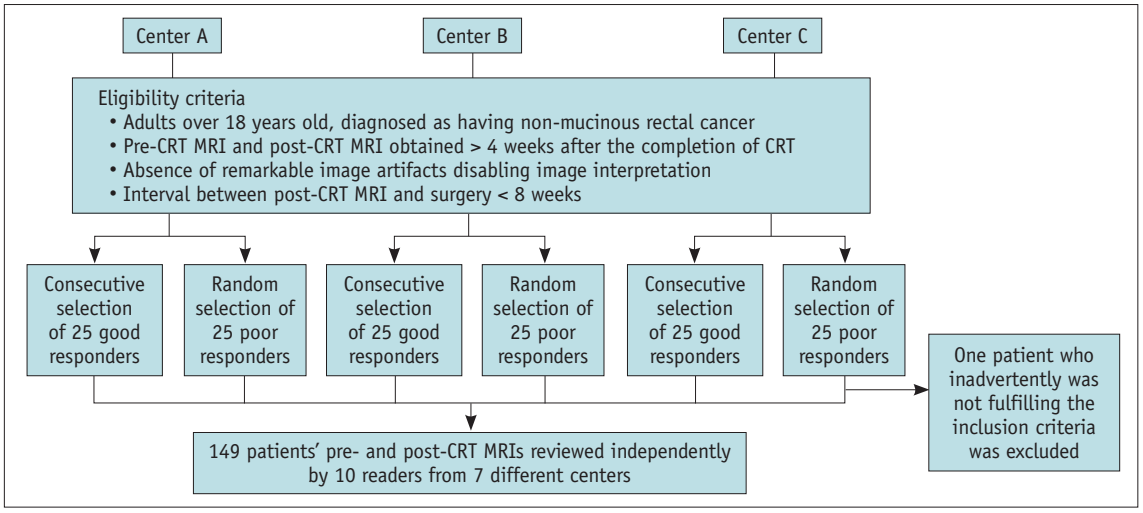

Fig. 1. Patient flow diagram. CRT = chemoradiotherapy

Table 2. Scan parameters of rectal magnetic resonance imaging

Parameter	Asan Medical Center		Kyungpook National University Chilgok Hospital		Severance Hospital	
Rectal gel	No	No	No	No	Yes	Yes
Magnetic field strength	3T	3T	1.5T	1.5T	3T	3T
Sequence	Fast spin-echo T2-weighted imaging	Diffusion-weighted imaging	Fast spin-echo T2-weighted imaging	Diffusion-weighted imaging	Fast spin-echo T2-weighted imaging	Diffusion-weighted imaging
Plane	Axial, coronal, sagittal, and oblique	Axial	Axial, coronal, sagittal, and oblique	Axial	Axial, coronal, sagittal, and oblique	Axial
Repetition time, ms	3300–7070	9500–12970	3920–6300	3100–3900	3800–5500	9500–12000
Echo time, ms	75–120	62–95	77–102	68	80–120	62–95
Flip angle, °	90–150	90	153–180	90	90–150	90
b-factor, s/mm^2	Not applicable	0 and 1000	Not applicable	0 and 1000	Not applicable	0, 300, and 1000
Field of view, mm	180 or 240	220	160, 170, 220	300	180 or 240	220
Matrix without interpolation	320–512	126 or 152	256–384	152	320–512	126–192
Slice thickness, mm	3	4	3	6	3	3–5
Slice gap, mm	0	0	0	0	0	0
Fat suppression	No	Yes	No	Yes	No	Yes

were instructed to initially review T2WI (for $mrCR_{T2W}$ and mrTRG) without referring DWI findings, separate sessions for the review of T2WI and DWI were not implemented. This approach was adopted because in daily practice, DWI is not typically interpreted as a standalone sequence but rather as an adjunctive sequence alongside T2WI. Following the review session, the readers were also prompted to provide narrative description of their criteria for $mrCR_{T2W}$, and how DWI was used.

Clinical and Pathologic Data Collection

Patient characteristics, including demographic findings; initial laboratory results; findings of tumor on pre-CRT MRI; details of neoadjuvant CRT (e.g., radiation dose and dates of initiation and completion); date of post-CRT MRI examination; date and type of surgery; and pathological findings (e.g., pCR status, pTRG grade, and stage), were collected from the electronic medical records of each center (Table 1).

https://doi.org/10.3348/kjr.2023.1213

Korean Journal of Radiology
KJR

Statistical Analysis

Of the 10 study readers, nine routinely utilized both T2WI and DWI in their daily practice and for this study, while one reader solely used T2WI (refer to the Results section for more details). Consequently, any analyses related to $mrCR_{overall}$ incorporated results from these nine readers, while the analyses concerning $mrCR_{T2W}$ and mrTRG included results from all 10 readers.

Inter-reader agreement for $mrCR_{overall}$, $mrCR_{T2W}$, and mrTRG (in its original 5-point scale, and in binary form of grade 1–2 vs. 3–5) were analyzed using multi-rater kappa. Additionally, Gwet's coefficients, which are known to be more robust to the paradoxes associated with kappa [33], were utilized.

We performed multivariable regression analyses to identify factors associated with inter-reader agreement for $mrCR_{overall}$ and $mrCR_{T2W}$. First, the degree of between-reader disagreement in the form of a "mean squared error (MSE)," defined as $\frac{1}{n}\sum$ (rating of each reader – average rating of all readers)^2 was calculated for each patient, and was used as the dependent variable. Second, "proportion of agreement," defined as the proportion of concurrent reading between any two readers among all possible reader pairs (e.g., 36 pairs for nine readers and 45 pairs for 10 readers) was obtained for each patient, and was used as the dependent variable. Multiple covariates were included in the regression as follows: age (year), sex, magnetic field strength (1.5T vs. 3T), use of rectal gel, tumor location (distance in cm from the anal verge), initial tumor size (cm), initial cT staging (≤ cT3b vs. ≥ cT3c, stratified according to the median and prognostic implication [34]), and interval between completion of CRT to post-CRT MRI (day). All continuous variables were analyzed in continuous form. The final regression model was obtained using backward elimination based on Akaike information criterion.

We also analyzed the sensitivity and specificity of MRI in predicting pCR using pathologic analysis of the surgical specimen as the reference standard. The results were obtained for reader pools, accounting for the correlation and variability among the readers using crossed random effects model, and for individual readers. Based on the fitted random effects model, we obtained 95% confidence intervals (CIs) via 1000 times of Monte-Carlo simulation. Additionally, we used Spearman correlation coefficient to confirm the presence of threshold effect (i.e., negative correlation between sensitivity and specificity), a specific cause of inter-reader variability [35]. We also analyzed correlation between reader experience with each of sensitivity and specificity.

All statistical analyses were performed using the R software package (version 4.1.1; R Foundation for Statistical Computing, Vienna, Austria). $P < 0.05$ was considered statistically significant.

RESULTS

Patients

Patient details are described in Table 1. Most patients presented with locally advanced cancer (cT3–4 or cN+ stages as assessed by the initial rectal MRI), while 19 (12.8%) patients exhibited cancers in lower clinical stages (cT1–2). For the 19 patients, CRT was performed for the following reasons: organ (sphincter and/or rectum) preservation due to tumor's proximity to the anus (n = 8), suspected nodal spread (n = 5), and for both of these reasons (n = 6). All patients underwent long-course CRT, consisting of 25 fractions for a total dose of 45–50 Gy. The median interval between completion of CRT and post-CRT MRI was 43 days (IQR, 39–50), and that between post-CRT MRI and surgery was 16 days (IQR, 8–32). There were 32 (21.5%), 43 (28.9%), 64 (43.0%), and 10 (6.7%) patients who exhibited pCR, near pCR, moderate regression, and minimal regression of the tumor, respectively, at pathologic analysis.

Readers' Interpretation Patterns

All readers used T2WI as the primary sequence for evaluating CRT response. All readers consistently considered mrTRG 1 as $mrCR_{T2W}$ and mrTRG 3–5 as non-$mrCR_{T2W}$. Six readers considered all cases of mrTRG 2 as $mrCR_{T2W}$, while the remaining four readers considered most but not all cases of mrTRG 2 as $mrCR_{T2W}$. A wide range of $mrCR_{T2W}$ rates (18.8%–40.3%) was observed across the readers (Supplementary Table 2, Figs. 2, 3). In evaluating $mrCR_{overall}$, nine readers used DWI as a supplement sequence, while one reader did not. In the nine readers who used DWI, their interpretation was modified by DWI in 0 to 12 patients (median, 3 patients) in individual readers, which accounted for 2.7% of all interpretations (36 of 1341 cases [149 patients x 9 readers]) (Supplementary Table 2). The modifications mostly (33/36) changed interpretations from $mrCR_{T2W}$ to non-$mrCR_{overall}$ (Supplementary Table 2), demonstrating the readers' conservative use of DWI.

https://doi.org/10.3348/kjr.2023.1213

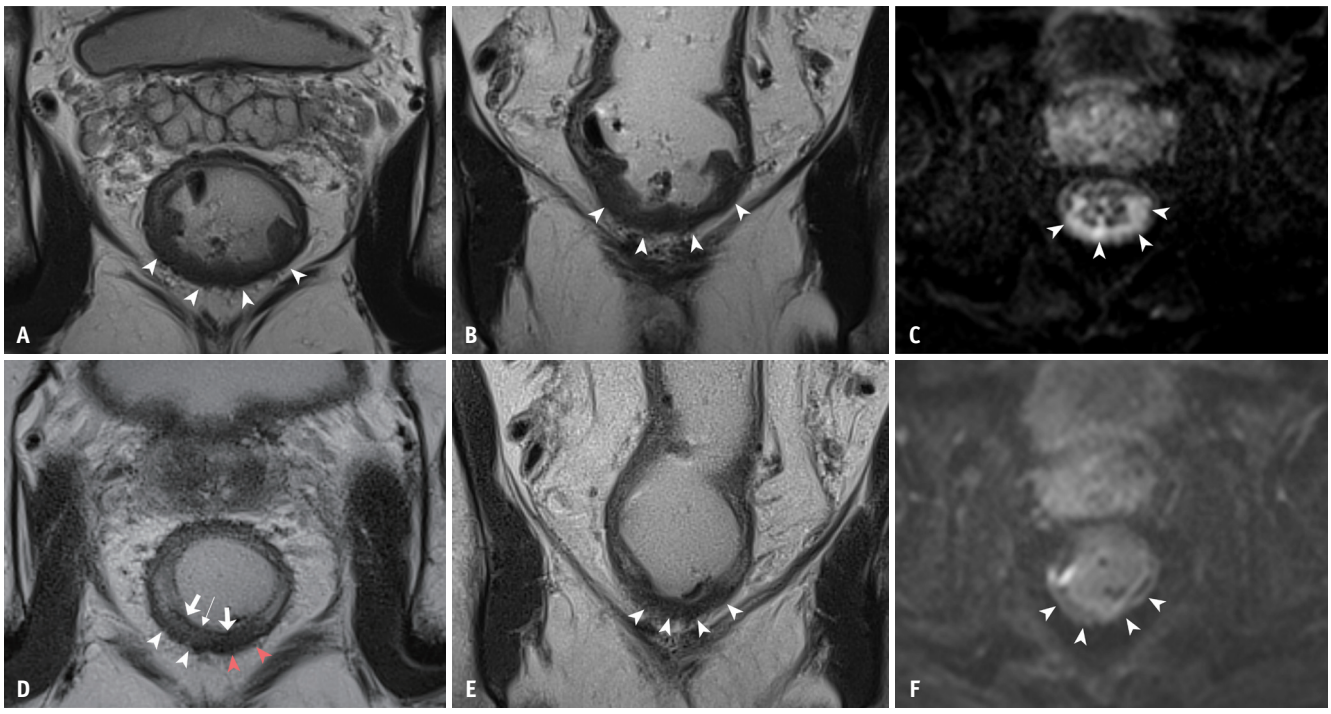

Fig. 2. A 57-year-old man who underwent CRT followed by low anterior resection for rectal cancer, which was pathologically confirmed to be of ypT2N0 staging with near complete response. **A-C:** On pre-CRT MRI, T2W oblique transverse **(A)** and oblique coronal **(B)** images show a cT3 rectal cancer, appearing as an ulcerofungating tumor with raised, rolled-up edges, located approximately at the 3 to 9 o'clock position (arrowheads). DWI **(C)** shows diffusion restriction of the tumor (arrowheads). **D-F:** On post-CRT MRI, T2W oblique transverse **(D)** and oblique coronal **(E)** images show residual thickening of the rectal wall in the tumor bed (white and red arrowheads) despite the decrease in tumor size. On the oblique transverse view **(D)**, a thin, markedly T2-hypointense innermost layer is observed (thick white arrows), creating a "split scar"-like appearance. However, the signal intensity of the underlying wall is not perfectly homogeneous, with a subtle gradation from lighter gray on the patient's right side (white arrowheads) to darker gray on the patient's left side (red arrowheads). Moreover, an equivocal presence of a tiny discontinuity was observed in the T2-hypointense innermost layer (thin white arrow). Similarly, on the oblique coronal view **(E)**, the tumor bed exhibits an overall dark gray signal, for which the readers' opinions were split regarding homogeneously as dark as the muscles in the pelvic wall vs. less dark with areas of subtle light-gray speckles. The post-CRT DWI **(F)** reveals a resolution of diffusion restriction in the tumor bed (arrowheads). Five readers considered the T2WI findings as $mrCR_{T2W}$, while the remaining five did not (non-$mrCR_{T2W}$). The five readers who rendered the non-$mrCR_{T2W}$ interpretation persistently rated the case as non-$mrCR_{overall}$ even after referring to DWI, likely indicating their conservative approach to the interpretation. CRT = chemoradiotherapy, T2W = T2-weighted, DWI = diffusion-weighted image, T2WI = T2-weighted image

Inter-Reader Agreement

The degree of inter-reader agreement for $mrCR_{overall}$ in terms of kappa was 0.54 (95% CI: 0.51, 0.57) for nine readers (after excluding one reader who did not use DWI). The kappa for $mrCR_{T2W}$ was 0.55 (0.52, 0.57). Meanwhile, the kappa for mrTRG in its original 5-point scale with ordinal weighting, and in the binary form (mrTRG 1–2 vs. 3–5) was 0.56 (0.49, 0.62) and 0.54 (0.51, 0.56), respectively (Table 3). The analysis using Gwet's coefficients was consistent with the analysis using kappa, although Gwet's coefficients yielded higher numerical values (Table 3).

The results of multivariable regression analysis regarding the factors associated with inter-reader agreement were consistent between the analyses using MSE and proportion of agreement (Table 4). In evaluating $mrCR_{overall}$, the use of rectal gel, initial tumor size, and initial cT stage showed significant independent association with inter-reader agreement (P, < 0.001 to 0.042). Larger tumor size and higher initial cT stage were associated with a higher inter-reader agreement (with lower MSE and higher proportion of agreement), while the use of rectal gel was associated with a lower inter-reader agreement. In evaluating $mrCR_{T2W}$, only the initial cT stage showed significant association with inter-reader agreement.

Sensitivity and Specificity for pCR

Pooled sensitivity and specificity of $mrCR_{overall}$ for predicting pCR across the nine readers (after excluding one reader who did not use DWI) were 61% (95% CI: 45%, 73%) and 82% (70%, 84%), respectively. Further details

https://doi.org/10.3348/kjr.2023.1213

Korean Journal of Radiology
KJR

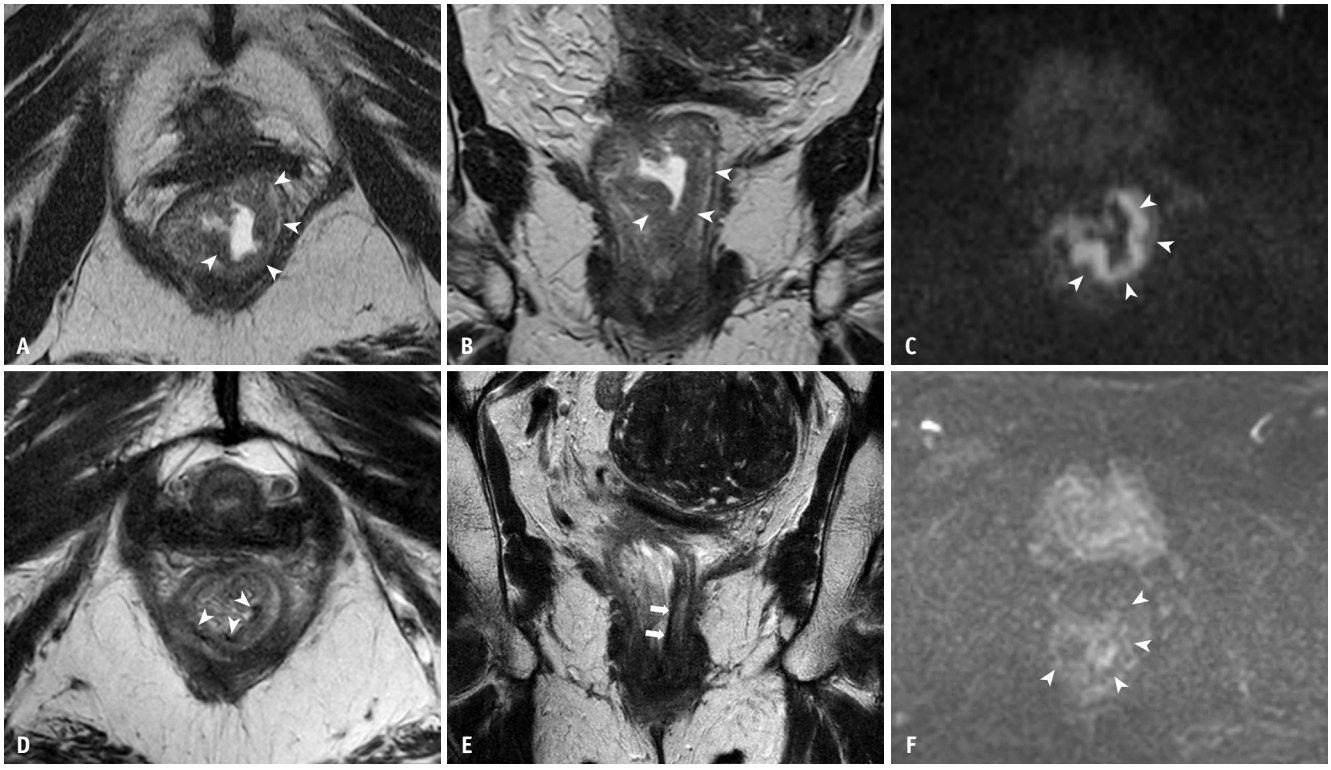

Fig. 3. A 63-year-old woman who underwent CRT followed by abdominoperineal resection for rectal cancer, which was pathologically confirmed to be of ypT2N0 staging with near complete response. **A-C:** On pre-CRT MRI, T2W oblique transverse **(A)** and oblique coronal **(B)** images show a cT2 rectal cancer at 1 to 9 o'clock position (arrowheads). DWI **(C)** shows diffusion restriction of the tumor (arrowheads). **D-F:** On post-CRT MRI, T2W oblique transverse **(D)** and oblique coronal **(E)** images show a remarkably decreased tumor size with dark linear fibrosis, but non-normalized rectal wall. A thin, markedly T2-hypointense innermost layer is noted, which seems to be continuous and split-scar-like on coronal view (arrows) but not on transverse view (arrowheads in **D**). Underlying rectal wall is not completely homogeneous in signal intensity. Post-CRT DWI **(F)** reveals resolution of diffusion restriction in the tumor bed (arrowheads). Six readers considered the case as $mrCR_{T2W}$, while the remaining four did not (non-$mrCR_{T2W}$). The four readers who rendered the non-$mrCR_{T2W}$ interpretation persistently rated the case as non-$mrCR_{overall}$ even after referring to DWI. CRT = chemoradiotherapy, T2W = T2-weighted, DWI = diffusion-weighted image

Table 3. Inter-reader agreement in assessing mrCR and mrTRG on post-CRT MRI

Task	Reader	Multi-rater kappa (95% CI)	Gwet's coefficient (95% CI)
$mrCR_{overall}$	9 readers*	0.54 (0.51, 0.57)	0.69 (0.62, 0.76)
$mrCR_{T2W}$	10 readers	0.55 (0.52, 0.57)	0.66 (0.59, 0.74)
mrTRG†‡	10 readers	0.56 (0.49, 0.62)	0.75 (0.72, 0.77)
$mrTRG_{binary}$†	10 readers	0.54 (0.51, 0.56)	0.62 (0.55, 0.69)

*Excluding one reader who did not use diffusion-weighted image, †A patient for which one reader was unable to assign a specific mrTRG grade was excluded from the analysis. $mrTRG_{binary}$ indicates mrTRG 1-2 vs. 3–5, ‡Ordinal weighting was used for calculating kappa and Gwet's coefficient.
mrTRG = MRI-based tumor regression grading, CRT = chemoradiotherapy, CI = confidence interval, T2W = T2-weighted

are provided in Supplementary Table 3. The results from individual readers are provided in Supplementary Table 4.

There was a statistically significant and strong negative correlation between the sensitivity and specificity of each reader (Table 5) (correlation coefficient, -0.718 to -0.963; *P*, < 0.001 to 0.019). There was no significant correlation between the reader's number of case experience and sensitivity (Table 5) (*P*, 0.108 to 0.89) or specificity (*P*, 0.179 to 0.78).

DISCUSSION

In this study, 10 readers from seven different general hospitals independently reviewed post-CRT MRI according to the interpretation methods used in their daily practice and mrTRG. Overall, the inter-reader agreement was moderate, with point estimates of kappa ranging from 0.54 to 0.56, and Gwet's coefficients ranging from 0.62 to 0.69. The pooled sensitivity and specificity of the nine readers in predicting pCR according to $mrCR_{overall}$ were 61% and 82%,

https://doi.org/10.3348/kjr.2023.1213

Table 4. Multivariable regression analysis for exploring factors associated with inter-reader agreement

Variables	Mean squared error				Proportion of agreement			
	$mrCR_{overall}$*		$mrCR_{T2W}$		$mrCR_{overall}$*		$mrCR_{T2W}$	
	Coefficient	*P*	Coefficient	*P*	Coefficient	*P*	Coefficient	*P*
Age, yr	…	…	…	…	…	…	…	…
Sex (women† vs. men)	…	…	…	…	…	…	…	…
Magnetic field strength (1.5T† vs. 3T)	…	…	…	…	…	…	…	…
Use of rectal gel (no† vs. yes)	0.0322	0.031	0.0217	0.157	-0.0725	0.031	-0.0481	0.157
Distance from the anal verge, cm	…	…	…	…	…	…	…	…
Initial tumor size, cm	-0.0105	0.042	-0.0092	0.081	0.0236	0.042	0.0205	0.081
Initial cT stage (≤ cT3b† vs. ≥ cT3c)	-0.0556	< 0.001	-0.0518	0.001	0.1250	< 0.001	0.1151	0.001
Interval from CRT to MRI, day	…	…	…	…	…	…	…	…

Backward elimination based on Akaike information criterion was used, and variable remaining in the final model are shown. Ellipsis indicates the variables that were eliminated before the final regression model.
*Analyzed for nine readers excluding one reader who did not use diffusion-weighted image, †Reference category.
T2W = T2-weighted, CRT = chemoradiotherapy

Table 5. Correlation among the sensitivity, specificity, and number of case experience of individual readers in predicting pathologic CR after chemoradiotherapy of rectal cancer

	Sensitivity vs. specificity		Sensitivity vs. reader experience		Specificity vs. reader experience	
	Correlation coefficient	*P*	Correlation coefficient	*P*	Correlation coefficient	*P*
Imaging diagnostic criteria: $mrCR_{overall}$*	-0.821	0.007	-0.571	0.108	0.492	0.179
Imaging diagnostic criteria: $mrCR_{T2W}$	-0.963	< 0.001	-0.213	0.55	0.201	0.58
Imaging diagnostic criteria: mrTRG 1-2†	-0.718	0.019	-0.049	0.89	0.104	0.78

Spearman correlation coefficient was used.
*Analyzed for nine readers, excluding one reader who did not use diffusion-weighted image, †A patient for which one reader was unable to assign a specific mrTRG grade was excluded.
CR = complete response, T2W = T2-weighted, mrTRG = MRI-based tumor regression grading

respectively.

While no previous study has employed the same study design as ours, several studies have reported inter-reader agreement in interpreting mrCR, providing a basis for comparison with our results. In studies that included only a few readers, the reported inter-reader agreement for mrTRG in terms of kappa statistics ranged from 0.20 to 0.84, showing heterogeneity [14-21]. A recent study that included 12 readers and 39 patients [13] reported kappa values ≤ 0.247 for inter-reader agreement in interpreting mrCR. In another study involving 22 readers and 90 patients [12], the kappa values for inter-reader agreement for mrCR ranged from 0.40 to 0.60 according to reader experience. While it is encouraging that our results did not lag behind those reported in the previous studies, the moderate degree of inter-reader agreement could nevertheless be further improved.

Our study provides several valuable insights into the factors associated with inter-reader agreement in the interpretation of post-CRT rectal MRI, a topic that has received limited attention. First, technical factors such as magnetic field strength and the use of rectal gel were non-critical. Interestingly, the use of rectal gel, if anything, was not advantageous. This finding is reassuring since it aligns with current practice recommendations, which advocate for the use of both 1.5T and 3T scanners, provided that high-resolution T2WI are acquired [8,24], without necessitating or even discouraging the use of rectal gel [27,28]. Second, a larger initial tumor size and a higher initial cT staging were significantly associated with a higher inter-reader agreement in evaluating $mrCR_{overall}$. This could be attributed to the readers' inclination to perceive that achieving pCR is more challenging in larger and more advanced tumors, which is likely informed by both individual experiences and existing literature [36,37]. Third, although the readers claimed to use largely similar criteria with only slight differences for $mrCR_{T2W}$, and to use DWI similarly as a supplementary sequence in a conservative manner, they seemed to have

https://doi.org/10.3348/kjr.2023.1213

Korean Journal of Radiology
KJR

varying standards for identifying CRT-related changes without residual viable tumor on MRI. This is evidenced by the wide range of mrCR rates across readers and strong negative correlation observed between sensitivities and specificities of individual readers.

In guiding readers through the interpretation of post-CRT MRI, several systems and expert consensus opinions are available, with some being more updated than the others [27,30,38,39]. Unlike mrTRG, several academic societies and expert groups, including the Society of Abdominal Radiology and the European Society of Gastrointestinal and Abdominal Radiology, recommend the acquisition of DWI as an ancillary sequence to T2WI and promote the combined use of T2WI and DWI when predicting pCR [6,27,28,38,39]. However, there remains a lack of clear consensus, supported by sufficient scientific evidence, regarding the precise manner in which findings of T2WI and DWI should be integrated. For instance, some studies [15,40-42] have proposed a conservative approach emphasizing oncologic safety, where only cases with findings suggestive of complete regression on both T2WI and DWI are considered as mrCR. In comparison, other studies used mathematical combination of T2WI and DWI scores [43,44]. Given these considerations, establishing a more comprehensive and standardized system for interpreting post-CRT rectal MRI, such as a RADS that encompasses DWI, may contribute to enhanced inter-reader agreement by mitigating variability stemming from differing individual thresholds. Such system should provide more detailed and clearer interpretive guidelines while considering disease probability or reader confidence (e.g., assigning a higher RADS score to indicate a higher probability of pCR).

While not covered in our study, in addition to efforts aimed at enhancing the reliability of interpreting T2WI and DWI, investigating potential benefits of other imaging techniques to augment inter-reader agreement may be valuable for future research. This might include incorporating post-contrast sequences or leveraging enhanced image quality and interpretive assistance provided by artificial intelligence, which are areas that remain largely unexplored.

Our results on the diagnostic performance of MRI for pCR were comparable to those in the literature, reaffirming the limited accuracy [5]. While reader experience is an important factor that affects diagnostic accuracy, our results did not reveal a significant association between the accuracy and degree of case experience. This is likely due to the fact that all 10 readers in our study already had a relatively sufficient level of experience, with a minimum of 80 cases. It is possible that improvement in performance reaches a plateau after surpassing a certain threshold of experience.

Our study had limitations. First, our study may be subject to biases inherent in retrospective research of a case-control design. Second, image review was performed in a single session with both T2WI and DWI available, aiming to simulate daily practice. Consequently, interpretation of one sequence may have inadvertently influenced the another. Third, we did not analyze disease other than primary tumor, such as nodal metastasis or extramural venous invasion. Radiologists may also consider the response of other disease component when evaluating primary tumor in practice. Therefore, the interpretation in this study may have been somewhat incomplete and artificial. Lastly, the results of endoscopy or digital rectal examination were made unavailable during image review in this study. The presence of such additional information in daily clinical setting may change the degree of inter-reader agreement.

In conclusion, 10 readers demonstrated moderate agreement in assessing CR on post-CRT MRI. Readers' varying standards on MRI interpretation (i.e., threshold effect), along with the use of rectal gel, initial tumor size, and initial cT stage, were significant factors associated with inter-reader agreement. Further standardization of post-CRT MRI interpretation would be needed.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2023.1213.

Availability of Data and Material

Data generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest

Joon Seok Lim and Seong Ho Park, who hold respective positions on the Editorial Board Member and Editor-in-Chief of the *Korean Journal of Radiology*, were not involved in the editorial evaluation or decision to publish this article. The remaining author has declared no conflicts of interest.

Author Contributions

Conceptualization: all authors. Data curation: Seung Hyun Cho, Nieun Seo, Jong Keon Jang. Formal analysis: Hae Young Kim, Seong Ho Park. Funding acquisition: Seong Ho

Park. Investigation: all authors. Methodology: Hae Young Kim, Seong Ho Park. Supervision: Seong Ho Park. Writing—original draft: Hae Young Kim, Seong Ho Park. Writing—review & editing: all authors.

ORCID IDs
Hae Young Kim
https://orcid.org/0000-0002-9508-4280
Seung Hyun Cho
https://orcid.org/0000-0001-7617-7302
Jong Keon Jang
https://orcid.org/0000-0002-2938-6635
Bohyun Kim
https://orcid.org/0000-0003-1157-415X
Chul-min Lee
https://orcid.org/0000-0001-7621-3377
Joon Seok Lim
https://orcid.org/0000-0002-0334-5042
Sung Kyoung Moon
https://orcid.org/0000-0003-4831-3439
Soon Nam Oh
https://orcid.org/0000-0003-2373-7024
Nieun Seo
https://orcid.org/0000-0001-8745-6454
Seong Ho Park
https://orcid.org/0000-0002-1257-8315

Funding Statement
This study was supported by a research fund from the Korean Society of Radiology through Radiology Imaging Network of Korea for Clinical Research (RINK-CR) and by the Korean Society of Abdominal Radiology.

REFERENCES

1. van der Valk MJM, Hilling DE, Bastiaannet E, Meershoek-Klein Kranenbarg E, Beets GL, Figueiredo NL, 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
2. Smith FM, Cresswell K, Myint AS, Renehan AG. Is "watch-and-wait" after chemoradiotherapy safe in patients with rectal cancer? *BMJ* 2018;363:k4472
3. Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo LJ, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010;11:835-844
4. Jayaprakasam VS, Alvarez J, Omer DM, Gollub MJ, Smith JJ, Petkovska I. Watch-and-wait approach to rectal cancer: the role of imaging. *Radiology* 2023;307:e221529
5. Jang JK, Choi SH, Park SH, Kim KW, Kim HJ, Lee JS, 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
6. Park SH, Cho SH, Choi SH, Jang JK, Kim MJ, Kim SH, et al. MRI assessment of complete response to preoperative chemoradiation therapy for rectal cancer: 2020 guide for practice from the Korean Society of Abdominal Radiology. *Korean J Radiol* 2020;21:812-828
7. Benson AB, Venook AP, Al-Hawary MM, Azad N, Chen YJ, Ciombor KK, et al. Rectal cancer, version 2.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2022;20:1139-1167
8. Bates DDB, Shaish H, Gollub MJ, Harisinghani M, Lall C, Sheedy SP. Multi-practice survey on MR imaging practice patterns in rectal cancer in the United States. *Abdom Radiol (NY)* 2022;47:28-37
9. Hong CW, Chernyak V, Choi JY, Lee S, Potu C, Delgado T, et al. A multicenter assessment of interreader reliability of LI-RADS version 2018 for MRI and CT. *Radiology* 2023;307:e222855
10. Hsieh SS, Cook DA, Inoue A, Gong H, Sudhir Pillai P, Johnson MP, et al. Understanding reader variability: a 25-radiologist study on liver metastasis detection at CT. *Radiology* 2023;306:e220266
11. Min JH, Lee MW, Park HS, Lee DH, Park HJ, Lim S, et al. Interobserver variability and diagnostic performance of gadoxetic acid-enhanced MRI for predicting microvascular invasion in hepatocellular carcinoma. *Radiology* 2020;297:573-581
12. El Khababi N, Beets-Tan RGH, Tissier R, Lahaye MJ, Maas M, Curvo-Semedo L, et al. Comparison of MRI response evaluation methods in rectal cancer: a multicentre and multireader validation study. *Eur Radiol* 2023;33:4367-4377
13. Yuval JB, Patil S, Gangai N, Omer DM, Akselrod DG, Fung A, et al. MRI assessment of rectal cancer response to neoadjuvant therapy: a multireader study. *Eur Radiol* 2023;33:5761-5768
14. Aker M, Boone D, Chandramohan A, Sizer B, Motson R, Arulampalam T. 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)* 2018;43:3213-3219
15. Lee MA, Cho SH, Seo AN, Kim HJ, Shin KM, Kim SH, et al. Modified 3-point MRI-based tumor regression grade incorporating DWI for locally advanced rectal cancer. *AJR Am J Roentgenol* 2017;209:1247-1255
16. Sclafani F, Brown G, Cunningham D, Wotherspoon A, Mendes LST, Balyasnikova S, et al. Comparison between MRI and pathology in the assessment of tumour regression grade in rectal cancer. *Br J Cancer* 2017;117:1478-1485
17. Siddiqui MR, Gormly KL, Bhoday J, Balyansikova S, Battersby NJ, Chand M, 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
18. van den Broek JJ, van der Wolf FS, Lahaye MJ, Heijnen LA, Meischl C, Heitbrink MA, et al. Accuracy of MRI in restaging locally advanced rectal cancer after preoperative chemoradiation. *Dis Colon Rectum* 2017;60:274-283
19. Voogt ELK, Nordkamp S, van Zoggel DMGI, Daniëls-Gooszen AW, Nieuwenhuijzen GAP, Bloemen JG, et al. MRI tumour regression grade in locally recurrent rectal cancer. *BJS Open* 2022;6:zrac033
20. Yoen H, Park HE, Kim SH, Yoon JH, Hur BY, Bae JS, et al. Prognostic value of tumor regression grade on MR in rectal cancer: a large-scale, single-center experience. *Korean J Radiol* 2020;21:1065-1076
21. Jang JK, Lee JL, Park SH, Park HJ, Park IJ, Kim JH, et al. Magnetic resonance tumour regression grade and pathological correlates in patients with rectal cancer. *Br J Surg* 2018;105:1671-1679
22. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106
23. Kim SH, Chang HJ, Kim DY, Park JW, Baek JY, Kim SY, et al. What is the ideal tumor regression grading system in rectal cancer patients after preoperative chemoradiotherapy? *Cancer Res Treat* 2016;48:998-1009
24. Gormly KL. High-resolution T2-weighted MRI to evaluate rectal cancer: why variations matter. *Korean J Radiol* 2021;22:1475-1480
25. Royal College of Radiologists. Recommendations for cross-sectional imaging in cancer management, second edition [accessed on November 10, 2023]. Available at: https://www.rcr.ac.uk/media/mv1hidxf/rcr-publications_recommendations-for-cross-sectional-imaging-in-cancer-management-second-edition-12-colon-rectum-and-anal-canal-cancer_april-2022.pdf
26. Cancer Council Australia. Clinical practice guidelines for the prevention, early detection and management of colorectal cancer [accessed on November 10, 2023]. Available at: https://www.cancer.org.au/clinical-guidelines/bowel-cancer/colorectal-cancer
27. Beets-Tan RGH, Lambregts DMJ, Maas M, Bipat S, Barbaro B, Curvo-Semedo L, et al. 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;28:1465-1475
28. Gollub MJ, Arya S, Beets-Tan RG, dePrisco G, Gonen M, Jhaveri K, 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 (NY)* 2018;43:2893-2902
29. Shin Y, Kim KW, Lee AJ, Sung YS, Ahn S, Koo JH, et al. A good practice-compliant clinical trial imaging management system for multicenter clinical trials: development and validation study. *JMIR Med Inform* 2019;7:e14310
30. Bhoday J, Smith F, Siddiqui MR, Balyasnikova S, Swift RI, Perez R, et al. Magnetic resonance tumor regression grade and residual mucosal abnormality as predictors for pathological complete response in rectal cancer postneoadjuvant chemoradiotherapy. *Dis Colon Rectum* 2016;59:925-933
31. Santiago I, Barata M, Figueiredo N, Parés O, Henriques V, Galzerano A, et al. The split scar sign as an indicator of sustained complete response after neoadjuvant therapy in rectal cancer. *Eur Radiol* 2020;30:224-238
32. Patel UB, Brown G, Rutten H, West N, Sebag-Montefiore D, Glynne-Jones R, et al. Comparison of magnetic resonance imaging and histopathological response to chemoradiotherapy in locally advanced rectal cancer. *Ann Surg Oncol* 2012;19:2842-2852
33. Quarfoot D, Levine RA. How robust are multirater interrater reliability indices to changes in frequency distribution? *Am Stat* 2016;70:373-384
34. Merkel S, Mansmann U, Siassi M, Papadopoulos T, Hohenberger W, Hermanek P. The prognostic inhomogeneity in pT3 rectal carcinomas. *Int J Colorectal Dis* 2001;16:298-304
35. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889-897
36. Tan Y, Fu D, Li D, Kong X, Jiang K, Chen L, et al. Predictors and risk factors of pathologic complete response following neoadjuvant chemoradiotherapy for rectal cancer: a population-based analysis. *Front Oncol* 2019;9:497
37. Lorimer PD, Motz BM, Kirks RC, Boselli DM, Walsh KK, Prabhu RS, et al. Pathologic complete response rates after neoadjuvant treatment in rectal cancer: an analysis of the national cancer database. *Ann Surg Oncol* 2017;24:2095-2103
38. Lee S, Kassam Z, Baheti AD, Hope TA, Chang KJ, Korngold EK, et al. Rectal cancer lexicon 2023 revised and updated consensus statement from the Society of Abdominal Radiology Colorectal and Anal Cancer Disease-Focused Panel. *Abdom Radiol (NY)* 2023;48:2792-2806
39. Fokas E, Appelt A, Glynne-Jones R, Beets G, Perez R, Garcia-Aguilar J, et al. International consensus recommendations on key outcome measures for organ preservation after (chemo) radiotherapy in patients with rectal cancer. *Nat Rev Clin Oncol* 2021;18:805-816
40. Jang JK, Lee CM, Park SH, Kim JH, Kim J, Lim SB, et al. How to combine diffusion-weighted and T2-weighted imaging for MRI assessment of pathologic complete response to neoadjuvant chemoradiotherapy in patients with rectal cancer? *Korean J Radiol* 2021;22:1451-1461
41. Lambregts DMJ, Delli Pizzi A, Lahaye MJ, van Griethuysen JJM, Maas M, Beets GL, et al. 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* 2018;61:328-337
42. Maas M, Lambregts DM, Nelemans PJ, Heijnen LA, Martens MH, Leijtens JW, et al. Assessment of clinical complete response

after chemoradiation for rectal cancer with digital rectal examination, endoscopy, and MRI: selection for organ-saving treatment. *Ann Surg Oncol* 2015;22:3873-3880
43. Hall WA, Li J, You YN, Gollub MJ, Grajo JR, Rosen M, et al. Prospective correlation of magnetic resonance tumor regression grade with pathologic outcomes in total neoadjuvant therapy for rectal adenocarcinoma. *J Clin Oncol* 2023;41:4643-4651
44. Chandramohan A, Siddiqi UM, Mittal R, Eapen A, Jesudason MR, Ram TS, et al. Diffusion weighted imaging improves diagnostic ability of MRI for determining complete response to neoadjuvant therapy in locally advanced rectal cancer. *Eur J Radiol Open* 2020;7:100223

https://doi.org/10.3348/kjr.2023.1213