Grading Crohn Disease Activity With MRI: Interobserver Variability of MRI Features, MRI Scoring of Severity, and Correlation With Crohn Disease Endoscopic Index of Severity

Jeroen A. W. Tielbeek[1]
Jesica C. Makanyanga[2]
Shandra Bipat[1]
Doug A. Pendsé[2]
C. Yung Nio[1]
Frans M. Vos[1,3]
Stuart A. Taylor[2]
Jaap Stoker[1]

[1]Department of Radiology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. Address correspondence to J. A. W. Tielbeek (j.a.w.tielbeek@amc.uva.nl).

[2]Centre for Medical Imaging, University College London, London, United Kingdom.

[3]Quantitative Imaging Group, Department of Imaging Science and Technology, Delft University of Technology, Delft, The Netherlands.

OBJECTIVE. The purpose of this article is to assess the interobserver variability for scoring MRI features of Crohn disease activity and to correlate two MRI scoring systems to the Crohn disease endoscopic index of severity (CDEIS).

MATERIALS AND METHODS. Thirty-three consecutive patients with Crohn disease undergoing 3-T MRI examinations (T1-weighted with IV contrast medium administration and T2-weighted sequences) and ileocolonoscopy within 1 month were independently evaluated by four readers. Seventeen MRI features were recorded in 143 bowel segments and were used to calculate the MR index of activity and the Crohn disease MRI index (CDMI) score. Multirater analysis was performed for all features and scoring systems using intraclass correlation coefficient (icc) and kappa statistic. Scoring systems were compared with ileocolonoscopy with CDEIS using Spearman rank correlation.

RESULTS. Thirty patients (median age, 32 years; 21 women and nine men) were included. MRI features showed fair-to-good interobserver variability (intraclass correlation coefficient or kappa varied from 0.30 to 0.69). Wall thickness in millimeters, presence of edema, enhancement pattern, and length of the disease in each segment showed a good interobserver variability between all readers (icc = 0.69, κ = 0.66, κ = 0.62, and κ = 0.62, respectively). The MR index of activity and CDMI scores showed good reproducibility (icc = 0.74 and icc = 0.78, respectively) and moderate CDEIS correlation (*r* = 0.51 and *r* = 0.59, respectively).

CONCLUSION. The reproducibility of individual MRI features overall is fair to good, with good reproducibility for the most commonly used features. When combined into the MR index of activity and CDMI score, overall reproducibility is good. Both scores show moderate agreement with CDEIS.

C rohn disease is a chronic inflammatory bowel disease that can cause a wide variety of symptoms. Several scoring systems that can grade disease activity are already well established in the management of luminal Crohn disease [1]. The Crohn disease endoscopic index of severity (CDEIS), histopathologic grading according to Borley et al. [2], and imaging scores such as that for perianal Crohn disease are increasingly used [3]. However, there is no universally accepted grading of Crohn disease activity, and we are left with a very important clinical problem: Can we grade disease activity with any method, and more importantly, how can we predict the outcome of medical therapy?

MRI might overcome this problem because it evaluates the bowel lumen, bowel wall, and extraenteric soft tissues without the use of ionizing radiation. Therefore, MRI is increasingly used to objectively assess Crohn disease activity and to guide management [4–7]. Numerous MRI features have been proposed as markers of disease activity, either alone or together in varying combinations [8–11]. Despite its promise, distinct interobserver variability has been reported for many of these MRI features [9, 12, 13].

Clearly, grading of MRI features in Crohn disease must be reproducible when used by different observers in different centers to be clinically useful. Increasing data suggest that robust MRI assessment of Crohn disease activity should be based on integrating several imaging features together, rather than relying on one or two individual findings [14]. The combination of selected MRI features into one strong scoring system could lead to a more objective, quantitative, and reproducible grading in the severity of Crohn disease and is, therefore, recommended.

Recently, two groups have developed a quantitative scoring system for Crohn disease activity: the MR index of activity and Crohn disease MRI index (CDMI) score [13, 15]. The MR index of activity has a reported high correlation to the CDEIS ($r = 0.80$; $p < 0.001$), whereas the CDMI score has a high correlation to a histopathology score (estimated acute inflammation score) (Kendall τ b = 0.48; $p = 0.002$). Therefore, either scoring system could be considered for assessing Crohn disease activity, but their reproducibility needs to be evaluated before wider clinical implementation. Furthermore, to our knowledge, no study has compared the accuracy of these two scoring systems in an external patient cohort.

The primary aim of this study was to assess the reproducibility of MRI features and scoring systems in patients with Crohn disease. The secondary aim was to correlate these scoring systems with the CDEIS in an external patient cohort.

Materials and Methods
Patients

Data from 33 consecutive patients with Crohn disease proven at histopathologic analysis were analyzed. These patients had taken part in a prospective single-center study comparing dynamic contrast-enhanced (DCE) MR enterography to ileocolonoscopy with CDEIS. The indication for ileocolonoscopy was clinical suspicion of relapsing Crohn disease. Exclusion criteria were age younger than 18 years, contraindications for MRI (including pacemakers, metallic implants, severe claustrophobia, and pregnancy), technical failure of a sequence, incomplete reference standard (CDEIS), and a negative diagnosis for Crohn disease. All patients had been recruited between February 2009 and November 2010 for assessment of Crohn disease activity. Furthermore, for all patients, ileocolonoscopy had been performed within a month of the MR enterography. The results of that study have been published previously [16].

Patient exclusion criteria for this study were nondiagnostic MR enterography image quality (i.e., the study was not of sufficient quality to determine disease activity, if present) as determined by one or more of the readers and an incomplete MR enterography scan protocol (i.e., not including T2-weighted single-shot fast spin-echo [FSE], T2-weighted fat-saturated single-shot FSE, or 3D T1-weighted contrast-enhanced sequences, which are mandatory for calculating the scoring systems). Per-segment exclusion criteria were resected bowel segments and insufficient distention or visibility (< 20% of the bowel adequately distended and visible) of a bowel segment, as determined by one of the readers.

The Crohn disease activity index (CDAI) score [17] and C-reactive protein levels were assessed in all patients. A CDAI score greater than 150 or a C-reactive protein level greater than 8 mg/L was considered as active disease.

For the previous study, ethical permission was obtained from the hospital medical ethics committee, and written informed consent was obtained from all patients. For the current study, informed consent was waived by the hospital medical ethics committee.

Finally, three of the 33 patients in the dataset of the prior study were excluded. Two observers assessed the quality of one MR enterography study (one patient) as nondiagnostic, and for two other MR enterography examinations, no T2-weighted fat saturation sequence was available (two patients). Thus, 30 patients (median age, 32 years; age range, 19–72 years; 21 women and nine men) were evaluated. The CDAI values showed that 47% ($n = 14$) of the patients had active disease. The C-reactive protein values showed that 50% ($n = 15$) of the patients had active disease. Baseline characteristics are shown in Table 1.

These 30 patients had 148 segments (in two patients only four segments were eligible after right hemicolectomy) of which five segments, all rectal, were excluded because of insufficient visibility, resulting in 143 evaluable segments. The remaining 143 segments were radiologically scored by the four observers.

MR Enterography Protocol

The protocol of the study has been published previously [16]. Patients fasted 4 hours before the examination and drank 1600 mL of mannitol (2.5%; Osmitrol, Baxter) solution 1 hour before the scan. Supine images were acquired using a 3-T MRI unit (Intera, Philips Healthcare) with a 16-channel torso phased-array body coil. Axial and coronal T2-weighted single-shot FSE sequences with and without fat saturation were acquired, followed by a coronal 3D T1-weighted spoiled gradient-echo sequence with fat saturation. After these series, 20 mg of butylscopolamine bromide (Buscopan, Boehringer Ingelheim) was IV administered, and a DCE-MRI sequence with 0.1 mL/kg bodyweight of gadobutrol (1.0 mmol/mL; Gadovist, Bayer Schering Pharma) was obtained. Ten seconds after the start of the dynamic sequence, 0.1 mL/kg bodyweight of gadobutrol (1.0 mmol/mL) was injected IV by bolus injec-

TABLE 1: Demographic Characteristics and Severity Indexes of the Study Population

Characteristic	Value
No. of patients	30
Men	9
Women	21
Age at time of imaging (y), median (IQR)	32 (26–45)
Disease duration (y), median (IQR)	9 (6–13)
Days between ileocolonoscopy and MR enterography, median (IQR)	7.5 (4.8–14)
Previous surgery, no. (%) of patients	16 (53)
Maintenance therapy, no. (%) of patients	22 (73)
Antitumor necrosis factor, no. (%) of patients	9 (30)
Steroids, no. (%) of patients	6 (20)
Purine-antagonists, no. (%) of patients	12 (40)
5-Aminosalicylic acid medications, no. (%) of patients	2 (7)
Methotrexate, no. (%) of patients	2 (7)
C-reactive protein level (mg/L), median (IQR)	8.3 (1.3–23.4)
Crohn disease activity index, median (IQR)	140.8 (80.5–237)
Crohn disease endoscopic index of severity	
Mean ± SD	5.4 ± 5.5
Median (IQR)	4.3 (0–23.4)
< 3.5, no. (%) of patients	12 (40)
3.5–7, no. (%) of patients	11 (36.7)
> 7, no. (%) of patients	7 (23.3)

Note—IQR = interquartile range.

TABLE 2: Sequences at 3 T Used to Assess All MRI Features [16]

Parameter	T2-Weighted Single-Shot Fast Spin-Echo		3D T1-Weighted Spoiled Gradient-Echo Sequence	
	Axial and Coronal	Axial and Coronal	Coronal	Axial
Fat saturation	No	Yes	Yes	Yes
TR/TE	516–758/65–118	1370–1450/70	1.87–2.19/1.0	1.87–2.19/1.0
Flip angle (°)	90	90	10	10
Slice thickness/gap (mm)	4/1	7/1	Not applicable	Not applicable
No. of slices	40	45	100	180
FOV (mm)	400 × 400	375 × 300	400 × 400 × 200	400 × 400 × 140
Matrix	256 × 256	288 × 288	192 × 192 × 100	208 × 208 × 70
Sensitivity encoding factor	2.5	2	1.5	2

tion (5 mL/s) through a 20-gauge IV catheter using an automated injection pump (Mallinckrodt Optistar, Liebel-Flarsheim). Injection of contrast medium was immediately followed by a bolus of 15 or 20 mL saline (5 mL/s), depending on the length of the contrast injection tube. The duration of the DCE-MRI sequence was 6 minutes. After these series, a second dose of 20 mg of butylscopolamine bromide was IV administered. Thereafter, contrast-enhanced axial and coronal 3D T1-weighted spoiled gradient-echo sequences with fat saturation were performed. All sequences were used for image analysis, except the DCE-MRI sequence.

Observers

Four readers from two tertiary centers in different countries with 18 years (700 MR enterography studies), 17 years (1100 MR enterography studies), 4 years (170 MR enterography studies), and 1 year (160 MR enterography studies), of experience in reading abdominal MRI evaluated the MRI scans using the axial and coronal T2-weighted single-shot FSE with and without fat saturation, coronal unenhanced, and axial and coronal contrast-enhanced 3D T1-weighted spoiled gradient-echo sequences (Table 2). All readers used a PACS (Impax 5.0, AGFA Healthcare, Agfa-Gevaert) workstation. All readers were unaware of the findings at the initial reading and the findings from ileocolonoscopy but were aware of patients' surgical history. The small bowel and the colon were divided into five segments: terminal ileum, right colon (cecum plus ascending colon), transverse colon, left colon (descending colon plus sigmoid), and rectum, so there could be a direct segment comparison between MRI and the CDEIS.

MRI Features

Seventeen different MRI features (Table 3) were evaluated by all readers. Features were selected according to the MRI features described in the literature and used by most abdominal radiologists as identified in an international inventory, together with those used in the two published scoring systems [13, 15, 18]. The most affected part of the segment was chosen for scoring.

The following MRI features were used to calculate the MR index of activity: mural thickness in millimeters, relative contrast enhancement, and the presence of edema and ulcers. These features have been proven to be significantly correlated to the CDEIS. The MR index of activity was calculated using the following formula: (1.5 × wall thickness in millimeters) + (0.02 × relative contrast enhancement) + (5 × edema) + (10 × ulceration) [13].

For the overall CDMI score, the following four features—mural thickness, mural T2 signal, perimural T2 signal, and mural T1 enhancement—were scored on a scale of 0 to 3, resulting in a maximum score of 12 [15]. These four features were selected because they were found to be significantly correlated with disease activity according to an endoscopic biopsy acute inflammatory score. In addition, the sum of the scores for mural thickness, mural T2 signal, perimural T2 signal, and contrast enhancement showed the highest accuracy [15]. Furthermore, the following features were assessed: abscess, comb sign, enlarged (> 1 cm) lymph nodes, fistulas, lymph node enhancement, pattern of mural enhancement, pseudopolyps, and total length of the disease in each segment.

The MRI features with regard to lymph nodes were scored per patient; all other features were assessed per segment. The readers used the same method as described in detail in the articles about the MR index of activity and the CDMI score [13, 15]. For calculating the relative contrast enhancement involved, we used the formula as described in Rimola et al. [13] [(WSI contrast-enhanced − WSI unenhanced) / WSI unenhanced] × 100 × (SD noise unenhanced / SD noise contrast-enhanced). Here, WSI is the wall signal intensity, SD noise unenhanced corresponds to the average of three SD of the signal intensity measured outside of the body before gadolinium-based contrast agent injection, and SD noise contrast-enhanced corresponds to the SD

of the same noise after gadolinium-based contrast agent administration [19]. Ulcerations (defined as deep depressions in the mucosal surface of a thickened segment) and the short-axis diameters of enlarged lymph nodes were assessed on contrast-enhanced 3D T1-weighted spoiled gradient-echo images with fat saturation.

Eight of the MRI features are common to the MR index of activity and CDMI score but were assessed using different definitions of abnormality according to the particular scoring system. Specifically, mural thickness was measured using an ordinal score (0–3) and as a continuous variable in millimeters using calipers on either single-shot FSE or spoiled gradient-echo sequences. T1 contrast enhancement was measured using an ordinal score (0–3) and using relative contrast enhancement [19]. Lymph nodes were assessed using an ordinal (0–3) score and a binominal score (yes/no). T2 signal was measured using an ordinal score (0–3) and a binominal score (edema; yes/no).

Reference Standard

Colonoscopy was performed after standard bowel preparation by either a gastroenterologist or a senior resident in gastroenterology under direct supervision of a gastroenterologist, using a standard colonoscope. The performing endoscopist was aware of the patient's history but was blinded to the MR enterography results. Segments were excluded from the analysis if they could not be scored during ileocolonoscopy. One of two gastroenterologists experienced in endoscopy of inflammatory bowel disease assessed the CDEIS [1]. A segmental CDEIS was calculated using the variables of deep ulceration (no = 0, yes = 12), superficial ulceration (no = 0, yes = 6), surface involved by disease (0–10), and ulcerated surface (0–10) for each of five bowel segments (terminal ileum, right colon, transverse colon, left colon, and rectum). The reference standard has been described in detail elsewhere [16]. In six patients, the terminal ileum could not be assessed during ileocolonoscopy because of a stenosis; there-

TABLE 3: MRI Features, by Category and Score

MRI Features	Score			
	0	1	2	3
Crohn disease MRI index features				
Mural thickness (mm)	1–3	> 3–5	> 5–7	> 7
Mural T2 signal	Normal bowel wall	Minor increase in signal intensity: bowel wall appears dark gray on fat-saturated images	Moderate increase in signal intensity: bowel wall appears light gray on fat-saturated images	Marked increase in signal intensity: bowel wall contains areas of white high signal approaching that of luminal content
Perimural T2 signal	Equivalent to normal mesentery	Increase in mesenteric signal but no fluid	Small fluid rim (≤ 2 mm)	Larger fluid rim (> 2 mm)
T1 enhancement	Equivalent to normal bowel wall	Minor enhancement: bowel wall signal intensity greater than normal small bowel but significantly less than nearby vascular structures	Moderate enhancement: bowel wall signal intensity increased but somewhat less than nearby vascular structures	Marked enhancement: bowel wall signal intensity approaches that of nearby vascular structures
MR index of activity features				
Wall thickness in millimeters				
Relative contrast enhancement				
Edema	Absent	Present		
Ulcers	Absent	Present		
Other features				
Comb sign	Absent	Present		
Abscess	Absent	Present		
Fistulas	Absent	Present		
Pseudopolyps	Absent	Present		
Lymph nodes, > 1 cm[a]	Absent	Present		
Lymph nodes, size and number[a]	Absent	Cluster < 1 cm	1 node > 1 cm	3 nodes > 1 cm
Lymph node enhancement[a]	Less than nearby vascular structure	Equivalent or greater than nearby vascular structure		
Mural enhancement pattern[b]	Not applicable	Homogeneous	Mucosal	Layered
Total length of disease in segment (cm)	0	0–5	5–15	> 15

Note—The rows for wall thickness and relative contrast enhancement are empty because these are quantitative values.

[a] Per patient.

[b] Enhancement pattern was classified as homogeneous (with all bowel wall enhancing equally), submucosal only (with only the innermost wall layer enhancing), or layered (with both inner wall and serosal bowel wall layers enhancing, with a central band of relatively reduced enhancement of the muscular layer).

fore, 137 segments were correlated to the segmental CDEIS scores.

The median time between colonoscopy and MR enterography was 7 days (interquartile range, 5–14 days). MRI and colonoscopy were not performed on the same day. The median CDEIS was 4.3 (interquartile range, 1.6–5.8).

Data Analyses

Several multirater analyses were performed for all features individually and for the overall MR index of activity and CDMI score to assess the interobserver agreement. For all ordinal data, a weighted kappa coefficient was calculated per two raters and eventually was pooled. For the binomial data, a multirater kappa coefficient was used, which was also calculated per two raters and pooled. For continuous data, a multirater intraclass correlation coefficient was determined.

In addition, the scores of both of the most experienced abdominal radiologists were analyzed post hoc. This was done to evaluate whether experience would positively influence the reproducibility values. Both the kappa and intraclass cor-

relation coefficient values interpretation was as follows: 0–0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, excellent [20].

For an overall correlation, we first calculated means of MR index of activity scores and CDMI scores per segment for all four observers and correlated these values with the CDEIS scores. Because the segmental scores were interpreted as continuous variables, the Spearman correlation was used to correlate the segmental CDEIS scores to the segmental MR index of activity and segmental CDMI scores. Correlation coefficient values were interpreted as follows: 0.0, not correlated; 0.2, weakly correlated; 0.5, moderately correlated; 0.8, strongly correlated; and 1.0, perfectly correlated. Statistical analysis was performed in Excel 2003 (Microsoft) using PASW statistics software (version 19, SPSS).

Results

MRI Features

All MRI features showed a fair to good interobserver variability (Table 4). Wall thickness measured in millimeters reported high-

est agreement of 0.69 (95% CI, 0.62–0.75). In addition to wall thickness measured in millimeters, the presence of edema, the pattern of enhancement, and the length of the disease in each segment showed a good interobserver variability between all readers (Figs. 1–3).

The results for measuring wall thickness in millimeters for experienced radiologists were significantly better than the multirater data and showed an excellent interobserver variability of 0.87 (95% CI, 0.82–0.90). The incidence of abscesses, fistulas, and pseudopolyps was very low in our study. Therefore, only the overall agreement among all readers (0.96, 0.94, and 0.99) is reported. Good interobserver variability was reported for abscesses and fistulas between the most experienced readers (0.66 and 0.66). Furthermore, the comb sign and lymph node assessment improved from fair to moderate agreement (0.51 and 0.58, respectively) for the experienced readers, suggesting that prior experience with MR enterography is an advantage in the assessment of these features.

TABLE 4: Multirater Kappa and Intraclass Correlation Coefficient Values

MRI Features	All Readers	Experienced Radiologists
Crohn disease MRI index features		
Mural thickness	0.59 (0.47–0.70)	0.74 (0.63–0.85)
Mural T2 signal	0.55 (0.44–0.66)	0.39 (0.27–0.52)
Perimural T2 signal	0.30 (0.15–0.45)	0.37 (0.20–0.55)
T1 enhancement	0.57 (0.46–0.68)	0.60 (0.50–0.70)
MR index of activity features		
Wall thickness in millimeters	0.69 (0.62–0.75)	0.87 (0.82–0.90)
Relative contrast enhancement	0.42 (0.33–0.51)	0.55 (0.42–0.65)
Edema	0.66 (0.59–0.72)	0.57 (0.41–0.74)
Ulceration	No ulceration was seen by all readers in 137 segments; overall agreement: 137/143 = 0.96	No ulceration was seen by the two readers in 138 segments; overall agreement: 138/143 = 0.97
Other features		
Comb sign[a]	0.39 (0.30–0.49)	0.51 (0.34–0.67)
Abscess	No abscess was seen by all readers in 137 segments; overall agreement: 137/143 = 0.96	0.66 (0.05–1.28)
Fistula	No fistula was seen by all readers in 135 segments; overall agreement: 135/143 = 0.94	0.66 (0.22–1.10)
Pseudopolyps	No pseudopolyp was seen by all readers in 141 segments; overall agreement: 142/143 = 0.99	No pseudopolyp was seen by all readers in 142 segments; overall agreement: 142/143 = 0.99
Lymph nodes (> 1 cm)[b]	0.35 (0.21–0.50)	0.58 (0.23–0.94)
Lymph nodes (size/number)[b]	0.36 (0.13–0.59)	0.55 (0.31–0.78)
Lymph node enhancement[b]	0.47 (0.32–0.61)	0.22 (−0.13 to 0.58)
Enhancement pattern	0.62 (0.49–0.75)	0.71 (0.59–0.83)
Length of disease in each segment	0.62 (0.51–0.73)	0.64 (0.53–0.74)

Note—Except for wall thickness in millimeters and relative contrast enhancement, data are kappa value (95% CI). For wall thickness in millimeters and relative contrast enhancement, data are intraclass correlation coefficient (95% CI).
[a]One reader did not score comb sign in all segments and was excluded from the analysis.
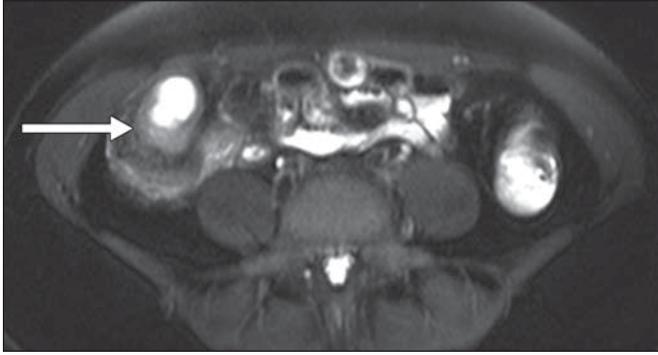[b]Lymph nodes were scored per patient.

Fig. 1—29-year-old woman with Crohn disease. Axial T2-weighted single-shot fast spin-echo image with fat saturation is shown. All readers reported presence of edema (*arrow*), moderate (grade 2) increase in signal intensity in bowel wall, and wall thickness of more than 7 mm.

MRI Scoring Systems

Reproducibility—The MR index of activity and CDMI scores showed good interobserver variability (0.74 and 0.78, respectively) in the assessment of Crohn disease activity for all readers (Table 5). There was a minimal increase in the reproducibility (0.80 and 0.81, respectively) for the MR index of activity and the CDMI scores between the two experienced readers only. For the segmental scores of the four observers, median MR index of activity was 4.90 (range, 1.22–35.62), and the median CDMI score was 0 (range, 0–12) (Fig. 4).

Correlation to the CDEIS—Both MR index of activity and CDMI scores correlated moderately with segmental CDEIS ($r = 0.51$ [95% CI, 0.38–0.63] and $r = 0.59$ [95% CI, 0.47–0.69], respectively).

Discussion

This study shows variable reproducibility of many individual MRI features advocated in the assessment of Crohn disease activity. Four features (wall thickness in millimeters, the presence of edema [yes/no], enhancement pattern [0–3], and length of the disease in each segment [0–3]) had good reproducibility, whereas extramural MRI features such as perimural T2 signal, comb sign, and lymph nodes showed only fair reproducibility. When individual features were combined into two scoring systems proposed in the literature (MR index of activity and CDMI), interobserver variability was good across four readers.

Our study has several strengths: four readers from two international expert centers assessed a large number of bowel segments using different features and two scoring systems. In addition, four specific MRI features were measured in two different ways within the two scoring systems [13, 15], and we applied both definitions to determine which method is most reproducible. The MRI scoring systems were correlated to the CDEIS per segment, an objective activity index in comparison with clinical and biochemical parameters [21]. Overall, we showed that the recently developed MRI scoring systems showed good-to-excellent reproducibility and moderate correlation to the CDEIS.

The reproducibility of several, mainly extramural, MRI features showed only fair reproducibility, and some authors have reported a higher interobserver variability [9, 13, 22] than in our study. Conversely, the variable interobserver agreement in our study is more in concordance with other data [12, 23–25]. An explanation of this distinction might be in the severity of the disease of the included patients. Severe disease is easier to diagnose than mild disease, because in the former, the MRI features are most pronounced [4]. Importantly, mural thickness, T1 contrast enhancement, and T2 wall signal indicating edema are considered important MRI features of activity [18]. These features are common to both the MR index of activity and the CDMI score, and, reassuringly, all showed moderate-to-good interobserver variability.

Although the aforementioned important MRI features may be considered as the basic elements of both systems, they are defined in different ways. In the CDMI score, only qualitative variables are used, whereas in the MR index of activity, predominantly quantitative data are extracted. Using quantitative data as in the MR index of activity score might lead to a more precise grading of disease activity, although it is more time consuming, which potentially limits the use of this score in the clinical setting. An example is the measurement of the relative contrast enhancement, where region of interest measurements are used. Region of interest–based measurements have a known poor interobserver variability [24]. In accordance, our study showed lower reproducibility of relative contrast enhancement (0.42), in comparison with grading T1 enhancement from 0 to 3 (0.57).

In addition to contrast enhancement, three other MRI features are defined in two different ways in the literature. The measurement of edema is essential in the management of Crohn disease to differentiate intestinal inflammation from fibrosis. Our study reported a higher reproducibility when edema was measured binomially (yes/no) rather than ordinally (0–3). Mural thickness measured in millimeters is not only more objective than using a qualitative variable, it also has higher reproducibility. The interobserver variability of lymph node measurement described in the development of the MR index of activity [13] and the CDMI score [15] showed similar fair interobserver agreement. These findings clarify how features can be most reproducibly measured and might result in a more consistent use in the future.

It is generally assumed that any new radiologic technique such as MR enterography has an associated learning curve for accurate interpretation. We therefore investigated whether experience might have influenced the reproducibility values. We had two experienced (700 or more MR enterography studies) and two less-experienced (170 or more MR enterography studies) readers. The assessment of just five of the tested MRI features showed improved reproducibility values when measured by experienced readers. Interobserver variability of enlarged lymph nodes (> 1 cm), lymph nodes (size and num-
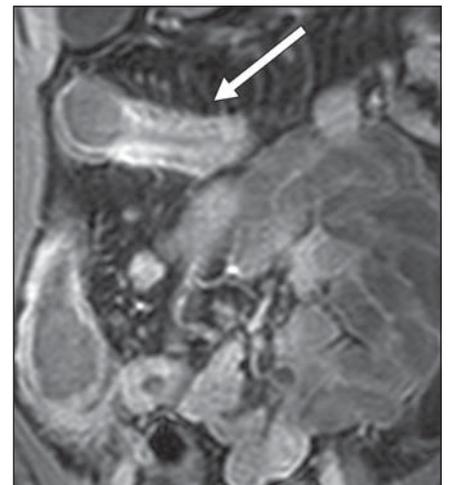

Fig. 2—19-year-old woman with Crohn disease. Coronal contrast-enhanced 3D T1-weighted spoiled gradient-echo image with fat saturation is shown. All readers reported marked (grade 3) and layered pattern of T1 enhancement (*arrow*) and wall thickness of more than 7 mm in transverse colon.
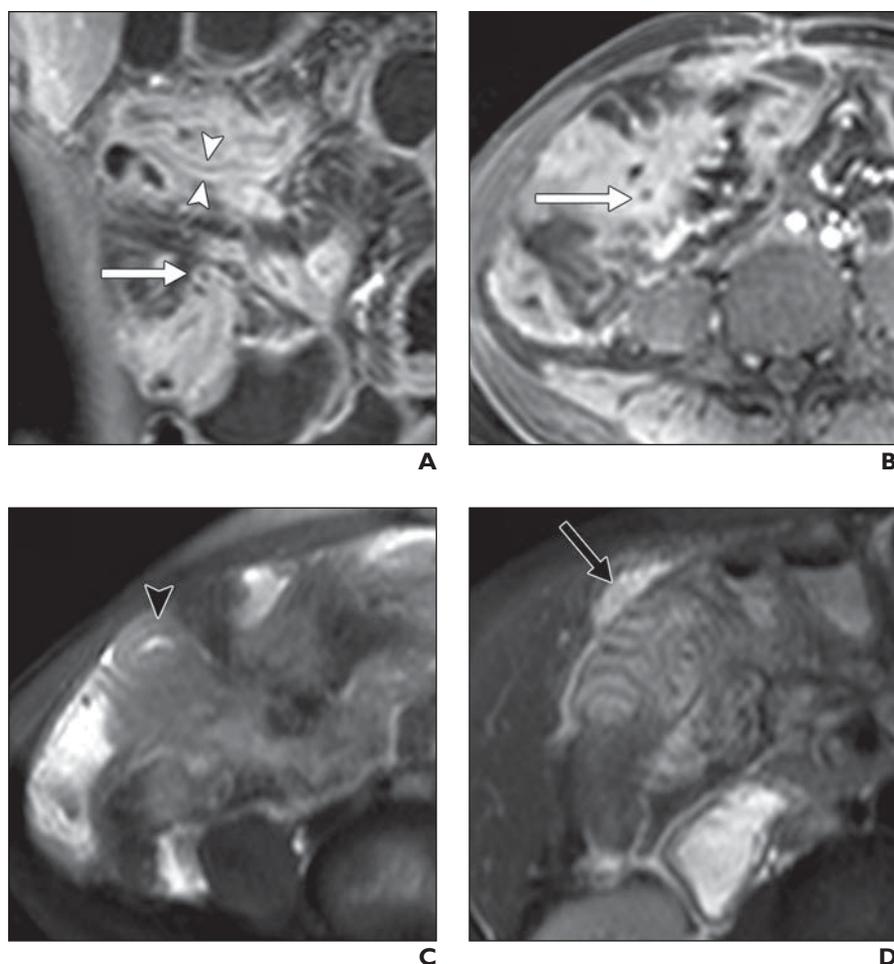
Fig. 3—35-year-old woman who previously underwent ileocecal resection. Colonoscopy showed mild terminal ileitis and mild colitis with superficial ulcerations.

A and **B,** On axial (**A**) and coronal (**B**) contrast-enhanced 3D spoiled gradient-echo images with fat saturation, three MR enterography readers noted presence of fistula (*arrow*, **A** and **B**) in neoterminal ileum and moderate (grade 2) mural T1 enhancement (*arrowheads*, **A**) with layered pattern of transverse colon.

C and **D,** Mild increase (grade 1) in T2 mural signal (*arrowhead*, **C**) and large perimural fluid rim (grade 3; *arrow*, **D**) were seen on axial T2-weighted single-shot fast spin-echo images with fat saturation.

TABLE 5: Interobserver Variability of Two MRI Scoring Systems

MRI Scoring System	All Readers	Experienced Radiologists
Crohn disease MRI index	0.78 (0.73–0.83)	0.81 (0.74–0.86)
MR index of activity	0.74 (0.68–0.79)	0.80 (0.73–0.85)

Note—Data are multirater intraclass correlation coefficients (95% CI).

ber; 0–3), comb sign, mural thickness (0–3), and mural thickness measured in millimeters increased from fair to moderate, moderate to good, or good to excellent, respectively. This could be because the less-experienced readers were not used to assessing these features.

One could argue that some MRI features may have shown a higher reproducibility when scored by experienced observers only. However, our data showed only a small increase in kappa or intraclass correlation coefficient values for only a few MRI features between experienced observers only. This is in accordance with findings of a previous study in which reproducibility of bowel-wall gadolinium enhancement measurements was determined [24].

To our knowledge, our study is the first to compare the reproducibility of multiple MRI features and two scoring systems and to describe the interobserver variability of similar MRI features measured in different ways. Although certain individual features (e.g.,

perimural T2 signal, ulcerations, and relative contrast enhancement) showed only fair interobserver variability, importantly, when combined together in both the CDMI score and the MR index of activity, the results showed good reproducibility.

The correlation to the CDEIS in our study, which is lower than that reported by Rimola et al. [13] for the MR index of activity in their study, might be explained by the different study protocols. We used a less-extensive method of contrast agent administration than was used to develop the MR index of activity, where warm water was retrogradely instilled into the colon. In addition, our study cohort primarily comprised patients with mild disease activity, whereas the MR index of activity was developed in a cohort including patients with more-severe disease activity. This may explain a lower correlation to the CDEIS and a low detection rate of ulcerations in our series than in the original article about the MR index of activity [13]. On the other hand, the correlation to endoscopic activity is in concordance with previous research [26, 27]. Furthermore, our protocol contained late phase IV contrast-enhanced series, which may have affected the evaluation of the contrast-enhanced series.

A number of limitations have to be acknowledged. The CDEIS is not a perfect reference standard because it assesses the mucosa only and gives little information on the transmural and extramural disease extent. However, endoscopy remains the reference standard for Crohn disease activity. We chose MR enterography as the contrast agent administration technique, because it is the most commonly used technique for bowel distention for patients with Crohn disease and is better accepted than MR enteroclysis [28]. Neither MR enterography nor MR enteroclysis is aimed at optimal colonic distention, although colonic distention will be obtained to a variable extent. In our study, sufficient colonic distention and visibility were achieved in all but five patients in which the rectum was inadequately visible. The MR index of activity was developed in a cohort using both MR enterography and rectal fluid administration. This difference in bowel preparation may, at least in part, explain the different correlation between the MR index of activity and CDEIS in this study as compared with the studies by the Barcelona group that introduced this score [13, 29]. Recent articles have reported that motility can be changed in affected small-bowel locations in Crohn disease [30, 31]. However, our protocol did not
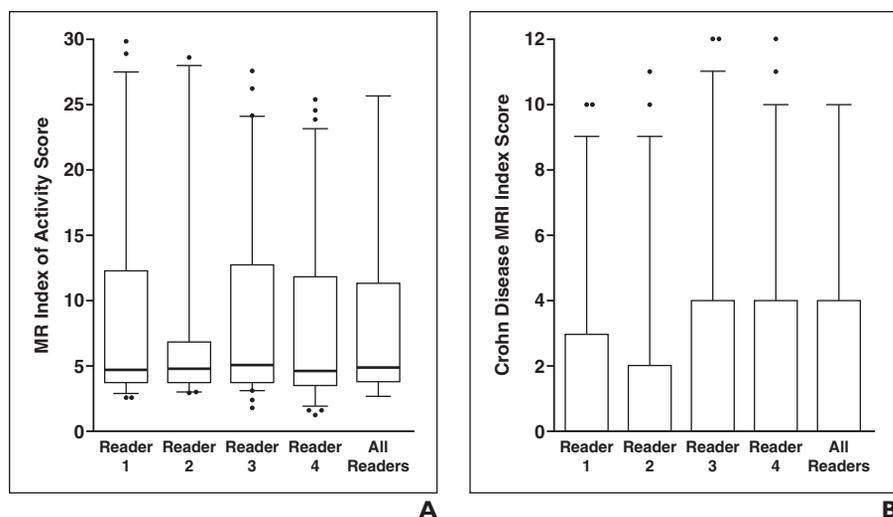
Fig. 4—Box-and-whisker plots of index activity scores.
A and **B**, Graphs show 2.5, 25, 50, 75, and 97.5 percentiles of 143 segmental MR index of activity scores (**A**) and 143 segmental Crohn disease MRI index scores (**B**) of all readers separately and combined. Median segmental score for all readers was 4.90 and 0, respectively.

contain cine MR motility series and, therefore, we could not study the scoring system developed by Girometti et al. [32].

Along with ulcerations, abscesses, fistulas, and pseudopolyps were rarely seen in our data. This is in line with the daily clinical experience in our tertiary referral centers and reflects the patient spectrum in our institutions. To accurately determine the interobserver variability of these features, analysis of a group of patients with larger disease severity might elucidate the reproducibility of these features.

We did not perform an intraobserver analysis, because the intraobserver variability is generally higher than the interobserver agreement, which is intuitive because one would expect an observer to agree more with himself or herself than with another reader. Another methodologic limitation might be that we only used MRI examinations obtained at 3 T, but we do not expect substantial differences in evaluation of the features, MR index of activity score, and CDMI score compared with 1.5 T. Indeed, one study has reported that 3 T is equally accurate as 1.5 T in the assessment of Crohn disease [33].

In summary, some commonly used MRI features have good reproducibility among four readers. Two recently developed scoring systems, the CDMI and MR index of activity scores, have good reproducibility and have moderate agreement with CDEIS. Additional research in a larger cohort of patients, including all disease stages and with more than

one reference standard, has to be performed before a global accurate MRI scoring system can be implemented in clinical trials and daily clinical practice.

References

1. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Thérapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). *Gut* 1989; 30:983–989
2. Borley NR, Mortensen NJ, Jewell DP, Warren BF. The relationship between inflammatory and serosal connective tissue changes in ileal Crohn's disease: evidence for a possible causative link. *J Pathol* 2000; 190:196–202
3. Van Assche G, Vanbeckevoort D, Bielen D, et al. Magnetic resonance imaging of the effects of infliximab on perianal fistulizing Crohn's disease. *Am J Gastroenterol* 2003; 98:332–339
4. Horsthuis K, Bipat S, Stokkers PCF, Stoker J. Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review. *Eur Radiol* 2009; 19:1450–1460
5. Pariente B, Peyrin-Biroulet L, Cohen L, Zagdanski A-M, Colombel J-F. Gastroenterology review and perspective: the role of cross-sectional imaging in evaluating bowel damage in Crohn disease. *AJR* 2011; 197:42–49
6. Sinha R, Verma R, Verma S, Rajesh A. MR enterography of Crohn disease. Part 1. Rationale, technique, and pitfalls. *AJR* 2011; 197:76–79
7. Sinha R, Verma R, Verma S, Rajesh A. MR enterography of Crohn disease. Part 2. Imaging and

pathologic findings. *AJR* 2011; 197:80–85
8. Punwani S, Rodriguez-Justo M, Bainbridge A, et al. Mural inflammation in Crohn disease: location-matched histologic validation of MR imaging features. *Radiology* 2009; 252:712–720
9. Maccioni F, Bruni A, Viscido A, et al. MR imaging in patients with Crohn disease: value of T2- versus T1-weighted gadolinium-enhanced MR sequences with use of an oral superparamagnetic contrast agent. *Radiology* 2006; 238:517–530
10. Gourtsoyiannis N, Papanikolaou N, Grammatikakis J, Papamastorakis G, Prassopoulos P, Roussomoustakaki M. Assessment of Crohn's disease activity in the small bowel with MR and conventional enteroclysis: preliminary results. *Eur Radiol* 2004; 14:1017–1024
11. Zappa M, Stefanescu C, Cazals-Hatem D, et al. Which magnetic resonance imaging findings accurately evaluate inflammation in small bowel Crohn's disease? A retrospective comparison with surgical pathologic analysis. *Inflamm Bowel Dis* 2011; 17:984–993
12. Ziech MLW, Bipat S, Roelofs JJTH, et al. Retrospective comparison of magnetic resonance imaging features and histopathology in Crohn's disease patients. *Eur J Radiol* 2011; 80:e299–e305
13. Rimola J, Ordás I, Rodriguez S, et al. Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. *Inflamm Bowel Dis* 2011; 17:1759–1768
14. Masselli G, Gualdi G. MR imaging of the small bowel. *Radiology* 2012; 264:333–348
15. Steward MJ, Punwani S, Proctor I, et al. Non-perforating small bowel Crohn's disease assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. *Eur J Radiol* 2012; 81:2080–2088
16. Ziech MLW, Lavini C, Caan MWA, et al. Dynamic contrast-enhanced MRI in patients with luminal Crohn's disease. *Eur J Radiol* 2012; 81:3019–3027
17. Best WR, Becktel JM, Singleton JW, Kern F. Development of a Crohn's disease activity index: National Cooperative Crohn's Disease Study. *Gastroenterology* 1976; 70:439–444
18. Ziech ML, Bossuyt PM, Laghi A, Lauenstein TC, Taylor SA, Stoker J. Grading luminal Crohn's disease: which MRI features are considered as important? *Eur J Radiol* 2012; 81:e467–e472
19. Semelka RC, Shoenut JP, Silverman R, Kroeker MA, Yaffe CS, Micflikier AB. Bowel disease: prospective comparison of CT and 1.5-T pre- and postcontrast MR imaging with T1-weighted fat-suppressed and breath-hold FLASH sequences. *J Magn Reson Imaging* 1991; 1:625–632
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174

21. Van Assche G, Dignass A, Panes J, et al. The second European evidence-based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *J Crohns Colitis* 2010; 4:7–27

22. Negaard A, Sandvik L, Mulahasanovic A, Berstad AE, Klöw N-E. Magnetic resonance enteroclysis in the diagnosis of small-intestinal Crohn's disease: diagnostic accuracy and inter- and intra-observer agreement. *Acta Radiol* 2006; 47:1008–1016

23. Jensen MD, Ormstrup T, Vagn-Hansen C, Østergaard L, Rafaelsen SR. Interobserver and intermodality agreement for detection of small bowel Crohn's disease with MR enterography and CT enterography. *Inflamm Bowel Dis* 2011; 17:1081–1088

24. Sharman A, Zealley I, Greenhalgh R, Bassett P, Taylor S. MRI of small bowel Crohn's disease: determining the reproducibility of bowel wall gadolinium enhancement measurements. *Eur Radiol* 2009; 19:1960–1967

25. Siddiki HA, Fidler JL, Fletcher JG, et al. Prospec- tive comparison of state-of-the-art MR enterography and CT enterography in small-bowel Crohn's disease. *AJR* 2009; 193:113–121

26. Florie J, Horsthuis K, Hommes DW, et al. Magnetic resonance imaging compared with ileocolonoscopy in evaluating disease severity in Crohn's disease. *Clin Gastroenterol Hepatol* 2005; 3:1221–1228

27. Oussalah A, Laurent V, Bruot O, et al. Diffusion-weighted magnetic resonance without bowel preparation for detecting colonic inflammation in inflammatory bowel disease. *Gut* 2010; 59:1056–1065

28. Negaard A, Paulsen V, Sandvik L, et al. A prospective randomized comparison between two MRI studies of the small bowel in Crohn's disease, the oral contrast method and MR enteroclysis. *Eur Radiol* 2007; 17:2294–2301

29. Rimola J, Rodriguez S, García-Bosch O, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. *Gut* 2009; 58:1113–1120

30. Froehlich JM, Waldherr C, Stoupis C, Erturk SM, Patak MA. MR motility imaging in Crohn's disease improves lesion detection compared with standard MR imaging. *Eur Radiol* 2010; 20:1945–1951

31. Menys A, Atkinson D, Odille F, et al. Quantified terminal ileal motility during MR enterography as a potential biomarker of Crohn's disease activity: a preliminary study. *Eur Radiol* 2012; 22:2494–2501

32. Girometti R, Zuiani C, Toso F, et al. MRI scoring system including dynamic motility evaluation in assessing the activity of Crohn's disease of the terminal ileum. *Acad Radiol* 2008; 15:153–164

33. Fiorino G, Bonifacio C, Padrenostro M, et al. Comparison between 1.5 and 3.0 Tesla magnetic resonance enterography for the assessment of disease activity and complications in ileo-colonic Crohn's disease. Dig Dis Sci 2013 Aug 1 [Epub ahead of print]

FOR YOUR INFORMATION

The comprehensive book based on the ARRS 2013 annual meeting categorical course on *Body MRI* is now available! For more information or to purchase a copy, see www.arrs.org.