Title

Magnetic resonance enterography features of small bowel Crohn's disease activity: an inter-rater reliability study of small bowel active inflammation in clinical practice setting

Journal

British Journal of Radiology, 92(1099)

ISSN

0007-1285

Authors

Tsai, Richard
Mintz, Aaron
Lin, Michael
et al.

Publication Date

2019-07-01

DOI

10.1259/bjr.20180930

Peer reviewed

BJR

https://doi.org/10.1259/bjr.20180930

Received: 29 October 2018 | Revised: 13 March 2019 | Accepted: 24 April 2019

Cite this article as:
Tsai R, Mintz A, Lin M, Mhlanga J, Chiplunker A, Salter A, et al. Magnetic resonance enterography features of small bowel Crohn's disease activity: an inter-rater reliability study of small bowel active inflammation in clinical practice setting. *Br J Radiol* 2019; **92**: 20180930.

FULL PAPER

Magnetic resonance enterography features of small bowel Crohn's disease activity: an inter-rater reliability study of small bowel active inflammation in clinical practice setting

[1]RICHARD TSAI, MD, [1]AARON MINTZ, [1]MICHAEL LIN, [1]JOYCE MHLANGA, [2]ADEETI CHIPLUNKER, [3]AMBER SALTER, [4]MATTHEW CIORBA, [4]PARAKKAL DEEPAK and [5]KATHRYN FOWLER

[1]Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, MO, United States
[2]Division of Gastroenterology. Washington University in St. Louis., St. Louis, MO, United States
[3]Division of Biostatistics, Washington University in St. Louis, St. Louis, MO, United States
[4]Washington University Inflammatory Bowel Diseases Center, St. Louis, MO, United States
[5]Department of Radiology, University of California, San Diego, San Diego, CA, United States

Address correspondence to: Dr Richard Tsai
E-mail: *r.tsai@wustl.edu*

Objective: The aim of this study is to determine the interrater agreement in a clinical practice environment for the most commonly used magnetic resonance enterography (MRE) features of Crohn's disease (CD).
Methods: CD patients with MRE's before and after treatment were retrospectively identified using search queries over a 7-year period (May 2017–September 2017). MRE features of CD comprising components of multiple CD scoring indices were scored by radiologists in the same segment of bowel. Agreement for nominal categorical and continuous variables was assessed using a κ and interclass correlation coefficients, respectively.
Results: 80 scans comprised the study population. Moderate interrater agreement was seen in both the pre- and post-treatment MRE's for presence of diffusion restriction (κ = 0.43, 0.48; pre- and post-treatment), stricturing disease (κ = 0.51, 0.52), overall degree of severity (κ = 0.49, 0.59). Substantial agreement was seen in pre- and post-treatment scans for length of involvement (interclass correlation coefficient = 0.67, 0.61). The presence of mucosal ulceration had no agreement (κ = −0.07, −0.042).
Conclusion: Many MRE features of active CD comprising the major CD scoring indices are reproducible when interpreted by non-CD focused abdominal radiologists. However, the presence of mucosal ulcerations had no agreement and may need more investigation before including this feature as a driver in therapeutic decision making.
Advances in knowledge: Demonstrates the unreliability of mucosal ulceration by non-CD focused abdominal radiologists, targeting a potential area for future education.
Key Points
The majority of MRE findings incorporated in to many CD scoring indices have fair to moderate inter-rater agreement even when read by non-MRE expert radiologists. Substantial agreement was seen in the length of involved bowel, but this feature is only incorporated in to one of the CD scoring indices. Presence of mucosal ulcerations had no interrater agreement in our study—a feature which is heavily weighted by several CD scoring indices. Research should be focused bridging those features which have poor interrater agreement.

INTRODUCTION

Crohn's disease (CD) is an inflammatory disease of the bowel that can affect any segment of the gastrointestinal tract from the mouth through the anus. The prevalence of CD is increasing worldwide,[1] with prevalence ranging from 319 persons per 100,000 in North America, to 322 persons per 100,000 in Europe. Advances in treatment of CD have focused on early immunosuppression or combination biological therapy to control inflammation and prevent complications and surgical intervention.[2]

Early medical intervention hinges on the ability of diagnostic tools to provide an accurate assessment of presence and the degree of bowel inflammation and presence of complications. In addition to identifying active disease, imaging can also stratify the patient into a disease

phenotype either by degree of transmural aggressiveness (structuring, penetrating, inflammatory) or location of bowel involved. While endoscopy is the current gold-standard for the evaluation of CD, endoscopy is limited by its invasive nature, inability to evaluate extraluminal pathology, traverse tight strictures, define anatomy of fistulae, or evaluate the mid small bowel.[3] Radiologic examinations such as CT enterography (CTE) and MR enterography (MRE) can image the bowel and extraluminal pathology not readily visualized at the time of endoscopy.[4,5] CTE and MRE have similar performance in the evaluation of CD, but due to the lack of ionizing radiation, MRE has an advantage in long-term, longitudinal evaluation of CD patients.[6–13] The European Society of Gastrointestinal and Abdominal Radiology recommends MRE as a first-line tool for the initial diagnosis and follow-up of patients with CD.[14,15]

Given the many advantages of imaging, there has been emerging clinical focus on developing and validating imaging treatment targets either to supplement or in some instances replace traditional endoscopic and symptom related scoring systems, both for initial disease activity assessment and response to therapy.[16,17] There are several potential multiparametric MRE scoring systems developed to suit this need, including: MR index of activity (MaRIA), MRE global score (MEGS), Clermont score, and London scoring systems.[8,9,15,18–22] These MRE systems have largely been validated within the context of controlled clinical trials by highly subspecialized radiologists that are specialists in inflammatory bowel disease. In the hands of select experts, these systems perform very well. The area under the receiver operating characteristic was 0.930 for the MaRIA study, 0.840 for the Clermont study, and 0.853 in the London study.[19,20,23]

The best imaging target/scoring system will require not only accuracy but a high level of reliability. To our knowledge, there have been relatively few studies documenting the reproducibility or inter reader reliability of the scoring systems or individual MRE features outside of the highly specialized centers.[24–28] This is a necessary step to allow for the adoption of noninvasive treatment targets, such as MRE, in clinical practice.

As such, our study aims to determine the inter-rater agreement and hence the reliability of the different MRE signs and features, before and after modification or initiation of medical therapy, that comprise the main CD scoring systems in a clinical practice environment to identify aspects of the lexicon that may require further revision/clarification prior to being incorporated in standardized reporting and scoring systems in clinical practice.

METHODS AND MATERIALS

Retrospective patient enrollment

After investigational review board approval and waiver of consent, a HIPAA-compliant, retrospective query of the radiology information system search tool was performed to discover all MR examinations performed for the evaluation of CD. MR examinations were discovered with the radiology information system tool using the search terms "MR enterography", "Crohn's disease" and cross-referenced with a database of all inflammatory bowel disease patients seen at a tertiary CD clinic. Each subsequent examination was reviewed to document that the examination was performed with the MRE protocol (detailed below). Inclusion criteria comprised those patients with confirmed CD by pathology or clinical history, an initial MRE before the initiation or alteration of a biologic agent, and subsequent MRE at least 3–6 months after initiation or alteration of a biologic agent. This population was selected to ensure an adequate representation of at least moderate disease activity and a relatively controlled cohort of patients reflecting that which might comprise a clinical study population. Patients were excluded if the MRE was non-diagnostic, no gadolinium contrast was administered, no diagnosis of CD could be confirmed, or if there was evidence of active penetrating disease, perianal disease, or isolated colonic involvement by MRI or by the clinical chart. The first 50 sequential patients that fulfilled these criteria were included in our study from May 2015 to September 2017.

Imaging protocol and interpretation

All imaging was performed at our institution in the supine position using a 1.5 or 3 Tesla MRI (Siemens) with a phased array surface coil. The patients fasted for 4 h prior to the examination. 2 bottles of 100 ml barium sulfate suspension 0.1% w/v (VoLumen, Bracco) were given to the patient by mouth approximately 45 min prior to the examination. A third bottle of 100 ml of the same suspension was given just prior to the MRE examination. Patients received 1 mg of glucagon administered intravenously while positioned on the table immediately prior to the start of imaging. Imaging was performed from the porta hepatis to the level of the iliac crests to include all small bowel. The parameters for the MRE sequences are provided in Table 1.

MRE examinations were independently interpreted by two radiologists with 4 and 10 years' experience in abdominal imaging (400 MRE's by one reader and 125 by the other). The interpreting radiologists were fellowship trained in abdominal imaging at two separate institutions, but are not focused experts in CD (>700 MRE's).[24] Current reporting at Mallinckrodt Institute of Radiology is not structured and no scoring systems are in place in practice. Hence, prior to interpretation, a training session was held to clarify grading of multiple parameters and the readers were provided with the reference articles for the different features.[19,24] The MRE examinations were known to the readers as patients with active CD that underwent treatment with a biologic agent. Otherwise, the reviewers were blinded to patient information and demographic data, laboratory data, colonoscopy and endoscopy results, and patient outcome.

All interpretations were performed on the Vitrea Vital picture arching and computing system (Vital Images). Radiologists chose the single worst segment of small bowel to grade. The readers were asked to choose the worst segment of inflamed small bowel based on general assessment rather than a defined criteria. MRE features (detailed in Table 2) assessed by the radiologists included portion of small bowel affected, length of bowel involved, degree of bowel wall thickening, presence of mucosal ulcerations, type of abnormal bowel wall enhancement, severity of abnormal bowel wall enhancement, degree of bowel wall edema, presence of small bowel diffusion restriction, presence of

Table 1. Magnetic Resonance enterography protocol.

Magnetic resonance enterography protocol
Coil: phased array
Coverage: liver hilum to urinary bladder
Contrast rate: 2 ml/s gadoterate meglumine (Dotarem)
Sequences (parameters provided for 1.5 T only)
• Transaxial diffusion-weighted echo planar imaging with FS (SPAIR); TR 6500 ms; TE 65 ms; ST 7 mm; matrix 156 × 192; *b*-values 50, 400, 800 s/mm^2
• Coronal/transaxial single-short turbo spin echo T_2 weighted imaging: TR: 1000–1300 ms; TE 80–90 ms; ST 6 mm; matrix 640 × 640 (coronal), 520 × 640 (axial)
• 2D transaxial GRE with FS: TR: 150–200 ms; TE 2–3 ms; ST 7 mm; matrix 208 × 256
• Pre-contrast transaxial
• Post-contrast transaxial
• Coronal fast low angle T_1 weighted imaging with FS (SPAIR): TR 3.5 ms; TE 1.2 ms; matrix 384 × 384; FS technique: SPAIR
• Pre-contrast coronal
• Post-contrast coronal x 2 (25 s delay after injection for first acquisition, 30 s delay between second and third acquisitions)
• Coronal/transaxial steady TrueFISP imaging with FS (SPAIR): TR 4 ms; TE 2 ms; ST 7 mm ; matrix 256 × 256 (coronal), 208 × 256 (transaxial)

2D, two-dimensional; FS, fat saturation; SPAIR, spectral attenuated inversion recovery; TE = echo time; TR = repetition time.

surrounding mesenteric edema, presence of mesenteric hyperemia, presence of local mesenteric lymphadenopathy (defined *a* ≥10 mm), presence of stricture in the disease segment, and presence of upstream bowel dilation in those MREs with small bowel strictures, as described in the reference articles.[19,24] Figures 1 and 2 demonstrate case examples from our study. These features were identified as those commonly used by radiologists and that comprise many of the elements of the aforementioned scoring systems (Table 3).

The segment of small bowel affected was identified as duodenum, jejunum, or ileum. Duodenum was defined through the ligament of treitz. An artificial boundary between jejunum and ileum was demarcated by a line from the liver hilum through the left lower quadrant. Length of bowel was measured in cm, using any sequence that profiled the bowel the best. Bowel wall measurements were made using any sequence that demonstrated small bowel cross-sectional thickness the most clearly. Mucosal ulcerations were defined as deep depressions in the mucosa.[15] Reviewing radiologists were asked to describe abnormal bowel wall enhancement (as compared to adjacent vascular structures) as mucosal/submucosal, homogeneous, or stratified (so-called targetoid appearance). Strictures were identified by their persistence across multiple time points on multiple sequential series. The presence of bowel dilation was defined as diameter greater than 3 cm.[24]

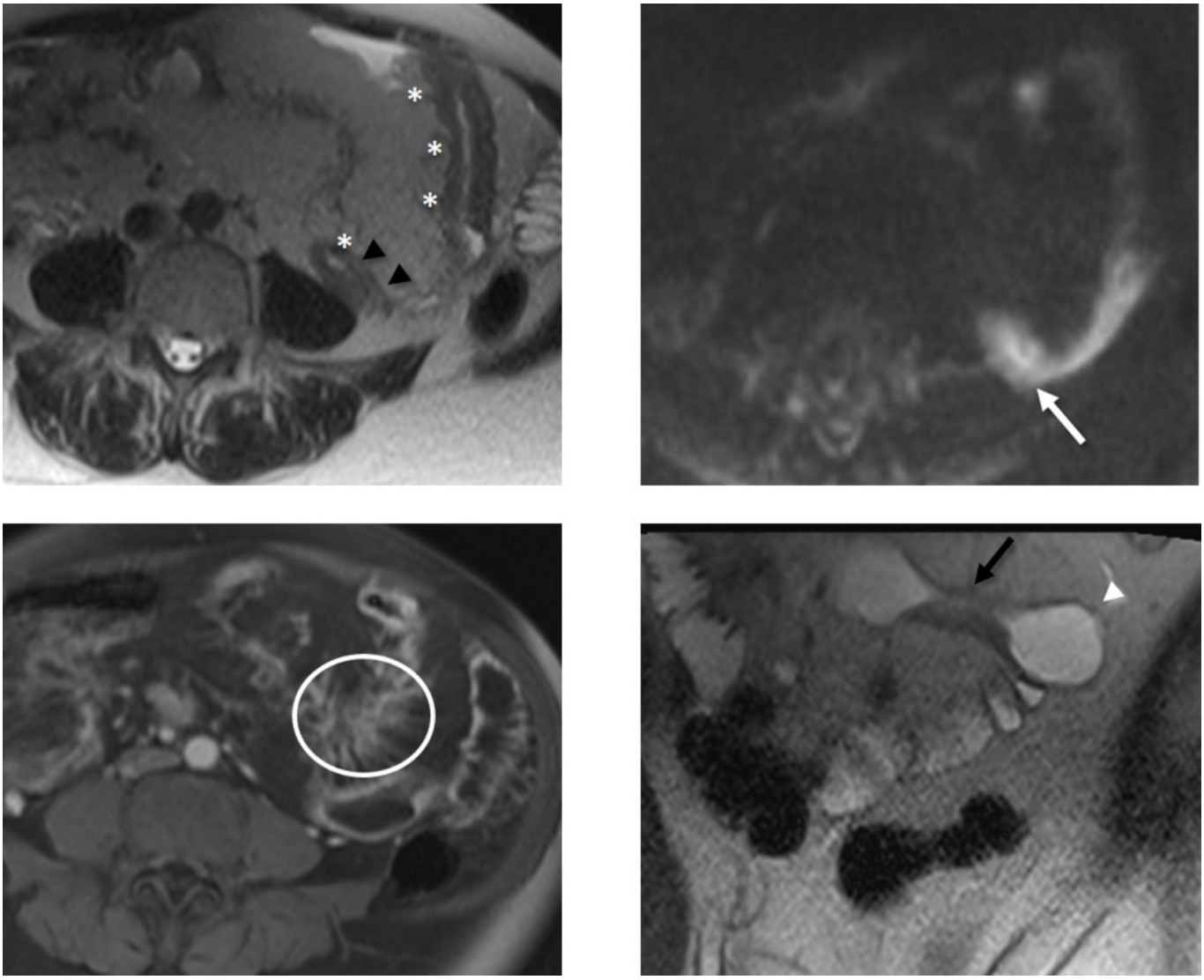
After evaluation of the MRE features, the radiologists were asked to give a general assessment of the severity of active bowel inflammation (mild, moderate, or severe); they did not provide a quantitative score. In order to establish baseline severity as interpreted by our readers, average and standard deviation for length of involvement and bowel wall thickness were calculated. An average of severity of edema and final impression was calculated from the findings provided by the two readers.

Table 2. MRI features by category and score.

MRI features	Normal/absent	Mild	Moderate	Severe
Mural T_2 signal	Normal bowel wall	Bowel wall appears gray	Bowel wall appears light gray	Bowel wall appears white or near fluid signal
T_1 enhancement	Normal bowel wall	Enhancement greater than normal bowel, but much less than vascular structures	Enhancement greater than normal bowel, but slightly less than vascular structures	Enhancement approaching or equal to adjacent vascular structures
Mucosal ulcerations (deep mucosal impressions)	Absent	Present		
Perimural, mesenteric edema (T_2 signal)	Absent	Present		
Mesenteric hyperemia (Comb sign)	Absent	Present		
Mesenteric lymphadenopathy (≥10 mm short-axis)	Absent	Present		
Small bowel diffusion restriction	Absent	Present		
Stricture (small bowel)	Absent	Present		
Upstream small bowel dilation >3 cm (if stricture is present)	Absent	Present		

Note—T_1 enhancement was also subcategorized into submucosal (inner most layer), layered (inner and outer layer), or homogeneous (all layers of bowel enhancing similarly)

Figure 1. A 60 year-old-man with Crohn's disease. MRE axial haste (a), diffusion-weighted images (b), and post-contrast T_1 weighted (c) images demonstrate a diffusion restricting (white arrow) segment of ileum with bowel wall thickening (asterisk), mesenteric hyperemia (circle), and moderate bowel wall edema (black arrowhead). Coronal haste (d) images demonstrate stricturing disease (black arrow) with upstream dilation (white arrowhead). Both readers agreed on the features for this segment of bowel. MRE, MR enterography.



Because 10 patients (20 cases) were excluded due to lack of agreement of the worst segment of small bowel, the readers were asked to re-review these cases at the conclusion of the study.

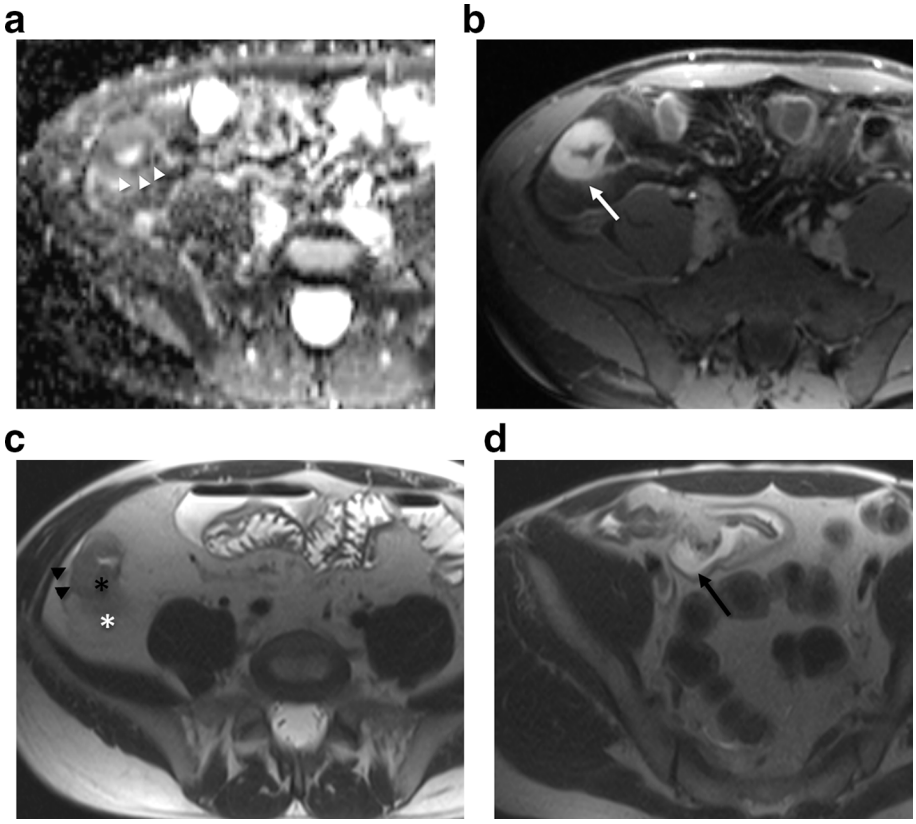
Statistical analysis

Categorical variables were summarized using frequencies. Continuous variables were summarized using mean (standard deviation, SD). Agreement between the readers for nominal categorical variables was assessed using a simple κ. A weighted κ was used to assess agreement between the raters for ordinal categorical variables. Interclass correlation coefficients (ICC) were calculated to assess agreement between raters for continuous variables.[29] κ and ICC values <0 were considered as having no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. All tests were two-sided and a significance level of 0.05 was used. SAS v. 9.4 was used to conduct data analyses (SAS Institute; Cary, NC).

RESULTS

556 patients with CD were seen in the IBD clinic at Washington University in St. Louis and had an MRE examination during the time frame from July 1, 2010 to September 1, 2017. 50 sequential patients, based on a sample of convenience and feasibility, meeting the inclusion and exclusion criteria were identified for analysis out of the first 269 charts reviewed. 126 patients were excluded due to the absence of a follow-up MRE examination, one patient due to a non-diagnostic examination, 49 patients due to the presence of penetrating disease, 16 due to normal MRE's, 7 due to isolated colonic involvement, 8 due to the absence of initiation or modification of a biologic agent, 2 due to multiple alterations in therapy between MRE,

Figure 2. A 30-year-old male with Crohn's disease. Axial T_2 weighted (a), apparent diffusion coefficient map (c), and post-contrast T_1 weighted (d) images of through the right lower quadrant demonstrate a markedly thickened loop of strictured terminal ileum (black asterisk), moderate bowel edema (black arrowhead), surrounding perimural edema (white asterisk), diffusion restriction (white arrowhead), and homogeneous enhancement (white arrow). Axial T_2 weighted image (b) through a slice more inferior demonstrate no upstream dilation (black arrow).



and 10 due to longer than 6 months between the change of therapy and subsequent MRE.

The average patient age was 44 years (sd ±15 years), with a majority female (52%). Of the included cohort, 98% (49/50) had isolated small bowel involvement with the remaining patient having large and small bowel involvement. In addition, 22% had perineal disease. 78%[30] had no past surgical bowel surgery, with 10%[5] having ileocolonic resections, 10%[5] with small bowel resections, and 2%[1] having total abdominal colectomy and end ileostomy. As interpreted by the two readers, the average length of bowel affected was 22 cm (SD 17 cm) and average bowel wall thickness was 8 mm (SD 2.4 mm). The readers found that 50% had mild, 36% had moderate, and 14% had severe degree of bowel wall enhancement and 52% had mild, 13% had moderate, and 6% had severe perimural edema. Finally, the readers found that 4% were radiologically normal and 38% had mild, 36% had moderate, and 22% had severe disease.

With regards to therapy, 75%[31] of patients were not treated with biologic agents at the time of initial MRE. At the time of the follow-up MRE, 94% (47) patients were treated with a biologic agent. Of the 50 patients who met inclusion criteria, the two readers disagreed on the location of the segment of bowel representing the worst inflammation in 10 patients (either the pre- or post-treatment scan). This left 40 patients (80 scans) to comprise our study. All 40 patients had the most severe disease in the ileum or neo-terminal ileum on both pre- and post-treatment scans as determined by both readers. Of these 40 patients, endoscopy was performed on 13 patients within 30 days of either pre- or post-treatment MRE. 8/13 patients demonstrated mucosal ulcerations by endoscopy.

Table 4 demonstrates the inter-reader agreement for the different MRE. The length of bowel involved had the highest agreement [ICC = 0.67, confidence interval; (CI) (0.46–0.81); ICC = 0.61, CI (0.37–0.77)] on the pre- and post-treament scans, respectively. Other findings with moderate agreement (κ or ICC 0.41–0.6) on the pre-treatment scans were thickness of diseased bowel [ICC = 0.56; CI (0.03–0.58)], type of enhancement pattern [κ = 0.41; CI (0.11–0.69)], mesenteric hyperemia [κ = 0.41; CI (0.28–0.7)], perimural edema [κ = 0.41; CI (0.11–0.69)], diffusion restriction [κ = 0.43; CI (0.15–0.71)], stricturing disease [κ = 0.51; (CI 0.24–0.79)], upstream dilation [κ = 0.5; CI (0.23–0.77)] and the final severity of the disease as rated by each reader [κ = 0.49; CI (0.28–0.7)]. For post-treatment scans, those features that had moderate agreement were mesenteric lymphadenopathy [κ = 0.44, CI (−0.002 to 0.90)], diffusion restriction [κ = 0.48, CI (0.21–0.78)], stricturing

Table 3. Scoring indices and MRE parameters

Scoring System	Abscess	ADC	Bowel Wall Edema	Degree of Enhancement	Enhancement Pattern	Fistula	Length	Mesenteric Hyperemia	Mesenteric Lymph nodes	Wall thickness	Peri-mural T2 Signal	StenosisUlcers
Clermont		x	x		x							x
London			x							x		
MaRIA			x	x						x		x
MEGS	X		x	x	x	x	x	x	x	x	x	x

ADC, apparent diffusion coefficient; MaRIA, Magnetic Resonance Index of Activity.

disease [κ = 0.52, CI (0.24–0.81)], and pattern of enhancement [κ = 0.45, CI (0.20–0.70)]. Several features demonstrated substantial agreement on the pre-treatment scan, but moderate agreement on the post-treatment scan, or vice versa; none of these features had significantly different agreements. The only feature that had no agreement (κ ≤ 0) on both pre- and post-treatment scans was the presence of mucosal ulcerations [κ = −.071, CI (−0.14 to −0.0003); κ = −0.042, CI (−0.11 to 0.03).

10 patients (20 cases) had bowel segments in which the readers could not agree on the most severely affected segment. These data are presented separately in Appendix A.

DISCUSSION

With shifting paradigms in treatment strategies for CD, the need for a noninvasive treatment target is increasingly recognized.[17,32] While several scoring systems have been proposed (Table 5) and validated in highly subspecialized circumstances (*i.e.* expert readers in clinical trials), the application of scoring systems has lagged behind in clinical practice. Despite, proposals by the Society of Abdominal Radiology CD Focus panel to promote standardized reporting,[33] MRE reporting and lexicon is still not standardized in many practice environments.

Similar to Tielbeek et al, our study demonstrated relatively better agreement for length of bowel wall involvement (κ = 0.62) *vs* ICC = 0.67/0.61 on pre- and post-treatment scans, respectively), degree of bowel wall thickening (κ = 0.59 *vs* ICC = 0.54/0.69), presence of mesenteric lymphadenopathy (κ = 0.35 *vs* κ = 0.39/0.44), and presence of mesenteric hyperemia (κ = 0.39 *vs* κ = 0.41/0.38) compared to the other MRE features of active CD. Two features common to all scoring systems are bowel edema and bowel wall thickness. Although bowel edema only demonstrated fair inter reader agreement (κ = 0.22/0.30) in our study, this feature demonstrated better agreement by the Tielbeek (κ = 0.66), Rimola (κ = 0.86) and Jairath (κ = 0.80) studies.[21,24,34] As stated before, wall thickness had moderate inter-rater agreement in ours and the Tielbeek study, but had substantial agreement by two other studies.[21,24,34] Tielbeek et al hypothesized these differences could be due to the severity of inflammatory bowel disease in their population; however, our study population is more similar to the population described by Rimola and Jairath, yet we too had lower inter-rater agreement. These similarities and differences in κ likely reflect the experience and specialization of the readers. The inter reader reliability was similar before initiation/modification of medical therapy as well as after initiation or modification of medical therapy. The findings are supportive of the previous studies that demonstrate CD activity can be correlated with radiologic features and CD scoring indices.[16,17,35,36]

Of note, the MRE protocol spanning the dates of data collection did not use fat-saturation T_2 weighted images which could explain the decreased agreement. In our clinical practice, we found that although increased T_2 signal in isolation may be confused with increased intramural fat deposition, often the

Table 4. Inter-rater agreement of MRE features of active Crohn's disease.

MRE feature	Pre-treatment κ/ICC	CI	Post-treatment κ/ICC	CI
Presence of mucosal ulcerations	−0.071	−0.14–0.003	−0.042	−0.11–0.03
Presence of Phlegmon/Abscess	1	1	1	1
Presence of surrounding mesenteric edema	0.41	0.17–0.65	0.38	0.043–0.72
Presence of mesenteric lymphadenopathy	0.39	0.11–0.67	0.44	−0.002–0.90
Presence of upstream dilation	0.5	0.23–0.77	0.32	0.018–0.62
Presence of diffusion restriction	0.43	0.15–0.71	0.48	0.21–0.78
Presence of stricturing disease	0.51	0.24–0.79	0.52	0.24–0.81
Comb sign	0.41	0.11–0.70	0.38	0.17–0.72
Degree of bowel edema (T_2 signal)	0.22	0.012–0.42	0.30	0.091–0.50
Degree of enhancement	0.28	0.096–0.46	0.095	−0.018–0.21
Degree of severity	0.49	0.28–0.70	0.59	0.39–0.78
Type of enhancement pattern	0.40	0.11–0.69	0.45	0.20–0.70
Bowel wall thickening	0.56*	0.30–0.58	0.69*	0.46–0.83
Length of involvement	0.67*	0.46–0.81	0.61*	0.37–0.77

CI, confidence interval; ICC, interclass correlation coefficient.
Note—Asterisk denotes ICC. For all other values, κ is used. Total number of patients = 40.

accompanying features such as perimural stranding, enhancement pattern, and less-defined borders of the increased signal could help differentiate the two findings. For the purposes of this study, the low *b*-value diffusion-weighted images served as a fat-saturated, T_2 weighted image; given the lower resolution of the diffusion images, this likely decreased sensitivity and there potentially could be intramural fat that was mistaken for bowel wall edema. However, T_2 signal is a qualitative feature, which may require better definition against an internal reference standard for consistent application.

It is important to note that in 20% of cases (10 out of the original 50) the segment identified as most severe was different between readers. For further analysis of agreement, these discordant cases were excluded from the primary analysis, but included in the appendix. As seen in Appendix A, there was severely decreased agreement in many of the MRE features, most notably in the more subjective findings. The features that seemed to perform the best across readers included presence of mesenteric lymphadenopathy and bowel wall thickening. This inter reader variability highlights a potential major pitfall in longitudinal consistency. Unlike other organs, *e.g.* the liver, the small bowel lacks major landmarks for detailing segmental location. Further research into ways to mitigate this pitfall in practice may focus on the use of screen saves and perhaps series/slice indicators for depicting the segment of bowel scored or even assigning patients to one reader.

The Society of Abdominal Radiology CD Focus panel suggests standardized reporting templates to include degree of bowel wall edema, perienteric stranding, bowel wall thickness, ulceration, and diffusion restriction in every MRE report.[33] Our results suggest that these features may require further education and better definition before widespread application in practice can yield precise and consistent results.

The use of diffusion-weighted imaging has gained attention in assessing disease severity and showed moderate agreement in our study [κ = 0.43, CI (0.15–0.71); κ = 0.48, CI (0.21–0.78)].[8,9,31,37,38] Perhaps more interestingly, the presence of mucosal ulcerations, which is weighted very heavily in the MaRIA, MEGS, and Clermont indices, showed almost no inter-reader agreement in our study(κ = −0.071), much less than those reported in the literature.[21,34] The presence of mucosal ulceration as defined, *i.e.* a "deep impression" of the mucosa, may not be specific enough to separate this feature from normal undulations of edematous small bowel by less experienced readers. It is worthwhile to note that the presence of mucosal ulcerations reported was rare (two seen by Reader 1 and four seen by Reader 2) and the low inter rater reliability may be due to its low prevalence in this study, even when 8/13 patients demonstrated mucosal ulcerations by endoscopy within 30 days of MRE. Of those eight patients with endoscopically identifiable mucosal ulcerations, only two were noted to have mucosal ulcerations on MRE.

When this feature was deemed present by one of our readers, the other reader interpreted the feature as absent. Our results were similar to those found by Tielbeek et al, in which no mucosal ulcerations were described by their interpreting radiologists. The absence of agreement on the presence of mucosal ulcerations in our patient cohort is surprising given the controlled population of patients, each who were deemed eligible for biological agents as a therapeutic choice based on clinical, imaging, and endoscopic findings. In a similarly diseased population, Jairath et al, demonstrated a much

higher prevalence and agreement for ulcerations [κ = 0.6, CI (0.45–0.72)], within the terminal ileum [mean κ = 0.51, CI (0.34–0.62), over all segments].[34] Based on our results, we would suggest additional clarification of this lexicon term and perhaps development of an educational pictorial atlas to help readers apply this feature more consistently in practice. Given the importance of mucosal ulcerations to the MaRIA score, this seems an essential step prior to widespread use for clinical trials or practice.

Additional limitations in our study should be acknowledged including the retrospective, observational nature which inherently introduces bias. Our inclusion criteria only documented those patients that had two subsequent MRE examinations—one examination prior to, and another after the initiation or modification of a biologic agent as this was part of a larger ongoing study. While this introduces some selection bias, it also provides a measure of control in the population and ensures an adequate frequency and severity of disease representation. In addition, patients were seen at a tertiary clinic specializing in inflammatory bowel disease. Patients with isolated colonic involvement were also excluded. These factors introduce selection bias in selecting patients that may not be representative of all clinical practice environments for CD.

Likewise, greater disease severity may overestimate the inter reader reliability as worse disease may be easier to diagnose.[39] Furthermore, although the readers were instructed to select the single worst segment of small bowel involved, all cases selected involved the ileum. For this reason, it is unclear if these findings can be extrapolated to other portions of the bowel, particularly the jejunum, where increased fold density may hinder the detection of abnormal bowel wall enhancement. We rarely saw mucosal ulceration in our study and given that only 2/8 patients were identified by either reader, this finding is probably a challenging observation to make in less experienced readers (<700 MRE's). However, the presence of mucosal ulcerations is a important branchpoint in prognosis, therapy, and disease response, and more research needs to be performed to bridge the gap between non-expert and expert CD readers.

Standardized terminology and scoring systems/diagnostic algorithms are intended to provide consistency and improved clarity of reporting. Several studies have shown that application of diagnostic algorithm can narrow the gap between expert and novice readers.[30,40] This should also be the goal for reporting of small bowel CD in order to facilitate useful radiological endpoints for clinical care and trials. Toward this end, future efforts may focus on clarifying the lexicon and investigating educational resources that might also improve final application of standardized terminology. Future research should also aim to validate the benefit of scoring systems, by assessing the impact on clinical decision making.

In conclusion, non-CD expert, subspecialty trained radiologists have moderate agreement for many of the MRE features of active CD that comprise the major severity scoring indices. The poor agreeability of mucosal ulcerations should give second thought on putting such heavy emphasis on this feature. Further clarification of lexicon is advised prior to widespread application in clinical practice or as a clinical trial endpoint.

REFERENCES

1. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on Systematic review. *Gastroenterology* 2012; **142**: e42.: 46: 54. doi: https://doi.org/10.1053/j.gastro.2011.10.001
2. Torres J, Mehandru S, Colombel J-F, Peyrin-Biroulet L, disease Crohn's. Crohn's disease. *The Lancet* 2017; **389**: 1741–55. doi: https://doi.org/10.1016/S0140-6736(16)31711-1
3. Balzola F, Cullen G, GT H, Russell RK, Wehkamp J. Endoscopic skipping of the distal terminal ileum in Crohn's disease can lead to negative results from ileocolonoscopy. *Inflammatory Bowel Disease Monitor* 2012;.
4. Deepak P, Fowler KJ, Fletcher JG, Bruining DH. Novel imaging approaches in inflammatory bowel diseases. *Inflamm Bowel Dis* 2018;.
5. Fowler KJ, Dassopoulos T, Raptis C. Magnetic resonance Enterography in the evaluation of Crohn's disease. *Inflammatory Bowel Diseases* 2014; **20**: 2179–88. doi: https://doi.org/10.1097/MIB.0000000000000132
6. Prezzi D, Bhatnagar G, Vega R, Makanyanga J, Halligan S, Taylor SA. Monitoring Crohn's disease during anti-TNF-α therapy: validation of the magnetic resonance enterography global score (MEGS) against a combined clinical reference standard. *Eur Radiol* 2016; **26**: 2107–17. doi: https://doi.org/10.1007/s00330-015-4036-1
7. Lee SS, Kim AY, Yang S-K, Chung J-W, Kim SY, Park SH, et al. Crohn disease of the small bowel: comparison of CT Enterography, MR Enterography, and small-bowel Follow-Through as diagnostic techniques. *Radiology* 2009; **251**: 751–61. doi: https://doi.org/10.1148/radiol.2513081184
8. Buisson A, Joubert A, Montoriol P-F, Ines DD, Hordonneau C, Pereira B, et al. Diffusion-weighted magnetic resonance imaging for detecting and assessing ileal inflammation in Crohn's disease. *Aliment Pharmacol Ther* 2013; **37**: 537–45. doi: https://doi.org/10.1111/apt.12201
9. Hordonneau C, Buisson A, Scanzi J, Goutorbe F, Pereira B, Borderon C, et al. Diffusion-weighted magnetic resonance imaging in ileocolonic Crohn's disease: validation of quantitative index of activity. *American Journal of Gastroenterology* 2014; **109**: 89–98. doi: https://doi.org/10.1038/ajg.2013.385
10. 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. *Inflammatory Bowel Diseases* 2011; **17**: 1081–8. doi: https://doi.org/10.1002/ibd.21534
11. Siddiki HA, Fidler JL, Fletcher JG, Burton SS, Huprich JE, Hough DM, et al. Prospective comparison of state-of-the-art Mr Enterography and CT Enterography

in small-bowel Crohn's disease. *American Journal of Roentgenology* 2009; **193**: 113–21. doi: https://doi.org/10.2214/AJR.08.2027

12. Fiorino G, Bonifacio C, Peyrin-Biroulet L, Minuti F, Repici A, Spinelli A, et al. Prospective comparison of computed tomography enterography and magnetic resonance enterography for assessment of disease activity and complications in ileocolonic Crohn's disease. *Inflammatory Bowel Diseases* 2011; **17**: 1073–80. doi: https://doi.org/10.1002/ibd.21533
13. Ippolito D, Invernizzi F, Galimberti S, Panelli MR, Sironi S. Mr enterography with polyethylene glycol as oral contrast medium in the follow-up of patients with Crohn disease: comparison with CT enterography. *Abdom Imaging* 2010; **35**: 563–70. doi: https://doi.org/10.1007/s00261-009-9557-0
14. Panes J, Bouhnik Y, Reinisch W, Stoker J, Taylor SA, Baumgart DC, et al. Imaging techniques for assessment of inflammatory bowel disease: joint ECCO and ESGAR evidence-based consensus guidelines. *J Crohn's Colitis* 2013; **7**: 556–85. doi: https://doi.org/10.1016/j.crohns.2013.02.020
15. Rimola J, Rodriguez S, Garcia-Bosch O, Ordas I, Ayala E, Aceituno M, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. *Gut* 2009; **58**: 1113–20. doi: https://doi.org/10.1136/gut.2008.167957
16. Ordás I, Rimola J, Rodríguez S, Paredes JM, Martínez-Pérez MJ, Blanc E, et al. Accuracy of magnetic resonance Enterography in assessing response to therapy and mucosal healing in patients with Crohn's disease. *Gastroenterology* 2014; **146**: 374–82. doi: https://doi.org/10.1053/j.gastro.2013.10.055
17. Deepak P, Fletcher JG, Fidler JL, Barlow JM, Sheedy SP, Kolbe AB, et al. Radiological response is associated with better long-term outcomes and is a potential treatment target in patients with small bowel Crohn's disease. *American Journal of Gastroenterology* 2016; **111**: 997–1006. doi: https://doi.org/10.1038/ajg.2016.177
18. Steward MJ, Punwani S, Proctor I, Adjei-Gyamfi Y, Chatterjee F, Bloom S, et al. Non-perforating small bowel Crohn's disease assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. *European Journal of Radiology* 2012; **81**: 2080–8. doi: https://doi.org/10.1016/j.ejrad.2011.07.013
19. Rimola J, Planell N, Rodríguez S, Delgado S, Ordás I, Ramírez-Morros A, et al. Characterization of Inflammation and Fibrosis in Crohn's Disease Lesions by Magnetic Resonance Imaging. *American Journal of Gastroenterology* 2015; **110**: 432–40. doi: https://doi.org/10.1038/ajg.2014.424
20. Coimbra AJF, Rimola J, O'Byrne S, Lu TT, Bengtsson T, de Crespigny A, et al. Magnetic resonance enterography is feasible and reliable in multicenter clinical trials in patients with Crohn's disease, and may help select subjects with active inflammation. *Aliment Pharmacol Ther* 2016; **43**: 61–72. doi: https://doi.org/10.1111/apt.13453
21. Rimola J, Ordás I, Rodriguez S, García-Bosch O, Aceituno M, Llach J, et al. Magnetic resonance imaging for evaluation of Crohn's disease. *Inflammatory Bowel Diseases* 2011; **17**: 1759–68. doi: https://doi.org/10.1002/ibd.21551
22. Makanyanga JC, Pendsé D, Dikaios N, Bloom S, McCartney S, Helbren E, et al. Evaluation of Crohn's disease activity: Initial validation of a magnetic resonance enterography global score (MEGS) against faecal calprotectin. *Eur Radiol* 2014; **24**: 277–87. doi: https://doi.org/10.1007/s00330-013-3010-z
23. Rimola J, Alvarez-Cofiño A, Pérez-Jeldres T, Ayuso C, Alfaro I, Rodríguez S, et al. Comparison of three magnetic resonance enterography indices for grading activity in Crohn's disease. *J Gastroenterol* 2017; **52**: 585–93. doi: https://doi.org/10.1007/s00535-016-1253-6
24. Tielbeek JAW, Makanyanga JC, Bipat S, Pendsé DA, Nio CY, Vos FM, et al. Grading Crohn disease activity with MRI: interobserver variability of MRI features, MRI scoring of severity, and correlation with Crohn disease endoscopic index of severity. *American Journal of Roentgenology* 2013; **201**: 1220–8. doi: https://doi.org/10.2214/AJR.12.10341
25. Church PC, Greer M-LC, Cytter-Kuint R, Doria AS, Griffiths AM, Turner D, et al. Magnetic resonance enterography has good inter-rater agreement and diagnostic accuracy for detecting inflammation in pediatric Crohn disease. *Pediatr Radiol* 2017; **47**: 565–75. doi: https://doi.org/10.1007/s00247-017-3790-4
26. Shih I-L, Lee T-C, Tu C-H, Chang C-C, Wang Y-F, Tseng Y-H, et al. Intraobserver and interobserver agreement for identifying extraluminal manifestations of Crohn's disease with magnetic resonance enterography. *Advances in Digestive Medicine* 2016; **3**: 174–80. doi: https://doi.org/10.1016/j.aidm.2015.05.004
27. 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**: 1008–16. doi: https://doi.org/10.1080/02841850600979071
28. Sharman A, Zealley IA, Greenhalgh R, Bassett P, Taylor SA. MRI of small bowel Crohn's disease: determining the reproducibility of bowel wall gadolinium enhancement measurements. *Eur Radiol* 2009; **19**: 1960–7. doi: https://doi.org/10.1007/s00330-009-1371-0
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74. doi: https://doi.org/10.2307/2529310
30. Davenport MS, Khalatbari S, Liu PSC, Maturen KE, Kaza RK, Wasnik AP, et al. Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. *Radiology* 2014; **272**: 132–42. doi: https://doi.org/10.1148/radiol.14131963
31. Kim K-J, Lee Y, Park SH, Kang B-K, Seo N, Yang S-K, et al. Diffusion-weighted MR Enterography for evaluating Crohn's disease. *Inflammatory Bowel Diseases* 2015; **21**: 101–9. doi: https://doi.org/10.1097/MIB.0000000000000222
32. Deepak P, Fletcher JG, Fidler JL, Bruining DH. Computed Tomography and Magnetic Resonance Enterography in Crohn's Disease: Assessment of Radiologic Criteria and Endpoints for Clinical Practice and Trials. *Inflamm Bowel Dis* 2016; **22**: 2280–8.
33. Bruining DH, Zimmermann EM, Loftus EV, Sandborn WJ, Sauer CG, Strong SA, et al. Consensus Recommendations for Evaluation, Interpretation, and Utilization of Computed Tomography and Magnetic Resonance Enterography in Patients With Small Bowel Crohn's Disease. *Radiology* 2018; **286**: 776–99. doi: https://doi.org/10.1148/radiol.2018171737
34. Jairath V, Ordas I, Zou G, Panes J, Stoker J, Taylor SA, et al. Reliability of Measuring Ileo-Colonic Disease Activity in Crohn's Disease by Magnetic Resonance Enterography. *Inflamm Bowel Dis* 2018; **24**: 440–9. doi: https://doi.org/10.1093/ibd/izx040
35. Huh J, Kim KJ, Park SH, Park SH, Yang S-K, Ye BD, , et al. Diffusion-weighted MR Enterography to monitor bowel inflammation after medical therapy in Crohn's disease: a prospective longitudinal study. *Korean J Radiol* 2017; **18**: 162. doi: https://doi.org/10.3348/kjr.2017.18.1.162
36. Buisson A, Pereira B, Goutte M, Reymond M, Allimant C, Obritin-Guilhen H, et al. Magnetic resonance index of activity (MaRIA) and Clermont score are highly

and equally effective MRI indices in detecting mucosal healing in Crohn's disease. *Digestive and Liver Disease* 2017; **49**: 1211–7. doi: https://doi.org/10.1016/j.dld.2017.08.033

37. Park SH. DWI at Mr Enterography for evaluating bowel inflammation in Crohn disease. *American Journal of Roentgenology* 2016; **207**: 40–8. doi: https://doi.org/10.2214/AJR.15.15862
38. Tielbeek JAW, Ziech MLW, Li Z, Lavini C, Bipat S, Bemelman WA, et al. Evaluation of conventional, dynamic contrast enhanced and diffusion weighted MRI for quantitative Crohn's disease assessment with histopathology of surgical specimens. *Eur Radiol* 2014; **24**: 619–29. doi: https://doi.org/10.1007/s00330-013-3015-7
39. 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**: 1450–60. doi: https://doi.org/10.1007/s00330-008-1287-0
40. Becker AS, Barth BK, Marquez PH, Donati OF, Ulbrich EJ, Karlo C, et al. Increased interreader agreement in diagnosis of hepatocellular carcinoma using an adapted LI-RADS algorithm. *European Journal of Radiology* 2017; **86**: 33–40. doi: https://doi.org/10.1016/j.ejrad.2016.11.004