

# Comparison of three magnetic resonance enterography indices for grading activity in Crohn's disease

Jordi Rimola<sup>1</sup> · Almudena Alvarez-Cofiño<sup>2</sup> · Tamara Pérez-Jeldres<sup>3</sup> · Carmen Ayuso<sup>1,5</sup> · Ignacio Alfaro<sup>4</sup> · Sonia Rodríguez<sup>1</sup> · Elena Ricart<sup>4,5</sup> · Ingrid Ordás<sup>4</sup> · Julián Panés<sup>4,5</sup> 

Received: 15 June 2016 / Accepted: 21 August 2016 / Published online: 6 September 2016  
© Japanese Society of Gastroenterology 2016

## Abstract

**Background** Magnetic resonance enterography (MRE) is an accurate examination for assessing activity in Crohn's disease (CD). Various MRE indices have been developed for that purpose, but have not been directly compared. The aim of the study was to compare the diagnostic accuracy of three MRE indices for detecting and grading disease activity in CD, using endoscopy as gold standard.

**Methods** MRE and ileocolonoscopies performed within 1 month in 43 patients with CD were analyzed. The magnetic resonance index of activity (MaRIA), Clermont, and London indices for each colonic segment and the terminal ileum were calculated. Simplified endoscopy score for CD (SES-CD) was considered the gold standard.

**Results** Two hundred and twenty-four intestinal segments were included in the analysis. According to the established cut-off points for detecting active disease using MaRIA,

Clermont, and London indices, the sensitivity of each index was 0.88, 0.89, and 0.71, and the specificity was 0.97, 0.78, and 0.99, respectively. The sensitivity for detecting ulcerations was 0.90 and 0.83 for the MaRIA and Clermont indices, respectively, with a specificity of 0.91 and 0.89. The AUROC curve for the MaRIA, Clermont, and London indices for detecting active disease was 0.92, 0.84, and 0.85, and for detecting ulcerations was 0.90 for the MaRIA, and 0.86 for Clermont index.

**Conclusions** The three MRE-based indices evaluated in the current study have high diagnostic accuracy for assessment of disease activity. The MaRIA index has the best operational characteristics for detecting not only disease activity but also for grading severity, which supports its use in clinical studies and clinical practice.

**Keywords** Crohn's disease · Diagnostic accuracy · Diffusion-weighted imaging · Disease activity · Disease complications · Magnetic resonance enterography

**Electronic supplementary material** The online version of this article (doi:10.1007/s00535-016-1253-6) contains supplementary material, which is available to authorized users.

✉ Julián Panés  
jpanes@clinic.cat

- <sup>1</sup> Department of Radiology, Hospital Clínic de Barcelona, University of Barcelona, Barcelona, Spain
- <sup>2</sup> Department of Radiology, Hospital Fundación Hospital de Jove, Gijón, Spain
- <sup>3</sup> Gastroenterology Department, Hospital San Borja Arriaran, Santiago, Chile
- <sup>4</sup> Gastroenterology Department, Hospital Clínic de Barcelona, IDIBAPS, University of Barcelona, C/Villarroel 170 Escala 7 Planta 4, 08036 Barcelona, Spain
- <sup>5</sup> Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Madrid, Spain

## Abbreviations

AUC	Area under the curve
CD	Crohn's disease
MRE	Magnetic resonance enterography
ROC	Receiver-operator characteristic
CI	Confidence interval
SES-CD	Simplified endoscopic score for Crohn disease
Anti-TNF	Anti-tumor necrosis factor

## Introduction

Ileocolonoscopy is the reference standard for objective evaluation of disease activity in Crohn's disease (CD) [1]. However, endoscopy only allows evaluation of the

mucosal surface precluding assessment of transmural inflammation and extraluminal complications. There is a growing interest in using magnetic resonance enterography (MRE) as part of the assessment of inflammatory CD lesions both in clinical practice and in research [2]. MRE could overcome some inherent limitations of ileocolonoscopy. First, MRE enables the assessment of disease activity out of the reach of the endoscope due to disease location (isolated small bowel disease) or technical reasons, which may facilitate inclusion of a broader population of patients in trials [3]. Second, inclusion of patients based on MRE criteria at baseline may result in a selection of a more homogeneous patient population, increasing the efficiency of the study. Third, MRE may improve safety results by identification of CD-related complications such as abscesses. Finally, as observed with endoscopy, assessment based on MRE would be a more reliable measure of changes in inflammatory status than symptom-based instruments such as the Crohn's disease activity index (CDAI).

In recent years, various MRE indices have been derived to evaluate disease activity in CD [4], but only the magnetic resonance index of activity (MaRIA) [5, 6] and London indices [7] have been constructed using a structured process.

More recently, the Clermont index, a new index similar to the MaRIA but including functional imaging, namely diffusion-weighted imaging (DWI), has been described, obtaining promising results [8]. Parallel to the development of the Clermont index, the same group reported the potential benefit of using a single quantitative measurement of diffusion, namely the apparent diffusion coefficient (ADC) for detecting segments with severe activity, i.e., segments with ulcers. The introduction of this measure in clinical trials would have the advantage of avoiding the use of gadolinium contrast, thus improving its acceptance. However, to select the most appropriate tool for implementation in clinical trials, a comparison of diagnostic performance between the three above-mentioned MRE indices would be necessary. Hence, the aim of this study was to compare the diagnostic accuracy of the three MRE indices for detecting activity and for categorizing the severity of inflammation.

## Patients and methods

This is a retrospective study performed at a single tertiary IBD referral center. The institutional review board approved the study and waived informed consent. The study was conducted according to the good clinical practice guidelines of the European Medicines Agency.

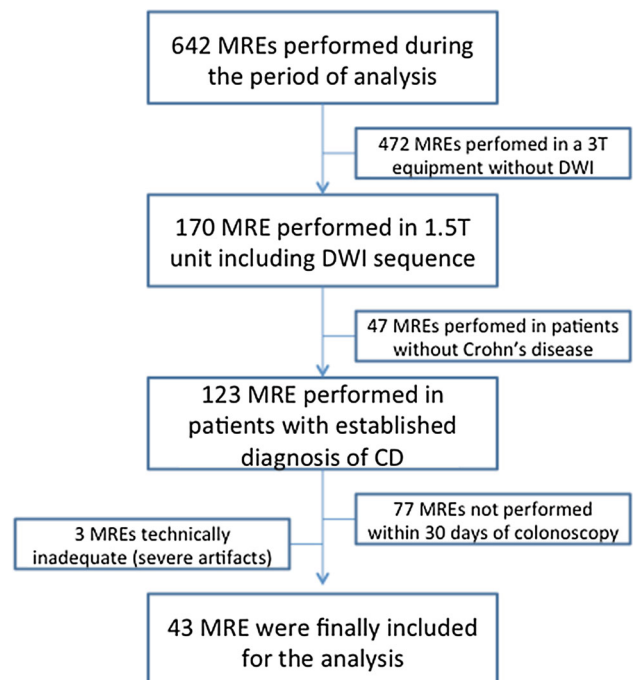
## Patient selection

Using the hospital radiology information system and patient databases, we identified eligible patients with established CD who underwent MRE scans between 2012 (introduction of routine clinical MRE protocol including DWI in a 1.5-T unit at our institution) and 2014, and who had an ileocolonoscopy performed within 1 month of MRE. Patients were included only if they did not receive any change in therapeutic interventions during this interval. Our routine MRE protocol remained with no changes during this period of time. The flow chart of patient selection is represented in Fig. 1. Patients were referred to the radiology department (MRE) for suspected activity of CD. Exclusion criteria included: inadequate MRE quality (i.e., important artifacts degrading image quality), and/or MRE not performed within 1 month of ileocolonoscopy.

Clinical disease activity based on the calculation of the Harvey–Bradshaw [9] index, laboratory tests, as well as concomitant therapy at the time of examination were collected based on patients' record files.

## Endoscopic data collection

Ileocolonoscopy was considered the reference standard for the evaluation of disease activity, extension, and severity. Patients followed a bowel cleansing protocol with oral ingestion of 1500–2000 ml of an isoosmotic polyethylene



**Fig. 1** Study flow diagram. *MRE* magnetic resonance enterography

glycol and electrolyte solution (Norgine Limited, Mid Glamorgan, United Kingdom) on the evening before the examination. Endoscopies were performed under anesthesia with propofol (Mayne Pharma, Madrid, Spain) and remifentanyl (GlaxoSmithKline, Madrid, Spain). The severity and extent of inflammatory lesions were evaluated in each colonic segment (ascending, transverse, descending, sigmoid, rectum) and in the terminal ileum using the simplified endoscopy score for Crohn's disease (SES-CD) [10], which is routinely assessed in all endoscopies in CD patients performed by two expert endoscopists (IO, ER). For SES-CD calculation, the endoscopic variables were as originally defined: ulcers (size and ulcerated surface), affected surface (including all inflammatory lesions in addition to ulcers), and strictures. For the purpose of this study, we scored separately the sigmoid and the descending colon, as in previous studies [5].

For each segment, a SES-CD  $\geq 2$  was considered as positive for active disease. In addition, based on endoscopic findings, each segment was categorically classified as inactive, mild-moderate activity when inflammatory lesions other than ulcers were identified, and as severe, when ulcers with a diameter  $>5$  mm were identified.

### Magnetic resonance enterography imaging

All MRE examinations were performed using a 1.5-T MR unit (Aera; Siemens Medical Solutions, Erlangen, Germany) by a trained technician acquiring MRE as described in previous studies [5]. In addition, diffusion-weighted imaging (DWI) was acquired in axial plane using three *b* values (50, 600, and 800 s/mm<sup>2</sup>) before the acquisition of T1 sequences. Technical characteristics of sequences are summarized in Supplementary Table 1.

### MRE image interpretation

MRE image analysis was performed by two-experienced radiologists (JR and AC). MRE images were read with Osirix MD DICOM viewers, an FDA-cleared Class II medical device, on Apple Macbook Pro computers.

Two different sets of images from each MRE examination were generated. Set 1 included T2 sequences, DWI and apparent diffusion coefficient (ADC) maps; set 2 included T2 sequences and T1 sequences. Reader 1 interpreted images of Set 1, and reader 2 interpreted images of Set2. Set 1 was used to calculate the Clermont and London indices, whereas Set 2 was used to calculate the MaRIA.

The three MRE indices of disease activity (MaRIA, Clermont, and London) were calculated in each segment according to the established formulas [5, 7, 8]. The cut-off points previously established for differentiating active from inactive were 7 for MaRIA, 8.4 for Clermont index, and 4.1 for London index. The cut-off points for severe inflammation were 11 for MaRIA and 12.5 for Clermont index. Additionally, a cut-off for detecting ulcers using the measurement of ADC values in each segment was previously defined as 1880 [11]. No cut-off value has been defined for the London index associated with severe inflammation. For Clermont index calculation and, in particular, for obtaining ADC measurements, readers used the average of three ROIs measurement placed over ADC maps for each segment. When the segment identification was difficult on ADC maps, readers used the images obtained at DWI sequence for guiding ROIs placement.

The presence of stenosis (defined as persistent luminal narrowing in area of CD with or without upstream dilation), fistula and abscesses/inflammatory masses detected in each set of images was recorded in a written form. In case of disagreement between readers in the detection of complications, a third radiologist with 10 years of experience in MRE (SR) acted as adjudicator. Radiologists were blinded to patient's symptoms, and to the results of endoscopic findings.

### Statistical analysis

Continuous variables are described as median and IQR, and categorical variables as absolute frequencies and percentages. The relationship between categorical variables was assessed by Chi-square test of independence computing the *p* value. McNemar test was used for comparison of sensitivity, specificity, and diagnostic accuracy between two different tests. Correlations of quantitative variables were tested using the Spearman rank test. Receiver-operator characteristic (ROC) area under the curve (AUC) was calculated to assess the ability of each MRE index to discriminate between active vs. inactive and severe active vs. non severe disease based on SES-CD. A random set ten MREs, representing approximately 25 % of the total of MREs, was read by both readers, and was used to calculate the interobserver agreement using the intra-class correlation coefficients (ICC) [12] for each scores and kappa statistics for binary classification of active vs. inactive and severe-active vs. non-severe active disease according to MaRIA and Clermont indices. A *p* value of less than 0.05 was considered significant. Analyses were performed using SPSS Statistics version 20.

## Results

### Demographic and clinical data

Forty-three patients fulfilled the inclusion criteria. Demographic and clinical characteristics are summarized in Table 1.

Ten patients (23.25 %) had incomplete ileocolonoscopy due to severe stenosis or impossibility to intubate the terminal ileum. By contrast, MRE could assess all potentially evaluable intestinal segments. Therefore, from the 256 segments potentially evaluable by endoscopy and MRE, 224 were examined by both techniques (31 ileal, 193 colonic) and represent the basis for establishing accuracy of the three MRE indices.

### Accuracy and performance characteristics of MRE indices for predicting active disease

From the 224 intestinal segments evaluated by ileocolonoscopy, 59 had a SES-CD  $\geq 2$ , and were categorized as having active disease; 23 corresponded to ileum and 36 to colonic segments.

According to pre-defined cut-off values for detecting activity, the sensitivities of MaRIA, Clermont and London indices for detecting a segmental SES-CD  $\geq 2$  were 0.88, 0.89, and 0.71, respectively, whereas the specificities were 0.96, 0.79, and 0.99, respectively.

Table 2 shows the results of the comparison of the three indices for detection of active disease. In terms of sensitivity MaRIA (88.1 %) and Clermont (89.9 %) indices, performed similarly, and were significantly superior to London index (71.2 %;  $p = 0.002$ , and  $p = 0.001$  respectively). MaRIA (97 %) and London (99.4 %) indices had the highest specificities, which were significantly higher to the specificity of the Clermont index (78.2 %,  $p < 0.0001$  for both comparisons). The MaRIA (94.6 %) and London (92 %) indices had the highest overall accuracies, which were significantly superior to Clermont index (81.3 %,  $p < 0.0001$ , and  $p < 0.007$ , respectively).

The area under the ROC curve for detecting a SES-CD  $\geq 2$  using MaRIA, Clermont, and London indices were 0.93, 0.84, and 0.85, respectively (Supplementary Figure S1).

Significant statistical differences were found between the different areas under ROC curves for detecting segments with active lesions for each index with MaRIA having significantly higher AUC compared to Clermont or London indices (Table 2).

The correlation of each index with the SES-CD was 0.68 (95 % CI 0.600–0.743) ( $p < 0.001$ ), 0.68 (95 % CI 0.609–0.749) ( $p < 0.001$ ) and 0.80 (95 % CI 0.753–0.846)

( $p < 0.001$ ) for the MaRIA, Clermont, and London indices, respectively. There were statistically significant differences between London and MaRIA correlations with SES-CD ( $p < 0.001$ ) and between London and Clermont correlations with SES-CD ( $p < 0.001$ ) but not between MaRIA and Clermont correlations with SES-CD ( $p = 0.75$ ).

### Detection of endoscopically severe disease by MRE

Severe disease, defined as the presence of ulcers  $> 5$  mm at endoscopy in a segment, was observed in 30 segments (Fig. 2). According to the pre-defined cut-off points for the detection of ulcers using MRE indices, 11 for the MaRIA, and 12.5 for the Clermont index, the sensitivities for detecting ulcerations were 0.90 and 0.83 and the specificities were 0.91 and 0.89, respectively. Although ADC is not an index itself, considering the previously reported cut-off point, we also included this metric measurement in our comparisons of diagnostic accuracy for the detection of ulcers. The sensitivity and specificity of ADC for detecting ulcerations at endoscopy considering the previously defined cut-off point of 1880 were 100 % and 12 %, respectively.

When comparing the sensitivities and specificities for detecting severe disease (presence of ulcers at endoscopy), no statistically significant differences were found between MaRIA and Clermont indices, but the MaRIA showed significantly higher accuracy than Clermont index. There were no significant differences on the sensitivities of the three tools but both MaRIA and Clermont indices had significantly superior specificity and accuracy compared to ADC alone (Table 3).

The area under ROC curve for detecting ulcers was 0.91 for MaRIA, 0.86 for the Clermont index, and 0.55 for ADC (Supplementary Figure S2). We found statistically significant differences in the area under ROC curves between ADC and MaRIA ( $p < 0.0001$ ), and ADC and Clermont index ( $p < 0.0001$ ), and a trend towards significant differences between MaRIA and Clermont indices ( $p = 0.052$ ) (Table 3). The London index has no pre-defined cut-off point for predicting the presence of ulcers at endoscopy, therefore it was not included in this part of the analysis.

### Receiver operating characteristic (ROC) curves for optimized cut-off points

In order to validate cut-off points for each index, including ADC, we determined the best cut-off for each index in our cohort of patients. Best cut-off points for predicting activity were 6.75 for MaRIA, 8.87 for Clermont index, 3.4 for London index, and 1368 for ADC, whereas for the

**Table 1** Demographic characteristics of 43 patients included in the study

Baseline characteristics	
Age, median (IQR)	39 (27–50)
Gender, <i>n</i> (%)	Male 17 (39 %)
	Female 26 (61 %)
Disease duration (years), median (IQR)	5.08 (1.25–17.5)
Smoking habit, <i>n</i> (%)	Smoker 17 (39 %)
	Non smoker 26 (61 %)
Days between MRE and endoscopy median (IQR)	8 (1–20)
Age at diagnosis (years), <i>n</i> (%)	
A1 (<16)	5 (12 %)
A2 (17–40)	31 (72 %)
A3 (>40)	7 (16 %)
Behavior, <i>n</i> (%)	
B1 (inflammatory)	20 (45.4 %)
B2 (stricturing)	7 (16.6 %)
B3 (penetrating)	16 (38 %)
Location no.(%)	
L1 (ileal)	18 (41.8 %)
L2 (colonic)	8 (18.6 %)
L3 (ileocolonic)	16 (37.2 %)
L4 (upper GI tract)	13 (30.2 %)
Perianal disease, <i>n</i> (%)	12 (27.9 %)
Previous resective surgery <i>n</i> (%)	14 (32.5 %)
Harvey Bradshaw Index, median (IQR)	6 (3–8)
SES-CD median (IQR)	6 (2–10)
Biomarkers, median (IQR)	
CRP (mg/dl)	0.61 (0.105–2.45)
ESR mm/h	18 (8–46)
Hemoglobin (g/l)	130.3 (110.3–140.3)
Platelets (10 <sup>9</sup> /l)	296 (243–389)
Albumin (g/l)	41 (35–44)
Active treatments at the time of inclusion no. (%)	
None	2 (4.7 %)
Anti-TNF	16 (38 %)
Immunomodulators (methotrexate/thiopurine)	17 (40.5 %)
Corticosteroids	9 (21.4 %)
AntiTNF + immunomodulators	12 (28.5 %)
Other treatments	9 (21.4 %)

Other treatment [Antibiotics (6) and study drugs (3)]

*IQR* interquartile range, *SES-CD* simplified endoscopic score for Crohn disease, *Anti-TNF* anti-tumor necrosis factor

diagnosis of severe disease the best cut-off points were 11.6 for MaRIA, 14.6 for Clermont index, 3.8 for London index and 1293 for ADC. Sensitivities, specificities and area under ROC curve for each index are reported in Table 4. The MaRIA best cut-off points calculated in the current population, are almost identical to those previously established using another equipment, and using the new

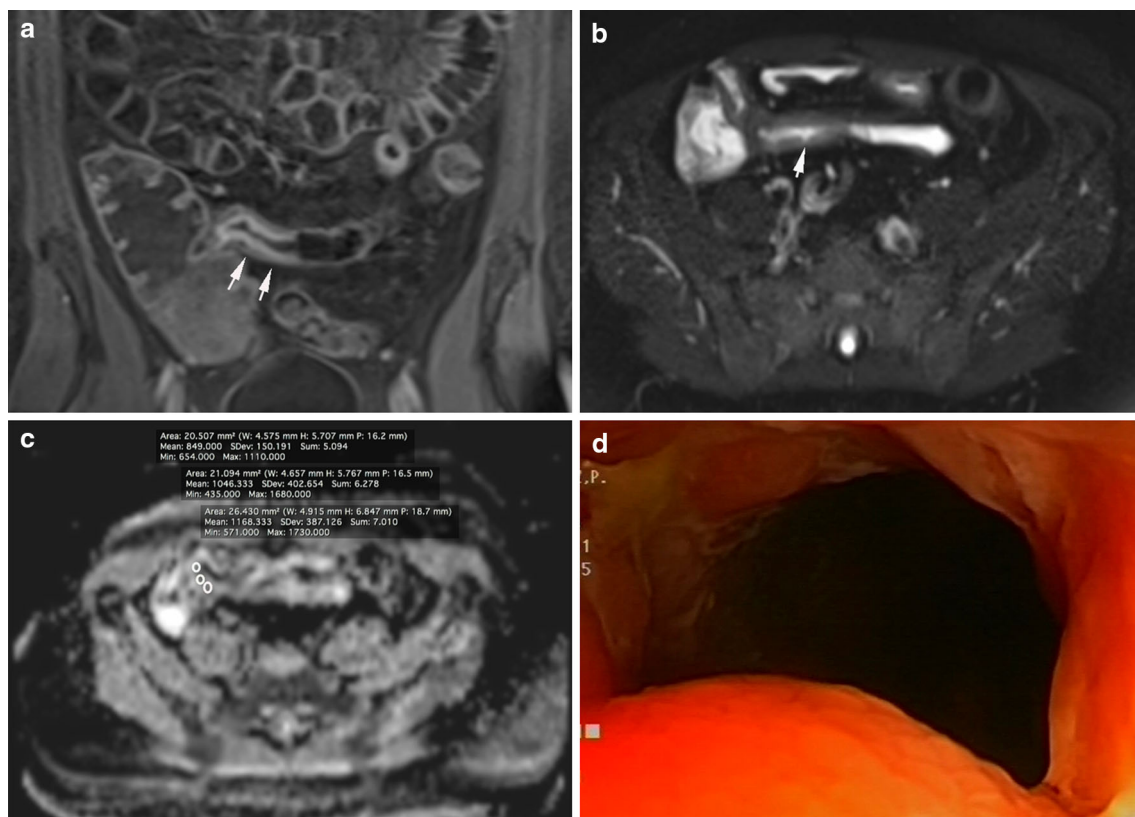
cut-off points had no effect on the diagnostic performance of this index. As for Clermont index, the cut-off points were somehow different from those previously described, but the effect of changing these points on diagnostic performance was minor. By contrast, the cut-off point for ADC was markedly different, and using the new cut-off considerably improved the performance of ADC.

**Table 2** Comparison of sensitivity, specificity and accuracy of each MRE index for detecting segments with active disease at endoscopy (SES-CD  $\geq 2$ )

	Sensitivity	Specificity	Accuracy	AUROC
MaRIA	52/59 (88.1 % [79–96])	160/165 (97 % [94–99 %])	212/224 (94.6 % [91–97])	0.930 (0.89–0.96)
Clermont	53/59 (89.9 % [82–97])	129/165 (78.2 % [72–84])	182/224 (81.3 % [76–86])	0.840 (0.78–0.88)
London	42/59 (71.2 % [59–82])	164/165 (99.4 % [98–100])	206/224 (92 % [88–95])	0.853 (0.80–0.89)
<i>p</i> values				
MaRIA vs. Clermont	n.s.	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
MaRIA vs. London	$p = 0.002$	n.s.	n.s.	$p = 0.0086$
London vs. Clermont	$p = 0.001$	$p < 0.0001$	$p = 0.0007$	n.s.

Values are expressed as fractions and in percentages. Values in brackets are 95 % confidence intervals

AUROC area under receiver operation characteristic curve; n.s. non-significant



**Fig. 2** MRE from a young male with active severe Crohn's disease in the terminal ileum. T1-weighted sequence after gadolinium injection (a) depicts wall thickening (8 mm) of the terminal ileum as well as presence of marked enhancement (arrows). T2-weighted sequence with fat saturation on axial plane (b), showed a high signal intensity (edema) of the same segment together with an ulceration (arrow).

ROI measurement over this segment on apparent diffusion coefficient map (c) showed low ADC values. Marked wall thickening and enhancement, presence of edema, ulcerations and low ADC values are indicative of severe inflammation on MRE. Endoscopy examination of the terminal ileal (d) confirmed the presence of active disease with ulceration in the same segment

### Interobserver agreement

The intra-rater agreement expressed by ICC was 0.70 (substantial) for MaRIA index ( $p < 0.001$ ), 0.65 (substantial) for Clermont index ( $p < 0.001$ ) and 0.77 (substantial)

for London index ( $p < 0.001$ ). London index obtained a substantial agreement for classifying segments as active or inactive ( $K = 0.73$   $p < 0.001$ ), MaRIA a moderate agreement ( $k = 0.60$   $p < 0.001$ ) and Clermont also moderate ( $K = 0.41$   $p = 0.005$ ). For classifying segments as severe

**Table 3** Comparison of sensitivities, specificities, and accuracy of MaRIA and Clermont indices and ADC for detecting ulcerations at endoscopy

	Sensitivity	Specificity	Accuracy	AUROC
MaRIA	27/30 (90 % [79–100])	177/194 (91.2 % [87–95])	204/224 (91.1 % [87–95])	0.90 (0.86–0.94)
Clermont	25/30 (83.3 % [70–96])	173/194 (89.2 % [85–93])	198/224 (88.4 % [84–92])	0.86 (0.81–0.90)
ADC	30/30 (100 %)	23/194 (11.9 % [7–16])	53/224 (23.7 % [18–29])	0.56 (0.49–0.62)
<i>p</i> value				
MaRIA vs. Clermont	n.s.	n.s.	<i>p</i> = 0.03	<i>p</i> = 0.0527
MaRIA vs. ADC	n.s.	<i>p</i> < 0.0001	<i>p</i> < 0.0001	<i>p</i> < 0.0001
ADC vs. Clermont	n.s.	<i>p</i> < 0.0001	<i>p</i> < 0.0001	<i>p</i> < 0.0001

Values expressed as fractions and in percentages. Values in brackets are 95 % confidence intervals

ADC apparent diffusion coefficient (quantification of diffusion), AUROC area under receiver operation characteristic curve, n.s. non-significant

**Table 4** Sensitivities, specificities and area under ROC curve for detecting active lesions and severe lesion according to optimized cut-off values in our cohort

	Optimal cut-off	Sensitivity	Specificity	AUROC
Active lesions				
MaRIA	6.7	89.83 (79.2–96.2)	95.76 (91.5–98.3)	0.93
Clermont	9.44	84.75 (73.0–92.8)	98.18 (94.8–99.6)	0.94
London	1.79	81.36 (69.1–90.3)	95.15 (90.7–97.9)	0.90
ADC	1405	89.83 (79.2–96.2)	63.64 (55.8–71)	0.83
Severe lesions				
MaRIA	12	86.67 (69.3–96.2)	92.27 (87.6–95.6)	0.94
Clermont	14.3	83.33 (65.3–94.4)	92.27 (87.6–95.6)	0.92
London	3.13	83.33 (65.3–94.4)	89.69 (84.5–93.6)	0.90
ADC	1244	73.33 (54.1–87.7)	78.87 (72.4–84.4)	0.83

Values expressed as % (95 % CI)

AUROC area under ROC, ADC apparent diffusion coefficient

active MaRIA index had substantial interobserver agreement ( $K = 0.66$   $p < 0.001$ ) whereas Clermont had moderate agreement ( $K = 0.59$   $p < 0.001$ ).

**Detection of complications using different combinations of MRE sequences**

In the context of clinical trials, and also in clinical practice, detection of complications is relevant for patient selection, as well as for safety assessments. A total of 247 segments were evaluated using set 1 (T2 sequence and DWI sequences) and set 2 images (T1 and T2 sequences), each set being evaluated by a different blinded reader. Set 1 detected 27 segments with stenosis. One of these segments was not confirmed on Set 2 when gadolinium sequences were added to T2 sequences, all the other stenotic lesions were confirmed and no additional stenotic lesions were identified using T1 and T2 sequences.

Set 1 detected seven fistulas, set 2 confirmed all these, and diagnosed four additional fistulizing lesions. Finally, six inflammatory masses/abscesses were identified on set 1

images and one additional inflammatory mass was identified using set 2 of images, corresponding to a small pelvic inflammatory mass next to the right ovary that was interpreted as a normal ovary due to similar characteristics on T2 and DWI).

**Discussion**

The selection of the most appropriate index for identifying patients with active disease or severe inflammation as target population, and at the same time, excluding those who have complications that might be unlikely to respond (e.g., stenosis), or even may aggravate the clinical condition (e.g., abscesses), is key in increasing efficiency of clinical trials and to personalize therapy in clinical practice. Also, the validity of a particular index with reproducible cut-off points for detection of activity and severity is a key property for its acceptability in clinical trials [13].

Our study shows that MaRIA index appears to have the best operating characteristics with regard to its overall

accuracy, which further supports its implementation as the preferred instrument for use in clinical trials. One aspect of our results that deserves attention is an apparent discrepancy between the accuracy of each index for classifying segments as active or severe active using endoscopy as gold standard and the correlation analysis between the endoscopic index of activity (SES-CD) and the values of the three indices studied. Although correlation between quantitative indices is an important aspect, for the purpose of this study, we aimed to determine which index is the most accurate for correctly identify those intestinal segments with active or with severe active inflammation. Thus, our conclusions are based on the assumption that, as it was mentioned above, for clinical trials a binary diagnosis process is usually applied for including or excluding patients in clinical trials and for outcome measures.

Another important aspect to consider is the generalizability of our findings in different cohorts of patients or equipment. In that regard, when we reassessed the optimal cut-off points for diagnosis of active disease and severe disease, for the MaRIA index we found almost equal cut-off points than in previous cohorts [2, 11] whereas for the Clermont index the variations were up to 15 % (i.e., 14.4 vs. 12.5 for detecting severe lesions) and 45 % for the ADC compared to the originally suggested values. Furthermore, MaRIA has been validated in an independent study by the group of investigators that originally described the index, and more importantly, by other groups, in the context of multicenter studies [2, 14].

The main differential characteristic of Clermont index, compared with the other two indices, is the inclusion of the quantification of the DWI sequence, namely ADC. In our study, we could not reproduce the results that have been shown in a recent publication [11] by using the pre-defined cut-off point for the ADC for detection of ulcers. However, when calculating the optimal cut-off point in our cohort, an ADC of 1293 resulted in an optimal cut-off for detecting ulcers and led to a markedly improved accuracy. Our interpretation of this variability is that the measurement of ADC probably is an equipment-dependent metric value, or highly dependent on technical parameters, representing a major caveat for its applicability as a reliable tool in clinical practice and research. In addition, in our cohort, ADC values for differentiating active segments from those having ulcers showed considerable overlapping values. Differences between the cut-off values for defining active disease described on original publication of London index and our study can be explained because London index was derived using surgical specimens whereas in this study we used endoscopy as gold standard, plausibly including less severe lesions.

When applying different indices, one pre-requisite for each index to be considered useful evaluative instrument is

to determine the agreement and reproducibility between readers that represents an important factor in the performance of each score in multicenter studies. Overall, all three indices obtained a moderate agreement when were read by the two readers. However, London index was somewhat better than the other two for classifying segments as active or inactive. It could be controversial to determine whether the selection of an index should be based on the accuracy or on its reproducibility. MaRIA and London index had similar ICC but London has somewhat better agreement for classifying segments as active or inactive. However, considering the higher sensitivity and AUROC of MaRIA over London and the fact that London index is only able to classify segments as active or inactive, but not its severity, we think that MaRIA index has better overall properties for its use in trials and clinical practice.

We limited our analysis to the three indices derived in a structured process based on adequate reference standard. We did not include other indices in our analysis because of their inherent limitations such those derived based on expert opinion [15, 16], or those derived using fecal calprotectin as reference standard [17] which has the caveat of a low sensitivity for detecting small bowel inflammation.

Finally, our analysis focused on the detection of CD related complications that commonly represent a contraindication for including patients in clinical trials. The accuracy of clinical assessment for the diagnosis of these complications is low [18, 19], and endoscopy misses penetrating complications and, frequently, can only provide an incomplete assessment of stenosis. Both, penetrating and obstructive lesions represent independent predictors of risk for bowel resection surgery in patients with CD, even in subjects under biological therapy [20, 21]. The inclusion of gadolinium sequences identified 11 patients with fistulas, whereas only 7 were detected by T2 sequences. In addition, one small abscess next to the ovary was misinterpreted as a normal ovary using T2 sequences. Therefore, although the number of complication is low, our data suggest that gadolinium may increase the detection rate of penetrating complications, and thus increase safety in the context of immunosuppressive therapies. In that regard, our results are in keeping with those published recently by a Korean group, where sequences with and without gadolinium were discordant in 2/7 segments with penetrating complications [22].

Our study has some limitations. First of all, it is an exploratory retrospective study including all patients fulfilling inclusion criteria but a formal statistical power calculation was not performed. Second, we did not include a routine distension of the colon with colonic water enema that has been suggested to be helpful for the assessment of colonic segments in CD. Third, we focused on what MRE index could be the ideal tool to for patient selection into



drug trials but another aspect to be addressed in the future is whether some index is superior for monitoring drug efficacy.

Although ileocolonoscopy is generally considered the most accepted gold standard for assessing inflammatory lesions in CD patients, previous studies as well as the current study highlight important limitations [2, 14, 21]. Importantly, ten patients (23.25 %) had incomplete ileocolonoscopy due to severe stenosis or impossibility to intubate the terminal ileum but all of them were evaluated by MRE. Of them, 8/10 terminal ileum and 1/9 colonic segments were classified as severe inflammation by both MaRIA and Clermont index. That reinforces the perception that MRE could be used as a primary tool in clinical trials in the short or mid-term.

In summary, among different MRE indices, MaRIA appears to have the better overall operational characteristics for its use in clinical trials. Our data suggests that gadolinium sequences on MRE examinations are still necessary to rule out penetrating complications, which may improve patient selection and therapeutic responses.

**Acknowledgments** This work has been supported in part by a grant from Leona and Harry Helmsley Charitable Trust.

#### Compliance with ethical standards

**Conflict of interest** Jordi Rimola serves as a consultant for Genentech, TiGenics and Roberts; Julian Panes serves as a consultant for Abbvie, Boehringer-Ingelheim, Galapagos, Genentech, Janssen, Pfizer, Takeda and TiGenix.

## References

1. Van Assche G, Dignass A, Reinisch W, et al. The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: special situations. *J Crohns Colitis*. 2010;4:63–101. doi:10.1016/j.crohns.2009.09.009.
2. Coimbra AJF, Rimola J, O'Byrne S, 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.
3. Samuel S, Bruining DH, Loftus EV, et al. Endoscopic skipping of the distal terminal ileum in Crohn's disease can lead to negative results from ileocolonoscopy. *Clin Gastroenterol Hepatol*. 2012;10:1253–9. doi:10.1016/j.cgh.2012.03.026.
4. Rimola J, Ordás I, Rodríguez S, et al. Imaging indexes of activity and severity for Crohn's disease: current status and future trends. *Abdom Imaging*. 2012;37:958–66.
5. Rimola J, Rodríguez S, García-Bosch O, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. *Gut*. 2009;58:1113–20.
6. Rimola J, Ordás I, Rodríguez S, et al. Magnetic resonance imaging for evaluation of Crohn's disease. *Inflamm Bowel Dis*. 2011;17:1759–68. doi:10.1002/ibd.21551.
7. 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;81:2080–8.
8. Hordonneau C, Buisson A, Scanzi J, et al. Diffusion-weighted magnetic resonance imaging in ileocolonic Crohn's disease: validation of quantitative index of activity. *Am J Gastroenterol*. 2014;109:89–98. doi:10.1038/ajg.2013.385.
9. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet*. 1980;1:514.
10. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. *Gastrointest Endosc*. 2004;60:505–12.
11. Buisson A, Hordonneau C, Goutte M, et al. Diffusion-weighted magnetic resonance entero-colonography is highly effective to detect ileocolonic endoscopic ulcerations in Crohn's disease. *Aliment Pharmacol Ther*. 2015;42:452–60. doi:10.1111/apt.13287.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–74.
13. Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. 2003;56:826–32.
14. Takenaka K, Ohtsuka K, Kitazume Y, et al. Comparison of magnetic resonance and balloon enteroscopic disease. *Gastroenterology*. 2014;147:334–42.
15. 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 I. *Acad Radiol*. 2008;15:153–64.
16. Hyun SB, Kitazume Y, Nagahori M, et al. Magnetic resonance enterocolonography is useful for simultaneous evaluation of small and large intestinal lesions in Crohn's disease. *Inflamm Bowel Dis*. 2011;17:1063–72. doi:10.1002/ibd.21510.
17. Makanyanga JC, Pendsé D, Dikaios N, 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.
18. Higgins PDR, Caoili E, Zimmermann M, et al. Computed tomographic enterography adds information to clinical management in small bowel Crohn's disease. *Inflamm Bowel Dis*. 2007;13:262–8. doi:10.1002/ibd.20013.
19. García-Bosch O, Ordás I, Aceituno M, et al. Comparison of diagnostic accuracy and impact of MRI and colonoscopy for the management of Crohn's disease. *J Crohns Colitis*. 2016;. doi:10.1093/ecco-jcc/jjw015.
20. Thia KT, Sandborn WJ, Harmsen WS, et al. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. *Gastroenterology*. 2010;139:1147–55.
21. Jauregui-Amezaga A, Rimola J, Ordás I, et al. Value of endoscopy and MRI for predicting intestinal surgery in patients with Crohn's disease in the era of biologics. *Gut*. 2015;64:1397–402. doi:10.1136/gutjnl-2014-308101.
22. Seo N, Park SH, Kim KJ, Kang BK, Lee Y, Ye BD, Park SH, et al. MR enterography for the evaluation of small-bowel inflammation in Crohn disease by using diffusion-weighted imaging without intravenous contrast material: a prospective noninferiority study. *Radiology*. 2016;278:762–72.