Original Article

The new simplified MARIA score applies beyond clinical trials: A suitable clinical practice tool for Crohn's disease that parallels a simple endoscopic index and fecal calprotectin

Joana Roseira[1,2] 🄳, Ana Rita Ventosa[2,3],
Helena Tavares de Sousa[1,2] and Jorge Brito[2,3]

Abstract

Background: A simplified magnetic resonance enterography (MRe) index (sMARIA) for Crohn's disease (CD) was recently developed and validated.

Objective: Our aims were (a) to assess sMARIA's accuracy in a sample other than the validation cohort; (b) to evaluate its correlation with a simpler endoscopy index (SES-CD) and fecal calprotectin (FC); and (c) to assess the need of an expert radiologist to reliably use sMARIA.

Methods: Patients with CD who underwent MRe, ileocolonoscopy and FC within 2–4 weeks had their MRe retrospectively reviewed by two blinded raters. Disease activity was evaluated through sMARIA, SES-CD and FC. sMARIA's accuracy, indices correlation, and interrater reliability were assessed.

Results: In total, 84 patients were included, comprising 420 intestinal segments evaluations. sMARIA ≥ 1 accurately identified segments with active disease (90% sensitivity, 98% specificity; area under the curve 0.94, 95% confidence interval (CI) 0.91–0.97; $p < 0.01$). sMARIA correlated with endoscopy, both for ileal and colonic segments (R $= 0.94$ and R $= 0.82$; $p < 0.01$). Per patient, there was a strong correlation between sMARIA, endoscopy (R $= 0.95$; $p < 0.01$) and FC (R $= 0.91$; $p < 0.01$). Interrater agreement was excellent (intraclass correlation coefficient 0.95; 95% CI 0.94–0.96; $p < 0.01$).

Conclusion: sMARIA accurately measured CD activity using SES-CD as standard of reference, and exhibited high correlation with a simple endoscopic index and a biomarker. The interrater reliability between a radiology resident and an expert was excellent.

Introduction

Magnetic resonance enterography (MRe) has an established role in the assessment of Crohn's disease (CD), from diagnosis to treatment monitoring, detection of disease complications, and for uncovering postoperative recurrence.[1–3] Several indices for grading disease using MRe have been developed, the Magnetic Resonance Index of Activity for CD (MARIA) being the best-characterised.[4,5] As this index has some

[1]Gastroenterology Department, Centro Hospitalar Universitário do Algarve, Portimão, Portugal
[2]ABC – Algarve Biomedical Center, Portimão, Portugal
[3]Radiology Department, Centro Hospitalar Universitário do Algarve, Portimão, Portugal

Corresponding author:
Joana Roseira, Gastroenterology Department, Centro Hospitalar Universitário do Algarve, Estrada do Poço Seco, 8500-338 Portimão, Portugal.
Email: jsr_roseira@hotmail.com; twitter @RoseiraJoana

limitations, a newly simplified version of the score (sMARIA) was recently developed and validated.[6] sMARIA is a categorical 4-variable score, that was shown to accurately measure CD activity, severity and response to therapy. The same group further explored sMARIA, demonstrating a reproducible performance, even without gadolinium-enhanced sequences.[7] These studies used the Crohn's Disease Endoscopic Index of Severity (CDEIS) to address sMARIA's accuracy and endoscopy correlation.[6,7] Though CDEIS has proven to be reliable and reproducible,[8,9] this is a time-consuming index, making it unsuitable for everyday practice. Hence, a complementary analysis using a simpler endoscopic index is desirable. Furthermore, to prove sMARIA's role as a monitoring tool, a biomarker correlation analysis would be appropriate. On the other hand, an essential aspect of indices is reliability.[10] In the original sMARIA publication, one of the limitations pointed by Ordás et al.[6] was that only experienced radiologists were involved in sMARIA development.

Our study aimed to assess sMARIA diagnostic accuracy, in a sample other than the validation cohort, using the Simple Endoscopic Score for CD (SES-CD) as the reference standard. We also evaluated sMARIA's correlation and agreement with everyday practice instruments such as a simpler endoscopy index and fecal calprotectin (FC) biomarker. To evaluate the required training for an acceptable use of sMARIA, we calculated the inter-rater reliability between a radiology resident and a senior abdominal radiologist.

Material and methods

Study design

This was a retrospective, 5-year cohort study (March 2015–March 2020). Through the Radiology department registries, all MRe studies performed since March 2015 were signalled. Subsequently, MRe studies performed for CD evaluation were identified, and these patients' endoscopic data, clinical data, and biochemistry assays were reviewed. Adult patients diagnosed with ileal, colonic or ileocolonic CD,[11] followed at our tertiary hospital, who underwent MRe, ileocolonoscopy and FC within a 2–4 week timeframe were eligible. Ileocolonoscopy and FC were considered within a maximum of 2 weeks prior or 2 weeks after MRe. Patients treated with over 15 mg of prednisolone daily, those whose pharmacological therapy was modified within the examinations timeframe, and patients whose MRe evaluations were considered of low quality were excluded.

Eligible MRe studies were retrospectively and independently evaluated by an expert radiologist with 13 years of experience in abdominal radiology (JB),

and by a final-year radiology resident (ARV) differentiating in abdominal radiology who had previously completed MRe training.

Patient characteristics

Demographics (age, gender) and clinical variables (Montreal classification for CD, time of diagnosis, surgical history, ongoing therapy) were collected.

Procedure and operative details

Disease activity was evaluated through sMARIA, SES-CD and FC.

MRe analysis: sMARIA. All MRe examinations were performed in the supine position in a 1.5 T magnet (Magnetom Symphony Tim System, Siemens, Erlangen, Germany) equipped with a two phased-array-6-elements coils, according to the institution standard acquisition protocol (Supplementary data file). All patients received orally 1000–1500 ml of iso-osmotic polyethylene glycol solution 1 h prior to MRe.

Image analysis was independently performed by two raters – one experienced radiologist and one radiology resident – blinded to clinical and endoscopic aspects. To allow comparison with the endoscopic score, five segments were considered: terminal ileum; right colon; transverse colon; left colon; and rectum. As defined by Ordás et al.,[6] the sMARIA in each segment was calculated by the following formula: $1 \times$ thickening $>3 \, mm + 1 \times edema + 1 \times fat \, stranding + 2 \times ulcers$, these four variables being categorised as absent or present (Figure 1). Hence, sMARIA ranges from 0 to 5 per segment and global sMARIA ranges from 0 to 25. Image interpretation focused on T2 weighted sequences to identify all sMARIA variables (thickening, mural edema, fat stranding and ulcers).

Ileocolonoscopy: SES-CD. All ileocolonoscopies were performed under deep or superficial sedation, by a single experienced inflammatory bowel disease endoscopist (HTS), following the standard protocol, which includes reporting CD lesions according to the SES-CD. According to SES-CD, five bowel segments are considered (terminal ileum, right colon, transverse colon, left colon, and rectum), and for each segment, four variables are scored from 0 to 3, as described elsewhere.[12] SES-CD per patient results from adding the five segments' individual scores. This allows to calculate ileal and colonic SES-CD as two separate scores.[13] Thus, SES-CD ranges from 0 to 12 per segment, and global SES-CD ranges from 0 to 60.[12] SES-CD between 0 and 2 is considered as inactive disease, 3–6 as mildly active, 7–15 as moderately active, and ≥ 16 as
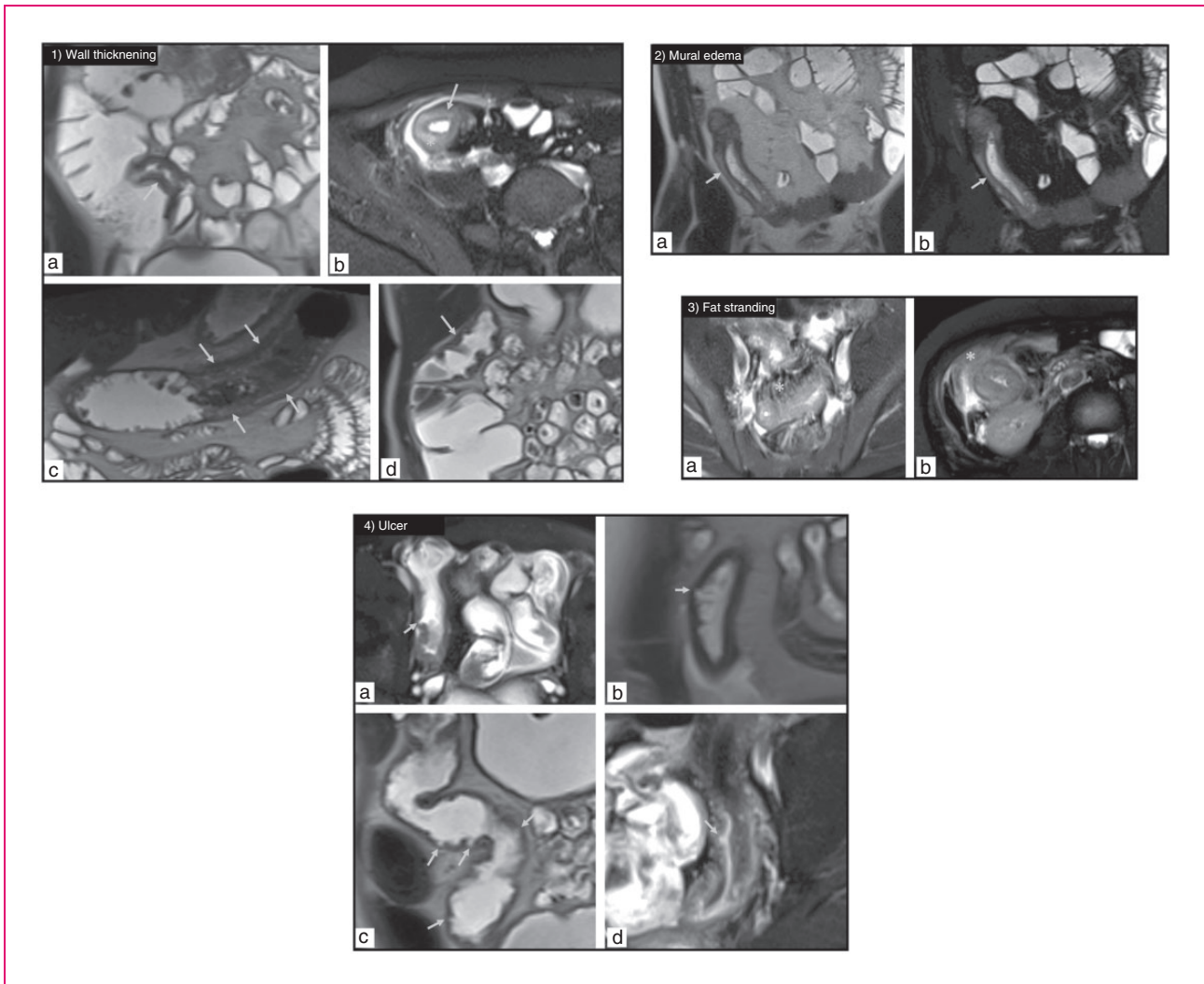
Figure 1. (1) Wall thickening – coronal T2-weighted (a) and axial T2-weighted fat suppressed (b) images show the terminal ileum with a thickened wall (arrow); coronal T2-weighted images depict wall thickening of the transverse colon (arrows in image (c)) and a slighter thickening of the ascending colon (arrow in image (d)). Note the intramural high signal intensity (star in image (b)) corresponding to inflammatory edema.
(2) Mural edema – coronal T2-weighted images without (a) and with fat suppression (b) show intramural high signal intensity in a thickened segment of terminal ileum (arrow), corresponding to inflammatory edema in the submucosal layer. Its presence in the fat-suppressed image assures it corresponds to edema.
(3) Fat stranding – axial T2-weighted fat suppressed images (a,b) show fat stranding (star) as a loss of the interface between the intestinal wall and the mesenteric (a) or mesocolic fat (b) due to the presence of edema and fluid. Note the thickening of the ileum (a) and the colon (b).
(4) Ulcers – axial T2-weighted fat-suppressed (a) and coronal T2-weighted (b) images depicting ulcers in the terminal ileum (arrows). Coronal T2-weighted image (c) shows ulcers in the transverse colon wall (arrows). Axial T2-weighted fat-suppressed image (d) depicts an ulcer in the sigmoid colon (arrow).

severely active disease.[14,15] In this study, patients were considered as having severe disease if global SES-CD was ≥16. In addition, a classification of severity on a segment basis was performed by considering the presence of severe lesions (ulcers).[6]

Concentration of FC. Sample extracts were performed at our hospital laboratory, and FC concentration (µg/g)

was then measured using a commercially available fluoroenzyme immunoassay (EliA Calprotectin). Peyrin-Biroulet et al.[16] underlined that multiple FC cut-off values have been described, but determining thresholds for fecal biomarkers to differentiate between different disease severities can be challenging. In our study, to prevent selection bias and systematic error, FC scoring was used as a continuous scale to describe disease activity.

Statistical considerations

All statistical analysis was performed using the Statistical Package for the Social Sciences, version 25.0 (IBM Corp., Armonk, New York, USA), and level of significance was established at 5%. An informal approach to test numerical variables normality was performed using the histogram evaluation, that showed a right-skewed distribution for all variables. Descriptive data were described as median and inter-quartile range for continuous variables, and categorical variables were summarised using absolute (n) and relative frequencies (%). Two decimals were used throughout the manuscript and three decimals were used to express *p*-values. Kruskall–Wallis test was used to compare different groups of subjects.

The senior abdominal radiologist reader set was used to evaluate sMARIA's diagnostic accuracy for the prediction of active disease and severe disease per segment, using the receiver operating characteristic (ROC) area under the curve (AUC), based on the endoscopic categorisation of active disease and ulcer observation. Analysis using the radiology resident reader set was also performed for comparison purposes.

Correlations between the sMARIA and SES-CD (per segment and per patient), and global sMARIA and FC were measured by the Spearman coefficient test. Kappa statistics test was applied for the evaluation of agreement between the sMARIA and endoscopy for binary classification of active and severe disease.

For the interrater reliability assessment, the intraclass correlation coefficient (ICC) was calculated to access total sMARIA per segment score agreement between the two raters. Sub-analysis was performed for each intestinal segment and for each sMARIA item using Kappa statistics test for each paired evaluation by the two raters.

Ethical considerations

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki, and received favourable opinion from the Algarve University Hospital Center Ethic Committee (30/11/2019). The authors retrospectively analysed data from March 2015 to March 2020. Thus, waiver of consent for this study was approved. All efforts were made to ensure confidentiality of the data.

Results

Patient characteristics

In total, 87 patients were eligible. Three MRe studies were excluded because of low quality (movement artifacts [*n* = 2]; suboptimal bowel distension [*n* = 1]).

Table 1. Patient demographics and clinical characteristics.

Characteristics	
Age, years; median (IQR)	27 (42–20)
Female/Male gender, *n* (%)	48 (57.14)/36 (42.86)
CD Montreal classification, *n* (%)	
Age at diagnosis	
A1 (under 16 years)	19 (22.62)
A2 (17–40 years)	63 (75.00)
A3 (over 40 years)	2 (2.38)
Disease location	
L1 (terminal ileum)	38 (45.24)
L2 (colon)	12 (14.29)
L3 (ileum plus colon)	34 (40.48)
Disease behaviour	
B1 (nonstricturing, nonpenetrating)	64 (76.19)
B2 (stricturing)	11 (13.09)
B3 (penetrating)	9 (10.71)
Disease duration, years; median (IQR)	7 (21–3)
Surgical history, *n* (%)	5 (5.95)
Treatment, *n* (%)	
No treatment	20 (23.81)
Steroids	2 (2.38)
Immunosuppressives	27 (32.14)
Biologics	35 (41.67)

IQR: interquartile range; CD: Crohn's disease.

Finally, 84 patients were included, comprising a total of 420 intestinal segments explored by MRe and endoscopy. Four patients had a history of ileocecal surgical resection prior to MRe. One patient had two ileal stricturoplasties that did not compromise segment evaluation. Table 1 presents patients' demographic and clinical characteristics.

Disease activity assessment

60 patients (71.43%) had active disease on endoscopy (SES-CD score ≥3), for a total of 85 segments with disease activity. As for severe disease, 12 patients (14.29%) had severe disease (SES-CD score ≥16), and severe lesions (ulcers) were identified in 28 segments. Endoscopic disease activity and severity were documented both for ileal (49 and 18 segments, respectively) and colonic segments (36 and 10 segments, respectively). Table 2 presents descriptive statistics on the sMARIA and FC assessments between patients with inactive, active or severe disease on endoscopy.

## Diagnostic accuracy of the sMARIA

To maximise the accuracy captured by the ROC curve, all points of the sMARIA score domain were dichotomously tested to identify the ideal cut-off point for the identification of intestinal segments with active and severe disease on endoscopy. sMARIA ≥1 accurately identified segments with active disease with 90%

Table 2. MRe and FC assessments according to disease activity on endoscopy.

Assessments, median (IQR)	Inactive	Active	Severe	*p*-value
sMARIA				
Per segment	0	2 (3–2)	3 (5–2)	0.000
Per patient	0	2 (3–1)	5 (8–3)	0.000
FC value (µg/g)				
Per patient	23 (11–60)	358 (102–497)	1001 (680–1819)	0.000

p-values based on Kruskal–Wallis test.

IQR: interquartile range; sMARIA: simplified Magnetic Resonance Index of Activity for Crohn's Disease; MRe: magnetic resonance enterography; FC: fecal calprotectin.
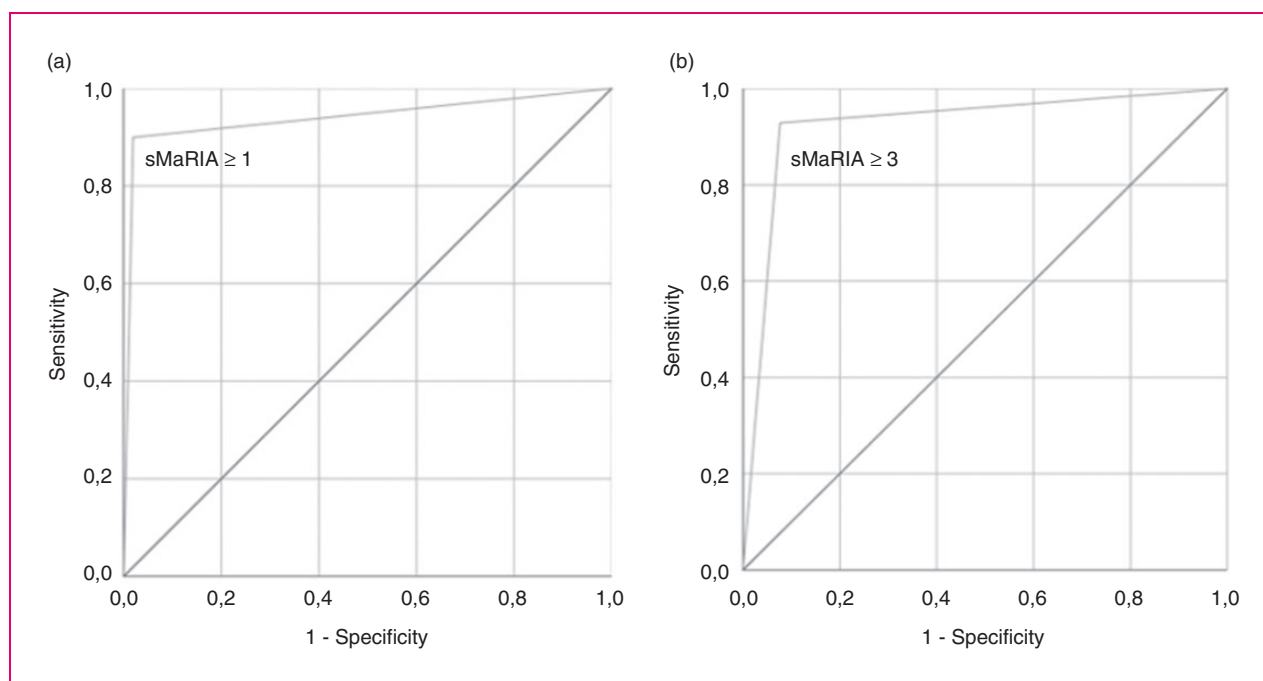
Figure 2. (a) ROC curve for the prediction of active disease; (b) ROC curve for the prediction of severe disease.

sensitivity (77 segments with sMARIA ≥ 1, out of 85 segments with SES-CD ≥ 3) and 98% specificity (AUC 0.94; 95% confidence interval [CI] 0.91–0.97; $p < 0.001$). As for severe disease, sMARIA ≥ 3 accurately identified segments with severe lesions (ulcers) with 93% sensitivity (26 segments with sMARIA ≥ 3 out of 28 segments with ulcers) and 92% specificity (AUC 0.93; 95% CI 0.89–0.97; $p < 0.001$) (Figure 2). sMaRIA's diagnostic accuracy separate sub-analysis for each intestinal segment is presented in Table 3. MRe overestimated severity in the rectum, and accuracy was not significant for the identification of severe disease in this segment. Still, the one patient with severe rectal lesions on endoscopy was adequately captured by a sMARIA ≥ 3. Also, endoscopic remission (SES-CD 0 to 2 per segment) was correctly identified by a sMARIA score <1 in 98.8% of cases (332 out of 336 segments with inactive disease).

This analysis was additionally performed for the radiology resident reader set, and accuracy was comparable between raters. sMARIA ≥ 1 and ≥ 3, rated by the radiology resident, accurately identified segments with active disease with 87% sensitivity and 98% specificity (AUC 0.92; 95% CI 0.88–0.97; $p < 0.001$) and severe disease with 89% sensitivity and 92% specificity (AUC 0.91; 95% CI 0.84–0.97; $p < 0.001$).

Correlation and agreement between scores

At the segment level, a significant correlation between sMARIA and SES-CD was observed for both ileal (R = 0.94; $p < 0.001$) and colonic segments (R = 0.82; $p < 0.001$). Also, there was an excellent correlation between global sMARIA and SES-CD (R = 0.95; $p < 0.001$) and between sMARIA and FC (R = 0.91; $p < 0.001$) per patient. The correlation between

Table 3. sMaRIA's diagnostic accuracy for the identification of active and severe disease on endoscopy, per intestinal segment.

sMARIA's diagnostic accuracy	AUC	CI 95%		p-value
		Lower limit	Upper limit	
Active disease on endoscopy				
Terminal ileum	0.96	0.92	0.99	0.000
Right colon	0.91	0.79	1	0.000
Transverse colon	0.93	0.84	1	0.000
Left colon	0.92	0.84	0.99	0.000
Rectum	0.89	0.76	1	0.000
Severe disease on endoscopy				
Terminal ileum	0.88	0.82	0.93	0.000
Right colon	0.94	0.83	1	0.000
Transverse colon	0.86	0.61	1	0.014
Left colon	0.95	0.91	0.99	0.002
Rectum	0.73	0.29	1	0.260

p-values based on the receiver operating characteristic (ROC) area under the curve (AUC) analysis.
CI: Confidence Interval; sMARIA: simplified Magnetic Resonance Index of Activity for Crohn's Disease.

sMARIA and FC remained equally strong when a separate analysis was performed for ileal, colonic and ileocolonic CD (R = 0.88, R = 0.92, R = 0.88, respectively; $p < 0.001$).

Per segment agreement between the sMARIA and endoscopy for binary classification of active disease was strong (K = 0.87; $p < 0.001$). Agreement for the identification of severe disease was moderate (K = 0.60; $p < 0.001$).

Interrater reliability assessment

Overall, interrater agreement between the radiology expert and the resident was excellent (ICC 0.95; 95% CI 0.94–0.96; $p < 0.001$). Sub-analysis per sMARIA variables, and per intestinal segment is detailed in Table 4. sMARIA variables were all found highly reliable between raters, and per intestinal segment analysis showed raters agreement proportions over 92%. Interrater agreement between the two readers was always superior for the terminal ileum (K = 0.88–1.0), and globally inferior, still moderate, for the transverse colon and rectum (K = 0.52–0.59).

Discussion

In this study, we conclusively showed sMARIA's high accuracy to predict CD activity and severity using SES-CD as the reference standard. In addition, we disclosed sMARIA's strong correlation with everyday practice tools such as a simple endoscopic score and a biomarker. Lastly, we demonstrated an overall, per item, and per segment sMARIA interrater adequate reliability, between an experienced radiologist and a resident.

This is the largest assessment on sMARIA's performance after its development study.[6] More recently, the

Table 4. Interrater agreement analysis between the two raters.

sMARIA items detection per segment	Kappa value	Agreement proportion	p-value
Thickening (>3 mm)	0.88	404/420	0.000
Terminal ileum	1	84/84	
Right colon	0.93	83/84	0.000
Transverse colon	0.69	80/84	0.000
Left colon	0.74	78/84	0.000
Rectum	0.59	79/84	0.000
Edema	0.87	404/420	0.000
Terminal ileum	0.95	82/84	0.000
Right colon	0.93	83/84	0.000
Transverse colon	0.58	80/84	0.000
Left colon	0.78	79/84	0.000
Rectum	0.64	80/84	0.000
Fat stranding	0.86	406/420	0.000
Terminal ileum	0.98	83/84	0.000
Right colon	0.84	82/84	0.000
Transverse colon	0.52	79/84	0.000
Left colon	0.79	80/84	0.000
Rectum	0.66	82/84	0.000
Ulcer	0.75	411/420	0.000
Terminal ileum	0.88	82/84	0.000
Right colon	0.55	81/84	0.000
Transverse colon	0.66	83/84	0.000
Left colon	0.66	82/84	0.000
Rectum	0.66	83/84	0.000

p-values based on Kappa statistics test.
sMARIA: simplified Magnetic Resonance Index of Activity for Crohn's Disease.

same group of authors addressed the score's performance without gadolinium-enhanced sequences in a 50-patient sample, reinforcing this instrument asset.[7] In addition, only a comment by Williet et al.[17] was published on this subject, confirming sMARIA's

correlation with the original MARIA, and the Clermont index. Williet et al. advocated that a sMARIA <1 could predict 'deep remission'. However, 'deep remission' was defined with no endoscopic evaluation. Thus, we believe that an independent external validation of the sMARIA was lacking.

To date, endoscopy and endoscopic scores remain the gold standard for the evaluation of CD.[18] Although CDEIS is considered the standard score, its calculation is complex and unsuitable for clinical practice.[13] SES-CD is simpler, significantly correlates with CDEIS,[12,19] and with MRe through the original MARIA score.[20–22] In our study, more than 400 intestinal segments were evaluated by ileocolonoscopy, with endoscopic assessment based on SES-CD. Our results conclusively demonstrated that sMARIA's diagnostic accuracy to identify disease remission, active disease and severe disease was excellent, using SES-CD as the reference index. Ordás et al.[6] showed that sMARIA ≥ 1 and ≥ 2 were the best cut-off values to identify active and severe disease, respectively, using CDEIS. Similarly, we demonstrated a sMARIA ≥ 1 and ≥ 3 as the optimal cut-off values to identify active and severe disease, using SES-CD. This difference in the cut-off value for identification of severe inflammation on MRe may be explained by a low number of segments with severe lesions ($n = 28$) in our study population.

The major driver for the sMARIA development was the demand for a quick and easy imaging index that could be generalised for clinical practice, without compromising its original accuracy.[6] Indeed, there is a growing need to replace invasive tools with non-invasive alternatives,[1] which should correlate with the endoscopy gold standard, intercorrelate themselves, and add to one another. Endoscopic skipping phenomenon requiring a beyond-the-mucosa assessment and FC elevation constitute examples of this concept. FC appears to be the most sensitive biomarker of intestinal inflammation in inflammatory bowel disease, but it provides no information on disease phenotype or complications.[1] In these scenarios, a MRe evaluation would be useful and complementary. Original MARIA and FC correlation was demonstrated for colonic CD.[23] To prove sMARIA's role as a clinical practice monitoring tool, a biomarker correlation was lacking. Our data demonstrated an excellent correlation between sMARIA and FC for ileal, colonic and ileocolonic CD.

Lastly, we explored sMARIA's interrater reliability between an experienced abdominal radiologist and a radiology resident in her last year of training. The original authors pointed out that sMARIA could require much training, and that this could limit the score's availability.[6] Thereupon, the same authors partly addressed this, showing a moderate to excellent agreement between a less-experienced radiologist and an expert.[7] Further ahead, we demonstrated a moderately good to perfect, per variable and per segment, interrater agreement between an experienced radiologist and a resident. The agreement was lower in distal colonic segments, probably due to suboptimal distension compared with the more proximal segments, as previously reported.[24] This may also explain sMARIA's lower diagnostic accuracy in the rectum.

Some limitations should be noted in this study. Its design did not allow the control of assessment timepoints, and this may limit the examination's association precision. To overcome this and prevent significant information bias, the interval between paired MRe and endoscopy and paired MRe and FC was considered only if within 2 weeks, as previously reported,[23] and patients with therapeutic interventions during this timeframe were excluded. Ideally, two senior radiologists' readings should have been included for diagnostic accuracy analysis. Finally, radiologists were not blinded to gadolinium-enhanced sequences, as this was a retrospective study and its use is currently recommended in guidelines.[1]

However, our study has a few strengths that should be highlighted. It encompasses a robust investigation, comprising more than 400 intestinal segments evaluated by MRe and ileocolonoscopy. It represents an independent sample and author contribution, as required for an unbiased external validation.[25,26] In addition, to the authors' knowledge, this is the first study validating sMARIA using everyday tools such as a simple endoscopic index and a biomarker. Also, we further analysed the score's reliability, evaluating interrater agreement between an expert radiologist and a resident. Finally, unlike the original authors,[6,7] luminal colonic contrast administration by a rectal catheter was not performed. As seldom reported,[22,27,28] this study shows an accurate colonic assessment may be performed using only oral luminal contrast.

In conclusion, our study demonstrated sMARIA to accurately predict CD activity using SES-CD as a reference and to correlate with a simple endoscopic index and a biomarker. Moreover, the interrater reliability between a radiology resident and an expert was excellent, supporting sMARIA as a suitable clinical practice instrument.

Ethics approval

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki, and received favourable opinion from the Algarve University Hospital Center Ethic Committee (30/11/2019).

Informed Consent

The authors retrospectively analysed data from March 2015 to March 2020. Thus, waiver of consent for this study was

approved. All efforts were made to ensure confidentiality of the data.

Declaration of conflicting interests

Funding

ORCID iD

Joana Roseira (ID) https://orcid.org/0000-0002-5098-8729

Supplemental material

Supplemental material for this article is available online.

References

1. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohn's Colitis* 2018; 13: 144–164K.

2. Sousa HT de, Brito J, Magro F. New cross-sectional imaging in IBD. *Curr Opin Gastroenterol* 2018; 34: 194–207.

3. Nehra AK, Sheedy SP, Wells ML, et al. Imaging findings of ileal inflammation at computed tomography and magnetic resonance enterography: What do they mean when ileoscopy and biopsy are negative? *J Crohns Colitis*. Epub ahead of print January 2020. DOI: 10.1093/ecco-jcc/jjz122.

4. Bernklev T, Moum B and Moum T. Quality of life in patients with inflammatory bowel disease: Translation, data quality, scaling assumptions, validity, reliability and sensitivity to change of the Norwegian version of IBDQ. *Scand J Gastroenterol* 2002; 37: 1164–1174.

5. Gower-Rousseau C, Sarter H, Savoye G, et al. Validation of the Inflammatory Bowel Disease Disability Index in a population-based cohort. *Gut* 2017; 66: 588–596.

6. Ordás I, Rimola J, Alfaro I, et al. Development and validation of a Simplified Magnetic Resonance Index of Activity for Crohn's Disease. *Gastroenterology* 2019; 157: 432–439.e1.

7. Capozzi N, Ordás I, Fernandez-Clotet A, et al. Validation of the Simplified Magnetic Resonance Index of Activity [sMARIA] without gadolinium-enhanced sequences for Crohn's Disease. *J Crohn's Colitis*. Epub ahead of print 2020. DOI: 10.1093/ecco-jcc/jjaa030.

8. Cellier C, Sahmoud T, Froguel E, et al. Correlations between clinical activity, endoscopic severity, and biological parameters in colonic or ileocolonic Crohn's disease. A prospective multicentre study of 121 cases. The Groupe d'Etudes Therapeutiques des Affections Inflammatoires Digestives. *Gut* 1994; 35: 231–235.

9. Modigliani R, Mary JY, Simon JF, et al. Clinical, biological, and endoscopic picture of attacks of Crohn's disease. Evolution on prednisolone. Groupe d'Etude Therapeutique des Affections Inflammatoires Digestives. *Gastroenterology* 1990; 98: 811–818.

10. Lange RT. Inter-rater Reliability. In: Kreutzer JS, DeLuca J, Caplan B (eds) *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer New York, p. 1348.

11. Satsangi J, Silverberg MS, Vermeire S, et al. The Montreal classification of inflammatory bowel disease: Controversies, consensus, and implications. *Gut* 2006; 55: 749–753.

12. 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: 505–512.

13. Koutroumpakis E and Katsanos KH. Implementation of the simple endoscopic activity score in Crohn's disease. *Saudi J Gastroenterol* 2016; 22: 183–191.

14. Algaba A, Linares PM, Fernandez-Contreras ME, et al. Relationship between levels of angiogenic and lymphangiogenic factors and the endoscopic, histological and clinical activity, and acute-phase reactants in patients with inflammatory bowel disease. *J Crohns Colitis* 2013; 7: e569–e579.

15. Sipponen T, Nuutinen H, Turunen U, et al. Endoscopic evaluation of Crohn's disease activity: Comparison of the CDEIS and the SES-CD. *Inflamm Bowel Dis* 2010; 16: 2131–2136.

16. Peyrin-Biroulet L, Panés J, Sandborn WJ, et al. Defining disease severity in inflammatory bowel diseases: Current and future directions. *Clin Gastroenterol Hepatol* 2016; 14: 348–354.e17.

17. Williet N, Jardin S and Roblin X. The Simplified Magnetic Resonance Index of Activity (MARIA) for Crohn's disease is strongly correlated with the MARIA and Clermont Score: An external validation. *Gastroenterology* 2020; 158: 282–283.

18. Papay P, Ignjatovic A, Karmiris K, et al. Optimising monitoring in the management of Crohn's disease: A physician's perspective. *J Crohns Colitis* 2013; 7: 653–669.

19. Daperno M, Comberlato M, Bossa F, et al. Inter-observer agreement in endoscopic scoring systems: Preliminary report of an ongoing study from the Italian Group for Inflammatory Bowel Disease (IG-IBD). *Dig Liver Dis* 2014; 46: 969–973.

20. 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.

21. Stoppino LP, Della Valle N, Rizzi S, et al. Magnetic resonance enterography changes after antibody to tumor necrosis factor (anti-TNF) alpha therapy in Crohn's disease: Correlation with SES-CD and clinical-biological markers. *BMC Med Imaging* 2016; 16: 1–9.

22. Takenaka K, Ohtsuka K, Kitazume Y, et al. Correlation of the endoscopic and magnetic resonance scoring

systems in the deep small intestine in Crohn's disease. *Inflamm Bowel Dis* 2015; 21: 1832–1838.

23. Somwaru AS, Khanijow V and Katabathina VS. Magnetic resonance enterography, colonoscopy, and fecal calprotectin correlate in colonic Crohn's disease. *BMC Gastroenterol* 2019; 19: 1–9.

24. Jairath V, Ordas I, Zou G, et al. Reliability of measuring ileo-colonic disease activity in Crohn's disease by magnetic resonance enterography. *Inflamm Bowel Dis* 2018; 24: 440–449.

25. Dyrvig AK, Kidholm K, Gerke O, et al. Checklists for external validity: A systematic review. *J Eval Clin Pract* 2014; 20: 857–864.

26. Collins GS, Ogundimu EO and Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. *Stat Med* 2016; 35: 214–226.

27. 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.

28. Rimola J, Alvarez-Cofino A, Perez-Jeldres T, et al. Comparison of three magnetic resonance enterography indices for grading activity in Crohn's disease. *J Gastroenterol* 2017; 52: 585–593.