

Original Article

Intestinal Ultrasound to Assess Disease Activity in Ulcerative Colitis: Development of a novel UC-Ultrasound Index

Steven Bots,^a Kim Nylund,^{b,c} Mark Löwenberg,^a Krisztina Gecse,^a Geert D'Haens^a

^aDepartment of Gastroenterology and Hepatology, Amsterdam University Medical Center, location AMC, The Netherlands ^bNational Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway ^cDepartment of Clinical Medicine, University of Bergen, Bergen, Norway

Corresponding author: Prof. Dr G. D'Haens, Department of Gastroenterology and Hepatology, Amsterdam University Medical Center, location AMC, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. Email: g.dhaens@amsterdamumc.nl

Abstract

Introduction: Intestinal ultrasound [IUS] is useful to assess inflammation in ulcerative colitis [UC] patients. We aimed to develop an ultrasonographic activity index using endoscopy as the reference standard.

Methods: Patients were included consecutively. IUS was performed within 3 weeks from endoscopy. IUS parameters and endoscopy were compared for each colonic segment [except the rectum]. The best parameters were used to construct a UC-IUS index, which was correlated with endoscopic disease activity using the Spearman's rank test.

Results: In 60 patients, 207 colonic segments were evaluated endoscopically. Bowel wall thickness [BWT] > 2.1 mm was optimal to discriminate between Mayo 0 and Mayo 1–3 (sensitivity 82.6%; specificity 93.0%; area under the curve [AUC] 0.910), a cut-off of 3.2 mm was optimal to discriminate between Mayo 0–1 and Mayo 2–3 [sensitivity 89.1%; specificity 92.3%; AUC 0.946] and BWT > 3.9 mm was optimal for detection of Mayo 3 [sensitivity 80.6%; specificity 84.1%; AUC 0.909]. The presence of colour Doppler signal [CDS] predicted active disease, stretches of CDS were associated with Mayo 2–3, lack of haustrations predicted active disease and fat wrapping was associated with severe disease. Inter- and intra-rater intraclass correlation for BWT was substantial. Inter-rater agreement for CDS was substantial and ranged from slight to substantial for haustrations. Intra-rater agreement for CDS was substantial and ranged from moderate to almost perfect for haustrations. The index showed strong correlation with endoscopic disease activity [Mayo: ρ 0.830; $p < 0.001$, UCEIS: ρ 0.759; $p < 0.001$].

Conclusion: We developed an UC-IUS index which showed strong correlation with endoscopic disease activity using internal validation. It is currently being validated in prospective studies.

Key Words: Intestinal ultrasound, inflammatory bowel disease, ulcerative colitis.

1. Introduction

Ulcerative colitis [UC] is a chronic inflammatory bowel disease characterized by relapsing and remitting episodes of inflammation

usually limited to the mucosal layer of the colon. Treatment targets for UC patients nowadays include patient-reported as well as endoscopic remission. Recently, endoscopy is increasingly being used to

guide treatment, because evidence suggests that mucosal healing [e.g. Mayo 0–1 activity] is associated with improved long-term outcomes.^{1,2} However, it is challenging to repeatedly perform colonoscopies to assess mucosal disease activity due to the high cost and burden for the patient.³ Hence, alternative and reliable non-invasive methods to assess disease activity are needed.

Blood tests such as the measurement of serum C-reactive protein [CRP], albumin and platelet counts have been evaluated, but these tests are not sufficiently sensitive or specific to reflect disease activity.^{4–7} Repeated measurement of faecal calprotectin [FCP] has been shown to accurately reflect the presence of disease activity.^{4,8} However, disease location, extent and severity cannot be adequately assessed with this technique. Intestinal ultrasound [IUS] is a rapid, efficient, non-invasive and relatively cheap imaging technique, which can also be performed in point-of-care settings. IUS has been reported to be accurate in the diagnosis of UC and can also be applied to determine the extent, severity and location of inflammation.^{9–12}

Therefore, IUS is an attractive tool for the assessment of disease activity in UC patients. So far, few studies have been performed to compare IUS with endoscopy.^{11,13–16} Additionally, studies evaluating responsiveness of IUS to a medication with known efficacy, validation to endoscopy and evaluating reliability are scarce.¹¹ In a previous systematic review, we showed that, although several IUS indices have been developed for the assessment of disease activity in UC patients, the methodology was suboptimal in most studies.⁹

Therefore, we aimed to develop an ultrasound activity index for the assessment of disease activity in patients with UC, using endoscopy as the reference standard.

2. Methods

2.1. Patient population

Adult UC patients undergoing endoscopy for evaluation of disease activity or surveillance were eligible for inclusion. Patients were consecutively included based on clinical Mayo score and Simple Clinical Colitis Activity Index [SCCAI, i.e. 20 quiescent disease, 40 active disease]. Patients underwent endoscopy and IUS along with FCP and serum CRP measurement within the shortest period possible with a maximum window of 3 weeks. If there was a change in treatment or symptoms between IUS and endoscopy, patients were excluded. IUS and endoscopy were not performed on the same day. At the time of endoscopy, the performing endoscopist was unaware of the IUS results and vice versa. Endoscopists and ultrasonographers were not blinded for clinical symptoms as would also be the case in a real-life clinical setting.

2.2. Ultrasound examinations

All the IUS examinations were performed by one of two investigators experienced in IUS [S.B. 3 years and K.N. 9 years of experience], with a Philips Epiq 5 ultrasound device using the C5-1 convex transducer and L12-5 linear transducer. Frequency, focus and gain settings were optimized to get the best images. The examination was performed after at least 4 h of fasting with the patient in the supine position. The large intestine was scanned beginning at the terminal ileum and further following its course to the rectum. The nine regions of the abdomen were also systematically scanned for the detection of enlarged lymph nodes and other possible pathology. Each colon segment scanned in B-mode was also examined with colour Doppler. The colour Doppler measurements were performed with standardized pre-sets with optimized wall filter, pulse repetition frequency, frequency, and velocity scale for registration of the slow flow in the gastrointestinal wall. Cine loops of each segment

in longitudinal sections were video-recorded in B-mode and colour Doppler mode.

2.3. Ultrasound parameters and measurements

The following IUS parameters were recorded during the procedure: bowel wall thickness [BWT], colour Doppler signal [CDS], image quality, normal or abnormal colonic haustrations, presence of fat wrapping [hyperechoic fat around the bowel], wall layer stratification [WLS], and presence of enlarged lymph nodes [short axis > 5 mm]. BWT was measured from, but not including, the central hyperechoic line of the lumen to the end of the outer hypoechoic margin of the wall [representing the muscularis propria]. All BWT measurements were performed in duplicate on longitudinal sections because it is easiest to notice the thickest wall section in longitudinal direction. CDS was divided into three categories: absent/single vessel [categorized as absent], spots or stretches of CDS. Image quality was categorized as good, average, low or uninterpretable. Assessment of image quality was based on the opinion of the ultrasonographer, as there is no validated index for this purpose. Normal colonic haustrations were defined as clearly visible haustrations or collapsed colonic folds with BWT < 2 mm. Abnormal colonic haustrations were defined as a clearly disrupted or tube-like in appearance. All measurements and image interpretations were performed by two observers [S.B. and K.N.] on the same cine loops to assess inter-rater variability. Additionally, 20 cases [five for each severity category] were randomly selected, re-anonymized and scrambled for a second interpretation by both observers to assess intra-rater variability. The time between first and second read was at least 3 months.

2.4. Endoscopy

Colonoscopy or sigmoidoscopy were performed according to standard procedures at our clinic by IBD experts. Endoscopic disease activity was scored using the Ulcerative Colitis Endoscopic Index of Severity [UCEIS] and the Mayo endoscopic sub-score for each segment.^{17,18} A Mayo endoscopic subscore of 1, 2 or 3 was considered as mild, moderate or severe disease, respectively. A UCEIS score of 4–5, 6–8 and 9–11 was considered as mild, moderate or severe disease, respectively.

2.5. Biomarkers

Blood samples were collected and analysed for CRP [mg/L] and stool samples collected were and analysed for FCP [$\mu\text{g/g}$] [Bühlmann fCal ELISA]. The upper limit of detection of the FCP test was 1800 $\mu\text{g/g}$. Samples were collected within 3 weeks before or after IUS as long as there was no change in treatment or clinical symptoms.

2.6. Clinical assessments

Medical history was assessed and information on the duration of ulcerative colitis, medical treatment, age, gender, weight, height and body mass index [BMI] was collected. At the IUS visit, symptom severity was scored using the SCCAI and Mayo score.^{17,19}

2.7. Sample size calculation

The sample size calculation was based on mean BWT in two patient groups, representing UC patients in remission [Group 1] and UC patients with active endoscopic disease [Group 2]. Based on literature data we assumed a colon wall thickness in healthy controls of a mean 1.1 mm [SD 0.3] and the cut-off of between normal and abnormal thickness of 2.0 mm.^{9,20} The colon wall in a heterogeneous group of UC patients with active disease was assumed to have a mean thickness of 4.5 mm [SD 1.3].^{9,11,15} For the sample size calculation this

resulted in a sample size of 20 in each group, which would offer 80% power to detect a difference in means of 0.9 mm assuming that the Group 1 standard deviation is 0.3 and the Group 2 standard deviation is 1.3 using a two-group Satterthwaite *t*-test with a 0.05 two-sided significance level. As we intended to study patients in multiple categories of disease activity, we intended to include 20 patients with quiescent disease and 40 with active disease.

2.8. Statistical analysis

Descriptive statistics were used to characterize the population. BWT, CDS, fat wrapping, WLS, haustration pattern and enlarged lymph nodes were compared with endoscopic findings for each segment except for the rectum. Normally distributed parameters were compared with unpaired *t*-tests. Categorical parameters were compared with logistic regression. Receiver operating characteristic [ROC] analysis was performed for BWT to determine optimal cut-offs. The most predictive parameters and cut-off values were used to construct a point-based UC-IUS index. The results obtained from the person actually performing IUS were used for this purpose. The index was

Table 1. Patient characteristics

Characteristic	<i>n</i> = 60
Male gender	28 [47%]
Age, years [median, IQR]	44 [30–54]
Height, cm [mean, SD]	176.4 [10.0]
BMI [mean, SD]	24.1 [3.2]
Medication use	
5-ASA	36 [60%]
Corticosteroids [oral/topical]	25 [42%]
Thiopurines	7 [12%]
Anti-TNF	8 [13%]
Vedolizumab	1 [2%]
Tofacitinib	3 [5%]
Tacrolimus [topical]	1 [2%]
Endoscopy results	
Mayo 0	16
Mayo 1	11
Mayo 2	15
Mayo 3	18
UCEIS < 4	15 [25%]
UCEIS 4–5	13 [22%]
UCEIS 6–8	15 [25%]
UCEIS 9–11	17 [28%]
Segments endoscopically explored	
Rectum	60 [100%]
Sigmoid	60 [100%]
Descending	58 [97%]
Transverse	49 [82%]
Ascending	40 [67%]
Total segments explored [excl. rectum]	207
Proctitis only	6 [10%]

BMI, body mass index; IQR, interquartile range; SD, standard deviation; UCEIS, ulcerative colitis endoscopic index of severity.

Table 2. Ultrasound image quality per segment

	Good	Average	Low	Uninterpretable
Rectum	12 [20.0%]	17 [28.3%]	23 [38.3%]	8 [13.3%]
Sigmoid	49 [81.7%]	10 [16.7%]	0 [0%]	1 [1.7%]
Descending	48 [80.0%]	11 [18.3%]	0 [0%]	1 [1.7%]
Transverse	45 [75.0%]	13 [21.7%]	2 [3.3%]	0 [0%]
Ascending	43 [71.7%]	15 [25.0%]	2 [3.3%]	0 [0%]

calculated for each patient and compared with the Mayo score and UCEIS score for each segment using the Spearman's rank correlation test. A value of 0.00–0.10 was considered as negligible correlation, 0.10–0.39 as weak correlation, 0.40–0.69 as moderate correlation, 0.70–0.89 as strong correlation and 0.90–1.00 as very strong correlation.²¹ Inter- and intra-rater agreement for categorical data was tested with Cohen's kappa statistics. A value of 0.0–0.20 was considered as slight agreement, 0.21–0.4 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement and 0.81–1.0 as almost perfect agreement.^{22,23} Inter- and intra-rater agreement for continuous BWT measurements was tested using intra-class correlation [ICC] statistics for average measurements. An ICC value of less than 0.50 was considered as poor agreement, a value of 0.50–0.75 as moderate agreement, a value of 0.75–0.90 as substantial agreement and a value of 0.90–1.00 as almost perfect agreement.²⁴ A *p*-value of < 0.05 was considered to be statistically significant. All statistical analyses were performed using SPSS 25.0 software [IBM].

2.9. Ethical approval and patient consent

This study was approved by the ethical committee of the Academic University Medical Center Amsterdam. All patients provided written informed consent prior to participation in this study.

3. Results

3.1. Patient population

A total of 60 UC patients were included. Patient characteristics are shown in Table 1. Sixteen patients were in complete endoscopic remission [Mayo 0] and 44 patients had active endoscopic disease [13 Mayo 1, 15 Mayo 2 and 18 Mayo 3]. Six patients had active proctitis only. In total, 207 colonic segments were explored at endoscopy [60, 58, 49 and 40 in sigmoid, descending, transverse and ascending colon, respectively]. IUS was performed within a median of 7 days (interquartile range [IQR] 5–11 days) from endoscopy, without change in treatment or symptoms in between. FCP samples were collected within a median of 2 days [IQR 0–4 days] from IUS.

3.2. Ultrasound

3.2.1. Image quality

Image quality for different colonic segments is shown in Table 2. Image quality in the rectum was average or higher in only 48.3% of patients. In 38.3% of patients image quality was low and in 13.3% the images were considered uninterpretable. Image quality was considered average or higher in 98.3% in the sigmoid and descending colon and 96.7% in the transverse and ascending colon.

3.2.2. BWT

Mean BWT was statistically different between Mayo 0 and Mayo 1 endoscopic activity [*p* < 0.001] and between Mayo 1 and Mayo 2 endoscopic activity [*p* < 0.001], but not between Mayo 2 and Mayo 3

[$p = 0.548$] [Figure 1]. A BWT cut-off of 2.1 mm was best to discriminate between inactive and active endoscopic disease activity [Mayo 0 vs Mayo 1–3] (sensitivity 82.6%; specificity 93.0%; area under the curve [AUC] 0.910). A BWT cut-off of 3.2 mm was best to discriminate between Mayo 0–1 and Mayo 2–3 endoscopic disease activity [sensitivity 89.1%; specificity 92.3%; AUC 0.946]. A BWT cut-off of 3.9 mm was best to discriminate between Mayo 0–2 and Mayo 3 endoscopic disease activity [sensitivity 80.6%; specificity 84.1%; AUC 0.909]. ROC curves are shown in Figure 3. For individual segments, a BWT cut-off of 2.1 mm was best to discriminate between Mayo 0 and Mayo 1–3 in the sigmoid [sensitivity 88.6%; specificity

88.0%; AUC 0.913], 2.5 mm in the descending [sensitivity 85.2%; specificity 87.1%; AUC 0.907], 1.75 mm in the transverse [sensitivity 88.9%; specificity 90.3%; AUC 0.944] and 2.6 mm in the ascending colon [sensitivity 75%; specificity 100%; AUC 0.903]. The mean difference in BWT in all segments for the two observers was 0.4 mm [SD 0.9; $p < 0.001$]. The inter-rater agreement for continuous BWT measurements was almost perfect (ICC 0.917; 95% confidence interval [CI] 0.853–0.948; $p < 0.001$). The intra-rater agreement for continuous BWT measurements was substantial [ICC 0.802; 95% CI 0.729–0.855; $p < 0.001$]. Based on the ROC cut-off points, the following categories for BWT were made: < 2 , 2.0–2.9, 3.0–3.9 and ≥ 4 mm. Sensitivity and specificity values were, respectively, 82.6% and 90% for 2 mm, 89.1% and 90.9% for 3 mm and 77.4% and 85.0% for 4 mm. Inter-rater agreement for these categories was moderate for the sigmoid [κ 0.53; $p < 0.001$], descending [κ 0.58; $p < 0.001$], transverse [κ 0.55; $p < 0.001$] and ascending colon [κ 0.43; $p < 0.001$]. Intra-rater agreement was substantial for the sigmoid [κ 0.68; $p < 0.001$], transverse [κ 0.63; $p < 0.001$] and ascending [κ 0.68; $p < 0.001$] colon and moderate for the descending colon [κ 0.59; $p < 0.001$]. The rectum was excluded from this analysis.

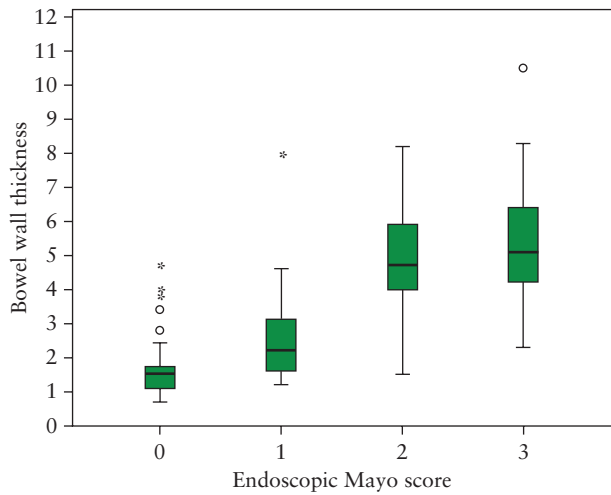


Figure 1. Mean bowel wall thickness for different Mayo scores

3.2.3. Colour Doppler signal

Examples of different CDS categories are shown in Figure 2. The presence of any CDS was associated with the presence of endoscopic disease activity (odds ratio [OR] 14.0; 95% CI 6.8–28.7; $p < 0.001$). The presence of any CDS was also associated with moderate to severe endoscopic activity as compared to mild or quiescent endoscopic activity [OR 14.9; 95% CI 7.3–30.4; $p < 0.001$] and stretches of CDS was more strongly associated with moderate to severe endoscopic activity [OR 22.3; 95% CI 7.3–67.8; $p < 0.001$]. The presence of

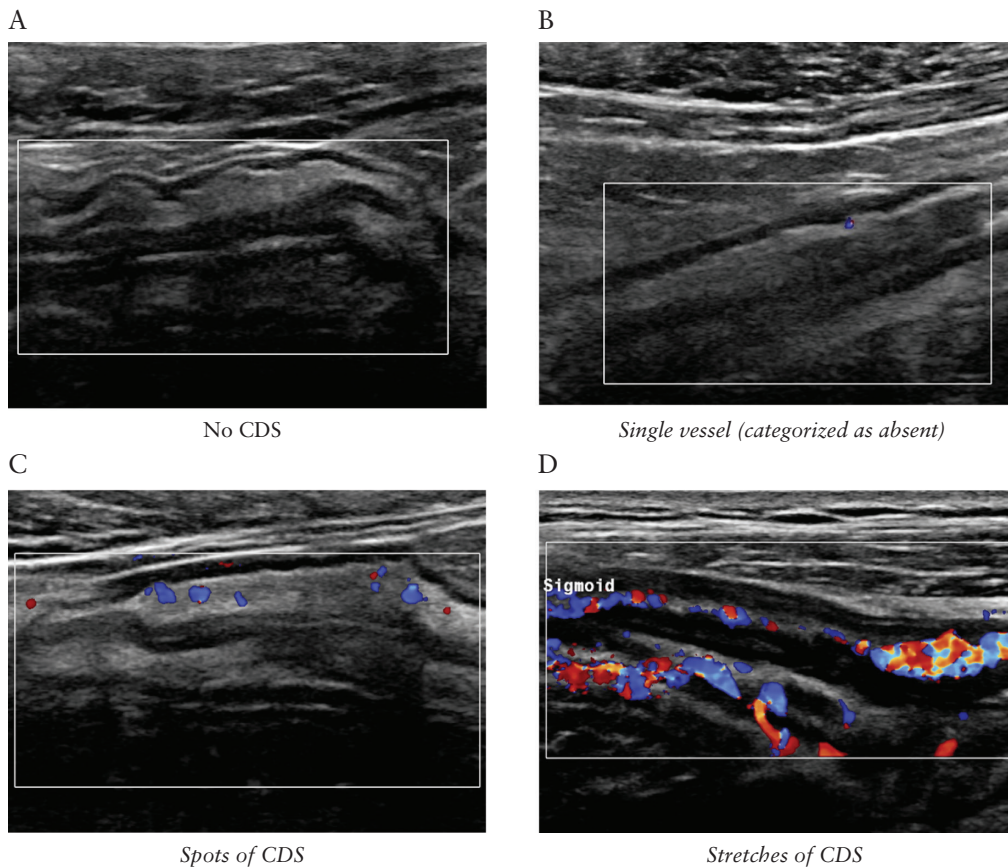


Figure 2. Categories of colour Doppler signal [CDS]. [A] no CDS; [B] single vessel [categorized as absent]; [C] spots of CDS; [D] stretches of CDS.

stretches of CDS also discriminated between moderate and severe endoscopic activity [OR 7.2; 95% CI 3.0–17.4; $p < 0.001$]. Inter-rater agreement was substantial for the sigmoid [κ 0.79; $p < 0.001$], descending [κ 0.78; $p < 0.001$], transverse [κ 0.75; $p < 0.001$] and ascending [κ 0.60; $p < 0.001$] colon. Intra-rater agreement was substantial for the sigmoid [κ 0.78; $p < 0.001$], descending [κ 0.69; $p < 0.001$], transverse [κ 0.60; $p < 0.001$] and ascending [κ 0.65; $p < 0.001$] colon.

3.2.4. Haustrations

Examples of normal and abnormal haustrations are shown in Figure 4. An abnormal haustration pattern was strongly associated with active endoscopic disease [OR 126.2; 95% CI 36.3–438.7; $p < 0.001$]. It was also associated, albeit to a lesser extent, with moderate to severe endoscopic disease [OR 100.7; 95% CI 35.0–290.1; $p < 0.001$]. Inter-rater agreement was substantial for the sigmoid [κ 0.69; $p < 0.001$] and descending colon [κ 0.61; $p < 0.001$], fair for the transverse colon [κ 0.36; $p = 0.004$] and slight for the ascending colon [κ 0.17; $p < 0.001$]. Intra-rater agreement was substantial for the sigmoid colon [κ 0.65; $p < 0.001$], moderate for the descending colon [κ 0.59; $p < 0.001$] and transverse [κ 0.52; $p < 0.001$] colon, and substantial for the ascending colon [κ 0.80; $p < 0.001$].

3.2.5. Fat wrapping

Fat wrapping was observed in 14/60 [23.3%] patients. The presence of fat wrapping was strongly associated with severe endoscopic disease [OR 34; 95% CI 6.0–191.8; $p < 0.001$]. Inter- and intra-rater

agreement for fat wrapping was not assessed because this could not be properly assessed using the available cine-loops.

3.2.6. Wall layer stratification

WLS was classified as normal in 55/60 [92%] patients in the sigmoid colon and the descending colon and in 56/60 [93%] patients in the transverse and the ascending colon. Because WLS was normal in most cases, an association between endoscopic disease activity was not assessed.

3.2.7. Lymph nodes

The presence of enlarged lymph nodes was observed in only 3/60 [5%] patients. An association between the presence of lymph nodes and disease activity could therefore not be assessed.

3.3. Biomarkers

In the patients without endoscopic activity the median FCP level was 48 $\mu\text{g/g}$ [IQR 33–180] and the median CRP level was 1.7 mg/L [IQR 0.6–3.0]. In the patients with endoscopic activity the median FCP level was 878 $\mu\text{g/g}$ [IQR 274–1800] and the median CRP level was 3.5 mg/L [IQR 1.6–10.4]. An FCP cut-off of 212 $\mu\text{g/g}$ [sensitivity 81.8%; specificity 81.2%; AUC 0.870] most accurately predicted endoscopic disease activity [Mayo 0 vs Mayo 1–3]. An FCP cut-off of 391 $\mu\text{g/g}$ [sensitivity 81.8%; specificity 81.5%; AUC 0.878] most accurately predicted moderate to severe endoscopic disease activity [Mayo 0–1 vs Mayo 2–3]. An FCP cut-off of 878 most accurately predicted severe endoscopic disease activity [sensitivity 83.3%; specificity 78.6%; AUC 0.867].

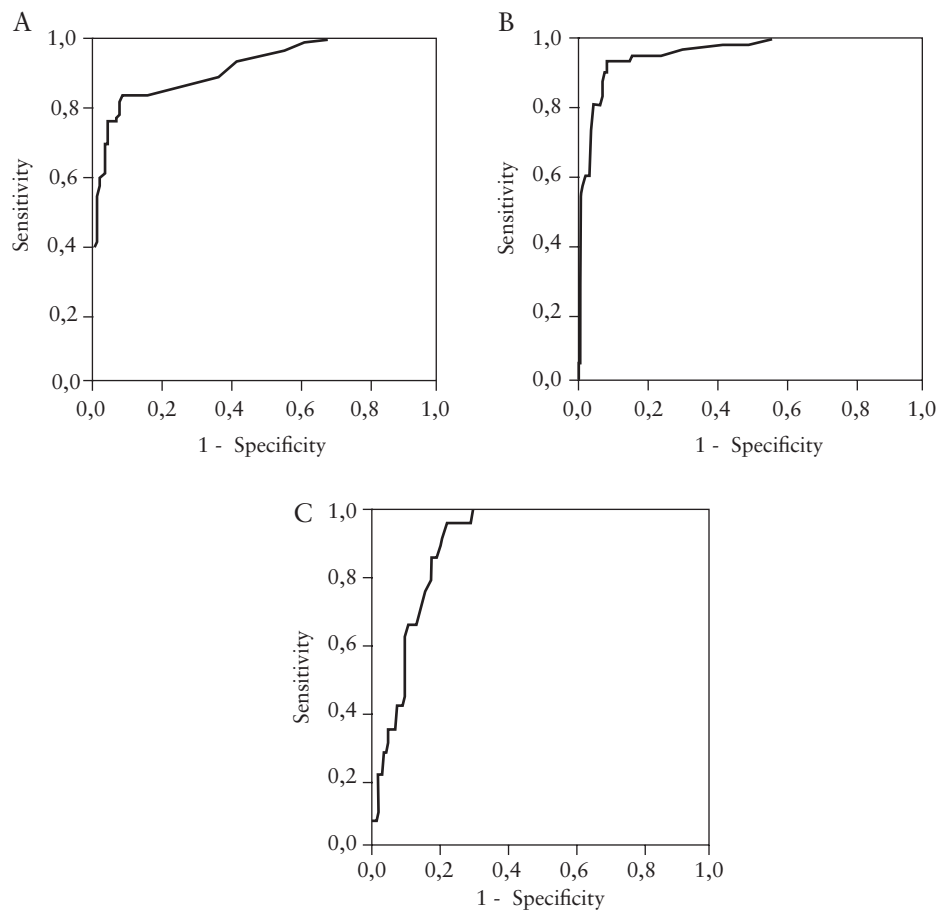


Figure 3. ROC curves for bowel wall thickness [BWT]. [A] ROC curve for BWT in Mayo 0 vs Mayo 1–3 segments. [B] ROC curve for BWT in Mayo 0–1 vs Mayo 2–3 segments. [C] ROC curve for BWT in Mayo 0–2 vs Mayo 3 segments.

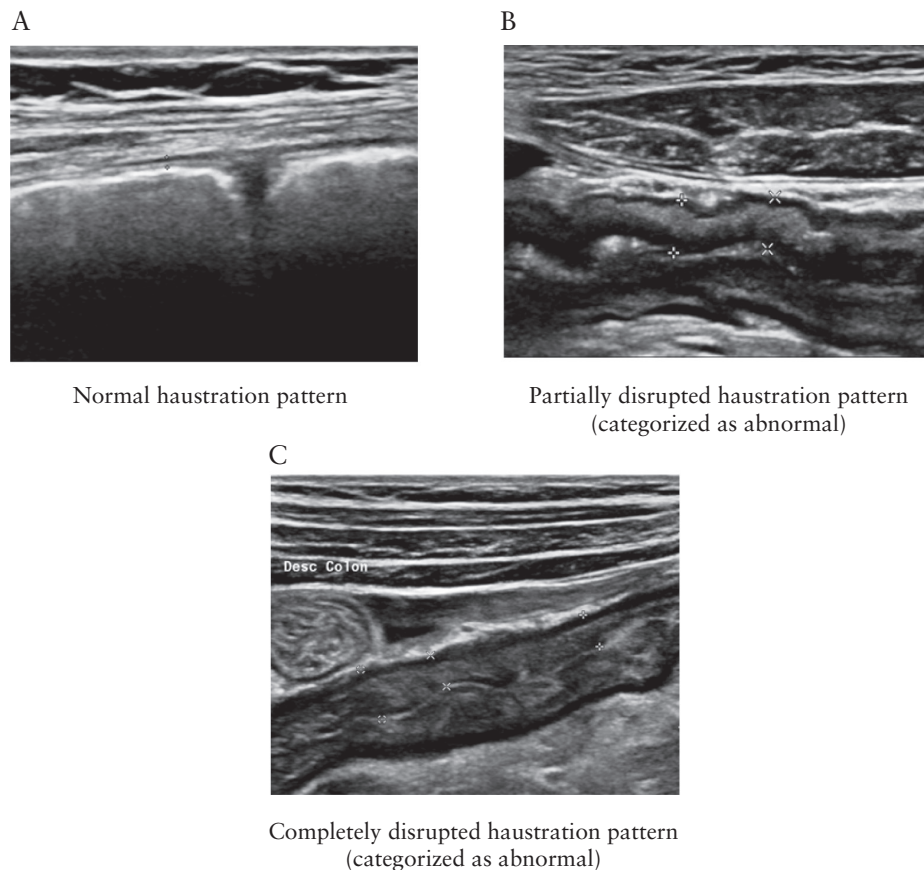


Figure 4. Haustration patterns. [A] normal haustration pattern; [B] partially disrupted haustration pattern [categorized as abnormal]; [C] completely disrupted haustration pattern [categorized as abnormal].

3.4. Combination of CDS and BWT for detection of disease activity

Sensitivity and specificity for detection of disease activity was tested for two combinations of CDS and BWT cut-offs. A combination of BWT > 2 mm or presence of CDS resulted in a sensitivity of 88% and a specificity of 84.3% for detection of disease activity in any colon segment except the rectum. A combination of BWT > 3 mm or presence of CDS with BWT < 3 mm resulted in a sensitivity of 81.5% and specificity of 87.8% for detection of disease activity in any colon segment. A combination of BWT > 2 mm and CDS resulted in a sensitivity of 58.7% and specificity of 96.5% and a combination of BWT > 3 mm and presence of CDS resulted in a sensitivity of 54.3% and specificity of 97.4%.

3.5. Combination of BWT and FCP for detection of disease activity

Sensitivity and specificity for detection of disease activity was tested for a combination of BWT and FCP cut-offs in 54 patients [Table 3]. Patients with proctitis only were excluded from this analysis. A combination of BWT > 2 mm or FCP > 200 µg/g resulted in a sensitivity of 94.9% and specificity of 66.7% for detection of active disease [i.e. > Mayo 0].

3.6. UC-IUS index

Based on the most predictive cut-offs and categories that were identified in the analysis, a point-based index was constructed. The index is detailed in Table 4. The score was calculated and compared per colon segment, excluding the rectum. Subsequently, the final scores

Table 3. Sensitivity and specificity for different combinations of BWT and FCP cut-offs

Combination	Sensitivity [%]	Specificity [%]
BWT > 2 mm or FCP > 200 µg/g	94.9	66.7
BWT > 2 mm and FCP > 100 µg/g	86.7	87.2
BWT > 2 mm and FCP > 200 µg/g	76.9	93.3
BWT > 2 mm and FCP > 300 µg/g	71.8	93.3
BWT > 2 mm and FCP > 400 µg/g	69.2	93.3

BWT, bowel wall thickness; FCP, fecal calprotectin.

were analysed for correlation with the UCEIS and endoscopic Mayo score. The IUS index showed strong correlation with the endoscopic Mayo score [$\rho = 0.830$; $p < 0.001$]. The index also showed strong correlation with the UCEIS index [$\rho = 0.759$; $p < 0.001$]. The final IUS score was also calculated for the second observer and compared with the other IUS score. The mean difference between observers for the final IUS score was 0.28 [SD 1.1; $p = 0.08$] The IUS score showed a strong correlation between observers [$\rho = 0.877$; $p < 0.001$].

4. Discussion

In this study, we developed a new IUS index for the grading of disease activity in UC patients. Endoscopy was used as the reference standard. the index showed strong correlation with endoscopic disease activity through internal validation in this cohort. The index is currently being validated and tested for sensitivity to change in UC patients receiving medical treatment.

Table 4. UC-IUS index

Parameters	Points [0–7]
Bowel wall thickness	
> 2 mm	1
> 3 mm	2
> 4 mm	3
Doppler signal	
Spots	1
Stretches	2
Abnormal haustrations	1
Fat wrapping	1

Several other IUS indices have been suggested for the assessment of disease activity in UC patients.^{9,11,13,14,16,25} The methodologies used in these earlier studies were different because, in most of them, the index parameters and cut-off values were defined before comparison with the reference standard.^{13,15,25} We based our inclusion of parameters and determination of cut-off values on a comparison with the endoscopic results, because it has been postulated that such an approach is optimal for the development of reliable diagnostic instruments.²⁶

Despite the methodological differences, there are obvious similarities between our novel index and other IUS indices for assessing disease activity in UC patients. Evidently, BWT and the presence of CDS are used as parameters in most indices. However, the cut-off values for BWT, CDS categories and other included parameters tend to differ between studies. Parente *et al.* used a predefined cut-off of 4mm for BWT and categorized CDS as present or absent.¹¹ Pascu *et al.* used a BWT cut-off of 3 mm and added increased CDS, loss of compressibility and loss of WLS as parameters.²⁵ Allocca *et al.* developed an index with BWT, CDS, WLS and presence of reactive lymph nodes as parameters.¹³ The cut-off values and included parameters were predefined but the index showed good correlation with the Mayo endoscopic subscore. To our knowledge, the only index that determined cut-off values and parameters based on endoscopy as the reference standard was developed by Civitelli *et al.* for assessing disease activity in paediatric UC patients.¹⁴ These authors developed an index with BWT, CDS, loss of WLS and presence of haustrations as parameters.

The variability in cut-off values and parameters included in different IUS indices shows that it is currently debatable which are best for the assessment of disease activity in UC patients. Additionally, it suggests that IUS is prone to variability in interpretation, as is the case with many diagnostic modalities. Therefore, we chose to construct a point-based score that is easy to use and thus less prone to variation. We did not mathematically weight the included factors because we believe this will make the score unnecessarily complicated. The amount a factor is weighted would be different in every cohort and one would probably need hundreds of patients to be able to accurately weight factors in a heterogeneous population.

Because diagnostic modalities are prone to variability in interpretation, it is important to assess inter-rater agreement. To our knowledge, only Allocca *et al.* investigated inter-rater agreement of IUS examinations in UC patients.¹³ In their study, all IUS examinations were performed by two ultrasonographers and the agreement between examiners for the overall IUS score was excellent. However, inter- and intra-rater agreement for the individual IUS parameters was not assessed. In our study, we investigated inter- and intra-rater agreement for individual IUS parameters by reading cine loops by

two investigators [S.B. and K.N.]. To our knowledge, this has not been reported before. For continuous BWT values, inter-rater agreement was excellent and intra-rater agreement was good, showing that BWT measurements are reproducible between and within observers. For the constructed categories of BWT, inter-rater agreement was good and intra-rater agreement ranged from moderate to good. The lower agreement in the BWT categories is probably a result of the fact that small differences in the continuous measurements could mean a difference in categories, thus potentially leading to lower agreement. Inter- and intra-rater agreement for CDS assessment was good, but fair or poor with regard to haustrations in the transverse and ascending colon. Another recent study showed poor inter- and intra-rater agreement for haustrations and moderate to good agreement for the other parameters.²⁷ This shows that assessing haustrations is probably the most difficult of the included parameters. Because abnormal haustrations were clearly associated with disease activity we decided to include it in the index. An ongoing validation study will have to show if this parameter should remain part of the index. We were unable to assess inter- and intra-rater agreement for fat wrapping as the recorded cine-loops were too short and stationary for this purpose. To properly assess fat wrapping you would need sweeping movement over a large area, which was only performed in live scanning. Inter- and intra-rater agreement could probably be improved with more experience and optimization of measurement definitions in the future. However, it is important to note that we used multiple categories for most parameters, which probably resulted in lower agreement. Another important factor could be that it is more difficult to assess certain parameters using cine-loops. It is to be expected that inter-rater agreement will decline when more investigators are involved in image interpretation. Nevertheless, correlation of the final score was strong between observers. This shows that a combination of parameters results in a more accurate overall assessment. It could also be that IUS interpretation may be more reliable when performing the examination than when only interpreting cine-loops. This is important to take into consideration, especially when considering the use of IUS in clinical trials that rely on central reading. The reliability of central reading of IUS examinations should therefore be investigated in future studies.

There are different technical aspects that are of importance when interpreting the results of this work and other comparable studies. For instance, IUS examinations are usually performed with a single US device in most studies. This is of importance for consistency when assessing parameters, such as CDS in the bowel wall. However, it is likely that there are differences in sensitivity of CDS measurements between US machines and US vendors. To our knowledge, a comparison of different US machines for measuring CDS in the bowel wall has not been conducted. Such a study would be of particular interest. Another potential issue when performing CDS measurements is the distance between the bowel wall and the US probe. Due to physical limitations of US, high-frequency colour Doppler does not penetrate as deep into the body as lower frequency colour Doppler due to attenuation, while low-frequency Doppler has lower spatial resolution. This could reduce the number of vessels detected in deeper lying bowel segments and result in undervaluation of disease activity.²⁸ Another factor that could potentially influence the presence of CDS is fibrosis of the bowel wall in patients with long-standing UC. To our knowledge, no studies have looked at the relationship between CDS and fibrosis in UC. However, two previous studies have indicated that fibrosis can result in reduced bowel wall vascularization and less CDS in Crohn's disease patients.^{29,30}

We also assessed the sensitivity and specificity for detection of disease activity when combining BWT with FCP measurements. Here, we show that sensitivity and specificity increases when combining these two parameters. Addition of FCP could therefore be of additive value in patients with minor findings on IUS, in order to better discriminate between quiescent and mild disease. FCP could also be useful for detection of proctitis in patients who have normal IUS findings in all colonic segments. Because FCP is already widely used, we believe that combining FCP and IUS for monitoring of disease activity should be of particular interest in clinical practice and in future studies.

Our study has several strengths. First, we used a systematic approach for determination of cut-off values and selection of IUS parameters with endoscopy as the reference standard. Second, ultrasonographers and endoscopists were blinded to the results of the other examination. Third, we assessed inter- and intra-rater agreement of the IUS parameters using cine-loops. Finally, the UC-IUS score was correlated with two different endoscopic scores. Our study also has some limitations. First, IUS examinations were not performed twice by different ultrasonographers. Second, there was no central reading of the endoscopy and third, we could not assess inter- and intra-rater agreement for all parameters [i.e. fat wrapping].

In conclusion, we have developed an UC-IUS index that showed strong correlation with endoscopic disease activity through internal validation in the same cohort. Addition of FCP increased the accuracy of detection of disease activity. We showed that IUS could be a reliable substitute for endoscopy for assessing disease activity in UC patients, except in patients with proctitis. Broad implementation of IUS could therefore reduce the need for endoscopy and may be especially useful for rapid [on the spot] detection of flares and for monitoring of treatment outcomes. Because this is a pilot study, the UC-IUS index should be validated in future studies and tested for sensitivity to change after medical treatment.

References

- Schnitzler F, Fidler H, Ferrante M, *et al.* Mucosal healing predicts long-term outcome of maintenance therapy with infliximab in Crohn's disease. *Inflamm Bowel Dis* 2009;15:1295–301.
- Colombel JF, Rutgeerts P, Reinisch W, *et al.* Early mucosal healing with infliximab is associated with improved long-term clinical outcomes in ulcerative colitis. *Gastroenterology* 2011;141:1194–201.
- Sharara AI, El Reda ZD, Harb AH, *et al.* The burden of bowel preparations in patients undergoing elective colonoscopy. *United European Gastroenterol J* 2016;4:314–8.
- De Vos M, Louis EJ, Jahnsen J, *et al.* Consecutive fecal calprotectin measurements to predict relapse in patients with ulcerative colitis receiving infliximab maintenance therapy. *Inflamm Bowel Dis* 2013;19:2111–7.
- Travis S, Satsangi J, Lémann M. Predicting the need for colectomy in severe ulcerative colitis: a critical appraisal of clinical parameters and currently available biomarkers. *Gut* 2011;60:3–9.
- Dignass A, Eliakim R, Magro F, *et al.* Second European evidence-based consensus on the diagnosis and management of ulcerative colitis part 1: definitions and diagnosis. *J Crohns Colitis* 2012;6:965–90.
- Magro F, Gionchetti P, Eliakim R, *et al.*; European Crohn's and Colitis Organisation [ECCO]. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders. *J Crohns Colitis* 2017;11:649–70.
- D'Haens G, Ferrante M, Vermeire S, *et al.* Fecal calprotectin is a surrogate marker for endoscopic lesions in inflammatory bowel disease. *Inflamm Bowel Dis* 2012;18:2218–24.
- Bots S, Nylund K, Löwenberg M, Gecse K, Gilja OH, D'Haens G. Ultrasound for assessing disease activity in IBD patients: a systematic review of activity scores. *J Crohns Colitis* 2018;12:920–9.
- Antonelli E, Giuliano V, Casella G, *et al.* Ultrasonographic assessment of colonic wall in moderate-severe ulcerative colitis: comparison with endoscopic findings. *Dig Liver Dis* 2011;43:703–6.
- Parente F, Molteni M, Marino B, *et al.* Are colonoscopy and bowel ultrasound useful for assessing response to short-term therapy and predicting disease outcome of moderate-to-severe forms of ulcerative colitis?: a prospective study. *Am J Gastroenterol* 2010;105:1150–7.
- Maconi G, Ardizzone S, Parente F, Bianchi Porro G. Ultrasonography in the evaluation of extension, activity, and follow-up of ulcerative colitis. *Scand J Gastroenterol* 1999;34:1103–7.
- Allocca M, Fiorino G, Bonovas S, *et al.* Accuracy of humanitas ultrasound criteria in assessing disease activity and severity in ulcerative colitis: a prospective study. *J Crohns Colitis* 2018;12:1385–91.
- Civitelli F, Di Nardo G, Oliva S, *et al.* Ultrasonography of the colon in pediatric ulcerative colitis: a prospective, blind, comparative study with colonoscopy. *J Pediatr* 2014;165:78–84.e2.
- Parente F, Molteni M, Marino B, *et al.* Bowel ultrasound and mucosal healing in ulcerative colitis. *Dig Dis* 2009;27:285–90.
- Ishikawa D, Ando T, Watanabe O, *et al.* Images of colonic real-time tissue sonoelastography correlate with those of colonoscopy and may predict response to therapy in patients with ulcerative colitis. *BMC Gastroenterol* 2011;11:29.
- Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med* 1987;317:1625–9.
- Travis SP, Schnell D, Krzeski P, *et al.* Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut* 2012;61:535–42.
- Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998;43:29–32.
- Nylund K, Hausken T, Ødegaard S, Eide GE, Gilja OH. Gastrointestinal wall thickness measured with transabdominal ultrasonography and its relationship to demographic factors in healthy subjects. *Ultraschall Med* 2012;33:E225–32.
- Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg* 2018;126:1763–8.
- McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–82.
- Cohen J. A coefficient of agreement for nominal scales Educational and psychological measurement. *Educational and Psychological Measurement* 1960;20:1:37–46.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
- Pascu M, Roznowski AB, Müller HP, Adler A, Wiedenmann B, Dignass AU. Clinical relevance of transabdominal ultrasonography and magnetic resonance imaging in patients with inflammatory bowel disease of the terminal ileum and large bowel. *Inflamm Bowel Dis* 2004;10:373–82.
- Whiting PF, Rutjes AW, Westwood ME, *et al.*; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- de Voogd FWR, Gecse K, Allocca M, *et al.* Inter-observer agreement of an expert panel for gastrointestinal ultrasound in ulcerative colitis. *Journal of Crohn's and Colitis*. 2020;14[Supplement 1]:S486–S7.
- Postema MK, Kotopoulos S, Jenderka, KV. Basic physical principles of medical ultrasound. *EFSUMB course book on ultrasound 2nd edition*. London: Latimer Trend & Company Ltd; 2018.
- Sasaki T, Kunisaki R, Kinoshita H, *et al.* Use of color Doppler ultrasonography for evaluating vascularity of small intestinal lesions in Crohn's disease: correlation with endoscopic and surgical macroscopic findings. *Scand J Gastroenterol* 2014;49:295–301.
- Ripollés T, Rausell N, Paredes JM, Grau E, Martínez MJ, Vizuete J. Effectiveness of contrast-enhanced ultrasound for characterisation of intestinal inflammation in Crohn's disease: a comparison with surgical histopathology analysis. *J Crohns Colitis* 2013;7:120–8.