

Original Article

Development and Validation of a Simple Ultrasound Activity Score for Crohn's Disease

Fredrik Sævik,^{a,b} Ragnar Eriksen,^c Geir Egil Eide,^{d,e} Odd Helge Gilja,^{a,b} Kim Nylund^{b,a}

^aDepartment of Clinical Medicine, University of Bergen, Bergen, Norway ^bNational Centre for Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway ^cDepartment of Medicine, Ålesund Hospital, Ålesund, Norway ^dDepartment of Global Public Health and Primary Care, University of Bergen, Bergen, Norway ^eCentre for Clinical Research, Haukeland University Hospital, Bergen, Norway

Corresponding author: Fredrik Sævik, MD, Department of Clinical Medicine, University of Bergen, Haukeland University Hospital, Jonas Lies vei, N-5021 Bergen, Norway. Tel: +47 40 01 39 10; Fax: + 47 55 97 29 50; Email: fredrik.savik@gmail.com.

Abstract

Background and Aims: To improve management of patients with Crohn's disease, objective measurements of disease activity are needed. Ileocolonoscopy is the current reference standard but has limitations that restrict repeated use. Ultrasonography is potentially useful for activity monitoring, but no validated sonographic activity index is currently in widespread use. Thus, we aimed to construct and validate a simple ultrasound score for Crohn's disease.

Methods: Forty patients were prospectively examined with ultrasound and endoscopy in the development phase. The Simple Endoscopic Score for Crohn's Disease [SES-CD] was used as a reference standard. Seven ultrasound variables [bowel wall thickness, length, colour Doppler, stenosis, fistula, stratification and fatty wrapping] were initially included, and multiple linear regression was used to select the variables that should be included in the final score. Second, the ultrasound data from each patient were re-examined for interobserver assessment using weighted kappa and intraclass correlation. Finally, the activity index was validated in a new cohort of 124 patients.

Results: Length, fistula and stenosis were excluded. The combination of the remaining variables provided a multiple correlation coefficient of $r=0.78$. Interobserver analysis revealed poor agreement for stratification and fatty wrapping and these were thus excluded. There was excellent interobserver agreement for the remaining score consisting of wall thickness and colour Doppler. In both patient cohorts, the ultrasound score correlated well with SES-CD [Development cohort: $\rho = 0.83$, $p < 0.001$, Validation cohort: $\rho = 0.78$, $p < 0.001$]. A receiver operating characteristic curve analysis revealed an area under the curve of 0.92 and 0.88 for detecting endoscopic activity and moderate endoscopic activity, respectively.

Conclusions: A simple ultrasound activity index for Crohn's disease consisting of bowel wall thickness and colour Doppler was constructed and validated and correlated well with endoscopic disease activity.

ClinicalTrials. gov ID: NCT03481751

Key Words: Crohn's disease; disease activity; monitoring; ultrasound; activity index

1. Introduction

Crohn's disease [CD] has a relapsing–remitting course which necessitates frequent follow-up examinations to monitor disease activity.¹ Disease management was previously guided by patient-reported symptoms, and treatment targets were based on symptom control. However, the patient's symptoms do not necessarily correspond to inflammatory activity,^{2,3} and current guidelines recommend that management should be based on objective evaluations.^{1,4} Furthermore, by moving therapeutic goals from clinical remission to objective endpoints, long-term patient outcomes seem to improve.⁵

Ileocolonoscopy is considered the reference standard method for determining disease status in CD.¹ The endoscopic scoring systems Crohn's Disease Endoscopic Index of Severity [CDEIS]⁶ and Simple Endoscopic Score for Crohn's Disease [SES-CD]⁷ are useful for standardizing measurements of disease activity.⁸ Although validated and reproducible, these scoring systems are complex and cumbersome to use in clinical practice,⁹ and even though ileocolonoscopy is an excellent tool for activity monitoring, it cannot be performed on a regular basis as it is invasive, is resource-intensive and causes considerable patient discomfort. As numerous follow-up examinations are required, simple non-invasive surrogate markers are needed.

Biomarkers such as C-reactive protein [CRP] and faecal calprotectin are well established in both primary work-up and disease monitoring.¹ Still, as they cannot depict disease location and have limited accuracy,¹⁰ additional tools are required.

Cross-sectional imaging modalities such as computed tomography [CT], magnetic resonance imaging [MRI] and ultrasound [US] are increasingly being used in CD management.¹ As most of the small bowel, subepithelial structures and peri-intestinal complications cannot be visualized by ileocolonoscopy, these methods provide important contributions to the overall assessment.⁴ In addition, they can be useful for activity monitoring, in which MRI and US are preferred methods, as CT causes radiation exposure.¹

The Magnetic Resonance Index of Activity [MaRIA]^{11,12} is a validated MRI-based activity index that correlates well with endoscopic activity in CD.¹³ In addition, a simplified derivative of MaRIA, as well as other MRI-based scoring systems with comparable diagnostic accuracies, are currently available.^{14–17} As the MRI-based approach may reduce the need for ileocolonoscopy examinations, its importance in patient management could increase significantly. Still, MRI is resource-intensive, requires specific preparations and has relatively low availability.

Gastrointestinal ultrasound [GIUS] has high diagnostic accuracy for detecting active CD,¹⁸ and in trained hands, it can make significant impact on clinical decision-making.¹⁹ Furthermore, as it is non-invasive, readily available and can be performed bedside, the modality seems well suited for bedside and frequent activity monitoring.²⁰ Still, interpretation of the GIUS findings may be influenced by the sonographer's level of experience. A standardized ultrasound activity index may simplify the interpretation of the sonographic findings, allowing for easier comparison between different examinations during follow-up. Although various sonographic activity scores are available,^{21–26} the methodology for development was shown to be insufficient in most studies,²⁷ and no index is in widespread clinical use. In this study, we aimed to develop, validate and assess inter-rater variability of a simple sonographic activity score for CD patients.

2. Materials and Methods

Study design

The study was separated into three phases, development, interobserver and validation, as similarly done by Daperno *et al.*⁷ In the development phase, we conducted a prospective single-centre study to identify and select the ultrasound variables that should be included in the activity index. In the second part, still images and cine loops from the included patients were re-examined to evaluate interobserver reliability. Finally, a scoring system consisting of significant variables with low interobserver variability was validated in a prospective multicentre study.

Patients

Forty CD patients scheduled for ileocolonoscopy as part of regular follow-up, including suspicion of active disease, remission, evaluation of treatment effect and relapse, were prospectively recruited for the development phase. Patient enrolment was performed at the Department of Medicine at Haukeland University Hospital in Bergen [Norway] from 2015 to 2017. In the validation phase, a new population of 124 patients with identical inclusion criteria was prospectively recruited at the Department of Medicine at Haukeland University Hospital and Ålesund Hospital from 2018 to 2019. Included patients were examined with ultrasonography, as well as clinical scoring and blood and stool sampling. Exclusion criteria were age less than 18 years, pregnancy, previous colectomy, isolated disease in the upper gastrointestinal tract or ongoing gastroenteritis.

Clinical and biochemical tests

Clinical assessment on the duration of CD, current medical and previous surgical treatment, age, sex, weight and height were obtained from the patients and/or access to medical records with consent, and each patient was phenotypically classified according to the Montreal classification.²⁸ Harvey Bradshaw Index [HBI]²⁹ was used to assess clinical disease activity, where clinical remission was defined as HBI < 5. Blood and stool samples were obtained within 1 week prior to or after the ultrasound examination. The biochemical markers haemoglobin [g/dL], leukocyte count [$10^9/L$], platelet count [$10^9/L$], CRP [mg/L] and albumin [g/L] were measured from blood samples, whereas calprotectin [mg/kg] was analysed from faecal samples.

Ileocolonoscopy

Ileocolonoscopy was performed as part of routine work by 20 endoscopists with a range of 1–30 years of experience. Disease activity measurements were assessed using the SES-CD in both development and validation phases and were calculated by the endoscopist immediately after the examination. The endoscopists were blinded to the results from the ultrasound examination but knew clinical data such as HBI, previous findings and Montreal classification. The SES-CD evaluates four endoscopic variables [ulcer size, ulcerated surface, affected surface and stenosis] in five bowel segments [rectum, left colon, transverse colon, right colon and the terminal ileum]. Each parameter yields a value of 0–3 according to its severity, and by summarizing the scores, segmental and total endoscopic activity can be quantified.⁷ Endoscopic remission was defined as SES-CD 0–2, as previously recommended.³⁰ In the same manner as in the development and validation of the SES-CD,⁷ we also included patients who had undergone ileocecal resection.

Ultrasound examination development phase

The ultrasound examinations were performed within 14 days prior to or after the endoscopic procedure, but not during bowel preparation or just after the ileocolonoscopy as the intestine usually collapses, making standardization more complicated. Furthermore, there were no changes in medical therapy between the examinations. Patients were examined after an overnight fast. All examinations were performed by one investigator [K.N.], who was blinded to the results from the ileocolonoscopy, using a Logiq E9 scanner [GE Healthcare], with low- [C1-6, 1–6 MHz] and high-frequency [9L, 5.5–9 MHz] probes for overview and detailed examinations of the bowel wall, respectively.

The large bowel scan was performed by following its course from the terminal ileum and further distally to the rectum, as previously recommended.³¹ In the development phase, seven well-established ultrasound variables (bowel wall thickness [BWT], length of affected segment, colour Doppler, stratification, fatty wrapping, fistula and stenosis) were selected and weighted according to current knowledge.³² The included parameters were evaluated in five ileocolonic segments [rectum, left colon, transverse colon, right colon and terminal ileum], each yielding a numerical value corresponding to its severity.

BWT measurements were performed on the anterior wall in longitudinal section as previously described,³¹ and two representative measurements were averaged. In affected bowel segments [BWT > 3 mm], the measurements were performed in the thickest section, while representative measurements were performed within healthy bowel segments without further standardization. The rectum was examined using a curvilinear probe because the bowel-segment is deeply located, and as the rectal wall is thicker on ultrasound,³³ a cut-off of 4 mm was used. Further classification of BWT was based on the authors' clinical experience as well as from previous studies.³⁴ The length of the affected segments was measured in centimetres. Colour Doppler was performed on segments with BWT exceeding 3 mm, using standardized scanning pre-sets. Colour Doppler measurements were not performed in the rectum due to reduced sensitivity at increased depths. The velocity scale was set to 5 cm/s, enabling registration of vessels with low velocities. Gain was turned up to a level where flash artefacts occurred, and then lowered until they disappeared. All Doppler acquisitions were performed during patient breath-hold. Evaluation of colour Doppler was performed using a modified version of that of Spalinger *et al.*³⁵ [Table 1], counting the number of Doppler signals per cm². Focal or diffuse loss of bowel wall stratification were defined as a limited hypoechoic disruption or total loss of bowel wall stratification, respectively. Fatty wrapping was defined as a hyperechoic mass encircling the affected bowel

loop, and stenosis as an intestinal section with total wall thickness greater than 3 mm, a narrow or closed off lumen, stiff appearance, and a lack of peristaltic movement with or without pre-stenotic dilatation [defined as > 2.5 mm]. Fistulas were defined as hypoechoic narrow tracts between the bowel loop and other tissues. The ultrasound variables with their corresponding definitions and score characteristics are presented in Table 1 [see also Supplementary Figure 1].

Inter-rater reliability

Before including the validation cohort, the ultrasound data from the development phase was reviewed by another examiner [F.S.] for assessing inter-rater reliability of the selected sonographic parameters. By using dedicated software (Phillips DICOM Viewer [Phillips Medical Systems] and Onis [DigitalCore, Co. Ltd]), each patient was re-evaluated and scored according to the activity index.

Validation phase

GIUS was performed identically as in the development phase, using a Logiq E9 scanner with equal probes at both centres [Haukeland University Hospital and Ålesund Hospital]. The ultrasound examinations were performed by two investigators [F.S. and R.E.], and interobserver assessments [F.S. and K.N.] were performed in a subgroup of patients. All investigators were blinded to the results from the corresponding endoscopic examinations. Each patient was scored using a dedicated score sheet.

Statistics

Descriptive statistics were used to analyse demographical data, and distribution was assessed by inspecting histograms as well as using the Shapiro-Wilk test.

Development phase

To identify the ultrasound variables that could be eligible for further analysis, a correlation analysis [Spearman's rank correlation] was performed between the SES-CD and each US variable. We also conducted a multiple linear regression analysis to select the ultrasound variables to be included in the activity index, using the SES-CD as the dependent variable. Ultimately, the final activity index was correlated with the SES-CD using Spearman's rank correlation.

Inter-rater reliability

Evaluation of inter-rater agreement were performed using weighted kappa [wκ] for the ultrasound variables as well as the calculated score. Additionally, intra-class correlation [ICC] and Bland-Altman³⁶ analyses were used to assess the level of agreement between the

Table 1. Ultrasound variables that were eligible for inclusion in the ultrasonographic activity score for Crohn's disease: the definition and score characteristics are presented for each variable.

Variable	Score			
	0	1	2	3
Bowel wall thickness (mm)	<3.0	3.0–4.9, or 4.0–4.9 mm [rectum]	5.0–7.9	≥ 8.0
Stenosis	None	Suspected [thickened wall with narrow lumen]	Suspected several per segment	Suspected with pre-stenotic dilatation [> 2.5 cm]
Length of affected segment (cm)	None	<5	5–10	>10
Colour Doppler score	No or single vessel	2–5 vessels per cm ²	>5 vessels per cm ²	
Stratification	Normal	Focal loss	Diffuse loss	
Fatty wrapping	Absent	Present		
Fistula	Absent	Present		

investigators for the calculated score. The data were further tested for fixed and proportional biases between the observers using a one-sided *t*-test and linear regression, respectively.

Validation phase

The US activity index was applied on a new patient cohort for validation, by testing its correlation [Spearman's rho] with the SES-CD.

Weighted kappa, ICC and Bland–Altman analyses were used to assess interobserver agreement.

Statistics of the entire cohort

The correlation between endoscopy, ultrasound, and clinical and biochemical variables was performed using Spearman's rho. The patients were dichotomized into active and inactive groups

Table 2. Clinical, biochemical and ultrasonographic characteristics of the 164 patients included in the study developing a simple ultrasonographic activity score for Crohn's disease in Bergen [Norway] from 2015 to 2017 [development] and Bergen and Ålesund from 2018 to 2019 [validation].

Variable	Development [range]	Validation [range]	Total population [range]
	<i>n</i> = 40	<i>n</i> = 124	<i>n</i> = 164
Endoscopic activity/remission	27/13	80/44	107/57
Sex, female/male	25/15	73/51	98/66
Age in years, median, range	37.5 [19, 83]	42.5 [18, 75]	41.5 [18,83]
Years of sickness, median, range	9 [0, 44]	10 [0, 40]	10 [0, 44]
Body mass index [BMI], kg/cm ² , median, range	24.5 [17.9–34.6]	24.2 [17.9–35.8]	24.2 [17.9–35.8]
Age at diagnosis, years, <i>n</i>			
<16	4	17	21
17–40	29	84	113
>40	7	23	30
Disease location, <i>n</i>			
Ileal [L1]	14	63	77
Colonic [L2]	8	19	27
Ileocolonic [L3]	18	42	60
Upper disease [L4]	1	8	9
Disease behaviour, <i>n</i>			
Non-stricturing, non-penetrating [B1]	11	61	72
Stricturing [B2]	16	54	70
Penetrating [B3]	13	9	22
Perianal involvement	2	20	22
Previous surgery, <i>n</i>	21	46	67
Concomitant treatment, <i>n</i>			
Aminosalicylate	5	15	20
Azathioprine	12	37	49
Methotrexate	3	9	12
Prednisolone	1	8	9
Budesonid	5	7	12
Infliximab	8	39	47
Adalimumab	6	11	17
Certolizumab	0	1	1
Vedolizumab	2	14	16
Biochemical tests, median, range			
Haemoglobin, g/dL	14.1 [9.7, 17.5]	13.8 [9.1, 16.3]	13.9 [9.1, 17.5]
Leukocyte count, 10 ⁹ /L	4.6 [2.2, 11.6]	6.4 [2.5, 15.9]	6.0 [2.2, 15.9]
Platelet count, 10 ⁹ /L	276.5 [175, 498]	272.0 [118, 513]	275.0 [118, 513]
Albumin, mg/L	45.5 [32, 52]	45.0 [34, 54]	45.0 [32, 54]
C-reactive protein, mg/L	2.0 [0, 43]	1.0 [1, 96]	1.5 [0, 96]
Calprotectin, mg/kg	74.0 [15, 2178]	39.8 [15, 1492]	42.0 [15, 2178]
Harvey Bradshaw Index [HBI], median, range	4 [0, 20]	2 [0, 12]	3 [0, 20]
Ultrasound variables, score > 0			
Bowel wall thickness	30	86	116
Colour Doppler	18	59	77
Stratification	12	19	31
Fatty wrapping	12	23	35
Stenosis	10	–	–
Length	30	–	–
Fistula	0	–	–

Definitions of the pathological ultrasound variables: bowel wall thickness > 3 mm, colour Doppler > 1 Doppler signal per cm², stratification as loss of normal echotexture, fatty wrapping as present, stenosis as thickened bowel wall [>3 mm] with narrowed lumen, length as present [any length of a pathological bowel wall segment], fistula as present.

at ileocolonoscopy, and receiver operating characteristics [ROC] curve analysis was performed to identify suitable cut-offs, as well as for calculation of diagnostic accuracy in terms of sensitivity and specificity.

Data analysis was performed using SPSS Statistics software version 25. The level of significance was $p \leq 0.05$.

Ethical consideration

The study was approved by the Regional Ethical Committee in Western Norway, and each patient signed informed consent before participation. The study is registered in ClinicalTrials.gov, number NCT03481751.

3. Results

Demographics

Forty and 124 patients were included in the development and validation cohorts, respectively. In the development cohort, 17 and 27 patients were in clinical and endoscopic activity, respectively. Corresponding numbers for the validation cohort were 29 and 80 patients. Further patient characteristics are presented in Table 2.

Development of activity index

We found no fistulas in the development cohort and this variable was excluded. The remaining six ultrasound variables [BWT, colour

Table 3. Contributions of the different ultrasound variables eligible for inclusion in the ultrasonographic activity score for Crohn's disease. The coefficient of multiple linear regression, multiple correlation [r] and Spearman correlation [rho] to SES-CD are presented based on 40 included patients in the development part of the study.

Variable	All variables	Without stenosis	Without fatty wrapping	Without stratification	Sum of all variables ^a	SUS-CD ^b
US variables						
BWT	1.053	1.053	1.113	1.385	-	-
Colour Doppler	1.929	1.934	2.097	2.353	-	-
Stratification	1.223	1.225	1.389	-	-	-
Fatty wrapping	1.267	1.275	-	-	-	-
Stenosis	0.013	-	-	-	-	-
Intercept	0.242	0.242	0.305	0.357	-	-
Multiple r	0.777	0.777	0.770	0.736	-	-
Rho						
SES-CD	0.892	0.885	0.872	0.829	0.881	0.832

Abbreviations: US, ultrasound; BWT, bowel wall thickness; SUS-CD, Simple Ultrasound Score for Crohn's Disease; SES-CD, Simple Endoscopic Score for Crohn's Disease.

^aSimple sum of bowel wall thickness, colour Doppler, stratification and fatty wrapping without including the regression equation.

^bSimple sum of bowel wall thickness and colour Doppler.

Table 4. Interobserver agreement between the investigators for calculating the activity index [wk and ICC] as well as for each included variable [wk] within the two cohorts [n = 40, and n = 124, respectively].

Variable	Weighted kappa		Intraclass correlation	
	Development cohort ^a	Validation cohort ^b	Development cohort ^a	Validation cohort ^b
	wk, 95% CI	wk, 95% CI	ICC, 95% CI	ICC, 95% CI
Activity index	0.82 [0.73, 0.91]	0.84 [0.69, 0.99]	0.95 [0.90, 0.97]	0.90 [0.78, 0.96]
BWT	0.81 [0.69, 0.93]	0.84 [0.70, 0.98]	-	-
Colour Doppler	0.93 [0.83, 1.00]	0.86 [0.66, 1.00]	-	-
Stratification	0.46 [0.24, 0.68]	-	-	-
Fatty wrapping	0.51 [0.22, 0.81]	-	-	-

Abbreviation: BWT, bowel wall thickness

^aRe-evaluation of the development cohort using dedicated software.

^bUltrasound examination performed independently by two investigators

Table 5. Scoring sheet for the Simple Ultrasound Score for Crohn's Disease [SUS-CD].

Variable	Ileum	Right colon	Transverse colon	Left colon	Rectum	Total
Bowel wall thickness [0–3]						
Colour Doppler score [0–2]						
					Score	

The severity of each variable in each segment is noted and further summed. The sum of the scores reflects the degree of disease activity. Ultrasound variables with their corresponding definitions and score characteristics are presented in Table 2. Colour Doppler measurements were not performed in the rectum due to reduced sensitivity at increased depths.

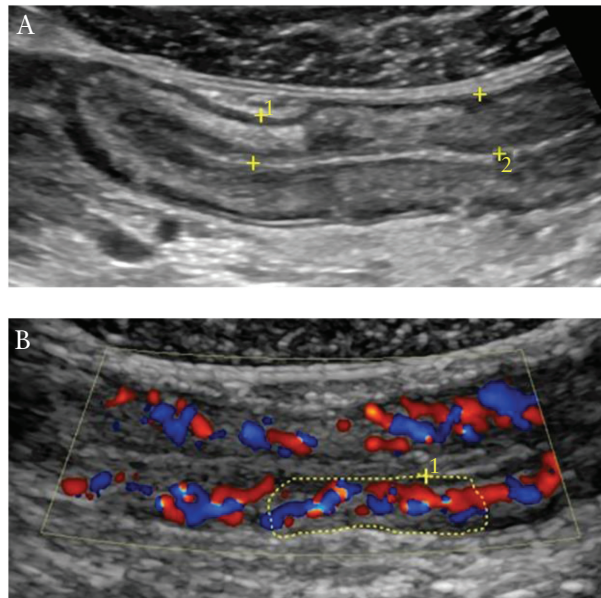


Figure 1. Examples of ultrasound variables included in the activity index. [A] Increased bowel wall thickness with normal stratification are demarcated between the yellow callipers. [B] Increased colour Doppler signals corresponding to a score of 2. Evaluation of colour Doppler was performed by counting the number of Doppler signals per cm² [within the demarcated area of the bowel wall] as demonstrated in the figure.

Doppler, stratification, fatty wrapping, length and stenosis] correlated with SES-CD and were included for further analyses using multiple linear regression. All variables were tested for multicollinearity, revealing a high intercorrelation [$r = 0.90$] between BWT and length as well as high variance inflation factors [BWT: 5.5, length: 5.7] and low tolerance values [BWT: 1.8, length: 1.7] of the aforementioned variables. Consequently, length was excluded to ensure the assumption of multicollinearity. The combination of the remaining five variables provided the highest multiple correlation coefficient [$r = 0.78$], explaining 60.4% of the variance of SES-CD. However, it was not reduced after exclusion of stenosis [$r = 0.78$]. Stenosis had only a minimal contribution [0.2%, $p = 0.987$] to the calculated model and was thus excluded from the final index. Only BWT [38.4%, $p = 0.004$] and colour Doppler [34.7%, $p = 0.006$] had significant contributions to the model, while less unique contributions were found for fatty wrapping [12.3%, $p = 0.329$] and stratification [22.6%, $p = 0.081$].

An equation for activity estimation could be derived from the multiple linear regression analysis: [BWT · 1.053] + [Colour Doppler · 1.934] + [Fatty wrapping · 1.275] + [Stratification · 1.225] + 0.242. However, the equation did not yield a better correlation with SES-CD than by summarizing the numerical values of the variables [$\rho = 0.885$ vs $\rho = 0.881$, respectively] [Table 3]. Accordingly, the sum of the four included ultrasonographic variables constituted the index in order to simplify the calculation.

The ultrasound score correlated well with SES-CD [$\rho = 0.88$, $p < 0.001$], as well as with HBI [$\rho = 0.41$, $p = 0.009$], CRP [$\rho = 0.62$, $p < 0.001$] and calprotectin [$\rho = 0.51$, $p = 0.004$].

Inter-rater reliability

By re-scoring the development cohort, the ultrasound activity index correlated well with SES-CD [$\rho = 0.85$, $p < 0.001$]. We also found good agreement between the investigators for calculating the activity

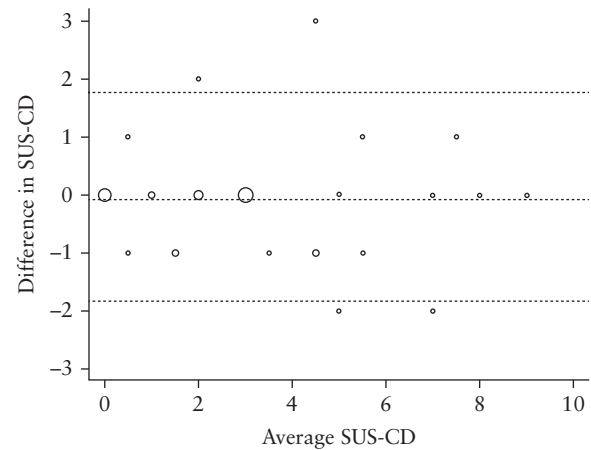


Figure 2. Bland–Altman plot displaying the agreement between the investigators [F.S. and K.N.] for the calculated activity index on 40 subjects. The size of the circles is proportional to the number of observations at the specific locations. The width of the agreement interval was within clinical acceptable limits. Interobserver analysis was performed after re-examining the development cohort.

index, as well as for BWT and colour Doppler. There was, however, poorer agreement for stratification and fatty wrapping [Table 4].

Stratification and fatty wrapping did not contribute significantly to the model, and as we revealed poor interobserver agreement, we further simplified the activity index by excluding these variables. Thus, the final ultrasound activity score, the Simple Ultrasound Score for Crohn's Disease [SUS-CD], was constituted by BWT and colour Doppler [Table 5 and Figure 1]. The SUS-CD correlated well with SES-CD [$\rho = 0.83$, $p < 0.001$], which was only slightly poorer than by using all four variables [$\rho = 0.88$, $p < 0.001$]. Furthermore, we found excellent interobserver agreement for calculating the simplified score [$\kappa = 0.82$, ICC = 0.95], and the Bland–Altman analysis revealed no fixed or proportional biases between the two investigators [Figure 2].

Validation phase

The SUS-CD correlated well with SES-CD [$\rho = 0.78$, $p < 0.001$] when applied on a new patient population. Corresponding values between the ultrasound score and HBI, CRP and calprotectin were 0.24, 0.42 and 0.43, respectively. A subgroup of 23 patients was examined by two investigators for interobserver analysis, revealing excellent agreement [Table 4], and no fixed or proportional biases between the sonographers were found using Bland–Altman analysis.

Score characteristics of the entire cohort

The SUS-CD correlated well with SES-CD [$\rho = 0.80$, $p < 0.001$] [Figure 3]. Poorer correlations were revealed for HBI [$\rho = 0.33$, $p < 0.001$], CRP [$\rho = 0.46$, $p < 0.001$] and calprotectin [$\rho = 0.51$, $p < 0.001$]. Similar findings were found between SES-CD and clinical and biochemical markers [HBI: $\rho = 0.37$, $p < 0.001$, CRP: $\rho = 0.44$, $p < 0.001$, calprotectin: $\rho = 0.54$, $p < 0.001$].

ROC curve analysis revealed an area under the curve of 0.92 for detecting endoscopic activity [SES-CD > 2] [Figure 4], where a cut-off of SUS-CD ≥ 1 yielded a sensitivity of 95.3% and specificity of 70.3%. Thus, remission could be defined as SES-CD = 0. The corresponding value for detecting moderate activity [SES-CD > 7] was

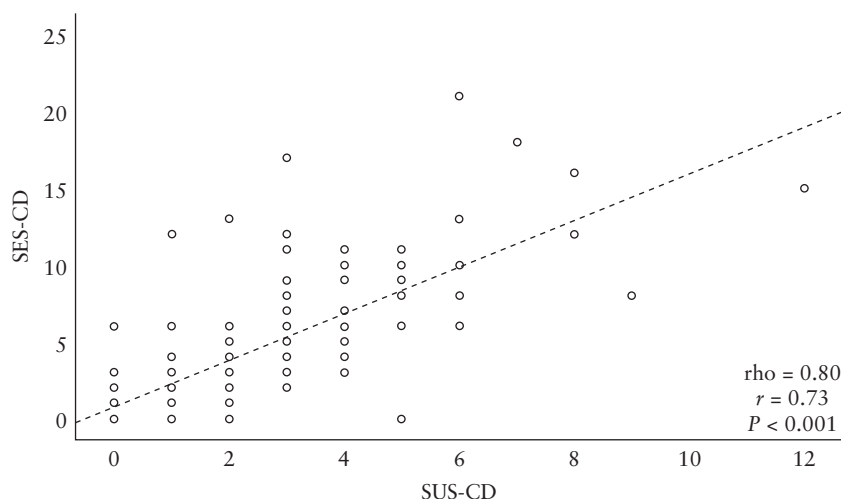


Figure 3. Correlation between endoscopy [SES-CD] and ultrasound score [SUS-CD] for the entire cohort. The regression line for SES-CD is depicted as $0.731 + 1.517 \times \text{SUS-CD}$, as well as Pearson's [r] and Spearman's rank correlations [rho].

an area under the curve of 0.88, where a cut-off of $\text{SUS-CD} \geq 3$ had a sensitivity and specificity of 88.5% and 69.0%, respectively.

4. Discussion

Ultrasonography is often used for follow-up examinations in patients with CD, as it is readily available, non-invasive, relatively inexpensive and offers high repeatability. Although focused bowel sonography may improve the clinical evaluation significantly,¹⁹ its usefulness in follow-up examinations has not been fully established, which may be due to operator dependency or lack of validated, reproducible scoring systems. In this study, we developed and validated a simple ultrasound activity index for measuring disease activity in CD. The ultrasound score correlates well with ileocolonoscopy in both patient cohorts, suggesting that it is a useful surrogate marker of endoscopic activity. Moreover, the activity index seems reproducible as low interobserver variability was revealed, and because we included a heterogeneous CD population, it seems applicable at different disease stages.

A standardized activity index may contribute to the initial assessment in CD; still, its optimal use is to monitor the same patient over time to determine whether the activity is decreasing or increasing. Although several attempts have been made to construct an ultrasonographic activity index for CD, the methodology for development was insufficient in most studies.²⁷ The current best method was developed and validated by Novak *et al.*²⁶ However, in the development phase, no endoscopic score was used, seven ulcerative colitis patients were included and it was conducted retrospectively. Moreover, in the validation phase, two endoscopic scoring systems were used, the cut-off for SES-CD was liberal [$\text{SES-CD} > 5$], and the ultrasonographer was not blinded to the results from the corresponding ileocolonoscopy. In our study, we sought to identify and overcome the limitations of previous scoring systems. The methodology for construction and validation of the SUS-CD was performed similarly as for the development of the SES-CD,⁷ as previously suggested.²⁷

MRI is commonly preferred in initial diagnosis as it is well suited for mapping extent and enables a better depiction of proximal and pelvic lesions than US.³⁷ Previous reports demonstrate its ability for grading disease activity³⁸ and several activity indices are available.³⁹ Still, due to low availability, it cannot be performed frequently.

Furthermore, in most MRI protocols, gadolinium-based contrast agents are required, which may deposit in brain tissues.⁴⁰ Although there are no currently proven harmful effects, it should be treated with caution and avoided when not necessary.⁴⁰ Thus, an MRI-based approach may be less applicable for frequent activity monitoring. Moreover, as a recent study found good agreement between US and MRI in guiding clinical decision-making,⁴¹ we suggest that GIUS should be used as a first-line tool during follow-up.

The components of the activity index were selected using multiple linear regression to identify significant contributing variables. The length of the affected segment had highest correlation with SES-CD and provided important predictive contributions to the model. Still, a substantial collinearity between length of the affected segment and BWT was found, and thus they cancelled each other out. We selected BWT to be incorporated in the activity score as it is considered the most important sonographic variable in CD and is considerably easier to measure than length. Stenosis had a minimal unique contribution to the model and was removed from the final score. On ultrasonography, it is detected as a segmental bowel wall thickening with a narrowed lumen and pre-stenotic dilatation. Increased BWT, as seen in active CD, could thus cancel out stenosis, which may explain its poor contribution to the model. Loss of bowel wall stratification and fatty wrapping had higher, but not significant, unique contributions to the model. They were included in the assessment of inter-rater reliability, but due to poor reliability they were excluded from the final score.

Re-examination of the development cohort showed good agreement between the investigators for BWT and colour Doppler. There was, however, poorer agreement for stratification and fatty wrapping. BWT and colour Doppler are quantifiable measurements, which could explain why these parameters were the easiest to reproduce. In contrast, stratification and fatty wrapping are more subjective quantities, which seems to affect reproducibility. These results are in concordance with previous reports.^{42,43} In both patient cohorts, the interobserver analyses revealed good agreement for calculating the activity index.

A clinically applicable activity score should be accurate, reproducible and easy to use. Although the ultrasound score had a slightly higher correlation with SES-CD when including all variables, stratification and fatty wrapping seem prone to different interpretation

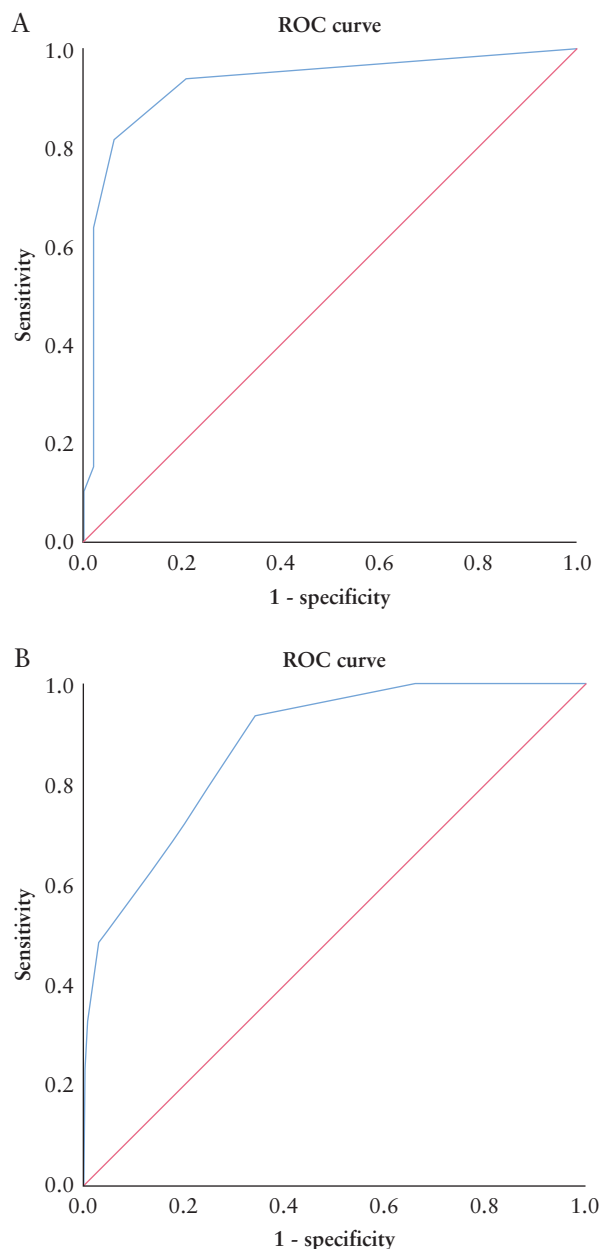


Figure 4. [A] ROC curve demonstration of the ability of SUS-CD to predict endoscopic activity defined as SES-CD > 2. By using a cut-off of SUS-CD ≥ 1 , a sensitivity of 95.3%, a specificity of 70.2% and an area under the curve of 0.92 are achieved. [B] ROC curve demonstration of the ability of SUS-CD to predict moderate endoscopic activity defined as SES-CD ≥ 7 . By using a cut-off of SUS-CD ≥ 3 , a sensitivity of 88.5%, a specificity of 69.0% and an area under the curve of 0.88 are achieved.

and were thus excluded from the final index. Accordingly, the final score is constituted by BWT and colour Doppler as they contributed significantly to the calculated model, correlated well with ileocolonoscopy and revealed excellent reproducibility. These results correspond to those of Novak *et al.*,²⁶ although differences in design and methodology exist. By excluding complications, length, stratification and fatty wrapping, the ultrasound score does not include some important features of CD. However, the trade-off yields a reliable, reproducible and easy-to-use tool during follow up. The

excluded parameters may instead serve as additional modifiers when present.

CD patients are regularly monitored to identify changes in disease activity, and at outpatient clinics, decision-making is usually limited to clinical assessment accompanied by biochemical tests. We found poor correlations between both the endoscopic and the ultrasonographic scores and clinical and biochemical markers, suggesting that the latter tests do not sufficiently reflect the degree of activity. The accuracy of clinical and biochemical activity reflected by HBI and CRP is limited, as previously demonstrated.^{2,44,45} Still, faecal calprotectin offers good diagnostic accuracy for detecting endoscopically active CD, although higher accuracies are achieved in ulcerative colitis.^{10,44} In our study, the patient compliance for delivering faecal samples was poor [40% missing data], and the majority of patients had only terminal ileitis, which could explain the disparities from previous reports.

Overall, there was a good correlation between SES-CD and SUS-CD, suggesting that the activity index may be a useful supplement to endoscopy during follow-up. Still, the ultrasound score has some limitations: it has not been developed for evaluating proximal small bowel lesions as these exceed the reach of ileocolonoscopy. Moreover, the ultrasound score may not be applicable in all patients with a generous body habitus due to reduced visibility and Doppler sensitivity.³¹ Vendor-specific differences in Doppler quality may potentially restrict widespread use of the ultrasound score. Still, this is a general limitation of the Doppler technique and is not limited to the SUS-CD score. Furthermore, mismatches between the modalities may occur as GIUS does not seem sensitive enough to detect aphthous lesions, and BWT could increase due to chronic changes after surgical resection. Missing segments due to surgical resections are not included in the score, which could give the patient a lower value. Finally, as increased BWT also occurs in fibrosis, the SUS-CD may erroneously classify fibrotic segments as inflamed lesions, putting patients at risk of receiving incorrect treatment. Thus, additional diagnostic methods may be necessary in the case of inappropriate treatment effects.

The study has some limitations. As ileocolonoscopy was performed as part of standard care, the examinations were conducted by several endoscopists, and no formal consensus regarding SES-CD calculation was performed. This may lead to differences in endoscopic assessments. Further, re-examination of the development cohort was performed on identical images and video loops, which may contribute to good results. Still, the interobserver analysis in the validation phase performed by two independent investigators revealed similar results. The present study comprises the development and validation of an ultrasound activity index using different patient cohorts. The scoring system should be validated by other groups and tested for responsiveness to changes in disease activity in future studies. Moreover, the SUS-CD should be compared with an MRI-based activity index, and it should be included in follow-up studies to determine its usefulness in clinical decision-making.

In conclusion, we have constructed and validated a simple ultrasound score for CD that correlates well with endoscopic activity. Implementation of the activity index may aid the monitoring and management of CD patients and could potentially reduce the need for endoscopic examinations.

Conference

Euroson 2017, Ljubljana, Slovenia; Euroson 2018, Poznan, Poland.

Funding

The work received no external funding. A travel grant for visiting the collaboration centre was approved.

Conflicts of Interest

F.S., R.E., G.E.E.: No declared conflicts of interest. O.H.G.: Has received speaker honoraria from the following companies: AbbVie, Bracco, Almirall, GE Healthcare, Takeda AS, Meda AS, Ferring AS and Allergan. He has served as consultant for Bracco, GE Healthcare, Takeda and Samsung. K.N.: Has served as speaker Takeda.

Author Contributions

K.N., F.S., O.H.G. and G.E.E. contributed to the concept and design of the study; F.S., R.E. and K.N. made substantial contributions to patient recruitment and acquisition of material; F.S. and K.N. made substantial contributions to analysis and interpretation of data; G.E.E. provided statistical support; F.S. drafted the manuscript; G.E.E., K.N., R.E. and O.H.G. critically revised the manuscript; all authors approved the final version of the manuscript.

Acknowledgments

We are indebted to the personnel at the Department of Medicine for helping with logistics and the gastroenterologists at both hospitals for providing reference standard assessments.

Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

References

- Maaser C, Sturm A, Vavricka SR, *et al.*; European Crohn's and Colitis Organisation [ECCO] and the European Society of Gastrointestinal and Abdominal Radiology [ESGAR]. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019;13:144–64.
- Peyrin-Biroulet L, Reinisch W, Colombel JF, *et al.* Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. *Gut* 2014;63:88–95.
- 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 Thérapeutique des Affections Inflammatoires Digestives. *Gastroenterology* 1990;98:811–8.
- Gomollón F, Dignass A, Annesse V, *et al.*; ECCO. 3rd European Evidence-based Consensus on the Diagnosis and Management of Crohn's Disease 2016: Part 1: Diagnosis and Medical Management. *J Crohns Colitis* 2017;11:3–25.
- Shah SC, Colombel JF, Sands BE, Narula N. Systematic review with meta-analysis: mucosal healing is associated with improved long-term outcomes in Crohn's disease. *Aliment Pharmacol Ther* 2016;43:317–33.
- Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Thérapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). *Gut* 1989;30:983–9.
- 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.
- Sturm A, Maaser C, Calabrese E, *et al.*; European Crohn's and Colitis Organisation [ECCO] and the European Society of Gastrointestinal and Abdominal Radiology [ESGAR]. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 2: IBD scores and general principles and technical aspects. *J Crohns Colitis* 2019;13:273–84.
- Walsh AJ, Bryant RV, Travis SP. Current best practice for disease activity assessment in IBD. *Nat Rev Gastroenterol Hepatol* 2016;13:567–79.
- Mosli MH, Zou G, Garg SK, *et al.* C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. *Am J Gastroenterol* 2015;110:802–19; quiz 820.
- Rimola J, Rodriguez S, Garcia-Bosch O, *et al.* Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease. *Gut* 2009;58:1113–20.
- Rimola J, Ordás I, Rodriguez S, *et al.* Magnetic resonance imaging for evaluation of Crohn's disease: validation of parameters of severity and quantitative index of activity. *Inflamm Bowel Dis* 2011;17:1759–68.
- Ordás I, Rimola J, Rodríguez S, *et al.* Accuracy of magnetic resonance enterography in assessing response to therapy and mucosal healing in patients with Crohn's disease. *Gastroenterology* 2014;146:374–82.e1.
- 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.
- 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.
- Oussalah A, Laurent V, Bruot O, *et al.* Diffusion-weighted magnetic resonance without bowel preparation for detecting colonic inflammation in inflammatory bowel disease. *Gut* 2010;59:1056–65.
- Buisson A, Joubert A, Montoriol PF, *et al.* Diffusion-weighted magnetic resonance imaging for detecting and assessing ileal inflammation in Crohn's disease. *Aliment Pharmacol Ther* 2013;37:537–45.
- Dong J, Wang H, Zhao J, *et al.* Ultrasound as a diagnostic tool in detecting active Crohn's disease: a meta-analysis of prospective studies. *Eur Radiol* 2014;24:26–33.
- Novak K, Tanyingoh D, Petersen F, *et al.* Clinic-based point of care transabdominal ultrasound for monitoring Crohn's disease: impact on clinical decision making. *J Crohns Colitis* 2015;9:795–801.
- Kucharzik T, Wittig BM, Helwig U, *et al.*; TRUST study group. Use of intestinal ultrasound to monitor Crohn's disease activity. *Clin Gastroenterol Hepatol* 2017;15:535–542.e2.
- Drews BH, Barth TF, Hänle MM, *et al.* Comparison of sonographically measured bowel wall vascularity, histology, and disease activity in Crohn's disease. *Eur Radiol* 2009;19:1379–86.
- 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.
- Paredes JM, Ripollés T, Cortés X, *et al.* Contrast-enhanced ultrasonography: usefulness in the assessment of postoperative recurrence of Crohn's disease. *J Crohns Colitis* 2013;7:192–201.
- Paredes JM, Ripollés T, Cortés X, *et al.* Non-invasive diagnosis and grading of postsurgical endoscopic recurrence in Crohn's disease: usefulness of abdominal ultrasonography and (99m)Tc-hexamethylpropylene amineoxime-labelled leucocyte scintigraphy. *J Crohns Colitis* 2010;4:537–45.
- Neye H, Voderholzer W, Rickes S, Weber J, Wermke W, Lochs H. Evaluation of criteria for the activity of Crohn's disease by power Doppler sonography. *Dig Dis* 2004;22:67–72.
- Novak KL, Kaplan GG, Panaccione R, *et al.* A simple ultrasound score for the accurate detection of inflammatory activity in Crohn's disease. *Inflamm Bowel Dis* 2017;23:2001–10.
- Bots S, Nylund K, Löwenberg M, Gecke 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.
- Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006;55:749–53.
- Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980;1:514.
- Vuitton L, Marteau P, Sandborn WJ, *et al.* IOIBD technical review on endoscopic indices for Crohn's disease clinical trials. *Gut* 2016;65:1447–55.

31. Nylund K, Maconi G, Hollerweger A, et al. EFSUMB recommendations and guidelines for gastrointestinal ultrasound. *Ultraschall Med* 2017;**38**:e1–e15.
32. Maconi G, Nylund K, Ripolles T, et al. EFSUMB recommendations and clinical guidelines for intestinal ultrasound (GIUS) in inflammatory bowel diseases. *Ultraschall Med* 2018;**39**:304–17.
33. 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.
34. 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.
35. Spalinger J, Patriquin H, Miron MC, et al. Doppler US in patients with Crohn disease: vessel density in the diseased bowel reflects disease activity. *Radiology* 2000;**217**:787–91.
36. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.
37. Taylor SA, Mallett S, Bhatnagar G, et al.; METRIC study investigators. Diagnostic accuracy of magnetic resonance enterography and small bowel ultrasound for the extent and activity of newly diagnosed and relapsed Crohn's disease (METRIC): a multicentre trial. *Lancet Gastroenterol Hepatol* 2018;**3**:548–58.
38. Puylaert CA, Tielbeek JA, Bipat S, Stoker J. Grading of Crohn's disease activity using CT, MRI, US and scintigraphy: a meta-analysis. *Eur Radiol* 2015;**25**:3295–313.
39. Rimola J, Alvarez-Cofiño A, Pérez-Jeldres T, et al. Comparison of three magnetic resonance enterography indices for grading activity in Crohn's disease. *J Gastroenterol* 2017;**52**:585–93.
40. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB; International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* 2017;**16**:564–70.
41. Allocca M, Fiorino G, Bonifacio C, et al. Comparative accuracy of bowel ultrasound versus magnetic resonance enterography in combination with colonoscopy in assessing Crohn's disease and guiding clinical decision-making. *J Crohns Colitis* 2018;**12**:1280–7.
42. Calabrese E, Kucharzik T, Maaser C, et al. Real-time interobserver agreement in bowel ultrasonography for diagnostic assessment in patients with Crohn's disease: an international multicenter study. *Inflamm Bowel Dis* 2018;**24**:2001–6.
43. Fraquelli M, Sarno A, Girelli C, et al. Reproducibility of bowel ultrasonography in the evaluation of Crohn's disease. *Dig Liver Dis* 2008;**40**:860–6.
44. Schoepfer AM, Beglinger C, Straumann A, et al. Fecal calprotectin correlates more closely with the Simple Endoscopic Score for Crohn's disease [SES-CD] than CRP, blood leukocytes, and the CDAL. *Am J Gastroenterol* 2010;**105**:162–9.
45. Ricanek P, Brackmann S, Perminow G, et al.; IBSEN II Study Group. Evaluation of disease activity in IBD at the time of diagnosis by the use of clinical, biochemical, and fecal markers. *Scand J Gastroenterol* 2011;**46**:1081–91.