Original Article

A Reliability Study: Strong Inter-Observer Agreement of an Expert Panel for Intestinal Ultrasound in Ulcerative Colitis

Floris De Voogd,[a,] **Rune Wilkens,[b,]** **Krisztina Gecse,[a]** **Mariangela Allocca,[c]** **Kerri Novak,[d]** **Cathy Lu,[d]** **Geert D'Haens,[a]** **Christian Maaser[e]**

[a]Amsterdam University Medical Center, Department of Gastroenterology and Hepatology, Amsterdam, The Netherlands [b]Gastrounit, Division of Medicine, Hvidovre Hospital, Copenhagen, Denmark [c]Department of Gastroenterology, Humanitas University, Milan, Italy [d]Division of Gastroenterology, Department of Medicine, University of Calgary, Alberta, Canada [e]Outpatient's Department of Gastroenterology, University Teaching Hospital Lueneburg Germany

Corresponding author: Dr F. de Voogd, Amsterdam University Medical Centers, Department of Gastroenterology and Hepatology, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands. Email: f.a.devoogd@amsterdamumc.nl

Abstract

Background: Intestinal ultrasound [IUS] is a promising and non-invasive cross-sectional imaging modality in the diagnosis and monitoring of ulcerative colitis [UC]. Unlike endoscopy, where standardized scoring for evaluation of disease activity is widely used, scoring for UC with IUS is currently unavailable. Therefore, we conducted a study to assess the reliability of IUS in UC among expert sonographists in order to identify robust parameters.

Methods: Thirty patients with both clinically active [25] and quiescent [five] UC were included. Six expert sonographers first agreed upon key IUS parameters and grading, including bowel wall thickness [BWT], colour Doppler signal [CDS], inflammatory fat [i-fat], loss of bowel wall stratification [BWS], loss of haustrations and presence of lymph nodes. Thirty video-recorded cases were blindly reviewed.

Results: Inter-observer agreement was almost perfect for BWT (intra-class correlation coefficient [ICC]: 0.96) and substantial for CDS [$\kappa = 0.63$]. Agreement was moderate for presence of lymph nodes [$\kappa = 0.41$] and fair for presence of i-fat [$\kappa = 0.36$], BWS [$\kappa = 0.24$] and loss of haustrations [$\kappa = 0.26$]. Furthermore, there was substantial agreement for presence of disease activity on IUS [$\kappa = 0.77$] and almost perfect agreement for disease severity [ICC: 0.93]. Most individual parameters showed a strong association with IUS disease activity as measured by the six readers.

Conclusion: IUS is a reliable imaging modality to assess disease activity and severity in UC. Important individual parameters such as BWT and CDS are reliable and could be incorporated in a future UC scoring index. Standardized acquisition and assessment of UC utilizing IUS with established reliability is important to expand the use of IUS globally.

Key Words: Intestinal ultrasound; ulcerative colitis; reliability; inter-observer agreement

1. Introduction

Ulcerative colitis [UC] is a chronic inflammatory bowel disease [IBD] affecting the mucosal and potentially submucosal layers of the colonic bowel wall, characterized by a relapse–remitting pattern.[1]

Close monitoring of disease activity is one of the key strategies in the treatment of UC to prevent relapse and disease progression. However, patient-reported symptoms do not always reflect true inflammation.[2] Therefore, biochemical parameters, endoscopic

assessment of disease activity and cross-sectional imaging techniques contribute to the close monitoring and treatment decision process in UC.[1,3] To facilitate frequent monitoring, the involved modalities ideally would be non-invasive, accurate, reliable and feasible. Intestinal ultrasound [IUS], which requires no preparation, exhibits these qualities and is therefore an appealing modality. In addition, IUS can directly visualize the bowel in a point-of-care setting, allows close follow-up and detects treatment response within the first 2–6 weeks.[4] Furthermore, it correlates well with both endoscopic outcomes and other cross-sectional imaging techniques, indicating the high accuracy of IUS.[5–7] Although interest in implementing IUS is increasing, actual use is currently limited.[8,9] A recent study found the assumed poor reliability and operator dependence as a predominant limiting factor.[8] However, data on the reliability of IUS and individual parameters are scarce.

Reliability depends on the variance between measurements of observers and is generally improved by clear definitions and a standardized way of scoring.[10] For both UC and Crohn's disease [CD], multiple endoscopic scoring indices have been created and validated with moderate to good inter-observer agreement.[11,12] For IUS multiple scoring indices are under development in CD and UC but none has been externally validated.[7,13,14] However, there is a pressing need for a scoring index because IUS is increasingly practised and could potentially be incorporated as an endpoint in clinical trials.[15]

In order to incorporate individual parameters in a future scoring index, the inter-observer agreement should be determined for all relevant parameters. The most important parameter assessed by IUS is bowel wall thickness [BWT].[7] Combined with colour Doppler signal [CDS], the presence of inflammatory fat [i-fat], lymphadenopathy, loss of bowel wall stratification [BWS] and loss of haustration, disease activity and severity can be determined.[16–18] A recent study in CD found moderate to substantial agreement among six expert raters for most of the aforementioned IUS parameters in CD.[19] Another study found good agreement between two raters to detect disease activity in UC patients.[5] However, further data on inter-observer agreement for IUS, especially in UC, are lacking.

Hence, the aim of this study was to determine inter-observer agreement for IUS parameters and overall disease activity in UC, thereby providing robust indicators to incorporate in a future scoring index.

2. Methods

2.1. Patient selection

In this prospective, observational study, 30 patients with a confirmed diagnosis of UC were recruited in three dedicated IBD centres [Germany, The Netherlands and Denmark]. All IUS examinations were performed as part of routine care and were collected and anonymized with patient approval. This study was assessed and approved by the medical ethics committee at the Amsterdam Medical Center. Twenty-five patients with clinical active disease (defined as a simple clinical colitis activity index [SCCAI] ≥ 5) and five patients with clinical quiescent disease were included. Patients with clinical active disease were only included if disease was not limited to the rectum in their medical history.

2.2. Ultrasonographic examinations

IUS acquisition was performed using a standardized protocol with different high-end ultrasound machines [Philips EPIQ 5G, General Electric Logic S8, Hitachi EUB-8500 and General Electric Logic E9] using an abdominal convex [1–8/1–6 MHz] and linear [3–18/2.5–8 MHz] probe. There was no specific preparation nor was fasting required. The patient was in the supine position and the complete colon was visualized starting with the sigmoid colon [SC] that was identified as the colonic segment overlying the left iliac vessels and iliopsoas muscle. A cine-loop of the SC was recorded in a longitudinal and cross-sectional scan plane for both B-mode and CDS covering both the intestine and the mesentery. Cine-loops of the SC were used for analysis as the SC is generally always involved in UC extending beyond proctitis, and location is very easy to reproduce owing to clear landmarks and is predominantly the worst segment in the absence of local treatment.[7] Subsequently, cine-loops were uploaded in DICOM format to an in-house build central reading platform and scored using DICOM-viewer software [RadiAnt DICOM Viewer, Medixant, or Horos]. Measurements were submitted to an electronic case reporting form.

2.3. Ultrasonographic parameters

In an online meeting before case-rating, consensus was reached on the cut-off values for all IUS parameters. BWT was measured from the lumen–mucosa interface up to the muscularis propria–serosa interface. Intestinal folds were avoided within the measurements. Measurement of BWT was defined as the mean of four measurements [to the nearest 0.1 mm] at the thickest part of the SC: two measurements in a longitudinal scan plane with at least 1 cm between the two measurements and two measurements in a cross-sectional scan plane with a minimum of 90° between the two measurements [Figure 1].

CDS was optimized for slow flow with a maximum velocity scale of 5–7 cm/s and was initially scaled up to its maximum and subsequently downscaled until noise fades away. Hypervascularization was graded as a modified Limberg score,[20] in four categories: [0] absent, [1] small spots [single vessels] within the wall, [2] long stretches within the wall, and [3] long stretches within the wall extending into
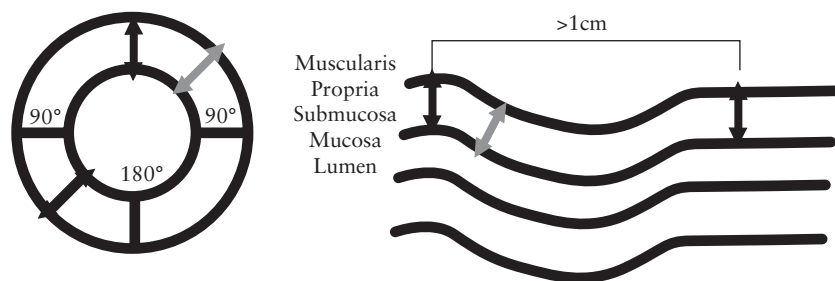
Figure 1. The black arrow indicates the first correct measurement, grey arrow indicates incorrect second measurement, and black arrow indicates correct second measurement.

the mesentery. BWS was graded as preserved, uncertain, focal loss [< 3 cm in length] or extensive loss of stratification. Haustrations and i-fat were graded as present, uncertain or absent. Lymph nodes were graded as absent or present and subsequently categorized as < 5.0 mm, between 5.0 and 9.9 mm or ≥ 10.0 mm as measured in the shortest axis [Table 1; Supplementary Figures 1–8 and Videos 1–4].

A BWT > 3.0 mm, CDS ≥ category 2, loss of BWS [both focal and extensive], loss of haustrations and presence of i-fat were all agreed to indicate disease activity [Table 1].

Furthermore, IUS disease activity and severity was scored by the six raters. IUS disease activity was scored as present when at least a BWT > 3 mm was measured together with another parameter being pathologic. Disease severity was scored using an 11-point Likert scale where 0 reflects normal bowel and 10 reflects most severe disease ever seen.

2.4. Ultrasonographic rating

Six raters from Canada, Denmark, Germany, Italy and The Netherlands centrally read all cine-loops of 30 patients blinded to clinical and endoscopic disease activity and the other raters' evaluation. There was ≥ 16 weeks between the collection of the cine-loops and central reading of cine-loops for all patients to ensure a sufficient wash-out period for the investigator who recorded the cineloops. All raters were sub-specialized in IBD, except for one GI-Fellow in final year [R.W.]. All investigators had at least 3 years of experience in IUS [C.M. and M.A. 15 years, R.W. 8 years, K.N. 7 years, C.L. and K.G. 3 years]. To assess difference in BWT inter-observer reliability for years of experience, we created three groups: group 1: 3 years of experience, group 2: 7–8 years of experience and group 3: 15 years of experience].

2.5. Statistics

Inter-observer reliability was analysed per parameter for the SC [SPSS Statistics, Version 25.0, 2017, IBM Corp. and Stata/SE 16.1, Stata Corp LP]. For all continuous parameters, the mean-rating [$k = 6$], absolute agreement of a two-way mixed effects intra-class correlation coefficient [ICC] was used.[21] For all categorical data, weighted Fleiss' kappa for multiple raters was computed. Both ICC and kappa statistics are interpreted as previously established.[22] For association with IUS disease severity, the mean value of the six raters was used for each IUS parameter, to compute Spearman's rank correlation coefficient. A P value < 0.05 was considered significant. The difference in BWT and ICC between groups was assessed using Kruskall–Wallis one-way ANOVA and a Mann–Whitney U test.

3. Results

From the 25 patients with active disease, nine had an acute severe UC and were admitted to the hospital for intravenous treatment, and three received a colectomy within a month after baseline IUS [Table 2].

3.1. Inter-observer reliability per IUS parameter

There was almost perfect reliability for BWT (ICC: 0.96, 95% confidence interval [CI]: 0.92–0.98] and BWT and ICC did not differ significantly according to years of experience [Figures 2 and 3]. CDS [all categories] showed substantial agreement [κ = 0.63, 95% CI: 0.49–0.77]. When CDS was divided into two broad categories [not present or present], there was an almost perfect agreement [κ = 0.83, 95% CI: 074–0.92]. Fair agreement was found for the presence of i-fat [κ = 0.36, 95% CI: 0.17–0.55]. Fair agreement was found for the presence of lymph nodes [κ = 0.36, 95% CI: 0.20–0.51] for the initial four categories and improved to a moderate agreement when dichotomizing this parameter to present or not present [κ = 0.41, 95% CI: 0.29–0.53]. There were no lymph nodes > 10 mm detected in this cohort. BWS [κ = 0.25, 95% CI 0.06–0.44] and loss of haustration [κ = 0.24, 95% CI: 0.13–0.35] showed a fair agreement.

3.2. Inter-observer reliability on overall IUS disease activity

There was substantial agreement for IUS-defined disease activity [Yes/No] among the six raters [κ = 0.77, 95% CI: 0.65–0.88]. For IUS disease severity [graded from 0 to 10] the agreement was almost perfect [ICC: 0.93, 95% CI: 0.87–0.96].

Mean BWT was significantly higher in the 25 patients with clinical active disease than in the group with clinically quiescent disease [4.6 ± 1.0 vs 2.1 ± 0.6 mm, $P < 0.0001$]. Furthermore, mean BWT showed a strong association with the mean IUS disease severity as scored by the six raters [ρ = 0.75, $P < 0.0001$] [Supplementary Figure 9]. For the other parameters, the correlation with IUS disease severity is shown in Supplementary Figure 9. Acute severe UC patients examined at admission were not rated worse than patients with active disease [5.8 ± 1.9 vs 5.3 ± 1.8 points, $p = 0.47$]. However, among the clinical active cases, those who underwent colectomy [$n = 3$] during admission had a trend towards increased IUS severity than those who did not receive a colectomy [7.3 ± 1.2 vs 5.2 ± 1.8 points, $p = 0.06$]. Mean BWT was not significantly increased in patients undergoing a colectomy [5.3 ± 1.1 vs 4.5 ± 1.0 mm, $p = 0.31$].

Table 1. Overview of IUS parameters, techniques and cut-off values

IUS parameter	Technique/categories	Pathologic
BWT	[2 × longitudinal plane + 2 × cross-sectional plane]/4	BWT > 3.0 mm
CDS	0: absent; 1: small spots [single vessels] within the wall; 2: long stretches within the wall; 3: long stretches extending into the mesentery	Grade 2 or 3
BWS	0: preserved; 1: uncertain [in doubt of disturbance of wall layers]; 2: focal loss [< 3 cm in length within the SC]; 3: extensive loss [≥ 3 cm in length within the SC]	Grade 2 or 3
Loss of haustration	0: preserved; 1: uncertain [in doubt of complete loss of haustrations]; 2: loss	Grade 2
i-fat	0: absent; 1: uncertain [in doubt of hyperechogenicity around the segment]; 2: present	Grade 2
Lymph nodes	0: absent; 1: present and < 5 mm in shortest axis; 2: present and 5–9.9 mm in shortest axis; 3: present and ≥ 10 mm in the shortest axis	Undetermined

IUS: intestinal ultrasound, BWT: bowel wall thickness, CDS: colour Doppler signal, BWS: bowel wall stratification, SC: sigmoid colon, i-fat: inflammatory fat.

4. Discussion

In our study, there was almost perfect agreement among expert raters for BWT, indicating high reliability for this important parameter. These findings are in line with a previous study in CD patients.[23] However, a more recent study in CD patients found a moderate interobserver agreement for BWT.[19] This difference in agreement could be affected by the technique of measuring and statistical analysis. In our study and the study by Fraquelli *et al.*, a mean of multiple measurements was calculated with a continuous outcome. Conversely in the study by Calabrese *et al.*, only one measurement was taken and reported as a dichotomous outcome [3–7 vs > 7 mm]. This might suggest that reliability improves when BWT is calculated by multiple measurements and reported as a continuous outcome.

In addition, we have shown equal intra-class correlation coefficients for 3, 7–8 and 15 years of experience. These results indicate that less experienced ultrasonographers perform equal BWT measurements as the more experienced ultrasonographers and that the most significant learning curve probably takes place in the first 3 years. Future studies should address this topic and investigate inter-observer agreement in ultrasonographers with less than 3 years of experience.

In the current study, hypervascularity shows substantial agreement, indicating good reliability for the four categories. Previous studies also incorporated hypervascularity to assess disease activity.[5,6,13] A recent large study found hypervascularity improves even

within 2 weeks after initiation of anti-inflammatory treatment and to be a marker of early treatment response.[4] Therefore, CDS improvement could be a reliable and responsive parameter in UC.

In our study, we identified several reliable IUS parameters. In addition, the overall agreement of presence and severity of disease on IUS was substantial to almost perfect, indicating high reliability in UC patients. Therefore, IUS is a promising and attractive cross-sectional imaging modality to frequently assess inflammation in UC with the ability to monitor treatment effect tightly. According to recent studies, IUS is a reliable technique to assess inflammation in IBD patients.[7,24] A systematic review in UC showed high accuracy for IUS to detect disease presence and extent when compared to endoscopic outcomes and other cross-sectional imaging techniques.[7] With regard to IUS disease severity, previous studies showed substantial similarity with endoscopic severity.[7] A next step towards incorporation of IUS in standard care is the development of a reliable and validated scoring index.[15]

In recently developed IUS disease activity scores for UC, BWT and CDS were the predominant parameters.[7,13] Although BWT is a reliable parameter, there is still debate on the most accurate cut-off value for BWT to distinguish active from quiescent disease. A recent systematic review concluded a BWT of 4.0 mm was most accurate[7] while other studies found lower cut-off values.[5,13,25,26] Determination of the most accurate cut-off value for BWT is of great importance to detect true inflammation and true quiescent disease. As the debate continues, BWT was interpreted as a continuous variable to prevent the bias of a predefined cut-off value in our reliability analysis.

In addition, BWS, loss of haustration and i-fat were also incorporated in previous scoring indices for UC.[7,13] Although these individual parameters showed fair agreement in the current study, these parameters should not be excluded from incorporation in a future scoring index. Although less prominent than in CD, proliferation of i-fat occurs in UC.[27,28] A recently developed IUS score in paediatric IBD incorporated presence of mesenteric i-fat to indicate severity of disease in both CD and UC.[29] Furthermore, proliferation of i-fat, loss of BWS and haustration improves after initiation of anti-inflammatory therapy in UC patients.[4]

The presence of mesenteric lymphadenopathy might also indicate disease activity and showed moderate agreement. These findings correspond with findings in previous reliability studies in CD.[19,23]

Table 2. Baseline characteristics

Characteristics	N = 30
Age, years [mean ± SD]	35.1 ± 15.1
Gender [Male]	18 [60%]
Clinical disease activity [SCCAI]	
• Remission	5 [17%]
• Active disease	25 [83%]
• Acute severe UC	9 [36%]
Disease location [E2/E3]	22 [73%]/8 [27%]

SD = standard deviation, SCCAI: Simple Clinical Colitis Activity Index, UC = ulcerative colitis, E2 = left-sided colitis, E3 = pancolitis.
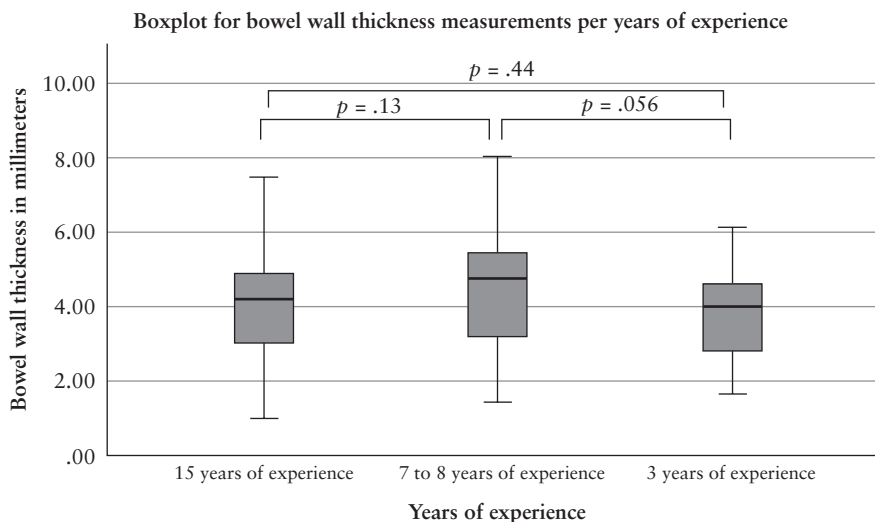
Figure 2. Bowel wall thickness [BWT] measurements per category years of experience [15 years of experience: mean BWT: 4.22 ± 1.58 mm, 7–8 years of experience: mean BWT: 4.47 ± 1.59 mm, 3 years of experience: mean BWT: 3.82 ± 1.26 mm].

Boxplot for intra-class correlation coefficients per yerars of experience
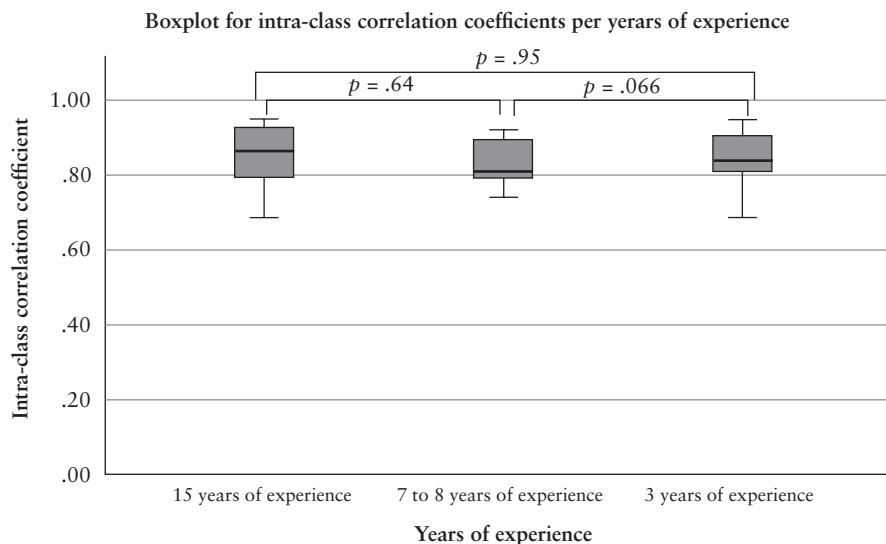
Figure 3. Intra-class correlation coefficients per category years of experience [15 years of experience: mean ICC: 0.84, 7–8 years of experience: mean ICC: 0.83, 3 years of experience: mean ICC: 0.84].

Furthermore, mesenteric lymphadenopathy correlates with endoscopic disease activity in UC[5] and improves when patients show clinical response on anti-inflammatory treatment, indicating a role for mesenteric lymphadenopathy in the assessment of treatment response in UC.[4] However, previous studies have not incorporated this parameter in a scoring index and studies analysing the role of presence of enlarged lymph nodes are scarce.[7,13] Although mesenteric lymphadenopathy might not be specific for the presence of UC, future studies should take the presence of enlarged lymph nodes into consideration as a potential marker to indicate disease activity and monitor disease presence.

Analysing inter-observer agreement for cross-sectional imaging techniques of the bowel has predominantly been studied for CD. For computed tomography, substantial to good inter-observer agreement was found for disease presence and individual parameters in CD.[30,31] For magnetic resonance enterography, fair to good agreement for individual parameters and good inter-observer agreement for multiple scoring indices was found.[32–34] Because we found substantial to almost perfect inter-observer agreement to detect disease activity, the reliability of IUS could be considered as at least as good as other cross-sectional imaging techniques.

This study has some limitations. We did not assess the correlation between IUS findings and endoscopic disease activity. However, previous studies have shown good correlation between BWT and endoscopic disease activity, and the primary aim of this study was to assess the reliability of IUS parameters in UC.[5,35,36] We only assessed the SC as this segment is most often and most severely involved in UC extending from the rectum.[7] However, diagnostic odds ratios for IUS differ among colonic segments and especially for the rectum in IBD.[37] Similarly, inter-observer agreement might also differ among colonic segments; hence our results cannot immediately be extrapolated to other colonic segments, especially for the rectum. Visualization of the rectum is more difficult than other colonic segments due to its location deep in the pelvis. Consequently, analysis of inter-observer agreement for disease activity in the rectum was not feasible and transabdominal IUS is not the preferred technique to assess proctitis.[3] However, transperineal IUS is an evolving area and could potentially be successfully used to evaluate proctitis and should be a focus of future studies.[38] Inter-observer agreement

was studied among sonographers with at least 3 years of experience and may vary more among less experienced sonographers. Although the six readers had different levels of expertise ranging from 3 to 15 years, future studies should consider including novice sonographers to determine the learning-curve. A general limitation of ultrasound is the operator-dependence in which poor acquisition of cine-loops negatively affects inter-observer agreement. To limit the bias of poor acquisition, recording was standardized by visualizing anatomical landmarks within the cine-loop.

In conclusion, IUS shows high reliability to determine disease activity and severity. Standardized assessment and scoring are of great importance to increase the use of IUS in UC and to reach high reliability of this non-invasive cross-sectional imaging modality. Future studies should focus on the development of an IUS scoring index, with BWT and CDS being reliable individual parameters to incorporate.

Funding

Conflicts of Interest

Author Contributions

F.V. [conceptualization, collection of data, methodology, statistical analysis, writing], R.W. [collection of data, central reading, methodology, statistical analysis, review final manuscript], K.G. [central reading, review final manuscript], M.A. [central reading, review final manuscript], K.N. [central reading, review final manuscript], C.L. [central reading, review final manuscript], G.D. [conceptualization, methodology, review final manuscript, supervision], C.M. [conceptualization, central reading, methodology, review final manuscript, supervision].

Supplementary Data

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

References

1. 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.

2. Peyrin-Biroulet L, Sandborn W, Sands BE, *et al*. Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. *Am J Gastroenterol* 2015;**110**:1324–38.

3. 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.

4. Maaser C, Petersen F, Helwig U, *et al*.; German IBD Study Group and the TRUST&UC study group; German IBD Study Group and TRUST&UC study group. Intestinal ultrasound for monitoring therapeutic response in patients with ulcerative colitis: results from the TRUST&UC study. *Gut* 2020;**69**:1629–36.

5. 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.

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.

7. Smith RL, Taylor KM, Friedman AB, Gibson RN, Gibson PR. Systematic review: Clinical utility of gastrointestinal ultrasound in the diagnosis, assessment and management of patients with ulcerative colitis. *J Crohns Colitis* 2020;**14**:465–79.

8. Gamboa HE, Molle-Rios Z, Anupindi SA. Underutilization of bowel ultrasound in North America in children with inflammatory bowel disease. *Digestive Diseases* 2020;**38**:390–7.

9. Maaser C, Kucharzik T, Gecse K. Is intestinal ultrasound ready to be used as standard monitoring tool in daily practice and as endpoint in clinical trials? *Journal of Crohn's and Colitis* 2021;**15**:1–2.

10. Kessler LG, Barnhart HX, Buckler AJ, *et al*.; QIBA Terminology Working Group. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res* 2015;**24**:9–26.

11. Travis SP, Schnell D, Krzeski P, *et al*. Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology* 2013;**145**:987–95.

12. 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**:969–73.

13. 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.

14. Sævik F, Eriksen R, Eide GE, Gilja OH, Nylund K. Development and validation of a simple ultrasound activity score for Crohn's disease. *J Crohns Colitis* 2021;**15**:115–24.

15. Kucharzik T, Maaser C, Maconi G. Do we need activity scores or simply clear criteria for intestinal ultrasound in ulcerative colitis? *J Crohns Colitis* 2018;**12**:1383–4.

16. Atkinson NSS, Bryant RV, Dong Y, *et al*. How to perform gastrointestinal ultrasound: anatomy and normal findings. *World J Gastroenterol* 2017;**23**:6931–41.

17. Bryant RV, Friedman AB, Wright EK, *et al*. Gastrointestinal ultrasound in inflammatory bowel disease: an underused resource with potential paradigm-changing application. *Gut* 2018;**67**:973–85.

18. Kucharzik T, Maaser C. Intestinal ultrasound and management of small bowel Crohn's disease. *Therap Adv Gastroenterol* 2018;**11**:1756284818771367.

19. 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.

20. Limberg B. Diagnosis of chronic inflammatory bowel disease by ultrasonography. *Z Gastroenterol* 1999;**37**:495–508.

21. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;**15**:155–63.

22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.

23. 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.

24. 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.

25. 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.

26. 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.

27. Eder P, Adler M, Dobrowolska A, Kamhieh-Milz J, Witowski J. The role of adipose tissue in the pathogenesis and therapeutic outcomes of inflammatory bowel disease. *Cells* 2019;**8**:628.

28. Kredel LI, Jödicke LJ, Scheffold A, *et al*. T-cell composition in ileal and colonic creeping fat–separating ileal from colonic Crohn's disease. *J Crohns Colitis* 2019;**13**:79–91.

29. Kellar A, Wilson S, Kaplan G, DeBruyn J, Tanyingoh D, Novak KL. The simple pediatric activity ultrasound score (SPAUSS) for the accurate detection of pediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2019;**69**:e1–6.

30. Booya F, Fletcher JG, Huprich JE, *et al*. Active Crohn disease: CT findings and interobserver agreement for enteric phase CT enterography. *Radiology* 2006;**241**:787–95.

31. Burlin S, Favaro LR, Bretas EA, *et al*. Using computed tomography enterography to evaluate patients with Crohn's disease: what impact does

examiner experience have on the reproducibility of the method? *Radiol Bras* 2017;**50**:13–8.

32. Tielbeek JA, Makanyanga JC, Bipat S, *et al*. Grading Crohn disease activity with MRI: interobserver variability of MRI features, MRI scoring of severity, and correlation with Crohn disease endoscopic index of severity. *AJR Am J Roentgenol* 2013;**201**:1220–8.

33. Jensen MD, Ormstrup T, Vagn-Hansen C, Østergaard L, Rafaelsen SR. Interobserver and intermodality agreement for detection of small bowel Crohn's disease with MR enterography and CT enterography. *Inflamm Bowel Dis* 2011;**17**:1081–8.

34. 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**:440–9.

35. Antonelli E, Giuliano V, Casella G, *et al*. Ultrasonographic assessment of colonic wall in moderate–severe ulcerative colitis: comparison with endoscopic findings. *Digestive and Liver Disease* 2011;**43**:703–6.

36. Migaleddu V, Scanu AM, Quaia E, *et al*. Contrast-enhanced ultrasonographic evaluation of inflammatory activity in Crohn's disease. *Gastroenterology* 2009;**137**:43–52.

37. Sagami S, Kobayashi T, Miyatani Y, *et al*. Accuracy of ultrasound for evaluation of colorectal segments in patients with inflammatory bowel diseases: a systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology* 2020.

38. Sagami S, Kobayashi T, Aihara K, *et al*. Transperineal ultrasound predicts endoscopic and histological healing in ulcerative colitis. *Alimentary Pharmacology & Therapeutics* 2020.