TITLE: Standardisation of intestinal ultrasound scoring in clinical trials for luminal Crohn's disease

SHORT TITLE: Intestinal ultrasound assessment of CD

AUTHORS: Thomas M. Goodsall[1] Vipul Jairath[2], Brian G. Feagan[2], Claire E. Parker[2], Tran M. Nguyen[2], Leonardo Guizzetti[2], Anil K. Asthana[3], Jakob Begun[4], Britt Christensen[3], Antony B. Friedman[3], Torsten Kucharzik[5], Andrew Lee[4], Peter J. Lewindon[4], Christian Maaser[5], Kerri L. Novak[6], Jordi Rimola[7], Kirstin M. Taylor[3], Stuart A. Taylor[8], Lauren S. White[4], Rune Wilkens[9], Stephanie R. Wilson[6], Emily K. Wright[3], Robert V. Bryant[1†,] Christopher Ma[6†]

† RVB and CM contributed equally as co-senior authors.

AUTHOR AFFILIATIONS:

[1]Adelaide, Australia

[2] London, Ontario, Canada

[3] Melbourne, Victoria, Australia

[4] Queensland, Australia

[5] Lueneburg, Germany

[6] Calgary, Alberta, Canada

[7] Barcelona, Catalonia-Spain

[8] London, United Kingdom

[9] Copenhagen, Denmark

AUTHOR AFFILIATIONS: Please see Appendix

CORRESPONDENCE: Dr Christopher Ma, MD, MPH, Assistant Professor, Division of Gastroenterology & Hepatology, Departments of Medicine & Community Health Sciences, Cumming School of Medicine, University of Calgary; 3280 Hospital Drive NW, Calgary, Alberta, Canada, T2N 4Z6

Tel: 587-434-5268

Email: christopher.ma@ucalgary.ca

Dr. Robert V. Bryant, MD, PhD, Head of IBD services, The Queen Elizabeth Hospital, 28 Woodville Road, Woodville South, 0511, SA, Australia

Tel: +61 08 8222 6000

Email: Robert.bryant@sa.gov.au

Word Count: 4419

ACKNOWLEDGEMENTS: None

LIST OF ABBREVIATIONS:

BWT, bowel wall thickness; CD, Crohn's disease; CDI, colour Doppler imaging; IBD, inflammatory bowel disease; IUS, intestinal ultrasound; RAND/UCLA, Research and Development / University of California, Los Angeles;

SUMMARY

Background: Intestinal ultrasound (IUS) is a valuable tool for assessment of Crohn's disease (CD). However, there is no widely accepted luminal disease activity index.

Aims: To identify appropriate IUS protocols, indices, items, and scoring methods for measurement of luminal CD activity, and integration of IUS in CD clinical trials.

Methods: An expert international panel of adult and paediatric gastroenterologists (n=15) and radiologists (n=3) rated the appropriateness of 120 statements derived from literature review and expert opinion (scale of 1-9) using modified RAND/UCLA methodology. Median panel scores of 1-≤ 3.5, >3.5-<6.5, and ≥6.5-9 were considered inappropriate, uncertain, and appropriate ratings, respectively. The statement list and survey results were discussed prior to voting.

Results: A total of 91 statements were rated appropriate with agreement after two rounds of voting. Items considered appropriate measures of disease activity were bowel wall thickness, vascularity, stratification, and mesenteric inflammatory fat. There was uncertainty if any of the existing IUS disease activity indices were appropriate for use in CD clinical trials. Appropriate trial applications for IUS included patient recruitment qualification when diseased segments cannot be adequately assessed by ileocolonoscopy and screening for exclusionary complications. At outcome assessment, remission endpoints including bowel wall thickness and vascularity, with or without mesenteric inflammatory fat, were considered appropriate. Components of an ideal IUS disease activity index were identified based upon panel discussions.

Conclusions: The panel identified appropriate component items and applications of IUS for CD clinical trials. Empiric evidence, and development and validation of an IUS disease activity index are needed.

KEYWORDS: Crohn's disease, inflammatory bowel disease, intestinal ultrasound, gastrointestinal ultrasound, disease activity index

SUMMARY WORD COUNT: 249 (excluding keywords)

INTRODUCTION

The management of Crohn's disease (CD) requires accurate and objective assessment of inflammatory activity to guide therapeutic decision making and disease monitoring.[1-3] While ileocolonoscopy is the reference standard for mucosal assessment in CD, limitations include invasiveness, patient tolerability, potential for procedural-related complications, inability to assess transmural or penetrating complications, and inability to evaluate disease beyond the reach of the colonoscope.[4] Non-invasive imaging, including computed tomography enterography, magnetic resonance enterography, and intestinal ultrasound (IUS), is increasingly used as a surrogate measure of luminal CD activity and complications in routine clinical practice. IUS has a similar degree of accuracy compared to colonoscopy, with a sensitivity and specificity of 75% to 90% and 75% to 100% respectively for active disease, and an area under the receiver operating curve of 0.94 for the colon and terminal ileum.[5] Additionally, IUS has several advantages over computed tomography enterography and magnetic resonance enterography including an absence of ionising radiation, no requirement for fasting or bowel preparation, higher levels of patient tolerance and better patient understanding of their disease process and monitoring.[5-8] IUS also provides transmural and extramural assessment of bowel inflammation, which may better predict risk of flare and likelihood of successful medication de-escalation when compared to ileocolonoscopy alone.[9]

Despite the reported accuracy of IUS, utilisation in clinical trials has been limited.[10] Concerns regarding operator dependence and reliability of IUS have been perpetuated by a lack of standardisation for sonographic evaluation of CD activity.[5] A recent systematic review identified 21 different inflammatory bowel disease (IBD) activity indices, that include different component items that vary in definition, assessment, grading, and weighting.[11] Additionally, the operating properties of the existing indices were found to lack rigorous validation, and notably few indices had undergone evaluation of inter- and intra-observer reliability and

responsiveness. Importantly, 92% of studies included in the systematic review were considered at high risk of bias.[11] Thus if IUS is to become accepted as a disease activity measure in clinical practice and in clinical trials, development and validation of a standardised IUS index is required. This effort should include the generation of index items with clear, data-driven definitions for scoring and appropriate weighting and prospective validation for reliability, construct validity, and responsiveness to change. As an initial step towards the development of a standardised IUS activity index for luminal CD disease activity, we aimed to identify appropriate IUS items for measurement of CD activity, as well as imaging protocols for both clinical trials and practice.

MATERIALS AND METHODS

Modified RAND/UCLA Appropriateness Methodology

The Research and Development/University of California, Los Angeles (RAND/UCLA) appropriateness methodology uses a modified Delphi panel approach to combine the best available scientific evidence with the collective judgement of content experts to develop a series of statements. The ultimate goal of this process is to assess agreement regarding statement appropriateness, without the requirement for a forced consensus.[12] This methodology has been effectively used to develop other IBD activity indices and to evaluate the face validity and feasibility of items.[13]

Development of the Modified RAND/UCLA Appropriateness Method List of Statements

The development of the initial list of statements was informed by a systematic review of IUS indices for IBD[11] and summary literature of individual items. The systematic review identified 26 studies, reporting on 21 ultrasound indices. Of these indices, 11 were CD-specific and included: Limberg score,[14] Ultrasound Activity Index for Crohn's Disease,[15] Lenze Score (Limberg derivative),[16] Maconi Score,[17] Contrast-enhanced Ultrasound Score,[18] Neye Score,[19]

Simple Ultrasonographic Score,[20] Paredes Postoperative Recurrence Score,[21] Paredes Contrast Enhanced Ultrasound Postoperative Recurrence Score,[22] Ramaswamy Score,[23] and the Ultrasound Lemann Index.[24] The initial list of survey items was reviewed by TMG, RVB, VJ, and CM.

Expert Panel

A multidisciplinary, international panel of experts was selected. Candidate panel members were identified based on clinical expertise in the performance of IUS and academic expertise with a publication record in the use of IUS for assessment of IBD. Eighteen panel members from six countries formed the final panel, which included 15 gastroenterologist sonographers affiliated with the Gastroenterological Network of Intestinal Ultrasound Australia and/or the International Bowel Ultrasound Group, and three radiologists with extensive experience in the interpretation of IUS images and academic expertise. All 18 panel members perform small intestine ultrasonography within their practice. Ten panel members routinely perform contrast enhanced ultrasonography whilst the other eight have experience but do not routinely use the technique.

Initial Panel Meeting and Survey Analysis

A moderated introductory videoconference was held to refine the initial list of statements, to identify additional relevant statements based on expert opinion, and instruct participants on the RAND methodology. The complete list of statements was then circulated via an online survey and panellists anonymously rated each item for appropriateness on a 9-point Likert scale (ranging from 1 [highly inappropriate] to 9 [highly appropriate]). Median panel scores determined item appropriateness where items with median scores in the 1 to ≤ 3.5 range without disagreement were considered inappropriate, those in the > 3.5 to < 6.5 range or any median score with disagreement were uncertain, and items with median scores in the ≥ 6.5 to 9 range without disagreement were considered appropriate. Disagreement amongst the panel was

present when six or more (approximately one-third) panellists rated the appropriateness of an item in both the lowest (1 to 3.5) and highest (6.5 to 9) three-point ranges. Results from the survey were summarised using median panel scores with the median absolute-deviation (MAD).

Second Panel Meeting and Survey

Results from the first round of panel voting were collated, distributed, and reviewed in a second moderated videoconference. The meeting focused on statements with disagreement and uncertainty. Panellists were encouraged to discuss and add rationale for individual responses, although a panel consensus was not required, consistent with RAND/UCLA appropriateness methods. The initial statement list was subsequently revised based upon the feedback obtained during the panel meeting and recirculated for a second round of voting. Median panel scores on the second list of statements were then calculated and distributed for review as previously described. All statistical analyses were performed using Stata (version 16.1, StataCorp LLC, College Station, TX, USA).

RESULTS

Overall Statement Appropriateness

Survey statements were grouped into the following domains: items related to standardising the general approach to IUS assessment of CD activity, specific IUS items for evaluation of luminal CD activity, existing IUS scoring indices for CD, considerations for the use of IUS in clinical trials, and paediatric considerations.

The initial draft survey included 106 statements and was amended following the introductory videoconference to 114 statements for the first round of survey voting. A total of 72% of statements (82/114) on the first survey were rated as appropriate without disagreement, 2% (2/114) were rated as inappropriate without disagreement and 26% (30/114) were rated as uncertain. Six additional statements were added following panel discussion of the first survey

results on the second moderated videoconference. A total of 76% (91/120) of the statements were rated as appropriate, 4% (5/120) were rated as inappropriate and 20% (24/120) were rated as uncertain. Rating distributions are available in **Supplementary Table 1**.

General Items and Standardised Approach to IUS

The panel determined that IUS should be performed, or directly supervised by, an accredited expert (***Table 1***). Fasting and bowel preparation were not considered necessary in most situations, although all three radiologists considered this an inappropriate statement (**Supplementary Table 2**). Routine IUS assessment of the colon, terminal ileum, and proximal small bowel was considered appropriate, and although technical factors may influence adequacy of image collection, the panel voted that rectal evaluation should be attempted in transabdominal IUS with documentation of image adequacy. Special techniques such as small intestine contrast ultrasonography and contrast-enhanced ultrasonography were not considered appropriate in routine assessment or for inclusion in a luminal CD activity index.

IUS Markers of Luminal CD Activity

Appropriate IUS markers of CD activity included bowel wall thickness (BWT), increased bowel wall vascularity, loss of bowel wall stratification, and mesenteric inflammatory fat (***Table 2***). Loss of small bowel peristalsis in a segment with increased BWT and/or increased Doppler vascularity, absence of colonic haustra coli, and the presence of disease complications such as an abscess or fistula were also considered appropriate. A visual analogue scale was considered appropriate to assess overall disease activity. Other potential findings were considered, although rated as uncertain with regard to their use as markers of disease activity included mesenteric lymphadenopathy and bowel wall compressibility. Uncertainty was in part due to discrepancy by specialty, with radiologists considering compressibility an appropriate marker of activity and mesenteric lymphadenopathy inappropriate with gastroenterologists voting vice versa. There was also discrepancy regarding whether lymph node diameter should

be measured on the short axis with gastroenterologists but not radiologists considering this appropriate. (**Supplementary Table 1**). The panel felt it was appropriate to consider BWT the most reliable marker of CD activity and should be weighted more than other parameters in an IUS index. The panel voted that BWT can only be accurately measured using a high frequency probe, and the average thickness from two or more separate measurements in the longitudinal and transverse planes was considered appropriate. A cut-off value of 3.0 mm was considered appropriate for distinguishing normal from pathologic small bowel and colon. Several statements addressed the measurement of increased bowel wall vascularity. Appropriate relevant items included the use of colour Doppler imaging (CDI) with a low velocity setting, and scoring according to a version of the modified Limberg score[25,26] that was further adapted based on panel recommendations to reduce subjectivity of assessment (***Table 2***). The panel was uncertain regarding the use of a semi-quantitative score (e.g., none, moderate, or severe), and binary (absent/present) measurement of bowel wall vascularity was considered inappropriate. For clinical trials (and to minimise variation in CDI signal strength, models, and probes), the use of consistent IUS equipment for serial assessments was considered appropriate. A loss of clearly demarcated mucosal, submucosal, and muscularis propria layers was considered an appropriate definition for abnormal bowel wall stratification, which the panel felt should include an assessment of submucosal prominence/thickening, and which should only be considered present when identified preferentially in longitudinal and cross-sectional images. Scoring of bowel wall stratification as present, focal (< 3 cm) or extensive (> 3 cm) loss was the only method surveyed that was rated as appropriate; binary (present/absent) or categorical (present/unclear/absent) methods for assessment were rated as uncertain. Appropriate methods for assessment of mesenteric inflammatory fat included binary (absent/present) scoring for changes in echogenicity of mesenteric fat surrounding a segment of thickened bowel wall and the presence of any fat wrapping around the associated segment

of bowel. The appropriateness of two alternative methods for scoring including a 3-point (absent, equivocal, present) or 4-point categorical scale (considering echogenicity, fat expansion, and fat wrap) was uncertain.

Categorical (absent, reduced, present, increased) assessment of loss of small bowel peristalsis in a segment with increased BWT and/or increased Doppler vascularity in the context of CD was considered appropriate. However, the panel specifically noted that changes in peristalsis alone are not specific for luminal disease activity and also occur in fibrostenotic CD. There are practical challenges to classifying peristalsis, and fasting status should be documented for evaluation of peristalsis. The value of bowel wall compressibility was rated as uncertain and the panel discussed challenges in standardising this item. Furthermore, tolerability was identified as an issue in patients with active disease.

Although uncertainty was present regarding the usefulness of mesenteric lymphadenopathy as a marker of disease activity, binary (absent/present) assessment of this finding when defined as lymph nodes greater than 4.0 mm in short axis diameter, located within the mesentery adjacent to an affected segment was considered appropriate. The panel extensively discussed the inclusion of disease complications such as abscess or fistula, as it was proposed that these are more suitable measures of disease severity rather than activity. However, binary (absent/present) assessment of disease complications was considered appropriate as useful markers given that they frequently occur in the setting of active inflammation. The panel rated continuous measurement of the length of a diseased segment as appropriate, although it was acknowledged that this may be technically challenging, particularly for long segments of inflammation. The panel was uncertain regarding the appropriateness of categorical reporting of diseased segment length when defined as discrete (< 1.0 cm), short (< 5.0 cm), and long (> 5.0 cm), however they also acknowledged that categorical reporting of diseased segment length

might be generally more reproducible than continuous measurement. Representative images of the items considered is presented in **Figure 1** in a patient with terminal ileitis.

Current IUS Scoring Indices

The appropriateness of existing IUS scoring indices for the assessment of luminal CD activity were considered (***Table 3***). The appropriateness of six of the indices for this purpose, including the modified Limberg score,[25,26] the Lenze score[16] (for distinguishing fibromatous, mixed or inflammatory disease), the Neye score[19] (a 4-point modification of the Limberg score), the Simple Ultrasonographic Score[20] (continuous algorithm of BWT and CDI), the Ultrasound Activity Index for CD,[15] and the Ultrasound Lemann index[24] (modified magnetic resonance imaging score validated for stricturing and penetrating disease), was rated as uncertain. Two indices (the Contrast-enhanced Ultrasound Score[18] and the Paredes Contrast Enhanced Ultrasound Postoperative Recurrence Score[22]) that include assessment based on the use of intravenous contrast were rated as inappropriate.

Considerations for Use of IUS in Clinical Trials

The panel discussed that centralised reading of IUS images has the potential to limit observer bias in the clinical trial setting. Quality standards for image acquisition, and the ideal qualities of an IUS index for use in CD clinical trials were assessed for appropriateness (***Table 4***).

For clinical trials, image optimisation with focal length set to the bowel segment of interest and depth set to demonstrate associated mesentery was rated as appropriate. The panel voted that two images (cross-sectional and longitudinal) for each bowel segment are required for appropriate measurement of BWT. Collection of one 3 to 5-second video-loop demonstrating colour Doppler bowel wall vascularity for each affected segment and one 10-second video-loop demonstrating distal ileal peristalsis was also considered appropriate. The panel was uncertain regarding the appropriateness of collecting short axis, long axis, and colour Doppler video loops for every segment due to feasibility considerations. However, the panellists agreed

that these may offer optimal centralised reader assessment of bowel wall stratification, mesenteric inflammatory fat, length of affected segment, and changes in peristalsis/colonic haustra coli in the clinical trial and research setting.

There was agreement that an ideal IUS activity index for luminal CD should be easy to calculate and output a single numeric score. Either an index calculation based on the most affected segment or based on all visualised segments was considered appropriate. The panel discussed that there were merits to both options: an index calculated from all visualised segments may better represent the overall disease activity, whereas an index based on the most affected segment potentially may be more responsive to change after effective medical therapy. An overall score based on the sum of all examined segments was considered appropriate, whereas there was uncertainty on the appropriateness of an average score based on dividing the sum of the segmental scores by the number of visualised segments.

At screening, the panel voted that use of IUS to qualify patients for trial recruitment among those with diseased segments that cannot be adequately assessed by ileocolonoscopy was appropriate, as was the use of IUS to exclude patients from trials who have complications such as abscesses. At outcome assessment, use of an IUS activity index for luminal CD to define response and remission was considered appropriate, as were remission outcomes defined as a combination of BWT normalisation (< 3.0 mm) and no bowel wall vascularity on CDI (either with and without mesenteric inflammatory fat). Another definition of remission considered appropriate included the absence of bowel wall vascularity without mesenteric inflammatory fat and BWT normalisation (even if the BWT remained > 3.0 mm). The panel discussed that in patients with longstanding disease, there is some evidence that BWT may not normalise to < 3.0 mm, even in the absence of endoscopic inflammation.[27,28]

A combination of the magnitude of decrease in BWT and colour Doppler activity was voted as appropriate for defining sonographic response. The panel noted that other additional items may

also be appropriate for defining IUS response, although they require validation and were not specifically included in the survey statements. Appropriate timepoints for assessing response to induction therapy included 8, 12, and 16 weeks whereas the appropriateness of earlier (4 or 6 weeks) timepoints was uncertain. For assessment of response to maintenance therapy, appropriate timepoints for assessment included both 26 and 52 weeks.

Paediatric Considerations

The panel felt it was appropriate to use the same general IUS considerations, item scoring methods, disease activity indices and response and remission criteria in both paediatric and adult patients, except for scoring methods for bowel wall compressibility and mesenteric lymphadenopathy. As previously mentioned, challenges in standardisation and patient tolerability likely influenced the appropriateness of bowel wall compressibility and mesenteric lymphadenopathy which may represent a normal variant in children (***Table 5***).

DISCUSSION

This panel of international, multidisciplinary experts determined there was uncertainty if any of the existing IUS indices were appropriate instruments for measuring luminal CD activity. Variation in the definitions of potential sonographic disease activity measures and the lack of a standardised and validated IUS disease activity index has limited the implementation of this imaging technique in both clinical trials and practice. Therefore, as an initial step towards developing a valid IUS index, we used the modified RAND/UCLA appropriateness methodology to determine the appropriateness of relevant imaging protocols and procedures, definitions and scoring methods for potential markers of luminal CD activity, and considerations related to integration of IUS in clinical trial study design and outcome assessment.

Whilst most general IUS items and statements regarding image acquisition protocols were considered appropriate, there was discussion regarding the necessity for rectal assessment and routine use of special IUS techniques. Transabdominal views of the rectum can be achieved using IUS, however image quality is dependent on available hardware, bladder filling, operator experience, and patient body habitus.[5,29] Rectal visualisation is appropriate when performing IUS to accurately assess CD activity, however the adequacy of views should be reported. Special techniques such as small intestine contrast ultrasonography and contrast enhanced ultrasound were recognised as valuable adjunctive tools, however they are most useful in assessing strictures and classifying penetrating complications.[30,31] Contrast enhanced ultrasound may also incrementally improve the accuracy of disease activity assessment when BWT and CDI are discordant.[18] Overall, technical challenges, additional time and cost, and minimal incremental benefit in accuracy make these techniques less appealing for routine assessment of inflammatory activity in CD.

Appropriate sonographic measures useful for evaluating luminal CD activity were identified, although uncertainty existed for specific scoring methods and indices. Heterogeneity in the literature, evolving concepts, and the need for empiric data likely contributed to the uncertainty regarding optimal scoring methods and cut-offs. Overall, the panel expressed a preference for the use of continuous measures to facilitate discrimination across a range of disease activity and maximise sensitivity to change after treatment. However, it was acknowledged that accurate continuous measurement may be technically challenging for some findings, such as the length of an affected segment, and categorical definitions may have better inter-rater reliability. A 3.0 mm cut-off for BWT was considered appropriate, however the panel acknowledged that this cut-off may reduce sensitivity in the small bowel compared to a 2.0 mm cut-off, and specificity in the rectosigmoid colon compared to a 4.0 mm cut-off.[10,11,32]

Assessment of bowel wall vascularity according to a modification of the Limberg score[25,26] was considered appropriate. Although a modification first applied by Drews *et al*.[26] in 2009 is widely applicable across different hardware,[14] the panel recommended that the subjectivity of intramural vascular signal assessment might be further reduced by the inclusion of the descriptors "circular" and "linear" to Grades 1 and 2 of the modified score. This adaptation was the only method considered appropriate for assessment of bowel wall vascularity. Concerns amongst panellists remained regarding the inter-rater reliability of a semi-quantitative scoring method, which was recently reported as moderate (Cohen's kappa = 0.6).[25,33] The value of collapsing grades 1 and 2, and the novel method proposed in this study to improve reliability of assessment of bowel wall vascularity will be determined in an ongoing study.

An appropriate definition for abnormal bowel wall stratification is an important outcome of this study. Heterogeneity and lack of clarity for defining bowel wall stratification may have contributed to the observed lack of reliability in previous studies and consequent exclusion of this items from some CD activity indices.[11] Two complementary components for scoring bowel wall stratification were considered appropriate: the first determining the categorical degree of loss of stratification and the second describing the length of abnormal stratification. Whilst this combination has not been previously described, the panel considered it appropriate recognising that assessment of the operating characteristics and empiric validation are required. Similarly, definitions of mesenteric inflammatory fat on IUS have been heterogeneous to date. In the panel discussion, neither mesenteric hyperechogenicity or fat wrapping alone were considered adequate, and the general term mesenteric inflammatory fat should be used to describe the mesenteric inflammatory changes observable with IUS. Whilst the panel felt that a binary grading system for this item was appropriate, two new grading scales proposed by Novak *et al*.

and Bhatnagar *et al*. (based on the METRIC study) were rated as uncertain but may demonstrate greater discriminatory power. [25,33-35]

Potential applications of IUS in clinical trials were also explored in this study. Beyond the value for patient screening, broad application of IUS-based outcome measures in clinical trials would require the development and validation of appropriate response and remission definitions. The magnitude of change necessary to define sonographic response was recently explored in an interim analysis of an IUS subgroup from the ustekinumab treat-to-target STARDUST trial.[36] In this analysis, a 25% reduction in BWT showed moderate agreement with endoscopic response.[10] The appropriateness of other IUS items in defining response or remission has not yet been investigated in prospective studies, and therefore, there was uncertainty among the panel with respect to a single consensus definition of IUS remission. Transmural healing, defined by normalisation of a previously affected bowel segment on cross-sectional imaging, has been shown to have prognostic importance beyond endoscopic remission alone.[9] However, whether an IUS-based definition of remission should require this level of stringency and if complete normalisation is a realistic treatment target with currently available medical therapies is unclear. The panel noted that in some cases, the BWT may never return to ≤ 3.0 mm due to fibrosis, yet this does not necessarily correlate with persistent luminal inflammation on endoscopy. Therefore, a remission definition capturing BWT normalisation (even if it remains > 3.0 mm) without bowel wall vascularity and mesenteric inflammatory fat was considered appropriate. Further prospective, longitudinal studies are required to understand how IUS activity and parameters correlate with objective long-term outcomes such as surgery, hospitalisation, or corticosteroid use, as well as biomarkers such as C-reactive protein and faecal calprotectin. The identified approach to IUS for clinical trials is applicable to clinical practice and there was very strong agreement that an activity index should be easily calculated as a single whole number, for ease of use and interpretation both in clinic and in

research settings (**Table 4**). Areas of uncertainty regarding interpretation and scoring of individual items, especially CDI and inflammatory fat, also highlights the need for caution in clinical practice pending further validation studies to elucidate the interpretation of these items. The panel identified uncertainty if any of the existing IUS indices were appropriate for measuring luminal CD activity. Parameters for development and validation of an ideal IUS activity index were also explored. The panel discussed that extraluminal complications of CD, although appropriate for measurement, should be reported separately, perhaps as indicators of disease severity or bowel damage, analogous to the Lemann index.[24] The total length of an affected segment provides important information for assessing response and treatment decision making in clinical practice; however the significance of this item in an activity score is yet to be determined. The panel also discussed using the most affected segment or all visualised segments in an IUS score. While there are advantages and disadvantages to each, ultimately an activity index should provide an overall sum as well as a segmental score to allow monitoring of segmental response and overall disease activity.

Our study has several strengths. First, we included an international interdisciplinary panel, from diverse research groups with a variety of expertise to address areas of uncertainty and inconsistency regarding the use of IUS for assessment of luminal CD activity. The level of discussion and engagement was high amongst this group, reflecting an important opportunity to refine and hone the included items. Secondly, panel voting and discussion was informed by an up-to-date systematic review as well as the most recent evidence from studies published between rounds of voting.[11,25,33,34,37] Limitations of this study include uncertainty on the appropriateness of some IUS component item definitions, response definitions, and definite structure of an ideal activity score for luminal CD. However, this is likely reflective of current gaps in the literature, heterogeneity in IUS definitions, and lack of standardisation of items incorporated in CD activity assessment. Methodologically, it is important to recognize that the

RAND/UCLA method does not force a panel consensus by majority vote, and permits items to be voted as uncertain with respect to appropriateness. Second, there were more gastroenterologists who perform IUS on the panel than radiologists, which may have biased responses on the approach to image acquisition and sonographic techniques. For example, although considered not routinely required in this study, contrast enhanced ultrasound is commonly performed in tertiary radiology centres, recognising that there is geographic variability in clinical practice. There were only four important areas of discrepancy between specialty, the most important being the requirement for fasting which reflects the point-of-care approach by gastroenterologists and is an area that would benefit from prospective validation. Additionally, the distribution of the panel reflects the predominant role of IUS as a point of care investigation performed by gastroenterologists and the need for applicability and generalisability, again, notwithstanding that operators of IUS vary by country and clinical setting.

In conclusion, this study employed a modified RAND/UCLA appropriateness methodology to define an evidence-based approach to standardise IUS assessment of luminal CD activity and determine areas of uncertainty and inconsistency of practice. These results will be used to inform future studies whose ultimate goals include the development and validation of an IUS-based CD activity score for implementation in clinical practice and clinical trials.

REFERENCES

1. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis.* 2019;13:144-164.
2. Peyrin-Biroulet L, Loftus EV Jr, Colombel JF, Sandborn WJ. The natural history of adult Crohn's disease in population-based cohorts. *Am J Gastroenterol.* 2010;105:289-297.
3. 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;110:1324-1338.
4. Walsh AJ, Bryant RV, Travis SP. Current best practice for disease activity assessment in IBD. *Nat Rev Gastroenterol Hepatol.* 2016;13:567-579.
5. Bryant RV, Friedman AB, Wright EK, et al. Gastrointestinal ultrasound in inflammatory bowel disease: an underused resource with potential paradigm-changing application. *Gut.* 2018;67:973-985.
6. Friedman A, Asthana A, Knowles S, Robbins A, Gibson P. Gastroenterologist-performed point-of-care gastrointestinal ultrasound improves patient understanding of disease activity, symptomatology, management decisions, and clinical outcomes. *J Crohns Colitis.* 2018;12:S406.
7. Goodsall TM, Noy R, Nguyen TM, Costello SP, Jairath V, Bryant RV. Systematic review: patient perceptions of monitoring tools in inflammatory bowel disease. [published online ahead of print January 24, 2020]. *J Can Assoc Gastroenterol.*
8. Miles A, Bhatnagar G, Halligan S, et al. Magnetic resonance enterography, small bowel ultrasound and colonoscopy to diagnose and stage Crohn's disease: patient acceptability and perceived burden. *Eur Radiol.* 2019;29:1083-1093.

9. Castiglione F, Imperatore N, Testa A, et al. One-year clinical outcomes with biologics in Crohn's disease: transmural healing compared with mucosal or no healing. *Aliment Pharmacol Ther.* 2019;49:1026-1039.

10. Kucharzik T, Wilkens R, Maconi G, et al. Intestinal ultrasound response and transmural healing after ustekinumab induction in Crohn's disease: week 16 interim analysis of the STARDUST trial substudy. *J Crohns Colitis.* 2020;14:S46-48.

11. Goodsall TM, Nguyen TM, Parker CE, et al. Systematic review: gastrointestinal ultrasound scoring indices for inflammatory bowel disease [published online ahead of print July 2, 2020]. *J Crohns Colitis.*

12. Fitch K, Bernstein SJ, Aguilar MD, et al. The RAND/UCLA appropriateness method user's manual. In: Santa Monica, California: RAND Corporation; 2001. [online]. https://www.rand.org/pubs/monograph_reports/MR1269.html. Accessed June 22, 2020.

13. Mosli MH, Feagan BG, Zou G, et al. Development and validation of a histological index for UC. *Gut.* 2017;66:50-58.

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

15. Futagami Y, Haruma K, Hata J, et al. Development and validation of an ultrasonographic activity index of Crohn's disease. *Eur J Gastroenterol Hepatol.* 1999;11:1007-1012.

16. Lenze F, Wessling J, Bremer J, et al. Detection and differentiation of inflammatory versus fibromatous Crohn's disease strictures: prospective comparison of 18F-FDG-PET/CT, MR-enteroclysis, and transabdominal ultrasound versus endoscopic/histologic evaluation. *Inflamm Bowel Dis.* 2012;18:2252-2260.

17. Maconi G, Carsana L, Fociani P, et al. Small bowel stenosis in Crohn's disease: clinical, biochemical and ultrasonographic evaluation of histological features. *Aliment Pharmacol Ther.* 2003;18:749-756.

18. Medellin-Kowalewski A, Wilkens R, Wilson A, Ruan J, Wilson SR. Quantitative contrast-enhanced ultrasound parameters in Crohn disease: their role in disease activity determination with ultrasound. *AJR Am J Roentgenol.* 2016;206:64-73.

19. 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.

20. 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;23:2001-2010.

21. 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;4:537-545.

22. Paredes JM, Ripolles T, Cortes X, et al. Contrast-enhanced ultrasonography: Usefulness in the assessment of postoperative recurrence of Crohn's disease. *J Crohns Colitis.* 2013;7:192-201.

23. Ramaswamy PK, Nagarajan KV, Yelsangikar A, Nagar A, Bhat A. Utility of bowel ultrasound in diagnosing disease activity in Crohn’s disease: Indian experience. *J Crohns Colitis.* 2019;13:S254.

24. Rispo A, Imperatore N, Testa A, et al. Bowel damage in Crohn's disease: direct comparison of ultrasonography-based and magnetic resonance-based Lemann index. *Inflamm Bowel Dis.* 2017;23:143-151.

25. Novak KL, Nylund K, Maaser C, et al. Expert consensus on optimal acquisition and development of the International Bowel Ultrasound Segmental Activity Score (IBUS-SAS): a reliability and inter-rater variability study on intestinal ultrasonography in Crohn's Disease [published online ahead of print October 24, 2020]. *J Crohns Colitis.*

26. Drews BH, Barth TF, Hanle MM, et al. Comparison of sonographically measured bowel wall vascularity, histology, and disease activity in Crohn's disease. *Eur radiol.* 2009;19:1379-1386.

27. Kim C, Park SH, Yang SK, et al. Endoscopic complete remission of Crohn disease after anti-tumor necrosis factor-α therapy: CT enterographic findings and their clinical implications. *AJR Am J Roentgenol.* 2016;206:1208-1216.

28. Rimola J, Alfaro I, Fernández-Clotet A, et al. Persistent damage on magnetic resonance enterography in patients with Crohn's disease in endoscopic remission. *Aliment Pharmacol Ther.* 2018;48:1232-1241.

29. Parente F, Greco S, Molteni M, et al. Role of early ultrasound in detecting inflammatory intestinal disorders and identifying their anatomical location within the bowel. *Aliment Pharmacol Ther.* 2003;18:1009-1016.

30. Pallotta N, Vincoli G, Montesani C, et al. Small intestine contrast ultrasonography (SICUS) for the detection of small bowel complications in Crohn's disease: a prospective comparative study versus intraoperative findings. *Inflamm Bowel Dis.* 2012;18:74-84.

31. Ripollés T, Martínez-Pérez MJ, Blanc E, et al. Contrast-enhanced ultrasound (CEUS) in Crohn's disease: technique, image interpretation and clinical applications. *Insights Imaging.* 2011;2:639-652.

32. Maaser C, Helwig U, Fischer I, Rath S, Kolterer S, Kucharzik T. Patient-reported outcomes (PRO-2) and intestinal ultrasound in ulcerative colitis patients: subanalysis of the TRUST&UC study cohort. *J Crohns Colitis.* 2020;14:S560-561.

33. Wilkens RT, Nylund K, Petersen F, et al. Expert consensus on acquisition and reporting of intestinal ultrasonography activity in Crohn's disease. A prospective inter-rater agreement study. *J Crohns Colitis.* 2020;14:S225-226.

34. Bhatnagar G, Rodriguez-Justo M, Higginson A, et al. Inflammation and fibrosis in Crohn's disease: location-matched histological correlation of small bowel ultrasound features [published online ahead of print June 20, 2020]. *Abdom Radiol.*

35. Taylor SA, Mallett S, Bhatnagar G, et al. 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;3:548-558.

36. Study of treat to target versus routine care maintenance strategies in Crohn's disease patients treated with ustekinumab (STARDUST). 2020. https://clinicaltrials.gov/ct2/show/NCT03107793.

37. Sævik F, Eriksen R, Eide GE, Gilja OH, Nylund K. Development and validation of a simple ultrasound activity score for Crohn's disease [published online ahead of print June 6, 2020] . *J Crohns Colitis.*

Appendix A: Author Affiliations

AUTHORS: Thomas M. Goodsall[1,2] Vipul Jairath[3,4], Brian G. Feagan[3,4], Claire E. Parker[4], Tran M. Nguyen[4], Leonardo Guizzetti[4], Anil K. Asthana[5], Jakob Begun[6], Britt Christensen[5], Antony B. Friedman[7], Torsten Kucharzik[8], Andrew Lee[9], Peter J. Lewindon[10], Christian Maaser[11], Kerri L. Novak[12], Jordi Rimola[13], Kirstin M. Taylor[7], Stuart A. Taylor[14], Lauren S. White[15], Rune Wilkens[16,17], Stephanie R. Wilson[18], Emily K. Wright[19], Robert V. Bryant[1,2,] Christopher Ma[12,20]

AFFILIATIONS:

[1]IBD Service, Department of Gastroenterology, The Queen Elizabeth Hospital, Adelaide, Australia.

[2]Faculty of Health Sciences, School of Medicine, University of Adelaide, Adelaide, Australia

[3] Division of Gastroenterology, Department of Medicine, University of Western Ontario, London, Ontario, Canada

[4] Alimentiv Inc. (formerly Robarts Clinical Trials, Inc.), London, Ontario, Canada

[5] Department of Gastroenterology, The Royal Melbourne Hospital, Melbourne, Victoria, Australia

[6] Mater Research, University of Queensland, Brisbane, Queensland, Australia

[7] Department of Gastroenterology, Alfred Health and Monash University, Melbourne, Victoria, Australia

[8] Department of Gastroenterology, University Teaching Hospital Lueneburg, Bögelstraße, Lueneburg, Germany

[9] Department of Gastroenterology, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia

[10] Department of Gastroenterology, Queensland Children's Hospital, University of Queensland, Brisbane, Australia

[11] Klinikum Lüneburg, Outpatient's Department of Gastroenterology, Lüneburg, Germany

[12] Division of Gastroenterology and Hepatology, Department of Medicine, University of Calgary, Calgary, Alberta, Canada

[13] Radiology Department IBD Unit, Hospital Clínic de Barcelona, Barcelona, Catalonia-Spain

[14] Centre for Medical Imaging, University College London, London, United Kingdom

[15] Sunshine Coast University Hospital, Birtinya, Queensland, Australia

[16] Gastrounit, Division of Medicine, Copenhagen University Hospital Hvidovre, Kettegaard Alle, Hvidovre, Copenhagen, Denmark
[17] Copenhagen Center for Inflammatory Bowel Disease in children, adolescents and adults, University of Copenhagen, Hvidovre Hospital, Copenhagen, Denmark
[18] Department of Radiology and Department of Medicine, Division of Gastroenterology, University of Calgary, Calgary, Alberta, Canada
[19] Department of Gastroenterology, St Vincent's Hospital, University of Melbourne, Melbourne, Victoria, Australia
[20] Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada

FIGURE LEGENDS

Figure 1. Cross sectional images demonstrating terminal ileitis

Left: B mode ultrasound image with increased wall thickness 5.1mm (wall thickness; Blue bracket, luminal interface; blue arrow, mucosa; star, submucosa; asterisk, muscularis propria; arrowhead) inflammatory fat with wrapping (yellow boarder), and lymph node (blue arrow heads)

Right: Colour Doppler imaging with short intramural signal (Modified Limberg score 1)

TABLES

Table 1. General items and standardised approach to intestinal ultrasound for assessing Crohn's disease activity

Item	Summary (MAD)	Rating
Gastrointestinal ultrasound should be performed or supervised by an expert gastroenterologist or radiologist with specific training in gastrointestinal ultrasound.	9 (0.41)	Appropriate
Gastrointestinal ultrasound should be performed using both a low frequency and high frequency probe.	8 (0.76)	Appropriate
A low frequency probe should be used to detect anatomy and gross pathology before changing to a high frequency probe.	8 (1.00)	Appropriate
Gastrointestinal ultrasound should be performed with a relaxed and fully supine patient.	8 (0.59)	Appropriate
A systematic approach should be taken to examine the whole intestine when performing gastrointestinal ultrasound.	9 (0.59)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the sigmoid colon.	9 (0.29)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the descending colon.	9 (0.29)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the transverse colon.	9 (0.31)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the ascending colon.	9 (0.29)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the caecum.	9 (0.31)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the terminal ileum.	9 (0.29)	Appropriate
Gastrointestinal ultrasound should routinely evaluate the proximal small bowel (i.e. proximal to the terminal ileum).	9 (0.73)	Appropriate
The rectum should be evaluated in transabdominal ultrasound if possible, and the adequacy of rectal views should be documented.	8 (1.12)	Appropriate
Gastrointestinal ultrasound does not require fasting in most situations.	9 (1.29)	Appropriate
Gastrointestinal ultrasound does not require bowel preparation in most situations.	9 (0.47)	Appropriate
When performing small intestine contrast ultrasound, the patient should be instructed to ingest 300 to 500 mL of polyethylene glycol 30 to 60 minutes before the procedure to increase the sensitivity and specificity of detection and characterisation of small intestinal lesions.	8 (1.24)	Appropriate
Small intestine contrast ultrasonography should be integrated into a Crohn's disease activity index.	2 (0.71)	Inappropriate
Contrast-enhanced ultrasonography with intravenous contrast is a useful adjunctive technique for identifying inflammatory and penetrating complications of Crohn's disease.	8 (1.18)	Appropriate
Routine use of contrast enhanced ultrasonography is not required in most situations because of limitations of the technique including time, expertise, and need for intravenous access.	8 (0.88)	Appropriate

Abbreviations: MAD, mean absolute-deviation from the median

Table 2. Intestinal ultrasound items for evaluation of luminal Crohn's disease activity

Item	Summary (MAD)	Rating
A Visual Analogue Scale (e.g., 0 to 100mm) should be used to assess overall disease activity.	7 (1.65)	Appropriate
A Visual Analogue Scale (e.g., 0-100mm) should be used to assess bowel wall stratification.	3 (1.88)	Inappropriate
Bowel wall thickness is a useful marker of disease activity.	9 (0.18)	Appropriate
Bowel wall thickness can only be accurately measured using a high frequency probe.	7 (1.71)	Appropriate
Bowel wall thickness should be measured from the interface of the intestinal contents and hypoechoic mucosa to the luminal margin of the hyperechoic serosa.	9 (0.41)	Appropriate
The same bowel wall thickness cut-off should be used for both the large and small bowel.	7 (1.41)	Appropriate
A bowel wall thickness of 3.0 mm should be used as a cut-off to distinguish normal from pathologic bowel in the colon.	7 (1.12)	Appropriate
A bowel wall thickness of 3.0 mm should be used as a cut-off to distinguish normal from pathologic bowel in the small bowel.	8 (1.18)	Appropriate
Bowel wall thickness is the most reliable marker of Crohn's disease activity.	8 (0.65)	Appropriate
Longitudinal and transverse bowel wall thickness should be measured separately.	8 (1.00)	Appropriate
When measuring bowel wall thickness, the average thickness from two or more measurements should be used.	8 (0.82)	Appropriate
Bowel wall thickness should be scored continuously as a value in millimetres (to one decimal place, e.g., 3.0 mm) in a GIUS index for assessing Crohn's disease.	8 (0.65)	Appropriate
Bowel wall thickness should be weighted more than other parameters in a GIUS index for assessing Crohn's disease.	8 (1.12)	Appropriate
The presence of increased bowel wall vascularity as measured by colour Doppler imaging is a useful marker of Crohn's disease activity.	8 (0.65)	Appropriate
Colour Doppler imaging should be used to measure bowel wall vascularity using a low velocity setting, with sensitivity calibrated by reducing the gain until artefactual signal is no longer present.	8 (0.71)	Appropriate
Bowel wall vascularity should be scored as a binary outcome (i.e., absent or present).	3 (1.59)	Inappropriate
Bowel wall vascularity should be scored semi-quantitatively as: None Moderate (visible vessels within bowel wall) Severe (visible vessels within bowel wall and extending into mesentery)	6 (1.59)	Uncertain
Bowel wall vascularity should be scored as: 0 = No blood flow on colour Doppler imaging 1 = Small, circular intramural vascular signal on colour Doppler imaging 2 = Longer linear intramural vascular signal on colour Doppler imaging 3 = Longer stretches of vascular signal with extension into mesentery on colour Doppler imaging	8 (1.29)	Appropriate
Consistent GIUS equipment must be used for baseline and post-treatment assessment in clinical trials of Crohn's disease.	8 (0.71)	Appropriate
Loss of bowel wall stratification is a useful marker of disease activity.	8 (0.94)	Appropriate
Abnormal bowel wall stratification should be defined as loss of clearly demarcated mucosal, submucosal, and muscularis propria layers.	8 (0.82)	Appropriate

Abnormal bowel wall stratification should only be considered present if it is identified in two views, preferably longitudinal and cross-sectional images.	7 (1.24)	Appropriate
Bowel wall stratification should be scored as a binary outcome (i.e., present or absent).	5 (1.71)	Uncertain
Bowel wall stratification should be scored as present, focal loss (< 3 cm), or extensive (≥ 3 cm).	7 (1.53)	Appropriate
Bowel wall stratification should be scored as present, unclear, or absent.	5 (2.06)	Uncertain
Bowel wall stratification should include an assessment of submucosal prominence/thickening.	7 (1.06)	Appropriate
Mesenteric inflammatory fat is a useful marker of disease activity.	8 (0.88)	Appropriate
Mesenteric inflammatory fat should be assessed based on changes in echogenicity of mesenteric fat surrounding a segment of thickened bowel wall and the presence of any fat wrapping around the associated segment of bowel.	8 (1.12)	Appropriate
Mesenteric inflammatory fat should be scored as a binary outcome (i.e., absent or present).	7 (1.47)	Appropriate
Mesenteric inflammatory fat should be scored using three categorical variables (i.e., absent, equivocal, or present).	5 (2.18)	Uncertain
Mesenteric inflammatory fat should be scored using the following categories: 0=Normal 1 = Focal hyperechoic without fat wrap (i.e., focal defined area of mesenteric fat of increased echogenicity without overall increase in volume of peri-mural fat) 2 = Stratified heterogeneous with fat expansion (i.e., overall increase in volume of peri-mural fat with maintained normal mesenteric stratification and no focal hyperechoic area) 3 = Focal hyperechoic with fat wrap (i.e., focal defined area of mesenteric fat of increased echogenicity with overall increase in volume of peri-mural fat)	5 (1.76)	Uncertain
Bowel wall compressibility is a useful maker of disease activity.	4 (1.59)	Uncertain
Bowel wall compressibility should be scored as a binary outcome (i.e. absent or present).	5 (1.24)	Uncertain
Mesenteric lymphadenopathy is a useful maker of disease activity.	6 (1.94)	Uncertain
Mesenteric lymphadenopathy should be defined as lymph nodes greater than 4.0 mm in short axis diameter that are located in the mesentery adjacent to an affected segment.	7 (1.94)	Appropriate
Mesenteric lymph nodes greater than 10.0 mm in short axis diameter should be considered pathologic.	8 (0.88)	Appropriate
Mesenteric lymphadenopathy should be scored as a binary outcome (i.e., absent or present).	8 (1.65)	Appropriate
Loss of small bowel peristalsis in a segment with increased bowel wall thickness and/or increased Doppler vascularity is a useful marker of disease activity.	7 (1.24)	Appropriate
Small bowel peristalsis (within the context of inflammatory Crohn's disease) should be scored using categorical variables (i.e., absent, reduced, present, or increased).	7 (1.29)	Appropriate
Small bowel peristalsis (within the context of inflammatory Crohn's disease) should be scored as a binary outcome (i.e., absent or present).	6 (2.06)	Uncertain
Fasting status of the patient should be documented for evaluation of peristalsis.	7 (1.12)	Appropriate
Absence of colonic haustra coli is a useful marker of disease activity.	7 (1.59)	Appropriate
Absence of colonic haustra coli should be scored as a binary outcome (i.e., absent or present).	7 (1.12)	Appropriate
Presence of complications such as abscess or fistula is a useful marker of disease activity.	8 (1.41)	Appropriate
Presence of complications such as abscess or fistula should be scored as a binary outcome (i.e., absent or present).	8 (1.18)	Appropriate
The total length of a diseased segment should be reported using three categorical variables	6 (1.82)	Uncertain

discrete= < 1.0 cm short = < 5.0 cm long = > 5.0 cm		
The total length of a diseased segment should be scored continuously as a value in centimetres (to one decimal place, e.g., 3.0 cm).	8 (1.94)	Appropriate

Abbreviations: GIUS, gastrointestinal ultrasound; MAD, mean absolute-deviation from the median

Table 3. Available IUS Scoring Indices for Crohn's disease

Item	Summary (MAD)	Rating
The Modified Limberg Score is an appropriate instrument for assessing Crohn's disease activity. 0 = Normal BWT and normal CDI 1 = Increased BWT and no CDI 2 = Increased BWT with short stretches of CDI 3 = Increased BWT with longer stretches of CDI 4 = Increased BWT and longer stretches of CDI extending into surrounding mesentery	6 (1.31)	Uncertain
The Contrast-Enhanced Ultrasound Score is an appropriate instrument for assessing Crohn's disease activity. Peak contrast enhancement: < 18.2 dB = Inactive 18.2-22.8 dB = Mild to moderate > 2.8 dB = Moderate to severe	3.5 (1.50)	Inappropriate
The Lenze Score is an appropriate instrument for assessing Crohn's disease activity. Fibromatous = Hyperechogenic wall thickening and Limberg 1 Mixed = Mixed hypo- and hyper-echogenic wall thickening and Limberg 2 Inflammatory = Hypoechogenic wall thickening and Limberg 3 or 4	5 (1.25)	Uncertain
The Neye Score is an appropriate instrument for assessing Crohn's disease activity. 1 (Inactive) = BWT < 5.0 mm and no vessels 2 (Mild activity) = BWT < 5.0 mm and 1-2 vessels OR BWT ≥ 5.0 mm and no vessels 3 (Moderate activity) = BWT < 5.0 mm and > 2 vessels OR BWT ≥ 5.0 mm and 1-2 vessels 4 (High activity) = BWT ≥ 5.0 mm and > 2 vessels	5 (1.19)	Uncertain
The Paraedes Contrast Enhanced Ultrasound Postoperative Recurrence Score is an appropriate instrument for assessing Crohn's disease activity. 0 = Normal BWT < 3.0 mm and CEUS enhancement < 34.5% 1 = BWT 3-5.0 mm with CEUS enhancement < 46% Recurrence 2 = BWT > 5.0 mm or CEUS enhancement > 46% Mod-severe recurrence 3 = BWT > 5.0 mm or CEUS enhancement > 70%, or presence of fistula	3.5 (1.38)	Inappropriate
The Simple Ultrasonographic Score is an appropriate instrument for assessing Crohn's disease activity. (A continuous algorithm of BWT and CDI)	6 (1.31)	Uncertain
The Ultrasound Activity Index for Crohn's Disease is an appropriate instrument for assessing Crohn's disease activity. A = Decreased compressibility and peristalsis with loss of haustrations but without bowel wall thickening (4.0 mm cut-off) B = Pathologic wall thickening and presence of BWS C = Pathologic wall thickening and loss of BWS	5 (1.31)	Uncertain
The Ultrasound Lemann Index is an appropriate instrument for assessing Crohn's disease activity. Small bowel – Stricturing:	5 (1.25)	Uncertain

Grade 1 = BWT > 3.0 mm or segmental enhancement without pre-stenotic dilatation Grade 2 = BWT > 4.0 mm or mural stratification without pre-stenotic dilatation Grade 3 = BWT > 4.0 mm, narrowed lumen, and fluid distended or echogenic content-filled loops proximal to thickened tract. Small bowel – Penetrating: Grade 2 = Deep transmural ulceration Grade 3 = Hypoechoeic duct-like structures with fluid or air content between intestine and skin, intestine or mesentery Colon – Stricturing: Grade 1 = BWT > 3.0 mm or segmental enhancement without pre-stenotic dilatation Grade 2 = BWT > 4.0 mm or mural stratification without pre-stenotic dilatation or < 50% of lumen Grade 3 = Stricture with pre-stenotic dilatation or > 50% of the lumen Colon – Penetrating: Grade 2 = Deep transmural ulceration Grade 3 = Phlegmon or any type of fistula		

Abbreviations: BWT, bowel wall thickness; BWS, bowel wall stratification; CDI, Color Doppler imaging; CEUS, contrast-enhanced ultrasound; MAD, mean absolute-deviation from the median

Table 4. Considerations for use of IUS in clinical trials

Item	Summary (MAD)	Rating
All images should be optimised with focal length set to the bowel segment of interest and depth set to demonstrate associated mesentery.	8 (0.69)	Appropriate
Two images with bowel wall measurements should be collected for each bowel segment: one cross-sectional and one longitudinal.	8 (0.94)	Appropriate
For each segment, three short video loops should be collected in short axis, long axis, and with colour Doppler imaging.	5.5 (1.94)	Uncertain
A 3-5 second video-loop demonstrating colour Doppler vascularity should be collected for each affected segment.	8 (1.00)	Appropriate
Images of lymph nodes with short axis measurements should be collected.	7 (1.56)	Appropriate
A 10 second video-loop demonstrating distal ileal peristalsis should be collected.	6.5 (1.25)	Appropriate
A gastrointestinal ultrasound score should be calculated based on all visualised segments.	8 (1.87)	Appropriate
A gastrointestinal ultrasound score should be calculated based on the most severely affected segment.	7 (1.60)	Appropriate
The most affected segment(s) before and after treatment should be captured for central reading.	8 (0.93)	Appropriate
The same segments before and after treatment should be captured for central reading.	8 (0.47)	Appropriate
A gastrointestinal ultrasound score should be calculated by summing the total score of each segment examined.	7 (1.87)	Appropriate
Each examined segment should be scored and then an overall activity score should be calculated as both the most affected segment and overall activity are important for assessing activity and determining treatment response.	8 (0.73)	Appropriate
A gastrointestinal ultrasound score should be calculated by dividing the sum of the individual segments by the number of segments explored.	5 (1.80)	Uncertain
A GIUS index of Crohn's disease activity should be designed to be easily calculated.	9 (0.60)	Appropriate
A gastrointestinal ultrasound index should be calculated with a single numeric score to indicate disease activity.	8 (0.80)	Appropriate
Extraluminal complications such as fistula or abscess should not be part of a luminal activity score.	8 (1.80)	Appropriate
Gastrointestinal ultrasound can be used as a screening tool to exclude patients from clinical trials if they have exclusionary criteria such as an abscess or fistula.	7 (0.87)	Appropriate
Gastrointestinal ultrasound can be used to qualify patients for clinical trials when diseased segments cannot be adequately assessed by ileocolonoscopy.	8 (0.73)	Appropriate
A gastrointestinal ultrasound activity index of Crohn's disease can be used to define remission.	8 (1.07)	Appropriate
A gastrointestinal ultrasound activity index of Crohn's disease can be used to define response.	8 (0.67)	Appropriate
When assessing response to induction therapy, gastrointestinal ultrasound should be done at: 4 weeks	6 (2.07)	Uncertain
When assessing response to induction therapy, gastrointestinal ultrasound should be done at: 6 weeks	5 (1.40)	Uncertain
When assessing response to induction therapy, gastrointestinal ultrasound should be done at: 8 weeks	7 (1.33)	Appropriate
When assessing response to induction therapy, gastrointestinal ultrasound should be done at: 12 weeks	8 (1.13)	Appropriate
When assessing response to induction therapy, gastrointestinal ultrasound should be done at: 16 weeks	7 (1.53)	Appropriate
When assessing response to maintenance therapy, gastrointestinal ultrasound should be done at: 26 weeks	8 (0.73)	Appropriate
When assessing response to maintenance therapy, gastrointestinal ultrasound should be done at: 52 weeks	9 (0.80)	Appropriate

Remission should be defined as a combination of bowel wall thickness normalisation (< 3.0 mm) and no bowel wall vascularity on colour Doppler imaging.	7 (1.33)	Appropriate
Remission should be defined as a combination of bowel wall thickness normalisation (< 3.0 mm), no bowel wall vascularity on colour Doppler imaging, and no mesenteric inflammatory fat.	7 (1.27)	Appropriate
Remission should be defined as bowel wall thickness normalization (<3.0mm) alone.	5 (1.93)	Uncertain
Remission should be defined as a combination of bowel wall thickness normalisation (even if > 3.0 mm), no bowel wall vascularity on colour Doppler imaging and no mesenteric inflammatory fat.	8 (1.27)	Appropriate
Sonographic remission should require complete resolution of mesenteric hyper-echogenicity and lymphadenopathy.	5 (1.67)	Uncertain
Response should be defined as a combination of the magnitude of the decrease in bowel wall thickness and colour Doppler activity.	8 (1.00)	Appropriate

Abbreviations: GIUS, gastrointestinal ultrasound; MAD, mean absolute-deviation from the median

Table 5. Paediatric Considerations

Item	Summary (MAD)	Rating
The same general considerations apply when performing gastrointestinal ultrasound in adult and paediatric Crohn's disease patients.	8 (0.59)	Appropriate
The same method for scoring bowel wall thickness should be used in both adult and paediatric patients.	8 (0.94)	Appropriate
The same method for scoring colour Doppler imaging should be used in both adult and paediatric patients.	8 (0.88)	Appropriate
The same method for scoring bowel wall stratification should be used in both adult and paediatric patients.	8 (0.71)	Appropriate
The same method for scoring mesenteric inflammatory fat should be used in both adult and paediatric patients.	8 (0.71)	Appropriate
The same method for scoring bowel wall compressibility should be used in both adult and paediatric patients.	6 (1.65)	Uncertain
The same method for scoring mesenteric lymphadenopathy should be used in both adult and paediatric patients.	5 (2.18)	Uncertain
The same method for scoring loss of small bowel peristalsis should be used in both adult and paediatric patients.	8 (1.06)	Appropriate
The same method for scoring absence of colonic haustra coli should be used in both adult and paediatric patients.	8 (0.94)	Appropriate
The same method for scoring presence of complications such as abscess or fistula should be used in both adult and paediatric patients.	8 (0.65)	Appropriate
The same gastrointestinal ultrasound index can be used in both adult and paediatric patients.	8 (1.00)	Appropriate
The same remission criteria can be used in both adult and paediatric patients.	8 (0.80)	Appropriate
The same response criteria can be used in both adult and paediatric patients.	8 (0.87)	Appropriate

Abbreviations: MAD, mean absolute-deviation from the median.

STATEMENT OF INTERESTS

Authors' declaration of personal interests:

TMG has received support through provision of an Australian Government research training program scholarship and grant support from Janssen.

VJ has received has received consulting fees from AbbVie, Eli Lilly, GlaxoSmithKline, Arena pharmaceuticals, Genetech, Pendopharm, Pfizer, Fresenius Kabi, Bristol Myers Squibb, Roche, Ferring, Sandoz, Merck, Takeda, Janssen, Alimentiv Inc (formerly Robarts Clinical Trials), Topivert, Celltrion, Mylan, Gilead; speaker's fees from Takeda, Janssen, Shire, Ferring, AbbVie, Pfizer.

BGF has received grant/research support from AbbVie Inc., Amgen Inc., AstraZeneca/MedImmune Ltd., Atlantic Pharmaceuticals Ltd., Boehringer-Ingelheim, Celgene Corporation, Celltech, Genentech Inc/Hoffmann-La Roche Ltd., Gilead Sciences Inc., GlaxoSmithKline (GSK), Janssen Research & Development LLC., Pfizer Inc., Receptos Inc./Celgene International, Sanofi, Santarus Inc., Takeda Development Center Americas Inc., Tillotts Pharma AG, and UCB; consulting fees from Abbott/AbbVie, Akebia Therapeutics, Allergan, Amgen, Applied Molecular Transport Inc., Aptevo Therapeutics, Astra Zeneca, Atlantic Pharma, Avir Pharma, Biogen Idec, BioMx Israel, Boehringer-Ingelheim, Bristol-Myers Squibb, Calypso Biotech, Celgene, Elan/Biogen, EnGene, Ferring Pharma, Roche/Genentech, Galapagos, GiCare Pharma, Gilead, Gossamer Pharma, GSK, Inception IBD Inc, JnJ/Janssen, Kyowa Kakko Kirin Co Ltd., Lexicon, Lilly, Lycera BioTech, Merck, Mesoblast Pharma, Millennium, Nestle, Nextbiotix, Novonordisk, Pfizer, Prometheus Therapeutics and Diagnostics, Progenity, Protagonist, Receptos, Salix Pharma, Shire, Sienna Biologics, Sigmoid Pharma, Sterna Biologicals, Synergy Pharma Inc., Takeda, Teva Pharma, TiGenix, Tillotts, UCB Pharma, Vertex Pharma, Vivelix Pharma, VHsquared Ltd., and Zyngenia; speakers bureau fees from Abbott/AbbVie, JnJ/Janssen, Lilly, Takeda, Tillotts, and

UCB Pharma; is a scientific advisory board member for Abbott/AbbVie, Allergan, Amgen, Astra Zeneca, Atlantic Pharma, Avaxia Biologics Inc., Boehringer-Ingelheim, Bristol-Myers Squibb, Celgene, Centocor Inc., Elan/Biogen, Galapagos, Genentech/Roche, JnJ/Janssen, Merck, Nestle, Novartis, Novonordisk, Pfizer, Prometheus Laboratories, Protagonist, Salix Pharma, Sterna Biologicals, Takeda, Teva, TiGenix, Tillotts Pharma AG, and UCB Pharma; and is the Senior Scientific Officer of Alimentiv Inc. (formerly Robarts Clinical Trials, Inc.).

CEP is an employee of Alimentiv, Inc. (formerly Robarts Clinical Trials, Inc.).

TMN is an employee of Alimentiv, Inc. (formerly Robarts Clinical Trials, Inc.).

LG employee of Alimentiv, Inc. (formerly Robarts Clinical Trials, Inc.).

AKA has no conflicts of interest to declare.

JB has received consulting fees from AbbVie, Pfizer, Bristol Myers Squibb, Roche, Ferring, Sandoz, Takeda, Janssen, Celltrion, Gilead; speaker's fees from Takeda, Janssen, Shire, Ferring, AbbVie, Pfizer.

BC has recieved Grant/Research support/Speaker fees from AbbVie, Ferring, Janssen, Takeda, Pfizer, Gilead, Chiesi, Celltrion and Novartis; research support, speaker fees and consultancy payments from AbbVie, Janssen, Takeda, Ferring, Pfizer, Gilead, Chiesi, Celltrion, and Novartis.

ABF has received research grants from Abbvie Pty Ltd, speaking or consulting fees from Abbvie Pty Ltd, Janssen-Cilag Pty Ltd, Takeda Pharmaceuticals Australia Pty Ltd and Shire Australia Pty Ltd and travel grants from Pfizer and Ferring.

TK has received Grant/Research support/Consulting fees or Speaker fees from AbbVie, Celgene, Celltrion, Falk Foundation, Ferring, Janssen, Gilead, Pfizer, Takeda, Roche.

AL has no conflicts of interest to declare.

PL has received speaker fees from Pfizer, AbbVie, Janssen-Cilag Pty, Ltd.

CM has received lecture fees and/or consultancy fees from AbbVie, Arena, Biogen, Celgene, Ferring, Dr. Falk Pharma GmbH, Gilead, Galapagos, Janssen, MSD Sharp & Dome GmbH, Pfizer, Roche, Takeda Pharma GmbH, Vifor.

KLN has received advisory board fees from AbbVie, Janssen, Pfizer, Ferring; Takeda; speaker's fees from AbbVie, Janssen; Pfizer and Takeda; and research support from AbbVie.

JR has received consulting fees from AbbVie, Janssen, Takeda, TiGenix and Alimentiv Inc. (formerly Robarts Clinical Trials, Inc.); speaker's fees from Janssen, Gilead, and Takeda; research grants from AbbVie and Genentech.

KT has no conflicts of interest to declare.

ST has received fees from Alimentiv Inc. (formerly Robarts Clinical Trials, Inc.); research support from the UCLH biomedical researcher centre; holds share options in Motilent and is a NIHR senior investigator.

LW has no conflicts of interest to declare.

RW has received grants/speaker fees from Pfizer, AbbVie, Janssen, Takeda.

SRW has received partial research support from Samsung and equipment support from Samsung, Siemens, and Philips.

EW has no conflicts of interest to declare.

RVB has received Grant/Research support/Speaker fees (all paid to employer for research support): AbbVie, Ferring, Janssen, Shire, Takeda, Emerge Health; shareholder in Biomebank.

CM has received consulting fees from AbbVie, Janssen, Pfizer, Alimentiv Inc. (formerly Robarts Clinical Trials, Inc.), and Takeda; and speaker's fees from AbbVie, Janssen, Pfizer, and Takeda.

Declaration of funding interest: Funding was received from Janssen for this study

AUTHORSHIP STATEMENT:

Guarantor of the article: Dr. Robert V. Bryant and Dr. Christopher Ma

Specific author contributions:

Development of the study concept and design: TMG, VJ, BGF, RVB, CM;

Study supervision: VJ, RVB;

Participation in the panel: AA, JB, BC, AF, TK, AL, PL, CM, KLN, JR, KT, ST, LW, RW, SRW, EW;

Data collection and analysis: LG, CEP;

Drafting of the manuscript: TMG, VJ, CEP, TMN, RVB, CM.

All authors performed critical revision of the manuscript for intellectual content and approved the final draft. All authors approved the final version of the manuscript including the authorship list.