**TITLE: Defining Component Items and Scoring Conventions for an Endoscopic Index in Crohn's Disease: An International Modified RAND/UCLA Appropriateness Study**

**AUTHORS:** Reena Khanna, MD, MSc[1*]; Christopher Ma, MD, MPH[2,3,4*]; Malcolm Hogan, MSc[4], Guangyong Zou, PhD[4,5,6]; Talat Bessissow, MD[7]; Brian Bressler, MD[8,9]; Jean-Frederic Colombel, MD[10]; Silvio Danese, MD, PhD[11]; Marco Daperno, MD[12]; James E. East, MD(Res) FRCP [13 + 13a]; Lawrence Hookey, MD[14]; Edward Loftus, Jr., MD[15]; John W. D. McDonald, MD[3]; Remo Panaccione, MD[1]; Laurent Peyrin-Biroulet, MD, PhD[16,17]; Matt Rutter, MD, PhD[18]; Bruce E. Sands, MD[19]; Severine Vermeire[20,21]; Julie Rémillard, MSc[4]; Stefanie McFarlane[4]; William J. Sandborn, MD[22]; Geert R. D'Haens, MD, PhD[3,23]; Brian G. Feagan, MD[1,4,5]; Vipul Jairath, MBChB, DPhil[1,4,5]

**AUTHOR AFFILIATIONS:**

[1]Department of Medicine, Division of Gastroenterology, University of Western Ontario, London, Ontario, Canada; [2]Division of Gastroenterology & Hepatology, Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; [3]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; [4]Alimentiv Inc, London, Ontario, Canada; [5]Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada; [6]Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; [7]Division of Gastroenterology, McGill University Health Centre, Montreal, Quebec, Canada; [8]Division of Gastroenterology, Department of Medicine, University of British Colombia, Vancouver, British Colombia, Canada; [9]St. Paul's Hospital, Vancouver, British Colombia; [10]The Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, New York,

USA; [11]Gastroenterology and Gastrointestinal Endoscopy Unit, IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy; [121]Division of Gastroenterology, Ospedale Mauriziano Umberto I, Torino, Italy; [13]Translational Gastroenterology Unit, Nuffield Department of Medicine, John Radcliffe Hospital, Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK; [13a] Division of Gastroenterology and Hepatology, Mayo Clinic Healthcare, London, UK; [14]Department of Medicine, Division of Gastroenterology, Queen's University, Kingston, Ontario, Canada; [15]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota, USA; [16]Nancy University Hospital, Department of Gastroenterology, Nancy, France; [17]Université de Lorraine, Inserm, Nutrition-Genetics and Environmental Risk Exposure (NGERE), Nancy, France; [18]Department of Gastroenterology, North Tees and Hartlepool NHS Foundation Trust, Stockton on Tees, UK; [19]Dr. Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [20]Department of Gastroenterology and Hepatology, University Hospitals Leuven, Leuven, Belgium; [21]Department of Chronic Diseases Metabolism and Ageing, Catholic University Leuven, Leuven, Belgium; [22]Division of Gastroenterology, University of California San Diego, La Jolla, California, USA; [23]Department of Gastroenterology and Hepatology, Amsterdam University Medical Centre, Academic Medical Centre, Amsterdam, the Netherlands

*CM and RK should be considered joint first authors

**CORRESPONDING AUTHOR**

Vipul Jairath, MBChB, DPhil

Department of Medicine, Division of Gastroenterology

Department of Epidemiology and Biostatistics

Western University

Suite 200, 100 Dundas Street

London, Ontario, N6A 5B6, Canada

Email: vjairath@uwo.ca

Phone: 519-685-8500

## DATA AVAILABILITY

All data relevant to the study are including in the article.

## ACKNOWLEDGMENT

The authors thank Leonardo Guizzetti, PhD, for assisting with the statistical analysis.

## FUNDING

JEE is funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health.

## CONFLICTS OF INTEREST

RK

CM has received consulting fees from AbbVie, Alimentiv, Amgen, AVIR Pharma Inc, BioJAMP, Bristol Myers Squibb, Celltrion, Ferring, Fresenius Kabi, Janssen, McKesson, Mylan, Takeda, Pendopharm, Pfizer, Roche, Sanofi; speaker's fees from AbbVie, Amgen, AVIR Pharma Inc,

Alimentiv, Bristol Myers Squibb, Ferring, Fresenius Kabi, Janssen, Takeda, Pendopharm, and Pfizer; royalties from Springer Publishing; research support from Ferring, Pfizer.

MH

GZ has received consulting fees from Alimentiv Inc.

TB

BB

JFC

SD

MD

JEE has served on clinical advisory board for Paion; has served on the clinical advisory board and has share options in Satisfai Health; and reports speaker fees from Falk, Jannsen and Medtronic.

LH

EL

JWDM

RP

LPB

MR

BES

SV

JR is an employee of Alimentiv Inc.

SM is an employee of Alimentiv Inc.

WJS

GRD has served as advisor for Abbvie, Ablynx, Alimentiv, Allergan, Amakem, Amgen, AM Pharma, Arena Pharmaceuticals, AstraZeneca, Avaxia, Biogen, Bristol Myers Squibb, Boehringer Ingelheim, Celgene/Receptos, Celltrion, Cosmo, Covidien/Medtronic, Ferring, DrFALK Pharma, Eli Lilly, Engene, Galapagos, Genentech/Roche, Gilead, Glaxo Smith Kline, Hospira/Pfizer, Immunic, Johnson and Johnson, Lycera, Medimetrics, Millennium/Takeda, Mitsubishi Pharma, Merck Sharp Dome, Mundipharma, Nextbiotics, Novo Nordisk, Otsuka, Pfizer/Hospira, Photopill, Prometheus laboratories/Nestle, Progenity, Protagonist, Salix, Samsung Bioepis, Sandoz, Seres/Nestle, Setpoint, Shire, Teva, Tigenix, Tillotts, Topivert, Versant and Vifor; received speaker fees from Abbvie, Biogen, Ferring, Johnson and Johnson, Merck Sharp Dome, Mundipharma, Norgine, Pfizer, Samsung Bioepis, Shire, Millennium/Takeda, Tillotts and Vifor.

BGF is a scientific advisory board member for AbbVie, Allergan, Amgen, AstraZeneca, Avaxia Biologics Inc., Boehringer-Ingelheim, Bristol-Myers Squibb, Celgene, Elan, Biogen, Ferring, Genentech–Roche, Janssen–Johnson & Johnson, Merck, Millennium, Nestlé, Novo Nordisk, Novartis, Pfizer, Prometheus, Protagonist, Receptos, Salix, Sigmoid Pharma, Takeda, Teva, TiGenix, Tillotts Pharma, and UCB Pharma; consulting fees from AbbVie, Actogenix, Akros, Albireo Pharma, Allergan, Amgen, AstraZeneca, Avaxia Biologics, Avir Pharma, Axcan, Baxter Healthcare, Biogen Idec, Boehringer Ingelheim, Bristol-Myers Squibb, Calypso Biotech, Celgene, Elan–Biogen, EnGene, Ferring, Genentech–Roche, GiCare Pharma, Gilead Sciences, Given Imaging, GlaxoSmithKline, Ironwood, Janssen Biotech–Centocor, Janssen–Johnson & Johnson, Kyowa Hakko Kirin, Eli Lilly, Merck, Mesoblast Pharma, Millennium, Nestlé, Novo Nordisk, Novartis, Pfizer, Prometheus, Protagonist, Receptos Salix, Sanofi, Shire, Sigmoid Pharma, Synergy Pharma, Takeda, Teva, TiGenix, Tillotts Pharma, UCB Pharma, Vertex, VHsquared, Wyeth, Zealand, and Zygenia; lecture fees from AbbVie, Janssen–Johnson & Johnson, Takeda,

and UCB Pharma, and grant support from AbbVie, Amgen, AstraZeneca, Bristol-Myers Squibb, Janssen Biotech–Centocor, Janssen–Johnson & Johnson, Pfizer, Receptos, Sanofi, and Takeda, and is the Senior Scientific Officer of Alimentiv Inc.

VJ has received has received consulting/advisory board fees from AbbVie, Alimentiv Inc, Arena Pharmaceuticals, Asahi Kasei Pharma, Asieris, AstraZeneca, Bristol Myers Squibb, Celltrion, Eli Lilly, Ferring, Flagship Pioneering, Fresenius Kabi, Galapagos, GlaxoSmithKline, Genentech, Gilead, Janssen, Merck, Mylan, Pandion, Pendopharm, Pfizer, Protagonist, Reistone Biopharma, Roche, Sandoz, Second Genome, Takeda, Teva, Topivert, Ventyx, and Vividion; and speaker's fees from Abbvie, Ferring, Galapagos, Janssen Pfizer Shire, Takeda, and Fresenius Kabi.

Alimentiv Inc. is an academic gastrointestinal contract research organization (CRO), operating under the Alimentiv Health Trust. Alimentiv Inc. provides comprehensive clinical trial services, precision medicine offerings, and centralised imaging solutions for endoscopy, histopathology, and other imaging modalities. The beneficiaries of the Alimentiv Health Trust are the employees of the enterprises it holds. None of the authors are beneficiaries of the Alimentiv Health Trust. BGF, CM, GRD, GZ, RK, RP, VJ, and WJS are consultants to Alimentiv Inc. and have a primary academic appointment; they do not hold equity positions or shares in Alimentiv Inc.

**AUTHORSHIP STATEMENT**

All authors have made substantial contributions and reviewed and approved the final manuscript.

Article guarantor: VJ

Conception and design: RK, WJS, GRD, BGJ, VJ

Analysis/interpretation: all authors

Manuscript drafting: CM, SM

Manuscript editing for important intellectual content: all authors

## ETHICS APPROVAL

Not required

## PATIENT CONSENT STATEMENT

Not applicable

## COUNTS

Manuscript: 3593 (no limit)

References: 27 (no limit)

Tables and Figures: 3 (no limit)

Supplemental Materials: 3 (no limit)

**STRUCTURED SUMMARY (247/250 WORDS)**

**Background and Aims**: Endoscopic assessment of disease activity is integral for evaluating treatment response in patients with Crohn's disease (CD). We aimed to define appropriate items for evaluating endoscopic activity and conventions for consistent endoscopic scoring rules in CD.

**Methods**: A two-round modified Research and Development/University of California Los Angeles appropriateness methodology study was conducted. A panel of 15 gastroenterologists used a 9-point Likert scale to rate the appropriateness of statements pertaining to the Simple Endoscopic Score for Crohn's Disease (SES-CD), Crohn's Disease Endoscopic Index of Severity (CDEIS), and additional items relevant to endoscopy scoring in CD. Each statement was voted as appropriate, uncertain, or inappropriate based on the median panel rating and presence of disagreement.

**Results**: Panellists voted that it is appropriate for all ulcers to contribute to endoscopic scoring in CD, including aphthous ulcers, ulcerations at a surgical anastomosis, and anal canal ulcers (scored in the rectum). Mucosal healing should reflect an absence of ulcers. Stenosis should be defined by an impassable narrowing, and when occurring at the junction between two segments, should be scored in the distal segment. The scoring of scarring and inflammatory polyps/pseudopolyps as affected mucosa was considered inappropriate. The optimal method for evaluating the total ulcerated and affected surface area and defining ulcer depth remains uncertain.

**Conclusions**: We outlined scoring conventions for the SES-CD and CDEIS, noting both scores have limitations. Therefore, we identified priorities for future research and steps for developing and validating a more representative endoscopic index in CD.

**KEYWORDS**

Crohn's disease; clinical trials; endoscopy; index; outcomes; ulcer; remission

## INTRODUCTION

Endoscopic evaluation is an integral component of disease activity assessment in patients with Crohn's disease (CD). Recognizing that symptoms are poorly correlated with objective measures of inflammation and that treating to symptomatic improvement alone does not change the natural history of CD, endoscopy has been increasingly adopted as a therapeutic endpoint.[1-3] In pivotal, registrational clinical trials, endoscopy is now routinely performed to qualify patients with active CD at enrolment and to assess for endoscopic response as a co-primary endpoint.[4] The use of endoscopy has also increased in clinical practice, with the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE)-II guidelines recommending endoscopic healing as the appropriate treatment target.[5] However, although endoscopy is more objective than symptom-based assessments, there are substantive limitations of the existing methods for scoring endoscopic disease activity.[6]

The two most commonly used endoscopic indices, the Simple Endoscopic Score for Crohn's Disease (SES-CD) and the Crohn's Disease Endoscopic Index of Severity (CDEIS), both involve the assessment of ulcerations, ulcerated and affected surface area, and the presence of stenosis, scored across the ileum and multiple colonic segments.[7,8] While the SES-CD and CDEIS have been shown to have high inter-observer reliability and responsiveness to treatments of established efficacy,[9,10] these indices have important limitations: (1) processes for handling missing or inaccessible segments are uncertain; (2) reliability and responsiveness are lower for the stenosis component compared to other items[11]; (3) thresholds for defining clinically relevant changes in endoscopic activity and endoscopic remission are uncertain[12]; (4) scoring conventions are not standardised for all index items (e.g., which types of lesions constitute affected surface area are

unclear); and (5) the total score is disproportionately impacted by the number of involved segments as opposed to the severity of findings in a single segment. Resultingly, the total SES-CD and CDEIS scores may inaccurately reflect overall endoscopic disease activity or inadequately capture response to treatment.

Given these limitations, we conducted a modified Research and Development/University of California, Los Angeles (RAND/UCLA) appropriateness methodology (RAM) study with international gastroenterologists specialising in CD endoscopy to (1) determine the endoscopic measures most appropriate for evaluating disease activity in CD; and (2) define the conventions that can be feasibly applied to facilitate consistent application of endoscopic scoring.

## MATERIALS AND METHODS

The RAM process is a widely accepted, evidence-based approach that employs a modified Delphi panel method to determine the appropriateness and face validity of a list of generated statements.[13] An international panel of 15 gastroenterologists from the United States, Canada, and Europe were invited to participate in the modified RAM process. Panellists were selected based on their expertise in CD endoscopy, which took precedence over geographic representation.

### Statement generation

The modified RAM used an online survey comprised of statements that were informed by prior systematic literature reviews[9,14] and expert opinion. The SES-CD and CDEIS (**Supplemental Tables 1 and 2**) were identified as the most used endoscopic instruments for the evaluation of CD activity, and statements relevant to the definitions and scoring of these indices and their component items were used to design the original survey. Additional items reported in the literature or unpublished items deemed relevant were also included (**Supplemental Table 3**).

### Panel meetings and surveys

All panel meetings were conducted over videoconference. In an initial meeting, the methodology and voting processes were summarised for all panellists, and the original list of statements for voting generated from the systematic reviews was discussed. Panellists were encouraged to comment on the clarity of these initial statements, provide feedback on uncertain wording, and to identify any additional potentially relevant items to be included. The list of statements was then revised according to the panellists' feedback and included in an online survey. Panellists rated the

*appropriateness* of each statement on a 9-point Likert scale (1, highly inappropriate; 5, uncertain; 9, highly appropriate).

For analysis, statements were classified as inappropriate, uncertain, or appropriate according to both the median panel rating and the presence of any disagreement. Inappropriate statements were defined by median ratings of 1 to 3 without disagreement; uncertain statements were defined by median ratings of 3.5 to 6.5 without disagreement or any median rating with disagreement; appropriate statements were defined by median ratings of 7 to 9 without disagreement. Disagreement was defined as at least five panellists rating the statement at each of the extreme ends of the scale (i.e., 1-3 and 7-9).

Anonymised group results were distributed to the panel, and the uncertain statements were specifically discussed in a second moderated videoconference. Panellists were encouraged to review each statement and discuss the rationale for their responses. Importantly, the modified RAM process does not *force* a consensus but aims to evaluate the appropriateness of each statement. After the discussion, the survey was revised and recirculated for a second round of voting by the panel. Panellists were encouraged to consider their original vote, the group vote, and the feedback from the discussions. For the final analysis, statement classifications were defined as described for the first-round survey. Median appropriateness ratings of the modified RAM panel were summarised, and the corresponding interquartile ranges were calculated.

## RESULTS

Of the 104 statements rated by panellists in the first-round survey, 36 (35%) statements were considered appropriate, 49 (47%) statements were uncertain (including 11 statements with disagreement), and 19 (18%) statements were inappropriate. In the second-round survey comprising 116 statements, 45 (39%) statements were voted as appropriate, 40 (34%) statements were uncertain (including 15 statements with disagreement), and 31 (27%) statements were inappropriate.

### Appropriateness of the SES-CD and component items

Appropriateness ratings of statements related to the SES-CD are summarised in **Table 1**. Different configurations of ulcer size scoring were considered, including both categorical and continuous measures. The panellists voted that the current schema for ulcer size (0=none, 1=aphthous ulcers [0.1-0.5 cm], 2=large ulcers [>0.5-2 cm], 3=very large ulcers [>2 cm]) is appropriate for scoring the largest visualised ulcer. The use of a greater number of categories to differentiate 0.5 to 2 cm ulcerations (separately score 0.5-1.0, 1.0-1.5, 1.5-2.0 cm ulcers) was voted as inappropriate, given the poor feasibility of implementing these categories in routine care and the substantial inter-rater variability in size estimation. The scoring of ulcers at the anastomotic rim irrespective of any suspicion regarding ischemic versus inflammatory pathogenesis was voted as appropriate, as was the scoring of anal canal ulcerations within the rectal segment. For the estimation of ulcerated surface, it was uncertain whether the current categories (0=none, 1=<10%, 2=10%-30%, 3=>30%) or a continuous evaluation on a 100-mm visual analogue scale (VAS) is more appropriate. This uncertainty was based on concerns regarding the subjectivity of ulcerated surface assessment, the difficulties in scoring ulcerated area in longer segments, and the variability of this assessment

when shorter versus deeper ileal intubation is performed. Furthermore, a VAS was considered difficult to independently validate. The panellists discussed that a comparison of operating properties among the ulcerated surface assessment permutations is required.

The panellists voted that the current categorical configuration (0=unaffected segment, 1=<50%, 2=50%-75%, 3=>75%) for scoring affected area is appropriate due to concerns regarding the accuracy of estimating the percentage of affected surface in a continuous fashion. Scoring the anus in the rectal segment was also deemed appropriate among the panel. In addition, the panellists discussed specific scenarios where scoring may be challenging due to anatomic location. To avoid double scoring and the resulting "over-inflation" of the total SES-CD score, the scoring of ulcerations at the *junction* of two colonic segments only in the most affected segment was voted as appropriate. However, the panellists also voted that very large ulcerations clearly extending over two segments should contribute to ulcer scoring in both segments. The panellists determined that the scoring of ulcerations affecting the ileocecal valve (ICV) in the distal colonic segment is appropriate. Ulcerations or affected surface should not be scored in a proximal segment for which an impassable narrowing precludes evaluation; this statement was deemed appropriate due to difficulties in accurately assessing the ulcerated and affected area when the segment is not fully visualized.

The panellists believed that an appropriate definition of *narrowing* is a clear decrease in lumen diameter (compared with the previous and subsequent bowel) that does not insufflate with air and may be associated with trauma when passing it; and that an appropriate definition of *multiple narrowings* is described by clearly repeated areas of narrowed diameter with intervening open

areas. An impassable *stenosis* should be defined as an impassable narrowing without dilation. It was agreed that stenosis at the junction between two segments (e.g., hepatic or splenic flexure) should be scored in only the distal segment. Similarly, stenosis of the ICV or anastomosis should be scored in the colonic segment distal to the stenosis.

**Appropriateness of the CDEIS and component items**

Appropriateness ratings of the statements related to the CDEIS are summarised in **Table 2**. While differentiating superficial versus deep ulcerations was deemed appropriate for the CDEIS, the operationalization of this was uncertain. Specifically, measuring the depth of an ulceration continuously was discussed but was believed to be infeasible, and panellists reviewed different definitions of deep ulcerations including visible cratering and the involvement of deeper layers beyond the mucosa (*i.e.,* involvement of the submucosa). The scoring of aphthous ulcerations as superficial ulcers in the CDEIS was voted as appropriate, as was the scoring of anal ulcers as part of the rectum and the inclusion and scoring of ulcerations at the anastomotic rim.

There was disagreement amongst panellists regarding whether the surface involved by the disease and the ulcerated surface, both measured using 10-cm linear scales, are the optimal descriptors of endoscopic disease activity. Although these descriptors have been used in historical studies and have face validity for endoscopic disease activity, some panellists expressed the ambiguity of these measures and the difficulty of reliably evaluating disease involvement using a 10-cm scale, particularly in some circumstances such as for clustered or patchy lesions. Components that should be included in the assessment of affected disease surface are listed under *Additional Items* in **Table 3** and summarised in the following section.

As with the SES-CD, the CDEIS statements deemed appropriate included the definition of stenosis as an impassable narrowing, the scoring of stenosis at the junction of two segments in the distal segment, the scoring of ulcerations across two segments once on the most affected side (unless very large ulcerations clearly involve multiple segments), and the scoring of ulcerations on the ICV in the right colon. The panel determined that patients participating in luminal inflammatory (not fibrostenotic) CD trials should not be considered eligible if a dilation is required at enrolment because there is a high risk of re-stenosis after dilation, there are variations between sites with respect to dilation ability and comfort, and differences in dilated anastomotic versus primary strictures. Panellists determined that any mucosal break within a stricture should be considered an ulcerated stricture, recognizing that determining the precise size and depth of ulcerations within a tight narrowing may be challenging.

**Items relevant to both indices (SES-CD and CDEIS) and additional items**

Appropriateness ratings of the statements applicable to both endoscopic indices and the additional items from the second-round survey are summarised in **Table 3**. There was agreement that pseudopolyps and scarring should not be considered for inclusion in the affected area for disease activity measures and that ulcerations limited to the top of an inflammatory pseudopolyp should not be considered part of the ulceration or ulcerated surface. In discussions, panellists believed that these items were not representative of active CD and should not be scored as such. There was uncertainty regarding appropriateness of the inclusion of oedema in the affected area score, given that this item may be challenging to consistently assess between observers.

For the scoring of ulcerations, the panel agreed on the appropriateness for local endoscopists to pause at the largest ulcer from each segment for a minimum of one second to facilitate accurate centralised reading in trial settings. The measurement of ulcer size using wire measurement tools was considered infeasible with regard to both cost and time, and the appropriateness was uncertain for benchmarking the lesion against open biopsy forceps given the variability among biopsy sizes.

Several issues pertaining specifically to the ileum and rectum were discussed. Panellists identified the inappropriateness of requiring rectal retroflexion in every case, given that retroflexion may not be safely performed in some scenarios (e.g., very deep or large rectal ulcers, narrow rectal vault). However, the panel discussed that careful evaluation of the rectum and anal canal should be performed in forward view if retroflexion is not performed. For the ileum, it was discussed that a minimum depth of intubation should be considered (10 cm), but the feasibility of measuring the depth of intubation in all cases was uncertain due to challenges that arise when there is looping of the colonoscope. The panel agreed that even aphthous ulcers within the ileum are important for measuring endoscopic disease activity, but there was disagreement on whether lesions visualised in the distal ileum or neoterminal ileum beyond an impassable stricture at the ICV or anastomosis should contribute to endoscopic scoring. Indeed, clear ulcerations may be visualised beyond an impassable stricture that contribute to symptoms and the patient's disease course. However, some panellists identified that scoring lesions in the ileum proximal to an impassable stricture may inaccurately reflect the proportion of the ileum/neoterminal ileum that is truly involved.

The panel agreed that the definition of healed mucosa should include an absence of ulceration. However, the panellists' votes identified the inappropriateness of including an absence of

inflammatory polyps, erythema, friability, or granularity in the definition of healed mucosa, which may be infeasible to achieve with current therapies. In addition, some of these components (e.g., erythema) are difficult to score or have poor reliability among observers, and inflammatory polyps may not reflect ongoing active CD. The panel also discussed the weighting of different segments, recognizing that all segments may not be visualised in all procedures. The use of "worst-score" imputation methods was deemed inappropriate for the SES-CD and CDEIS, given that such methods systematically increase the total score (and may be inappropriately used to meet minimum score criteria at trial enrolment). The heavier weighting of lesions in the rectum and ileum was voted as inappropriate and uncertain, respectively. Although it was identified that these lesions may be associated with a greater burden of symptoms and a negative impact on prognosis, inter-rater differences in scoring in these segments (with unique considerations with respect to visualization as previously identified) would be compounded with increased weighting. The panel also discussed whether all colonic segments should be pooled, with items scored across the entire colon rather than in multiple colonic segments. There was uncertainty regarding this method, highlighting that it should be evaluated in future research.

The panellists agreed that the quality of bowel preparation and washing should be graded for each procedure and that a minimum percentage of the mucosa should be visualised for scoring to be performed. A VAS was also suggested for assessing video quality in clinical trials with centralised endoscopy scoring because of the potential poor quality of the video recordings, even with sufficient bowel preparation.

## DISCUSSION

The accurate assessment of endoscopic disease activity in CD is paramount for clinical decision making as part of a treat-to-target strategy and for evaluating the efficacy of novel therapies as a coprimary endpoint in clinical trials. Although endoscopic scoring systems for CD have been developed and used for the approval of advanced therapies, there are substantive limitations of both the SES-CD and CDEIS. Here, we used a modified RAND/UCLA appropriateness methodology to develop guidance on the component items that have face validity for evaluating endoscopic CD activity. We also aimed to standardise scoring conventions to minimise variability in the application of the SES-CD and CDEIS (particularly in challenging but commonly encountered clinical scenarios such as stricturing CD) and to identify areas for future research, including steps for developing the next iteration of a validated endoscopic CD index.

The panellists discussed at length which endoscopic lesions constitute active CD. Ulcerations are clearly a hallmark finding in CD, which is reflected in the definition of mucosal healing that has historically incorporated the absence of ulcers.[10,12,15-17] The size, depth, and extent of ulceration were considered to be appropriate components of an endoscopic activity index in CD, although several limitations should be noted. First, accurately estimating ulcer size or ulcerated surface area can be challenging, and therefore, the panel voted to use categorical scoring rather than continuous estimates of these parameters. Second, evaluating three-dimensional ulcer depth from a two-dimensional image or video may be difficult. While frankly penetrating ulcers are easy to distinguish, the differentiation of more subtle cases requires an operational definition of “deep” versus “superficial,” which was not provided in the development of the original CDEIS.[8] Several different descriptors of “deep” ulcers were considered, including involvement beyond the

muscularis mucosa (as defined in peptic ulcer disease), appearance of a visible crater, and depth beyond that of a closed biopsy forceps.[18,19] Additional research is required to determine which of these definitions may be most reliable and feasible for application in clinical care.

In addition to the size, depth, and extent of ulcerations, the panel also discussed the relative merits of scoring aphthous ulcerations or erosions and whether ulcers in specific segments such as the rectum or ileum should be weighted more heavily. Generally, the panel believed that all ulcerations should contribute to an endoscopic score of CD activity, given that aphthous ulcers, anal canal ulcers, and ulcers along a surgical anastomosis can all be associated with symptoms and/or disease progression.[20,21] The exception to this guidance is for isolated ulcers on the cap of an inflammatory pseudopolyp, which are considered to reflect the re-epithelization and regeneration processes involved in healing rather than active inflammatory CD.[22] Several clinical trial programs have suggested that ileal and rectal ulcerations, especially very large ulcerations >2 cm, are less likely to heal with medical therapy, irrespective of treatment class.[23,24] This observation underlies the premise from a post hoc analysis of three CD trials, where individual component items were differentially weighted in a modification of the SES-CD.[25]

Although weighting is conceptually advantageous for addressing the potential differences in healing between segments, the limitations of weighting should also be considered. First, the weighting of individual components may compound the differences in scoring of less reliable items. For example, the presence of a non-passable stenosis is weighted more heavily in the modified SES-CD but is also the least reliable item in both the SES-CD and CDEIS.[11] The ability for a stenosis to be passed is highly variable and dependent on factors such as insufflation, pressure

applied by the endoscopist, and type of colonoscope used; and morphologically, strictures are better characterised on cross-sectional imaging than on endoscopy.[26] Second, the influence that weighting different components has on responsiveness is unclear because the relative weights have been conditioned on endoscopic remission defined by the SES-CD, rather than on differentiating changed versus unchanged patients in response to treatments of known efficacy.

The SES-CD and CDEIS are based on scoring component items in multiple colonic and ileal segments, which captures the extent of disease. However, scoring by segment may also be limited by missing data from segments that are not visualised due to an impassable stricture, technical challenges preventing procedure completion, poor bowel preparation, or prior surgery. Missing data from non-visualised segments is problematic in trial settings when different segments may be visualised at baseline compared to follow-up. The panel considered different methods for handling missing segments, including the imputation of scores, calculation of the SES-CD or CDEIS "as observed," or the analysis of only matching segments visualised both before and after treatment. Each approach has potential advantages and drawbacks, although we have previously shown empirically that the use of a worst-score imputation method is likely to increase the variance and may bias the treatment effects.[27] The panel agreed that only clearly visualised segments should contribute to scoring because of the reliance of both the ulcerated and affected surface components on accurate visualization of each segment.

Our study has some important strengths. We used a validated modified RAM process and convened a panel of gastroenterologists with expertise in CD endoscopy to address inherent challenges of the SES-CD and CDEIS. The scoring conventions that we addressed are relevant to

clinical and research settings. However, we also acknowledge some key limitations. First, several statements were based on the panellists' opinions rather than empiric evidence. These statements highlight important areas for future research, such as how to best define ulcer depth and how to accurately estimate the percentage of affected or ulcerated surface. In the future, artificial intelligence computerized evaluation may more precisely and reliably estimate affected area, although this requires further validation. Second, the functional application of our recommendations may differ between clinical care and trial settings. For instance, although we do not recommend scoring segments proximal to an impassable stricture, partially visualised disease activity should not be ignored in routine clinical care.

In summary, we have developed conventions for consistently scoring the SES-CD and CDEIS, and we have defined appropriate items with face validity for assessing endoscopic disease activity in CD. Both existing scores have shortcomings, and there remains a need to develop and validate an instrument that is feasible for use in routine clinical care and for defining eligibility criteria in clinical trials and that is responsive to change after treatment with effective medical therapies. Our study has established the framework for developing such an instrument through our identification of the key components that reflect endoscopic activity and the areas of uncertainty that require further research to optimise the corresponding operating properties.

**REFERENCES**

1. Peyrin-Biroulet L, Reinisch W, Colombel JF, et al. Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. Gut 2014;63:88-95.
2. Khanna R, Bressler B, Levesque BG, et al. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. Lancet 2015;386:1825-1834.
3. Neurath MF, Travis SP. Mucosal healing in inflammatory bowel diseases: a systematic review. Gut 2012;61:1619-1635.
4. Abreu MT, Sandborn WJ, IOIBD Defining Endpoints Biomarkers in Inflammatory Bowel Disease Writing Group. Defining endpoints and biomarkers in inflammatory bowel disease: moving the needle through clinical trial design. Gastroenterology 2020;159:2013-2018.
5. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: an update on the selecting therapeutic targets in inflammatory bowel disease (STRIDE) initiative of the International Organization for the Study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. Gastroenterology 2021;160:1570-1583.
6. Khanna R, Ma C, Jairath V, Vande Casteele N, Zou G, Feagan BG. Endoscopic assessment of inflammatory bowel disease activity in clinical trials. Clin Gastroenterol Hepatol 2020.
7. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004;60:505-512.

8. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 1989;30:983-989.

9. Khanna R, Nelson SA, Feagan BG, et al. Endoscopic scoring indices for evaluation of disease activity in Crohn's disease. Cochrane Database Syst Rev 2016:CD010642.

10. Vuitton L, Marteau P, Sandborn WJ, et al. IOIBD technical review on endoscopic indices for Crohn's disease clinical trials. Gut 2016;65:1447-1455.

11. Khanna R, Zou G, D'Haens G, et al. Reliability among central readers in the evaluation of endoscopic findings from patients with Crohn's disease. Gut 2016;65:1119-1125.

12. Ma C, Hanzel J, Panaccione R, et al. CORE-IBD: a multidisciplinary international consensus initiative to develop a core outcome set for randomized controlled trials in inflammatory bowel disease. Gastroenterology 2022.

13. Fitch K, Bernstein SJ, Aguilar MD, et al. *The RAND/UCLA Appropriateness Method User's Manual.* Santa Monica, CA: RAND Corporation; 2001.

14. Khanna R, Bouguen G, Feagan BG, et al. A systematic review of measurement of endoscopic disease activity and mucosal healing in Crohn's disease: recommendations for clinical trial design. Inflamm Bowel Dis 2014;20:1850-1861.

15. Colombel JF, Sandborn WJ, Reinisch W, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. N Engl J Med 2010;362:1383-1395.

16. Rutgeerts P, Van Assche G, Sandborn WJ, et al. Adalimumab induces and maintains mucosal healing in patients with Crohn's disease: data from the EXTEND trial. Gastroenterology 2012;142:1102-1111 e1102.

17. Peyrin-Biroulet L. Mucosal healing in Crohn's disease and ulcerative colitis. Gastroenterol Hepatol (N Y) 2020;16:206-208.

18. Lee JM, Lee KM. Endoscopic diagnosis and differentiation of inflammatory bowel disease. Clin Endosc 2016;49:370-375.

19. Narayanan M, Reddy KM, Marsicano E. Peptic ulcer disease and helicobacter pylori infection. Mo Med 2018;115:219-224.

20. Bachour SP, Shah RS, Lyu R, et al. Mild neoterminal ileal post-operative recurrence of Crohn's disease conveys higher risk for severe endoscopic disease progression than isolated anastomotic lesions. Aliment Pharmacol Ther 2022;55:1139-1150.

21. Allez M, Lemann M, Bonnet J, Cattan P, Jian R, Modigliani R. Long term outcome of patients with active Crohn's disease exhibiting extensive and deep ulcerations at colonoscopy. Am J Gastroenterol 2002;97:947-953.

22. Politis DS, Katsanos KH, Tsianos EV, Christodoulou DK. Pseudopolyps in inflammatory bowel diseases: Have we learned enough? World J Gastroenterol 2017;23:1541-1551.

23. Narula N, Wong ECL, Aruljothy A, et al. Ileal and rectal ulcer size affects the ability to achieve endoscopic remission: a post hoc analysis of the SONIC trial. Am J Gastroenterol 2020;115:1236-1245.

24. Dulai PS, Singh S, Vande Casteele N, et al. Should we divide Crohn's disease into ileum-dominant and isolated colonic diseases? Clin Gastroenterol Hepatol 2019;17:2634-2643.

25. Narula N, Wong ECL, Colombel JF, et al. Predicting endoscopic remission in Crohn's disease by the modified multiplier SES-CD (MM-SES-CD). Gut 2022;71:1078-1087.

26. Rieder F, Bettenworth D, Ma C, et al. An expert consensus to standardise definitions, diagnosis and treatment targets for anti-fibrotic stricture therapies in Crohn's disease. Aliment Pharmacol Ther 2018;48:347-357.

27. Ma C, Khanna R, Guizzetti L, Zou G, Feagan BG, Jairath V. Methods for handling missing segments in Crohn's disease clinical trials: analysis from the EXTEND trial. Gut 2022;71:447-448.

## TABLES

### Table 1. Voting summary of SES-CD items

| Statement | Median (IQR) |
|---|---|
| Ulcers (0=none; 1=aphthous ulcers [diameter 0.1-0.5 cm]; 2=large ulcers [diameter 0.5-2 cm]; 3=very large ulcers [diameter >2 cm]) is an optimal descriptor to assess endoscopic disease activity | 7 (5, 9) |
| Ulcers at the anastomotic rim should be scored as ulcers, regardless of whether or not there is suspicion that these are ischemic | 7 (5, 8) |
| Anal ulcers, as well as all lesions that can be visualised on retroflexion, should be scored as part of the rectum | 8 (7, 9) |
| Ulcers are optimally measured with a greater number of categories to describe ulcer size (e.g., <0.5 cm, 0.5-1 cm, 1-1.5 cm, 1.5-2 cm, >2 cm) | 3 (2, 5) |
| Ulcers noted on side-by-side anastomoses should be allocated to the segment with the greatest disease burden | 7 (3, 8) |
| Ulcer severity should be scored on a 10 cm linear scale from 0 to 10, for future item development | 5.5 (3, 8) |
| Ulcer size should be scored on a 10 cm linear scale from 0 to 10, for future item development | 5 (3, 8) |
| Ulcers should be scored on a linear scale with anchors very small to very large, for future item development | 3 (3, 5) |
| Ulcerated surface (0=none, 1=<10%, 2=10%-30%, 3=>30%) is an optimal descriptor to assess endoscopic disease activity | 6 (5, 8) |
| Ulceration across a narrowing at the junction of two colonic segments, with minimal extension into either segment will be scored ONCE, on the side that appears most affected | 8 (7, 9) |
| Ulcerations on the colonic surface of the IC valve, not in the ileum itself, should be scored as right colon | 8 (7, 9) |
| If ulceration extends clearly over a long distance over the two segments, then ulceration can be scored for two segments | 8 (8, 9) |
| Scoring ulceration or affected surface beyond an impassable stricture should not be performed. In this case, impassable narrowing (SES-CD subscore 3) can be assigned to the distal segment except in the ileum/ileocolonic stricture where the narrowing score (3) would be applied to the ileum. The statement above is an ideal SES-CD optimization | 7 (7, 8) |
| Aphthous ulcerations should contribute to the ulcerated surface in the SES-CD | 8 (7, 8) |
| Percentage of ulcerated surface is optimally scored as a continuous variable | 5 (3, 6) |
| Ulcerated surface is optimally measured on a 100-mm VAS from 0=no ulceration to 100=whole surface ulcerated | 6 (3, 8) |
| Affected surface (0=unaffected segment, 1=<50%, 2=50%-75%, 3=>75%) is an optimal descriptor to assess endoscopic disease activity | 7 (5, 7) |
| Percentage of affected surface is optimally scored as a continuous variable | 3 (3, 6) |
| Affected surface is optimally measured on a 100 mm VAS from 0=no affected surface to 100=whole surface affected | 6 (3, 7) |

| Statement | Median (IQR) |
|---|---|
| Stenosis (0=none, 1=single, can be passed, 2=multiple, can be passed, 3=cannot be passed) is an optimal descriptor to assess endoscopic disease activity | 6.5 (6, 7) |
| "Narrowing" should be defined as a clear decrease in the diameter of the lumen, compared to the previous and subsequent bowel, which does not insufflate with air, and may or may not be associated with trauma when passing it. Multiple narrowings are clearly repeated areas of narrowed diameter with intervening open areas. "Stenosis" is defined as impassable narrowing without dilation, using an adult scope. The statement above is an ideal SES-CD optimization | 7 (7, 8) |
| Stenosis at the junction between two segments (such as hepatic flexure or splenic flexure) should be scored in only the distal segment | 8 (7, 8) |
| Anal stenosis should be scored as part of the rectum | 8 (6, 8) |
| The anus should be included with the rectum for the propose of scoring | 7 (5, 8) |
| Stenosis noted at side-by-side anastomoses should be allocated to the right colon | 6 (4, 8) |
| If dilation is required to pass a stricture, that patient would not be eligible for entry into the clinical trial. (Reasons include probability or re-stenosis that can occur after dilation, variable ability of local endoscopist to dilate strictures, differences in probability of dilation of primary versus anastomotic strictures). If dilation is required to pass stricture after the trial starts, then the area could be scored above the dilated stenosis if it can be intubated and adequately visualised. The statement above is an ideal SES-CD optimization | 7 (7, 8) |
| The optimal method to account for missing or not-visualised segments in the SES-CD is to: | |
| Impute the score for the worst visualised segment | 3 (2, 5) |
| Average the score of visualised segments | 5 (3, 6) |
| Weight the score by dividing by the number of visualised segments | 7 (5, 7) |
| Only score same segments seen on original scope | 7 (5, 8) |

Abbreviations: IC, ileocecal; IQR, interquartile range; SES-CD, Simple Endoscopic Score for Crohn Disease; VAS, visual analogue scale.

Green indicates statements voted as appropriate (without disagreement); yellow indicates statements voted as uncertain (without disagreement); orange indicates statements voted as any rating (with disagreement); and red indicates statements voted as inappropriate (without disagreement).

**Table 2. Voting summary of CDEIS items**

| Statement | Median (IQR) |
|---|---|
| Deep ulceration (present/absent) is an optimal descriptor to assess endoscopic disease activity | 7 (4, 7) |
| Deep ulceration can be identified by the presence of depth noted during the colonoscopy | 5 (2, 7) |
| Superficial ulceration is defined as the absence of depth during the colonoscopy | 7 (3, 7) |
| Superficial ulceration (present/absent) is an optimal descriptor to assess endoscopic disease activity | 6 (2, 7) |
| Superficial and deep ulceration should be determined by depth of lesion | 7 (6, 8) |
| Surface involved by the disease (measured in cm, for partially explored segments and for the ileum, the 10 cm linear scale represents the surface effectively explored), is an optimal descriptor to assess endoscopic disease activity | 6 (3, 8) |
| Ulcerated surface (measured in cm, for partially explored segments and for the ileum, the 10 cm linear scale represents the surface effectively explored), is an optimal descriptor to assess endoscopic disease activity | 6 (3, 7) |
| This is an ideal CDEIS optimization: stenosis should be defined as impassable narrowing | 8 (7, 8) |
| Stenosis at the junction between two segments (such as the hepatic flexure or splenic flexure) should be scored in only the distal segment. Strictures that are scored by digital examination would be excluded from the scoring | 7 (7, 8) |
| Stenosis at the junction between two segments (such as hepatic flexure or splenic flexure, should be scored in both segments | 2 (2, 3) |
| Ulceration across a narrowing at the junction of two colonic segments, including flexures with minimal extension into either segment should be scored ONCE on the side that appears most affected. Note: Ulceration on the colonic surface of the IC valve, not in the ileum itself, should be scored as right colon. Note: if ulceration extends clearly over a long distance over the two segments, then ulceration can be scored for two segments. Note: Any mucosal break in a stricture is an ulcerated stricture. The above statement is an ideal CDEIS optimization | 8 (7, 8) |
| Any mucosal break in a stricture is an ulcerated stricture | 7 (5, 8) |
| More than two mucosal breaks in a stricture in an ulcerated stricture | 3 (2, 6) |
| Any mucosal break, except aphthous ulceration is an ulcerated stricture | 5 (3, 7) |
| If dilation is required to pass a stricture, that patient should not be eligible for entry into the clinical trial. (Reasons include probability of re-stenosis that can occur after dilation, variable ability of local endoscopists to dilate strictures, differences in probability of dilation of primary versus anastomotic strictures). If dilation is required to pass stricture after the trial starts, then the area could be scored above the dilated stenosis if it can be intubated and adequately visualised. The above statement is an ideal CDEIS optimization | 7 (7, 8) |
| Ulcers at the anastomotic rim should be scored as ulcers, regardless of whether or not there is suspicion that these are ischemic | 7 (5, 8) |

| Statement | Median (IQR) |
|---|---|
| Ulcers noted on side-by side anastomoses should be allocated to the segment with the greatest disease burden | 7 (6, 8) |
| Aphthous ulcerations should be scored as superficial ulcers in CDEIS because we believe they cannot be neglected. (This is an issue for the score and merits a rule to include them as ulcers.) | 8 (7, 8) |
| Anal ulcers, as well as all lesions that can be visualised in case of retroflexion, should be scored as part of the rectum given that they can cause symptoms and may respond to therapy, and are relevant to disease | 8 (7, 8) |
| Anal stenosis should be scored as part of the rectum | 8 (7, 8) |
| Stenosis noted at side-by side anastomoses should be allocated to the right colon | 7 (5, 8) |
| The optimal method to account for missing or non-visualised segments in the CDEIS is to: | |
| Impute the score from the worst visualised segment | 3 (2, 3) |
| Average the score of visualised segments | 4 (2, 6) |
| Weight the score by dividing by the number of visualised segments | 7 (5, 7) |
| Only score same segments seen on original scope | 7 (5, 8) |

Abbreviations: CDEIS, Crohn's Disease Endoscopic Index of Severity; IC, ileocecal; IQR, interquartile range.

Green indicates statements voted as appropriate (without disagreement); yellow indicates statements voted as uncertain (without disagreement); orange indicates statements voted as any rating (with disagreement); and red indicates statements voted as inappropriate (without disagreement).

**Table 3. Voting summary of items for both endoscopic indices (SES-CD and CDEIS) and additional items**

| Statement | Median (IQR) |
|---|---|
| ***Both Indices (SES-CD and CDEIS)*** | |
| Ulceration on the top of the inflammatory polyps should not be counted as ulcers or ulcerated surface | 8 (7, 8) |
| If visible past a stenosis, but the endoscopist is not able to put the scope through, any bowel visualised beyond this point should contribute to the score | 5 (2, 7) |
| If visible past a stenosis, but the endoscopist is not able to put the scope through, any clearly observed ulceration should contribute to the score even if the scope was not put through the opening | 4.5 (2, 7) |
| If visible past a stenosis, but endoscopist is not able to put scope through, only clearly observed ileal ulceration should contribute to the score even if the scope was not put through the opening | 4.5 (2, 7) |
| A minimum of 10 cm of ileum must be observed to describe percentage of surface area involved | 6 (3, 8) |
| The central reader should be asked to confirm "is there an adequate amount of segment to score?" | 6 (4, 7) |
| Rectal lesions must be observed with retroflex | 3 (2, 7) |
| Rectal lesions should be allocated higher weighting due to the poor prognosis | 3 (2, 5) |
| Local endoscopist should be instructed to take a still photo of each ulcer to provide a sufficient pause in video recording | 3 (2, 7) |
| Local endoscopist should be instructed to pause at largest ulcers in each segment for at least one second | 7 (7, 8) |
| Biopsy forceps should be standardly used to assess size of a lesion | 5 (3, 6) |
| A wire measurement tool should be standardly used to assess size of lesion | 3 (1, 5) |
| Quality (prep, washing, etc.) should be graded for each procedure and a minimum percentage of the mucosa must be visualised for scoring to be performed (e.g., add minimum rule, such as BBPS above 6, or can see at least 95% of mucosa) | 7 (5, 8) |
| Unless both insertion and withdrawal phases of a given segment are visualised, stenosis cannot be scored | 6 (5, 7) |
| There should be a minimum number of segments visualised if more than 1 segment is missing/not visualised | 5 (3, 8) |
| Insufflation with air or water must occur to determine a stenosis cannot be passed | 7 (6, 8) |
| ***Additional Items*** | |
| Presence of ulcer (yes/no) should be explored as a possible descriptor to assess endoscopic disease activity | 7 (6, 8) |
| Ulcer severity measured on a 100-mm VAS should be explored as a possible descriptor to assess endoscopic disease activity | 6 (3, 7) |
| Ulcers are optimally measured as none/mild/moderate/severe | 3 (2, 5) |
| Ulcers are optimally scored in each segment as; no lesion, rare lesion, sparse lesion, diffuse lesion, or complete involvement | 3 (2, 4) |

| Statement | Median (IQR) |
|---|---|
| Affected surface should include: | |
| Scarring | 1 (1, 2) |
| Inflammatory polyps | 1 (1, 5) |
| Oedema | 5 (2, 7) |
| Erythema | 7 (5, 8) |
| Scoring the original CDEIS validation items: The following types of lesions should be reconsidered for inclusion in a new disease activity scale: | |
| Pseudopolyps | 2 (1, 3) |
| Healed ulcerations (whitish area with a 'ground glass' appearance) | 2 (1, 2) |
| Frank erythema (plaques, bands, or diffuse) (slight or moderate erythema should be neglected) | 6 (2, 7) |
| Assessing aphthous ulcers in the ileum is important for measuring endoscopic disease activity | 7 (7, 8) |
| Aphthous ulcers in the ileum should be quantified on a categorical scale (e.g., no aphthous ulcers, few aphthous ulcers, extensive aphthous ulcers, large superficial and deep ulcers, ulcerated narrowing) | 6 (3, 8) |
| Ulcers in the ileum should contribute a greater weight to an overall endoscopic disease activity score | 5 (2, 6) |
| Ileal lesion should be given greater weight than colonic lesions | 4 (2, 5) |
| Ulcers in the ileum should be assessed on a 100-point VAS scale from 0=no ulcers to 100=maximum number of ulcers | 3 (2, 7) |
| Ulcers in the ileum should be assessed on a 0-4 scale where 0=no ulcers and 4=maximum number of ulcers | 3 (2, 7) |
| Ulcers in the ileum should be assessed on an ordinal scale (e.g., <10, 10-30, >30) | 6 (5, 7) |
| Ulcers in the ileum should be assessed as categories mild, moderate, severe | 5 (3, 6) |
| Ulcers in the ileum should be scored with the Rutgeerts score | 3 (2, 6) |
| Other unique items should be assessed in the ileum | 2.5 (2, 4) |
| The presence or absence of nodular hyperplasia should be recorded | 2 (2, 3) |
| Rectal lesions should contribute a greater weight to an overall endoscopic disease activity score | 3 (2, 5) |
| Rectal lesion should be given greater weight than colonic lesions | 3 (2, 6) |
| Rectal and right colonic lesions should be given greater weight than colonic lesions | 3 (2, 3) |
| All other colonic segments and items assessed on a 100-point VAS should be included for exploratory analysis | 5 (2, 7) |
| All other colonic segments and items assessed on a 0-4 scale should be included for exploratory analysis | 5 (2, 6) |
| All other colonic segments and items assessed on an ordinal scale should be included for exploratory analysis | 5 (3, 5) |
| All other colonic segments and items assessed as categories mild, moderate, severe should be included for exploratory analysis | 5 (3, 5) |
| Other unique items should assess colonic disease | 2.5 (1, 5) |
| Lesions (ulcer, erythema, etc.) should be graded as present/absent | 3 (2, 6) |

| Statement | Median (IQR) |
|---|---|
| Lesions, (ulcers, erythema, etc.) should be graded on a scale of 0-3 | 3 (2, 5) |
| Affected surface area is not required, since it adds too much subjectivity to scoring | 6 (3, 8) |
| Segments should only be scored in a follow-up assessment if they were observed during the eligibility (baseline) procedure | 7 (2, 8) |
| Healed mucosa is optimally described as absence of ulceration | 7 (6, 8) |
| Healed mucosa is optimally described as absence of inflammatory polyps | 2 (2, 3) |
| Healed mucosa is optimally described as an absence of erythema | 3 (2, 4) |
| Healed mucosa is optimally described as an absence of friability | 3 (2, 6) |
| Healed mucosa is optimally described as the presence of granularity | 2 (2, 3) |
| Ulcers and/or erosion on (or immediately adjacent to) post-inflammatory polyps should be explicitly excluded | 6 (4, 8) |
| Any mucosal break with a yellow/white base of any depth is a small ulcer | 7 (6, 8) |
| The largest ulcer should bear the burden of the score | 5 (3, 7) |
| An assessment of visible absence or presence of disease activity after a stricture should be included for exploratory index development | 6 (4, 7) |
| A bowel preparation quality score should be routinely assessed. (If a preference or recommendation, please state in comments, e.g., Aronchick, BBPS, OBPQS, etc.) | 8 (6, 8) |
| A bowel preparation quality statement (such as optimal/suboptimal) should be further defined, such as "adequate to not miss 5 cm ulcer" or "adequate to not miss 1 cm ulcer" | 6 (2, 8) |

Abbreviations: BBPS, Boston Bowel Preparation Scale; CDEIS, Crohn's Disease Endoscopic Index of Severity; IQR, interquartile range; OBPQS, Ottawa Bowel Preparation Quality Scale; SES-CD, Simple Endoscopic Score for Crohn Disease; VAS, visual analogue scale.
Green indicates statements voted as appropriate (without disagreement); yellow indicates statements voted as uncertain (without disagreement); orange indicates statements voted as any rating (with disagreement); and red indicates statements voted as inappropriate (without disagreement).

**SUPPLEMENTAL MATERIAL**

**Supplemental Table 1. Scoring of the SES-CD items**

| Item[†] | SES-CD Scoring | | | |
|---|---|---|---|---|
| | **0** | **1** | **2** | **3** |
| Presence and size of ulcers | None | Aphthous ulcers (diameter, 0.1-0.5 cm) | Large ulcers (diameter, 0.5-2 cm) | Very large ulcers (diameter >2 cm) |
| Extent of ulcerated surface | None | <10% | 10%-30% | >30% |
| Extent of affected surface | Unaffected segment | <50% | 50%-75% | >75% |
| Presence and type of narrowing | None | Single, can be passed | Multiple, can be passed | Cannot be passed |

Abbreviation: SES-CD, Simple Endoscopic Score for Crohn's Disease.

[†] Each item is scored in the ileum, right colon, transverse colon, left colon, and rectum; the total score for each item is summed to produce the total SES-CD score.

**Supplemental Table 2. Scoring and calculation of the CDEIS**

| | Ileum | Right Colon | Transverse Colon | Sigmoid and Left Colon | Rectum | Total |
|---|---|---|---|---|---|---|
| **Section explored** | Yes/no | Yes/no | Yes/no | Yes/no | Yes/no | n |
| **Deep ulceration** | 0, absent; 12, present | 0, absent; 12, present | 0, absent; 12, present | 0, absent; 12, present | 0, absent; 12, present | Total 1 |
| **Superficial ulceration** | 0, absent; 6, present | 0, absent; 6, present | 0, absent; 6, present | 0, absent; 6, present | 0, absent; 6, present | Total 2 |
| **Surface involved by CD** | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | Total 3 |
| **Ulcerated surface** | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | VAS range, 0% (best) to 100% (worst) | Total 4 |
| | | | **Total 1 + Total 2 + Total 3 + Total 4:** | | | A |
| | | | **No. of segments totally or partially explored:** | | | n |
| | | | **A/n** | | | B |
| | | | **Ulcerated stenosis:** 0, absent; 3, present | | | C |
| | | | **Nonulcerated stenosis:** 0, absent; 3, present | | | D |
| | | | **Total CDEIS score:** | | | B + C + D |

Abbreviations: CD, Crohn's disease; CDEIS, Crohn's Disease Endoscopic Index of Severity; VAS, visual analogue scale.

**Supplemental Table 3. List of additional items identified through the RAM**

| Item | Scoring | | | | | |
|---|---|---|---|---|---|---|
| | **Ileum** | **Right Colon** | **Transverse Colon** | **Sigmoid and Left Colon** | **Rectum** | **Overall** |
| Severity of ulcers | None, mild, moderate, or severe | | | | | – |
| Severity of ulcers VAS | 0.0 to 10.0 cm | | | | | 0.0 to 10.0 cm |
| Presence of erythema (on insertion of endoscope) | Absent or present | | | | | – |
| Presence of lesions (ulcers, erythema) | Absent or present | | | | | – |
| Severity of lesions (ulcers, erythema) | 0 = no lesions, 1 = mild, 2 = moderate, or 3 = severe | | | | | – |
| Lesion severity VAS | 0.0 to 10.0 cm | | | | | 0.0 to 10.0 cm |
| Lesion type | None, rare, sparse, diffuse, or complete involvement | | | | | – |
| Aphthous ulcers | None, few, or extensive | – | – | – | – | – |
| Aphthous ulcers | None, 1-10, or >10 | – | – | – | – | – |
| Aphthous ulcers VAS | 0.0 to 10.0 cm | – | – | – | – | – |
| Presence of fistula | Absent or present | | | | | – |
| Presence of ischemic ulcers | Yes or no | | | | | – |
| Overall presence of ulcers | – | | | | | Absent or present |
| Overall ulcer severity VAS | – | | | | | 0.0 to 10.0 cm |
| Overall size of largest ulcer VAS | – | | | | | 0.0 to 10.0 cm |
| Presence of stenosis with visible disease activity after segment? | – | | | | | Absent or present |
| Presence of anastomosis | – | | | | | Absent or present |
| Presence of stenosis at anastomosis | – | | | | | Absent or present |

| Item | Scoring | | | | | |
|---|---|---|---|---|---|---|
| | **Ileum** | **Right Colon** | **Transverse Colon** | **Sigmoid and Left Colon** | **Rectum** | **Overall** |
| Can at least 10 cm of ileum be observed to describe percentage of surface area involved? | – | | | | | Yes or no |
| Presence of ulceration across a narrowing at the junction of two colonic segments? | – | | | | | Absent or present |

Abbreviations: RAM, Research and Development/University of California, Los Angeles appropriateness methodology; VAS, visual analogue scale.