CORE-IBD: A Multidisciplinary International Consensus to Develop a Core Outcome Set for Randomized Controlled Trials in Inflammatory Bowel Disease

The CORE-IBD Collaborators:

Christopher Ma[1], Jurij Hanzel[2], Remo Panaccione[3], William J. Sandborn[4], Geert R. D'Haens[5], Vineet Ahuja[6], Raja Atreya[7], Charles N. Bernstein[8], Peter Bossuyt[9], Brian Bressler[10], Robert V. Bryant[11], Benjamin L. Cohen[12], Jean-Frederic Colombel[13], Silvio Danese[14], Axel Dignass[15], Marla C. Dubinsky[16], Phillip Fleshner[17], Richard B. Gearry[18], Stephen B. Hanauer[19], Ailsa Hart[20], Paulo G. Kotze[21], Torsten Kucharzik[22], Peter L. Lakatos[23], Rupert W. Leong[24], Fernando Magro[25], Julian Panés [26], Laurent Peyrin-Biroulet[27], Zhihua Ran[28], Miguel Regueiro[29], Siddharth Singh[30], Antonino Spinelli[31], Hillary A. Steinhart[32], Simon P. Travis[33], Janneke Van der Woude[34], Bruce Yacyshyn[35], Takayuki Yamamoto[36], Matthieu Allez[37], Willem A. Bemelman[38], Amy Lightner[39], Edouard Louis[40], David T. Rubin[41], Ellen J. Scherl[42], Corey A. Siegel[43], Mark S. Silverberg[44], Severine Vermeire[45], Claire E. Parker[46], Stefanie C. McFarlane[46], Leonardo Guizzetti[46], Michelle I. Smith[46], Niels Vande Casteele[47], Brian G. Feagan[46,48,49], and Vipul Jairath[46,48,49]

Affiliations:

[1]Division of Gastroenterology and Hepatology, Departments of Medicine & Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; Alimentiv Inc. London, Ontario, Canada.

[2]Department of Gastroenterology, University Medical Center Ljubljana, Ljubljana, Slovenia; Alimentiv Inc. London, Ontario, Canada.

[3]Inflammatory Bowel Disease Unit, Division of Gastroenterology and Hepatology, University of Calgary, Calgary, Alberta, Canada.

[4]Division of Gastroenterology, University of California San Diego, La Jolla, CA, USA

[5]Department of Gastroenterology and Hepatology, Amsterdam University Medical Centers, Amsterdam, the Netherlands

[6]Department of Gastroenterology All India Institute of Medical Sciences New Delhi India.

[7]Medical Department I, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.

[8]Department of Internal Medicine, University of Manitoba, 820 Sherbrook Street, Winnipeg, Manitoba, R3A 1R9 Canada.

[9]Imelda GI clinical research center, Imelda general hospital, Bonheiden, Belgium.

[10]Department of Medicine, Division of Gastroenterology, St Paul's Hospital, University of British Columbia.

[11]IBD Service, Department of Gastroenterology and Hepatology, The Queen Elizabeth Hospital Adelaide, South Australia, Australia; School of Medicine, Faculty of Health Sciences University of Adelaide, Adelaide, South Australia, Australia.

[12]Cleveland Clinic Department of Gastroenterology, Hepatology, and Nutrition, Cleveland, Ohio.

[13]Division of Gastroenterology; Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[14]Department of Gastroenterology and Endoscopy, Università Vita Salute San Raffaele, Milano, Italy; Department of Gastroenterology and Endoscopy, San Raffaele Hospital, Milano, Italy.

[15]Department of Medicine I, Agaplesion Markus Hospital, Goethe-University, Frankfurt am Main, Germany.

[16]Department of Medicine, Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, New York.

[17]Division of Colorectal Surgery, Cedars-Sinai Medical Center, Los Angeles, California.

[18]Christchurch Hospital, Canterbury District Health Board, Canterbury, New Zealand; University of Otago, Christchurch, New Zealand.

[19]Department of Medicine (Gastroenterology and Hepatology), Northwestern University, Chicago, Illinois.

[20]Department of Gastroenterology, St Mark's Hospital and Academic Institute, Harrow, UK.

[21]Health Sciences Postgraduate Program Professor, IBD Outpatient Clinics, Colorectal Surgery Unit, Catholic University of Paraná (PUCPR), Curitiba, Brazil.

[22]Department of Gastroenterology, Lüneburg Hospital, University of Hamburg, Lüneburg, Germany.

[23]The University of Sydney, Sydney, NSW, Australia; Concord Repatriation General Hospital, Sydney, NSW, Australia.

[24]Department of Gastroenterology and Hepatology, IBD Services, Concord Repatriation General Hospital.

25 Department of Biomedicine, Unit of Pharmacology and Therapeutics, Faculty of Medicine, University of Porto, Porto, Portugal. Department of Clinical Pharmacology, São João University Hospital Center (CHUSJ), Porto, Portugal. Faculty of Medicine, University of Porto, Portugal. Center for Health Technology and Services Research (CINTESIS), Porto, Portugal.

26 Hospital Clinic Barcelona, IDIBAPS, CIBERehd, Barcelona, Spain.

27 Hépato-gastro-entérologie and Inserm U954, Lorraine University, Nancy, France.

28 Division of Gastroenterology and Hepatology, Key Laboratory of Gastroenterology and Hepatology, Ministry of Health, Shanghai Inflammatory Bowel Disease Research Center, Renji

Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai Institute of Digestive Disease, Shanghai, China.

29 Department of Gastroenterology, Hepatology, and Nutrition, Cleveland Clinic, Cleveland, USA.

30 Division of Gastroenterology, Department of Medicine, University of California, San Diego, La Jolla, California, USA.

31Department of Biomedical Sciences, Humanitas University, Pieve Emanuele - Milan, Italy; IRCCS Humanitas Research Hospital, Rozzano - Milan, Italy.

32 Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, Canada; Division of Gastroenterology, Department of Medicine, University of Toronto, Toronto, Canada.

33 Translational Gastroenterology Unit, John Radcliffe Hospital, Oxford, United Kingdom.

34 Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, The Netherlands.

35 Department of Internal Medicine, University of Cincinnati, Cincinnati, Ohio.

36 Department of Materials Science and Biotechnology, Graduate School of Science and Engineering, Ehime University, Matsuyama, Japan.

37 Department of Gastroenterology, Hôpital Saint-Louis, APHP, INSERM U1160, Université de Paris, Paris, Île-de-France, France.

38 Department of Surgery, Amsterdam UMC, Location AMC, Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam, The Netherlands.

39 Department of Colorectal Surgery, Cleveland Clinic, Cleveland, OH 44195, United States.

40 Hepato-Gastroenterology and Digestive Oncology Department, University and Centre Hospitalier Univestitaire (CHU) Liège, Liège, Belgium.

41 University of Chicago Medicine, Inflammatory Bowel Disease Center, Chicago, IL, USA.

42 Jill Roberts Center for IBD, Division of Gastroenterology and Hepatology, Weill Cornell Medicine.

43 Inflammatory Bowel Disease Center, Section of Gastroenterology and Hepatology, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA.

44 Mount Sinai Hospital Inflammatory Bowel Disease Centre, Toronto, ON, Canada.

45 Department of Gastroenterology and Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium.

46 Alimentiv Inc.

47 Department of Medicine, University of California San Diego, La Jolla, CA, USA

48 Departments of Medicine, Division of Gastroenterology, Western University, London, ON, Canada

49 Department of Epidemiology and Biostatistics, Western University; London, ON, Canada

Short Title:	A Core Outcome Set for IBD RCTs
Word Counts:	
Abstract Word Count:	260/260
Manuscript Word Count:	6241
Number of Tables:	2
Number of Figures:	3
Supplemental Files:	5
Keywords:	core outcome set; efficacy; endpoint; Crohn's disease; patient-reported; randomized; trials; ulcerative colitis; outcomes
Writing Support:	None

Author Contributions: CM, VJ conceptualized or designed the study; **CM, JH, CEP, SM, LG, MS** performed data acquisition, analysis, or interpretation; **CM, CEP, SM, BGF, VJ** drafted the manuscript; all authors revised the manuscript critically for important intellectual content. All authors approved the final version of the manuscript.

Conflicts of Interest:

CM: consulting fees from AbbVie, Alimentiv, Amgen, AVIR Pharma Inc, BioJAMP, Bristol Myers Squibb, Celltrion, Ferring, Fresenius Kabi, Janssen, McKesson, Mylan, Takeda, Pendopharm, Pfizer, Roche; speaker's fees from AbbVie, Amgen, AVIR Pharma Inc, Alimentiv, Ferring, Janssen, Takeda, and Pfizer; research support from Ferring, Pfizer.

JH:

RP:

WJS:

GDH:

VA:

RA:

CB:

PB:

BB:

RB:

BC:

JFC:

SD:

AD:

MD:

PF:

RG:

SH:

AH:

PK:

TK:

PL:

RL:

FM:

JP:

LPB:

ZR:

MR:

SS:

AS:

HS:

ST:

JVW:

BY:

TY:

MA:

WB:

AL:

EL:

DR:

ES:

CS:

MS:

SV:

CEP: is an employee of Alimentiv Inc.

SM: is an employee of Alimentiv Inc.

LG is an employee of Alimentiv Inc.

MS: is an employee of Alimentiv Inc.

NVC:

BGF:

VJ: has received has received consulting/advisory board fees from AbbVie, Alimentiv Inc., Arena pharmaceuticals, Asahi Kasei Pharma, Asieris, Astra Zeneca, 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 centralized imaging solutions that include endoscopy, histopathology, and magnetic resonance imaging. 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. CM, WJS, GD, NVC, BGF and VJ are consultants to Alimentiv Inc. and have a primary academic appointment; they do not hold equity positions or shares in Alimentiv Inc.

Correspondence:
Dr. Christopher Ma, MD, MPH
Assistant Professor
Division of Gastroenterology & Hepatology

Departments of Medicine and Community Health Sciences
Cumming School of Medicine, University of Calgary
6D61 Teaching Research Wellness Building
3280 Hospital Drive NW
Calgary, Alberta, Canada T2N 4Z6
Tel: 587-434-5268
Email: christopher.ma@ucalgary.ca

Dr. Vipul Jairath MB ChB DPhil
Professor of Medicine
Departments of Medicine and Epidemiology and Biostatistics
Western University
Suite 200, 100 Dundas Street
London, Ontario, Canada N6A 5B6
Tel: 519-685-8500
Fax: 519-663-3658
Email: vjairath@uwo.ca

Funding: This study was supported by a Canadian Institutes for Health Research Fellowship award

Acknowledgments: The authors would like to acknowledge Franck Carbonnel, Subrata Ghosh, Dermot McGovern, and Mamoru Watanabe for completing one round of the Delphi survey, and Saba Turk for providing technical support for this project.

Abstract

Background: Endpoints to determine the efficacy and safety of medical therapies for Crohn's disease (CD) and ulcerative colitis (UC) are evolving. Given the heterogeneity in current outcome measures, harmonizing endpoints in a core outcome set (COS) for randomized controlled trials (RCTs) is a priority for drug development in inflammatory bowel disease (IBD).

Methods: Candidate outcome domains and outcome measures were generated from systematic literature reviews and patient engagement surveys and interviews. An iterative Delphi process was conducted to establish consensus: panelists anonymously voted on items using a 9-point

Likert scale and feedback was incorporated between rounds to refine statements. Consensus meetings were held to ratify the outcome domains and core outcome measures. Stakeholders were recruited internationally, and included gastroenterologists, colorectal surgeons, and clinical trialists.

Results: A total of 235 patients and 53 experts participated. Patient-reported outcomes, quality of life, endoscopy, biomarkers, and safety were considered core domains; histopathology was included as an additional domain for UC. In CD, there was consensus to use the PRO2 (abdominal pain, stool frequency), Crohn's Disease Activity Index, Simple Endoscopic Score for CD, C-reactive protein (CRP), fecal calprotectin, and a co-primary endpoint of symptomatic remission and endoscopic response. In UC, there was consensus to use the adapted 9-point Mayo Clinic Score, fecal urgency, Robarts Histopathology Index, fecal calprotectin, and a composite primary endpoint including both symptomatic and endoscopic remission. Safety outcomes should be reported using the Medical Dictionary for Regulatory Activities.

Conclusions: This multidisciplinary collaboration involving patients and clinical experts has produced the first COS that can be applied to RCTs of CD and UC.

Introduction

The inflammatory bowel diseases (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), are chronic, progressive, and disabling disorders of the gastrointestinal tract.[1, 2] The current paradigm for medical therapy focuses on controlling inflammatory activity. While the armamentarium of treatment options for IBD has substantially expanded over the past several decades with the adoption of biologic and novel small molecule therapies, these developments have not resulted in transformational efficacy. Clinical remission rates remain low, and most patients do not achieve complete endoscopic, radiographic, histologic, or biomarker-based definitions of remission with currently available agents.[3] This therapeutic gap underscores the importance of continued efforts to improve drug development in CD and UC, as better understanding of biological pathways of disease provide both new targets and personalized medicine strategies.

The approval of new therapies relies on data from robust randomized controlled trials (RCTs), which have increased in both size and complexity over time.[4] Advances in our understanding of CD and UC have also resulted in the evolution of study designs for these indications. For example, there is increasing recognition that symptom-based measurements alone are insensitive and poorly specific for assessing disease activity, resulting in a paradigm shift towards normalizing objective measures of inflammation.[5-7] Trials now routinely incorporate endoscopic evaluation for qualifying patients at enrolment and for assessing efficacy. Beyond endoscopic mucosal appearance, the added value of targeting aspirational goals such as histologic remission in UC or transmural healing in CD is under evaluation.[8, 9] At the same time, there has been an emphasis at the regulatory level to more accurately capture the patient experience using validated patient-reported outcomes (PROs).[10, 11] Given the evolving landscape of treatment endpoints and the rapid development of novel therapies, standardizing what, how, and when to measure key efficacy and safety outcomes in IBD trials is a research priority.[12]

Historically, insufficient attention has been paid to the standardized assessment of outcome measures in IBD trials.[13, 14] For example, multiple versions of the Mayo Clinic Score (MCS) are currently used in phase 2 and 3 trials with no universally accepted convention having been defined. The development of a core outcome set (COS) for use in IBD RCTs will increase the relevance of clinical research for multiple stakeholder groups, reduce heterogeneity in outcome reporting, and enhance the quality of evidence synthesis.[15] A COS is a consensus-derived *minimum* set of outcomes that should be measured and reported in all trials of a given disease.[16] While these core outcomes should always be assessed, it should also be recognized that a COS is not restrictive and that investigators are encouraged to explore other endpoints of interest, in addition to those included in the COS.

The CORE-IBD consensus was a multiple phase program, that included patients, gastroenterologists, colorectal surgeons, methodologists, and clinical trialists, which aimed to develop the first international consensus-based COS for use in CD and UC RCTs.

Methods

Registration and Scope

The CORE-IBD initiative is registered with Core Outcome Measures in Effectiveness Trials (COMET)[17] and was conducted in accordance with recommendations outlined in the COMET handbook and the Core Outcome Set-STAndards for Development (COS-STAD).[16, 18, 19] The study protocol was previously published.[15] All patient-related activities were approved by the ethics committee at the University of Calgary (REB20-1827).

The scope of this COS is for use in RCTs of pharmacologic therapies for adult patients (≥18 years) with CD or UC. We primarily considered luminal CD and excluded trials of patients with specific phenotypes such as pouchitis or perianal fistulizing CD. Recognizing that RCTs will typically be conducted in different settings, with different available resources and operational support, and choosing different outcomes compared to 'real-world' registries or non-randomized, prospective cohorts, we focused on outcomes relevant for RCTs and acknowledge that the measures recommended may not be feasible due to cost or operational considerations. Interventions involving surgical treatment were outside the scope of this COS. Furthermore, this COS may not apply to pediatric patients, where outcomes unique to this population, such as growth failure, are measured.

Overview of COS Development

A multiple phase approach was used to develop the CORE-IBD consensus (**Figure 1**). First, candidate outcomes that have been previously measured in IBD RCTs were identified in a series of comprehensive systematic literature reviews and organized into outcome domains. Second, patient engagement surveys and qualitative interviews were conducted to prioritize outcome domains of importance for patients and to identify any additional endpoints that may not have been captured in existing studies. Third, a comprehensive list of outcome measures developed in phases 1 and 2 was evaluated by an international panel of multidisciplinary experts

in a two-round Delphi survey. Finally, virtual ratification meetings were held to vote on the final outcomes included in the COS.

Phase 1: Outcome Identification through Systematic Review

Outcomes relevant to the scope of the COS were identified in previously published systematic literature reviews,[13, 14] and searches were updated to 2021. In summary, MEDLINE, EMBASE, and the Cochrane CENTRAL Library were searched without language restrictions to identify placebo controlled RCTs evaluating medical interventions (including corticosteroids, immunosuppressants, mesalamine compounds, biologics, and novel small molecules) for the treatment of adult patients with CD or UC. Trials of postoperative CD recurrence, complementary therapies, devices, surgical interventions, and hospitalized patients were excluded. A comprehensive inventory and definitions of trial endpoints were extracted. Individual endpoints were categorized into outcome domains, including patient-reported, endoscopic, histologic, radiographic, biomarker, and composite outcomes of efficacy. Results from these systematic reviews are not reported in this manuscript.

Phase 2: Patient Engagement Surveys and Interviews

Patients were engaged in the COS development through both online surveys and semi-structured interviews. An anonymous survey consisting of semi-structured and single selection multiple choice responses was used to assess patient perceptions and preferences of different outcome domains relevant to IBD care. This survey was distributed using an IBD email list-server and available online, to capture a broad range of responses from a diverse population of adult patients with IBD, including patients of different ages, disease durations, disease activity statuses (e.g., symptomatic patients and those in remission), and treatment experiences (e.g., patients previously exposed to corticosteroids, immunosuppressants, and biologics).

The online survey was developed with input from patient advocates, clinicians, and IBD nurse specialists, and the language and format of the survey was piloted before being distributed (**Supplemental Appendix 1**). The survey consisted of 15 questions, capturing demographic characteristics, IBD phenotype, current and past treatment history, patient assessment of remission status, patient experience with different methods of disease assessment, and perceived importance of different outcome domains of treatment efficacy. Open-ended free-text responses aimed at identifying other measures of treatment efficacy, beyond those identified in the systematic review, were also included.

Second, semi-structured interviews were conducted with patients who had previously participated in an RCT. For feasibility, patients were purposively sampled from the IBD Trials Unit at the University of Calgary, Canada, and patients were enrolled until thematic saturation was achieved. All interviews were conducted in English, using a topic guide to identify the patient's lived experiences with IBD, benefits and harms of IBD-related treatment, specific experience in the RCT, and outcomes they believed to be relevant and important to include in IBD trials. Narrative data were indexed and mapped to a thematic framework to summarize key points and outcomes of priority.

Phase 3: Delphi Panel

A comprehensive list of outcomes identified in phases 1 and 2 were incorporated in a two-round Delphi survey in phase 3. For each Delphi survey, a minimum sample size of 30 respondents was targeted. A diverse pool of gastroenterologists, colorectal surgeons, methodologists, and clinical trialists, who brought a broad range of clinical knowledge, RCT-related experience, and geographical diversity was identified and invited to participate by the lead (CM) and senior investigator (VJ). Minimum requirements for participation included expertise in IBD trial conduct or outcome assessment, as reflected by metrics such as authorship of at least 25 publications related to IBD, involvement in at least two IBD clinical trials (either as an

investigator or through input into the trial design), or clinical expertise as demonstrated by being the medical or surgical leads of a dedicated IBD center.

The Delphi method used, which has been endorsed for COS development, allows panelists to anonymously derive consensus through multiple rounds of sequential questionnaires.[16] After each of the two electronic voting rounds, the group responses and each panelists' individual responses were provided in a feedback summary.[20] The survey was anonymously conducted to avoid any single personality from dominating the consensus and to facilitate broader international participation.. All electronic questionnaires were pilot tested for clarity before distribution.

Participants were asked to rate each outcome on a scale from 1 to 9, based upon the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group definitions.[21] Panelists were specifically instructed to rate the most important outcomes highly (7-9 range), and downgrade outcomes of lesser importance (scores of 1-3 indicate an outcome that is not important for inclusion). Scores of 4-6 were used to indicate an outcome of some importance that is not critical for inclusion. All outcomes had free-text entry options for participants to provide clarifying statements or identify additional outcomes of interest. Responses were collated, and descriptive statistics were used to summarize the scoring and distribution of the entire panel. Individualized feedback reports were provided to panelists with the group score, individual panelist score, and any additional comments from the panel.

Based upon panelist feedback regarding the length of the survey, only outcomes for which ≥50% of panelists voted in the 7-9 range were carried forward from the first round to the second round of voting. The panelists were asked to rescore these outcomes on a 1-9 Likert scale with insights from the group taken into consideration. Outcomes for which ≥70% of panelists scored in the 7-9 range and <15% of panelists scored in the 1-3 range during the second round of voting were decided *a priori* to have met consensus for inclusion.

Phase 4: Ratification Meetings

Consensus ratification meetings were held by videoconference on November 3 and November 6, 2021. The criteria for COS inclusion were reviewed for each. All items from Round 2 were reviewed, with a focus on potentially contentious items rated in the 7-9 range by 60%-80% of panelists. Items under consideration for inclusion were discussed, and arguments for and against inclusion were synthesized in a summary document. After discussion, panelists re-voted on these candidate items in an anonymous third round, where voting was simplified to “Include in the COS”, “Do not include in the COS”, or “Unsure”. Similar to phase 3, outcomes receiving ≥70% of votes in the ‘Include’ category and <15% of votes in the ‘Do not include’ category were ratified for final inclusion in the COS.

Results

Participants

Demographics of the patient participants and expert panelists are summarized in **Supplemental Table 1**. A total of 235 patients (52% female) with IBD completed the online engagement survey (response rate 47.0%). Almost all patients (226/235, 96.1%) had received at least one form of medical therapy for IBD. A total of 37 patients (15.7%) had previously participated in an RCT. Almost all patients reported having previously undergone disease assessment by colonoscopy (222/235, 94.5%), with half of the patients having undergone a colonoscopy within the past year (114/235, 48.5%). Two thirds of patients routinely underwent either stool-based testing (154/235, 65.6%) or imaging-based monitoring (164/235, 69.8%). A total of 22 IBD patients (50% female) who had previously participated in an IBD RCT underwent qualitative interviews. The median age of prior trial participants was 54.5 years (interquartile range 40.75, 60.75 years) and half of the participants had UC (11/22, 50.0%).

Of the 76 experts invited, 53 (69.7%) experts from 17 countries representing North and South America, Europe, and Asia-Pacific participated in the Delphi panel. Most panelists practiced in academic hospitals (51/53, 96.21%), typically in mixed roles as clinician-researchers (46/53, 86.8%), and had more than 20 years of experience in IBD care (38/53, 71.7%).

Patient Engagement Surveys and Semi-Structured Interviews

Patients rated the relative importance of six different outcome domains (symptoms, quality of life, endoscopy, histology, biomarkers, and imaging), both in the context of a short term (within three months of starting treatment) and long-term (one year after starting treatment) duration, on a 9-point Likert scale. Consistently, patients rated improvement in IBD-related symptoms and quality of life as the most important measures of efficacy (**Figure 2**). Improvements in abdominal pain and stool frequency were commonly reported as the most important aspects of symptomatic improvement, as were associated symptoms such as control over bowel movements, urgency,

continence, and stool form. Improvements in quality of life were captured in both the online survey and semi-structured interviews, characterized most notably by general well-being and a holistic sense of “feeling well”, and the ability to function in usual everyday activities while living as “normal” a life as possible. Achievement of endoscopic improvement in the long-term was voted as very important by 70.6% (166/235) of patients, yet approximately 10% of patients rated achievement of endoscopic, histologic, biomarker, or imaging outcomes as not important (ratings of 1-3 on a 9-point Likert scale). In semi-structured interviews, almost all participants were satisfied with their clinical trial experience although the burden of frequent visits and long, complex surveys was highlighted. Most participants expressed that they understood why both subjective and objective measures of disease were being measured in trials. Symptoms that were identified as potentially important in both interviews and from the qualitative responses included depression, anxiety, fatigue, sleep quality, appetite, nausea, and the ability to eat a normal diet. All these items were included in the expert Delphi survey.

Delphi Survey Results and Outcome Domains

A total of 475 statements were included in the first-round survey, and 228 of these statements were voted as important for inclusion in a COS by >50% of the panel and included in round 2. In the ratification vote, 111 candidate statements rated in the 7-9 range by 60%-80% of panelists during round 2 were discussed and re-voted on. A summary of voting results from the final round of the Delphi survey and from the ratification meetings are presented in **Supplemental Tables 2** and **3**.

In both CD and UC, there was consensus that core domains of treatment efficacy should include PROs, quality of life, endoscopy, and biomarkers (**Figure 3**). In UC, there was also consensus that histopathology should be a core outcome domain. Radiographic and histopathology outcomes in CD and health care utilization overall were not considered core domains for measurement in clinical trials.

Core Outcomes for Crohn's Disease Trials

Patient Reported Outcomes and Composite Indices

Core outcomes for CD RCTs are summarized in **Table 1**. There was consensus that symptomatic outcomes in CD RCTs should be defined using a PRO measure that incorporates stool frequency and abdominal pain, with symptoms assessed for at least 7 days before a study visit (excluding the day before, of, and after a colonoscopy) to capture potential symptom variability. Although other symptoms such as liquid stool frequency, nocturnal bowel movements, urgency, cramping, nausea, vomiting, extraintestinal manifestations, general well being, depression, anxiety, fatigue, sleep quality, sexual dysfunction, and need for dietary modifications were considered, the panel discussed that the correlation between these symptoms and objective disease activity may be poor and some of these symptoms may be challenging to define or quantify.[22] It was acknowledged that symptoms such as fatigue and mental wellness are important to patients and will be assessed in some RCTs, although they do not constitute core outcomes that must be measured in every trial. The panel also recognized that a fully validated PRO is not yet available although several are in development[23-25] and many different PRO or symptom-based indices were considered. In the interim, only the PRO2 incorporating abdominal pain and stool frequency was voted as a core outcome.[26, 27] A definition of remission based on the PRO2 with a mean abdominal pain score ≤1 and stool frequency subscore ≤2.8 was discussed at length. This definition identified a similar proportion of clinical remitters when compared to the Crohn's Disease Activity Index (CDAI) in phase 3 trials of risankizumab.[28] However, a consensus threshold was not reached for this definition as it was felt that the operating properties of PRO2 are still being evaluated in continuing RCTs.

The CDAI[29] was included as a core composite outcome measure with response defined by a reduction of >100 points compared to baseline. Limitations of the CDAI, including its complexity of calculation, applicability to routine clinical care, weighting towards stool frequency,

and poor correlation with endoscopy were discussed by the panel.[30] However, given that the CDAI has been used in all registrational CD trials for modern therapies, it was considered that this instrument should continue to be measured as a core outcome while validated PROs are developed. Additionally, the panel discussed that the CDAI has been used to define relatively homogeneous disease populations for trial enrolment. A consensus definition of CDAI-based remission was not agreed upon. Although a CDAI score of <150 has historically been used to define clinical remission, some panelists felt that this definition is poorly specific and could be improved with the addition of a >70 to 100-point reduction from baseline. However, other panelists contended that such a caveat would reduce the sensitivity of the remission threshold, and that a reduction would already be captured based on the inclusion requirements of most moderate-to-severe CD trials, which use a CDAI of 220-450 as an enrolment criterion.

Endoscopic Outcomes

The Simple Endoscopic Score for CD (SES-CD) was endorsed over the Crohn's Disease Endoscopic Index of Severity (CDEIS) given the simplicity of calculation.[31, 32] There was consensus that endoscopic remission should be defined by absence of any ulcerations in all segments. Defining endoscopic remission using SES-CD ≤2, which would capture absence of ulcerations, was considered an aspirational target. There was consensus to use this threshold for isolated ileal CD. The panel voted that endoscopic response should be defined by a >50% reduction in SES-CD compared to baseline. Both endoscopic response and endoscopic remission should be measured in all induction and maintenance trials. There was consensus that endoscopic outcomes should be reported at 52 weeks for maintenance trials, although no consensus on the timing of endoscopy during induction was reached. Assessing endoscopy too early in induction trials risks missing delayed improvement, although panelists highlighted that this must be balanced against a prolonged induction period leading to higher rates of patient dropout and potential prolonged exposure to fixed corticosteroid dosing regimens.

Biomarker Outcomes

There was consensus that both C-reactive protein (CRP) and fecal calprotectin should be measured and reported in CD induction and maintenance trials.[33] Although other biomarkers such as the Endoscopic Healing Index were considered, limitations with respect to availability, cost, or accuracy precluded their inclusion as a core outcome for measurement in all trials.[34] Biomarker-based remission in CD trials should include a CRP <5 mg/L. There was no consensus on the threshold definition for fecal calprotectin-based remission although multiple cut-offs ranging from 50 to 250 µg/g were considered. In panel discussions, potential variability in fecal calprotectin concentrations with disease location and extent were contributing factors to the lack of consensus.[35, 36] However, a fecal calprotectin reduction >50% among those with an elevated calprotectin at baseline was considered a core outcome. There was no consensus on timing for measurement of biomarker endpoints in either induction or maintenance. Although biomarkers were considered an informative surrogate measure of pharmacodynamic effects, the panel discussed that biomarker assessment should be balanced against the potential patient burden for collection, which may be increasingly relevant for virtual trial visits. There was support for measuring biomarkers at all visits during induction and every three months during maintenance, although the threshold for consensus was not reached.[6]

Configuration of Outcomes in CD Trials

There was consensus that remission in CD trials should be defined using coprimary endpoints of symptomatic remission and endoscopic response, which captures both the patient experience and objective assessment of disease activity. There was considerable discussion and ultimately, consensus was reached that corticosteroid-free remission should also be reported in induction trials. Although corticosteroid dosing is typically fixed during the induction period of a trial, there was discussion about the potential to design trials that allow for early corticosteroid tapering, which would increase the sensitivity of detecting treatment efficacy and would favor

highly efficacious agents.[37] Endoscopic remission was considered but felt to be too stringent an endpoint to be used as the primary outcome.

In maintenance trials, there was consensus to measure coprimary symptomatic and endoscopic response, remission, corticosteroid-free remission, endoscopic response and remission, biomarker-defined remission, and sustained remission (defined by remission at enrolment in the maintenance trial and at every study visit thereafter). There was consensus that worsening symptoms, as assessed by PROs in combination with either worsening endoscopy or biomarkers are required for defining loss of response. Need for rescue corticosteroids or surgery were not considered adequate definitions of loss of response because of the substantial heterogeneity in how clinical decision-making influences these endpoints.

The panel discussed that the timing of measuring outcomes in induction and maintenance CD trials will vary depending on the mechanism of action. Generally, 9-12 weeks was felt to be an appropriate duration for induction studies and there was consensus that 52 weeks was an appropriate time point to measure maintenance outcomes for both re-randomization and treat-through study designs.

Core Outcomes in Ulcerative Colitis Trials

Patient Reported Outcomes and Composite Indices

Core outcomes for UC RCTs are summarized in **Table 1**. There was consensus that symptomatic remission and response in UC trials should be defined using a PRO encompassing rectal bleeding and stool frequency as the hallmark symptoms. Additionally, there was agreement that fecal urgency should be captured as a core outcome given patient input on the debilitating nature of this symptom.[38] Other symptoms were considered, however, abdominal pain, cramping, nausea, vomiting, loss of appetite, stool consistency, and tenesmus or sensation of incomplete evacuation were not voted as critical for inclusion. Nocturnal bowel movements were discussed as a potential marker of pathology, yet the threshold for inclusion was not met.

Experts agreed that the adapted 9-point MCS, comprising rectal bleeding, stool frequency, and endoscopic appearance, should be the core composite measure of efficacy, as opposed to the full 12-point MCS, which includes physician global assessment (PGA).[39, 40] This was driven by concerns regarding the reliability and reproducibility of the PGA and is not congruent with the notion of a patient reported outcome measure. However, it was discussed that the full MCS has been the benchmark for drug development in UC over the past 30 years, and that the adapted 9-point MCS could be calculated *post hoc* from the full MCS parameters. A consensus definition of response or remission based on the adapted MCS was not reached, although experts agreed that symptomatic remission should be defined by a rectal bleeding subscore of 0 and stool frequency subscore of ≤1. The panel recognized that several reasons may contribute to patients not achieving normalization of stool frequency even while in endoscopic remission, such as decreased rectal compliance or overlapping functional bowel disorders associated with diarrhea.[41] A definition of remission based upon an adapted MCS≤2 with stool frequency subscore of 0 or 1 (and no greater than baseline), rectal bleeding subscore of 0, and modified Mayo endoscopic subscore (mMES) of 0 or 1 was the closest definition to reaching consensus.

Endoscopic Outcomes

There was consensus that sigmoidoscopy should be used to assess endoscopic outcomes in UC RCTs, with scoring based on the worst affected segment. The panel discussed that in some instances, colonoscopy may be required at enrolment to exclude dysplasia in patients with long-standing disease. However, several arguments supported the use of sigmoidoscopy for outcome measurement: 1) current endoscopic scores were developed based on sigmoidoscopic examination; 2) there are substantive practical advantages with respect to need for bowel preparation, time, and cost of sigmoidoscopy, especially in trials that require multiple endoscopies; and 3) most patients have the most severe disease activity in the distal rectosigmoid, which is representative of the more proximal colon.[42]

Both the mMES and UC Endoscopic Index of Severity (UCEIS) were considered.[43, 44] The mMES excludes 'mild friability' and scores any friability as mMES=2.[45] This modification is used almost exclusively in contemporary UC RCTs, although the panel recognized that the dynamic range of the score (0 to 3) is narrow and differences between mMES=0 and 1 are subtle. The score may be overly restrictive for outcome assessment if ulcerations have improved but not completely healed and all other components as well as healing extent improve (still scored as mMES=3 in this scenario). The UCEIS has advantages of scoring individual component items (vascularity, bleeding, and erosions and ulcerations), with a broader range of scores for assessment of responsiveness after therapy. However, some features such as friability are not captured, and practitioners may be less familiar with this score in day-to-day practice.

There was consensus that endoscopic remission and response should be measured and reported in both induction and maintenance UC RCTs, with endoscopy at 9-12 weeks in induction trials and 52 weeks in maintenance studies. The panel voted that endoscopic remission should be defined by an mMES=0 and that response should be defined by an mMES reduction ≥1 compared to baseline. The previous definition of endoscopic remission (mMES=0 or 1) is now termed endoscopic improvement in trials of moderate-to-severely active UC that require a baseline mMES≥2 for enrolment.[46] The panel voted to use the term 'endoscopic response' based on a reduction of ≥1 point in the mMES, which could also be used in trials of mild-to-moderately active UC where endoscopic enrolment requirements vary.

Histopathology Outcomes

Histopathology was considered a core outcome domain in UC induction and maintenance RCTs. There was consensus to use the Robarts Histopathology Index (RHI) as a validated score in the trial setting.[47] The ordinal Nancy Index[48] was considered a highly practical and validated index for use in routine clinical practice; however, the wider range and continuous nature of the RHI were discussed as advantages for demonstrating responsiveness where sample sizes are

relatively small, such dose finding studies. The absence of neutrophilic inflammation was an important determinant of histologic remission: both an RHI<3 without neutrophils and a Geboes score <3.0 without neutrophilic inflammation in the epithelium achieved consensus for defining histologic remission. There was no consensus on measuring histologic response, as the panel felt that the change in histology scores from baseline constituting a meaningful response have not yet been defined, and baseline histologic activity is not used as an entry criterion for most RCTs.[49, 50]

Biomarker Outcomes

Fecal calprotectin was recognized as the most important biomarker for assessing inflammatory activity. There was discussion that the operating properties of fecal calprotectin depend on the threshold chosen, and there was consensus that biomarker-defined remission should be based on a fecal calprotectin <150 µg/g. Biomarker response was similarly defined as in CD by a reduction of >50% compared to baseline among those with an elevated fecal calprotectin at baseline. There was consensus that fecal calprotectin should be measured at 9-12 weeks in induction trials, and then measured again at 24- and 52-weeks during maintenance therapy.

Configuration of Outcomes

Symptomatic and endoscopic remission were voted to be the core components of a composite primary endpoint in UC induction trials. Additional outcomes that were voted as core induction endpoints included the composite outcome of clinical and endoscopic response, clinical remission/response, endoscopic remission/response, and histologic remission. Corticosteroid-free remission was not voted as a core outcome in UC induction trials, although it was considered important during maintenance, as was sustained remission (defined using a similar definition as in CD). Panel members expressed the importance of clear instructions for corticosteroid dosing

during both induction and maintenance given the highly sensitive nature of UC symptoms to corticosteroids. There were lengthy conversations about the most appropriate definition of "corticosteroid-free remission"; no individual definition met the threshold for consensus, although the definition preferred by most panelists was withdrawal of systemic corticosteroids for at least 12 consecutive weeks before the final study visit during maintenance. Mucosal healing, defined by both endoscopic and histologic remission, was considered a core outcome in maintenance studies. There was also support for reporting loss of response in UC maintenance trials, defined by worsening symptoms and either worsening endoscopy or biomarkers. Although a relatively infrequent occurrence, the need for colectomy was voted as a core endpoint in UC trials.

The panel discussed that therapies for UC should achieve early symptomatic response, although 4-8 weeks may be too early depending on the mechanism of action. There was consensus to measure induction endpoints at 9-12 weeks. For maintenance trials, a week 52 outcome was considered appropriate for both re-randomization and treat-through designs.

Safety Outcomes

There was consensus that safety outcomes should be reported in all IBD trials, including all serious adverse events (SAEs) and AEs occurring in >5% of the trial population. The panel voted that a common terminology should be used to describe AEs in all IBD RCTs, and there was consensus to use the MedDRA, which is a clinically validated international terminology supported by regulatory agencies.[51]

Discussion

This international, multidisciplinary collaborative effort has developed the first consensus COS for standardizing outcome reporting in RCTs of pharmacologic therapies for the treatment of IBD. Through multiple rounds of voting and engagement with patients, methodologists, and disease experts, several core outcome domains were selected, including PROs, quality of life, endoscopy, biomarkers, safety, and histopathology for UC. Additionally, we identified appropriate measurement tools for standardizing disease activity assessment, definitions for individual core outcomes, and trial configurations best suited for induction and maintenance studies. Selecting appropriate endpoints for use in RCTs is critical because their operating properties are key determinants of trial efficacy and safety, ultimately driving our ability to efficiently identify new agents in a time when RCT recruitment is increasingly challenging. The choice of outcomes used in pivotal trials also shapes clinical practice since they are considered by payers when determining relative cost-effectiveness, which consequently influences health policy decisions.[52] This first iteration of the CORE-IBD COS will improve the quality of research in IBD, minimize reporting bias, standardize endpoint definitions, and serve as the impetus for additional research to address unanswered questions for the field.

In assessing outcomes used in CD and UC RCTs, it was evident that there has been a tremendous evolution in endpoints over time, driven by several factors. First, there has been increasing recognition that symptoms alone are neither sufficiently sensitive nor specific for assessing mucosal inflammation, resulting in a greater focus on achieving objective measures of remission. In clinical care, this has been practically applied using treat-to-target approaches that emphasize endoscopic and biomarker targets in addition to symptomatic response. Our work in developing this COS is complementary to recent guidelines from the Selecting Therapeutic Targets in IBD (STRIDE)-II group, as many of the core outcome domains overlap with short- (symptomatic), intermediate- (biomarker), and long-term (endoscopy) targets in clinical care.[53] Second, it should be recognized that drug development has been highly influenced by regulatory

guidance, which has also evolved over time. Some examples include: 1) the increasing emphasis placed by regulators on capturing how patients feel, function, and survive using PROs; 2) the exclusion of the PGA in UC assessment, resulting in adoption of the 9-point adapted MCS; and 3) changes in what constitutes mucosal healing, which now incorporates both endoscopic and histologic remission in UC.[54] All these concepts have been captured in our COS, although we emphasized to panelists that outcome measures felt to be important for assessment did not have to map precisely onto current regulatory recommendations.

Given the changes in outcome measures over time, it is unsurprising that consensus definitions could not be reached on several endpoints, highlighting the uncertainty that exists even among experts. Notably, a consensus threshold for defining clinical remission using PRO2 in CD and clinical remission or response using the adapted MCS in UC was not achieved, with panel discussions focusing on the need for data from continuing clinical trials to fully characterize the operating properties of these measures. Indeed, the PRO2 is only an interim tool, as no fully validated disease-specific PROs developed according to regulatory standards have been used as the primary endpoint in an IBD RCT to-date. Disease-specific PROs such as the CD and UC-PRO Signs and Symptoms diary and the Symptoms and Impacts Questionnaire for CD and UC were both considered for inclusion, however, further validation work is required before these tools can be included in a COS.[23-25] In contrast, the PRO2 and the adapted 9-point MCS have both been successfully used in recent phase 3 programs, with similar effect sizes demonstrated compared to the CDAI and full 12-point MCS.[55] An important limitation of the PRO2 and adapted MCS is their inability to capture potentially important disease-related symptoms beyond stool frequency, abdominal pain, or rectal bleeding, such as depression, dietary changes, fatigue, sleep disturbance, incontinence, limitations in functioning, and overall well-being. These were emphasized as important by patients, and additional research is required to ensure that these outcomes can be measured in a reliable, meaningful, and interpretable way.

Several potentially novel innovations in study design for IBD RCTs were identified in the development this COS. First, the panel discussed at length the appropriate handling of corticosteroids, particularly during induction. Historically, corticosteroid dosing was fixed during induction to minimize confounding the interpretation of therapeutic effects. However, the emergence of endoscopy as a coprimary or composite primary outcome, the potential risk of overlooking mucosal improvement on early endoscopy, and the desire to mitigate potential corticosteroid-related AEs have contributed to a growing appetite for earlier corticosteroid tapering, even during induction. The panel stressed that clear tapering rules would need to be applied and assessing corticosteroid-free remission in induction would be more sensitive for identifying highly effective agents. Second, although rectal bleeding and stool frequency have been the hallmark symptoms associated with UC, fecal urgency was added as a core outcome for RCTs. Urgency was consistently identified in patient surveys as a debilitating symptom, which is associated with incontinence, social impairment, anxiety, depression, and reduced quality of life. Additionally, a recent study in the IBD Partners research network identified that urgency was associated with increased risk of hospitalization, corticosteroid use, and colectomy.[56] Measuring urgency in addition to rectal bleeding and stool frequency may better capture the patient's disease experience in UC RCTs.

The development of this COS underscored several important areas of research priority. For example, radiographic endpoints were notably not included as a core outcome domain in CD, despite computed tomography, magnetic resonance imaging, and ultrasound being important tools in clinical practice for evaluating disease activity.[57] Delineating the role of transmural healing, which could be associated with better long-term outcomes in CD, may change the prioritization of this domain for future COS iterations.[58-60] Second, it was evident that better instruments for measuring disease activity need to be developed, particularly for assessing endoscopy. Limitations to the SES-CD such as weighting towards extent of disease, inability to capture ulcer

depth, and uncertainty regarding the most appropriate method to analyze missing segments were acknowledged.

Our study has several strengths. We used a rigorous, mixed methods approach consistent with COMET recommendations to develop this COS.[16] This allowed us to include different stakeholders, capture patient perspectives, and garner insights from a large panel of internationally recognized experts in IBD with a wealth of both research and clinical experience. The Delphi method has been endorsed for generating a consensus in COS exercises, and iterative rounds of feedback allowed panelists to consider a broad range of viewpoints while voting.[18] However, we also acknowledge some important limitations. First, as previously discussed, there were several outcomes that did not reach the threshold for consensus inclusion, which likely reflects evolution in endpoint definitions over time. Second, we cannot exclude the possibility of voter fatigue given the length of the surveys. However, the survey length was necessary to capture the full range of possible outcome measures in IBD RCTs. To mitigate fatigue, we provided projected times for survey completion, allowed panelists to complete their responses over multiple sittings, only carried forward items that reasonably would be included in a COS (according to defined rules used in other COS development programs[61]), and organized the statements into outcome domains. Although we had originally planned to present the panelists with the statements in random order, the length and complexity of the questions made voting extremely difficult on pilot testing. Third, patients were asked to prioritize the outcome domains of importance and identify other relevant outcomes, but they did not provide input on specific measurement tools because these decisions were primarily based on technical factors. For example, the specific considerations of whether histologic remission should be defined by absence of neutrophilic inflammation was deemed to be less relevant to patients as compared to the overall importance of histologic assessment. For feasibility, patients were recruited from Canada and were all English speaking. Therefore, cultural differences in outcome priority may not have been captured.

In conclusion, we have developed the first internationally guided minimum set of core outcomes for use in RCTs of pharmacologic therapies in adult patients with IBD that captures the evolution of endpoints over the past several decades. Adoption of this COS will improve the quality of evidence synthesis and reduce heterogeneity in outcome reporting. We anticipate this will be the first iteration of a COS for IBD trials, as several key areas of research priority were highlighted in our panel discussions. Additional work to validate both existing and novel instruments for measuring disease activity will shape the next IBD COS.

References

1. Kobayashi T, Siegmund B, Le Berre C, et al. Ulcerative colitis. Nat Rev Dis Primers 2020;6:74.
2. Torres J, Mehandru S, Colombel JF, et al. Crohn's disease. Lancet 2017;389:1741-1755.
3. Baumgart DC, Le Berre C. Newer Biologic and Small-Molecule Therapies for Inflammatory Bowel Disease. N Engl J Med 2021;385:1302-1315.
4. Khanna R, Zou G, Feagan BG. Evolution of the Randomized Controlled Trial in Inflammatory Bowel Disease: Current Challenges and Future Solutions. Inflamm Bowel Dis 2018;24:2155-2164.
5. Hindryckx P, Baert F, Hart A, et al. Clinical trials in luminal Crohn's disease: a historical perspective. J Crohns Colitis 2014;8:1339-50.
6. Colombel JF, Panaccione R, Bossuyt P, et al. Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial. Lancet 2017;390:2779-2789.
7. 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;386:1825-34.
8. Geyl S, Guillo L, Laurent V, et al. Transmural healing as a therapeutic goal in Crohn's disease: a systematic review. Lancet Gastroenterol Hepatol 2021;6:659-667.
9. Yoon H, Jangi S, Dulai PS, et al. Incremental Benefit of Achieving Endoscopic and Histologic Remission in Patients With Ulcerative Colitis: A Systematic Review and Meta-Analysis. Gastroenterology 2020;159:1262-1275 e7.
10. Bojic D, Bodger K, Travis S. Patient Reported Outcome Measures (PROMs) in Inflammatory Bowel Disease: New Data. J Crohns Colitis 2017;11:S576-S585.

11. Burke LB, Kennedy DL, Miskala PH, et al. The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. Clin Pharmacol Ther 2008;84:281-3.
12. D'Haens G, Feagan B, Colombel JF, et al. Challenges to the design, execution, and analysis of randomized controlled trials for inflammatory bowel disease. Gastroenterology 2012;143:1461-9.
13. Ma C, Hussein IM, Al-Abbar YJ, et al. Heterogeneity in Definitions of Efficacy and Safety Endpoints for Clinical Trials of Crohn's Disease: A Systematic Review. Clin Gastroenterol Hepatol 2018;16:1407-1419 e22.
14. Ma C, Panaccione R, Fedorak RN, et al. Heterogeneity in Definitions of Endpoints for Clinical Trials of Ulcerative Colitis: A Systematic Review for Development of a Core Outcome Set. Clin Gastroenterol Hepatol 2018;16:637-647 e13.
15. Ma C, Panaccione R, Fedorak RN, et al. Development of a core outcome set for clinical trials in inflammatory bowel disease: study protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. BMJ Open 2017;7:e016146.
16. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. Trials 2017;18:280.
17. COMET.
18. Kirkham JJ, Davis K, Altman DG, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. PLoS Med 2017;14:e1002447.
19. Kirkham JJ, Gorst S, Altman DG, et al. COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). Trials 2015;16:373.
20. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. PLoS Med 2011;8:e1000393.

21. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395-400.
22. Gracie DJ, Williams CJ, Sood R, et al. Poor Correlation Between Clinical Disease Activity and Mucosal Inflammation, and the Role of Psychological Comorbidity, in Inflammatory Bowel Disease. Am J Gastroenterol 2016;111:541-51.
23. Dulai PS, Jairath V, Khanna R, et al. Development of the symptoms and impacts questionnaire for Crohn's disease and ulcerative colitis. Aliment Pharmacol Ther 2020;51:1047-1066.
24. Higgins PDR, Harding G, Revicki DA, et al. Development and validation of the Ulcerative Colitis patient-reported outcomes signs and symptoms (UC-pro/SS) diary. J Patient Rep Outcomes 2017;2:26.
25. Higgins PDR, Harding G, Leidy NK, et al. Development and validation of the Crohn's disease patient-reported outcomes signs and symptoms (CD-PRO/SS) diary. J Patient Rep Outcomes 2017;2:24.
26. Khanna R, Zou G, D'Haens G, et al. A retrospective analysis: the development of patient reported outcome measures for the assessment of Crohn's disease activity. Aliment Pharmacol Ther 2015;41:77-86.
27. Lewis JD, Rutgeerts P, Feagan BG, et al. Correlation of Stool Frequency and Abdominal Pain Measures With Simple Endoscopic Score for Crohn's Disease. Inflamm Bowel Dis 2020;26:304-313.
28. D'Haens GR, Colombel JF, Bossuyt P, et al. 775a Risankizumab Induction Therapy in Patients With Moderate-to-Severe Crohn’s Disease with Intolerance or Inadequate Response to Conventional and/or Biologic Therapy: Results from the Phase 3 ADVANCE Study. Gastroenterology 2021;161:e28.
29. Best WR, Becktel JM, Singleton JW, et al. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976;70:439-44.

30. 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.

31. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004;60:505-12.

32. 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;30:983-9.

33. Ma C, Battat R, Parker CE, et al. Update on C-reactive protein and fecal calprotectin: are they accurate measures of disease activity in Crohn's disease? Expert Rev Gastroenterol Hepatol 2019;13:319-330.

34. D'Haens G, Kelly O, Battat R, et al. Development and Validation of a Test to Monitor Endoscopic Activity in Patients With Crohn's Disease Based on Serum Levels of Proteins. Gastroenterology 2020;158:515-526 e10.

35. Kennedy NA, Jones GR, Plevris N, et al. Association Between Level of Fecal Calprotectin and Progression of Crohn's Disease. Clin Gastroenterol Hepatol 2019;17:2269-2276 e4.

36. Gecse KB, Brandse JF, van Wilpe S, et al. Impact of disease location on fecal calprotectin levels in Crohn's disease. Scand J Gastroenterol 2015;50:841-7.

37. George J, Singh S, Dulai PS, et al. Corticosteroid-free remission vs overall remission in clinical trials of moderate-severe ulcerative colitis and Crohn's disease. Inflamm Bowel Dis 2020;26:515-523.

38. Rubin D, Panaccione R, Potts Bleakman A, et al. P028 Communicating Needs and Features of IBD Experiences (CONFIDE) Survey: Patient and Healthcare Professional Perspectives on Experience of Ulcerative Colitis Symptoms. Am J Gastroenterol 2021;116:S7.

39. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med 1987;317:1625-9.

40. Lewis JD, Chuai S, Nessel L, et al. Use of the noninvasive components of the Mayo score to assess clinical response in ulcerative colitis. Inflamm Bowel Dis 2008;14:1660-6.

41. Ma C, Sandborn WJ, D'Haens GR, et al. Discordance Between Patient-Reported Outcomes and Mucosal Inflammation in Patients With Mild to Moderate Ulcerative Colitis. Clin Gastroenterol Hepatol 2020;18:1760-1768 e1.

42. Colombel JF, Ordas I, Ullman T, et al. Agreement Between Rectosigmoidoscopy and Colonoscopy Analyses of Disease Activity and Healing in Patients With Ulcerative Colitis. Gastroenterology 2016;150:389-95 e3.

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

44. Travis SP, Schnell D, Krzeski P, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). Gut 2012;61:535-42.

45. Scherl EJ, Pruitt R, Gordon GL, et al. Safety and efficacy of a new 3.3 g b.i.d. tablet formulation in patients with mild-to-moderately-active ulcerative colitis: a multicenter, randomized, double-blind, placebo-controlled study. Am J Gastroenterol 2009;104:1452-9.

46. Khanna R, Ma C, Jairath V, et al. Endoscopic Assessment of Inflammatory Bowel Disease Activity in Clinical Trials. Clin Gastroenterol Hepatol 2020.

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

48. Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, et al. Development and validation of the Nancy histological index for UC. Gut 2017;66:43-49.

49. Magro F, Doherty G, Peyrin-Biroulet L, et al. ECCO Position Paper: Harmonization of the Approach to Ulcerative Colitis Histopathology. J Crohns Colitis 2020;14:1503-1511.
50. Ma C, Sedano R, Almradi A, et al. An International Consensus to Standardize Integration of Histopathology in Ulcerative Colitis Clinical Trials. Gastroenterology 2021;160:2291-2302.
51. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf 1999;20:109-17.
52. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.
53. 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.
54. US Food and Drug Administration. Ulcerative Colitis: Clinical Trial Endpoints Guidance for Industry, 2016.
55. Vermeire S, Danese S, Zhou W, et al. OP23 Efficacy and safety of upadacitinib as induction therapy in patients with Moderately to Severely Active Ulcerative Colitis: Results from phase 3 U-ACCOMPLISH study. Journal of Crohn's and Colitis 2021;15:S021-S022.
56. Sninsky J, Barnes E, Zhang X, et al. P075 Urgency and its Association with Quality of Life and Clinical Outcomes in Ulcerative Colitis Patients. Official journal of the American College of Gastroenterology | ACG 2021;116:S19-S20.
57. 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.

58. Danese S, Vermeire S, D'Haens G, et al. Treat to target versus standard of care for patients with Crohn's disease treated with ustekinumab (STARDUST): an open-label, multicentre, randomised phase 3b trial. Lancet Gastroenterol Hepatol 2022.

59. Danese S, Sandborn WJ, Colombel JF, et al. Endoscopic, Radiologic, and Histologic Healing With Vedolizumab in Patients With Active Crohn's Disease. Gastroenterology 2019;157:1007-1018 e7.

60. Fernandes SR, Rodrigues RV, Bernardo S, et al. Transmural Healing Is Associated with Improved Long-term Outcomes of Patients with Crohn's Disease. Inflamm Bowel Dis 2017;23:1403-1409.

61. Ma C, Schoepfer AM, Safroneeva E, et al. Development of a Core Outcome Set for Therapeutic Studies in Eosinophilic Esophagitis (COREOS): An International Multidisciplinary Consensus. Gastroenterology 2021;161:748-755.

Tables and Figures Legend

Table 1. Core outcomes in Crohn's disease and ulcerative colitis randomized controlled trials

Figure 1. Development of a core outcome set for IBD randomized controlled trials

Figure 2. Patient rating of outcome domains of importance for assessing treatment efficacy in the short (A) and long (B) term

Figure 3. Core outcome domains for inclusion in IBD randomized controlled trials

Supplemental Table 1. Patient and expert panel demographics

Supplemental Table 2. Voting results from Delphi round 2 and ratification vote for Crohn's disease related core outcomes

Supplemental Table 3. Voting results from Delphi round 2 and ratification vote for ulcerative colitis related core outcomes

Table 1. Summary of core outcomes in Crohn's disease and ulcerative colitis randomized controlled trials

Domain	Crohn's Disease	Ulcerative Colitis
Configuration of Outcomes	• The coprimary endpoint of symptomatic remission and endoscopic response should be used in CD trials • Induction outcomes: coprimary symptomatic remission and endoscopic response, symptomatic response, corticosteroid-free remission, endoscopic response, and biomarker remission/response • Maintenance outcomes: coprimary symptomatic remission and endoscopic response, symptomatic and endoscopic remission, corticosteroid free remission, sustained remission, endoscopic remission/response, biomarker remission • Loss of response should be defined by worsening symptoms and either worsening endoscopy or biomarkers • Outcome timing: 9-12 weeks (induction), 52 weeks (maintenance or treat-through designs)	• The composite endpoint of symptomatic and endoscopic remission should be used in UC trials • Response should be defined by the composite of symptomatic and endoscopic response in UC trials • Induction outcomes: composite symptomatic and endoscopic remission, composite symptomatic and endoscopic response, endoscopic response, endoscopic remission, histologic remission, combined clinical and biomarker remission • Maintenance outcomes: clinical response and remission (composite outcomes), corticosteroid-free remission, sustained remission, loss of remission/response, endoscopic remission/response, histologic remission, mucosal healing (combined endoscopic and histologic remission), biomarker remission • Loss of response should be defined by worsening symptoms and either endoscopy or biomarkers • Outcome timing: 9-12 weeks (induction), 52 weeks (maintenance or treat-through designs)
PROs, Symptom-based Measures, and Composite Indices	• A PRO for CD should include stool frequency and abdominal pain • CDAI should be used as a composite outcome measure • Clinical response should be defined by CDAI reduction >100 points compared to baseline	• A PRO for UC should include rectal bleeding, stool frequency, and fecal urgency • The adapted 9-point MCS (including rectal bleeding, stool frequency, and mMES) should be used in UC trials • Symptomatic remission should be defined by rectal bleeding subscore = 0 and stool frequency subscore ≤1
Endoscopic Outcomes	• Endoscopic outcomes should be assessed using the SES-CD • Endoscopic response should be defined by >50% reduction in SES-CD vs. baseline • Endoscopic remission should be defined by absence of ulcerations in all segments • Endoscopic remission in isolated ileal CD should be defined by SES-CD ≤2	• Endoscopic outcomes should be assessed by flexible sigmoidoscopy in UC trials • Scoring should be based on the most affected segment • Endoscopic remission should be defined as mMES = 0 • Endoscopic response should be defined by reduction in mMES ≥1 from baseline

Domain	Crohn's Disease	Ulcerative Colitis
	• Missing segments should be reported at baseline and after treatment • Endoscopic response should be measured in induction trials • Endoscopic remission and response should be measured in maintenance trials at 1 year	• Endoscopic response and remission should be measured in UC induction trials at 9-12 weeks • Endoscopic response and remission should be measured in UC maintenance trials at 52 weeks
Histopathology	• Not voted as a core domain for CD	• Histopathology should be scored using the RHI • Histologic remission should be defined by RHI <3 with absence of neutrophils (or Geboes score <3.0 with no neutrophilic inflammation in the epithelium) • Histologic remission should be measured in both induction and maintenance trials
Biomarker Outcomes	• CRP and fecal calprotectin should be measured in CD induction and maintenance trials • Biomarker remission should be defined in CD trials by CRP < 5mg/L • Biomarker response should be defined by >50% reduction in CRP and fecal calprotectin among those with elevated levels at baseline	• Fecal calprotectin should be measured in UC induction and maintenance trials • Biomarker remission should be defined by UC trials by CRP < 5mg/L or fecal calprotectin < 150 µg/g • Biomarker response should be defined by >50% reduction in fecal calprotectin among those with elevated levels at baseline

Abbreviations: CD Crohn's disease; CDAI Crohn's Disease Activity Index; CRP C-reactive protein; MCS Mayo Clinic Score; mMES modified Mayo Endoscopic Subscore; PRO patient reported outcome; RHI Robarts Histopathology Index; SES-CD Simple Endoscopic Score Crohn's Disease; UC ulcerative colitis

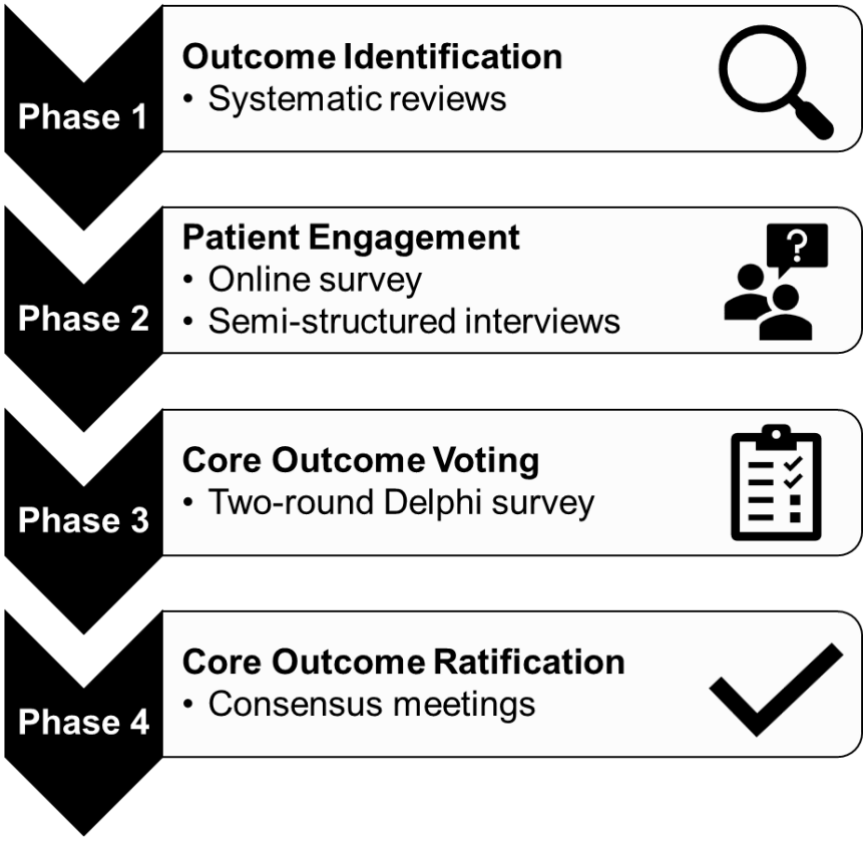

Figure 1. Development of a core outcome set for IBD randomized controlled trials

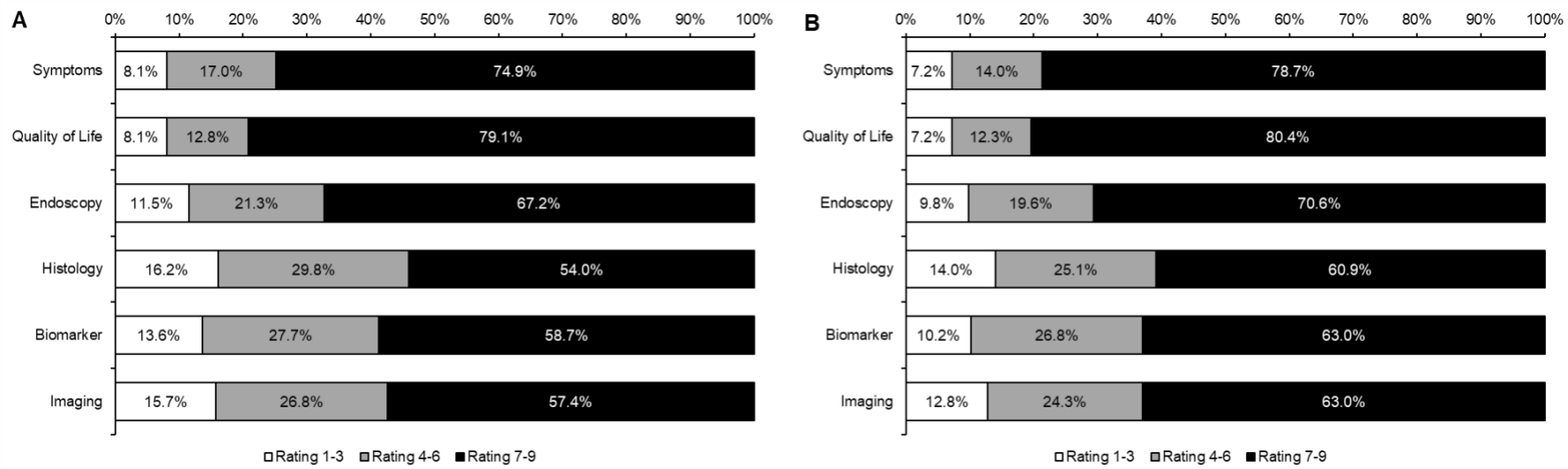

Figure 2. Patient rating of outcome domains of importance for assessing treatment efficacy in the short (A) and long (B) term.

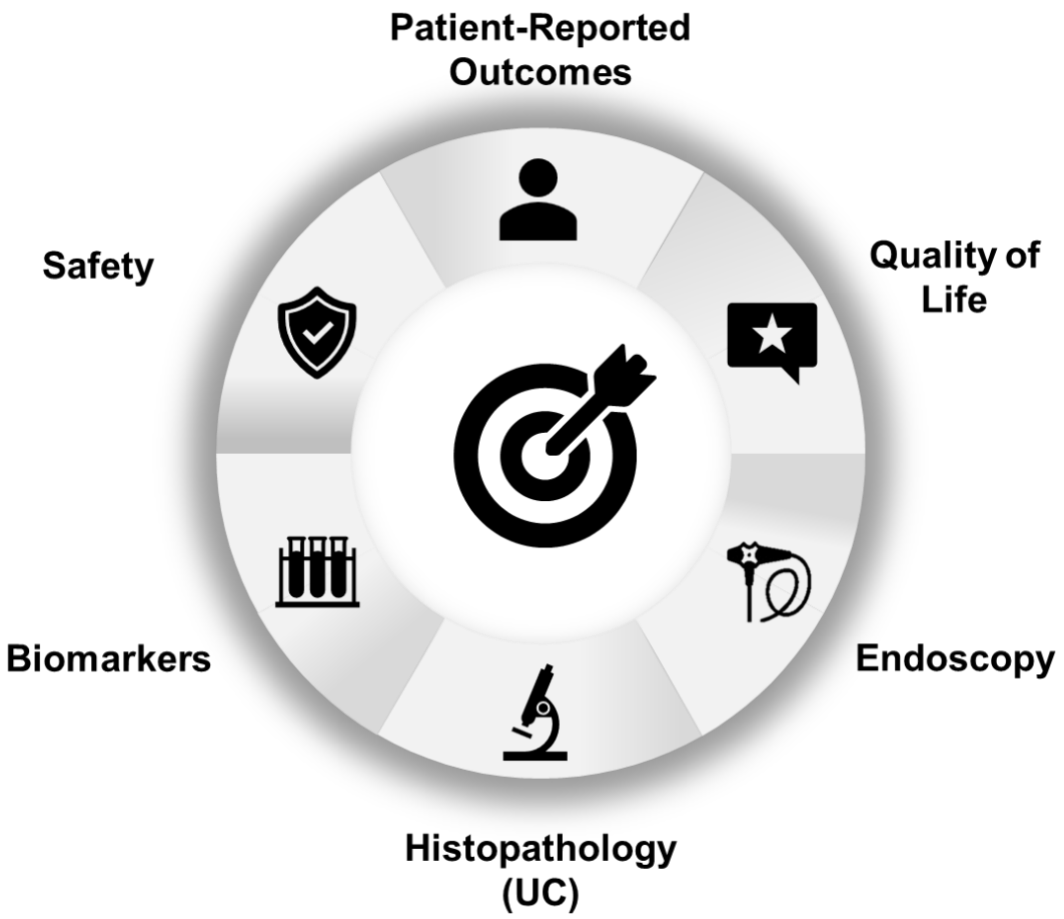

Figure 3. Core outcome domains for inclusion in IBD randomized controlled trials