

The Inflammatory Bowel Disease Symptom Inventory: A Patient-report Scale for Research and Clinical Application

Kathryn A. Sexton, PhD,^{*,†} John R. Walker, PhD,^{*,†} Laura E. Targownik, MD,^{*,‡} Lesley A. Graff, PhD,^{*,†} Clove Haviva, MA,^{*,§} Brooke E. Beatie, MA,[§] Sarah K. Petty, MA,[§] Matthew T. Bernstein, MA,^{†,§} Harminder Singh, MD,[‡] Norine Miller, RN,^{*} and Charles N. Bernstein, MD^{*,‡,◉}

Objectives: Existing measures of inflammatory bowel disease (IBD) symptoms are not well suited to self-report, inadequate in measurement properties, insufficiently specific, or burdensome for brief or repeated administration. We aimed to develop a patient-reported outcome measure to assess a broader range of IBD symptoms.

Methods: The IBD Symptoms Inventory (IBDSI) was developed by adapting symptom items from existing clinician-rated or diary-format inventories; after factor analysis, 38 items were retained on 5 subscales: bowel symptoms, abdominal discomfort, fatigue, bowel complications, and systemic complications. Participants completed the IBDSI and other self-report measures during a clinic visit. A nurse administered the Harvey Bradshaw Index (HBI) for Crohn's disease (CD) or the Powell-Tuck Index (PTI) for ulcerative colitis (UC), and a gastroenterologist completed a global assessment of disease severity (PGA).

Results: The 267 participants with CD (n = 142) or UC (n = 125), ages 18 to 81 (M = 43.4, SD = 14.6) were 58.1% female, with a mean disease duration of 13.9 (SD = 10.5) years. Confirmatory factor analysis supported the 5 subscales. The total scale and subscales showed good reliability and significant correlations with self-report symptom and IBD quality of life measures, the HBI, PTI, and PGA.

Conclusions: The IBDSI showed strong measurement properties: a supported factor structure, very good internal consistency, convergent validity, and excellent sensitivity and specificity to clinician-rated active disease. Self-report HBI and PTI items, when extracted from this measure, produced scores comparable to clinician-administered versions. The 38-item IBDSI, or 26-item short form, can be used as a brief survey of common IBD symptoms in clinic or research settings.

Key Words: symptoms, inflammatory bowel disease, measurement properties, self report

INTRODUCTION

The availability of reliable and accurate assessment measures is important in both the research and provision of specialized clinical services. A number of useful clinician-administered measures have been developed for Crohn's disease (CD)^{1,2} and ulcerative colitis (UC),^{3,4} and these have facilitated research and clinical work. In their 2016 review, Walsh et al⁵ reviewed activity measures for CD and UC recently and recommended the use of separate and independently validated

indices for clinical, endoscopic, histological, and quality of life (QOL) assessment. They identified 8 measures for UC and 8 for CD that focus primarily on symptoms and indicate that many others include both symptom ratings and other data in the same measure. They also note that some of the symptom measures may be used independently by the patient. A review of the wording of the measures, however, suggested that in many cases, the medical language involved would be unfamiliar to some patients, and the rating scales are inconsistent within and across measures. The review concluded that there

Received for publications September 4, 2018; Editorial Decision December 20, 2018.

From the departments of ^{*}Inflammatory Bowel Disease Clinical and Research Centre; [†]Clinical Health Psychology; [‡]Internal Medicine; and [§]Psychology, University of Manitoba, Winnipeg, Manitoba, Canada

Conflicts of Interest: CB has served on advisory boards of Abbvie Canada, Ferring Canada, Janssen Canada, Napo Pharmaceuticals, Pfizer Canada, Shire Canada, Takeda Canada and has consulted to 4D Pharma and Mylan Pharmaceuticals. He has received unrestricted educational grants from Abbvie Canada, Janssen Canada, Shire Canada, and Takeda Canada. He has been on speaker's bureau of Ferring Canada and Shire Canada.

LET is on the speaker's panel for Janssen Canada, Takeda Canada, and Pfizer Canada and on the advisory boards for Takeda Canada, Abbvie Canada, and Janssen Canada.

HS is on the advisory board Merck Canada, Pendopharm, Ferring Canada and investigator-initiated research grant Merck Canada. All other authors have no conflicts to report.

Author Contribution: KAS, JRW, CNB, and LAG contributed to the study concept and design. KAS, NM, CH, BEB, SKP, and MTB contributed to the acquisition

of data. KAS, JRW, CNB, LET, LAG, and BEB contributed to the analysis and interpretation of data. KAS, JRW, CNB, LET, and LAG contributed to the drafting of the manuscript. KAS, CNB, LET, JRW, and LAG contributed to the critical revision of the manuscript for important intellectual content and statistical analysis.

Supported by: KAS was supported by a post-doctoral fellowship from the Canadian Institutes of Health Research and the Canadian Association of Gastroenterology. LET received grant support from Pfizer Canada and Abbvie Canada.

Address correspondence to: Charles N. Bernstein, Department of Internal Medicine (Gastroenterology), University of Manitoba, 804F-715 McDermot Avenue, University of Manitoba, Winnipeg, MB R3E3P4, Canada. E-mail: Charles.Bernstein@umanitoba.ca.

© 2019 Crohn's & Colitis Foundation. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

doi: 10.1093/ibd/izz038

Published online 27 March 2019

is a pressing need for validation of the measures used to assess CD and UC.

To date there has been much less work on the development of patient-reported outcome measures (PROMs) for IBD. Patient-reported outcome measures have a variety of advantages. Patients are in a unique position to observe their own symptoms over time and capture experiences that cannot easily be observed by the clinician. In fact, clinician-administered measures often involve a process where the clinician asks the patient to describe their symptoms and even suggests categories for responding. Patient-reported outcome measures are less costly to use and may be used when the clinician is not present to record responses. The use of self-report measures may encourage patient engagement in assessment and treatment. Finally, PROMs often correlate well with clinician-administered measures, as is the case with the Powell-Tuck Index (PTI, 4) and the Simple Clinical Colitis Activity Index¹ for UC and the Harvey Bradshaw Index (HBI, 2) for CD.

Existing IBD symptoms measures, both clinician-administered and self-report, have often been designed without the benefit of core methods now used in the development of assessment measures, such as exploratory and confirmatory factor analysis. Further, the wording of items and rating scales are often not consistent with current development approaches. Many rating scales restrict variability by offering too few rating options and suffer from unequal intervals between anchor points. Additionally, existing scales often assesses an insufficient breadth of the symptoms commonly experienced in CD and UC or use too few items to sample each type of symptom adequately for the scale to have good measurement properties. Generally, existing scales have shown inadequate evidence of reliability and validity.⁶

A 2017 systematic review of patient-reported outcome measures in IBD identified 20 measures.⁷ Four of these were single-item measures; many of the others were developed by using items from previously used clinician-rated measures. Some of the measures assessed specific symptoms while others assessed global disease activity or the sense that the disease was well-controlled. Many had good evidence for correlation with clinician-rated measures. In many cases, the relationship between the PROM and measures of inflammation or endoscopic findings was modest.^{8–11} This fits with recent evidence that patients may have active symptoms with little evidence of inflammation and vice versa.^{12, 13} Our earlier research¹³ indicated a bidirectional association between anxiety and depression (emotional distress) and symptom measures. Being ill is an especially stressful situation. Many patients also attribute exacerbations in symptoms to changes in exposure to dietary factors.^{14, 15} The authors⁷ identified a need for better evaluated PROM with good evidence for validity and reliability.

We undertook to develop a PROM of symptoms that could be used in both CD and UC among patients with no special knowledge of the use of medical language to describe

symptoms. As these conditions are associated with a wide range of symptoms which may vary in presentation among patients over time, we planned for the measure to provide subscale scores for logical groupings of symptoms to better characterize each patient's experience. Many of the clinician-reported IBD measures that have been commonly used were developed many years ago and were not founded on appropriate scale development and validation principles for symptom measures. The approach in the development of this new measure considered current recommendations for the development of symptom measures (Consensus-based Standards for the Selection of Health Measurement Instruments, COSMIN).¹⁶

This article reports on the development and validation of a new IBD Symptom Inventory (IBDSI) for patients with CD or UC, assessing its measurement properties and concordance with other patient and clinician assessments of IBD symptoms and disease status. A goal was to have a range of items addressing symptoms in a variety of areas that are common concerns in IBD.

METHODS

Development of the Measure

Given the goal of facilitating comparability to symptom measures used previously, we started by considering symptoms assessed in 2 measures used widely in research on IBD: the HBI for CD and the PTI for UC. While the indices overlap considerably in content, they differ in the rating scales used and the anchor points on the scales. This poses a significant problem for research seeking to compare these patient groups on disease severity, naturalistic course of the disease, or treatment outcomes. In practice, the uneven rating scales are also problematic when following symptom reports over time or following interventions, as it is not clear how to establish the amount of change that is clinically meaningful across disease types and different levels of disease severity. In creating the new measure, we used 4- or 5-point ratings scales and revised wording to better reflect equal intervals across the rating points, with the aim of producing a measure with stronger measurement characteristics. The questions were modified for self-report administration using everyday nontechnical language.

Problematic items were specifically addressed. That is, there has been some concern in previous research about the bowel frequency HBI item: “number of liquid stools per day) ___” as compared with the bowel frequency item on the PTI: “less than 3 (rated 1), 3 to 6 (rated 2), and more than 6 (rated 2).” For the HBI, if a respondent has frequent liquid bowel movements, as might be the case in patients who have had past bowel resections, especially ileocecal resections, their resulting clinical index score could be quickly driven up into the active range without there being active disease, as the scoring for this item results in a much greater weight than the rating on the other items. The item and scoring were revised as follows:

“The number of loose/liquid bowel movements or diarrhea I had most days was: ___”, which was then converted into a 5-point scale of symptom severity. With this approach, liquid and stool frequency do not have a disproportionate influence on the total scale score.

The HBI also has an item asking the clinician to assess the patient for an abdominal mass by palpating the abdomen. The response alternatives are “none, dubious, definite, and definite and tender.” We asked: “Has your doctor ever told you that you have a mass or a lump in your abdomen?” with a “yes/no” response. This item allows comparison to the HBI but is not included in the total IBDSI score. A “yes” response is followed by the question: “In the past two weeks, I had a lump in my abdomen,” with a 4-point rating scale for severity, which is included in the scoring.

To broaden the content of the measure, a pool of 21 additional items were compiled from existing clinician-rated or diary-format symptom inventories and from the experience of clinicians on the research team who specialize in the treatment of IBD. These items cover concepts such as fatigue, gas and bloating, urgency of bowel movements, soiling, difficulties with weight, fever, and sleeping. These complement items from the HBI/PTI that assess general health, abdominal pain, consistency of bowel movements, and IBD-related complications, overall, aiming to reflect symptoms across a variety of areas that are common concerns in IBD. The compiled 18 modified questions from the HBI and PTI and the additional questions were administered to patients in clinic (ie, a total of 39 questions with 42 responses).

In developing the measure, we used the COSMIN standards for the development of health-related patient-reported outcome measures.¹⁶ This approach involves the assessment of measurement qualities including: A) internal consistency, B) reliability, C) measurement error, D) content validity, E) structural validity, F) hypothesis testing, G) cross-cultural validity, H) criterion validity, I) responsiveness, and J) interpretability. All of these qualities cannot be measured in a single study. This article describes the assessment on items A to E, H, and J. Criteria F and I will be assessed in a current study by our group assessing IBD symptoms on a biweekly basis over a 1-year period. The 1-year period will allow the assessment of changes in symptoms that come with adding or removing specific treatments. Criterion G, cross cultural validity, will be assessed in a future study involving other language groups, such as French and Spanish language samples.

The 21 additional items were gathered in discussions with gastroenterologists (CNB, LET, and HS), clinical nurse specialist (NM), and a psychologist with extensive experience in the treatment of persons with IBD (LAG). These items were then assessed in pilot work with the extensive list of items. Of the 21 new questions, 17 non-redundant questions were retained after an evaluation of item reliabilities, response frequencies, item discriminability, and exploratory factor analysis (data not

shown). Two questions on fistulas (one from the HBI/PTI, and one added question) were combined into one question with subparts, and one PTI question was separated into 2 different ratings (nausea and vomiting), leaving a total of 35 questions with 38 responses. One of these questions was retained for its relevance in inferring a HBI/PTI score but was not included in the final sum of the IBDSI as it was conceptually redundant with other items that provided superior ratings. Similarly, all components of the combined items were retained for purposes of clarifying the question or response and to ensure an HBI/PTI score could be accurately derived, but some subparts are not scored because of item redundancy.

Five IBD symptom subscales were identified in the factor analysis including: 1) bowel symptoms (10 questions; 9 are scored), 2) abdominal discomfort (11 questions), 3) fatigue (6 questions), 4) bowel complications (3 questions), and 5) systemic complications (5 questions). The fatigue subscale includes 6 items and assesses difficulties with feeling tired, being in poor perceived health, and experiencing weight loss. The 2 remaining subscales assess IBD-related complications: the bowel complications subscale includes items rating the severity of any fistulas, anal fissures, or abscesses that are present, whereas the systemic complications subscale is composed of items assessing arthralgia, uveitis/iritis, erythema nodosum, aphthous ulcers/aphthous stomatitis, and pyoderma gangrenosum.

The Inflammatory Bowel Disease Symptom Inventory (IBDSI-LF, 38 items) was intended to comprehensively address IBD symptoms, including uncommon symptoms that would be clinically important. A 26-item short form (IBDSI-SF) was derived from the core modified HBI and PTI items and from additional representative subscale items to facilitate brief or repeated administrations while still providing for the factor scores. The short form eliminates some symptoms, particularly some of the less commonly experienced IBD-related complications. In contrast to the long form, all items are scored to maximize the reliability of this shorter version. The IBDSI-LF and IBDSI-SF can be found in Supplementary Appendices A and B, respectively. To enable direct comparisons with measures in long-standing use in past research and clinical trials, a self-report HBI and PTI score can be extracted directly from the items on the IBDSI-LF and inferred from the IBDSI-SF (see instructions in Supplementary Appendix C). The HBI/PTI items omitted from the short form were the IBD-related complications that are infrequently endorsed (uveitis, erythema nodosum, aphthous ulcers, pyoderma gangrenosum, and anal fissure). Two complications were retained in the short form: the presence of a fistula, given its pronounced predictive utility with respect to overall disease status in CD, and difficulty with arthralgia (described as “joint pain”) given its more frequent endorsement and higher prevalence.

Rating approaches for each item are presented in Appendix A and B of the online supplementary material. While most items are scored from 0 to 3 or 0 to 4, for some items adopted

from the HBI and PTI this scale was adjusted to better reflect discontinuities between anchor points (eg, “none,” “mild,” and “prolonged/severe” were extended to “none,” “mild,” “moderate,” “prolonged/severe”). Similarly, items with frequency counts that could be very wide-ranging (eg, counts of number of days with bowels symptoms) were coded into a 5-point scale to prevent high counts from being unduly weighted compared with other items. Thus, after coding, the total score on the measure may be calculated as a simple sum across the ratings. If a small number of items has been left blank (33% or less), the total score may be prorated to adjust for the missing items. In the case of items from the bowel and systemic complications subscales, as many of these symptoms are uncommon, it was determined that if items are missing they should not be counted into the total score.

For subscale scores, to accommodate the different number of items on the bowel symptoms, abdominal discomfort, and fatigue symptoms subscales, the mean is used. These mean scores may be used to compare the subscales on the same metric. If a small number of items on a subscale has been left blank (25% or less), the average subscale score may be calculated by dividing by the number of valid responses. In the case of the bowel and systemic complications subscales, as many of these symptoms are uncommon, it is best just to calculate an average score for the subscale by dividing by the total number of questions posed.

To evaluate the clarity and readability of the IBDSI-LF, a patient panel of 11 participants (7 women, 4 men) independently reviewed the final version of the IBDSI-LF as part of pilot testing for a larger study (see the description of sample 2). Ten of eleven patients agreed that overall the questions were clear and easy to answer. Some specific concerns were noted regarding 5 items (items 3, 11, 18, 24, and 35); these were considered in developing the scoring approach to these items. Specifically, one patient noted that on item 3 the number of bowel movements differed day to day, making it difficult to report an “average”; however, difficulty recalling an “average” number of bowel movements is made less relevant by the coding of this variable into a 5-point Likert scale. One patient noted that items 11 and 24, which ask about tiredness and energy level, were highly similar; although this is very helpful for the reliability of the subscale, we did take advantage of this by retaining only item 11 in the short form. One patient was confused about the breadth of situations being queried in item 18, in which one might be “finding it hard to get things done,” but we judged that the vagueness of this question was helpful in getting an overall rating of a person’s functioning and level of fatigue. Finally, one patient noted that the presence of any fistula could be considered very severe and expressed concern about the usefulness of rating its severity. The final scoring approach to this item was to score only the more concrete rating of “fistula activity” and to incorporate the “yes/no” answer regarding fistulas into the coding approach so that any fistula nonetheless receives at least 1 point toward the total score.

Thus, overall, patients found the IBDSI-LF items to be clear and understandable.

Study 1: Validation Sample

Participants were recruited from the tertiary care outpatient IBD clinic at the Health Sciences Centre, the largest academic and clinical hospital in the central Canadian province of Manitoba, based in Winnipeg, with an enrollment period from July 2014 to May 2015. Eligible adults with a diagnosis of either CD or UC were approached and provided with a letter describing the study and inviting them to participate. Interested patients who provided written informed consent to complete the self-report measures and to use information collected by the clinical nurse specialist and the gastroenterologist in the usual course of patient care were enrolled in the study. The self-report measures were collected before the patient saw the clinical nurse specialist and then the gastroenterologist. The data collected in study 1 were reviewed by the clinical nurse specialist after they had independently completed the HBI for CD or the PTI for UC. The nurse ensured that participants had not skipped pages or missed major sections of the survey. Participants in study 1 could skip over a question if they did not understand it. This study was approved by the University of Manitoba Research Ethics Board.

Self-Report Measures

Manitoba inflammatory bowel disease index

The Manitoba Inflammatory Bowel Disease Index (MIBDI) assesses symptom activity over the previous 6 months with a validated 6-level single item scale.¹⁷ In this study, IBD symptom frequency was assessed over the previous 3 months. The question posed was: In the past 3 months, my disease has been: 1) constantly active (symptoms every day), 2) often active (symptoms most days), 3) sometimes active (symptoms on some days, for instance 1–2 days/week), 4) occasionally active (symptoms 1–2 days/month), 5) rarely active (symptoms only a few days of the past 3 months), and 6) I was well in the past 3 months (what I consider a remission or absence of symptoms). We reversed this scoring so that higher scores indicate more frequent symptoms. When categorizing symptoms as active or inactive, those who reported symptoms over the previous 3 months as being at least sometimes active (a few days every other week) were defined to have active symptoms.

The inflammatory bowel disease questionnaire

The Inflammatory Bowel Disease Questionnaire (IBDQ)¹⁸ is the most commonly used and extensively validated disease-specific quality of life measure used in IBD research. The 32 items assess 4 areas: bowel symptoms, emotional health, systemic symptoms, and social functioning. Higher scores indicate higher quality of life.

Cohen perceived stress scale

The 4-item short form of the Cohen Perceived Stress Scale (CPSS)¹⁹ assesses the degree to which the person feels they are experiencing stress in the past month.

Demographic information

Participants provided information on their age, sex, current marital status (eg, single, married/common-law, divorced, separated, widowed), and educational background (eg, number of grades completed, and number of years of post-secondary university or business/technical college education if applicable).

Clinical history

Participants reported their diagnosis (confirmed by medical chart review), year of diagnosis, past experience with and date of any IBD-related surgeries, use of prednisone and/or biologic (anti-TNF) agents in the last 3 months, and smoking.

Clinician-administered Measures

Harvey Bradshaw index for Crohn's disease

The 12-item instrument Harvey Bradshaw Index for Crohn's disease (HBI)² measures disease-specific symptoms over the previous week. Scores ≥ 5 on the HBI are consistent with active symptoms. This measure or the PTI described later on were administered by a clinical nurse specialist with extensive experience with IBD.

Powell-Tuck index for ulcerative colitis

The Powell-Tuck Index for ulcerative colitis (PTI) is a 16-item instrument⁴ that assesses symptoms specific to UC and related complications over the previous week. Scores ≥ 5 on the PTI indicate active symptoms.

Gastroenterologist global assessment

For the Gastroenterologist Global Assessment, the gastroenterologist provided a global assessment of the severity of disease activity on a 4-point Likert scale, where 0 indicates inactive (in remission); 1, mildly active; 2, moderately active; and 3, severely active disease. Further, the examining gastroenterologist provided a rating of disease status as either 0, "not experiencing a symptom flare currently," or 1, "flaring currently (IBD symptoms are worse than usual in any way)."

Study 2: Test-retest Reliability Sample

The data used to assess the test-retest reliability of the IBDSI-SF were collected from a population-based cohort (nonoverlapping with study 1) recruited for a study involving repeated assessment of the experiences of persons with IBD over a 1-year period. This also afforded the opportunity to confirm the reliability of the short form when administered independently per current guidelines in the development of

short forms.²⁰ In this study, the IBDSI-SF was administered through survey software every 2 weeks, and the IBDSI-LF was administered at baseline, 6 months, and 1 year later. The online survey was set up so that respondents had to completely answer earlier questions before subsequent questions could be viewed and answered. Data collected in that study at weeks 2, 4, 10, and 26 were used to evaluate test-retest reliability. This study was also approved by the University of Manitoba Research Ethics Board.

Data Analysis

The COSMIN checklist of criteria for measures of patient-reported outcomes²¹ was followed to evaluate the measurement characteristics of the IBDSI-LF and IBDSI-SF. We specifically assessed the following items from the COSMIN checklist: construct validity, internal consistency (of scale total and subscales), test-retest reliability, amount of missing data, convergent validity, concurrent validity, known groups validation, and criterion-related validity. Construct validity refers to the degree to which a test or scale measures what the latent variable (in this case, IBD symptoms) purports to measure. To this end, we undertook a confirmatory factor analysis of the IBDSI-LF subscales and of the IBDSI-SF factor structure. We assessed internal consistency of the scale and subscales through the Cronbach α measurement, through intra-class correlation coefficients, and an examination of item-total intercorrelations.

We assessed the convergent validity of the IBDSI-LF by examining associations among the new measure and its subscales and the self-rated MIBDI, the nurse-rated HBI for CD or PTI for UC, the gastroenterologist's global assessment, and the self-rated Cohen's Perceived Stress Scale. We expected that the IBDSI-LF would be strongly correlated with the MIBDI, the HBI or PTI, and the gastroenterologist's global assessment and less strongly related to the CPSS. While the other measures are all focused on assessing symptoms, the CPSS assesses a different concept, which has been shown to be related to symptoms but not as strongly as other symptom measures.^{7, 8} Concurrent validity was examined by comparing IBDSI-LF scores of those with and without active symptoms, as reported on the MIBDI, evaluated on the HBI or PTI, or as assessed by a gastroenterologist's global assessment. Known groups validity was explored by comparing scores across CD and UC diagnostic groups.

Criterion validity was evaluated by comparing active symptom status based on the recommended cutoff scores for the HBI for CD and the PTI for UC (criterion standard) with the IBDSI-LF and IBDSI-SF. For each symptom scale, we computed sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) as compared with the criterion standard. In addition, we used receiver operating curve (ROC) analysis to identify the best cut point in this IBD sample for predicting the presence of active symptoms by maximizing the Youden J index (sensitivity + specificity - 1),²² in which sensitivity and specificity are balanced.

When items from any of the scales were left unanswered by respondents, we prorated the score on that scale if at least two thirds of the elements were completed. In contrast, for the factor analyses, any participant with missing data on the IBDSI-LF was not included, as any replacement value for an individual item could bias the findings.

RESULTS

Demographic characteristics of study 1 and study 2 samples are described in [Table 1](#) and [Table 2](#), respectively. We do not have the precise response rate, but it is estimated that more than 90% of those approached to complete the symptom survey agreed to do so. In study 1, survey data were available for 267 participants, 125 with a diagnosis of CD and 142 with UC, ages 18 to 81 years ($M = 43.4$, $SD = 9.5$), 58.1% female. The mean disease duration was 13.9 ($SD = 10.5$) years. The study 2 sample included 154 respondents, 69.5% female, 65.6% with CD, with a mean age of 42.7 years ($SD = 12.6$) and a mean disease duration of 14.8 years ($SD = 10.3$).

TABLE 1. Demographic and Disease Characteristics of Participants in Study 1

Variable	Crohn's Disease (CD)	Ulcerative Colitis (UC)
	N = 142	N = 125
Sex: Female, n (%)	64 (45.1%)	78 (54.9%)
Marital status, n (%)		
Married/common-law	93 (66.0%)	78 (62.4%)
Separated/divorced/ widowed	8 (5.7%)	17 (13.6%)
Single/never married	40 (28.4%)	30 (24.0%)
Mean number of years of education	14.2 (SD = 2.7)	14.3 (SD = 2.8)
Mean age, in years	54.7 (SD = 13.0)	56.9 (SD = 13.1)
Mean age at diagnosis, in years	31.6 (SD = 12.7)	35.2 (SD = 13.2)
<17 years, n (%)	34 (24.3%)	18 (15.0%)
17–39 years, n (%)	85 (60.7%)	74 (61.7%)
≥40 years, n (%)	21 (15.0%)	28 (23.3%)
Previous IBD-related surgery, n (%)	78 (56.5%)	3 (2.5%)
Current smoker, n (%)	27 (19.3%)	13 (10.7%)
Medication use in previous 3 months, n (%)		
Prednisone	9 (6.7%)	14 (11.7%)
Biologic agent	60 (43.5%)	24 (20.0%)
Manitoba IBD Index, proportion active (symptoms ≥ a few days every other week), n (%)	26 (18.7%)	19 (15.6%)

Confirmatory Factor Analysis of the IBDSI

[Table 3](#) shows the factor loadings of the IBDSI-LF items on the 5 proposed subscales. The bowel symptoms subscale is composed of 10 items (9 of them scored) and assesses difficulties with frequent bowel movements, urgency, fever, blood in stool, and soiling. The abdominal discomfort subscale, composed of 11 items, assesses difficulties with abdominal pain, any abdominal mass, gas, nausea and vomiting. Although the earlier administration of the scale asked about nausea and vomiting in a single item, these items were separated in the final version of the scale; findings from study 2 suggest this was warranted, in part due to the correlation between the items ($r = 0.28$).

Goodness of fit indices for the confirmatory factor analysis were all within recommended guidelines for good fit²¹ and are as follows: Comparative Fit Index (CFI) = 0.95, Bentler-Bonnet Non-Normed Fit Index (NNFI) = 0.94; Standardized Root Mean-square Residual (SRMR) = 0.08; Root Mean-square Error of Approximation (RMSEA) = 0.06 (90% confidence interval [CI], 0.06–0.07).

The IBDSI-SF retains 3 of these 5 factors: bowel symptoms, abdominal discomfort, and fatigue. Two complications were retained in the short form due to their more frequent endorsement: arthralgia is scored with the abdominal discomfort factor in the short form, and fistula ratings are scored on the bowel symptoms factor. Goodness of fit indices for the confirmatory factor analysis of the short form upon independent administration in sample 2 (N = 140) are as follows: Comparative Fit Index (CFI) = 0.95, Bentler-Bonnet Non-Normed Fit Index (NNFI) = 0.94; Standardized Root Mean-square Residual (SRMR) = 0.08; Root Mean-square Error of Approximation (RMSEA) = 0.07 (90% CI, 0.05–0.08). All factor loadings of the IBDSI-SF (see [Supplementary Table S1](#)) on the 3 retained factors were significant at $P < 0.05$.

Measurement Properties of the IBDSI

Reliability

The IBDSI-LF items showed very good internal consistency ($\alpha = 0.93$), a moderate intraclass correlations coefficient (ICC = 0.29; 95% CI, 0.25–0.34), and very good corrected item-total correlations (see [Table 3](#)). Internal consistency of subscales for the IBDSI-LF are reported in [Table 4](#). The bowel symptoms, abdominal discomfort, and fatigue factors showed very good to excellent internal consistency. The 2 complication factors showed lower but acceptable internal consistency, which was not unexpected given that there are fewer items sampling these factors and that these symptoms are less frequently endorsed. Despite their lower base rate, however, the presence of these symptoms is clinically significant, and as factor analyses showed a decrement in the overall goodness of fit when combining these 8 items, they were thus retained as 2 distinct factors. Subscale intercorrelations are presented in [Table 3](#) and support the conceptualization of these 5 symptom facets as distinct but related.

TABLE 2. Demographic and Disease Characteristics of Participants in Study 2

Characteristic	Value (N = 154)
Sex: Female, n (%)	107 (69.5%)
European-Canadian ethnicity, n (%)	130 (84.4%)
Marital status, n (%)	
Married/common-law	97 (63.0%)
Separated/divorced/widowed	22 (14.3%)
Single/never married	35 (22.7%)
Mean number of years of education	15.7 (SD = 3.1)
Mean age at enrollment, in years	42.7 (SD = 12.6)
Mean disease duration, in years	14.8 (SD = 10.3)
Mean age at diagnosis, in years	28.1 (SD = 11.6)
A1: <17 years, n (%)	24 (15.6%)
A2: 17–39 years, n (%)	104 (67.5%)
A3: ≥40 years, n (%)	25 (16.2%)
Proportion with Crohn's disease, n (%)	101 (65.6%)
Proportion with a comorbid chronic/medical condition, n (%)	90 (58.4%)
Previous IBD-related surgery, n (%)	53 (34.4%)
IBD-related hospitalization in the past year, n (%)	19 (12.3%)
Current smoker, n (%)	27 (17.5%)
Medication use in previous 3 months, n (%)	
Prednisone	4 (2.6%)
Biologic agent	49 (31.8%)
Manitoba IBD Index, proportion active (symptoms ≥ a few days every other week), n (%)	93 (60.3%)

Measurement properties of the IBDSI-SF are presented in Table 5. The IBDSI-SF showed comparable internal consistency ($\alpha = 0.92$, sample 1; $\alpha = 0.90$, sample 2), a moderate ICC = 0.27 (95% CI, 0.22–0.33; sample 2), and corrected item-total correlations ranging from $r = 0.21$ to $r = 0.67$ (sample 2; see Supplementary Table S1). The test-retest reliability of the IBDSI-SF for administrations 2 weeks and 2 months apart in study 2 in a computer-administered format is also shown in Table 5. These correlations provide strong evidence of stability over the short-term. Furthermore, the test-retest correlations are progressively more moderate over somewhat longer periods (eg, see the IBDSI-SF at 2 months and the IBDSI-LF at 6 months, shown in Table 4), suggesting that the measure is likely also appropriately sensitive to change. Finally, independent administration the IBDSI-SF in sample 2 showed subscale intercorrelations comparable to the long form (see Table 5), again suggesting that the factors capture a good balance of related but distinct symptom groups.

Missing data on individual questions

A large amount of missing data on particular questions in a measure suggests that respondents find that question difficult to understand or to answer. We assessed the number of

respondents who started the questions in the IBDSI-LF but left some questions unanswered. Overall, there were very little missing data on the IBDSI-LF. Twenty-one out of 39 items had missing data from >1% of the respondents, but only 4 items (4, 5, 29, and 35a) had a relatively high percent (>2%) of missing data (5.6% of respondents did not respond to item 4 about liquid stools; 3% of respondents did not respond to item 5 on stool consistency; and 2.2% of respondents did not respond to items 29 and 35a on IBD complications). Of the 26 participants who said “yes” to item 35a and should have gone on to answer items 35b through 35d, 0%, 0.4%, and 0.4% missed them, respectively. There was thus no indication from missing data that there were any particular items that were difficult to answer. Of the 267 who responded to at least 1 IBDSI-LF item, 81.3% answered all items, 12.7% missed 1, and only 6.0% (16 participants) missed more than 1 item. Three of these individuals were missing more than 25% of the items on the IBDSI and so were excluded from the analyses.

Convergent validity

Table 6 shows the correlations between the IBDSI-LF and the clinician and self-administered measures. The IBDSI-LF correlated highly with the clinician-administered HBI and PTI ($r = 0.86$, $r = 0.85$) and correlated well with the clinician-rated global assessment of disease status ($r = 0.70$). The lower correlations with the gastroenterologist's assessment are not that surprising given the more limited range of what is evaluated by the specialist on that measure relative to the clinical indices (the HBI and PTI). The IBDSI-LF was also strongly associated with more frequent IBD symptoms scores on the MIBDI and the bowel symptoms subscale of the IBDQ (IBDQ-BS). The report of more symptoms on the IBDSI-LF was also generally associated with self-reports of higher stress and lower quality of life, as evidenced by the significant correlations with the CPSS and IBDQ overall quality of life measure across disease groups and across the various IBDSI-LF subscales. The CPSS assesses experiences that are influenced by symptoms but are quite different, so the lower correlation with the CPSS relative to other GI symptoms measures is consistent with what would be predicted, thus providing some evidence of the IBDSI-LF's discriminant validity.

The 3 main subscales of the IBDSI-SF also showed good evidence of convergent validity within both diagnostic groups. Summary subscale scores assessing bowel symptoms, abdominal discomfort, and fatigue showed strong associations with the Gastroenterologist's global assessment and clinician-rated disease status in CD ($r = 0.60$, $r = 0.57$, $r = 0.58$ and $r = 0.50$, $r = 0.42$, $r = 0.51$, respectively) and in UC ($r = 0.82$, $r = 0.59$, $r = 0.58$ and $r = 0.73$, $r = 0.46$, $r = 0.41$, respectively). The self-reported IBDQ-BS subscale and MIBDI likewise showed similarly high correlations with the IBDSI-SF in both diagnostic groups ($r = -0.86$ and $r = 0.66$, respectively in CD, and $r = -0.92$ and $r = 0.68$, respectively in UC). Regardless of diagnostic

TABLE 3. Subscales (Factor Structure) of the Inflammatory Bowel Disease Symptom Inventory Long Form

Subscale	Item	Item retained in short form (item # in short form)	Item retained in short form (item # in short form)							
			<i>rcorr</i>	<i>r</i>	I	II	III	IV	V	<i>E</i>
I. Bowel Symptoms			0.65	0.84		0.58	0.58	0.34	0.30	
	3. On average, the number of bowel movements I had each day was...	Yes (i3)	0.53		0.63					0.77
	4. The number of loose/liquid bowel movements or diarrhea I had most days was...	Yes (i4)	0.62		0.71					0.71
	5. My stool consistency was generally... (administered but not scored in long form)	Yes (i5) (scored in short form only)								
	6. I noticed blood in my stool	Yes (i6)	0.38		0.43					0.90
	13. Losing control of bowel movements	Yes (i14)	0.63		0.81					0.59
	15. Waking because of urge to have bowel movements	Yes (i16)	0.61		0.68					0.73
	17. Urgency of bowel movements	Yes (i18)	0.68		0.78					0.63
	20. Urge for bowel movement despite empty bowels	No	0.65		0.74					0.68
	22. Soiling underwear	Yes (i21)	0.58		0.75					0.66
	26. Fever over 37.8°C or 100°F	No	0.54		0.55					0.84
II. Abdominal Discomfort			0.72	0.87			0.72	0.29	0.40	
	2. I had abdominal pain	Yes (i2)	0.70			0.88				0.48
	7. I had loss of appetite	Yes (i7)	0.63			0.66				0.75
	8. I had nausea	Yes (i8)	0.44			0.50				0.87
	9. I had vomiting	Yes (i9)	—			—				—
	10. I had tenderness in my abdomen when touched	Yes (i10)	0.68			0.83				0.55
	12. Excessive bowel gas (farting)	Yes (i13)	0.40			0.35				0.94
	14. Abdominal bloating	Yes (i15)	0.63			0.64				0.77
	16. Waking because of abdominal pain	Yes (i17)	0.64			0.74				0.67
	19. Difficulty releasing gas	Yes (i20)	0.60			0.54				0.84
	23. Abdominal cramps	Yes (i22)	0.73			0.90				0.43
	27. In the past week, I had a lump in my abdomen	Yes (i11b)	0.28			0.37				0.93
III. Fatigue			0.73	0.86				0.28	0.41	
	1. My health was...	Yes (i1)	0.73				0.70			0.72
	11. Feeling fatigued or tired and worn out	Yes (i12)	0.64				0.90			0.44
	18. Finding it hard to get things done	Yes (i19)	0.71				0.82			0.58
	21. Feeling generally unwell	No	0.75				0.75			0.66
	24. Having low energy	No	0.66				0.92			0.39
	25. Having trouble maintaining or gaining weight	No	0.44				0.44			0.90
IV. Bowel Complications			0.31	0.47						0.33
	33. Anal fissure (a tear around the anus)	No	0.34					0.49		0.87
	34. Abscess (a localized collection of pus surrounded by inflamed or infected tissue—inside the body or on the skin)	No	0.29					0.77		0.64
	35a. Do you have a fistula? (a fistula is any internal or external path from the bowel to other organs or to the outside skin)... 35c. Has the fistula been active (sore, swollen, or draining) during the past week?	Yes (scored with Bowel Symptoms) (i24a,b)	0.18					0.59		0.81
	35d. Is this a new fistula you have developed in the past week? (not scored)	No								

TABLE 3. continued

Subscale	Item	Item retained in short form (item # in short form)	<i>r</i> _{corr}	<i>r</i>	I	II	III	IV	V	<i>E</i>
V. Systemic Complications			0.43	0.54						
	28. Joint pain (arthralgia/arthritis)	Yes (scored with Abdominal Discomfort factor) (i23)	0.43						0.57	0.82
	29. Inflammation of the coloured part of the eyes (uveitis/iritis)	No	0.23						0.44	0.90
	30. Tender red lumps on shins or arms (erythema nodosum)	No	0.24						0.44	0.90
	31. Small painful sores or ulcerations on the skin (aphthous ulcers/aphthous stomatitis)	No	0.26						0.37	0.93
	32. Painful sores or ulcerations on the skin (pyoderma gangrenosum)	No	0.19						0.47	0.88

Note. N = 235. *r*_{corr}, corrected (jack-knife) item-total correlations or factor correlations with the total score. All factor loadings from the confirmatory factor analysis of the IBDSI-LF are significant at *P* < 0.05. *r*, zero-order Pearson correlations; subscale intercorrelations are also zero-order Pearson *r*'s. All correlations are significant at *P* < 0.001. *E*, standardized error variance (unique item variance not explained by common factors).

TABLE 4. Descriptive Statistics and Measurement Properties of the IBDSI-LF

IBDSI-LF Scale	No. Items Scored	α	Total IBD Sample			F, comparing CD vs. UC	Test retest reliability (over 6 months)
			CD		UC		n = 144 (Sample 2)
			N = 264	n = 140	n = 124		
<i>Total Score</i>	34 ^a	0.93	Mean (SD), 0–88.0	Mean (SD), 0–83.0	Mean (SD), 0–88.0	3.22 <i>P</i> < 0.10	<i>r</i> = 0.63 <i>P</i> < 0.001
<i>I. Bowel Symptoms</i>	9 ^a	0.88	0.6 (0.7) 0–3.4	0.6 (0.6) 0–3.0	0.6 (0.8) 0–3.4	0.09 <i>P</i> > 0.05	<i>r</i> = 0.67 <i>P</i> < 0.001
<i>II. Abdominal Discomfort</i>	11	0.87	0.6 (0.6) 0–2.7	0.6 (0.6) 0–2.6	0.5 (0.5) 0–2.7	6.84 <i>P</i> < 0.01	<i>r</i> = 0.49 <i>P</i> < 0.001
<i>III. Fatigue</i>	6	0.88	1.0 (0.9) 0–3.7	1.1 (0.9) 0–3.5	0.9 (0.8) 0–3.7	2.09 <i>P</i> > 0.05	<i>r</i> = 0.61 <i>P</i> < 0.001
<i>IV. Bowel Complications</i>	3	0.55	0.2 (0.5) 0–3.3	0.3 (0.6) 0–3.3	0.1 (0.4) 0–2.7	5.36 <i>P</i> < 0.05	<i>r</i> = 0.77 <i>P</i> < 0.001
<i>V. Systemic Complications</i>	5	0.61	0.3 (0.5) 0–2.8	0.4 (0.5) 0–2.2	0.3 (0.4) 0–2.8	4.00 <i>P</i> < 0.05	<i>r</i> = 0.62 <i>P</i> < 0.001

Note. The total score is calculated as a sum of the coded items. Subscale scores are calculated as the average of items on that scale. ^aItem 5 which reports on stool consistency is not included in the total or subscale score in the long form because of redundancy with item 4 which reports on the number of loose or liquid bowel movements. The question is retained so that a HBI or PTI score can be extracted. *E*, residual error variances (item-specific variability).

TABLE 5. Descriptive Statistics and Measurement Properties of the IBD-SF

IBDSI-SF Scale	# Items	α	Total IBD Sample 1		CD Sample 1		UC Sample 1		F, CD vs UC	r with I	r with II	r with III	Test-retest reliability (over 2 weeks)	Test-retest reliability (over 2 months)
			Mean (SD), Min-Max	N = 264	Mean (SD), Min-Max	n = 140	Mean (SD), Min-Max	n = 124						
Total Score	24	0.92	15.3 (13.3) 0-67.0	16.6 (12.5) 0-60.0	13.8 (14.0) 0-67.0	2.98 P < 0.10	0.81 P < 0.001	0.88 P < 0.001	0.75 P < 0.001	0.83 P < 0.001	0.74 P < 0.001	r = 0.83 P < 0.001	r = 0.74 P < 0.001	
I. Bowel Symptoms ^a	9	0.85	0.7 (0.7) 0-3.6	0.7 (0.7) 0-2.8	0.7 (0.8) 0-3.6	0.18 P > 0.10	—	—	—	0.88 P < 0.001	0.78 P < 0.001	r = 0.88 P < 0.001	r = 0.78 P < 0.001	
II. Abdominal & Bodily Discomfort ^b	12	0.87	0.6 (0.6) 0-2.8	0.6 (0.6) 0-2.7	0.5 (0.5) 0-2.8	7.10 P < 0.01	.46 P < 0.001	—	—	0.79 P < 0.001	0.71 P < 0.001	r = 0.79 P < 0.001	r = 0.71 P < 0.001	
III. Fatigue	3	0.82	1.1 (0.9) 0-3.7	1.2 (0.9) 0-3.3	1.0 (0.9) 0-3.7	1.65 P > 0.05	.44 P < 0.001	.62 P < 0.001	—	0.71 P < 0.001	0.56 P < 0.001	r = 0.71 P < 0.001	r = 0.56 P < 0.001	

Note. The total score is calculated as a sum of the coded items. Subscale scores are calculated as the average of items on that scale. ^aIncludes item 35 reporting on fistula activity, which in the long form is part of the Bowel Complications subscale. ^bIncludes item 28 reporting on arthralgia/arthritis, which in the long form is part of the Systemic Complications subscale. r = Pearson correlations.

TABLE 6. IBDSI-LF Total Score and Subscale Correlations With Other Patient- and Clinician-Reported Measures (N ≤ 264)

	IBDSI					
	Total	I	II	III	IV	V
MIBDI	0.67 ^d	0.61 ^d	0.59 ^d	0.57 ^d	0.27 ^d	0.26 ^d
HBI ^a	0.86 ^d	0.79 ^d	0.68 ^d	0.70 ^d	0.35 ^d	0.48 ^d
PTI ^b	0.85 ^d	0.78 ^d	0.77 ^d	0.68 ^d	0.26 ^c	0.37 ^d
IBDQ	-0.89 ^d	-0.73 ^d	-0.80 ^d	-0.83 ^d	-0.27 ^d	-0.43 ^d
IBDQ-BS	-0.89 ^d	-0.80 ^d	-0.83 ^d	-0.70 ^d	-0.31 ^d	-0.42 ^d
GGA	0.70 ^d	0.72 ^d	0.57 ^d	0.58 ^d	0.28 ^d	0.21 ^c
CPSS	0.50 ^d	0.35 ^d	0.45 ^d	0.50 ^d	0.20 ^c	0.27 ^d

Note. All correlations significant at P < 0.001. IBDSI-LF, Inflammatory Bowel Disease Symptom Inventory Long Form; Total, IBDSI total score; Factor I, Bowel Symptoms; Factor II, Abdominal Discomfort; Factor III, Fatigue; Factor IV, Bowel Complications; Factor V, Systemic Complications; MIBDI, Manitoba IBD Index, higher scores indicate more frequent symptoms; HBI, Harvey Bradshaw Clinical Index for Crohn's disease, higher scores on the HBI indicate more frequent/severe symptoms; PTI, Powell-Tuck Clinical Index for ulcerative colitis, higher scores on the PTI indicate more frequent/severe symptoms; IBDQ, Inflammatory Bowel Disease Questionnaire, higher scores indicate higher quality of life; IBDQ-BS, Inflammatory Bowel Disease Questionnaire Bowel Symptoms subscale, higher scores indicate less difficulty with IBD symptoms; GGA, Gastroenterologist's Global Assessment of Disease Severity; CPSS, Cohen's Perceived Stress Scale-4 item short form. ^an = 128; ^bn = 118. ^cP < 0.01; ^dP < 0.001.

group, correlations with the 2 IBD complication subscales of the IBDSI-SF would be expected to be lower because there are fewer items sampling the complication factors and fewer participants endorsing these symptoms; 69.7% and 81.6% of participants with CD and UC, respectively, reported no bowel complications, and 40.1% and 45.6% of participants with CD and UC, respectively, reported no systemic complications. Nonetheless, these items and subscales were retained as they assess clinically important symptoms.

Concurrent validity

Analyses of variance comparing the IBDSI-LF total scores between groups with active vs inactive symptoms based on Gastroenterologist's global assessment, HBI or PTI cutoffs, or against self-reported symptom frequencies (MIBDI) are presented in Table 7. The IBDSI-LF showed significant step-wise differences in symptomatic disease activity with increasing disease severity on all measures.

Known groups validation

Scores on the IBDSI-LF and IBDSI-SF were compared across CD and UC diagnostic groups (see Tables 4 and 5). Participants with CD, which often runs a more severe course due to more widespread inflammation, showed significantly higher scores on the IBDSI Abdominal Discomfort subscale in both the long and short form and reported more IBD-related Bowel Complications and Systemic Complications on the long form

TABLE 7. Analysis of Variance Comparing IBDSI-LF Scores for those With Active or Inactive Symptoms for the Full IBD Sample and Within Each Disease Type

Measure	Level of Active Disease	Total IBD		F, comparing disease activity levels within each measure	CD		UC	
		N ≤ 264			n ≤ 140		n ≤ 122	
		Mean (SD)	95% CI		Mean (SD)	95% CI	Mean (SD)	95% CI
Harvey-Bradshaw Index for CD	Active symptoms (≥5)	—	—	126.63 P < 0.001	37.3 (15.8) 32.5–42.2	—	—	—
	Inactive (<5)	—	—	—	12.4 (9.1) 10.5–14.5	—	—	—
Powell-Tuck Index for UC	Active symptoms (≥5)	—	—	115.14 P < 0.001	—	—	34.6 (18.0) 28.4–40.8	—
	Inactive (<5)	—	—	—	—	—	9.4 (7.8) 7.8–11.1	—
Gastroentero-logist Global Assessment	Severely active	62.0 (22.7) 38.2–85.8	—	82.99 P < 0.001	83.0 _a	—	57.8 (22.6) 29.7–85.9	—
	Moderately active	38.5 (15.7) 33.3–43.7	—	—	41.7 (13.1) 36.0–47.4	—	33.6 (18.4) 23.4–43.8	—
	Mildly active	26.2 (13.9) 22.6–29.9	—	—	24.4 (11.4) 20.5–28.3	—	29.0 (17.0) 21.7–36.4	—
	Inactive (in remission)	11.1 (10.1) 9.5–12.7	—	—	13.6 (11.9) 10.9–16.3	—	8.6 (7.0) 7.0–10.2	—
Gastroentero-logist Rating of Disease Status	Flaring currently (symptoms worse than usual)	35.4 (19.4) 30.5–40.2	—	109.57 P < 0.001	36.1 (16.7) 30.4–41.7	—	34.5 (22.6) 25.9–43.1	—
	Not flaring currently	13.3 (12.1) 11.5–15.1	—	—	15.5 (13.0) 12.7–18.2	—	11.0 (10.6) 8.7–13.3	—
Manitoba Inflammatory Bowel Disease Index	(1) Constantly active (symptoms every day)	41.2 (23.2) 32.7–49.7	—	43.24 P < 0.001	39.2 (21.9) 27.1–51.3	—	43.0 (24.9) 29.7–56.3	—
	(2) Often active (symptoms most days)	31.4 (14.1) 26.8–35.2	—	—	34.7 (13.8) 29.1–40.2	—	26.8 (13.5) 20.3–33.3	—
	(3) Sometimes active (symptoms on some days, for instance 1–2 days/week)	22.0 (13.8) 17.7–26.4	—	—	23.6 (13.7) 17.6–29.7	—	20.2 (14.1) 13.3–27.0	—
	(4) Occasionally active (symptoms 1–2 days/month)	17.1 (10.2) 13.5–20.8	—	—	17.6 (10.4) 13.0–22.2	—	16.0 (10.2) 8.7–23.3	—
	(5) Rarely active (symptoms only a few days of the past 3 months)	13.4 (10.0) 10.4–16.9	—	—	14.4 (10.5) 9.9–18.9	—	11.9 (9.4) 6.7–17.1	—
	(6) I was well in the past 3 months (what I consider a remission or absence of symptoms)	6.2 (5.5) 4.9–7.5	—	—	7.6 (5.6) 5.5–9.7	—	5.2 (5.3) 3.5–6.8	—
	≤3	30.6 (18.4) 27.2–34.0	—	124.30 P < 0.001	31.8 (17.1) 27.4–36.1	—	29.3 (19.9) 23.8–34.7	—
	>3	10.8 (9.6) 9.2–12.4	—	—	13.0 (10.2) 10.7–15.3	—	8.3 (8.3) 6.3–10.3	—

_an = 1.

than did participants with UC. Comparisons of IBDSI-LF descriptive statistics of patients with CD or UC reporting active symptoms on other self-report and clinician-administered measures of IBD symptom activity are also presented in [Table 7](#).

Criterion-related validity

Identifying active symptoms with the IBDSI-LF. In CD, the commonly used metric for identifying active disease on the HBI is a score of HBI ≥ 5. This was used to establish a cutoff score on

the IBDSI-LF and IBDSI-SF that could similarly denote active symptoms of disease. Using Receiver Operating Characteristics (ROC) methods, a cutoff score of >24 on the long form IBDSI was shown to have a specificity of 89% and a sensitivity of 84% for identifying active disease on the HBI (area under the curve: 0.93; 95% CI, 0.89–0.98).

In UC, a cutoff of PTI ≥ 5 has likewise shown good specificity and sensitivity in identifying active disease. Receiver operating curve analysis found a cutoff score of >17 on the IBDSI-LF demonstrated 87% specificity and 91% sensitivity for PTI ≥ 5 (area under the curve: 0.95; 95% CI, 0.92–0.99). The area under the curve for the IBDSI-LF plotted against the HBI and PTI is shown in [Supplementary Fig. S1A and S1B](#), respectively.

Identifying active symptoms using the IBDSI-SF. These analyses were repeated with the 26-item short form of the IBDSI. For CD, a cutoff score of >14 on the IBDSI-SF showed a specificity of 81% and a sensitivity of 93% for a score of HBI ≥ 5 (area under the curve: 0.94; 95% CI, 0.89–0.98). Similarly for UC, a cutoff score of >13 on the IBDSI-SF showed a specificity of 88% and a sensitivity of 94% for a PTI score ≥ 5 (area under the curve: 0.97; 95% CI, 0.94–0.99). [Supplementary Figure S2](#) depicts the areas under the curve for the IBDSI-SF.

Thus, both the long form and short form of the IBDSI were able to match the established cutoffs for active disease status on the HBI and PTI with very good specificity and sensitivity.

Comparability of the Modified Self-report HBI/PTI and Clinician-administered Versions

To assess the IBDSI's ability to retain comparability of assessment with existing studies employing the HBI and PTI, we extracted the modified self-report HBI and PTI items and calculated scores on the HBI or PTI for participants with CD or UC, respectively, using the original scoring systems. The modified self-report HBI and PTI had moderate internal consistencies (Cronbach's $\alpha = 0.74, 0.75$) comparable to or better than the clinician-administered versions ($\alpha = 0.49, 0.72$) with which they were highly correlated ($r = 0.86$ for HBI, $r = 0.90$ for PTI; $P < 0.001$). Both versions produced good agreement when categorizing patients with clinically significant (≥ 5) symptoms ($\kappa = 0.68$ for HBI, $\kappa = 0.80$ for PTI; $P < 0.001$).

[Supplementary Table 2](#) provides a flow chart of the development of the IBDSI.

DISCUSSION

A strength of the IBDSI is that it was developed and evaluated following current recommendations for measures of patient-rated outcomes.^{9,13} This patient-report measure showed strong measurement properties including good factor structure and effective sampling of the range of symptom experience in

patients with IBD. A longer (more comprehensive) version was developed, along with a shorter version. The goal in developing the shorter version was to have a measure suitable for repeated assessments in the context of research or clinical work after a more comprehensive assessment has been done. The IBDSI-LF has 38 items, and the IBDSI-SF has 26 items. The short form eliminates symptoms that are clinically significant but rare and thus are less important to assess in repeated assessments.

Many existing clinician-administered symptom measures were developed specifically for CD or for UC, with some overlapping symptoms but nonetheless some differences in item content and in the response metrics. It has therefore been difficult for researchers and clinicians alike to develop consistent criteria for symptom activity and to make comparisons in the experience of symptoms between persons with UC and CD. To this end, the IBDSI offers a significant advantage as this measure may be applied in both CD and UC, with parallel metrics to allow direct IBD subtype comparison. We recognize that including an item like fistulas does not apply to persons with UC; however, it is an important complication for persons with CD and, hence, is the only variable queried that would not apply across both diseases. However, we have established different cutoffs for disease activity for each disease.

We were also cognizant of the significant advantages of having the same measure used consistently across studies and over time, which allows for comparisons of symptom severity and degree of change in symptoms over time, in both descriptive/epidemiological studies and in clinical trials. The findings of study 1 indicate that the new measure has strong construct validity including high correlations with the measures that were used in its development and other measures of patient symptoms. There is good internal consistency for the overall measure and the 3 main subscales. Study 2 indicates acceptable levels of test-retest reliability for the short form overall measure and the 3 main subscales. Both the long and short forms provide detailed descriptions of symptoms that are of concern to patients but take only about 5 or 3 minutes, respectively, to complete. The 3 main subscales (eg, bowel symptoms, abdominal discomfort, and fatigue) allow for the evaluation of the influence of treatment and changes in the disease course on each symptom factor. This has not been possible with most previous measures used in the area.^{5,7}

The Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) program of the International Organization for the Study of Inflammatory Bowel Diseases²⁴ has used a consensus process to agree upon 12 recommendations for UC and CD. The target for UC was “clinical/patient-reported outcome (PRO) remission (defined as resolution of rectal bleeding and diarrhea/altered bowel habit) and endoscopic remission (defined as a Mayo endoscopic subscore of 0–1). Histological remission was considered as an adjunctive goal. Clinical/PRO remission was also agreed upon as a target for CD and defined as resolution of abdominal pain and diarrhea/altered bowel habit; and

endoscopic remission, defined as resolution of ulceration at ileocolonoscopy, or resolution of findings of inflammation on cross-sectional imaging in patients who cannot be adequately assessed with ileocolonoscopy. Biomarker remission (normal C-reactive protein (CRP) and calprotectin) was considered as an adjunctive target.” A recent review documents the work to put the STRIDE targets into practice and challenges in implementing the standards.²⁵ The new IBDSI would be an excellent measure to establish patient-reported remission in both CD and UC.

Given the modest association at times of inflammation and IBD symptoms, another advantage of the IBDSI is that it allows an assessment of symptoms over time that may not respond to changes in treatment of inflammation.^{25–27} Often an exacerbation in symptoms is attributed by patients to stress, anxiety, depression,¹³ or dietary factors.^{14,15} Further study in these areas may assist in the development of approaches to managing IBD symptoms. Patients may also require assistance in managing bowel symptoms, abdominal discomfort, and fatigue. As an example, a recent study of fatigue in an IBD cohort found that fatigue increases steadily over time in a population-based community sample.²⁷ More research will be needed to assess the value of this measure in hypothesis testing (such as considering the impact of new treatments on symptoms) and responsiveness to change in symptoms.

Limitations and Directions for Future Research

This study has limitations which should be considered. Due to the cross-sectional nature of study 1, we were not able to assess the test-retest reliability of the full-scale measure over periods shorter than 6 months nor the responsiveness of the measure to changes in symptom status, as recommended in the COSMIN standards.¹⁰ There is a high degree of overlap between the content of the long form and the short form, so we expect that test-retest reliability will also be high when it is assessed for the longer form. The observed pattern of correlations across the 2 versions thus far does suggest that the IBDSI-LF and IBDSI-SF are both stable measurement tools responsive to changes in patient symptoms. That is, test-retest correlations were very good at 2 weeks and 2 months (on the short form) and moderate at 6 months (on the long form). We plan to further evaluate the sensitivity of the measures to change by evaluating changes in scores related to significant changes in clinical status of patients monitored closely over time in a recent cohort study.

It would be optimal for clinical indices in IBD to be validated against endoscopy findings. While it is a limitation that this study did not involve endoscopic assessment, the IBDSI nonetheless was found to demonstrate a high degree of sensitivity and specificity for active disease cutoffs on existing clinician-administered measures that were established against endoscopic disease activity markers. It will nonetheless be helpful in

future research to evaluate the relationship between this patient-reported outcome measure, endoscopy findings, and measures of inflammation. That said, most clinical indices do not line up well with endoscopic, radiologic, or other biological indices of disease activity. Further, we have previously shown the poor correlation between symptoms and active inflammation.¹²

Due to the nature of the study samples, we were not able to assess the cross-cultural validity of the scale or the validity in new onset, younger patients. Future research will need to investigate the measurement properties of the IBDSI in a broader range of IBD patients. Our plan is to develop French and Spanish language versions of the measures and to encourage other translation efforts to facilitate its use in other cultural and language groups.

In future work on construct validity, it would be helpful to evaluate the correlation between this scale and other PRO measures such as the PROMIS Project gastrointestinal symptom scales.²⁶ It will be informative to explore the utility of the different symptom clusters (eg, bowel symptoms, abdominal discomfort, and fatigue) in predicting various clinical outcomes, such as inflammation, IBD-related complications, seeking medical assistance for symptom exacerbations, treatment responsiveness, patient well-being, perceived stress, and symptoms of anxiety and depression. It will also be informative to track the naturalistic variability in these symptoms and their trajectories over time relative to other disease processes and outcomes. Existing clinician-administered measures of the symptom experience have focused heavily on the bowel symptoms associated with IBD; yet preliminary correlations with the IBDSI subscales and quality of life measures suggest it may be abdominal discomfort and fatigue that drive the patient experience. Inclusion of the IBDSI in clinical trials or naturalistic studies may thus expand our understanding of the impact and patient experience of IBD.

Summary

Our study findings suggest that the IBDSI is a reliable and valid measure of the fuller patient experience of IBD symptoms than has previously been available. It has the advantage of using language that is more accessible to the general (nonmedical) population than existing clinician-administered tools and still captures the essential clinical information at least as well as current widely used clinician-administered clinical indices such as the HBI and PTI. It allows for direct comparisons to these measures in previously published studies or past clinical assessments. It has strong measurement properties, having used sound scale development methodologies. Overall, the IBDSI is a promising new tool for the assessment of patient-reported symptom experiences in IBD—in clinical and research settings alike. Furthermore, a short form, comparable in reliability and validity, allows for efficient tracking of these symptoms over time. Use of this measure could allow clinicians to score and follow symptom reports as a means to quantify the impact of interventions.

SUPPLEMENTARY DATA

Supplementary data is available at *Inflammatory Bowel Diseases* online.

ACKNOWLEDGMENTS

The authors would like to thank Linda Rogala and Meloney Ivekovic for their assistance with this study. Study 2 was supported by funding from the Canadian Institutes of Health Research.

REFERENCES

- Best WR, Bechtel JM, Singleton JW, et al. Development of a Crohn's disease activity index. National cooperative Crohn's disease study. *Gastroenterology*. 1976;70:439–444.
- Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet*. 1980;1:514.
- Walmsley RS, Ayres RC, Pounder RE, et al. A simple clinical colitis activity index. *Gut*. 1998;43:29–32.
- Powell-Tuck J, Bown RL, Lennard-Jones JE. A comparison of oral prednisolone given as single or multiple daily doses for active proctocolitis. *Scand J Gastroenterol*. 1978;13:833–837.
- Walsh AJ, Bryant RV, Travis SP. Current best practice for disease activity assessment in IBD. *Nat Rev Gastroenterol Hepatol*. 2016;13:567–579.
- Alrubaiy L, Rikaby I, Sageer M, et al. Systematic review of the clinical disease severity indices for inflammatory bowel disease. *Inflamm Bowel Dis*. 2015;21:2460–2466.
- de Jong MJ, Huibregtse R, Masclee AAM, et al. Patient-reported outcome measures for use in clinical trials and clinical practice in inflammatory bowel diseases: a systematic review. *Clin Gastroenterol Hepatol*. 2018;16:648–663.e3.
- Schoepfer AM, Beglinger C, Straumann A, et al. Fecal calprotectin correlates more closely with the simple endoscopic score for Crohn's disease (SES-CD) than CRP, blood leukocytes, and the CDAI. *Am J Gastroenterol*. 2010;105:162–169.
- af Björkstén CG, Nieminen U, Turunen U, et al. Surrogate markers and clinical indices, alone or combined, as indicators for endoscopic remission in anti-TNF-treated luminal Crohn's disease. *Scand J Gastroenterol*. 2012;47:528–537.
- Falvey JD, Hoskin T, Meijer B, et al. Disease activity assessment in IBD: clinical indices and biomarkers fail to predict endoscopic remission. *Inflamm Bowel Dis*. 2015;21:824–831.
- Boon GJ, Day AS, Mulder CJ, et al. Are faecal markers good indicators of mucosal healing in inflammatory bowel disease? *World J Gastroenterol*. 2015;21:11469–11480.
- Targownik LE, Sexton KA, Bernstein MT, et al. The relationship among perceived stress, symptoms, and inflammation in persons with inflammatory bowel disease. *Am J Gastroenterol*. 2015;110:1001–1012; quiz 1013.
- Sexton KA, Walker JR, Graff LA, et al. Evidence of bidirectional associations between perceived stress and symptom activity: a prospective longitudinal investigation in inflammatory bowel disease. *Inflamm Bowel Dis*. 2017;23:473–483.
- Kakodkar S, Mutlu EA. Diet as a therapeutic option for adult inflammatory bowel disease. *Gastroenterol Clin North Am*. 2017;46:745–767.
- Hou JK, Lee D, Lewis J. Diet and inflammatory bowel disease: review of patient-targeted recommendations. *Clin Gastroenterol Hepatol*. 2014;12:1592–1600.
- Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737–745.
- Clara I, Lix LM, Walker JR, et al. The Manitoba IBD index: evidence for a new and simple indicator of IBD activity. *Am J Gastroenterol*. 2009;104:1754–1763.
- Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology*. 1989;96:804–810.
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983;24:385–396.
- Smith GT, McCarthy DM, Anderson KG. On the sins of short-form development. *Psychol Assess*. 2000;12:102–111.
- Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international delphi study. *Qual Life Res*. 2010;19:539–549.
- Ruopp MD, Perkins NJ, Whitcomb BW, et al. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J*. 2008;50:419–430.
- Hu L, Bentler P. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol Meth*. 1998;3:424–453.
- Peyrin-Biroulet L, Sandborn W, Sands BE, et al. Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. *Am J Gastroenterol*. 2015;110:1324–1338.
- Darr U, Khan N. Treat to target in inflammatory bowel disease: an updated review of literature. *Curr Treat Options Gastroenterol*. 2017;15:116–125.
- Graff LA, Clara I, Walker JR, et al. Changes in fatigue over 2 years are associated with activity of inflammatory bowel disease and psychological factors. *Clin Gastroenterol Hepatol*. 2013;11:1140–1146.
- Williet N, Sandborn WJ, Peyrin-Biroulet L. Patient-reported outcomes as primary end points in clinical trials of inflammatory bowel disease. *Clin Gastroenterol Hepatol*. 2014;12:1246–1256.e6.
- Spiegel BM, Hays RD, Bolus R, et al. Development of the NIH patient-reported outcomes measurement information system (PROMIS) gastrointestinal symptom scales. *Am J Gastroenterol*. 2014;109:1804–1814.