Adv Ther (2024) 41:3678–3705
https://doi.org/10.1007/s12325-024-02923-9

ORIGINAL RESEARCH

Content Validity and Psychometric Evaluation of the Crohn's Symptom Severity (CSS) Questionnaire in Patients with Moderately to Severely Active Crohn's Disease

Edouard Louis · Wan-Ju Lee · Leighann Litcher-Kelly · Sarah Ollis · Emma Pranschke · Kristina Fitzgerald · Ana Paula Lacerda · Ezequiel Neimark · Yuri Sanchez Gonzalez · Julian Panés

Received: March 12, 2024 / Accepted: June 6, 2024 / Published online: August 6, 2024

ABSTRACT

***Introduction*:** Individuals living with Crohn's disease (CD) experience burdensome symptoms. As such, it is important to measure CD symptom severity in clinical research. The goal of this study was to evaluate the content validity, psychometric performance, and score interpretability of a new patient-reported instrument, the Crohn's Symptom Severity (CSS) questionnaire, among adolescents and adults with moderately to severely active CD.

***Methods*:** Cognitive debriefing interviews ($N = 30$; $n = 20$ adults, $n = 10$ adolescents) were conducted to evaluate the content validity of the CSS. Additionally, the CSS scores were evaluated for reliability and validity using data from a phase 3 randomized clinical trial of risankizumab (NCT03105128; $N = 850$). Meaningful within-patient change (MWPC) thresholds were estimated using anchor-based methods.

***Results*:** All interview participants ($n = 30/30$, 100.00%) reported the CSS was easy to complete and most participants ($n = 28/29$, 96.55%) reported that the CSS was relevant to their experience of CD. Among the clinical trial subjects ($N = 850$) the following was found for the CSS: mostly acceptable item–total correlations (0.26–0.79); weak to moderate inter-item correlations ($r = 0.07$–0.57), good internal consistency (Cronbach's $\alpha = 0.76$–0.87); intraclass correlation coefficients ranged from 0.48 to 0.70, not consistently exceeding the acceptable range for test–retest reliability (0.70); acceptable convergent validity and known-groups results; and demonstrated sensitivity to change. Analyses supported an MWPC estimate of 6–11 points.

***Conclusions*:** This study supports use of the CSS for measuring CD symptoms and sleep impact among adolescents and adults aged 16 and older with moderately to severely active CD in clinical research.

***Trial Registration*:** NCT03105128 (registration date 4 April 2017).

Keywords: Crohn's disease; Symptoms; Qualitative interviews; Psychometrics; Score interpretation

E. Louis (✉)
University of Liège, Liège, Belgium
e-mail: Edouard.Louis@ulg.ac.be

W.-J. Lee · K. Fitzgerald · A. P. Lacerda · E. Neimark · Y. Sanchez Gonzalez
AbbVie Inc, Chicago, IL, USA

L. Litcher-Kelly · S. Ollis · E. Pranschke
Adelphi Values, Boston, MA, USA

J. Panés
Hospital Clínic de Barcelona, Barcelona, Spain

Key Summary Points

Thirty qualitative interviews with adolescents and adults (ages 15–75 years old) with moderately to severely active Crohn's disease (CD) confirmed that the Crohn's Symptom Severity (CSS) questionnaire is easily interpreted and that the concepts measured are relevant and comprehensive of patients' experience with CD symptoms.

Results of the psychometric evaluation of the CSS, which used data from a phase 3 randomized clinical trial, suggest that the scores produced are construct valid and sensitive to change in a clinical trial setting.

This research provides evidence that in a clinical trial setting a score change between 6 and 11 points of improvement on the CSS may represent meaningful within-patient change (MWPC), which aligned with improvements in clinical remission/response and endoscopic remission/response, as well as improvements in patient-reported quality of life.

INTRODUCTION

Crohn's disease (CD) is a chronic inflammatory bowel disease (IBD) characterized by relapsing and remitting transmural inflammation in any portion of the gastrointestinal tract [1]. As of 2017, it was estimated that there were between six and eight million cases of IBD globally [2, 3]. CD symptoms can be burdensome and include diarrhea, abdominal pain, fatigue, malaise, anemia, weight loss, and rectal bleeding and may lead to other impacts such as sleep disruptions [4–7].

The Crohn's Symptom Severity (CSS) questionnaire was originally developed by AbbVie for use in clinical trials for the purpose of assessing CD-related symptoms and sleep impact for adults with mild to severe CD. The questionnaire was developed to be a disease-specific qualitative assessment of a wider range of symptoms specific to CD, compared to the Crohn's Disease Activity Index (CDAI) [8] or the Harvey–Bradshaw Index (HBI) [9], which both measure patient-reported stool frequency and abdominal pain score over the past 7 days and 24 h, respectively; and the Inflammatory Bowel Disease Questionnaire (IBDQ) [10], which assesses the impacts of a very broad range of CD or ulcerative colitis (UC) symptoms on quality of life over the past 2 weeks. The CSS was developed to determine the most frequent and relevant CD signs and symptoms with impact on patients' quality of life.

The goals of the research described in this paper were to describe results from both (1) a qualitative interviews study with adolescent and adult patients with moderately to severely active CD to evaluate the content validity (i.e., readability, relevance, comprehensibility, and comprehensiveness) of the CSS and (2) a quantitative analyses of phase 3 randomized clinical trial data involving patients with moderately to severely active CD aged 16–80 years (NCT03105128; hereafter identified as ADVANCE) to evaluate the psychometric performance of the CSS and establish score interpretation guideline (MWPC) estimates for the CSS.

METHODS

Instrument

The CSS questionnaire, originally called the Crohn's Disease Symptom Questionnaire (CDSQ), was developed in line with best practices described in the US Food and Drug Administration (FDA) patient-reported outcome guidance document [11]. This included (1) a targeted literature review (N=22 articles), (2) interviews with clinical experts in gastroenterology (N=4), and (3) focus groups (N=4 focus groups; 20 participants total) and individual interviews (n=2) with adults with mild to severe CD in the USA. After qualitative data from clinical expert and patient interviews were collected and analyzed, the most relevant symptoms and impacts (i.e., frequently reported and appropriate for patient-based measurement) were selected for inclusion in the draft questionnaire. Experts (N=3) confirmed the relevance and clarity of the

draft CSS, and their feedback on specific items was considered along with results from the cognitive debriefing interviews with adult patients with mild to severe CD in the USA (N=15) to make revisions and finalize the questionnaire.

The final CSS includes 14 items that measure CD-related symptoms (13 items) and sleep (1 item), with a recall period of "the past week." Specifically, item concepts assessed by the CSS are bowel movements, passing gas, abdominal pain, nausea, loss of appetite, bloating, diarrhea, blood in stool, constipation, vomiting, stomach growling, feeling tired/lacking energy, joint pain, and difficulty sleeping. All items utilize a five-point verbal rating scale (VRS) ranging from 1 ("not at all") to 5 ("very much") for items 1–8 and 1 ("never") to 5 ("always") for items 9–14. A total score for the CSS is calculated by summing the individual item scores. CSS scores can range from 14 to 70, where higher scores indicate greater disease symptoms and impact (i.e., worse health-related quality of life).

Evaluation of Content Validity: Qualitative Cognitive Debriefing Interviews

Open-ended, semi-structured cognitive debriefing interviews were conducted with patients with moderately to severely active CD to evaluate the comprehensibility, relevance, and comprehensiveness of the CSS for use in clinical research. The study protocol and all study documents were approved by a centralized independent review board.

Participants

Adolescents (ages 15–17) and adults (ages 18–80) were recruited from four clinical sites in the USA. Participants were to be fluent in US English and have a clinician-confirmed diagnosis of moderately to severely active CD to be eligible for the study. All inclusion and exclusion criteria are in Table 1. Health Insurance Portability and Accountability Act (HIPAA) authorization and informed consent or parental permission and assent was obtained for all interview participants.

Conduct of Qualitative Interviews

Qualitative interviews were conducted by trained interviewers in person or by telephone following a semi-structured interview guide. Participants completed the CSS on an electronic device or by using PDF screenshots of the instrument while "thinking aloud" about the process they used to arrive at each answer. This process helped to identify any words, terms, or concepts within the instrument that they did not understand or did not interpret as intended. Following the "think-aloud" exercise, participants were asked additional questions designed to evaluate the content of the CSS, including its comprehensibility, relevance, and comprehensiveness. Interviews were audio-recorded, following participants' verbal consent, and subsequently transcribed and anonymized.

Qualitative Coding and Data Analysis

Each transcript underwent qualitative coding to organize and catalogue participants' feedback and responses on the instructions, items, and response options of the CSS. All transcripts were coded in ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany) and then further analyzed to identify themes or inform any necessary modifications to the questionnaire.

Evaluation of Psychometric Performance and Interpretability of Scores of the CSS: Analysis of Phase 3 Clinical Trial Data

Data from the ADVANCE phase 3 clinical program of risankizumab [12] were used for the psychometric and score interpretation analyses. The study protocol was approved by ethical review committees, and all participants in this research provided informed consent for their participation (or assent and caregiver consent for individuals under the age of 18).

Trial participants completed the CSS using an electronic tablet device at the clinical site unless they were unable or not comfortable to come to the site as a result of the coronavirus disease 2019 (COVID-19) pandemic; in these instances, the questionnaire was administered

Table 1 Qualitative interview inclusion and exclusion criteria

The study inclusion criteria were as follows:
Provided Health Insurance Portability and Accountability Act (HIPAA) authorization, informed consent to participate for adult participants, and parental permission and assent for adolescent participants
At least 15 years of age and ≤ 80 years old
Physician-confirmed diagnosis of moderate to severe Crohn's disease
Fluent in US English (i.e., able to speak, read, write, and comprehend)
Willing and able to participate in one 90-min, face-to-face or telephone interview
The study exclusion criteria were as follows:
Had a clinical diagnosis of fulminant colitis and/or toxic megacolon
Had an ostomy, ileoanal pouch, or symptomatic bowel stricture
Enrolled in an investigational drug study or participated in such a study within 30 days of entry into this study
Had a condition or situation (e.g., comorbid condition, cognitive impairment, substance/alcohol abuse) that, in the opinion of the clinician, would interfere with his/her participant in a 90-min interview and/or impact the data collected *For adolescents only:* Had physical developmental issues or delays that would have indicated potential physical immaturity

via a telephone-based interview. Participants (N=850) completed the CSS during clinic visits at baseline and weeks 4, 8, 12, 16, 20, and 24.

Supplemental Instruments

In addition to the CSS, participants in the ADVANCE program completed the CDAI [8], IBDQ [10], 36-Item Short Form Survey Version 2 (SF-36v2®) [13], five-level version of the EQ-5D (EQ-5D-5L) [14], Patient Global Impression of Change (PGIC), Patient Global Impression of Severity (PGIS), and Work Productivity and Activity Impairment Questionnaire for CD (WPAI:CD) [15–18] during the trials. Scores from these assessments supported the psychometric and score interpretation evaluation of the CSS. These study assessments are described in Table 2.

Analyses

Scores on the CSS were evaluated for item–total and inter-item correlations, reliability (internal consistency using Cronbach's α and test–retest reliability using the intraclass correlation coefficient [ICC]), validity (convergent and discriminant validity, known-groups methods), and sensitivity to change. Item–total correlations were examined by generating correlation coefficients between the individual items of the CSS and the CSS total score at week 12. Inter-item correlations characterizing the extent to which scores on one item in a scale relate to scores produced by all other items within that same scale were analyzed at week 12. Strong correlation was defined as ≥0.70 to ≤0.90, a moderate correlation as ≥0.30 to <0.70, and a weak correlation as <0.30 [19]. Cronbach's α for the CSS total score was calculated for evaluation of internal consistency for which an α ≥0.70 was considered acceptable [20]. Test–retest reliability was assessed among three subsets of participants from the psychometric analysis population. The three subsets were (1) participants who chose the same score on the PGIS at baseline and week 4, (2) participants who chose the same score on the PGIS at week 4 and week 12, and (3) participants who selected "not changed" since before treatment began on the PGIC at week 4 and week 12. ICCs were considered acceptable if they exceeded 0.70 [21, 22].

Analyses were also conducted to establish MWPC estimates for the CSS total score. The CSS total score was anchored to categories defined by PGIS and PGIC change scores between baseline and week 12. A one- to

Table 2 Reference measures for psychometric and score interpretation analyses

Measure	Concepts measured	Recall period	Scoring
Disease activity assessments			
Crohn's Disease Activity Index (CDAI)[a] [8]	Abdominal pain, stool frequency, and general well-being, in addition to presence of other CD conditions, treatment for diarrhea, abdominal mass, hematocrit, and body weight/standard weight	Past 7 days	Summing the weighted individual scores of eight items; the score ranges from 0 to 600[b]
Impacts and quality of life assessments			
Work Productivity and Activity Index WPAI for CD (WPAI:CD)[c] [17, 18]	Impact of disease on absenteeism, presenteeism, productivity loss, and activity impairment	Past 7 days	Six items; scores are expressed as impairment percentages, adjusting for hours worked according to the WPAI score algorithm[b]
Inflammatory Bowel Disease Questionnaire (IBDQ)[c] [10]	Dimensions of symptoms and health-related quality of life: bowel symptoms, systemic symptoms, social function, emotional function	Past 2 weeks	32 items split into four domains; each domain score can be computed with scores ranging from 10 to 70, 5 to 35, 12 to 84, and 5 to 35; total scores range from 32 to 224[d]
36-Item Short Form (SF-36v2®)[c] [13]	Physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health problems	Past 4 weeks	36 items that, when scored, can be aggregated into two summary responses: the Physical and Mental Component summary scores; each item is scored on a 0–100 range; scores represent the percentage of total possible scores; items in the same scale are then averaged together
Work Productivity and Activity Index WPAI for CD (WPAI:CD)[c] [17, 18]	Impact of disease on absenteeism, presenteeism, productivity loss, and activity impairment	Past 7 days	Six items; scores are expressed as impairment percentages, adjusting for hours worked according to the WPAI score algorithm[b]

Table 2 continued

Measure	Concepts measured	Recall period	Scoring
Global assessments			
The Patient Global Impression of Severity (PGIS)[c] [15, 16]	Severity of the overall symptoms due to CD	Past week	One item rated on a seven-point verbal rating scale from "absent" to "very severe"[b]
The Patient Global Impression of Change (PGIC)[c] [15, 16]	Overall change in their CD symptoms	Since before treatment began	One item rated on a seven-point verbal rating scale ranging from "very much improved" to "very much worse"
General health and utilities assessment			
Five-level EQ-5D (EQ-5D-5L)[c] [14]	Mobility, self-care, usual activities, pain/discomfort, anxiety/depression, and self-evaluated health status	Current health (i.e., "today")	EQ-5D descriptive system: Each of the five dimensions includes five levels[b] EQ-5D visual analogue scale (VAS): A vertical VAS ranging from 0 (worst health you can imagine) to 100 (best health you can imagine)

CD Crohn's disease

[a]Composite measure (patient, clinician, laboratory)

[b]Higher scores indicate greater disease severity or greater negative impact

[c]Patient-rated measure

[d]Higher scores indicate better outcomes

two-point improvement on the seven-level PGIS and the responses of "minimally improved" and "much improved" on the seven-level PGIC were used as the main definitions of improvement for the two anchors. The defined anchor groupings for the PGIS and PGIC were supported by patient input from qualitative interviews conducted with adolescents and adults with CD. In addition to the primary anchor-based analyses, the following supportive analyses were conducted to aid in interpretation: empirical cumulative distribution functions (eCDFs), probability density functions (PDFs), and receiver operating characteristic (ROC) curves. Estimates for MWPC were further evaluated by stratifying responders versus non-responders on the CSS by scores on other clinical and patient-reported quality-of-life measures.

All analyses along with timepoints analyzed and rules for interpretation (e.g., thresholds for acceptability) for each analysis have been included in Table 3.

RESULTS

Qualitative Cognitive Debriefing Interviews

Participant Demographics

Thirty individuals with clinician-confirmed moderate to severe CD ($n=20$ adults 18–80 years of age and 10 adolescents 15–17 years of age) participated in the interviews. Participants were recruited between January 2020 and September 2020 from clinical sites in Chicago, Illinois; New Orleans, Louisiana; St. Louis, Missouri; and Los Angeles, California. Participants' ages ranged from 15.1 to 75.4 years (mean = 36.6 years [standard deviation (SD) = 19.2]). Nineteen participants ($n=19/30$, 63.33%) had moderate CD, and 11 participants ($n=11/30$, 36.67%) had severe CD, as determined by their clinicians during interview screening. Additional demographic and health information is provided in Table 4.

Cognitive Debriefing Results

Each instruction, item, and response option was interpreted as intended by at least 76.67% of participants. All participants ($n=30/30$, 100.00%) interpreted the CSS instructions as intended and most participants ($n=28/30$, 93.33%) interpreted the CSS recall period ("the past week") as intended. Specifically, the two participants who were determined to have misinterpreted the recall period provided responses to CSS items in reference to the previous 5–8 days, rather than an exact week (7 days). Each concept (i.e., the sign, symptom, or impact that each item is designed to measure) and response option was interpreted as intended by at least 25 of the 30 participants (≥ 83.33%). Results for each of the CSS items are further summarized in Table 5 regarding participant interpretations and experience with each item concept.

All participants ($n=30/30$, 100.00%) reported the CSS was easy to complete, and most participants ($n=28/29$, 96.55%) reported that the CSS was relevant to their experience of CD overall. Each item concept was experienced by at least 79.17% of participants, either within the 7-day recall period or prior to the recall period. Items 10 (diarrhea), 11 (blood in your stool), and 13 (vomit or throw up) were most frequently reported as the most important questions to ask individuals with CD.

When asked if there were any additional symptoms or impacts that should be included in the instrument, 10 participants suggested 14 additional concepts for inclusion (whether you have had surgery, comorbid conditions, feeling cold, having a fever, presence of mouth ulcers, anal fissures, bone pain, feet swelling, losing one's voice, gynecological complications, headaches, watery eyes, mental state/mental health, and cramping). However, no concept was suggested as missing by more than one participant, and several suggestions are not considered to be appropriate for assessment in a patient-reported symptom questionnaire. These findings suggest that there are not any particularly common or salient CD symptoms missing from the CSS.

Table 3 Summary of psychometric and score interpretation analyses for CSS scores

Analysis	Description	Timepoint(s)
Item–total correlation	Evaluate the extent to which each item within the questionnaire correlates to the domain or total score. An item–total correlation ≥ 0.3 will be considered as acceptable [24]	Week 12
Inter-item correlation	Evaluate how highly items within the CSS correlate to each other to understand conceptual overlap. A strong correlation was defined as ≥ 0.70 to ≤ 0.90, a moderate correlation as ≥ 0.30 to < 0.70, and a weak correlation as < 0.30 [19]	Week 12
Internal consistency reliability	Evaluate the degree to which individual items are measuring the same general concept. Cronbach's α coefficient typically ranges from 0 to 1.0, with higher estimates indicative of stronger internal consistency among items. Internal consistency was considered to be met if $\alpha \geq 0.70$ [20]. Cronbach's α was calculated with each item removed from its respective score to assess the impact	Baseline and week 12
Test–retest reliability	Evaluate the degree to which items produce stable, reliable scores under similar conditions For CD, stability was defined in two ways: participants who selected (1) the same response on the PGIS at baseline and week 4 or baseline and week 12 and (2) "not changed" since before treatment began for CD symptoms on the PGIC at week 4 and week 12 Test–retest reliability ICCs were calculated with a two-way mixed effects model without interaction (ICC[3A,1]). An ICC of 0.70 or greater was used as evidence of acceptable test–retest reliability for a scale [21, 22]	Baseline, week 4, and week 12
Convergent/discriminant validity	Evaluate the degree to which scores produced by CSS correlate with scores produced by other measures that should theoretically be associated with them. A strong correlation defined as ≥ 0.70 to ≤ 0.90, moderate correlation as ≥ 0.30 to < 0.70, and a weak correlation as < 0.30 [19] Concurrent assessments were the CDAI, PGIS, IBDQ, SF-36v2®, EQ-5D-5L, and WPAI:CD	Week 12
Known-groups methods	Evaluate the degree to which scores produced by the CSS are capable of distinguishing among groups hypothesized a priori as being clinically distinct Known groups were defined using the CDAI, IBDQ, and PGIS. The classification values for the CDAI were < 150 for remission versus ≥ 150 for non-remission, and for the IBDQ total score, ≥ 170 for remission versus < 170 for non-remission	Week 12

Table 3 continued

Analysis	Description	Timepoint(s)
Sensitivity to change	Evaluate the degree to which the change of scores produced by the CSS change in concert with the change of scores produced by other concurrent measures (same as those listed in the convergent/discriminant validity row above)	Baseline and week 12
Anchor-based methods	Estimate the MWPC for the CSS total score. Anchors were the PGIS and PGIC. Estimates were supplemented by eCDF curves, PDF curves, and ROC curves	Baseline and week 12
Distribution-based methods	Estimate the clinically important difference between groups for the CSS total score MCID1: 0.5 of an SD [25] at baseline MCID2: SEM SEM = SD at baseline $*$ $\sqrt{1 - \text{reliability}}$[a]	Baseline
Evaluation of MWPC estimates via clinical and quality-of-life measures	Comparison clinical and quality of life outcomes between the groups of CSS responders vs. non-responders (via MWPC estimates) between induction baseline and week 12 CDAI clinical remission < 150 CDAI clinical response (reduction of CDAI ≥ 100 points from baseline) Endoscopic response (decrease in SES-CD > 50% from baseline (or for subjects with isolated ileal disease and a baseline SES-CD of 4, at least a 2-point reduction from baseline), as scored by central reviewer) Endoscopic remission (SES-CD ≤ 4 and at least a 2-point reduction versus baseline and no subscore greater than 1 in any individual variable, as scored by a central reviewer) IBDQ total score ≥ 170 Change of IBDQ ≥ 16 Mean change in EQ-5D VAS Mean change in SF-36 Physical Component Summary and SF-36 Mental Component Summary Mean change in WPAI scores (work time missed, impairment at work, overall work impairment, activity impairment)	Week 12

CD Crohn's disease, *CDAI* Crohn's Disease Activity Index, *CSS* Crohn's Symptom Severity, *eCDF* empirical cumulative distribution function, *EQ-5D-5L* five-level EQ-5D, *IBDQ* Inflammatory Bowel Disease Questionnaire, *ICC* intraclass correlation coefficient, *MCID* minimal clinically important difference, *MWPC* meaningful within-patient change, *PDF* probability distribution function, *PGIC* Patient Global Impression of Change, *PGIS* Patient Global Impression of Severity, *ROC* receiver operating characteristic, *SD* standard deviation, *SEM* standard error of measurement, *SES-CD* Simple Endoscopic Score for Crohn's disease, *SF-36v2*® 36-Item Short Form Survey Version 2, *VAS* visual analogue scale, *WPAI:CD* Work Productivity and Activity Impairment Questionnaire for Crohn's Disease

[a]Cronbach's α at baseline

Psychometric Evaluation Using Phase 3 Clinical Trial Data

A total of 850 patients from the ADVANCE study were included in the psychometric and score interpretation analysis. This sample size is considered sufficient for these analyses [23]. Participants' ages ranged from 16 to 79 years (mean = 37.5; SD = 13.3); 45.88% of the sample was female.

Score Distributions

Quality of completion for the psychometric analysis population was high across the timepoints, with the number of participants with missing data ranging from 14 to 37. In general, respondents used the entire range of the response scale for the CSS items and item and total scores trended toward improvement over time. Items 11 (blood in your stool), 12 (constipated), and 13 (vomit or throw up) demonstrated floor effects at baseline (> 40% of participants endorsing that they "never" experienced the symptom over the 7-day recall period). Refer to Fig. 1 for all score distributions at baseline and week 12.

Item–Total and Inter-item Correlations

The magnitude of the item–total correlations between each item and the total CSS score ranged between 0.26 and 0.79, which were over the thresholds for acceptable item–total correlations (i.e., ≥ 0.3 [24]) with the exception of the relationship between item 12 (constipated) and the total score (r = 0.26).

Inter-item correlations were weak to moderate (r = 0.07–0.57) across items; this can be expected for a multi-symptom questionnaire. The strongest correlation was observed between item 3 (abdominal pain) and item 4 (felt tired or lacking energy) (r = 0.57). Inter-item correlations can be found in Table 6.

Reliability

Overall Cronbach's α for the CSS total score ranged from 0.76 to 0.87 from baseline to week 12, which provides support for acceptable internal consistency of the questionnaire's items. Removal of any individual item did not substantially improve internal consistency reliability.

For test–retest reliability, ICCs did not consistently exceed the acceptable threshold for test–retest reliability. The ICC comparing the CSS total score among patients defined as stable using the PGIS between baseline and week 4 was 0.48 (95% confidence interval [CI] 0.08–0.69). Between week 4 and week 12, ICC = 0.70 (95% CI 0.61–0.76) for PGIS stable patients. Among PGIC stable patients between week 4 and week 12, ICC = 0.58 (95% CI 0.43–0.68).

Validity

All concurrent measures correlated with the CSS in the hypothesized directions, and the strengths of associations between the CSS and most concurrent measures were as anticipated. The CSS total score correlated more strongly with patient-reported symptom-related measures, such as the PGIS, IBDQ bowel symptom and systemic symptom domains, and IBDQ total scores compared to more distal measures (i.e., EQ-5D-5L, SF-36v2®, WPAI:CD) and outcomes that included clinician-reported items (i.e., CDAI), as presented in Table 7.

The CSS additionally exhibited validity according to known-groups analysis, as CSS scores successfully differentiated between groups of clinically distinct patients. CSS total scores demonstrated a 10.09-point difference between groups classified as remission versus non-remission on the CDAI, an 11.95-point difference in remission versus non-remission groups using the IBDQ, and a monotonic decrease by PGIS group, all of which were statistically significant (p < 0.001). Results from the known-groups analysis can be referenced in Fig. 2.

Table 4 Participant- and clinician-reported demographic and health information

Characteristic	Total sample ($N = 30$) *n* (%)[a]
Age[b]	
Range (minimum–maximum)	15.1–75.4
Mean (SD)	36.60 (19.20)
Sex[b]	
Female	16 (53.23%)
Male	14 (46.67%)
Race	
White	22 (73.33%)
Hispanic	6 (20.00%)
Black or African American	2 (6.67%)
Spanish/Hispanic/Latino[c]	
Not Spanish/Hispanic/Latino	24 (80.00%)
Mexican/Mexican American, Chicano	5 (16.67%)
Hispanic, unspecified	1 (3.33%)
Time since diagnosis (in years)[b]	
Range (minimum–maximum)	1–45
Mean (SD)	9.70 (10.21)
Clinician-reported severity of CD (moderate or severe)[b]	
Moderate	19 (63.33%)
Severe	11 (36.67%)
Current medication use[b,d]	
5-ASA (e.g., mesalamine, sulfasalazine)	10 (33.33%)
Advanced therapy (e.g., biologics)	11 (36.67%)
Immunomodulators (e.g., azathioprine or 6-mercaptopurine)	7 (23.33%)
Corticosteroids	13 (43.33%)
Other: Pantoprazole	1 (3.33%)
Other: B12	1 (3.33%)
Education (adults, $n = 20$)[c]	
High school diploma or GED	2 (6.67%)
Some college or associate degree	8 (26.67%)
College or university degree	5 (16.67%)

Table 4 continued

Characteristic	Total sample ($N = 30$) n (%)[a]
Graduate or professional degree	4 (13.33%)
Other: Technical school	1 (3.33%)
Education (adolescents, $n = 10$)[c]	
9th grade	2 (6.67%)
10th grade	1 (3.33%)
11th grade	2 (6.67%)
12th grade	5 (16.67%)
Living situation[c]	
Living with family or friends	27 (90.00%)
Living alone	3 (10.00%)
Annual household income (adults, $n = 20$)[c]	
Under $25,000	1 (3.33%)
$25,000 to $49,999	7 (23.33%)
$50,000 to $74,999	6 (20.00%)
$75,000 to $99,999	3 (10.00%)
$100,000 and above	3 (10.00%)
Work status[c]	
Working full-time	11 (36.67%)
Student	11 (36.67%)
Working part-time	3 (10.00%)
Unemployed	2 (6.67%)
Retired	1 (3.33%)
On disability	1 (3.33%)
On workers' compensation	1 (3.33%)
Other health conditions[c,d]	
None	20 (66.67%)
Depression/anxiety	3 (10.00%)
Diabetes: type II	2 (6.67%)
High blood pressure	2 (6.67%)
Rheumatoid arthritis	2 (6.67%)
Arthritis	1 (3.33%)

Table 4 continued

Characteristic	Total sample ($N=30$) *n* (%)[a]
Cancer: lung cancer	1 (3.33%)
Fibromyalgia	1 (3.33%)
High cholesterol	1 (3.33%)
Ovarian cysts	1 (3.33%)
Thyroid disease	1 (3.33%)
Uterine fibroids	1 (3.33%)
Wolff–Parkinson–White syndrome	1 (3.33%)

5-ASA 5-aminosalicylic acid, *CD* Crohn's disease, *SD* standard deviation

[a]Unless otherwise indicated

[b]Clinician-reported information

[c]Participant-reported information

[d]Not mutually exclusive

Sensitivity to Change

The magnitude of correlations was observed to be moderate to strong between the CSS change score and change scores on most concurrent measures ($r=0.35$–0.70). Correlations between change in scores for the CSS and more conceptually similar measures (e.g., the SF-36v2® Physical Component Summary) were strongest. The EQ-5D-5L self-care and mobility domains, which were more conceptually dissimilar to the CSS overall measurement constructs (compared to the other concurrent measures), correlated weakly with the CSS change score ($r=0.20$ and $r=0.25$, respectively), which was anticipated.

Additionally, change scores of most items on the CSS had moderate correlations ($r\geq0.50$ for most items) with the changes in the CSS total score from baseline to week 12, except for item 12 (constipated; $r=0.21$). This was likely due to the limited endorsement of item 12 by trial participants (i.e., at baseline, 69.06% of participants indicated they "never" experienced constipation during the recall period). Results suggest that changes in individual CSS items contribute proportionally to the overall change score. In other words, there is no major concern that certain items are solely responsible for changes observed in the CSS total score over time.

Score Interpretation

Results from anchor-based methods and supportive analyses suggested estimates of MWPC between – 6 and – 11 points in the CSS (Tables 8, 9, 10). Participants who achieved the CSS MWPC (i.e., at least 6 points improvement) had statistically significantly ($p<0.001$) higher rates of CDAI clinical remission/response and endoscopic remission/response, as well as greater improvements in patient-reported quality-of-life measures from baseline to week 12 (Table 11). This finding supports the clinical relevance of the estimated MWPC threshold.

Figure 3 presents the eCDF plots for the CSS change score distributions in the ADVANCE study between baseline and week 12, grouped by change categories in PGIS responses during the same timepoints. Patients reporting any kind of worsening were grouped together in this analysis. Figure 3 shows that, overall, the CSS change score distributions were distinct by anchor groups, though there was some overlap between "no change" and "worsened" groups at the upper end of the change score distribution.

Table 5 Cognitive debriefing summary table: items

CSS item	Participant interpretation of item concept *n* (%)	Participant has experienced item concept *n* (%)
Item 01: During the past week, were your bowel movements more frequent than usual?	As intended 27/30 (90.00%) Not as intended 3/30 (10.00%) One participant interpreted bowel movements as passing gas and/or defecating One participant was unfamiliar with the term bowel movements One participant interpreted bowel movements as pain while using the restroom	25/26 (96.15%)[a]
Item 02: During the past week, did you pass gas more than usual?	As intended 29/30 (96.67%) Not as intended 1/30 (3.33%) One participant interpreted passing gas as inclusive of burping	24/28 (85.71%)[a]
Item 03: During the past week, did you have abdominal pain?	As intended 27/29 (93.10%)[a] Not as intended 2/29 (6.90%)[a] One participant was unfamiliar with the term abdominal One participant interpreted abdominal pain as pain below the stomach	28/28 (100.00%)[a]
Item 04: During the past week, have you felt tired or lacking energy?	As intended 30/30 (100.00%)	30/30 (100.00%)
Item 05: During the past week, did you feel nauseated?	As intended 25/30 (83.33%) Not as intended 5/30 (16.67%) Two participants interpreted nausea as throwing up Two participants interpreted nausea as dizziness One participant Interpreted nausea as feeling generally sick, including having diarrhea and feeling tired Of the participants who did not interpret the item concept as intended almost all ($n = 4/5$, 80.00%) were adolescents	22/24 (91.67%)[a]
Item 06: During the past week, did you experience loss of appetite?	As intended 30/30 (100.00%)	27/30 (90.00%)

Table 5 continued

CSS item	Participant interpretation of item concept *n* (%)	Participant has experienced item concept *n* (%)
Item 07: During the past week, did you have joint pain?	As intended 24/28 (85.71%)[a] Not as intended 4/28 (14.29%)[a] Two participants interpreted joint pain as muscle pain Two participants interpreted joint pain as pain all over the body	19/24 (79.17%)[a]
Item 08: During the past week, did you have difficulty sleeping?	As intended 29/30 (96.67%) Not as intended 1/30 (3.33%) One participant attributed difficulty sleeping to reasons other than CD	26/30 (86.67%)
Item 09: During the past week, did you experience bloating?	As intended 25/27 (92.59%)[a] Not as intended 2/27 (7.41%)[a] One participant attributed bloating to overeating One participant was unfamiliar with the term bloating	24/24 (100.00%)[a]
Item 10: During the past week, did you have diarrhea?	As intended 24/26 (92.31%)[a] Not as intended 2/26 (7.69%)[a] Two participants interpreted diarrhea as extremely frequent bowel movements	22/24 (91.67%)[a]
Item 11: During the past week, did you have blood in your stool?	As intended 28/30 (93.33%) Not as intended 2/30 (6.87%) One participant was unfamiliar with the term stool One participant interpreted the item as bleeding while urinating or defecating	23/27 (85.19%)[a]
Item 12: During the past week, were you constipated?	As intended 30/30 (100.00%)	25/30 (83.33%)
Item 13: During the past week, did you vomit or throw up?	As intended 28/28 (100.00%)[a]	25/28 (89.29%)[a]
Item 14: During the past week, did your stomach growl or make gurgling sounds?	As intended 27/28 (96.43%)[a] Not as intended 1/28 (3.57%)[a] One participant attributed stomach growling or gurgling to hunger	27/28 (96.43%)[a]

CSS Crohn's Symptom Severity

[a]Results are calculated from varying denominators based on the total number of participants who were determined to have an evaluable response available for analysis and interpretation

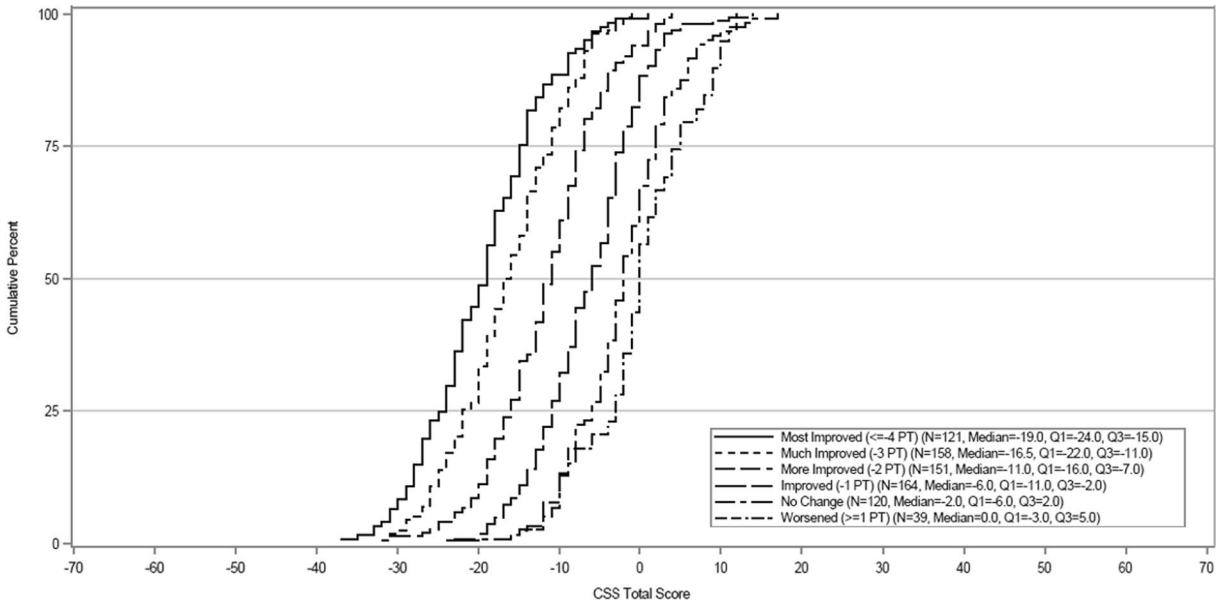

Fig. 1 Score distributions of CSS items and total score at baseline and week 12 for the ADVANCE study

The change score distributions presented in Fig. 3 and the percentile change presented in Table 10 align with the mean-change anchor-based estimates.

Figure 4 presents kernel-smoothed PDF curves for the changes in CSS change-score distributions, based on PGIS change categories between baseline and week 12. PDF curves show that improvement groups align with the (negative) change scores presented in the eCDF plots (Fig. 3) and anchor-based analyses.

Figure 5 shows the ROC curve for participants with a 1-point change on the PGIS between baseline and week 12. The thresholds suggested by Youden's *J* were more sensitive to the degree of change and ranged between changes of 7 and 11 points on the CSS. The sum-of-squares thresholds were less variable and ranged between 7 and 9 points.

DISCUSSION

This research supports the content validity, psychometric performance, and score interpretation of the CSS in patients with moderately to severely active CD ages 16–80 years. Specifically, results from qualitative cognitive debriefing interviews demonstrated that patients could read and interpret the CSS as intended, that the concepts measured in the CSS are relevant, and that the CSS comprehensively measures patients' experience with CSS symptoms. Further, psychometric evaluation of the clinical trial data suggests that the CSS is valid and sensitive to change, providing evidence supporting the use of the CSS in this population. Preliminary MWPC estimates for the CSS total score (6–11 points improved) were indicated for use in clinical research, which demonstrated the clinical relevance including improvements in clinical remission/response and endoscopic remission/response, as well as greater improvements in patient-reported quality-of-life.

While these results comprehensively evaluated the CSS through qualitative patient interviews and psychometric analyses involving a large sample of patients from a clinical trial with moderately to severely active CD, there were some limitations. First, select CSS items had weak inter-item correlations, which may suggest minimal conceptual overlap within the instrument. However, given the range of CD symptoms and impacts captured in the CSS,

Table 6 Item–total and inter-item Spearman correlations for the CSS at ADVANCE week 12 ($N = 819$)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14
CSS total score	0.62	0.58	0.79	0.75	0.59	0.63	0.57	0.63	0.69	0.62	0.41	0.26	0.33	0.64
Item 1. Bowel movements more than usual	1.00	–	–	–	–	–	–	–	–	–	–	–	–	–
Item 2. Pass gas more than usual	0.44	1.00	–	–	–	–	–	–	–	–	–	–	–	–
Item 3. Abdominal pain	0.50	0.43	1.00	–	–	–	–	–	–	–	–	–	–	–
Item 4. Felt tired or lacking energy	0.43	0.35	0.57	1.00	–	–	–	–	–	–	–	–	–	–
Item 5. Feel nauseated	0.34	0.27	0.47	0.45	1.00	–	–	–	–	–	–	–	–	–
Item 6. Experience loss of appetite	0.35	0.27	0.51	0.50	0.53	1.00	–	–	–	–	–	–	–	–
Item 7. Joint pain	0.24	0.26	0.41	0.48	0.31	0.30	1.00	–	–	–	–	–	–	–
Item 8. Difficulty sleeping	0.29	0.30	0.43	0.51	0.39	0.40	0.43	1.00	–	–	–	–	–	–
Item 9. Experience bloating	0.38	0.43	0.54	0.43	0.37	0.34	0.34	0.35	1.00	–	–	–	–	–
Item 10. Diarrhea	0.46	0.27	0.45	0.37	0.27	0.36	0.24	0.26	0.35	1.00	–	–	–	–
Item 11. Blood in your stool	0.24	0.19	0.31	0.24	0.17	0.25	0.15	0.22	0.23	0.23	1.00	–	–	–
Item 12. Constipated	0.07	0.11	0.17	0.15	0.14	0.14	0.11	0.19	0.21	−0.10	0.18	1.00	–	–
Item 13. Vomit or throw up	0.16	0.14	0.24	0.20	0.42	0.36	0.11	0.18	0.19	0.17	0.16	0.13	1.00	–
Item 14. Stomach growl or gurgling	0.34	0.39	0.44	0.36	0.28	0.30	0.28	0.28	0.50	0.44	0.17	0.12	0.16	1.00

CSS Crohn's Symptom Severity Scale

Table 7 Spearman correlation coefficients between CSS total score and concurrent assessments at baseline, week 4, and week 12 for the ADVANCE study

Instrument/score	Hypothesized relationship to CSS	CSS total score		
		Baseline ($N = 850$)	Week 4 ($N = 841$)	Week 12 ($N = 819$)
Convergent validity				
CDAI	++	0.29 (+)	0.59 (++)	0.69 (++)
PGIS	++	0.57 (++)	0.69 (++)	0.78 (+++)
IBDQ total score	++	− 0.65 (++)	− 0.75 (+++)	− 0.81 (+++)
IBDQ—Bowel Symptom Domain score	++	− 0.72 (+++)	− 0.80 (+++)	− 0.84 (+++)
IBDQ—Emotional Function Domain score	++	− 0.53 (++)	− 0.61 (++)	− 0.69 (++)
IBDQ—Social Function Domain score	++	− 0.48 (++)	− 0.60 (++)	− 0.65 (++)
IBDQ—Systemic Symptom Domain score	++	− 0.62 (++)	− 0.77 (+++)	− 0.80 (+++)
SF-36v2® Physical Component Summary score	++	− 0.46 (++)	− 0.59 (++)	− 0.69 (++)
EQ-5D-5L Mobility	++	0.36 (++)	0.36 (++)	0.37 (++)
EQ-5D-5L Self-care	++	0.23 (+)	0.27 (+)	0.28 (+)
EQ-5D-5L Usual activities	++	0.44 (++)	0.54 (++)	0.55 (++)
WPAI:CD Presenteeism (impairment at work/reduced on-the-job effectiveness)	++	0.42 (++)	0.54 (++)	0.57 (++)
WPAI:CD Work productivity loss (overall work impairment/absenteeism plus presenteeism)	++	0.41 (++)	0.51 (++)	0.55 (++)
Discriminant validity				
EQ-5D-5L Anxiety/depression	+	0.34 (++)	0.41 (++)	0.45 (++)
EQ-5D-5L Pain/discomfort	+	0.54 (++)	− 0.60 (++)	0.71 (++)
EQ-5D VAS	+	− 0.44 (++)	− 0.58 (++)	− 0.65 (++)
WPAI:CD Absenteeism	+	− 0.33 (++)	− 0.41 (++)	− 0.38 (++)
SF-36v2® Mental component summary score	+	− 0.41 (++)	− 0.46 (++)	− 0.55 (++)

CD Crohn's disease, *CDAI* Crohn's Disease Activity Index, *CSS* Crohn's Symptom Severity Scale, *EQ-5D-5L* five-level EQ-5D, *EQ-5D-VAS* EQ-5D Visual Analogue Scale, *IBDQ* Inflammatory Bowel Disease Questionnaire, *PGIS* Patient Global Impression of Severity, *SF-23v2*® 36-Item Short-Form Survey, version 2, *WPAI:CD* Work Productivity and Activity Impairment Questionnaire: Crohn's Disease

+++ = strong relationship (> 0.70 but ≤ 0.90), ++ = moderate relationship (> 0.30 but ≤ 0.70), + = weak relationship (≤ 0.30)

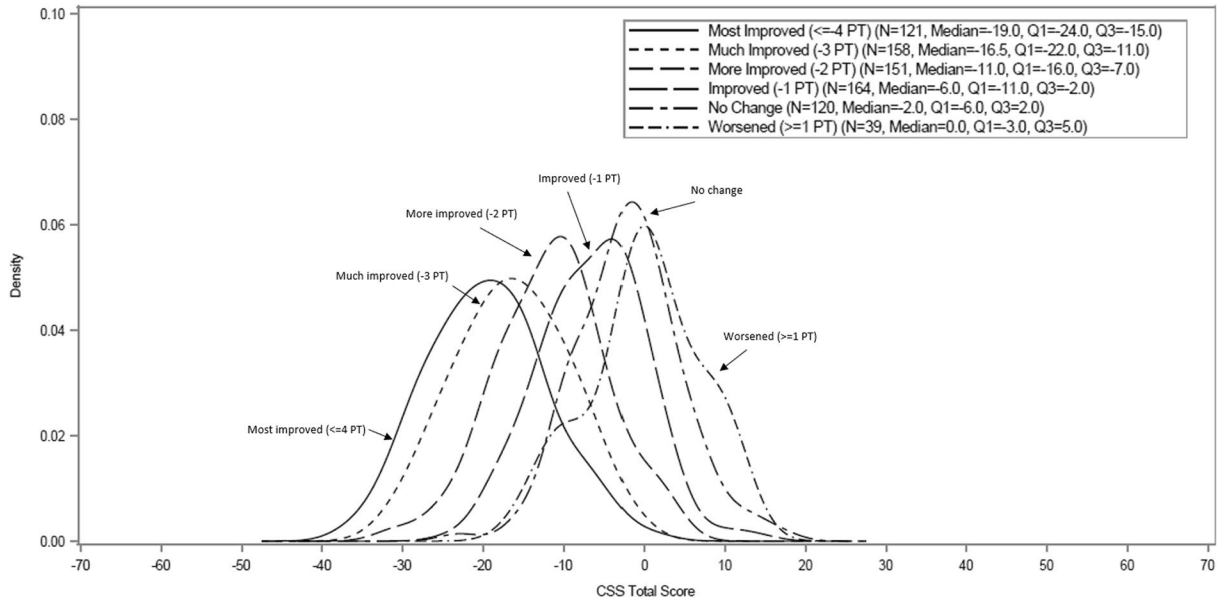

Fig. 2 Known-groups comparisons for CSS total score at week 12 for the ADVANCE study. *CDAI* Crohn's Disease Activity Index, *CI* confidence interval, *CSS* Crohn's Symptom Severity Scale, *IBDQ* Inflammatory Bowel Disease Questionnaire, *n* sample size, *PGIS* Patient Global Impression of Severity

and the fact that the CSS includes CD symptoms beyond gastrointestinal symptoms, it is expected that certain items would show weaker inter-item correlations. This result by itself does not necessarily reflect a weakness of the CSS as a measure, but rather shows the broad range of symptoms captured in the measure and distinctness in the items assessed, as expected in a multi-item symptom measure.

In addition, ICC estimates for test–retest reliability did not consistently exceed acceptable thresholds for test–retest reliability; however, the setting and timepoints in which test–retest analyses were conducted may have resulted in the lower ICC estimates. Ideally, test–retest analysis is conducted when the only change between repeated assessments is the passage of time, and any other factor that could produce systematic change in the sample (e.g., treatment) is absent or controlled. In the context of a randomized phase 3 study, this was not possible. Additionally, timepoints to use for test–retest analyses should be close (e.g., 1–2 weeks apart), as significant change is not expected over a shorter period. However, the timepoints available for this analysis (baseline, week 4, and week 12) were 1–2 months apart. Given limitations in the available data, the observed ICCs should be interpreted with caution, and likely underestimate the reliability of the scores, and future studies should explore this further in a more controlled setting.

While the qualitative interviews demonstrated that all items of the CSS assess experiences relevant to this population (i.e., all symptoms and impacts were said to be experienced by ≥ 79.17% of the sample), the score distributions from the clinical trial analyses, the baseline scores of items assessing feeling nauseated (item 5), blood in stool (item 11), constipation (item 12), and vomiting (item 13) in the trial reflected that 35.06%, 42.47%, 69.05%, and 77.29%, respectively, did not experience those symptoms within the past week; this may be due to the symptoms being more episodic in nature (e.g., representing

Table 8 Anchor-based estimates of CSS score MWPC by PGIS-stratified anchor categories from baseline to week 12 for the ADVANCE study

Score	Change in PGIS (baseline to week 12)	*N*	Baseline mean (SD)	Week 12 mean (SD)	Mean change (SD)	Within-group *p* value[a]	Between-groups *p* value[b]
CSS total score	Most improved (4-point or greater improvement)	121	42.38 (6.93)	22.87 (6.11)	− 19.51 (7.41)	< 0.001	< 0.001*
	Much improved (3-point improvement)	158	40.08 (7.23)	23.67 (5.42)	− 16.41 (6.88)	< 0.001	
	More improved (2-point improvement)	151	39.12 (8.31)	27.54 (7.31)	− 11.58 (6.87)	< 0.001	
	Improved (1-point improvement)	164	38.95 (8.33)	32.50 (8.50)	− 6.45 (6.45)	< 0.001	
	No change	120	38.63 (8.24)	36.63 (9.52)	− 2.00 (6.45)	< 0.001	
	Worsened (1-point or greater deterioration)	39	38.26 (7.78)	38.28 (7.96)	0.03 (6.88)	0.913	

CD Crohn's disease, *CSS* Crohn's Symptom Severity Scale, *MWPC* minimal within-person change, *PGIS* Patient Global Impression of Severity, *SD* standard deviation

[a]The within-group *p* value is from a Wilcoxon signed rank test on change scores at each level of PGIS (7-level) response

[b]The between-groups *p* value is from a Kruskal–Wallis testing distributional shift in change scores between PGIS (7-level) response groups

bowel obstruction). However, even with low endorsement of the symptoms at baseline, the items showed improvement over time. Further, during the qualitative interviews, blood in stool and vomiting were reported to be among the most important questions to ask individuals with CD, and nearly all participants with data reported experiencing nausea ($n = 22/23$; 95.65%) and constipation ($n = 25/28$; 89.29%) within or outside the recall period. Considering this and that internal consistency reliability between all items and the total score was considered acceptable, no changes to the CSS scoring (i.e., removal of items) were deemed necessary. For future research, consideration of the target patient population and the symptoms that those patients experience (e.g., the severity of the disease experience) should inform the evaluation of the future scoring.

CONCLUSIONS

CD symptoms can be impactful and burdensome for patients. It is highly important to consider patients' experiences of their CD symptoms and impacts in clinical research using

Table 9 Anchor-based estimates of CSS score MWPC by PGIC-stratified anchor categories from baseline to week 12 for the ADVANCE study

Score	PGIC at week 12	*N*	Baseline mean (SD)	Week 12 mean (SD)	Mean change (SD)	Within-group *p* value[a]	Between-groups *p* value[b]
CSS total score	Very much improved (PGIC = 1)	119	38.87 (7.24)	20.62 (4.98)	− 18.25 (7.34)	< 0.001	< 0.001*
	Much improved (PGIC = 2)	290	39.70 (8.14)	25.82 (6.58)	− 13.88 (7.97)	< 0.001	
	Minimally improved (PGIC = 3)	196	39.56 (7.75)	31.70 (7.62)	− 7.86 (7.61)	< 0.001	
	Not changed (PGIC = 4)	112	39.30 (8.74)	35.37 (8.61)	− 3.94 (7.37)	< 0.001	
	Worsened (PGIC > 4)	53	41.64 (7.14)	41.91 (8.09)	0.26 (6.24)	0.894	

CD Crohn's disease, *CSS* Crohn's Symptom Severity Scale, *MWPC* minimal within-person change, *PGIC* Patient Global Impression of Change, *SD* standard deviation

[a]The within-group *p* value is from a Wilcoxon signed rank test on change scores at each level of PGIC response

[b]The between-groups *p* value is from a Kruskal–Wallis testing distributional shift in change scores between PGIC (seven-level) response groups

Table 10 Percentile change in CSS total scores from baseline to week 12 by PGIS response groups per empirical cumulative distribution function curve for ADVANCE

Change in CSS total score	Baseline to week 12 PGIS (7-level) anchor category							
	≥ 3-pt improvement	2-pt improvement	1-pt improvement	No change	1-pt deterioration	2-pt deterioration	3-pt deterioration	Total
N	279	151	164	120	33	5	1	753
Mean (SD)[a]	− 17.75 (7.27)	− 11.58 (6.87)	− 6.45 (6.45)	− 2.00 (6.45)	− 0.33 (6.87)	2.40 (8.02)	0.00 (N.A.)	− 10.62 (9.35)
10th percentile[b]	− 27.00	− 20.00	− 15.00	− 10.00	− 10.00	− 10.00	0.00	− 23.00
25th percentile[b]	− 23.00	− 16.00	− 11.00	− 6.00	− 3.00	1.00	0.00	− 17.00
Median (50th percentile)[b]	− 18.00	− 11.00	− 6.00	− 2.00	0.00	2.00	0.00	− 10.00
75th percentile[b]	− 13.00	− 7.00	− 2.00	2.00	4.00	9.00	0.00	− 4.00
90th percentile[b]	− 8.00	− 3.00	1.00	6.00	9.00	10.00	0.00	1.00

CSS Crohn's Symptom Severity, *N.A.* not applicable, *PGIS* Patient Global Impression of Severity, *SD* standard deviation

[a]Mean (SD) for the change score on the CSS between baseline and week 12 for each anchor category

[b]Change score for CSS is presented associated with each percentile group

Table 11 Responder evaluation of CSS total scores using an estimated meaningful within-person change of ≤ − 6 points ($N = 819$)

Instrument	Responder		Non-responder		Difference in mean score	*p* value
	N	% or mean (SD)	*N*	% or mean (SD)		
CDAI clinical remission	343	73.13%	126	26.87%	N.A.[a]	< 0.001
CDAI clinical remission	388	82.73%	81	17.27%	N.A.[a]	< 0.001
CDAI clinical response	280	85.89%	46	14.11%	N.A.[a]	< 0.001
Endoscopic response	204	81.60%	46	18.40%	N.A.[a]	< 0.001
Endoscopic remission	143	85.63%	24	14.37%	N.A.[a]	< 0.001
IBDQ ≥ 170	316	84.49%	58	15.51%	N.A.[a]	< 0.001
Change of IBDQ ≥ 16	477	82.96%	98	17.04%	N.A.[a]	< 0.001
EQ-5D VAS	522	73.15 (17.24)	240	56.14 (20.11)	17.01 (18.19)	< 0.001
SF-36v2® Physical Component Summary	522	48.79 (7.41)	239	42.19 (7.97)	6.61 (7.59)	< 0.001
SF-36v2® Mental Component Summary	522	47.71 (9.74)	239	41.28 (11.16)	6.43 (10.21)	< 0.001
WPAI:CD Work time missed (%)	307	15.62 (27.98)	131	27.12 (31.07)	− 11.50 (28.94)	< 0.001
WPAI:CD Impairment while working (%)	289	23.60 (21.88)	120	44.42 (26.56)	− 20.82 (23.35)	< 0.001
WPAI:CD Overall work impairment (%)	307	34.27 (30.12)	131	56.83 (31.42)	− 22.56 (30.52)	< 0.001
WPAI:CD Activity impairment (%)	522	29.54 (25.32)	239	50.50 (27.19)	− 20.96 (25.92)	< 0.001

CDAI Crohn's Disease Activity Index, *CSS* Crohn's Symptom Severity, *EQ-5D-5L* Five-level EQ-5D, *IBDQ* Inflammatory Bowel Disease Questionnaire, *N.A.* not applicable, *SD* standard deviation, *SF-36v2®* 36-Item Short Form Survey Version 2, *VAS* visual analogue scale, *WPAI:CD* Work Productivity and Activity Impairment Questionnaire for Crohn's Disease

[a]Difference in percentages were not calculated, as they were analyzed with chi-squared tests against the null hypothesis of equal percentages

patient-reported outcome measures, as those experiences are often most reliably reported via self-report. Overall, this study supports use of the CSS for measuring CD symptoms and sleep impact among adolescents and adults aged 16 and older with moderately to severely active CD in clinical research.

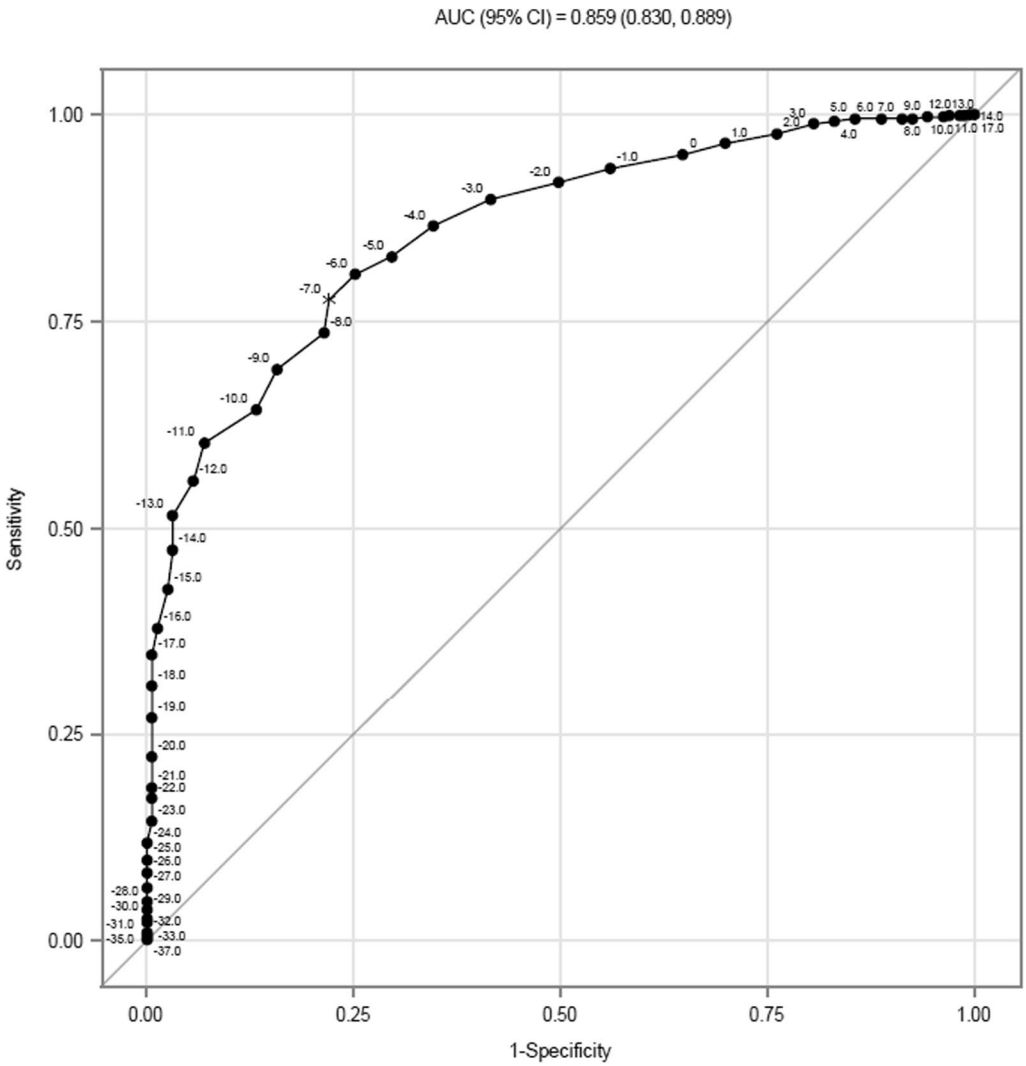

Fig. 3 Empirical cumulative distribution function for change in CSS total score by PGIS change response groups from baseline to week 12 for the ADVANCE study ($N = 819$). *CSS* Crohn's Symptom Severity Scale, *PGIS* Patient Global Impression of Severity, *PT* points

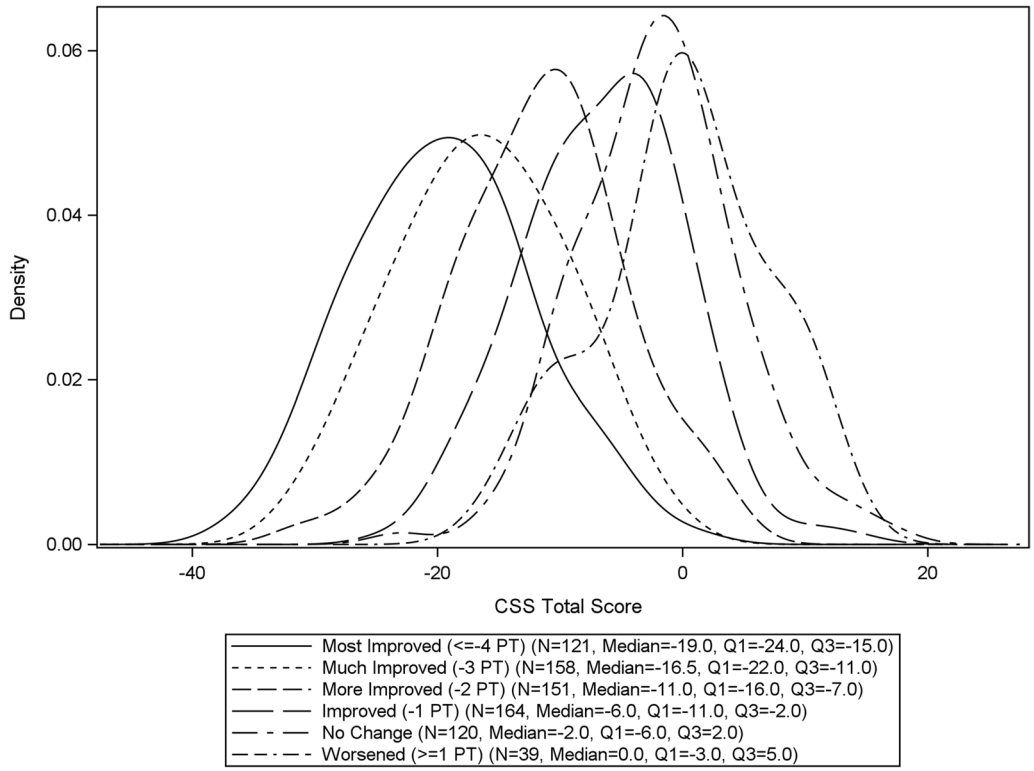

Fig. 4 Probability density function for change in CSS by PGIS change response groups from baseline to week 12 for the ADVANCE study ($N = 819$). *CSS* Crohn's Symptom Severity Scale, *PGIS* Patient Global Impression of Severity, *PT* points

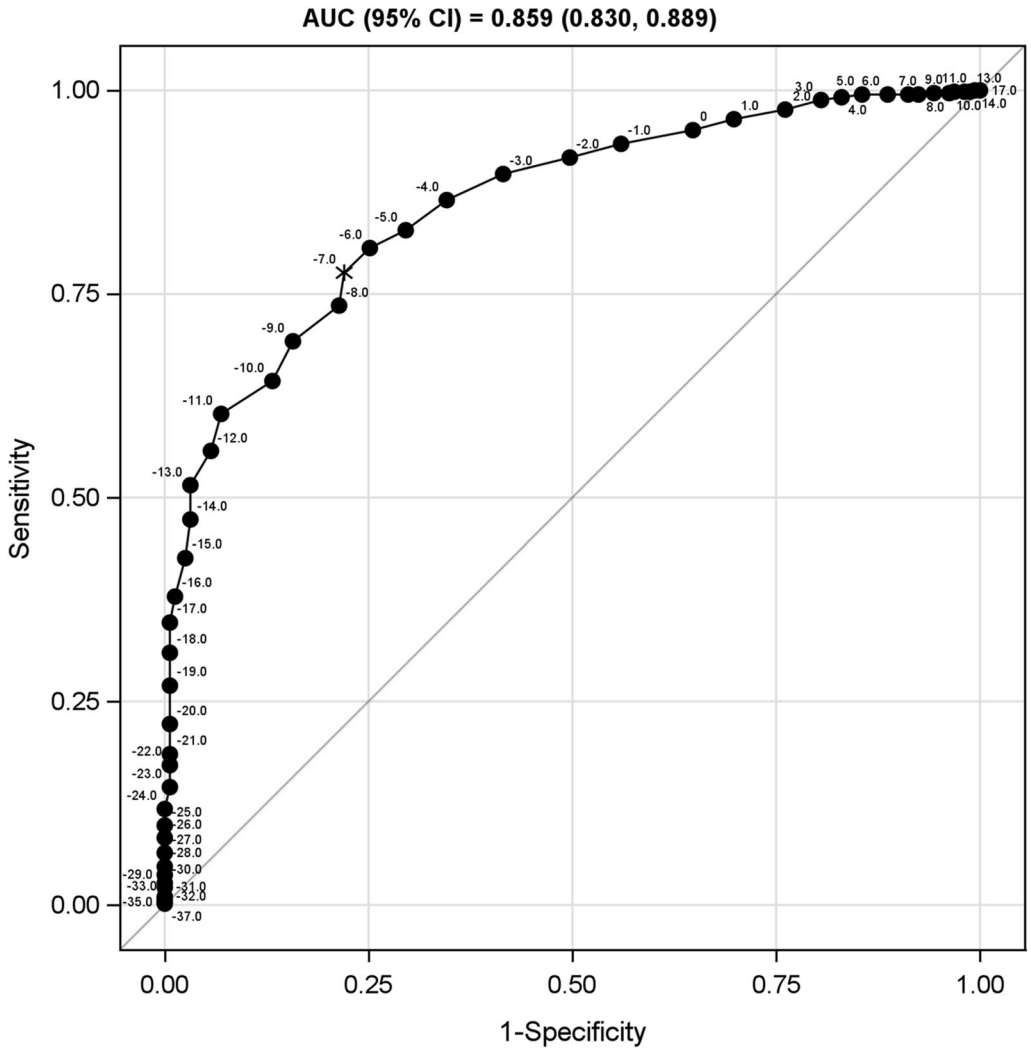

Fig. 5 Receiver operating characteristic curve for CSS total score by PGIS 1-point improvement from baseline to week 12 for the ADVANCE study. *AUC* area under the curve, *CI* confidence interval, *CSS* Crohn's Symptom Severity, *PGIS* Patient Global Impression of Severity, *SE* sensitivity, *SP* specificity, *ROC* received operating characteristic. [1] Youden's *J*. [2] Point on the ROC that minimizes the sum of squares

ACKNOWLEDGEMENTS

The authors would like to thank Alan Shields, Sylvia Su, Ethan Arenson, Doug Huntington, and Alejandro Moreno-Koehler for their important contributions to the completion of this research.

Medical Writing, Editorial, and Other Assistance. Writing support was provided by Sylvia Su of Adelphi Values and funded by AbbVie, Inc.

Author Contributions. AbbVie (Wan-Ju Lee, Kristina Fitzgerald, Ana Paula Lacerda, Ezequiel Neimark, and Yuri Sanchez Gonzalez) and Adelphi Values authors (Leighann Litcher-Kelly, Sarah Ollis, and Emma Pranschke) supported conception of the study design, conduct of research, and analysis. All authors (from AbbVie, Adelphi Values, and external authors [Edouard Louis, Julian Panés]) were involved in data collection and interpretation of data. All authors had access to the data results and participated in the development, review, and approval of this manuscript. All authors have approved the final version of the article, including the authorship list.

Funding. AbbVie, Inc. provided funding for the research described in this manuscript and the journal's Rapid Service and Open Access fees. AbbVie participated in the study design, research, data collection, analysis and interpretation of data, writing, reviewing, and approving the publication. No honoraria or payments were made for authorship.

Data Availability. AbbVie is committed to responsible data sharing regarding the clinical trials we sponsor. This includes access to anonymized, individual, and trial-level data (analysis data sets), as well as other information (e.g. protocols, clinical study reports, or analysis plans), as long as the trials are not part of an ongoing or planned regulatory submission. This includes requests for clinical trial data for unlicensed products and indications. These clinical trial data can be requested by any qualified researchers who engage in rigorous, independent, scientific research, and will be provided following review and approval of a research proposal, Statistical Analysis Plan (SAP), and execution of a Data Sharing Agreement (DSA). Data requests can be submitted at any time after approval in the USA and Europe and after acceptance of this manuscript for publication. The data will be accessible for 12 months, with possible extensions considered. For more information on the process or to submit a request, visit the following link: https://vivli.org/ourmember/abbvie/ then select "Home".

Declarations

Conflict of Interest. Edouard Louis has received research grants from Janssen, Pfizer, and Takeda; received educational grants from AbbVie, Janssen, MSD, and Takeda; received speaker fees from AbbVie, Falk, Ferring, Hospira, Janssen, MSD, Pfizer, and Takeda; served on advisory boards for AbbVie, Arena, Celgene, Ferring, Galapagos, Gilead, Hospira, Janssen, MSD, Pfizer, and Takeda; and served as a consultant for AbbVie. Wan-Ju Lee, Yuri Sanchez Gonzalez, Ana Paula Lacerda, Kristina Fitzgerald, and Ezequiel Neimark are full-time employees of AbbVie and may hold AbbVie stock and/or stock options. At the time of this manuscript's development, Leighann Litcher-Kelly, Sarah Ollis, and Emma Pranschke were employed by Adelphi Values LLC, which received payment from AbbVie Inc. to support the research activities presented in this publication. Emma Pranschke is currently employed by Analysis Group. Julian Panés received financial support for research from AbbVie and Pfizer; consultancy fees/honorarium from AbbVie, Arena, Athos, Atomwise, Boehringer Ingelheim, Celgene, Celltrion, Ferring, Galapagos, Genentech/Roche, GlaxoSmithKline, Janssen, Mirum, Morphic, Nestlé, Origo, Pandion, Pfizer, Progenity, Protagonist, Revolo, Robarts, Takeda, Theravance and Wasserman; reports payment for lectures including service on speaker bureau from Abbott, Ferring, Janssen, Pfizer and Takeda; and reports payment for development of educational presentations from Abbott, Janssen, Pfizer Roche and Takeda.

Ethical Approval. The study protocols and all study documents were approved by ethical review committees. All participants in this research provided informed consent for their participation (or assent and caregiver consent for individuals under the age of 18).

REFERENCES

1. Gohil K, Carramusa B. Ulcerative colitis and Crohn's disease. Pharm Ther. 2014;39(8):576–7.

2. GBD 2017 Inflammatory Bowel Disease Collaborators. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol. 2020;5(1):17–30.

3. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390(10114):2769–78.

4. Panaccione R. Mechanisms of inflammatory bowel disease. Gastroenterol Hepatol (N Y). 2013;9(8):529–32.

5. National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention. Inflammatory bowel disease (IBD). 2022. https://www.cdc.gov/ibd/index.htm. Accessed 18 Aug 2022.

6. 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(11):1047–66.

7. 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. 2018;2(1):24.

8. Best WR, Becktel JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology. 1976;70(3):439–44.

9. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. Lancet. 1980;1(8167):514.

10. Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology. 1989;96(3):804–10.

11. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Office of Communications, Division of Drug Information; 2009.

12. D'Haens G, Panaccione R, Baert F, et al. Risankizumab as induction therapy for Crohn's disease: results from the phase 3 ADVANCE and MOTIVATE induction trials. Lancet. 2022;399(10340):2015–30.

13. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473–83.

14. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.

15. Panes J, Otley A, Sanchez Gonzalez Y, et al. Ulcerative Colitis-Symptom Questionnaire: valid for use in adults with moderately to severely active ulcerative colitis. Dig Dis Sci. 2023;86:2318–32.

16. Loftus EV Jr, Ananthakrishnan AN, Lee W-J, et al. Content validity and psychometric evaluation of the functional assessment of chronic illness therapy-fatigue (FACIT-Fatigue) in patients with

Crohn's disease and ulcerative colitis. PharmacoEcon Open. 2023;7:823–40.

17. Reilly MC, Gerlier L, Brabant Y, Brown M. Validity, reliability, and responsiveness of the work productivity and activity impairment questionnaire in Crohn's disease. Clin Ther. 2008;30(2):393–404.

18. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4(5):353–65.

19. Hinkle DE, Jurs SG, Wiersma W. Applied statistics for the behavioral sciences. 2nd ed. Boston: Houghton Mifflin; 1988.

20. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297–334.

21. Bland JM, Altman DG. Cronbach's alpha. BMJ. 1997;314(7080):572.

22. Weir J. Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.

23. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed 1 Dec 2022.

24. Nunnally JC. The assessment of reliability. In: Bernstein I, editor. Psychometric theory. New York: McGraw Hill; 1994. p. 248–92.

25. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41(5):582–92.