PharmacoEconomics - Open (2023) 7:823–840
https://doi.org/10.1007/s41669-023-00419-w

ORIGINAL RESEARCH ARTICLE

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

Edward V. Loftus Jr[1] · Ashwin N. Ananthakrishnan[2] · Wan-Ju Lee[3] · Yuri Sanchez Gonzalez[3] · Kristina A Fitzgerald[3] · Kori Wallace[3] · Wen Zhou[3] · Leighann Litcher-Kelly[4] · Sarah B. Ollis[4] · Sylvia Su[4] · Silvio Danese[5]

Accepted: 30 April 2023 / Published online: 9 June 2023

Abstract

Background Patients with Crohn's disease (CD) or ulcerative colitis (UC) frequently experience fatigue, although it is often overlooked in medical research and practice.

Aims To explore patients' experience of fatigue and evaluate content validity, psychometric properties, and score interpretability of the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT–Fatigue) in patients with CD or UC.

Methods Concept elicitation and cognitive interviews were conducted with participants aged ≥ 15 years with moderately-to-severely active CD ($N = 30$) or UC ($N = 33$). To evaluate psychometric properties (reliability and construct validity) and interpretation of FACIT–Fatigue scores, data from two clinical trials were analyzed [ADVANCE (CD): $N = 850$; U-ACHIEVE (UC): 248]. Meaningful within-person change was estimated using anchor-based methods.

Results Almost all interview participants reported experiencing fatigue. Over 30 unique fatigue-related impacts were reported per condition. The FACIT–Fatigue was interpretable for most patients. FACIT–Fatigue items had good internal consistency (Cronbach's α 0.86–0.88 for CD and 0.94–0.96 for UC); the total score displayed acceptable test–retest reliability (intraclass correlation coefficients > 0.60 for CD and > 0.90 for UC). FACIT–Fatigue scores had acceptable convergent validity with similar measures. A 7–10 point improvement for CD and 4–9 point improvement for UC on the FACIT–Fatigue total score may represent meaningful improvements.

Conclusions These results highlight the importance of fatigue among adolescents and adults with CD or UC and provide evidence that the FACIT–Fatigue is content valid and produces reliable, valid, and interpretable scores in these populations. Care should be taken if using the questionnaire with adolescents who may be less familiar with the word "fatigue."
Clinical trial registration numbers NCT03105128 (date of registration: 4 April 2017) and NCT02819635 (date of registration: 28 June 2016).

1 Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are characterized by chronic relapsing and remitting inflammation of the gastrointestinal (GI) tract. While patients with CD and UC experience similar symptoms [1–3], the two conditions differ with respect to the location and extent of inflammation in the GI tract. CD can affect any portion of the GI tract and presents as transmural inflammation involving all tissue layers of the bowel wall, while UC is manifested as diffuse mucosal inflammation of the colon and/or rectum [4].

One impactful, yet often overlooked, symptom of CD and UC is fatigue [4–7]. The prevalence of fatigue among CD and UC patients, based on previous literature, ranges from approximately 24% to 87% [7–11]. The pooled prevalence of fatigue in CD and UC was 47% based on a random-effects meta-analysis, compared with only 5% in healthy individuals. For individuals with active disease, the pooled prevalence of fatigue was 72%, compared with 47% for those in remission [7]. Common risk factors for fatigue in CD and UC include sleep disturbance, anxiety, depression, vitamin and mineral deficiencies, and anemia [7]. Fatigue is associated with impaired health-related quality of life, including high disability, decreased physical function, and negative effects on work productivity [7]. The prevalence and burden of fatigue demonstrates its importance for consideration when developing interventions for CD and UC [7].

Fatigue, however, is not conceptually straightforward—it is a multifaceted symptom that may be experienced in various ways, both within and across individuals. As

Extended author information available on the last page of the article

Key Points for Decision Makers

Fatigue is an important symptom experienced by individuals with moderately-to-severely active CD and UC, with substantial impact to their lives.

The FACIT–Fatigue is comprehensible, relevant, and comprehensive for assessing fatigue and fatigue-related impacts for use with individuals with moderately-to-severely active CD and UC.

FACIT–Fatigue is a reliable and valid tool that is sensitive to change, and an improvement in the FACIT–Fatigue total score of 7–10 points or 4–9 points may be clinically meaningful for patients with moderately-to-severely active CD and UC, respectively, in a clinical trial setting.

such, it is important to consider an assessment approach accounting for diverse aspects of fatigue.

One common instrument to measure the patient experience of fatigue is the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT–Fatigue), which assesses concepts related to the severity and impacts of fatigue over the past 7 days [12]. While there are many questionnaires that could be considered for the evaluation of fatigue, and even some that assess fatigue specifically among those with inflammatory bowel disease (IBD), such as the Inflammatory Bowel Disease Questionnaire (IBDQ) [13], the FACIT–Fatigue captures varied fatigue-related concepts across its 13 items, allowing for a nuanced assessment (as opposed to a single item broadly evaluating fatigue).

While the FACIT–Fatigue was originally developed to assess anemia-related fatigue in cancer patients [14–16], and has been widely used in a variety of disease areas, the appropriateness of the FACIT–Fatigue for use in CD and UC has also been supported by prior publications [11, 17]. Previous studies on IBD have used the FACIT–Fatigue to evaluate the severity of fatigue in patients with IBD and have reported mean scores that are associated with moderate or severe fatigue [17–19], while lower fatigue was associated with improvements in other symptoms related to treatment [5, 20]. In addition, previous research has aimed to understand the reliability, validity, and sensitivity to change of the FACIT–Fatigue in CD and UC [17, 20–22].

To expand on this previous research and to confirm that the FACIT–Fatigue (13-item version) is appropriate for use in registrational clinical trials for CD and UC, the current study aimed to evaluate the content validity (readability, relevance, comprehensibility, and comprehensiveness), psychometric performance, and score interpretation of the FACIT–Fatigue for adolescents and adults with clinician-confirmed, moderately-to-severely active CD or UC.

2 Methods

2.1 Evaluation of Content Validity

Trained interviewers conducted open-ended, semi-structured qualitative interviews with adolescents and adults in the USA with moderately-to-severely active CD or UC to (1) identify, describe, and substantiate the important and relevant CD- and UC-related fatigue experiences and impacts (concept elicitation) and (2) evaluate the readability, comprehensibility, relevance, and comprehensiveness of the FACIT–Fatigue (cognitive debriefing). The study protocol and all study documents were approved by a centralized independent review board.

2.1.1 Participants

Adolescents (i.e., 15–17 years of age) and adults (i.e., ≥ 18 years of age) with a clinician-confirmed diagnosis of moderately-to-severely active CD or UC were recruited from clinical sites within the USA. The study inclusion/exclusion criteria specified, among other criteria, that Health Insurance Portability and Accountability Act (HIPAA) authorization, informed consent to participate for adult participants, and parental permission and assent for adolescent participants be obtained. Participants ≥ 15 years of age and ≤ 80 years old, fluent in US English, and with a clinician-confirmed diagnosis of moderate-to-severe CD or UC were considered for inclusion in the study. Full inclusion and exclusion criteria can be found in Supplementary Table 1 in the Online Resource 1: Supplementary tables and figures.

2.1.2 Conduct of Qualitative Interviews

The qualitative interviews were conducted either in person or over the phone using a semi-structured interview guide. During the interviews, participants were first asked to briefly discuss whether they have experienced fatigue related to their CD or UC and to describe how their disease-related fatigue felt and how it impacted their lives. Participants were subsequently asked to complete the FACIT–Fatigue either on an electronic device or by using screenshots of the questionnaire while "thinking aloud" about the process they used to arrive at each answer to identify any words, terms, or concepts within

the questionnaire that they did not understand or did not interpret as intended. Following the "think-aloud" process, participants were asked additional questions designed to evaluate the content of the FACIT–Fatigue, including its comprehensibility, readability, relevance, and comprehensiveness. Interviews were audio recorded, following participants' verbal consent, and subsequently transcribed and anonymized.

2.1.3 Qualitative Coding and Data Analysis

Data were analyzed separately for CD and UC. Each transcript underwent qualitative coding to organize and catalog participants' CD- or UC-related fatigue experience, feedback, and responses on the instructions, items, and response options of the FACIT–Fatigue. All transcripts were coded in ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany); coded quotations were then further reviewed and aggregated by themes in study findings tables.

2.2 Evaluation of Psychometric Performance and Interpretability of Scores of the FACIT–Fatigue

Data from two clinical programs [risankizumab for CD Phase 3 (NCT03105128; hereafter identified as ADVANCE) [23] and upadacitinib for UC Phase 2b (NCT02819635; hereafter identified as U-ACHIEVE) [24] were used for the psychometric and score interpretation analyses. The following properties of the FACIT–Fatigue scores were evaluated for CD and UC separately: quality of completion of the questionnaire, total score distribution, reliability [internal consistency using Cronbach's alpha and test–retest reliability using intraclass correlation coefficient (ICC)], and validity (convergent and discriminant validity, known-groups methods, and sensitivity to change). Further, analyses were conducted to establish meaningful within-person change (MWPC) estimates for the FACIT–Fatigue total score.

During the clinical trials, participants completed the FACIT–Fatigue questionnaire (along with other questionnaires/assessments) during clinic visits using an electronic tablet device. The methodology of these trials is described fully elsewhere [23–25].

2.2.1 Participants

Scores on the FACIT–Fatigue and other clinical outcome assessments (both patient reported and clinician reported) were included in the psychometric and score interpretation analyses for CD and UC separately. Analysis populations were drawn from those participants in each clinical trial who were randomized to an active treatment or placebo group, but required that participants had scores on the FACIT–Fatigue from one or more clinic visits. For CD, participants with scores on the FACIT–Fatigue at baseline, week 4, and/or week 12 of the clinical trial were included in the psychometric analyses. For UC, participants with scores on the FACIT–Fatigue at baseline, week 2, week 4, and/or week 8 were included. While all participants in the psychometric analyses were randomized, scores from all treatment groups were collapsed for all analyses (i.e., analyses did not examine the differences between treatment and placebo groups).

2.2.2 Assessments

The version of the FACIT–Fatigue (version 4) under evaluation includes 13 items measuring fatigue and its impacts. All items include the same 5-point verbal rating scale ranging from 0 ("not at all") to 4 ("very much") with a recall period of the "past 7 days." Items 7 (i.e., "I have energy") and 8 (i.e., "I am able to do my usual activities") express a positive connotation, while other items on the 5-point scale reflect a negative connotation (e.g., "I am too tired to eat") related to fatigue. To calculate a total score, the scores of all items except for items 7 and 8 are reversed by subtracting the response from 4. After reversing the items (specifically, items 1–6 and 9–13), a total score is calculated by summing the individual item scores. The FACIT–Fatigue (version 4) has a maximum score of 52 and a minimum score of 0, where a higher score equates to less fatigue. Item numbers will be referenced as 1–13; however, the associated item bank numbers are as follows: item 1 (HI7), item 2 (HI12), item 3 (An1), item 4 (An2), item 5 (An3), item 6 (An4), item 7 (An5), item 8 (An7), item 9 (An8), item 10 (An12), item 11 (An14), item 12 (An15), and item 13 (An16).

In addition to the FACIT–Fatigue, scores of other assessments were used to support the analysis. The Crohn's Disease Activity Index (CDAI) and the Crohn's Symptoms Severity Questionnaire (CSS) were used for CD, and the Adapted Mayo Score and the Ulcerative Colitis Symptoms Questionnaire (UC-SQ) were used for UC as disease-specific assessments. Impact and quality-of-life assessments, namely the Work Productivity and Activity Impairment Questionnaire (WPAI: CD and WPAI: UC), the IBDQ, and the 36-Item Short Form Survey Version 2 (SF-36v2®), were used for both CD and UC. The five-level EQ-5D (EQ-5D-5L) was used for CD/UC as a general health and utilities assessment. In addition, a disease-specific Global Impression of Change (PGIC) was used in both CD and UC, while a disease-specific Patient Global Impression of Severity (PGIS) was used for CD only. The PGIC and PGIS items asked patients to rate the change in, or the severity of, their

Table 1 Reference measures for psychometric and score interpretation analyses (CD and UC)

Measure	Concepts measured	Included in CD and/or UC analyses	Recall period	Scoring
Disease activity assessments				
Crohn's Disease Activity Index (CDAI)[a] [31]	Abdominal pain, stool frequency, and general well-being, in addition to presence of other conditions, treatment for diarrhea, abdominal mass, hematocrit, and body weight/standard weight	CD	The past 7 days	Summing the weighted individual scores of 8 items[b]
Adapted Mayo Score Scoring System[c] [32]	Stool frequency, rectal bleeding, and endoscopy results	UC	Current state	Subscores from each of the three domains each ranging from 0 to 3; total score ranges from 0 to 9[b]
Symptom assessments				
Crohn's Symptoms Severity Questionnaire (CSS)[d]	CD-related gastrointestinal and nongastrointestinal symptoms (frequency and intensity)	CD	The past 7 days	14 items with overall symptom score calculated by combining ratings from individual items (possible scores range from 14 to 70)[b]
UC Symptom Questionnaire (UC-SQ)[d]	UC-related gastrointestinal symptoms and nongastrointestinal symptoms	UC	The last week	17 items with the total score calculated by summing the rating for each item[b]
Impacts and quality of life assessments				
Work Productivity and Activity Index (WPAI) for CD or UC (WPAI: CD and WPAI: UC)[d] [33, 34]	Impact of disease on absenteeism, presenteeism, productivity loss, and activity impairment	CD and UC	The 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)[d] [13]	Dimensions of symptoms and health-related quality of life: bowel symptoms, systemic symptoms, social function, emotional function	CD and UC	The 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[e]
36-Item Short Form Fatigue (SF-36v2®)[d] [35]	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	CD and UC	The 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
Global assessments				
The Patient Global Impression of Severity (PGIS)[d]	Severity of the overall symptoms due to CD	CD	The past week	One item rated on a 7-point verbal rating scale from "Absent" to "Very severe"[a]
The Patient Global Impression of Change (PGIC)[d]	Overall change in their UC or CD symptoms	CD and UC	Since before treatment began	One item rated on a 7-point verbal rating scale ranging from "Very much improved" to "Very much worse"[a]

Table 1 (continued)

Measure	Concepts measured	Included in CD and/or UC analyses	Recall period	Scoring
General health and utilities assessment				
Five-level EQ-5D (EQ-5D-5L)[d] [36]	Mobility, self-care, usual activities, pain/discomfort, and anxiety/depression Self-evaluated health status	CD and UC	Current health (i.e., "today")	EQ-5D descriptive system: Each of the five dimensions includes fives levels[a] EQ-5D visual analog scale (VAS): A vertical VAS ranging from 0 (worst health you can imagine) to 100 (best health you can imagine)[a]

CD Crohn's disease, *UC* ulcerative colitis

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

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

[c]Clinician-rated measure

[d]Patient-rated measure

[e]Higher scores indicate better outcomes

CD or UC symptoms overall. All assessments are described in Table 1.

2.2.3 Statistical Analyses

Descriptive characteristics, reliability, validity, and score interpretation analyses were conducted on the clinical trial datasets for CD (ADVANCE) and UC (U-ACHIEVE) separately. Table 2 summarizes each set of analyses and benchmarks for determining acceptable results, where appropriate.

3 Results

3.1 Qualitative Research for Content Validity in CD and UC

3.1.1 Participant Demographics

Between January 2020 and September 2020, 30 individuals with CD and 33 individuals with UC who met all the study inclusion and none of the exclusion criteria (as described in Supplementary Table 1 in the Online Resource 1: Supplementary tables and figures) were recruited from four US clinical sites. Participants with clinician-confirmed moderate-to-severe CD ($n = 20$ adults and $n = 10$ adolescents) and UC ($n = 22$ adults and $n = 11$ adolescents) participated in qualitative hybrid concept elicitation and cognitive debriefing interviews. Participant demographic and health information are provided in Table 3.

3.1.2 Fatigue Experience

The open-ended concept elicitation portion of the interviews confirmed that almost all participants experienced fatigue due to CD or UC (CD: $n = 30/30$, 100.0%; UC: $n = 32/33$, 97.1%). Participants described fatigue as: "a feeling of physical or mental tiredness that comes in waves, exhaustion, weakness, lethargy, lack of motivation, lack of energy, wanting to go to bed or relax, sleepiness, being worn out, sluggish or slow, being unalert, lifeless, achy, and/or feeling drained." When asked to rate fatigue on a 0–10 numeric rating scale with 0 = not bothersome at all to 10 = most bothersome, CD participants reported a mean bother rating of 6.8 [median = 7, interquartile range (IQR) = 5–8] and UC participants reported a mean bother rating of 6.2 (median = 6.5, IQR = 5–8).

Fatigue was additionally reported to impact quality of life. Participants reported fatigue-related impacts ($n = 32$ unique impacts for CD and $n = 33$ unique impacts for UC) across nine domains (activities of daily living, cognitive function, emotional function, leisure activities, physical activities, relationships, sleep, social function, and work/

Table 2 Summary of psychometric analyses for FACIT–Fatigue scores (CD and UC)

Analysis	Description	Timepoint(s) for CD	Timepoint(s) for UC
Quality of completion	Evaluate missing scores for the FACIT–Fatigue items and total score	Baseline, week 4, and week 12	Baseline, week 2, and week 8
Item and total score distributions	Descriptive statistics to summarize FACIT–Fatigue item and total scores, as well as to describe the ceiling/floor effects in the target measure. While there is no generally agreed-upon threshold that defines a floor or ceiling effect, items were considered to have an extreme floor/ceiling effect if more than 40% of participants endorsed the lowest or highest option, respectively [37]	Baseline and week 12	Baseline and week 8
Inter-item correlation	Evaluate how highly items within the FACIT–Fatigue 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 [38]	Week 12	Week 8
Item-total correlation	Evaluate the extent to which each item contributes to the total score. The threshold for acceptable item-total correlations was ≥ 0.3 [26]	Week 12	Week 8
Internal consistency reliability	Evaluate the degree to which individual items are measuring the same general concept. Cronbach's alpha 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 $\alpha \geq 0.70$ [39]. Cronbach's alpha was calculated with each item removed from its respective score to assess the impact	Baseline and week 12	Baseline and week 8
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) "no change" for CD symptoms on the PGIC at week 4 and week 12. For UC, stability was defined in two ways: (1) participants who selected "no change" for UC symptoms on the PGIC at week 2 or (2) no change between baseline and week 2 on scores for UC-SQ item 6 (tired, lacking energy during the past week). 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 [40, 41]	Baseline, week 4, and week 12	Baseline and week 2

Table 2 (continued)

Analysis	Description	Timepoint(s) for CD	Timepoint(s) for UC
Convergent/discriminant validity	Evaluate the degree to which scores produced by FACIT–Fatigue 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 [38] For CD, concurrent assessments were the CDAI, PGIS, IBDQ, SF-36v2®, EQ-5D-5L, and WPAI: CD For UC, concurrent assessments were the Adapted Mayo Score, EQ-5D-5L, IBDQ, SF-36v2®, UC-SQ, and WPAI: UC	Week 12	Week 8
Known-groups methods	Evaluate the degree to which scores produced by the FACIT–Fatigue are capable of distinguishing among groups hypothesized a priori as being clinically distinct For CD, known groups were defined using the CDAI, IBDQ, and PGIS. The classification values for the CDAI were < 150 for remission versus ≥ 150 for nonremission, and for the IBDQ total score, ≥ 170 for remission versus < 170 for nonremission For UC, known groups were defined using the Adapted Mayo Score, UC-SQ item 6 (felt tired or lacking energy during the past week), and IBDQ item 2 (feeling of fatigue or being tired and worn out)	Week 12	Week 8
Sensitivity to change	Evaluate the degree to which the change of scores produced by the FACIT–Fatigue change in concert with the change of scores produced by other concurrent measures (same as those listed in the convergent validity above)	Baseline and week 12	Baseline and week 8
Anchor-based methods	Estimate the MWPC for the FACIT–Fatigue total score. For CD, anchors were the PGIS and PGIC, and for UC, the anchor was the PGIC. Estimates were supplemented by eCDF curves, PDF curves, and ROC curves	Baseline and week 12	Baseline and week 8
Distribution-based methods	Estimate the clinically important difference between groups for the FACIT–Fatigue total score MCID1: 0.5 of an SD [42] at Baseline; MCID2: SEM: $\text{SEM} = \text{SDatBaseline} * \sqrt{1 - \text{reliability}}$ [a]	Baseline	Baseline

CD Crohn's disease, *CDAI* Crohn's Disease Activity Index, *eCDF* empirical cumulative distribution function, *EQ-5D-5L* Five-level EQ-5D, *FACIT–Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *IBDQ* Inflammatory Bowel Disease Questionnaire, *ICC* intraclass correlation coefficient, *MCID* minimal clinically important difference, *MWPC* meaningful within-person 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, *SF-36v2®* 36-Item Short Form Survey Version 2, *UC* ulcerative colitis, *UC-SQ* Ulcerative Colitis Symptoms Questionnaire, *WPAI: CD* Work Productivity and Activity Impairment Questionnaire for Crohn's Disease

[a]Cronbach's alpha at baseline

school). The most frequently reported CD-related fatigue impacts were lack of motivation ($n = 14$, 46.7%), limited social interactions ($n = 11$, 36.7%), difficulty starting things ($n = 11$, 36.7%), and limitation to household chores ($n = 11$, 36.7%). The most frequently reported UC-related impacts were limited physical activity ($n = 18$, 56.3%), limitation to household chores ($n = 12$, 37.5%), limited social interactions ($n = 10$, 30.3%), and limited productivity at work or school ($n = 9$, 28.1%).

Table 3 Patient interviews: Participant-and clinician-reported demographic and health information (CD and UC)

Characteristic	CD total sample ($N = 30$) n (%)[a]	UC total sample ($N = 33$) n (%)[a]
Age, in years[b]		
Range (minimum–maximum)	15.1–75.4	15.1–80.1
Mean (SD)	36.6 (19.2)	38.3 (19.8)
Sex[b]		
Female	16 (53.3%)	17 (51.5%)
Male	14 (46.7%)	16 (48.5%)
Spanish/Hispanic/Latino[c]		
Not Spanish/Hispanic/Latino	24 (80.0%)	32 (97.0%)
Mexican/Mexican American, Chicano	5 (16.7%)	–
Hispanic, unspecified	1 (3.3%)	–
Puerto Rican	–	1 (3.0%)
Race[c]		
White	22 (73.3%)	24 (72.7%)
Hispanic	6 (20.0%)	–
Black or African American	2 (6.7%)	7 (21.2%)
Asian	–	1 (3.0%)
Unspecified	–	1 (3.0%)
Education (adults)[c;d]		
High school diploma or GED	2 (6.7%)	3 (9.1%)
Some college or associate degree	8 (26.7%)	5 (15.2%)
College or university degree	5 (16.7%)	10 (30.3%)
Graduate or professional degree	4 (13.3%)	4 (12.1%)
Other: Technical school	1 (3.3%)	–
Education (adolescents)[c;d]		
9th grade	2 (6.7%)	–
10th grade	1 (3.3%)	2 (6.1%)
11th grade	2 (6.7%)	2 (6.1%)
12th grade	5 (16.7%)	7 (21.2%)
Clinician-reported severity of CD or UC (moderate or severe)[b]		
Moderate	19 (63.3%)	19 (57.6%)
Severe	11 (36.7%)	14 (42.4%)
Current medication use[b]		
5-ASA (e.g., mesalamine, sulfasalazine)	10 (33.3%)	20 (60.6%)
Advanced therapy (e.g., biologics)	11 (36.7%)	13 (39.4%)
Immunomodulators (e.g., azathioprine or 6-mercaptopurine)	7 (23.3%)	5 (15.2%)
Corticosteroids	13 (43.3%)	21 (63.6%)
Time since diagnosis (in years)[b]		
Range (minimum–maximum)	1–45	1-35
Mean (SD)	9.7 (10.2)	6.0 (5.8)
Other health conditions[c;e]		
None	20 (66.7%)	15 (45.5%)
Depression/anxiety	3 (10.0%)	2 (6.1%)
Diabetes: Type II	2 (6.7%)	2 (6.1%)
High blood pressure	2 (6.7%)	7 (21.2%)
Rheumatoid arthritis	2 (6.7%)	–
High cholesterol	1 (3.3%)	5 (15.2%)
Unspecified	–	6 (18.2%)
Disease extent for UC only[f]		
E1 (proctitis)	–	24 (72.7%)

Table 3 (continued)

Characteristic	CD total sample ($N = 30$) n (%)[a]	UC total sample ($N = 33$) n (%)[a]
E2 (distal to splenic flexure)	–	12 (36.4%)
E3 (proximal to splenic flexure)	–	16 (48.5%)

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

[a]Unless otherwise indicated

[b]Clinician-reported information

[c]Participant-reported information

[d]Not mutually exclusive

[e]Other health conditions were reported by a one participant with CD or UC each: arthritis, cancer (lung cancer), fibromyalgia, ovarian cysts, thyroid disease, uterine fibroids, Wolff–Parkinson–White syndrome, migraine headaches

[f]Extent of disease based upon endoscopy findings: (E1: disease limited to rectum with or without sigmoid involvement; E2: UC present in descending colon up to, but not proximal to, splenic flexure; E3: evidence of UC proximal to the splenic flexure) [43]

3.1.3 Cognitive Debriefing of FACIT–Fatigue

During the interviews, participants completed the FACIT–Fatigue and, across items, 40.0–90.0% of participants with CD and 33.3–84.8% of participants with UC reported experiencing the assessed concepts within the recall period (Supplementary Table 2 in the Online Resource 1: Supplementary tables and figures). During the cognitive debriefing of the FACIT–Fatigue, ≥ 70.0% of CD participants and ≥ 93.0% of UC participants interpreted the instructions, recall period, items, and response options as intended, demonstrating the readability and comprehensibility of the questionnaire. The primary interpretation issue for CD was that six participants ($n = 6$, 20%), mostly adolescents ($n = 5/6$, 83.3%), reported being unfamiliar with the term "fatigue" or did not interpret the concept as intended (e.g., interpreted as nausea, dizziness, pain, sickness, or the act of taking a break/resting). Participants who did not understand the word "fatigue" were ultimately provided a definition for the purpose of understanding whether it was relevant experience. Overall, all CD and UC participants ($n = 30$ and $n = 33$, 100.0%, respectively) reported that the FACIT–Fatigue items measured concepts that are relevant to their experience, which was further supported by the spontaneous descriptions provided in the concept elicitation portion of the interviews. Participants were asked to report whether anything relevant to their fatigue experience was missing from the questionnaire, while there were multiple suggestions, the only concept mentioned by more than one person per condition was mental health/mental fatigue (reported by $n = 6$ with CD and $n = 1$ with UC)—this was described as an impact to one's mental "state" or "mental health effects" due to fatigue which can lead to changes in one's thinking process or mood.

3.2 CD Psychometric Evaluation and Score Interpretation

A total of 850 patients from the ADVANCE study were included in the psychometric and score interpretation analysis. Participants' ages ranged from 16 to 79 years [mean = 37.5 years; standard deviation (SD) = 13.3 years]; slightly less than half (45.9%) of the sample was female (Table 4).

3.2.1 CD: Quality of Completion and Score Distribution

Quality of completion for the FACIT–Fatigue for the CD psychometric analysis population was high. At least 95.0% of participants had complete data across all analysis timepoints. The mean FACIT–Fatigue total score was 25.09 at baseline ($n = 836$, SD = 11.29), 31.07 ($n = 825$, SD = 11.81) at week 4, and 34.98 ($n = 778$, SD = 11.87) at week 12. While item 10 (too tired to eat) and item 11 (need help doing usual activities) demonstrated floor effects at baseline (> 40% of participants endorsing the lowest option, "not at all," indicating no experience of the symptom within the recall period), improvement for these impacts was still demonstrated over time for the sample.

3.2.2 CD: Inter-item and Item-Total Correlations

Inter-item correlations using data from week 12 were moderate to strong ($r = 0.32$–0.90). The strongest correlation was between item 5 (trouble starting things) and item 6 (trouble finishing things) ($r = 0.90$ at week 12). Inter-item correlations for the ADVANCE study at week 12 are summarized in Supplementary Table 3 in the Online Resource 1: Supplementary tables and figures. For item-total correlations, the

magnitude between each item in the FACIT–Fatigue scale and the FACIT–Fatigue total score across analysis timepoints ranged between $r = 0.57$ and $r = 0.92$.

3.2.3 CD: Reliability

Score reliability for the FACIT–Fatigue questionnaire was assessed in two ways: internal consistency reliability and test-retest reliability. For the first type of reliability, Cronbach's α for the FACIT–Fatigue total score ranged from 0.86 to 0.88 from baseline to week 12, exceeding standard thresholds for acceptable internal consistency. Removal of any item did not markedly improve internal consistency reliability.

Among participants who were considered "stable" (i.e., those that either chose the same score on the PGIS for disease activity at baseline and week 4 or week 4 and week 12, or selected "no change" on the PGIC at week 4 and week 12, depending on the timepoint being evaluated), the ICCs for the FACIT–Fatigue total score ranged from 0.63 [95% confidence interval (CI) 0.50–0.72] to 0.73 (95% CI, 0.64–0.79). Interpretively, the ICCs ranged from slightly below to minimally above the a priori-defined acceptable threshold of 0.70.

3.2.4 CD: Convergent and Discriminant Validity

The correlations between the FACIT–Fatigue total score and scores on almost all concurrent measures were either as strong as or stronger than expected in the correct directions, demonstrating evidence of convergent validity. Weaker correlations were hypothesized between the FACIT–Fatigue and the measures used for discriminant validity analyses (EQ-5D-5L, SF-36v2®, IBDQ, and WPAI: CD), but moderate relationships were observed, indicating fatigue is more strongly related to the constructs assessed by the discriminant measures than expected. Detailed results are presented in Table 5, summarizing the hypothesized relationships between the FACIT–Fatigue and other measures and the observed correlations between them.

3.2.5 CD: Known-Groups Analysis

FACIT–Fatigue scores were strongly differentiated between clinically distinct groups on all measures, as expected. More specifically, FACIT–Fatigue total scores demonstrated a 10.6–11.2 point difference between groups classified as remission versus nonremission on the CDAI and a 14.9–16.6 point difference in remission versus nonremission groups using the IBDQ, and there was a monotonic decrease in the total score by PGIS group as severity improved; all comparisons were statistically significant ($p < 0.001$) (Table 6).

Table 4 Clinical trial sample demographics (ADVANCE and U-ACHIEVE)

Characteristic	CD	UC
Age (years)		
N	850	248
Mean (SD)	37.5 (13.3)	42.3 (14.1)
Median	34.0	41
Min–max	16.0–79.0	18.0–75.0
Gender (*n*, %)		
Male	390 (45.9%)	99 (39.9%)
Female	460 (54.1%)	149 (60.1%)

max maximum, *min* minimum, *SD* standard deviation

3.2.6 CD: Sensitivity to Change

Moderate-to-strong correlations were observed (0.430–0.701) between the FACIT–Fatigue change score and change scores on concurrent measures from baseline to week 12, except for the EQ-5D-5L mobility and self-care domains and the WPAI: CD work-time-missed domain. These three domains were weakly correlated with the FACIT–Fatigue change score (0.287, 0.236, and 0.269, respectively). Stronger correlations were observed between FACIT–Fatigue change scores and conceptually similar measures such as the SF-36v2 Physical Component Summary and IBDQ item 2 (0.624, 0.658, and 0.701, respectively).

3.2.7 CD: Interpretation of Scores—Anchor-Based Methods and Supportive Analyses

For CD, FACIT–Fatigue total score improved in parallel with each PGIS and PGIC anchor level. One and two point improvements on the PGIS generated a total score change of 6.17–10.03 on the FACIT–Fatigue (Table 7). Anchor groups on the PGIC ("minimally improved" or "much improved") generated a 6.83–13.63 point improvement in FACIT–Fatigue total score (Table 8).

Considering additional evidence from empirical cumulative distribution function (eCDF) curves and probability distribution function (PDF) curves, it is reasonable to conclude that a change on the FACIT–Fatigue of 7–10 points may indicate meaningful improvement (i.e., an estimate of MWPC). Based on receiver operating characteristic (ROC) curves, a 9 point change in the total score is recommended as a threshold for determining meaningful improvement

Table 5 Spearman correlation coefficients between FACIT–Fatigue total score and other assessments at baseline, week 4, and week 12 for the ADVANCE study for CD

Instrument/score	Hypothesized relationship to FACIT–Fatigue	FACIT–Fatigue total score		
		Baseline ($N = 850$)	Week 4 ($N = 841$)	Week 12 ($N = 819$)
CDAI	++	−0.31 (++)	−0.48 (++)	−0.56 (++)
PGIS	++	−0.58 (++)	−0.62 (++)	−0.64 (++)
IBDQ total score	++	0.72 (+++)	0.77 (+++)	0.82 (+++)
SF-36v2® Physical component summary score	++	0.59 (++)	0.63 (++)	0.70 (++)
EQ-5D-5L Mobility	++	− 0.44 (++)	− 0.42 (++)	− 0.43 (++)
EQ-5D-5L Self-care	++	−0.33 (++)	−0.31 (++)	−0.32 (++)
EQ-5D-5L Usual activities	++	−0.58 (++)	−0.61 (++)	−0.61 (++)
WPAI: CD presenteeism (impairment at work/reduced on-the-job effectiveness)	++	−0.47 (++)	−0.59 (++)	−0.57 (++)
WPAI: CD work productivity loss (overall work impairment/absenteeism plus presenteeism)	++	−0.47 (++)	−0.58 (++)	−0.52 (++)
WPAI: CD activity impairment	++	−0.60 (++)	−0.66 (++)	−0.66 (++)
EQ-5D-5L anxiety/depression	+	−0.41 (++)	−0.54 (++)	−0.53 (++)
EQ-5D-5L Pain/discomfort	+	−0.51 (++)	−0.55 (++)	−0.60 (++)
EQ-5D-VAS	+	0.53 (++)	0.61 (++)	0.64 (++)
WPAI: CD absenteeism	+	−0.33 (++)	−0.41 (++)	−0.38 (++)
SF-36v2® mental component summary score	+	0.57 (++)	0.63 (++)	0.73 (++)

CD Crohn's disease, *CDAI* Crohn's Disease Activity Index, *EQ-5D-5L* Five-level EQ-5D, *EQ-5D-VAS* EQ-5D Visual Analog Scale, *FACIT–Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *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)

(Supplementary Table 4 and Supplementary Figs. 1–3 in the Online Resource 1: Supplementary tables and figures).

3.3 UC Psychometric Evaluation and Score Interpretation

A total of 248 patients from the U-ACHIEVE study were included in the psychometric analysis. Participants' ages ranged from 18 to 75 years (mean = 42.3 years, SD = 14.1 years); 60% of the sample was female (Table 4).

3.3.1 UC: Quality of Completion and Score Distribution

Quality of completion for the UC psychometric analysis population was high. At least 92.6% of the population had complete data across analysis timepoints. The mean FACIT–Fatigue total score was 28.7 at baseline ($n = 238$, SD = 11.7) and 38.4 ($n = 214$, SD = 11.4) at week 8.

In general, respondents used the entire range of the response scale for the FACIT–Fatigue items across assessment timepoints, and scores trended toward improvement over time.

3.3.2 UC: Inter-item and Item-Total Correlations

Week 8 correlations between FACIT–Fatigue items were moderate to strong in magnitude (from 0.49 to 0.94). Similar to the CD results, the highest correlations were between items 5 (trouble starting things) and 6 (trouble finishing things) for all timepoints. The lowest correlations were between items 8 (usual activities) and 9 (sleep during the day), which were negatively correlated due to the reverse score direction of the items. Item-total correlations across analysis timepoints ranged between 0.52 and 0.86, exceeding the threshold for acceptable item-total correlations of ≥ 0.3 [26].

3.3.3 UC: Reliability

Overall, Cronbach's α ranged from 0.94 to 0.96 from baseline to week 8, exceeding the standard thresholds for acceptable internal consistency. Removing items for the UC dataset did not improve the α coefficient at any of the timepoints.

Among participants who (1) selected "no change" on the PGIC at week 2 or (2) selected the same response at both baseline and week 2 on UC-SQ item 6 (tired, lacking energy

Table 6 Known-groups comparisons for FACIT–Fatigue total score at week 12 for the ADVANCE study for CD ($N = 841$)

Comparison group	*n*	Mean (SD)	Median	*p*-Value
CDAI				
Remission: CDAI < 150	326	41.41 (8.50)	43.50	< 0.001*
Non-remission: CDAI ≥ 150	429	30.22 (11.72)	31.00	
IBDQ total score				
IBDQ total score ≥ 170 remission	380	43.45 (6.17)	45.00	< 0.001*
IBDQ total score <170 nonremission	398	26.88 (10.24)	27.00	
PGIS				
0 Absent: No symptoms	67	46.45 (4.45)	47.00	< 0.001*
1 Minimal: Can be easily ignored without symptoms	203	42.04 (8.17)	44.00	
2 Mild: Can be ignored with effort	163	36.02 (9.23)	37.00	
3 Moderate: Cannot be ignored but does not influence my daily activities	123	34.59 (9.39)	36.00	
4 Moderately severe: Cannot be ignored and occasionally limits my daily activities	129	27.50 (9.67)	27.00	
5 Severe: Cannot be ignored and often limits my concentration on daily activities	66	21.02 (9.89)	19.50	
6 Very severe: Cannot be ignored and markedly limits my daily activities	22	15.86 (12.65)	12.00	

CD Crohn's disease, *CDAI* Crohn's Disease Activity Index, *FACIT–Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *IBDQ* Inflammatory Bowel Disease Questionnaire, *SD* standard deviation, *WPAI: CD* Work Productivity and Activity Impairment Questionnaire: Crohn's Disease

* p-value <0.05

Table 7 Anchor-based estimates FACIT–Fatigue score by PGIS-stratified anchor categories from baseline to week 12 for the ADVANCE study for CD

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]
FACIT–Fatigue total score	Most improved (4 point or greater improvement)	120	20.09 (10.44)	41.90 (8.95)	21.81 (10.54)	< 0.001*	< 0.001*
	Much improved (3 point improvement)	158	25.62 (10.55)	39.82 (8.74)	14.20 (9.61)	< 0.001*	
	More improved (2 point improvement)	**148**	**25.87 (10.61)**	**35.90 (10.45)**	**10.03 (8.88)**	**< 0.001***	
	Improved (1 point improvement)	**161**	**26.14 (11.23)**	**32.30 (11.26)**	**6.17 (7.74)**	**< 0.001***	
	No change	118	26.64 (12.18)	27.54 (12.85)	0.90 (7.64)	0.042*	
	Worsened (1 point or greater deterioration)	39	25.08 (11.99)	25.38 (10.89)	0.31 (11.69)	0.630	

CD Crohn's disease, *FACIT–Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *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 (seven level) response

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

Bolding represents the primary PGIS anchor definitions for evaluation of the change in the FACIT-Fatiguescore: "1 point improvement" and "2 point improvement"

* p-value <0.05

during the past week), the FACIT–Fatigue total showed adequate test–retest reliability (ICC = 0.908 and ICC = 0.945 for PGIC and UC-SQ item 6 definitions of stability, respectively). All items demonstrated acceptable test–retest reliability for one or both stability definitions.

3.3.4 UC: Convergent and Discriminant Validity

Among the UC population, the correlations between the FACIT–Fatigue total score and scores on almost all concurrent measures were either as strong as or stronger than

Table 8 Anchor-based estimates FACIT–Fatigue score by PGIC-stratified anchor categories from baseline to week 12 for the ADVANCE study for CD

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]
FACIT–Fatigue total score	Very much improved (PGIC = 1)	118	26.41 (11.90)	44.44 (7.50)	18.03 (11.54)	< 0.001*	< 0.001*
	Much improved (PGIC = 2)	**288**	**24.77 (10.76)**	**38.40 (9.58)**	**13.63 (10.20)**	**< 0.001***	
	Minimally improved (PGIC = 3)	**192**	**25.33 (11.28)**	**32.17 (10.34)**	**6.83 (9.34)**	**< 0.001***	
	No change (PGIC = 4)	110	24.75 (11.96)	27.60 (12.31)	2.85 (8.23)	< 0.001*	
	Worsened (PGIC > 4)	52	23.44 (11.80)	22.21 (11.23)	−1.23 (8.53)	0.451	

CD Crohn's disease, *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

Bolding represents the primary PGIC anchor definitions for evaluation of the change in the FACIT-Fatiguescore: "Much Improved" or "Minimally improved"

* p-value <0.05

expected in the correct directions, demonstrating evidence of convergent validity. Exceptions were the self-care domain of the EQ-5D-5L and the Adapted Mayo Score, which had weaker than expected correlation with the FACIT–Fatigue total score in the UC sample. Weak correlations were hypothesized between the FACIT–Fatigue and the measures used for discriminant validity analyses (EQ-5D-5L, SF-36v2®, IBDQ, and WPAI: UC), but moderate relationships were observed, indicating fatigue is more strongly related to the constructs assessed by the discriminant measures than expected. Detailed results are presented in Table 9, summarizing the hypothesized relationships between the FACIT–Fatigue and other measures, and the observed correlations between them.

3.3.5 UC: Known-Groups Analysis

FACIT–Fatigue total scores demonstrated a 0.9–3.6 point difference between groups classified as remission versus nonremission on the Adapted Mayo Score, which was not statistically significant at week 2 but was statistically significant at week 8 (Table 10).

3.3.6 UC: Sensitivity to Change

Moderate-to-strong correlations were observed (0.45–0.76) between the FACIT–Fatigue change score and change scores on concurrent measures, except for the EQ-5D-5L mobility domain, EQ-5D-5L self-care domain, and WPAI:UC work-time-missed domains. These were weakly correlated with the FACIT–Fatigue change score (0.20, 0.20, and 0.35, respectively). Stronger correlations were observed between FACIT–Fatigue change scores and conceptually similar measures such as the SF-36v2 Physical Component Summary and IBDQ Total Score (0.61 and 0.76, respectively).

3.3.7 UC: Interpretation of Scores—Anchor-Based Methods and Supportive Analyses

The FACIT–Fatigue total score improved in parallel with each PGIC anchor level. Anchor groups on the PGIC ("minimally improved" or "much improved") generated a total score change of 4.24–9.33 points for phase 2b U-ACHIEVE substudy 1 (Table 11).

Considering all additional evidence (results for eCDF, PDF, and ROC curves are in Online Resource 1: Supplementary tables and figures), a change on the FACIT–Fatigue between 4 and 9 points can be recommended as meaningful improvement thresholds (i.e., an estimate of MWPC). Results are presented in Supplementary Table 5 and Supplementary Figs. 4–6 in the Online Resource 1: Supplementary tables and figures.

4 Discussion/limitations

This study demonstrated the importance and relevance of fatigue among IBD patients as well as the validity of the FACIT–Fatigue in measuring fatigue in moderate-to-severe IBD. To the knowledge of the authors, these were the first studies to thoroughly evaluate the content validity of the FACIT–Fatigue among adolescent and adult patients with moderately-to-severely active CD and UC separately.

Table 9 Spearman correlation coefficients between FACIT–Fatigue total score and other assessments at baseline, week 2, and week 8 for UC for U-ACHIEVE substudy 1 for UC

Assessment/score	Hypothesized relationship to FACIT–Fatigue	FACIT–Fatigue total score		
		Baseline ($N = 248$)	Week 2 ($N = 248$)	Week 8 ($N = 231$)
Adapted Mayo Score	+	−0.19 (+)	–	−0.38 (++)
IBDQ Bowel symptoms	+	0.61 (++)	0.61 (++)	0.59 (++)
IBDQ Systemic symptoms	+	0.78 (+++)	0.80 (+++)	0.73 (+++)
IBDQ Emotional function	+	0.68 (++)	0.71 (+++)	0.73 (+++)
IBDQ Social function	+	0.64 (++)	0.65 (++)	0.66 (++)
EQ-5D-5L Mobility	++	−0.34 (++)	−0.36 (++)	−0.40 (++)
EQ-5D-5L Self-care	++	−0.26 (+)	−0.27 (+)	− 0.30 (+)
EQ-5D-5L Usual activities	++	−0.66 (++)	−0.68 (++)	−0.70 (+++)
EQ-5D-5L Pain/discomfort	+	−0.58 (++)	−0.57 (++)	−0.59 (++)
EQ-5D-5L Anxiety/depression	+	−0.55 (++)	−0.58 (++)	−0.54 (++)
EQ-5D-VAS	+	0.60 (++)	0.63 (++)	0.71 (+++)
SF-36v2® Physical component summary	+	0.62 (++)	0.64 (++)	0.65 (++)
SF-36v2® Mental component summary	+	0.70 (+++)	0.67 (++)	0.73 (+++)
UC-SQ	+	−0.68 (++)	−0.66 (++)	−0.66 (++)
WPAI: UC absenteeism (work time missed)	+	−0.45 (++)	− 0.43 (++)	−0.32 (++)
WPAI: UC presenteeism (impairment at work)	+	−0.58 (++)	−0.71 (+++)	−0.66 (++)
WPAI: UC overall work impairment	+	−0.63 (++)	−0.70 (+++)	−0.62 (++)
WPAI: UC activity productivity	+	−0.58 (++)	−0.65 (++)	−0.72 (+++)

EQ-5D-5L Five-level EQ-5D, *EQ-5D-VAS* EQ-5D Visual Analog Scale, *FACIT–Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *IBDQ* Inflammatory Bowel Disease Questionnaire, *SF-23v2*® 36-Item Short-Form Survey, version 2, *UC* ulcerative colitis, *UC-SQ* Ulcerative Colitis Symptoms Questionnaire, *WPAI: UC* Work Productivity and Activity Impairment Questionnaire: Ulcerative Colitis

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

Results from the qualitative study provide details on the experience of fatigue and evidence supporting the use of the FACIT–Fatigue in these populations.

Specifically, during concept elicitation and cognitive debriefing interviews, all but one participant [i.e., over 98% of participants ($n = 62/63$)] reported experiencing fatigue. Participants reported cognitive, physical, social, work/school, sleep, leisure, emotional, and relationship impacts, as well as impacts to activities of daily living due to fatigue. Patients rated fatigue above a 6.0 on a scale from 0 (not at all bothersome) to 10 (most bothersome). In general, patients interpreted the FACIT–Fatigue content as intended and reported the questionnaire was relevant and comprehensive. However, a subset of CD patients ($n = 6/63$, 9.5%, mostly adolescents) reported being unfamiliar with the term "fatigue" (although upon receiving clarification, all of them reported experiencing it). This result poses a limitation to the content validity of the questionnaire for use with adolescents. Researchers may consider providing a definition of fatigue when administering the FACIT–Fatigue, especially within adolescent populations. If individuals are not provided with a definition of the term, there is risk that they may not be responding about the intended concept. This study was limited in sample size and greater research may be needed to better understand the extent of adolescent understanding of fatigue and whether it may vary between CD and UC.

Further, analyses using two clinical trial datasets demonstrated that the scores on the FACIT–Fatigue had (1) good internal consistency and acceptable test–retest reliability and (2) acceptable convergent and discriminant validity, known-groups results, and sensitivity to change. These results aligned with findings from Tinsley et al. [17] where the reliability and validity of the FACIT–Fatigue was evaluated among adults with CD or UC. While floor effects were observed for items 10 and 11 among CD patients at baseline, improvements in item-level scores still occurred over time. It may be that the floor effects observed at baseline do not indicate restriction in the FACIT–Fatigue response scale but likely reflect participants' actual levels of fatigue-related impact. Overall, although the FACIT–Fatigue was not developed for use in IBD, results from this research provide evidence that fatigue is relevant and highly important to CD and UC patients and that the FACIT–Fatigue has sound psychometric properties in this patient population.

This study was also the first to summarize thresholds for MWPC for the FACIT–Fatigue total score for patients with moderately-to-severely active CD or UC and provides

Table 10 Known-groups comparisons for FACIT–Fatigue total score at week 2 and week 8 for U-ACHIEVE substudy 1 for UC

Comparison group	*N*	Mean (SD)	Median	*p*-Value[a]
Week 2 ($N = 248$)				
Adapted Mayo Score				
Clinical remission	28	34.5 (11.7)	34.5	0.378
Nonremission	173	33.6 (12.1)	36.0	
UC-SQ Item 6. During the past week, have you felt tired or lacking energy?				
Not at all	12	49.3 (1.6)	49.5	< 0.001*
A little bit	37	41.7 (7.0)	44.0	
Somewhat	28	34.4 (7.7)	35.0	
Quite a bit	23	23.2 (7.6)	26.0	
Very much	16	12.9 (7.7)	13.5	
IBDQ Item 2. How often has the feeling of fatigue or of being tired and worn out been a problem for you during the last 2 weeks?				
All of the time	19	13.1 (7.0)	13.0	< 0.001*
Most of the time	35	22.4 (8.8)	22.0	
A good bit of the time	45	27.8 (8.1)	27.0	
Some of the time	52	36.4 (7.7)	38.0	
A little of the time	49	41.4 (7.0)	43.0	
Hardly any of the time	29	45.6 (4.5)	47.0	
None of the time	9	45.0 (7.8)	49.0	
Week 8 ($N = 248$)				
Adapted Mayo Score				
Clinical remission	27	41.6 (10.0)	44.0	0.041
Nonremission	167	38.0 (11.5)	42.0	
UC-SQ Item 6. During the past week, have you felt tired or lacking energy?				
Not at all	25	49.0 (4.1)	50.0	< 0.001*
A little bit	55	41.9 (5.3)	42.0	
Somewhat	22	37.7 (8.7)	39.5	
Quite a bit	13	26.5 (7.9)	24.0	
Very much	10	16.6 (10.8)	11.5	
IBDQ Item 2. How often has the feeling of fatigue or of being tired and worn out been a problem for you during the last 2 weeks?				
All of the time	11	17.0 (13.6)	12.0	< 0.001*
Most of the time	15	21.2 (9.3)	19.0	
A good bit of the time	28	30.7 (9.4)	30.0	
Some of the time	41	39.1 (7.1)	39.0	
A little of the time	54	40.9 (6.6)	42.0	
Hardly any of the time	43	45.6 (5.1)	47.0	
None of the time	22	49.4 (3.2)	50.0	

Clinical remission on the Adapted Mayo is defined as having a stool frequency subscore of 0 or 1 and not greater than the baseline score and rectal bleeding subscore of 0 and endoscopic subscore of 0 or 1; nonremission is defined as individuals not in the remission state

FACIT–Fatigue Functional Assessment of Chronic Illness Therapy–Fatigue, *IBDQ* Inflammatory Bowel Disease Questionnaire, *SD* standard deviation, *UC* ulcerative colitis, *UC-SQ* Ulcerative Colitis Symptoms Questionnaire

[a] *p*-Values are from a Mann–Whitney *U* test for comparison of means between two groups and a Kruskal–Wallis test for more than two groups

* p-value <0.05

guides for interpreting within-patient score changes in this population. A variety of quantitative methods were employed to generate reliable estimates, and during the qualitative interviews, patients described meaningful changes in fatigue, which provided context to aid in the interpretation of quantitative MWPC estimates. MWPC estimates differed by condition, and the threshold was higher for the analyses of the CD clinical trial data compared with the UC data (7–10 point improvement for CD and 4–9 point improvement for UC). This aligns with the qualitative results; the degree of bother of fatigue was generally higher among CD patients than UC patients, which may explain such differences. The ranges of MWPC estimates for the FACIT–Fatigue can be used as a guide for future clinical research for each condition, but it should be noted that if the target patient population or other trial characteristics (e.g., number of study weeks, administration schedule) differs, MWPC estimates may be impacted as well. Therefore, these ranges are presented to a starting place or aid in developing endpoints based on the FACIT–Fatigue scores for future research in CD and UC. After triangulating all evidence including ROC and eCDF graphs, an improvement of 9 and 5 points on the FACIT–Fatigue for CD and UC, respectively, could be considered for future clinical research in these target patient populations. These MWPC estimates were applied in the ADVANCE and phase 3 U-ACHIEVE trials, which demonstrated significantly greater portions of patients in treatment groups versus placebo were responders [27–30].

Data for psychometric evaluation of the FACIT–Fatigue came from two randomized clinical trials. The sample sizes for psychometric analyses were large, and data were of high quality. Patients had little missing data on the FACIT–Fatigue across timepoints, and data were captured in a standardized way during the trials.

While relying on data from clinical trials provided large samples of quality data, the test–retest analysis populations were limited because, due to the nature of clinical research, there were challenges to identifying a subgroup of "stable" patients (i.e., patients with unchanging disease status) across timepoints to include in analyses primarily due to the long duration between assessment timepoints and the administration of treatment. ICCs for CD ranged from slightly below to minimally above the threshold for acceptability. Given constraints in the available data, the ICCs for CD and UC should be interpreted with caution, as they likely underestimate the reliability of the scores. Using a fatigue-specific anchor (UC-SQ item 6) to identify stable patients in UC yielded higher ICCs. Additional research should evaluate the test–retest reliability of the FACIT–Fatigue in CD and UC patients using a larger sample of stable patients and define stability specific to fatigue.

Table 11 Anchor-based estimates FACIT–Fatigue score by PGIC stratified anchor categories for U-ACHIEVE substudy 1 for UC (baseline to week 8)

Score	PGIC at week 12[c]	*N*	Baseline Mean (SD)	Week 12 Mean (SD)	Mean change (SD)	Within-group *p*-value[a]	Between-groups *p*-value[b]
U-ACHIEVE substudy 1							
FACIT–Fatigue total score	Very much improved (PGIC = 1)	59	27.69 (12.91)	44.64 (8.96)	16.95 (11.40)	< 0.001*	< 0.001*
	Much improved (PGIC = 2)	**64**	**30.94 (10.66)**	**40.27 (7.73)**	**9.33 (8.44)**	**< 0.001***	
	Minimally improved (PGIC = 3)	**38**	**28.05 (11.55)**	**32.29 (12.74)**	**4.24 (7.66)**	**< 0.001***	
	No change (PGIC = 4)	29	31.34 (11.57)	35.17 (10.19)	3.83 (6.09)	0.002*	
	Worsened (PGIC = 5)	7	29.14 (12.38)	30.29 (11.93)	1.14 (6.36)	0.500	
	Much worsened (PGIC = 6)	8	24.88 (12.92)	22.38 (15.35)	−2.50 (4.90)	0.188	

FACIT–Fatigue Functional Assessment of Chronic Illness Therapy–Fatigue, *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

[c]There were no individuals who endorsed "Very much worsened" (PGIC = 7)

Bolding represents the primary PGIC anchor definitions for evaluation of the change in the FACIT-Fatiguescore: "Much Improved" or "Minimally improved"

* p-value <0.05

Additional paths forward for future research may include comparing the FACIT–Fatigue with other multi-item fatigue scales, evaluating the FACIT–Fatigue among individuals with mild disease, and evaluating the differences in how fatigue impacts quality of life between CD and UC.

5 Conclusions

This study demonstrated the relevance and importance of fatigue to patients' disease experience in moderately-to-severely active CD and UC. Cognitive debriefing results provided evidence that the FACIT–Fatigue can be interpreted as intended and has comprehensive coverage of fatigue and fatigue-related impacts among adults with moderately-to-severely active CD or UC; however, some caution may be needed when using this questionnaire with adolescents who may have difficulty understanding the word "fatigue." Psychometric analyses indicate that the scores generated by the FACIT–Fatigue demonstrate acceptable reliability, construct validity, and sensitivity to change. Further, a 7–10 point improvement for CD and a 4–9 point improvement for UC on the FACIT–Fatigue total score may represent meaningful improvements and may serve as context for future research in individuals with moderate-to-severe CD and UC.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41669-023-00419-w.

Acknowledgements The authors would like to thank Alan Shields, Roger Lamoureux, Emma Pranschke, Ryan Hoffman, Naomi Suminski, Doug Huntington, Jeff Mcdonald, and Alejandro Moreno-Koehler for their important contributions to the completion of this research.

Declarations

Authorship statement EVL is acting as the submission's guarantor. AbbVie and Adelphi Values authors collaborated on the conception of the study design, conduct of research, data collection, analysis, and interpretation of data. All authors had access to the data results and participated in the development, review, and approval of this manuscript. No honoraria or payments were made for authorship. All authors have approved the final version of the article, including the authorship list.

Declaration of personal interests Edward Loftus has received financial support for research from AbbVie, Bristol-Myers Squibb, Celgene, Genentech, Gilead, Gossamer Bio, Janssen, Pfizer, Receptos, Robarts Clinical Trials, Takeda, Theravance, and UCB; and received consultancy fees from AbbVie, Amgen, Arena, Boehringer Ingelheim, Bristol-Myers Squibb, Calibr, Celgene, Eli Lilly, Fresenius Kabi, Genentech, Gilead, Gossamer Bio, Iterative Scopes, Janssen, Morphic, Ono Pharma, Pfizer, Protagonist, Scipher Medicine, Sun Pharma, Surrozen, Takeda, and UCB. Ashwin Ananthakrishnan has received funding from the US National Institutes of Health, served on scientific advisory boards for Gilead and received research funding from Pfizer. Wan-Ju Lee, Yuri Sanchez Gonzalez, Kristina Fitzgerald, Kori Wallace, and Wen Zhou are full-time employees of AbbVie and may hold AbbVie stock and/or stock options. Leighann Litcher-Kelly, Sarah Ollis, and Sylvia Su are employed by Adelphi Values LLC, which received payment from AbbVie Inc. to support the research activities presented in this publication. Silvio Danese has received consultancy fees from AbbVie, Allergan, Amgen, AstraZeneca, Biogen, Boehringer Ingelheim, Celgene, Celltrion, Ferring, Gilead, Hospira, Janssen, Johnson & Johnson, MSD, Mundipharma, Pfizer, Roche, Sandoz,

Takeda, TiGenix, UCB, and Vifor Pharma. No honoraria or payments were made for authorship.

Declaration of funding interests This work was supported by AbbVie, Inc. AbbVie participated in the study design, research, data collection, analysis and interpretation of data, writing, reviewing, and approving the publication. Writing support was provided by Naomi Suminski and Emma Pranschke of Adelphi Values and funded by AbbVie, Inc.

Data availability statement The data that support the findings of this research may be available from AbbVie, but restrictions apply to the availability of these data, which were used under license for the current research, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of AbbVie.

Ethics approval Ethical approval for the interview study was obtained via a centralized independent review board. Ethical approval for the clinical trials was obtained separately.

Consent to participate All interview and clinical trial participants provided consent (or assent with parental/guardian consent, as appropriate).

Consent for publication (from patients/participants) Not applicable.

Code availability May be available upon reasonable request.

References

1. Danese S, Fiocchi C. Ulcerative colitis. N Engl J Med. 2011;365(18):1713–25.
2. National Institute of Diabetes and Digestive and Kidney Diseases. Ulcerative Colitis. 09/2014; https://www.niddk.nih.gov/health-information/digestive-diseases/ulcerative-colitis. Accessed 29 Mar 2022.
3. Panaccione R. Mechanisms of inflammatory bowel disease. Gastroenterol Hepatol. 2013;9(8):529–32.
4. Gohil K, Carramusa B. Ulcerative colitis and Crohn's disease. P T. 2014;39(8):576–7.
5. Villoria A, García V, Dosal A, et al. Fatigue in out-patients with inflammatory bowel disease: prevalence and predictive factors. PLoS ONE. 2017;12(7): e0181435.
6. Klusmann B, Fleer J, Trovote KA, et al. Trajectories of fatigue in inflammatory bowel disease. Inflamm Bowel Dis. 2021;27(12):1919–30.
7. D'Silva A, Fox DE, Nasser Y, et al. Prevalence and risk factors for fatigue in adults with inflammatory bowel disease: a systematic review with meta-analysis. Clin Gastroenterol Hepatol. 2022;20(5):994–1009.
8. Bager P, Befrits R, Wikman O, et al. Fatigue in out-patients with inflammatory bowel disease is common and multifactorial. Aliment Pharmacol Ther. 2012;35(1):133–41.
9. Borren NZ, van der Woude CJ, Ananthakrishnan AN. Fatigue in IBD: epidemiology, pathophysiology and management. Nat Rev Gastroenterol Hepatol. 2019;16(4):247–59.
10. Jelsness-Jørgensen LP, Bernklev T, Henriksen M, Torp R, Moum BA. Chronic fatigue is more prevalent in patients with inflammatory bowel disease than in healthy controls. Inflamm Bowel Dis. 2011;17(7):1564–72.
11. Nocerino A, Nguyen A, Agrawal M, Mone A, Lakhani K, Swaminath A. Fatigue in inflammatory bowel diseases: etiologies and management. Adv Ther. 2020;37(1):97–112.
12. FACIT Fatigue Scale (Version 4). 1997; http://www.ser.es/wp-content/uploads/2015/03/FACIT-F_INDICE.pdf. Accessed 29 Mar 2022.
13. 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.
14. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. J Pain Symptom Manage. 1997;13(2):63–74.
15. Meysami A, Yari H, Haghighat S, Montazeri A. The Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue Subscale): validity and reliability of the Iranian version. Oncol Res Treat. 2017;40(12):789–93.
16. FACIT.org. FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue. https://www.facit.org/measures/FACIT-F. Accessed 29 Mar 2022.
17. Tinsley A, Macklin EA, Korzenik JR, Sands BE. Validation of the functional assessment of chronic illness therapy-fatigue (FACIT-F) in patients with inflammatory bowel disease. AlimentPharmacol Ther. 2011;34(11–12):1328–36.
18. Conley S, Proctor DD, Jeon S, Sandler RS, Redeker NS. Symptom clusters in adults with inflammatory bowel disease. Res Nurs Health. 2017;40(5):424–34.
19. Williet N, Sarter H, Gower-Rousseau C, et al. Patient-reported outcomes in a French Nationwide Survey of inflammatory bowel disease patients. J Crohns Colitis. 2017;11(2):165–74.
20. Loftus EV, Feagan BG, Colombel JF, et al. Effects of adalimumab maintenance therapy on health-related quality of life of patients with Crohn's disease: patient-reported outcomes of the CHARM trial. Am J Gastroenterol. 2008;103(12):3132–41.
21. Farrell D, Artom M, Czuber-Dochan W, Jelsness-Jørgensen LP, Norton C, Savage E. Interventions for fatigue in inflammatory bowel disease. Cochrane Database Syst Rev. 2020;4(4):Cd012005.
22. Saraiva S, Cortez-Pinto J, Barosa R, et al. Evaluation of fatigue in inflammatory bowel disease—a useful tool in daily practice. Scand J Gastroenterol. 2019;54(4):465–70.
23. 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.
24. Sandborn WJ, Ghosh S, Panes J, et al. Efficacy of upadacitinib in a randomized trial of patients with active ulcerative colitis. Gastroenterology. 2020;158(8):2139-2149.e2114.
25. Ghosh S, Sanchez Gonzalez Y, Zhou W, et al. Upadacitinib treatment improves symptoms of bowel urgency and abdominal pain, and correlates with quality of life improvements in patients with moderate to severe ulcerative colitis. J Crohn's Colitis. 2021;15:2022.

26. Thompson B, Vacha-Haase T. Psychometrics is datametrics: the test is not reliable. Educ Psychol Measur. 2000;60(2):174–95.
27. Danese S, Tran J, D'Haens G, et al. OP08 The effects of maintenance therapy with upadacitinib on abdominal pain, bowel urgency, and fatigue in patients with moderately to severely active Ulcerative Colitis: Phase 3 U-ACHIEVE maintenance results. J Crohn's Colitis. 2022;16(Supplement_1):i008-i010.
28. Danese S, D'Haens G, Rubin D, et al. Patients with ulcerative colitis report improvements in abdominal pain, bowel urgency, and fatigue with 8-week upadacitinib treatment in two phase 3 trials: U-ACHIEVE and U-ACCOMPLISH. United Eur Gastroenterol J. 2021;9(S8):25.
29. Peyrin-Biroulet L, Louis E, Ghosh S, et al. DOP88 Effect of risankizumab on patient-reported outcomes in patients with Crohn's Disease who had an inadequate response or intolerance to conventional and/or biologic treatments: Results from phase 3 MOTIVATE and ADVANCE trials. J Crohn's Colitis. 2021;15(1):S119–21.
30. Peyrin-Biroulet L, Ghosh S, Lee SD, et al. Effect of risankizumab on health-related quality of life in patients with Crohn's disease: results from phase 3 MOTIVATE, ADVANCE and FORTIFY clinical trials. Aliment Pharmacol Ther. 2023;57(5):496–508.
31. 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.
32. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med. 1987;317(26):1625–9.
33. 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.
34. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4(5):353–65.
35. 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.
36. 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.
37. Wang L, Zhang Z, McArdle JJ, Salthouse TA. Investigating ceiling effects in longitudinal data analysis. Multivar Behav Res. 2009;43(3):476–96.
38. Hinkle DE, Jurs SG, Wiersma W. Applied statistics for the behavioral sciences. 2nd ed. Boston: Houghton Mifflin; 1988.
39. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(3):297–334.
40. Bland JM, Altman DG. Cronbach's alpha. BMJ. 1997;314(7080):572.
41. Weir J. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.
42. 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.
43. Basson M. How is ulcerative colitis (UC) defined based on the extent of disease? https://www.medscape.com/answers/183084-13754/how-is-ulcerative-colitis-uc-defined-based-on-the-extent-of-disease. Accessed 29 Mar 2022.

Authors and Affiliations

Edward V. Loftus Jr[1] · Ashwin N. Ananthakrishnan[2] · Wan-Ju Lee[3] · Yuri Sanchez Gonzalez[3] · Kristina A Fitzgerald[3] · Kori Wallace[3] · Wen Zhou[3] · Leighann Litcher-Kelly[4] · Sarah B. Ollis[4] · Sylvia Su[4] · Silvio Danese[5]

✉ Edward V. Loftus Jr
loftus.edward@mayo.edu

[1] Mayo Clinic College of Medicine and Science, Rochester, MN, USA

[2] Massachusetts General Hospital, Boston, MA, USA

[3] AbbVie Inc, Chicago, IL, USA

[4] Adelphi Values, Boston, MA, USA

[5] Humanitas University, Milan, Italy