Adv Ther (2025) 42:1044–1060
https://doi.org/10.1007/s12325-024-03081-8

ORIGINAL RESEARCH

The Urgency Numeric Rating Scale: Psychometric Evaluation in Adults with Crohn's Disease

Marla Dubinsky · Aisha Vadhariya · Sylvia Su · Xian Zhou · Frederick Durand · Claudine Clucas · Larissa Stassek · Ariane K. Kawata · Simon Travis

Received: August 17, 2024 / Accepted: November 22, 2024 / Published online: December 18, 2024

ABSTRACT

Introduction: Bowel urgency has recently been recognized as a Crohn's disease (CD) symptom that substantially impacts patients' quality of life. The Urgency NRS is a single-item patient-reported outcome measure assessing bowel urgency severity in the past 24 h (0–10 scale). We aimed to evaluate the psychometric properties of the Urgency Numeric Rating Scale (NRS) in adults with moderately to severely active CD and to estimate thresholds for meaningful improvement and bowel urgency remission.

Methods: Psychometric analyses used pooled data from the Phase 3 VIVID-1 study of mirikizumab, where participants with CD completed the Urgency NRS and other assessments. The Patient Global Rating of Severity (PGRS) and Patient Global Impression of Change (PGIC) were used as primary anchors to estimate Urgency NRS thresholds representing meaningful improvement and remission.

Results: The Urgency NRS showed good test–retest reliability in participants who were stable based on PGRS and PGIC. It was moderately correlated with similar assessments and weakly correlated with endoscopic/laboratory assessments. It differentiated between participant subgroups varying in disease severity and quality of life based on PGRS and other assessments. It was sensitive to change, as Urgency NRS improvements during the trial differed between most PGRS change and PGIC categories. A 3–5-point reduction on the Urgency NRS represented meaningful improvement and a score of ≤ 2 represented remission.

Conclusion: The Urgency NRS demonstrated strong psychometric properties in the VIVID-1 population of moderately to severely active CD. Analyses also suggested meaningful improvement and remission thresholds.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12325-024-03081-8.

M. Dubinsky
Dr. Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

A. Vadhariya (✉) · S. Su · F. Durand
Eli Lilly and Company, 893 S Delaware St., Indianapolis, IN 46225, USA
e-mail: vadhariya_aisha@lilly.com

X. Zhou
Syneos Health, Morrisville, NC, USA

C. Clucas
Evidera, London, UK

L. Stassek · A. K. Kawata
Evidera, Wilmington, NC, USA

S. Travis
Kennedy Institute, Translational Gastroenterology Unit and Biomedical Research Centre, University of Oxford, Oxford, UK

Trial Registration: Clinicaltrials.gov, NCT03926130. Registered 23 April 2019, https://clinicaltrials.gov/study/NCT03926130.

Keywords: Bowel urgency; Crohn's disease; Psychometric evaluation; Urgency Numeric Rating Scale

Key Summary Points

Why carry out this study?

Bowel urgency has recently been recognized as a common symptom of Crohn's disease (CD) that significantly impacts patients' lives

The Urgency Numeric Rating Scale (NRS) is a single-item patient-reported outcome measure assessing severity of bowel urgency in the previous 24 h on an 11-point scale

This study evaluated the psychometric properties of the Urgency NRS in moderately to severely active CD and estimated thresholds for meaningful change and remission for bowel urgency

What was learned from the study?

The Urgency NRS demonstrated reliability, validity, and responsiveness in a clinical trial of moderately to severely active CD; a 3–5-point reduction in the Urgency NRS score could indicate clinically meaningful improvement of bowel urgency in this patient population, and a score of ≤ 2 could represent remission

These findings support the suitability of the Urgency of NRS for assessing treatment benefits in patients with CD during clinical trials

INTRODUCTION

Crohn's disease (CD) is a chronic inflammatory bowel disease (IBD) that can affect any segment of the gastrointestinal tract [1, 2]. Common symptoms include diarrhea, abdominal pain and cramping, weight loss, fatigue, and fever [1, 2]. Up to half of patients may also experience extraintestinal manifestations, including in joints, skin, and eyes [2]. Treatment goals have evolved, and current consensus incorporates short-term (clinical response), medium-term (clinical remission and normalization of biomarkers), and long-term targets [endoscopic healing and normalized quality of life (QoL)] [2, 3]. Additionally, disease clearance has emerged as an overall goal, and there is an increasing focus on profiling patient determinants to ensure QoL [3–5].

Bowel urgency, the sudden and urgent need to have a bowel movement, is widely recognized to be among the most common and disruptive symptoms of ulcerative colitis (UC), another IBD [7–10]. As such, bowel urgency assessment is included in current clinical guidelines for UC [11, 12]. The multi-country CONFIDE survey recently highlighted the significant impact that bowel urgency and related fecal incontinence can have on patients with moderately to severely active CD [13, 14]. Unsurprisingly, patients identify bowel urgency control as one of the most valued treatment objectives [15–17]. However, clinical guidelines in CD do not incorporate bowel urgency assessment [18, 19], and this symptom has often been overlooked by physicians [7, 13, 20]. The impact of bowel urgency is exacerbated because it can also affect other aspects of daily life. For example, physical activity is known to have a positive impact on QoL, but CD symptoms like bowel urgency can make it difficult to participate in and experience benefits from physical activity [6].

Patient-reported outcome measures (PROs) capture direct input from patients about their symptom experience [21]. It is important to assess bowel urgency, which can only be accurately reported by patients themselves, using PRO measures that produce valid, reliable, and interpretable scores.

The 11-point Urgency Numeric Rating Scale (NRS) is a single-item PRO measure assessing bowel urgency severity over the past 24 h [22]. It was initially developed for use in UC [22] and its psychometric properties have been thoroughly studied in that patient population [23]. While qualitative evidence and a pilot quantitative

survey have subsequently supported the validity and reliability of the Urgency NRS for assessing bowel urgency in CD [24], its responsiveness in CD has not yet been confirmed and the Urgency NRS score representing bowel urgency remission and score change representing meaningful improvement have not been explored quantitatively. Further, the Urgency NRS measurement properties have not yet been confirmed with clinical trial data of CD or specifically in severely to moderately active CD. This study used data from a Phase 3 trial to evaluate these in moderately to severely active CD and to estimate thresholds for meaningful change and bowel urgency remission.

METHODS

Study Population and Design

The present analysis used data from VIVID-1 (NCT03926130), a Phase 3, randomized, active- and placebo-controlled study assessing the safety and efficacy of mirikizumab in adults (18–80 years old) with CD. Participants were required to have a confirmed diagnosis of CD for ≥ 3 months prior to baseline and moderately to severely active disease defined by average daily very soft or liquid stool frequency of ≥ 4 and/or average daily abdominal pain ≥ 2 (on a 4-point scale ranging from none to severe) at baseline with endoscopic evidence of mucosal inflammation based on a simple endoscopic score for Crohn's disease (SES-CD) score of ≥ 7 (or ≥ 4 for those with isolated ileal disease). Participants with intolerance, inadequate response, or loss of response to prior conventional or biologic therapies (or both) were randomly assigned in a 6:3:2 ratio to mirikizumab, ustekinumab, or placebo for a total of 52 weeks. The study was compliant with the Declaration of Helsinki and Council for International Organizations of Medical Sciences International Ethical guidelines and the International Conference on Harmonization guidelines on Good Clinical Practice. Informed consent forms and protocols received ethical review board approval before study initiation. Participants provided written informed consent before any procedures were performed.

Assessments

Urgency NRS

The Urgency NRS is a single-item PRO measure for rating the severity of bowel urgency experienced in the past 24 h using an 11-point NRS anchored at 0 = "no urgency" and 10 = "worst possible urgency." The Urgency NRS was completed daily from screening to Week 52 using an electronic diary. Weekly average scores for analyses were calculated by averaging the most recent 7 days in the 12 days prior to each visit, with at least 4 days of non-missing values. A missing data analysis conducted at the assessment level justified the use of this scoring rule, as up to 3 days of missing data when taking the weekly average did not impact score stability or significantly increase variance of the overall weekly averages.

Patient Global Rating of Severity (PGRS)

The PGRS [25] is a single-item instrument assessing participants' rating of their overall CD symptom severity over the past 24 h (1 = "none" to 6 = "very severe"). The PGRS was completed daily from screening visit to week 52. At baseline (week 0) and weeks 4, 12, 16, and 52, weekly average scores were calculated using the most recent 7 days in the 12 days prior to each visit, with at least 4 days of non-missing values, and they were rounded to the nearest integer for analysis.

Patient Global Impression of Change (PGIC)

The PGIC [25] is a single-item instrument designed to assess participants' rating of changes in CD symptoms at a given time point compared to how they were before they started taking the study drug (1 = "very much better," 4 = "no change," and 7 = "very much worse"). The PGIC was administered at weeks 4, 8, 12, and 52.

Inflammatory Bowel Disease Questionnaire (IBDQ)

The IBDQ is a 32-item instrument that measures four domains of participants' lives over the past 2 weeks: bowel function, systemic symptoms, emotional function, and social function. Item scores (1 = "a very severe problem" to 7 = "not a problem at all") are summed to compute a total score (range, 32–224, with higher scores indicating better QoL), as well as scores for each of the four domains. Following permission for use, the IBDQ was administered at baseline, week 12, and week 52. Analyses included the IBDQ total score, IBDQ Bowel Symptom domain score (calculated as the sum of items 1, 5, 9, 13, 17, 20, 22, 24, 26, and 29, per developer scoring instructions [26]), and item 1 score (which specifically measures bowel frequency: "How frequent have your bowel movements been during the last 2 weeks?"). IBDQ response was defined as a ≥ 16-point increase (improvement) from baseline in IBDQ total score, and IBDQ remission was defined as a total score of ≥ 170 [27, 28].

Crohn's Disease Activity Index (CDAI)

The CDAI is an 8-item instrument including three patient-reported items (abdominal pain, stool frequency, and general well-being) and five clinician-reported/laboratory items (physical signs and hematocrit). The patient-reported items were administered daily, and the clinician-reported/laboratory assessments were completed at all study visits except screening, week 2, and week 6. The CDAI score ranges 0–600, with a higher score indicating more severe disease. CDAI remission was defined as a total CDAI score of < 150 [29–31] and CDAI response was defined as a reduction of ≥ 100 points in the CDAI score versus baseline and/or CDAI remission [32, 33].

SES-CD

The SES-CD is a clinician-reported instrument based on four endoscopic variables assessed across five bowel segments (scale: 0–3 per variable and bowel segment) [34]. The sum of the scores for each endoscopic variable (0–11 for stenosis and 0–15 for the rest) leads to an overall range of 0–56, with higher scores indicating more severe disease. The SES-CD was completed at screening, week 12, and week 52.

Laboratory Assessments

The inflammatory markers high-sensitivity C-reactive protein (hsCRP) from blood samples and fecal calprotectin from stool samples were assessed at baseline and at weeks 4, 8 (only hsCRP), 12, 16, 28, 44, and 52.

Psychometric Analyses

The present psychometric analyses used individual participant data from VIVID-1 pooled across treatments. All variables were summarized descriptively. Statistical analyses were performed using SAS® version 9.4 (SAS Institute, Cary, NC, USA).

Test–retest Reliability

The test–retest reliability of the Urgency NRS was examined by calculating intraclass correlation coefficients (ICCs) between baseline and week 4 in subgroups of stable participants. ICCs were calculated using a two-way mixed effects model with patient as a random effect, time as a fixed effect, and no interaction [35]. The score differences between both timepoints were also calculated and compared using paired *t*-tests between the stable participants. Stable participants were defined as either those with an unchanged weekly average PGRS score between baseline and week 4 or those with a response of "no change" on the PGIC at week 4 (PGIC = 4). ICCs ≥ 0.70 indicated good test–retest reliability and ICCs > 0.90 indicated excellent reliability [36].

Convergent and Discriminant Validity

The convergent and discriminant validity of the Urgency NRS was evaluated at baseline, week 12, and week 52. Convergent validity was assessed by calculating Spearman correlations between the Urgency NRS and PGRS, IBDQ total score,

IBDQ Bowel Symptom domain, and CDAI. Discriminant validity was assessed by calculating Spearman correlations between the Urgency NRS and SES-CD, hsCRP, and fecal calprotectin. Based on their absolute values, correlations were considered weak ($|r|< 0.30$), moderate ($|r|$: 0.30 to < 0.70), or strong ($|r|\geq 0.70$) [37]. Moderate to strong correlations were expected between the Urgency NRS and the PGRS, IBDQ Bowel Symptom domain, the IBDQ total score, and CDAI. Weak correlations were expected between the Urgency NRS and SES-CD, hsCRP, and fecal calprotectin.

Known-groups Validity

The known-groups validity of the Urgency NRS was assessed at baseline, week 12, and week 52 by comparing Urgency NRS scores between subgroups based on PGRS categories (level of response), IBDQ Bowel Symptom domain score (median split), and IBDQ total score (median split). Average Urgency NRS scores were compared between individual subgroups using analysis of variance (ANOVA) with Scheffé's correction for post hoc pairwise comparisons [38]. Analyses included the Urgency NRS score as the dependent variable and the known-group variable as the independent variable. Participants with more severe CD symptoms based on these other measures were expected to report greater bowel urgency severity (higher Urgency NRS scores).

Ability to Detect Change (Responsiveness)

Spearman correlations for change scores from baseline to week 12 and week 52 were calculated between the changes in Urgency NRS scores and corresponding changes in potential anchor measures [PGRS, PGIC, IBDQ Bowel Symptom domain, IBDQ item 1 (bowel movement frequency), IBDQ total score, and CDAI score]. The responsiveness of Urgency NRS was also evaluated by comparing mean changes in the Urgency NRS scores from baseline to weeks 12 and 52 with changes between select pre-defined anchor groups in the same period using one-way analysis of covariance (ANCOVA) with Scheffé's correction for post hoc pairwise comparisons [38] and controlling for the Urgency NRS score at baseline. The anchor groups were pre-defined by PGRS average change, PGIC categories, IBDQ Bowel Symptom domain response (≥ 8-point increase from baseline), IBDQ response, IBDQ remission, CDAI response, and CDAI remission.

Thresholds for Meaningful Improvement and Remission

An anchor-based approach was used to estimate the Urgency NRS change from baseline that would constitute a meaningful within-patient improvement as well as the Urgency NRS score at week 12 or 52 that would correspond to bowel urgency remission (a state with "no to very mild" bowel urgency). The PGRS and PGIC were used as primary anchors. IBDQ item 1 (bowel movement frequency), IBDQ Bowel Symptom domain response, and CDAI response were used as supplemental anchors to determine the threshold for meaningful improvement. IBDQ item 1 (bowel movement frequency), IBDQ remission, and CDAI remission were used as supplemental anchors to determine the threshold for bowel urgency remission. Mean, standard deviation (SD), and percentile groups (10th, 25th [quartile 1], 50th [quartile 2, median], 75th [quartile 3], and 90th) for Urgency NRS scores and score changes were reported for each of the anchor groups.

Anchor-based methods were supplemented with cumulative distribution function (CDF) and probability density function (PDF) plots, estimated using kernel density estimation curves, to examine the distribution of Urgency NRS scores or score changes by PGRS and PGIC categories. The CDF or PDF were shown across a range of possible responder definitions defined by PGRS and PGIC.

RESULTS

Baseline Demographic and Clinical Characteristics

The present analysis included 1065 study participants. The participants' mean age was 36.2 years

(SD: 13.0). Seventy-one percent of the participants were white and 55% were male. At baseline, 31% of participants were using corticosteroids (Table 1).

Distribution of Assessment Scores

The mean Urgency NRS score at baseline was 6.6 (SD: 2.1). The mean score decreased to 4.2 (SD: 2.7) at week 12 and to 2.9 (SD: 2.5) at week 52 (Fig. 1), indicating improvement of bowel urgency severity. The full range of responses (0.0–10.0) was recorded at all timepoints, and no ceiling or floor effects were observed. Very few responses were missing: 0.1% at baseline, 0.5% at week 12, and 2.4% at week 52. Score distribution by PGRS at all time-points is presented in Figure S1 of the ESM. Descriptive statistics for all other assessments are presented in Table S1 of the ESM.

Test–retest Reliability

A total of 417 participants were defined as stable based on the PGRS. For these participants, the ICC for the Urgency NRS score between baseline and week 4 was 0.86. A total of 267 participants were defined as stable based on the PGIC. For these participants, the ICC was 0.78 (Table 2).

Convergent and Discriminant Validity

At baseline, the Urgency NRS score was moderately correlated with PGRS (r = 0.48), IBDQ Bowel Symptom domain (r = − 0.37), IBDQ total score (r = − 0.31), and CDAI (r = 0.48). The Urgency NRS score was weakly correlated with endoscopic (SES-CD [r = 0.12]) and laboratory assessments (hsCRP [r = 0.07] and fecal calprotectin [r = 0.03]) (Fig. 2; Table S2 of the ESM). At weeks 12 and 52, the correlations were larger overall than at baseline, but still within the same range. Relationships between the Urgency NRS score and other assessments were generally consistent with the pre-specified hypotheses. Of note, convergent and discriminant validity results were similar across disease locations (Table S3 of the ESM).

Table 1 Demographic and clinical characteristics at baseline

Characteristic	*n* = 1065
Age (years), mean (SD)	36.2 (13.0)
Sex, *n* (%)	
Male	587 (55.1)
Female	478 (44.9)
Race, *n* (%)	
White	753 (70.7)
Black or African American	23 (2.2)
Asian	264 (24.8)
American Indian/Alaska Native	6 (0.6)
Multiple	4 (0.4)
Missing	15 (1.4)
Age at CD diagnosis (years)	
Mean (SD)	29.3 (12.0)
< 17 years, *n* (%)	113 (10.6)
17 to 40 years, *n* (%)	749 (70.3)
> 40 years, *n* (%)	202 (19.0)
Missing, *n* (%)	1 (0.1)
Duration of CD, *n* (%)	
< 1 year	147 (13.8)
≥ 1 to < 5 years	392 (36.8)
≥ 5 years	525 (49.3)
Missing	1 (0.1)
Disease location, *n* (%)	
Ileal	113 (10.6)
Colonic	422 (39.6)
Ileal-colonic	530 (49.8)
Corticosteroid use, *n* (%)	325 (30.5)

CD Crohn's disease, *SD* standard deviation

Known-groups Validity

Participants with lower (better) PGRS scores tended to report lower Urgency NRS scores

(indicating less severe bowel urgency) at baseline, week 12, and week 52 (Fig. 3; Tables S4–S6 of the ESM). At baseline, significant pairwise comparisons were observed for PGRS responses of "mild" versus "moderate," "very mild," "mild," and "moderate" versus "severe," and "none" through "severe" versus "very severe" ($p < 0.05$) (Table S4 of the ESM). All pairwise comparisons at week 12 (all $p < 0.001$) and nearly all comparisons at week 52 ($p < 0.05$) were statistically significant (Tables S5 and S6 of the ESM).

Urgency NRS scores also tended to be lower among participants with higher IBDQ Bowel Symptom domain scores and among participants with higher IBDQ total scores (indicating better QoL) at all timepoints (Tables S7 and S8 of the ESM).

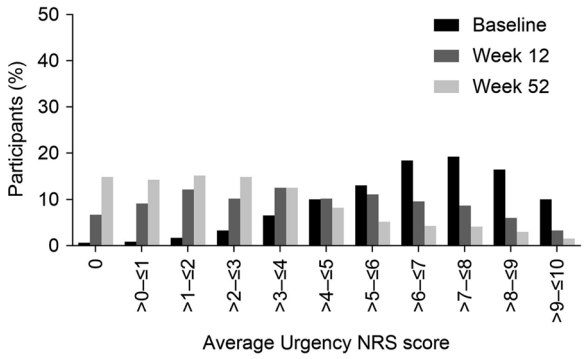

Fig. 1 Urgency NRS score distributions. An Urgency NRS score of 0 represents no urgency and a score of 10 the worst possible urgency. *NRS* Numeric Rating Scale

Responsiveness

Urgency NRS score changes from baseline to weeks 12 and 52 were moderately to strongly correlated with changes in PGRS, IBDQ Bowel Symptom domain score, IBDQ total score, IBDQ item 1 score, and CDAI, as well as with PGIC score ($|r| = 0.34–0.72$; Table S9 of the ESM).

Greater improvements in PGRS from baseline to weeks 12 and 52 corresponded with greater improvements on the Urgency NRS from baseline to weeks 12 and 52. Differences in mean Urgency NRS change were statistically significant between PGRS change categories (Table 3). Participants with lower PGIC categories, IBDQ response or remission, IBDQ Bowel Symptom domain score response, and CDAI clinical response or remission at weeks 12 and 52 showed greater improvements on the Urgency NRS (Tables S10 and S11 of the ESM).

Threshold for Meaningful Improvement

The median estimate of Urgency NRS score change at week 12 was – 3.7 for participants with a 2-point decrease from baseline on the PGRS, – 5.6 for those with a ≥ 3-point decrease on the PGRS, – 3.0 for those with a PGIC

Table 2 Test–retest reliability of the Urgency NRS

	n	Baseline mean (SD)	Week 4 mean (SD)	Mean difference	ICC
Stable on PGRS[a]					
Urgency NRS score	418	6.4 (2.18)	6.0 (2.26)	– 0.5****	0.86
Stable on PGIC[b]					
Urgency NRS score	267	6.7 (2.19)	6.0 (2.48)	– 0.6****	0.78

****All comparisons of week 4 versus baseline scores reached a significance level of $p < 0.0001$ (paired *t*-test)

[a]Participants whose weekly average PGRS score (rounded to the nearest integer) did not change between baseline and week 4

[b]Participants who reported "no change" on the PGIC at week 4 (PGIC=4)

ICC intraclass correlation coefficient, *NRS* Numeric Rating Scale, *PGIC* Patient Global Impression of Change, *PGRS* Patient Global Rating of Severity, *SD* standard deviation

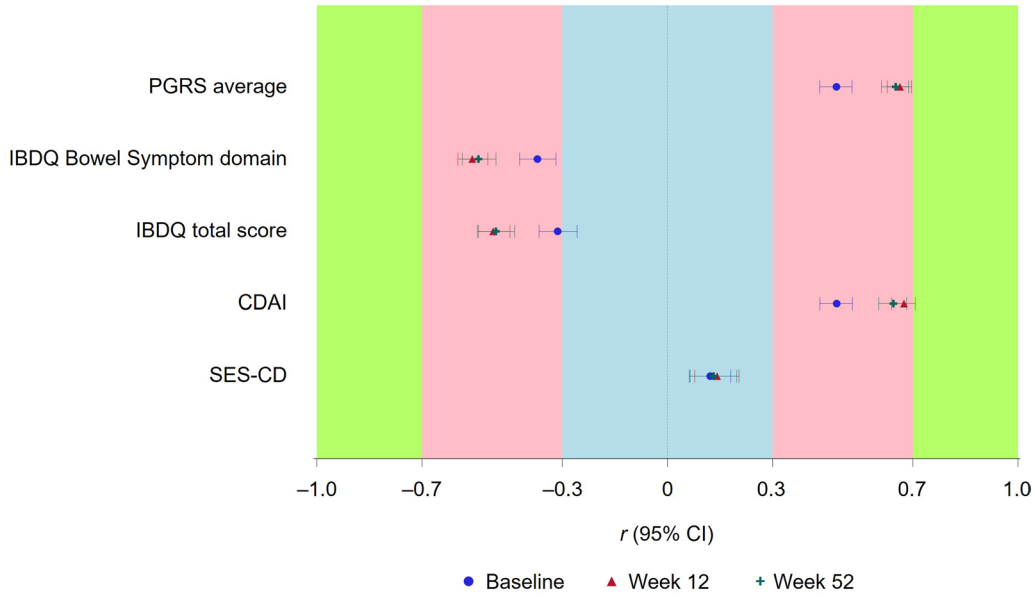

Fig. 2 Convergent and discriminant validity of the Urgency NRS at baseline, week 12, and week 52. Based on their absolute values, correlations were considered weak ($|r|$< 0.30), moderate ($|r|$: 0.30– < 0.70), or strong ($|r|$≥ 0.70). *CDAI* Crohn's Disease Activity Index, *CI* confidence interval, *IBDQ* Inflammatory Bowel Disease Questionnaire, *NRS* Numeric Rating Scale, *PGRS* Patient Global Rating of Severity, *SES-CD* Simple Endoscopic Score for Crohn's Disease

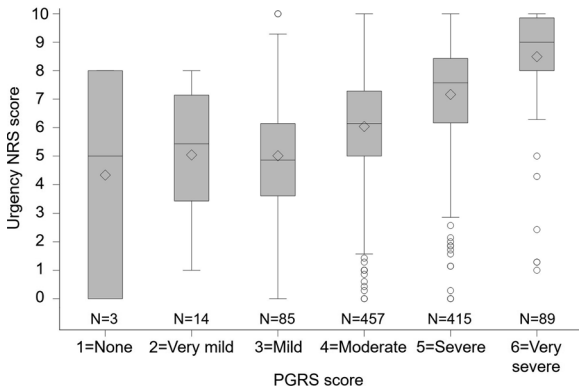

Fig. 3 Known-groups validity of the Urgency NRS by PGRS at baseline. The sample size was small for PGRS "none" and "very mild" responses. *NRS* Numeric Rating Scale, *PGRS* Patient Global Rating of Severity

response of "much better," and – 4.6 for those with a PGIC response of "very much better." At week 52, the median Urgency NRS score change estimate was – 4.0 for participants with a 2-point decrease on the PGRS, – 5.6 for those with a ≥ 3-point decrease on the PGRS, – 3.4 for those with a PGIC response of "much better," and – 5.1 for those with a PGIC response of "very much better" (Table 4). The supplemental anchors IBDQ item 1, CDAI response, and IBDQ Bowel Symptom domain response supported a threshold of 3–5 points, with median change in Urgency NRS scores ranging from – 2.9 to – 3.7 at week 12 and – 3.4 to – 4.9 at week 52 for these anchor groups (Table S12 of the ESM). The CDF and PDF plots based on PGRS change and PGIC, which supplemented anchor-based analyses, showed good differentiation between improvement categories and distinguished participants who experienced and did not experience meaningful improvement based on the anchors (Figs. S2–S5 of the ESM).

Threshold for Bowel Urgency Remission

For participants in the PGRS "very mild" category, the median Urgency NRS score was 2.0 at weeks 12 and 52. For participants in the PGIC "much better" category, the median Urgency

Table 3 Responsiveness of the Urgency NRS by PGRS change from baseline to weeks 12 and 52

PGRS change from baseline	≥ 4-point improvement	3-point improvement	2-point improvement	1-point improvement	No change	≥ 1-point worsening
Urgency NRS score change from baseline to week 12	*n* = 47	*n* = 126	*n* = 222	*n* = 315	*n* = 268	*n* = 36
LSM (SE)	− 5.8 (0.3)	− 4.8 (0.2)	− 3.5 (0.1)	− 2.0 (0.1)	− 0.7 (0.1)	0.9 (0.3)
LSMD (CI)						
vs. ≥ 4-point improvement		− 1.0 (− 2.0, 0.0)	− 2.3 (− 3.2, − 1.3)****	− 3.8 (− 4.7, − 2.9)****	− 5.1 (− 6.0, − 4.1)****	− 6.7 (− 8.0, − 5.4)****
vs. 3-point improvement			− 1.3 (− 2.0, − 0.6)****	− 2.8 (− 3.4, − 2.2)****	− 4.1 (− 4.7, − 3.5)****	− 5.7 (− 6.8, − 4.6)****
vs. 2-point improvement				− 1.5 (− 2.0, 1.0)****	− 2.8 (− 3.3, − 2.3)****	− 4.4 (− 5.5, − 3.4)****
vs. 1-point improvement					− 1.3 (− 1.8, − 0.8)****	− 2.9 (− 3.9, − 1.8)****
vs. no change						− 1.6 (− 2.6, − 0.6)****
Urgency NRS score change from baseline to week 52	*n* = 97	*n* = 222	*n* = 215	*n* = 205	*n* = 106	*n* = 13
LSM (SE)	− 5.8 (0.2)	− 5.0 (0.1)	− 3.8 (0.1)	− 2.5 (0.1)	− 1.2 (0.2)	0.4 (0.5)
LSMD (CI)						
vs. ≥ 4-point improvement		− 0.8 (− 1.6, − 0.1)*	− 2.0 (− 2.7, − 1.2)****	− 3.3 (− 4.1, − 2.5)****	− 4.6 (− 5.5, − 3.7)****	− 6.2 (− 8.0, − 4.3)****
vs. 3-point improvement			− 1.1 (− 1.7, − 0.6)****	− 2.5 (− 3.1, − 1.9)****	− 3.7 (− 4.5, − 3.0)****	− 5.3 (− 7.1, − 3.6)****
vs. 2-point improvement				− 1.3 (− 1.9, − 0.7)****	− 2.6 (− 3.3, − 1.9)****	− 4.2 (− 5.9, − 2.4)****
vs. 1-point improvement					− 1.3 (− 2.0, − 0.5)****	− 2.9 (− 4.6, − 1.1)****
vs. no change						− 1.6 (− 3.4, 0.2)

*$p < 0.05$, ****$p < 0.0001$; the pairwise LSMD (CI) and p values were adjusted using Scheffe's correction

CI confidence interval, *LSM* least square mean, *LSMD* least square mean difference, *NRS* Numeric Rating Scale, *PGRS* Patient Global Rating of Severity, *SD* standard deviation, *SE* standard error

NRS score was 3.1 at week 12 and 2.6 at week 52 (Table 5). The supplemental anchors supported an Urgency NRS score threshold of ≤ 2. The median Urgency NRS score for participants with an IBDQ item 1 score of 7 (normal, no increase in frequency of bowel movements) was 3.0 at week 12 and 1.4 at week 52; the median Urgency NRS score for those with an IBDQ item 1 score of 6 (slight increase in frequency of bowel movements) was 3.2 at week 12 and 2.9 at week 52. The median Urgency NRS score was 2.7 at week 12 and 1.8 at week 52 for participants in the IBDQ remission group, and 2.7 at week 12 and 2.0 at week 52 for those in the CDAI remission group (Table S13 of the ESM). The CDF and PDF plots showed good differentiation of Urgency NRS scores between PGRS and PGIC categories and distinguished participants who achieved and did not achieve remission based on the anchors used (Figs. S6–S9 of the ESM).

Table 4 Anchor-based estimates for change in Urgency NRS from baseline to weeks 12 and 52

Anchors	Point change in Urgency NRS from baseline to week 12							Point change in Urgency NRS from baseline to week 52						
	n	Mean (SD)	Percentile					*n*	Mean (SD)	Percentile				
			10th	25th (Q1)	50th (Q2)	75th (Q3)	90th			10th	25th (Q1)	50th (Q2)	75th (Q3)	90th
PGRS change														
≥ 3-point decrease	173	− 5.2 (2.3)	− 7.7	− 6.9	− 5.6	− 3.7	− 2.0	319	− 5.4 (2.3)	− 8.3	− 7.0	− 5.6	− 3.9	− 2.0
2-point decrease	222	− 3.5 (1.9)	− 5.6	− 4.9	− 3.7	− 2.4	− 1.1	215	− 3.9 (2.1)	− 6.4	− 5.3	− 4.0	− 2.4	− 1.1
1-point decrease	315	− 1.9 (2.0)	− 4.4	− 3.1	− 1.8	− 0.7	0.3	205	− 2.3 (2.1)	− 5.1	− 3.7	− 2.3	− 1.0	0.0
No change	268	− 0.7 (1.5)	− 2.6	− 1.4	− 0.4	0.0	0.9	106	− 1.0 (2.0)	− 3.4	− 2.3	− 0.9	0.3	1.5
≥ 1-point increase	36	0.9 (1.9)	− 0.7	0.1	1.0	2.3	3.0	13	− 0.0 (1.7)	− 2.7	− 1.1	0.0	0.7	1.7
PGIC														
1 = very much better	154	− 4.5 (2.5)	− 7.7	− 6.4	− 4.6	− 2.6	− 1.1	280	− 4.8 (2.5)	− 8.0	− 6.7	− 5.1	− 3.0	− 1.1
2 = much better	325	− 3.1 (2.2)	− 6.1	− 4.6	− 3.0	− 1.4	− 0.3	349	− 3.5 (2.5)	− 6.7	− 5.0	− 3.4	− 1.6	− 0.4
3 = a little better	308	− 1.7 (2.2)	− 4.7	− 3.1	− 1.6	− 0.1	0.6	146	− 2.5 (2.6)	− 6.0	− 4.7	− 2.3	− 0.7	0.5
4 = no change	159	− 1.0 (2.2)	− 4.1	− 2.0	− 0.4	0.0	1.3	40	− 2.3 (2.5)	− 5.3	− 4.0	− 2.9	− 0.1	0.4
5 = a little worse	33	− 0.6 (2.3)	− 4.4	− 1.8	− 0.1	0.4	2.4	12	− 1.6 (2.4)	− 5.6	− 2.9	− 1.1	0.1	0.9
6 = much worse	21	− 0.7 (1.5)	− 2.4	− 0.9	− 0.4	0.2	0.6	3	2.0 (2.5)	0.1	0.1	1.1	4.9	4.9

Table 4 continued

Anchors	Point change in Urgency NRS from baseline to week 12							Point change in Urgency NRS from baseline to week 52						
	n	Mean (SD)	Percentile					*n*	Mean (SD)	Percentile				
			10th	25th (Q1)	50th (Q2)	75th (Q3)	90th			10th	25th (Q1)	50th (Q2)	75th (Q3)	90th
7 = very much worse	8	0.4 (2.1)	− 2.1	− 1.6	0.7	2.3	3.0	4	− 3.8 (3.1)	− 6.6	− 6.3	− 4.4	− 1.3	0.3

NRS Numeric Rating Scale, *PGIC* Patient Global Impression of Change, *PGRS* Patient Global Rating of Severity, *Q* quartile, *SD* standard deviation

DISCUSSION

The prevalence and impact of bowel urgency in CD have only recently been recognized [7, 13, 20]. The assessment of bowel urgency has rarely been included as an endpoint in CD clinical trials, and it is not currently incorporated into CD clinical guidelines [13, 18, 19]. Bowel urgency is also missing from most indices measuring clinical disease activity, including the widely used CDAI [9, 29, 39]. CD remission may not imply absence of bowel urgency, which can remain in inactive disease [40]. Conversely, patients with moderate disease may only have mild bowel urgency, highlighting the importance of assessing this symptom with responsive and reliable PRO measures [13]. The Urgency NRS measures bowel urgency severity, which may help monitor the improvement or worsening of this key symptom. Previous work evaluated the Urgency NRS measurement properties [23, 24]. However, its psychometric properties in CD had not yet been confirmed using clinical trial data; in particular, evidence was needed to support its responsiveness and identify score thresholds representing meaningful improvement as well as bowel urgency remission.

Data from the Phase 3 VIVID-1 trial of adults with moderately to severely active CD were used to evaluate the test–retest reliability, construct validity, known-groups validity, and responsiveness of the Urgency NRS, as well as to estimate Urgency NRS thresholds representing meaningful improvement and remission of bowel urgency. Overall, the Urgency NRS demonstrated excellent measurement properties in the analysis population. There were no extreme floor or ceiling effects at baseline, week 12, or week 52. Urgency NRS scores decreased (improved) during the study, consistently with the general trend of improvement in CD symptoms observed during the trial.

The Urgency NRS demonstrated strong test–retest reliability (ICC ≥ 0.70) in participants defined as stable based on the PGRS and PGIC. This is consistent with Dubinsky *et al.*'s pilot CD study, which found moderate to strong test–retest reliability of the Urgency NRS using the same measures to define stable participants

Table 5 Anchor-based estimates for Urgency NRS scores at weeks 12 and 52

Anchors	Urgency NRS score at week 12							Urgency NRS score at week 52						
	n	Mean (SD)	Percentile					*n*	Mean (SD)	Percentile				
			10th	25th (Q1)	50th (Q2)	75th (Q3)	90th			10th	25th (Q1)	50th (Q2)	75th (Q3)	90th
PGRS change														
1 = none	107	1.1 (1.5)	0.0	0.0	0.4	1.4	3.0	226	1.2 (1.7)	0.0	0.0	0.4	1.7	3.7
2 = very mild	211	2.8 (2.2)	0.1	1.1	2.0	4.0	6.0	234	2.1 (1.7)	0.0	1.0	2.0	3.0	4.4
3 = mild	274	3.8 (2.1)	1.0	2.3	3.5	5.0	6.9	222	3.5 (2.1)	1.0	2.1	3.1	4.4	6.4
4 = moderate	289	5.4 (2.1)	2.6	4.0	5.6	6.9	8.1	134	4.8 (2.3)	1.7	3.1	4.9	6.6	7.9
5 = severe	109	6.9 (2.1)	3.7	6.0	7.2	8.1	9.0	41	7.1 (2.2)	3.6	6.1	7.4	8.5	9.3
6 = very severe	26	9.1 (1.1)	7.9	8.1	9.5	10.0	10.0	2	8.8 (1.7)	7.6	7.6	8.8	10.0	10.0
PGIC														
1 = very much better	154	2.2 (2.2)	0.0	0.0	1.4	3.3	5.0	280	1.7 (2.0)	0.0	0.0	1.1	2.7	4.3
2 = much better	325	3.4 (2.3)	0.5	1.6	3.1	4.9	6.4	349	3.0 (2.3)	0.1	1.1	2.6	4.1	6.4
3 = a little better	308	4.7 (2.5)	1.3	2.7	4.9	6.6	8.0	146	4.2 (2.5)	1.0	2.3	3.8	6.0	7.4
4 = no change	160	6.0 (2.5)	2.7	4.7	6.3	7.8	8.8	40	4.5 (3.1)	0.4	2.2	4.0	7.0	8.8
5 = a little worse	33	6.3 (2.5)	2.0	5.6	6.9	8.0	9.0	12	4.9 (2.3)	1.7	3.1	5.6	6.9	7.0
6 = much worse	21	6.8 (3.0)	0.6	5.4	8.0	8.9	9.2	3	9.1 (0.9)	8.3	8.3	9.0	10.0	10.0
7 = very much worse	8	7.6 (2.8)	1.3	7.1	8.1	9.3	10.0	4	2.3 (4.5)	0.0	0.0	0.1	4.6	9.0

NRS Numeric Rating Scale, *PGIC* Patient Global Impression of Change, *PGRS* Patient Global Rating of Severity, *Q* quartile, *SD* standard deviation

[24]. Although Urgency NRS scores differed between baseline and week 4 for these stable populations in our study, the difference was only ~ 0.5 points, which was not considered clinically relevant.

Hypotheses around construct validity were met at all analyzed timepoints. The Urgency NRS score showed moderate correlations with measures of similar concepts, and weak correlations with endoscopic and laboratory assessments, supporting its convergent and discriminant validity. These results are also consistent with the pilot CD study, where the Urgency NRS showed construct validity using the PGRS [24]. The Urgency NRS showed evidence of construct validity in all disease locations, indicating that it can be used to measure bowel urgency irrespective of disease location.

The current analysis also demonstrated the known-groups validity of the Urgency NRS, as it was able to discriminate between groups of participants defined based on PGRS, IBDQ Bowel Symptom domain, and IBDQ total scores. In line with these findings, in the pilot study of CD, more severe Urgency NRS scores were associated with more severe responses on other assessments, including the PGRS (although the pilot study used collapsed PGRS categories due to the small sample size) [24].

This study analyzed the responsiveness of the Urgency NRS for the first time in CD. The Urgency NRS was sensitive to detect change in symptoms, with improvements in this measure showing moderate to large correlations with changes in other measures; statistically significant greater improvement in Urgency NRS also corresponded to improvement in anchor groups, supporting the ability of the Urgency NRS to detect change in bowel urgency severity in CD.

It is important to assess what constitutes clinically meaningful improvement in bowel urgency considering the patient perspective. Anchor-based estimates using clinical trial data support a reduction of 3–5 points in Urgency NRS score as representing meaningful improvement in patients with moderately to severely active CD. This is consistent with previous estimates from qualitative interviews, which suggested that meaningful improvement in bowel urgency would be represented by a 3–5-point reduction in patients with moderately to severely active CD who had an Urgency NRS score ≥ 5 (similar to the mean baseline score in the present study) [24]. Additionally, estimates from our study support an Urgency NRS score of ≤ 2 points as representing bowel urgency remission in the analysis population. Although Urgency NRS thresholds representing bowel urgency remission have not been previously suggested, which is consistent with evidence that patients may still experience some level of bowel urgency even if they are considered to have inactive disease [24, 40].

Although the findings presented are generally consistent with analyses of moderately to severely active UC, where the Urgency NRS also demonstrated strong psychometric properties [20], differences between CD and UC should be considered. For instance, previous estimates for UC suggested a similar threshold for meaningful improvement (≥ 3-point improvement in the Urgency NRS) but a lower threshold for remission than the present study (Urgency NRS score of ≤ 1) [23]. This highlights the need to define thresholds that are specific to a particular patient population.

These analyses used a large dataset derived from a Phase 3 clinical trial and confirmed the measurement properties of the Urgency NRS using a variety of global and disease-specific instruments. Further, this study evaluated the psychometric properties of the Urgency NRS in CD over 52 weeks, while the previous survey pilot study only analyzed data over 2 weeks [24]. However, the current analysis also has potential limitations. Although participants in the VIVID-1 trial were recruited across multiple countries [41], and thus represent varied demographics, most study participants were white or Asian. Therefore, these results may not be generalizable to patients with CD of other races. Also, use of weekly average scores for the Urgency NRS may not translate into clinical practice, where patient follow-up tends to be less frequent than in clinical trial settings and a daily assessment may be less practical. Finally, Covid-19 may have worsened bowel urgency and other symptoms, either directly or via decreased treatment adherence due to fear of infection [42, 43], but it is unlikely to impact psychometric results.

CONCLUSIONS

Recent evidence highlights that bowel urgency can impact QoL in CD. The Urgency NRS demonstrated strong psychometric properties in VIVID-1, supporting its use as a reliable and construct-valid measure of bowel urgency in adults with moderately to severely active CD. A 3–5-point reduction represented clinically meaningful improvement and a score of ≤ 2 represented remission of bowel urgency.

Medical Writing/Editorial Assistance. Pablo Izquierdo, PhD, of PPD (a Thermo Fisher company) provided medical writing support, which was funded by Eli Lilly and Company and conducted in accordance with Good Publication Practice (GPP3) guidelines (http://www.ismpp.org/gpp3).

Author Contributions. Marla Dubinsky, Aisha Vadhariya, Frederick Durand, and Simon Travis contributed to the interpretation of data and critically reviewed the manuscript. Sylvia Su and Xian Zhou contributed to study design, data analysis, and interpretation. Claudine Clucas, Larissa Stassek, and Ariane K. Kawata contributed to the design of the psychometric analysis plan and the analysis and interpretation of the results. All authors read and approved the final manuscript.

Funding. This study was funded by Eli Lilly and Company. Co-authors employed by Eli Lilly and Company provided input into the design of the psychometric analysis plan and participated in the analysis and interpretation of the results. The journal's Rapid Service Fee and Open Access Fee were funded by Eli Lilly and Company.

Data Availability. Eli Lilly and Company provides access to all individual participant data collected during the study, after anonymization. Data are available to request 6 months after the indication studied has been approved in the USA and EU and after primary publication acceptability, whichever is later. No expiration date of data requests is currently set once data are made available. Access is provided after a proposal has been approved by an independent review committee identified for this purpose and after receipt of a signed data sharing agreement. Data and documents, including the study protocol, statistical analysis plan, and study report will be provided in a secure data sharing environment. For details on submitting a request, see the instructions provided at [www.vivli.org].

Declarations

Conflict of Interest. Marla Dubinsky reports being a consultant for AbbVie, Arena, Astra Zeneca, BMS, Boehringer Ingelheim, Eli Lilly and Company, Genentech, Janssen, Johnson and Johnson Innovative Medicine, Merck, Pfizer, Prometheus Biosciences, Prometheus Laboratories, Sphyre, and Takeda. Aisha Vadhariya, Sylvia Su, and Frederick Durand are employees and shareholders of Eli Lilly and Company. Xian Zhou is an employee of Syneos Health, which received funding from Eli Lilly and Company in connection with this study. Claudine Clucas, Larissa Stassek, and Ariane K. Kawata are employees of Evidera, which received funding from Eli Lilly and Company in connection with this study. Simon Travis has received grants from AbbVie, BUHLMANN Diagnostics, ECCO, Eli Lilly and Company, Ferring Pharmaceuticals, International Organization for the Study of Inflammatory Bowel Disease, Janssen, Merck Sharp & Dohme, Norman Collison Foundation, Pfizer, Procter & Gamble, Schering-Plough, Takeda, UCB Pharma, Vifor Pharma, and Warner Chilcott; and reports other disclosures from Abacus Pharma, AbbVie, Actial Farmaceutica, ai4gi, Alcimed, Allergan, Amgen, Aptel, Arena Pharmaceuticals, Asahi Kasei Pharma, Aspen Pharmacare, Astellas, AstraZeneca, Atlantic Pharmaceuticals, Barco, Biocare Medical, Biogen, BL Pharma, Boehringer Ingelheim, Bristol Myers Squibb, BUHLMANN Diagnostics, Calcico Therapeutics, Celgene, Cellerix, Cerimon Pharmaceuticals, ChemoCentryx, Chiesi, Cisbio, Comcast, Coronado Biosciences, Cosmo Pharmaceuticals, Dr. Falk Pharma, Ducentis BioTherapeutics, Dynavax Technologies, Elan Pharma,

Eli Lilly and Company, Enterome, Equillium, Ferring Pharmaceuticals, Galapagos NV, Genentech/Roche, Genzyme, Gilead Sciences, GlaxoSmithKline, Glenmark Pharmaceuticals, Grünenthal, GW Pharmaceuticals, Immunocore, Indigo Biosciences, Janssen, Lexicon Pharmaceuticals, Medarex, MedTrix, Merck, Merrimack Pharmaceuticals, Mestag Therapeutics, Millennium Pharmaceuticals, Neovacs, Novartis, Novo Nordisk, NPS Pharmaceuticals/Nycomed, Ocera Therapeutics, OPTIMA Pharma, Origin Pharma, Otsuka, Palau Pharma, Pentax Medical, Pfizer, PharmaVentures, Phesi, Phillips Pharma Group, Procter & Gamble, Pronota, Proximagen, Resolute Pharma, Robarts Clinical Trials, Sandoz, Santarus, Satisfai Health, Sensyne Health, Shire, Sigmoid Pharma, Sorriso Pharmaceuticals, Souffinez, SynDermix, Synthon, Takeda, Theravance Biopharma, TiGenix, Tillotts Pharma AG, Topivert, Trino Therapeutics with Wellcome Trust, TxCell SA, UCB Pharma, Vertex Pharma, VHsquared, Vifor Pharma, Warner Chilcott, and Zeria Pharmaceutical.

Ethical approval. The study was compliant with the Declaration of Helsinki and Council for International Organizations of Medical Sciences International Ethical guidelines and the International Conference on Harmonization guidelines on Good Clinical Practice. Informed consent forms and protocols received ethical review board approval before study initiation. Participants provided written informed consent before any procedures were performed.

REFERENCES

1. Cushing K, Higgins PDR. Management of Crohn disease: a review. JAMA. 2021;325(1):69–80.
2. Dolinger M, Torres J, Vermeire S. Crohn's disease. Lancet. 2024;403(10432):1177–91.
3. Turner D, Ricciuto A, Lewis A, D'Amico F, Dhaliwal J, Griffiths AM, et al. STRIDE-II: an update on the selecting therapeutic targets in inflammatory bowel disease (STRIDE) initiative of the International Organization for the Study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. Gastroenterology. 2021;160(5):1570–83.
4. Colombel JF. Disease clearance in inflammatory bowel disease. Gastroenterol Hepatol. 2021;17(5):233–5.
5. van der Have M, van der Aalst KS, Kaptein AA, Leenders M, Siersema PD, Oldenburg B, Fidder HH. Determinants of health-related quality of life in Crohn's disease: a systematic review and meta-analysis. J Crohns Colitis. 2014;8(2):93–106.
6. Gravina AG, Pellegrino R, Durante T, Palladino G, D'Onofrio R, Mammone S, et al. Inflammatory bowel diseases patients suffer from significant low levels and barriers to physical activity: The "BE-FIT-IBD" study. World J Gastroenterol. 2023;29(41):5668–82.
7. Ha C, Naegeli AN, Lukanova R, Shan M, Wild R, Hennessy F, et al. Rectal urgency among patients with ulcerative colitis or Crohn's disease: analyses from a global survey. Crohns Colitis 360. 2023;5(4):otad052.
8. Dubinsky M, Bleakman AP, Panaccione R, Hibi T, Schreiber S, Rubin D, et al. Bowel urgency in ulcerative colitis: current perspectives and future directions. Am J Gastroenterol. 2023;118(11):1940–53.
9. Pakpoor J, Travis S. Why studying urgency is urgent. Gastroenterol Hepatol. 2023;19(2):95–100.
10. Dawwas GK, Jajeh H, Shan M, Naegeli AN, Hunter T, Lewis JD. Prevalence and factors associated with fecal urgency among patients with ulcerative colitis and Crohn's disease in the study of

a prospective adult research cohort with inflammatory bowel disease. Crohns Colitis 360. 2021;3(3):otab046.

11. Magro F, Gionchetti P, Eliakim R, Ardizzone S, Armuzzi A, Barreiro-de Acosta M, et al. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. Part 1: definitions, diagnosis, extra-intestinal manifestations, pregnancy, cancer surveillance, surgery, and ileo-anal pouch disorders. J Crohns Colitis. 2017;11(6):649–70.

12. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG clinical guideline: ulcerative colitis in adults. Am J Gastroenterol. 2019;114(3):384–413.

13. Schreiber S, Hunter Gibble T, Panaccione R, Rubin DT, Travis S, Hibi T, et al. Patient and health care professional perceptions of the experience and impact of symptoms of moderate-to-severe Crohn's disease in US and Europe: results from the cross-sectional CONFIDE study. Dig Dis Sci. 2024;69(7):2333–44.

14. Jairath V, Hunter Gibble T, Potts Bleakman A, Chatterton K, Medrano P, McLafferty M, et al. Patient perspectives of bowel urgency and bowel urgency-related accidents in ulcerative colitis and Crohn's disease. Inflamm Bowel Dis. 2024;30:izae044.

15. Casellas F, Herrera-de Guise C, Robles V, Navarro E, Borruel N. Patient preferences for inflammatory bowel disease treatment objectives. Dig Liver Dis. 2017;49(2):152–6.

16. Louis E, Ramos-Goni JM, Cuervo J, Kopylov U, Barreiro-de Acosta M, McCartney S, et al. A qualitative research for defining meaningful attributes for the treatment of inflammatory bowel disease from the patient perspective. Patient. 2020;13(3):317–25.

17. Schoefs E, Vermeire S, Ferrante M, Sabino J, Lambrechts T, Avedano L, et al. What are the unmet needs and most relevant treatment outcomes according to patients with inflammatory bowel disease? A qualitative patient preference study. J Crohns Colitis. 2023;17(3):379–88.

18. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE. ACG clinical guideline: management of Crohn's disease in adults. Am J Gastroenterol. 2018;113(4):481–517.

19. Torres J, Bonovas S, Doherty G, Kucharzik T, Gisbert JP, Raine T, et al. ECCO guidelines on therapeutics in Crohn's Disease: medical treatment. J Crohns Colitis. 2020;14(1):4–22.

20. Travis S, Potts Bleakman A, Dubinsky MC, Schreiber S, Panaccione R, Hibi T, et al. The communicating needs and features of IBD experiences (CONFIDE) Study: US and European patient and health care professional perceptions of the experience and impact of symptoms of moderate-to-severe ulcerative colitis. Inflamm Bowel Dis. 2024;30(6):939–49.

21. US Food and Drug Administration: Patient-focused drug development: collecting comprehensive and representative input. Guidance for industry, food and drug administration staff, and other stakeholders; 2020. https://www.fda.gov/media/139088/download. Accessed Oct 2024

22. Dubinsky MC, Irving PM, Panaccione R, Naegeli AN, Potts-Bleakman A, Arora V, et al. Incorporating patient experience into drug development for ulcerative colitis: development of the Urgency Numeric Rating Scale, a patient-reported outcome measure to assess bowel urgency in adults. J Patient Rep Outcomes. 2022;6(1):31.

23. Dubinsky MC, Shan M, Delbecque L, Lissoos T, Hunter T, Harding G, et al. Psychometric evaluation of the Urgency NRS as a new patient-reported outcome measure for patients with ulcerative colitis. J Patient Rep Outcomes. 2022;6(1):114.

24. Dubinsky MC, Delbecque L, Hunter T, Harding G, Stassek L, Moses RE, Lewis JD. Validation of the bowel urgency numeric rating scale in patients with Crohn's disease: results from a mixed methods study. Qual Life Res. 2023;32(12):3403–15.

25. US Food and Drug Administration: Patient-focused drug development: Incorporating clinical outcome assessments into endpoints for regulatory decision-making. Guidance for industry, food and drug administration staff, and other stakeholders; 2023. https://www.fda.gov/media/166830/download. Accessed Oct 2024

26. Guyatt G, Mitchell A, Irvine EJ, Singer J, Williams N, Goodacre R, Tompkins C. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology. 1989;96(3):804–10.

27. Irvine EJ. Quality of life of patients with ulcerative colitis: past, present, and future. Inflamm Bowel Dis. 2008;14(4):554–65.

28. Irvine EJ, Zhou Q, Thompson AK. The Short Inflammatory Bowel Disease Questionnaire: a quality of life instrument for community physicians managing inflammatory bowel disease. CCRPT Investigators. Canadian Crohn's Relapse Prevention Trial. Am J Gastroenterol. 1996;91(8):1571–8.

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

30. Vermeire S, Schreiber S, Petryka R, Kuehbacher T, Hebuterne X, Roblin X, et al. Clinical remission in patients with moderate-to-severe Crohn's disease treated with filgotinib (the FITZROY study): results from a phase 2, double-blind, randomised, placebo-controlled trial. Lancet. 2017;389(10066):266–75.

31. Wong ECL, Dulai PS, Marshall JK, Jairath V, Reinisch W, Narula N. Resolution of dominant patient-reported outcome at end of induction predicts clinical and endoscopic remission in Crohn's disease. Aliment Pharmacol Ther. 2022;55(9):1151–9.

32. Abbass M, Cepek J, Parker CE, Nguyen TM, MacDonald JK, Feagan BG, et al. Adalimumab for induction of remission in Crohn's disease. Cochrane Database Syst Rev. 2019;2019(11):CD012878.

33. Peyrin-Biroulet L, Arkkila P, Armuzzi A, Danese S, Guardiola J, Jahnsen J, et al. Comparative efficacy and safety of infliximab and vedolizumab therapy in patients with inflammatory bowel disease: a systematic review and meta-analysis. BMC Gastroenterol. 2022;22(1):291.

34. Daperno M, D'Haens G, Van Assche G, Baert F, Bulois P, Maunoury V, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc. 2004;60(4):505–12.

35. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.

36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.

37. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. 5th ed. Boston: Houghton Mifflin; 2003.

38. Scheffé H. The analysis of variance. Hoboken: Wiley; 1959.

39. Sandborn WJ, Feagan BG, Hanauer SB, Lochs H, Lofberg R, Modigliani R, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. Gastroenterology. 2002;122(2):512–30.

40. Farrell D, McCarthy G, Savage E. Self-reported symptom burden in individuals with inflammatory bowel disease. J Crohns Colitis. 2016;10(3):315–22.

41. National Library of Medicine: NCT03926130. A study of mirikizumab (LY3074828) in participants with Crohn's disease (VIVID-1); 2023. https://www.clinicaltrials.gov/study/NCT03926130. Accessed Oct 2024

42. Wu J, Fang Y, Bai B, Wu Y, Liu Q, Hu J, et al. Clinical characteristics and the risk factors for the exacerbation of symptoms in patients with inflammatory bowel disease during the COVID-19 pandemic. Front Med. 2024;11:1404880.

43. Pellegrino R, Pellino G, Selvaggi F, Federico A, Romano M, Gravina AG. Therapeutic adherence recorded in the outpatient follow-up of inflammatory bowel diseases in a referral center: damages of COVID-19. Dig Liver Dis. 2022;54(10):1449–51.