

Evaluation and performance of a newly developed patient-reported outcome instrument for diarrhea-predominant irritable bowel syndrome in a clinical study population

Leticia Delgado-Herrera, Kathryn Lasch, Bernhardt Zeiher, Anthony J. Lembo, Douglas A. Drossman, Benjamin Banderas, Kathleen Rosa, Christopher Lademacher and Rob Arbuckle

Abstract

Background: To evaluate the psychometric properties of the newly developed seven-item Irritable Bowel Syndrome – Diarrhea predominant (IBS-D) Daily Symptom Diary and four-item Event Log using phase II clinical trial safety and efficacy data in patients with IBS-D. This instrument measures diarrhea (stool frequency and stool consistency), abdominal pain related to IBS-D (stomach pain, abdominal pain, abdominal cramps), immediate need to have a bowel movement (immediate need and accident occurrence), bloating, pressure, gas, and incomplete evacuation.

Methods: Psychometric properties and responsiveness of the instrument were evaluated in a clinical trial population [ClinicalTrials.gov identifier: NCT01494233].

Results: A total of 434 patients were included in the analyses. Significant differences were found among severity groups ($p < 0.01$) defined by IBS Patient Global Impression of Severity (PGI-S) and IBS Patient Global Impression of Change (PGI-C). Severity scores for each Diary and Event Log item score and five-item, four-item, and three-item summary scores were calculated. Between-group differences in changes over time were significant for all summary scores in groups stratified by changes in PGI-S ($p < 0.05$), two of six Diary items, and three of four Event Log items; a one-grade change in PGI-S was considered a meaningful difference with mean change scores on all Diary items -0.13 to -0.86 [standard deviation (SD) $0.79-1.39$]. Similarly, for patients who reported being 'slightly improved' (considered a clinically meaningful difference) on the PGI-C, mean change scores on Diary items ranged from -0.45 to -1.55 (SD $0.69-1.39$). All estimates of clinically important change for each item and all summary scores were small and should be considered preliminary. These results are aligned with the previous standalone psychometric study regarding reliability and validity tests.

Conclusions: These analyses provide evidence of the psychometric properties of the IBS-D Daily Symptom Diary and Event Log in a clinical trial population.

Keywords: diarrhea-predominant irritable bowel syndrome, IBS Global Assessment of Improvement, IBS Patient Global Impression of Severity, psychometric, responsiveness

Received: 29 November 2016; revised manuscript accepted: 5 July 2017.

Introduction

Irritable bowel syndrome (IBS) is a functional gastrointestinal disorder characterized by abdominal

pain or discomfort associated with altered bowel habits, including constipation, diarrhea, or both.¹ Efforts to study the potential benefit of treatment

Ther Adv Gastroenterol

2017, Vol. 10(9) 673–687

DOI: 10.1177/
1756283X17726018

© The Author(s), 2017.
Reprints and permissions:
[http://www.sagepub.co.uk/
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)

Correspondence to:
Leticia Delgado-Herrera
Astellas Pharma Global
Development, Inc., One
Astellas Way, Northbrook,
IL 60062, USA
[leticia.delgado-herrera@
astellas.com](mailto:leticia.delgado-herrera@astellas.com)

Kathryn Lasch
Pharmerit International,
Newton, MA, USA

Bernhardt Zeiher
Christopher Lademacher
Astellas Pharma Global
Development, Inc.,
Northbrook, IL, USA

Anthony J. Lembo
Beth Israel Deaconess
Medical Center, Boston,
MA, USA

Douglas A. Drossman
Drossman Center for the
Education and Practice
of Biopsychosocial
Care, UNC Center for
Functional GI and Motility
Disorders, University
of North Carolina at
Chapel Hill, Chapel Hill,
NC, USA, and Drossman
Gastroenterology PLLC,
Chapel Hill, NC, USA

Benjamin Banderas
Kathleen Rosa
Adelphi Values, Boston,
MA, USA

Rob Arbuckle
Adelphi Values, Bollington,
Cheshire, UK

of patients with IBS and diarrhea-predominant IBS (IBS-D) have been hindered by a lack of well validated symptom severity scales that meet current standards for patient-reported outcome (PRO) instruments^{2,3} and a reliable biologic marker for IBS.⁴ In turn, this presents a challenge for development of optimal study designs and endpoints for clinical trials assessing drug impact on IBS,⁴ thereby hindering development of drugs that provide benefit to patients.

A patient-reported instrument for IBS-D has recently been developed and validated^{5,6} in line with the guidance from the US Food and Drug Administration (FDA) with respect to development of both PROs and drugs to treat IBS.^{2,3} The instrument is a PRO measure that includes a seven-item IBS-D Daily Symptom Diary (hereafter referred to as the Diary), in which patients record the severity of symptoms over the previous 24 h, and a four-item IBS-D Symptom Event Log (hereafter referred to as the Event Log), which is completed following each bowel movement to assess information regarding individual bowel movements.⁵ The 'Diary' collects information on abdominal pain, stomach pain, abdominal cramps, abdominal bloating, frequency of gas and accidents; and the 'Event Log' on each individual bowel movement, the immediacy of need, and the consistency of the bowel movement. It was developed by a multidisciplinary research team for use as a clinical trial endpoint in IBS-D and its content validity was evaluated during development using five qualitative research studies.⁶ Evidence from another observational study supported an initial evaluation of the psychometric properties of the instrument.⁷

In this article, we report additional analyses using data from a patient population in an interventional clinical trial to further evaluate the psychometric properties of the instrument in the specific context of a clinical trial population. In particular, we evaluated clinical validity (known groups analysis), responsiveness (ability to detect change), quality of completion, testing of test-retest reliability and evaluation of construct validity through examining floor and ceiling effects and inter-item correlations, and the interpretation of scores in the context of clinically important change (responder definition).

The psychometric results obtained from both the standard standalone psychometric study⁷ and the

LX1033 negative phase II data reported in this manuscript were evaluated by clinician experts in the field of IBS. This is the first IBS-D specific patient-reported outcome (PRO) developed in accordance with the US FDA PRO guidance to be used in an interventional study. This tool measures all concepts that are clinically relevant and important to patients in a manner that patients understand and can respond to, such as abdominal pain, bowel function, and bloating.

The experts' review of the findings and recommendation was considered in the context of the data generated during the development of the new instrument, including the LX1033 phase II clinical study.

This is the third in a series of papers describing the development and validation of a new measure of disease-specific symptom severity for IBS-D. This paper provides new data from a phase II multicenter, randomized, double-blind, placebo-controlled, multiple-dose study to determine efficacy and safety in patients with IBS-D. In this interventional study the qualitative results were supported and psychometric findings were replicated, the validity of the PRO instrument showed good test-retest reliability and good separations of groups differing in self-reported symptom severity.

Materials and methods

IBS-D Symptom Diary and Event Log

Details of the development of the Diary and Event Log have been described previously.⁵⁻⁷ Briefly, the initial development of the instrument involved several iterative rounds of in-depth qualitative research, including concept elicitation focus groups and cognitive interviews to support evaluation of content validity. This qualitative research involved the identification, clarification, and refinement of items, instructions, and response scales to assess symptom concepts such as diarrhea, immediacy of need, bloating/pressure, among others.⁶

In the Marquis (2014) paper,⁶ in which the qualitative assessment of the Bristol Stool Form Scale (BSFS) and an adapted BSFS was tested in the IBS-D population, it was found that patients' spontaneous descriptions of stool form and consistency did not optimally correspond with those

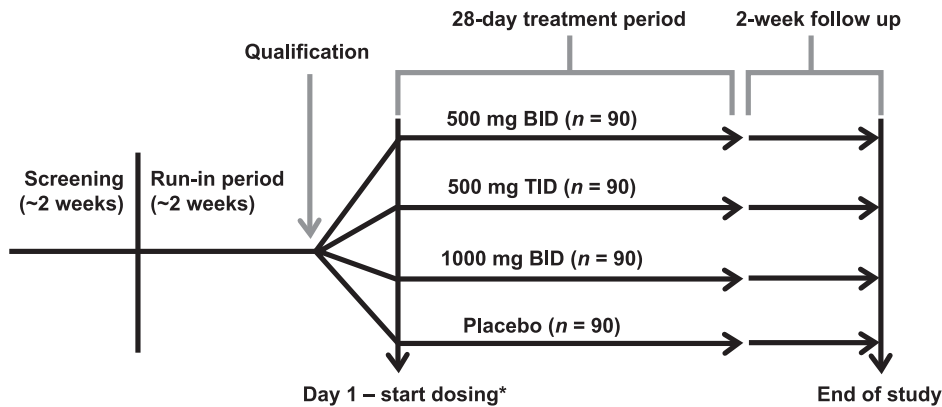


Figure 1. Clinical trial design. Dosing numbers were targets and do not reflect the number of patients recruited for each treatment arm. BID, twice daily; TID, three times daily.

*Day 1 may occur on same day as qualification visit or 3 to 5 days after, if assigned study drug is not onsite.

of the two stool scales tested, one of which is mostly widely used in this setting. A new stool form and consistency scale applicable to IBS-D in the clinical setting was developed and cognitively tested.⁸ The final seven-item Diary includes items assessing abdominal pain, stomach pain, abdominal cramps, abdominal pressure, bloating, frequency of passing gas, and bowel control (see Appendix Figure A). The response option for pain, cramps, and bloating is a 0–10 numerical rating scale ranging from 0 (low severity) to 10 (high severity). Frequency of passing gas is rated on a five-point verbal descriptor scale ranging from ‘None of the time’ to ‘All of the time’. Response categories for the daily diary used an 11-point numeric rating scale to assess severity; while the event log used a five-point or six-point ordinal (Likert-type) scale to assess the number of times the bowel movement occurred.⁶ The item for bowel control (accidents) requires a dichotomous ‘yes/no’ answer. Items thought to be more accurately and reliably recorded at the time of occurrence were included in the Event Log (see Appendix Figure B). This included the date and time of occurrence, immediacy of need (on a five-point scale from ‘No immediate need’ to ‘Extreme immediate need’), stool consistency (measured with the Astellas Stool Form Scale (ASFS)), and a dichotomous ‘yes/no’ item for incomplete bowel evacuation.

Study population

The instrument was included as an exploratory endpoint in a phase II, multicenter, randomized, double-blind, placebo-controlled, multiple-dose

clinical trial involving 434 patients with IBS-D⁹ [ClinicalTrials.gov identifier: NCT01494233]. The trial design consisted of four periods: a screening period (approximately 2 weeks), a run-in period (15–20 days), the blinded treatment period (28 days), and the follow-up period (14 days) (Figure 1).

During the run-in period, patients were allowed rescue medication: two or fewer doses of up to 4 mg each of loperamide per week, and previous doses of bulking agents (psyllium products, fiber tablets) with a stable regimen (no significant change within the prior 3 months). Eligible patients were randomized 1:1:1:1 to receive one of three doses of LX1033 [an investigational drug; 1000 mg twice daily ($n = 90$); 500 mg three times daily ($n = 90$); 500 mg twice daily ($n = 90$)], or matching placebo ($n = 90$) for 28 days. The study was conducted in accordance with the Declaration of Helsinki, Good Clinical Practice, International Conference on Harmonization guidelines and all applicable laws and regulations. All procedures were approved by appropriate central or local institutional review boards that provided ethical oversight for the study and all participating patients provided written informed consent.

Eligible patients were men and women aged from 18 to 70 years, diagnosed with IBS-D based on the following criteria:

- (1) Fulfilment of Rome III criterion for at least 3 months with symptom onset for at least 6 months prior to diagnosis.

- (2) Recurrent abdominal pain or discomfort (the latter defined as an uncomfortable sensation not described as pain) for at least 3 days/month during the last 3 months associated with two or more of the following: improvement with defecation, onset associated with an increased frequency of stool, and onset associated with a change in form (appearance) of stool (i.e. loose/watery) for at least 2 days per week with at least one stool that had a consistency of type 6 or 7 on the BSFS¹⁰ for each of the 2 weeks during the run-in period.
- (3) Weekly average of worst abdominal pain in the past 24 h score of greater than 3.0 using a 0 (no pain) to 10 (worst pain) point scale for each of the 2 weeks during the run-in period.
- (4) Normal structural evaluation of the colon (by air contrast barium enema, virtual colonoscopy, or endoscopic colonoscopy) within 5 years prior to screening for patients over 50 years or patients who had alarm symptoms (e.g. anemia, clinically significant unexplained weight loss, family history of colorectal cancer, rectal bleeding, etc.).

Exclusion criteria included the inability to discontinue any current drug therapy for IBS (except bulking agents) for 14 days immediately prior to the run-in period, during the run-in period, and for the duration of the study; an abdominal pain score rated at least 7 for more than 5 days/week during the run-in period; and concomitant use of opioid analgesics, or other drugs that specifically affect bowel motility, unless otherwise specified in the protocol.

Assessments

The schedule of assessments from the clinical trial protocol is shown in Table 1. All patients submitted paper diary recordings for the run-in, treatment, and follow-up periods or until early discontinuation. In addition to Diary and Event Log assessments, patients were also required to complete the IBS Patient Global Impression of Change (IBS PGI-C; weekly) and IBS Patient Global Impression of Severity (IBS PGI-S; weekly). The IBS PGI-C is a single-item rating scale that asks patients, 'Compared to the way you felt before you entered the study, have your IBS symptoms over the past 7 days been: (1)

'Substantially worse'; (2) 'Moderately worse'; (3) 'Slightly worse'; (4) 'No change'; (5) 'Slightly improved'; (6) 'Moderately improved'; or (7) 'Substantially improved'.' The IBS PGI-S is a single-item rating scale that asks patients 'Please rate your IBS symptoms for the past 7 days on a scale of 0–10, with 0 representing no symptoms and 10 the worst symptoms imaginable.'

In addition to the IBS PGI-S, two versions of the IBS Patient Global Impression of Change (PGI-C) were completed during this study: one that assessed change since the previous day at day 15 (PGI-C-Day) and another that asked patients to rate how the past week compared with the first week on the study at end of the study (PGI-C-Week).

Analyses

For analyses of known group comparisons, all eligible patients (without regard to treatment assignment) who completed the patient questionnaires on the first and last day of the 14-day run-in period and had recorded data for at least 10 days during that period were included in the cross-sectional analysis population.

For the Diary items, symptom summary scores were created as the mean of responses to items one to five of the Diary (abdominal pain, stomach pain, abdominal cramps, abdominal pressure, and bloating; 'frequency of gas' item not included). In addition, four-item (stomach pain omitted) and three-item (stomach pain and bloating omitted) summary scores were also assessed. All three summary scores were included in analyses to provide insights regarding the relative validity of each summary score to help identify which item grouping was most informative. Scoring was conducted *via* a multifaceted approach, using qualitative data, psychometric analysis, item response theory, and review by experts. Items excluded from scoring were not removed from the algorithm.

Quality of completion was assessed for the Diary, and missing responses were described for days 1–15 on all questionnaires received. This included the number and percentage of missing items per patient and missing response per item, and number of patients with at least one missing item. For the Event Log, missing responses were described

Table 1. Schedule of assessments from the clinical trial protocol.

Tolerance (days)	Screening (-2 weeks) N/A	Run in (-2 weeks)*	Qualification	Day 1 [§] N/A	Treatment week 1 \pm 1	Treatment week 4 \pm 1	2-week follow-up \pm 2
WPAI-IBS				X [‡]	X	X	
IBS-QOL				X [‡]		X	
HADS				X [‡]			
IBS Adequate Relief, IBS Severity, and IBS PGI-C [§]		X		X	—————→		
Daily Diary (electronic)		X			—————→		
Daily Symptom and Event Diary (paper) ^{,¶}		X			—————→		
Administration of LX1033				X	—————→		

* Between 15 and 20 days to allow evaluation of Diary data over 14 consecutive days.
[§]Day 1 may occur same day as qualification visit or 3–5 days after, if assigned study drug is not available onsite.
[‡]Predose.
^{||}To be completed weekly by patient (see Appendix B of protocol).
^{||}Diaries should be completed daily for the duration of the run-in, qualification, treatment, and follow-up periods (see Appendix B of protocol).
[¶]Paper diary is separate from daily electronic diary.
 IBS, irritable bowel syndrome; IBS PGI-C, IBS Patient Global Assessment of Change; IBS-QOL, IBS Quality of Life Questionnaire; HADS, Hospital Anxiety and Depression Scale; WPAI-IBS, Work Productivity and Activity Impairment – IBS specific.

for day 1 and day 15 for all bowel movements reported (i.e. all bowel movements for which a date is indicated in the Event Log) and included the number and percentage of missing items per bowel movement and missing response per item, as well as the number of bowel movements with at least one missing item. No imputation was conducted.

Three test-retest reliability analyses were performed: daily scores were tested between days 14 and 15, using patients classified as stable by the PGI-C Day; weekly scores were tested for day 1-7 to day 22-28, using patients defined as stable based on their response to the PGI-C Week; and patients with equivalent PGI-S on any two of the five PGI-S days were used to assess test-retest reliability (see Table 1 for the schedule of assessments). This approach was used to potentially increase the number of stable patients in the analyses. Test-retest reliability was evaluated between the scores collected from the two days that patients had the same PGI-S.

Floor and ceiling effects were investigated for each score. For item scores, an item was considered to have a ceiling effect if the percentage of responses in the highest response category was over 100/(number of response options on an item); and a floor effect if the percentage of responses in the lowest response category was over 100/(number of response options on an item).

Inter-item correlations were evaluated among the items of the IBS-D Daily Symptom Diary at day 1 and week 1. The content of items that correlated extremely highly with one another (>0.80) was reviewed as the items may have been so similar in content that they were capturing redundant information.

Evaluation of known group validity involved a cross-sectional comparison of scores on an instrument among groups that would be expected to differ in the constructs being assessed.¹¹ For the known groups analyses, weekly average scores on the Diary summary scores and the Event Log were assessed for weeks 1 and 4 treatment periods relative to values on the IBS PGI-C and IBS PGI-S scores. Categorical values for the IBS PGI-C were 'Substantially worse', 'Moderately worse', 'Slightly worse', 'No change', 'Slightly improved', 'Moderately improved', and 'Substantially improved'. The IBS PGI-S scores were stratified

as '0-3', '4-6', and '7-10'. Comparisons were tested using one-way analysis of variance (ANOVA).

For analyses of responsiveness and interpretation, patients who were included in the cross-sectional analysis population and also provided follow-up data at weeks 1 and 4 of the treatment period and the 2-week follow-up period on the Diary and Event Log were included in the longitudinal analyses. For this assessment, changes in Diary and Event Log scores between week 2 of the run-in and week 4 of the treatment period were categorized as 'Improved', 'No change', and 'Worse' based on separate analyses of changes in their PGI-S responses, and IBS PGI-C responses. For the PGI-S, patients with a two-grade or higher decrease were categorized as 'Improved'; patients with a less than two-grade change were categorized as 'No change'; and patients with a two-grade or higher increase were categorized as 'Worse'. For the IBS PGI-C, patients were categorized by responses on that single item at week 4 as 'Improved' ('Moderately improved' or 'Substantially improved'), 'No change' ('No change'), or 'Worse' ('Moderately worse' or 'Substantially worse'). In each case, Student *t* tests were performed to compare changes from baseline in mean Diary and Event Log scores within these change groups and ANOVA was used to compare changes in Diary and Event Log scores between those groups. Both single-item scores and the five-, four- and three-item IBS-D Daily Diary summary scores were compared among those change groups (see above for details of the composition of each summary score).

To aid future interpretation of the IBS-D Daily Diary and Event Log scores, one anchor-based and two distribution-based approaches were used to help understand the level of change that would be considered clinically important.¹² In the anchor-based analysis, changes in the IBS PGI-S scores and the IBS PGI-C scores were used to define change groups. A minimal important difference was defined as a one-grade reduction on the IBS PGI-S scores from run-in week 2 to treatment week 4, or a response of 'Slightly improved' on the IBS PGI-C. In addition, and as a sensitivity analysis, a response of 'Moderately improved' on the IBS PGI-C was also analyzed. In the first distribution-based method, a 0.5 standard deviation (SD) was used as the threshold for a minimally important difference.^{12,13} In the second

distribution-based approach, we used the standard error of measurement (SEM),¹⁴ calculated as:

$$\text{SEM} = \text{SD at baseline} \times \sqrt{1 - \text{reliability}}$$

[reliability was defined as the intraclass correlation coefficient (ICC) between run-in week 2 and treatment week 1 in patients with no change in IBS PGI-S scores between day 1 and treatment week 1].

Results

A total of 434 patients were included in the study. Patient demographics are shown in Table 2.

Quality of completion

The quality of completion for the IBS-D Daily Symptom Diary met acceptability criteria, with less than 5% of responses missed by patients on any one day of the 14-day run-in period, and less than 10% of patients missing one or more items on any one day of week 1 or week 2 of the run-in period.

Floor and ceiling effects

Between 12% and 18% of subjects reported symptoms at the floor of measurement at day 1 and day 15, except for 'Frequency of gas' (5–8%), and between 1% and 2.5% of subjects reported symptoms at the ceiling of measurement, except for 'Frequency of gas' (4% at day 1 and 1.5% at day 15). Floor effects are less of a concern in this instance as they represent absence of a symptom on that given day, which is characteristic of the disease, in particular, flare fluctuations seen with IBS-D. Therefore, these levels of floor and ceiling effects are considered acceptable and not of concern. For future trials, patients with more severe ratings on average should be considered if a target effect is desired.

Inter-item correlations

With the exception of correlations with the 'Frequency of gas' item, moderate to high inter-item correlations were observed among all items: day 1 (range 0.674–0.904). The inter-item correlations were highest between items measuring severity of abdominal and stomach pain ($r = 0.904$ for the daily report at day 1 and 0.942 for

Table 2. Patient demographics.

Demographics	n (%) (N = 434)
Mean age (SD)	44.6 (13.5)
Gender	
Female	297 (68.4%)
Male	136 (31.3%)
Missing	1 (0.2%)
Ethnicity	
Not Hispanic or Latino	385 (88.7%)
Hispanic or Latino	45 (10.4%)
Unknown/missing	3 (8.7%)
No response	1 (0.2%)
Race	
White/Caucasian	383 (88.2%)
Black/African American	36 (8.3%)
American Indian/Alaskan Native	5 (1.2%)
Asian	4 (0.9%)
Native Hawaiian/other Pacific Islanders	2 (0.5%)
Other	5 (1.2%)

SD, standard deviation.

the week 1 mean score), suggesting potential redundancy for those two items in particular, as patients consider these two items to be the same, although abdominal pain, cramps, and pressure need further investigation.⁷ These two items were also correlated at over 0.80 with abdominal cramps and abdominal pressure for both day 1 and the week 1 mean scores. The results indicate that the items assessing gas and stomach pain should be omitted when using the instrument in future clinical trials.

Test-retest reliability

All symptom items on the IBS-D Daily Symptom Diary except 'Frequency of gas' met the threshold for test-retest reliability (ICC ≥ 0.70 ; range 0.778–0.834). For the Symptom Event Log, the mean number of events also met or surpassed the threshold of 0.70; however, mean immediacy, consistency, and emptying scores did not, as these symptoms are highly variable.⁷

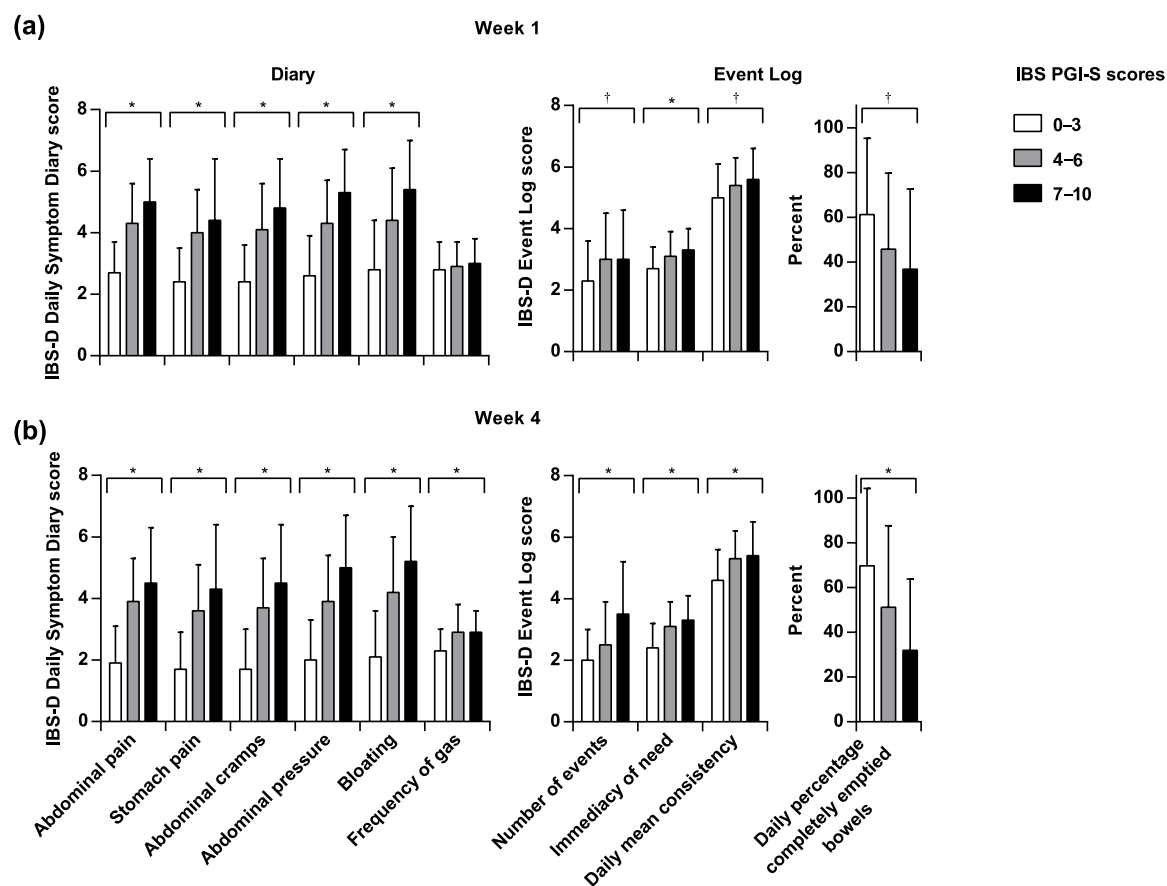


Figure 2. Known groups analysis using IBS PGI-S scores. IBS PGI-S scores indicating more severe symptoms (0–3, least severe; 4–6, moderately severe; 7–10, most severe) were associated with increasing severity on the Diary and Event Log at weeks 1 (a) and 4 (b). IBS PGI-S scores indicating more severe symptoms (0–3, least severe; 7–10, most severe) were associated with increasing severity on the Diary and Event Log at weeks 1 (a) and 4 (b). Between-group differences for each item except for ‘Frequency of gas’ at week 1 were significant. Note that lower scores indicate less severe symptoms. * $p < 0.001$; † $p < 0.01$ (ANOVA). ANOVA, analysis of variance; IBS-D, diarrhea-predominant irritable bowel syndrome; IBS PGI-S, IBS Patient Global Impression of Severity.

Clinical validity

Known groups analyses. The construct validity of the Diary and Event Log was evaluated by a ‘known groups’ comparison approach, in which patients were stratified according to severity based on the single-item IBS PGI-C or IBS PGI-S rating scales at weeks 1 and 4 of the treatment period.¹¹

IBS PGI-S. The IBS PGI-S scores were grouped as ‘0–3’ (least severe), ‘4–6’, and ‘7–10’ (most severe). Scores for all items on both the Diary and Event Log were higher among patients of higher IBS severity levels at weeks 1 and 4 of the treatment period, with statistically significant differences ($p < 0.001$) in the scores among all the groups, except for ‘Frequency of gas’ at week 1 [Figure 2(a, b)].

Five-item, four-item, and three-item Diary summary scores. The known groups approach was also used to compare summary scores of the Diary that included five-item (omitting ‘Frequency of gas’), four-item (also omitting ‘Stomach pain’), and three-item (also omitting ‘Bloating’) versions. Differences among severity groups defined by either approach were statistically significant at both weeks 1 and 4 of the treatment period for all three versions of the summary score [Figures 3(a,b)]. In addition, the mean values for all three versions of the Diary summary scores were greater in those groups identified as having more severe symptoms using the IBS PGI-S scores at weeks 1 and 4, and using the IBS PGI-C at week 1. The one exception was that, at week 4, mean values for all versions of the Diary

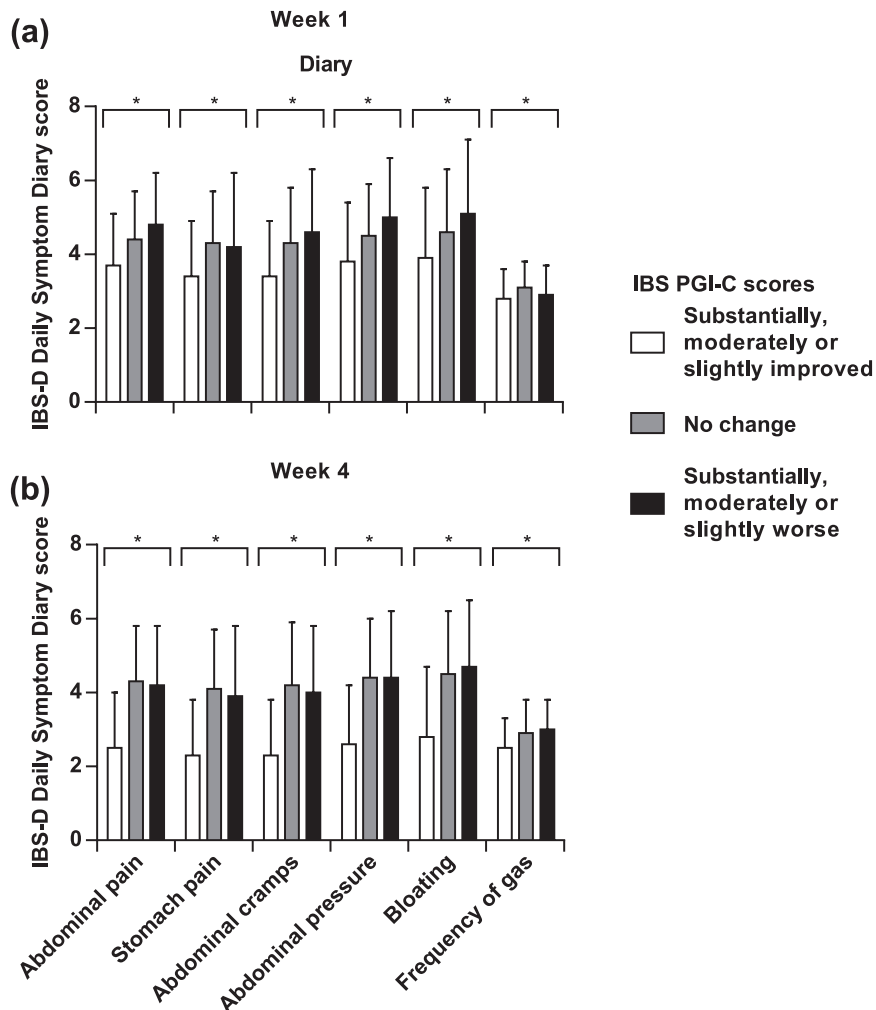


Figure 3. Known groups analysis of five-item, four-item, and three-item summary scores of the Diary using the IBS PGI-S and IBS PGI-C scores at weeks 1 (a) and 4 (b). * $p < 0.05$ [ANOVA]. ANOVA, analysis of variance; IBS-D, diarrhea-predominant irritable bowel syndrome; IBS PGI-C, IBS Patient Global Impression of Change; IBS PGI-S, IBS Patient Global Impression of Severity.

summary scores in the group identified using the IBS PGI-C as having ‘Substantially, moderately, or slightly worse’ symptoms were slightly less than those identified as having ‘No change’.

Responsiveness. The ability of the instrument to detect changes over time was evaluated by comparing changes in Diary and Event Log scores among responder groups defined by changes in the IBS PGI-S and IBS PGI-C scores between run-in week 2 and treatment week 4.

IBS PGI-S. Patients were stratified into three groups: ‘Improved’, ‘No change’, and ‘Worsened’ based on two-grade or higher improvements, less than two-grade change, or two-grade or higher worsening. For each item of both the Diary and Event Log except ‘Daily percentage completely

emptied bowels’, the pattern of mean changes in scores was consistent with their assignment to these responder groups based on the IBS PGI-S scores [Figure 4(a)]. Specifically, the greatest improvements in Diary and Event Log scores for each item were observed in those assigned to the ‘Improved’ group based on their IBS PGI-S scores. Between-group differences were significant for all Diary items and all Event Log items [Figure 4(a); $p < 0.05$, ANOVA]. Within-group changes over time from baseline in mean scores were significant for the ‘Improved’ groups on most items of the Diary and Event Log, and for the ‘No change’ groups on most items of the Diary ($p < 0.05$, Student’s t test).

For responder groups defined using the IBS PGI-C, the three groups of ‘Improved’, ‘No change’, and ‘Worsened’ were based on responses of

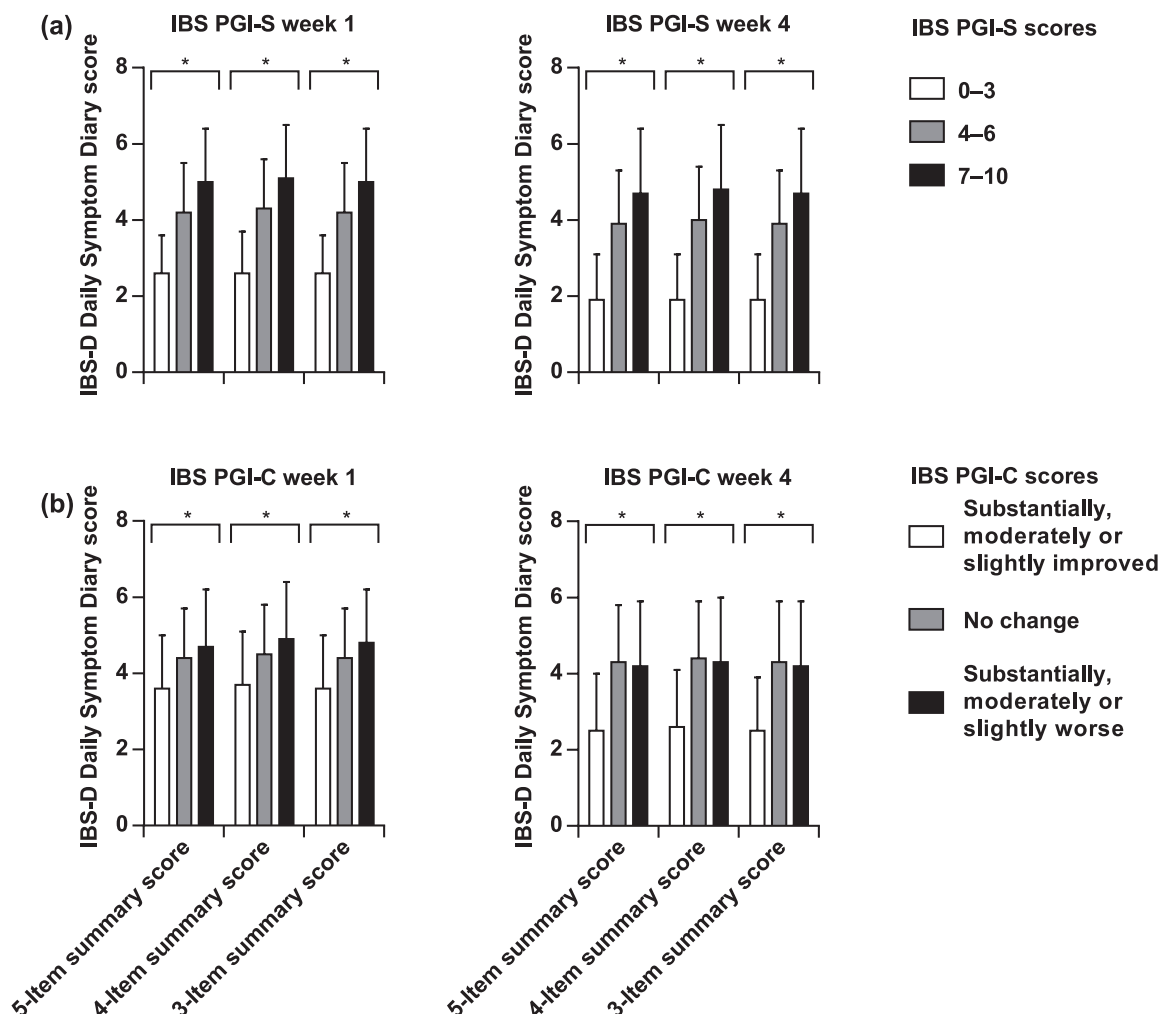


Figure 4. Responder analysis. The relative distributions of mean scores for almost every item on the Diary and Event Log, except for 'Daily percentage of completely emptied bowels', were consistent with their assignment to responder groups of 'Improved', 'No change', and 'Worsened' using either the IBS PGI-S (a) or IBS PGI-C (b). Note that negative values for changes in scores indicate improvement. *Between-group $p < 0.05$ (ANOVA); †within-group (change from baseline) $p < 0.05$ (Student's t test). ANOVA, analysis of variance; IBS-D, diarrhea-predominant irritable bowel syndrome; IBS PGI-C, IBS Patient Global Impression of Change; IBS PGI-S, IBS Patient Global Impression of Severity.

'Moderately improved' or 'Substantially improved', 'No change', and 'Moderately worse' or 'Substantially worse', respectively, at treatment week 4 compared with run-in week 2. The relative mean change in scores for almost every item on the Diary and Event Log were consistent with their assignment to each responder group [Figure 4(b)]. Between-group differences were significant for all items on the Diary and Event Log ($p < 0.05$, ANOVA). Within-group changes from baseline in mean scores were significant for the 'Improved' groups on all items of the Diary and Event Log, and for the 'No change' groups on 5 of the 10 total items ($p < 0.05$, Student's t test).

Five-item, four-item, and three-item Diary summary scores. These were also assessed using the responder groups defined by the IBS PGI-S or IBS PGI-C scores. For all three versions of the Diary summary score, the relative mean changes in the scores reflected their assignment to each responder group for both IBS PGI-S and IBS PGI-C [Figure 5(a, b)]. In the analyses using either IBS PGI-S or IBS PGI-C, between-groups differences were significant for all three versions ($p < 0.05$, ANOVA), and within-group change from baseline was significant for the 'Improved' and 'No change' groups ($p < 0.05$, Student's t test).

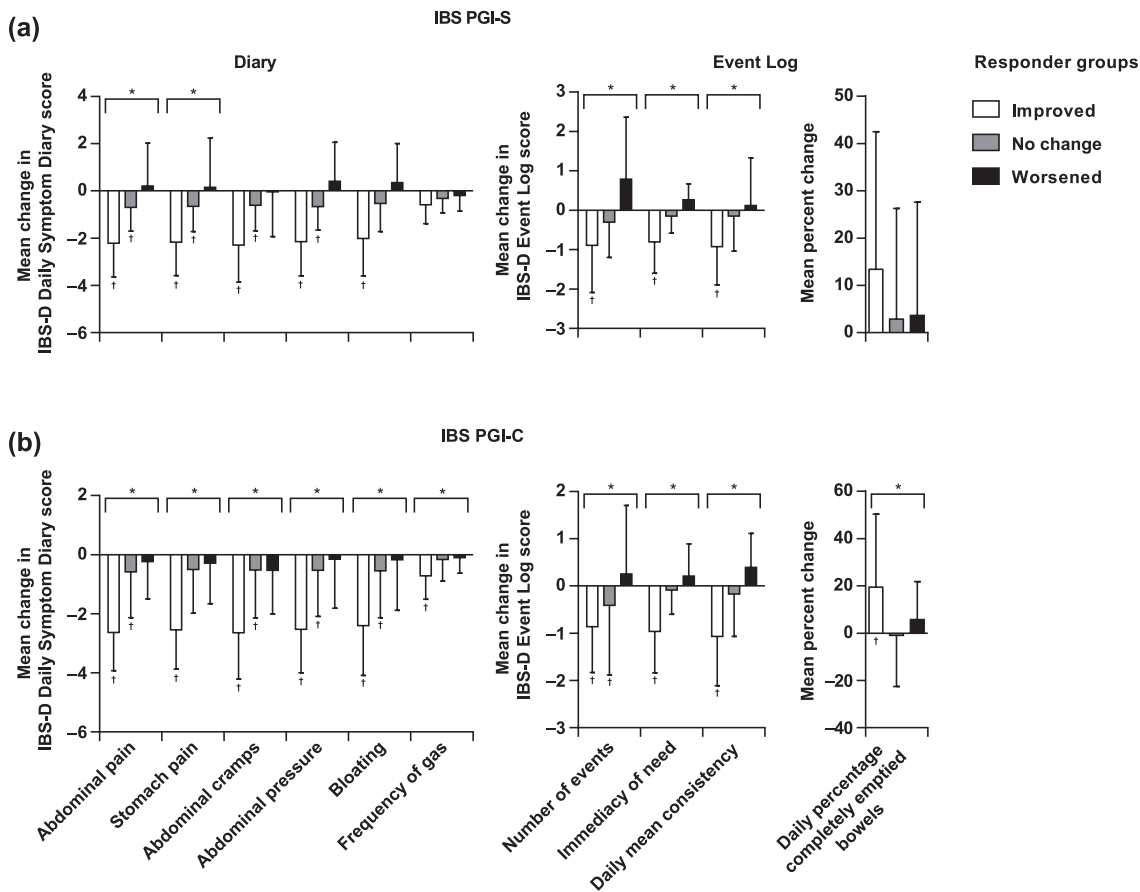


Figure 5. Responder analysis of Diary summary scores. The relative distribution of five-item, four-item, and three-item mean summary scores on the Diary were consistent with their assignment to responder groups of 'Improved', 'No change', and 'Worsened' using either the IBS PGI-S (a) or IBS PGI-C (b). Note that negative values for changes in scores indicate improvement. *Between-group $p < 0.05$ (ANOVA); †within-group (change from baseline) $p < 0.05$ (Student's t test). ANOVA, analysis of variance; IBS-D, diarrhea-predominant irritable bowel syndrome; IBS PGI-C, IBS Patient Global Impression of Change; IBS PGI-S, IBS Patient Global Impression of Severity.

Preliminary estimation of clinically important changes

To define the level of change in scores that can be considered clinically important for the instrument, an anchor-based approach was employed using changes in the IBS PGI-S and IBS PGI-C between run-in week 2 and treatment week 4 as anchors. For the IBS PGI-S, a one-grade reduction was defined as being a meaningful difference, whereas for the IBS PGI-C, patients who reported being 'Slightly improved' were considered to have experienced a clinically important change. Using these definitions, similar preliminary estimates of clinically important change were observed for all items on the Diary and Event Log (except 'Percentage of completely emptied bowels') and all three versions of the Diary summary score (Table 3). The

mean preliminary estimates of clinically important change ranged from -0.13 to -1.55 points (SD range 0.69 – 1.39), and the mean clinically important change for the average daily percentage of completely emptied bowels was 9.38 (SD 26.24) using IBS PGI-S as an anchor and 10.85 (SD 26.80) using the IBS PGI-C as an anchor.

Two distribution-based approaches were also used to estimate the clinically important change: 0.5 of a SD or the SEM as thresholds (see Methods). Again, similar clinically important change estimates were observed using either approach for all items on the Diary and Event Log (except 'Percentage of completely emptied bowels') and all three versions of the Diary summary score (Table 3). The mean clinically

Table 3. Interpretation of scores.

Score	Anchor-based estimations Mean (SD)		Distribution-based estimations Mean (SD)	
	IBS PGI-S <i>n</i> = 38	IBS PGI-C <i>n</i> = 73	0.5 SD <i>n</i> = 330	SEM <i>n</i> = 330
Diary score (mean)				
Abdominal pain	-0.79 (1.31)	-1.44 (1.36)	0.55	0.62
Stomach pain	-0.76 (1.21)	-1.47 (1.38)	0.65	0.61
Abdominal cramps	-0.79 (1.34)	-1.55 (1.39)	0.66	0.68
Abdominal pressure	-0.86 (1.39)	-1.44 (1.35)	0.70	0.61
Bloating	-0.81 (1.69)	-1.26 (1.37)	0.86	0.60
Frequency of gas	-0.13 (0.79)	-0.45 (0.69)	0.39	0.37
Summary score (five item)	-0.80 (1.30)	-1.43 (1.26)	0.58	0.53
Summary score (four item)	-0.81 (1.34)	-1.42 (1.27)	0.61	0.54
Summary score (three item)	-0.81 (1.29)	-1.48 (1.30)	0.57	0.56
Event Log average daily score				
Number of events	-0.54 (1.33)	-0.77 (1.26)	0.77	0.58
Immediacy of need (mean)	-0.16 (0.73)	-0.51 (0.66)	0.35	0.28
Consistency (mean)	-0.34 (0.81)	-0.72 (0.82)	0.38	0.45
Percentage of completely emptied bowels	9.38 (26.24)	10.85 (26.80)	16.75	10.61

IBS-D, diarrhea-predominant irritable bowel syndrome; IBS PGI-C, IBS Patient Global Impression of Change; IBS PGI-S, IBS Patient Global Impression of Severity; SD, standard deviation; SEM, standard error of measurement.

important change estimates ranged from 0.35 to 0.77. The mean clinically important change for the average daily percentage of completely emptied bowels was 16.75 using the 0.5 SD threshold and 10.61 using the SEM threshold.

As a sensitivity analysis with regard to the IBS PGI-C, an anchor of ‘moderately improved’ was also analyzed as being clinically important (Table 4). The mean preliminary estimates of clinically important change (except ‘Percentage of completely emptied bowels’) ranged from -0.62 to -2.32 (SD range 0.78–1.49). The mean clinically important change for the average daily percentage of completely emptied bowels was 21.39 (SD 30.86).

Discussion

The IBS-D Daily Symptom Diary and Event Log is an instrument, recently developed in line with the FDA guidance, which has the potential to address the need for assessment of the effects of

therapies in clinical trials. The data presented here represent the first evidence of the psychometric properties in a clinical trial population that reflects the targeted context of use. Importantly, this represents the first evidence of the ability of the instrument scores to detect changes over time in an interventional study. To date, evaluation of the psychometric properties and concurrent validity of the instrument compared with other instruments, including the IBS Severity Scoring System¹⁵ and the IBS Quality of Life Questionnaire has been performed using data from an observational study.⁷ The findings reported here are highly consistent with those preliminary findings, but this time from a clinical trial population. First, both quality of completion and the floor and ceiling effects were similar to those seen in the original validation study.⁸ The prespecified criteria for floor and ceiling effects were quite stringent in this observational setting for the purpose of psychometric testing. This is critical for a measure intended for use in a clinical trial in which changes over the course of treatment are primary or secondary endpoints. It is

Table 4. Interpretation of scores: sensitivity analysis on IBS PGI-C.

Score	Anchor-based estimations
	Mean (SD)
	IBS PGI-C <i>n</i> = 61
Diary score (mean)	
Abdominal pain	-2.32 (1.29)
Stomach pain	-2.25 (1.31)
Abdominal cramps	-2.26 (1.45)
Abdominal pressure	-2.25 (1.23)
Bloating	-2.11 (1.49)
Frequency of gas	-0.62 (0.78)
Summary score (five item)	-2.24 (1.22)
Summary score (four item)	-2.24 (1.23)
Summary score (three item)	-2.28 (1.23)
Event Log average daily score	
Number of events	-0.80 (0.99)
Immediacy of need (mean)	-0.71 (0.76)
Consistency (mean)	-0.95 (0.97)
Percentage of completely emptied bowels	21.39 (30.86)

IBS-D, diarrhea-predominant irritable bowel syndrome; IBS PGI-C, IBS Patient Global Impression of Change; SD, standard deviation.

assumed that the clinical threshold for study inclusion in future studies will screen subjects at the floor from inclusion in the study.

Similarly, the test–retest reliability results reported here were consistent with previous results and were above the suggested threshold criteria ($ICC \geq 0.70$).

The instrument also demonstrated preliminary responsiveness to longitudinal change as evaluated using changes in the IBS PGI-S and IBS PGI-C scores. The results of those analyses provide strong evidence that the instrument is able to separate patients whose condition is ‘improved’ from those whose condition is unchanged or worsened. In contrast, the evidence in relation to separating the unchanged and worsened groups was mixed; often there was little numerical difference between those two groups. However, this is

perhaps not surprising given the limited change observed, and may reflect the presence of a placebo effect.

Regarding quality of completion, no one item has notably higher missing data than any other item. This suggests that the missing data were random and did not reflect any problems with acceptability, relevance, or comprehension. Nevertheless, a study by Mujagic and colleagues reported that end-of-day diaries resulted in higher abdominal pain and flatulence scores compared with day-average scores in patients with IBS.¹⁶

One of the major strengths of the IBS-D Symptom Diary and Event Log is that it provides rigorous assessment of all important symptoms of IBS-D. The aspects of IBS-D assessed by this instrument provide more specific information regarding a patient’s condition or response to therapy compared with a single-item rating system such as the IBS PGI-C. Nevertheless, the ultimate usefulness of an instrument such as this involves a balance between comprehensiveness and feasibility/respondent burden; such issues are particularly important for instruments that are completed daily/per event, such as this one. On initial evaluation of psychometric validity, item level results for the ‘Frequency of gas’ item were relatively weak and suggested that this item may not be closely related to severity. Here, a five-item summary score omitting the ‘Frequency of gas’ item was evaluated and results from analyses with that version were equally strong, suggesting that this item may be dispensable. Given that the four-item and three-item summary scores for the Diary also provided remarkably similar results, there may be room to streamline the instrument further. In fact, a previous analysis suggests that ‘Stomach pain’, which was omitted from the four-item summary score, is a strong candidate for deletion as it is so strongly related to ‘Abdominal pain’.⁷

Astellas engaged a group of gastroenterology experts to understand the relevance and clinical applicability of the instrument. The panel was probed about bloating and its potential of being multidimensional. The panel stated that bloating was observed in clinics and seemed very important to patients from the clinical perspective; therefore, it was recommended not to delete the bloating item, but instead to keep it and utilize it as an exploratory endpoint. Another example is the three

items associated with abdominal pain (abdominal cramps, pain, and pressure), which Astellas thought should be evaluated as a single score, and the experts concurred. Upon a complete review of the results, the experts were supportive of the relevance, clinical applicability of the instrument, and conceptual framework with regard to items included and their respective domains.

The results presented here provide evidence that the findings of the earlier psychometric evaluation can be replicated in a clinical trial population, thus supporting the instrument's use in that setting. Inter-item correlations showed a logical pattern of correlations that was consistent with that demonstrated in the observational study.⁷ The instrument was again able to consistently distinguish among patients who differed in their self-rated severity levels based on other assessments, thus supporting its clinical validity. This provides evidence that the instrument is assessing symptoms of IBS-D that are salient and important to patients and, again, this is consistent with findings from the observational study.⁷ The instrument also demonstrated preliminary responsiveness to longitudinal change as evaluated using changes in the IBS PGI-S and IBS PGI-C scores. Change scores and distribution-based analyses were also used to provide preliminary estimates of the level of changes in scores that can be considered clinically important and meaningful. Sensitivity analyses utilizing a higher level of improvement (i.e. 'Moderately improved' on the IBS PGI-C) demonstrated that mean change scores incrementally increased, which further demonstrates the instrument's ability to change in concert with other measures. Further, the sensitivity analyses force the discussion of what changes should be considered clinically meaningful and what changes are meaningful to the patient. That is, what is seen as a lack of clinical improvement (i.e. a one-point change) may be an important improvement to the patient. Given that limited change was seen in this study, these initial estimates should be interpreted with caution.

The current analysis has some limitations. The lack of a well established 'gold-standard' assessment of IBS-D severity means that any anchor-based analyses to estimate minimal clinically important difference (MCID) have limitations and should be interpreted with caution. Moreover, the regulatory recommended primary endpoint and this instrument revealed no difference in the response rates between the investigational drug

and placebo. Thus, caution should be applied in interpreting the responsiveness and anchor-based analyses given that they are based on a study with an unsuccessful intervention. Nevertheless, the analyses evaluating validity and reliability provide strong evidence that the psychometric properties established in an observational study are equally strong in the context of use of an interventional clinical trial.

In addition, the responder analyses need to be interpreted with caution as they are based on a study in which changes over time were limited. The appropriateness of the clinically important change values will ultimately be strengthened by replication of these analyses in a study that includes successful therapeutic intervention.

Despite those caveats, there is reason to be optimistic about the validity of the instrument since positive results of psychometric analyses are now documented in several distinct study populations.⁵⁻⁷ Thus, this instrument may provide a useful tool to assess the efficacy of potential IBS-D therapies. Furthermore, it may prove useful in a clinical setting because it will provide clinicians with a means to more accurately assess changes in the status of IBS-D in patients and therefore enable improved treatment options.

In summary, this new instrument provides a valid and reliable assessment of IBS-D severity, as reported in the previous qualitative and psychometric papers.⁵⁻⁷ The results of this interventional study provide further evidence that this instrument has potential for measuring the effect of treatment in clinical trials, in particular, when the desired effects are reported and known only to the patients. It is recommended that the anchor-based analyses performed to define MCID are further evaluated in a study that includes successful therapeutic intervention. Finally, the instrument may also prove useful in the real-world clinical setting, because it will provide clinicians with a means to more accurately assess changes in the status of IBS-D in patients and, therefore, provide valuable information to evaluate treatment options.

Acknowledgements

Medical writing assistance was provided by Ed Parr, PhD, CMPP, and Gill Sperrin, CBiol, MRSB, CMPP, of Envision Pharma Group, which was funded by Astellas. Analysis was performed by Jeffrey McDonald, MS, of Adelphi Values.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

L. Delgado-Herrera and C. Lademacher and B. Zeiher are employees of Astellas. K. Lasch is an employee of Pharmerit International. A. Lembo has received consultancy fees from AstraZeneca, Ironwood/Forest, Prometheus, and Salix outside the submitted work. D.A. Drossman has received consultancy fees from Shire and Salix. R. Arbuckle, B. Banderas, and K. Rosa are employees of Adelphi Values (formerly Mapi Values), which was under contract by Astellas for this work; K Rosa has a patent pending.

References

1. Longstreth GF, Thompson WG, Chey WD, *et al.* Functional bowel disorders. *Gastroenterology* 2006; 130: 1480–1491.
2. U.S. Department of Health and Human Services Food and Drug Association. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>, 2009.
3. U.S. Department of Health and Human Services Food and Drug Association. Guidance for industry irritable bowel syndrome – clinical evaluation of drugs for treatment. www.fda.gov/downloads/Drugs/Guidances/UCM205269.pdf, 2012.
4. Trentacosti AM, He R, Burke LB, *et al.* Evolution of clinical trials for irritable bowel syndrome: issues in end points and study design. *Am J Gastroenterol* 2010; 105: 731–735.
5. Lasch K, Delgado-Herrera L, Kothari S, *et al.* The irritable bowel syndrome-diarrhea (IBS-D) daily symptom diary and event log: a newly developed patient-reported outcome (PRO) measure. *Gastroenterology* 2011; 140: 612.
6. Marquis P, Lasch KE, Delgado-Herrera L, *et al.* Qualitative development of a patient-reported outcome symptom measure in diarrhea-predominant irritable bowel syndrome. *Clin Transl Gastroenterol* 2014; 5: e59.
7. Rosa K, Delgado-Herrera L, Zeiher B, *et al.* Psychometric assessment of the IBS-D daily symptom diary and symptom event log. *Qual Life Res* 2016; 25: 3197–3208.
8. Lasch K, Delgado-Herrera L, Tesler Waldman L, *et al.* Development of a new instrument to assess stool form and consistency in irritable bowel syndrome with diarrhea. *Gastroenterol Hepatol* 2016; 4: 00084.
9. Lexicon Pharmaceuticals I. Lexicon completes phase 2 study of LX1033 in IBS-D 2014, <http://www.lexpharma.com/news/press-releases/2309-lexicon-completes-phase-2-study-of-lx1033-in-ibs-d.html>.
10. Lewis SJ and Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol* 1997; 32: 920–924.
11. Hays RD, Anderson R and Revicki DA. Quality of life assessment. In: Staquet M, Hays RD and Fayers PM (eds) *Clinical trials methods and practice*. New York: Oxford University Press, 1998.
12. Hays RD, Farivar SS and Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005; 2: 63–67.
13. Turner D, Schunemann HJ, Griffith LE, *et al.* The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010; 63: 28–36.
14. Thompson B. *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association, 2004.
15. Francis CY, Morris J and Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. *Aliment Pharmacol Ther* 1997; 11: 395–402.
16. Mujagic Z, Leue C, Vork L, *et al.* The experience sampling method – a new digital tool for momentary symptom assessment in IBS: an exploratory study. *Neurogastroenterol Motil* 2015; 27: 1295–1302.

Visit SAGE journals online
[journals.sagepub.com/
home/tag](http://journals.sagepub.com/home/tag)

 SAGE journals