Title

Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) Gastrointestinal Symptom Scales

Journal

The American Journal of Gastroenterology, 109(11)

ISSN

0002-9270

Authors

Spiegel, Brennan MR

Hays, Ron D

Bolus, Roger

et al.

Publication Date

2014-11-01

DOI

10.1038/ajg.2014.237

Peer reviewed

Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) Gastrointestinal Symptom Scales

Brennan M.R. Spiegel, MD, MSHS, RFF, FACG, AGAF[1,2,3,4], Ron D. Hays, PhD[4,5], Roger Bolus, PhD[2], Gil Y. Melmed, MD, MS[1], Lin Chang, MD[5,6], Cynthia Whitman, MPH[2], Puja P. Khanna, MD, MPH[7], Sylvia H. Paz, PhD[4], Tonya Hays, MS[4], Steve Reise, PhD[8] and Dinesh Khanna, MD, MSc[7]

OBJECTIVES: **The National Institutes of Health (NIH) Patient-Reported Outcomes Measurement Information System (PROMIS®) is a standardized set of patient-reported outcomes (PROs) that cover physical, mental, and social health. The aim of this study was to develop the NIH PROMIS gastrointestinal (GI) symptom measures.**

METHODS: **We first conducted a systematic literature review to develop a broad conceptual model of GI symptoms. We complemented the review with 12 focus groups including 102 GI patients. We developed PROMIS items based on the literature and input from the focus groups followed by cognitive debriefing in 28 patients. We administered the items to diverse GI patients (irritable bowel syndrome (IBS), inflammatory bowel disease (IBD), systemic sclerosis (SSc), and other common GI disorders) and a census-based US general population (GP) control sample. We created scales based on confirmatory factor analyses and item response theory modeling, and evaluated the scales for reliability and validity.**

RESULTS: **A total of 102 items were developed and administered to 865 patients with GI conditions and 1,177 GP participants. Factor analyses provided support for eight scales: gastroesophageal reflux (13 items), disrupted swallowing (7 items), diarrhea (5 items), bowel incontinence/soilage (4 items), nausea and vomiting (4 items), constipation (9 items), belly pain (6 items), and gas/bloat/flatulence (12 items). The scales correlated significantly with both generic and disease-targeted legacy instruments, and demonstrate evidence of reliability.**

CONCLUSIONS: **Using the NIH PROMIS framework, we developed eight GI symptom scales that can now be used for clinical care and research across the full range of GI disorders.**

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at http://www.nature.com/ajg

Am J Gastroenterol advance online publication, 9 September 2014; doi:10.1038/ajg.2014.237

INTRODUCTION

Patients typically seek health care because they experience symptoms. This is especially true in gastroenterology where most digestive disorders initially present with symptoms rather than biochemical abnormalities alone. To fully describe the illness experience of gastrointestinal (GI) patients, providers must elicit, measure, and interpret patient symptoms as part of their clinical evaluation (1,2).

[1]Department of Gastroenterology, Cedars-Sinai Medical Center, Los Angeles, California, USA; [2]Cedars-Sinai Center for Outcomes Research and Education, Los Angeles, California, USA; [3]Department of Gastroenterology, VA Greater Los Angeles Healthcare System, Los Angeles, California, USA; [4]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, California, USA; [5]Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA; [6]Gail and Gerald Oppenheimer Family Center for Neurobiology of Stress, David Geffen School of Medicine at UCLA, Los Angeles, California, USA; [7]Division of Rheumatology, University of Michigan, Ann Arbor, Michigan, USA; [8]Department of Psychology, UCLA, Los Angeles, California, USA. **Correspondence:** Brennan M.R. Spiegel, MD, MSHS, RFF, FACG, AGAF, Department of Gastroenterology and Health Services, Cedars-Sinai Medical Center, Pacific Theatre Building, 116 N. Robertson Blvd, 4th Floor, Los Angeles, California 90048, USA or Dinesh Khanna, MD, MSc, University of Michigan Scleroderma Program, Division of Rheumatology/Department of Internal Medicine, 300 North Ingalls Street, Suite 7C27, Ann Arbor, Michigan 48109, USA. E-mail: bspiegel@ucla.edu or khannad@med.umich.edu

Received 24 December 2013; accepted 24 June 2014

Patient-generated reports, also known as patient-reported outcomes (PROs), capture the patients' illness experience in a structured format and may help providers understand symptoms from the patients' perspective (1). PROs measure any aspect of health directly reported by the patient (e.g., physical, emotional, or social symptoms) and can help to direct care and improve clinical outcomes (3–9). When clinicians systematically collect patient-reported data in the right place at the right time, PRO measurement can effectively aid in detection and management of conditions (3,4), improve satisfaction with care (5), and enhance the patient–provider relationship (5–9).

The National Institutes of Health (NIH) launched the Patient-Reported Outcomes Measurement Information System (PROMIS®) in 2004 with the goal of developing, evaluating, and disseminating a toolbox of publicly available item banks capable of measuring PROs across the breadth and depth of the human illness experience (www.nihpromis.gov) (10). Moreover, PROMIS measures are designed for either traditional paper-and-pencil or electronic modes of data collection. The NIH PROMIS vision is to create highly efficient and short questionnaires that are feasible to implement in busy clinical systems while preserving reliability and validity. PROMIS is a system that offers the potential for establishing common-language benchmarks for symptoms across conditions and identifying clinical thresholds for action and meaningful improvement or decline.

In the field of gastroenterology, patients, providers, investigators, and regulators are interested in using PROs to guide clinical decision making (1), conduct clinical research (1), and achieve drug approval (11). Over the past two decades, investigators have developed over 100 disease-targeted PROs that measure a range of GI symptoms (12). However, the field remains in need of a standardized, rigorously developed, electronically administered set of PROs that span the breadth and depth of GI symptoms, and can be used across all GI disorders for clinical and research purposes.

This paper describes content and cross-sectional construct validation of the NIH PROMIS GI symptom scales using data from diverse GI patients and members of the general population (GP).

METHODS

Study overview and objectives

We sought to develop and evaluate a new set of PROMIS GI symptom scales that capture the breadth and depth of physical symptoms associated with the GI system. We designed the scales to be applicable to both the GP and patients with a defined GI illness. The scales were designed to be *system targeted* for GI overall rather than *disease targeted*; there are already over 100 disease-targeted scales in GI (12). To develop the PROMIS GI symptom scales, we followed published criteria for qualitative and quantitative development of NIH PROMIS measures with oversight from the NIH PROMIS Steering Committee (10,13,14). The study involved three phases conducted over a 4-year period: (i) development of candidate items (phase 1), (ii) qualitative item review (phase 2), and (iii) quantitative psychometric testing (phase 3). We describe the methods for each phase in the sections, below.

Phase 1: item development

Systematic literature review . We performed a structured search to identify English-language PROs across all luminal diseases and other illnesses that directly affect the GI tract (e.g., systemic sclerosis—a "non-GI" condition that affects GI function). Next, we developed a search strategy that targeted studies of English-language PROs that measure GI symptoms and abstracted individual items from each PRO to develop a comprehensive item library. Then, we developed "bins" to categorize items describing GI symptoms, and used this to assess a framework for GI symptom reporting, similar to one developed previously for irritable bowel syndrome (IBS) (15) and in line with the process supported by the NIH PROMIS network (14). After binning items into defined categories, we "winnowed" items that were similar, leaving only items that covered unique symptom attributes. We presented our results to an expert panel consisting of three gastroenterologists with PRO expertise that provided feedback and identified additional PROs and candidate items (William Chey (University of Michigan), Douglas Drossman (University of North Carolina), and Jan Irvine (University of Toronto)). We previously reported the extended methods and results of this search that culminated in the "GI-PRO database"—a publicly available search engine to identify extant GI PROs (http://www.researchcore.org/gipro/) (12).

Focus groups. In order to gain insights from patients about their GI-related symptoms, we conducted 12 disease-specific focus groups. We conducted the groups at the University of California Los Angeles (UCLA) and the West Los Angeles Veteran Administration (WLAVA) campuses between 13 November 2010 and 12 February 2011. Subjects were eligible if they were diagnosed by a physician with gastroesophageal reflux disease, inflammatory bowel disease (IBD), IBS, or systemic sclerosis (SSc); these conditions span the breadth and depth of GI symptoms. We next recruited participants across gender, ethnicity, and education levels and identified patients through recruitment from the GI clinics at UCLA, WLAVA, and Cedars-Sinai Medical Center. Additional participants were recruited through flyers distributed around UCLA clinics and through online advertisements using Craigslist. Before the focus groups, we developed a guide with patient instructions, open-ended think-aloud exercises, and scripted probes. An experienced moderator led each group with assistance from a co-facilitator (refer to **Supplementary Appendix A** online for the moderator's guide).

Each focus group lasted ~90 min and consisted of 6 to 12 participants (average 8 per group). The interviews were audiotaped and transcribed for analysis. There were three focus groups for each of the four GI disorders.

We asked patients to describe their illness experience in their own words and without prompting. Through group interaction, we identified common and unique language used to describe GI symptoms and their attributes. We conducted multiple groups to ensure that interactions of a single group did not bias any one conclusion and to provide greater generalizability.

Qualitative data collection and analysis methods

We analyzed the transcribed focus group text using ATLAS.ti software (ATLAS.ti Scientific Software Development, Berlin, Germany)—a qualitative analysis program that allows coding of patient language and classification of vocabulary into major and minor concepts. The evaluation process included generation of key words, phrases, and quotes regarding GI symptoms. To be considered credible, concepts needed to be raised in an unsolicited manner by more than one participant in a single group and by participants in more than one group. We used ATLAS.ti to generate code count histograms within major and minor symptom concepts, and developed a symptom network among concepts to depict a framework describing the breadth and depth of GI symptoms.

Phase 2: qualitative item review

Developing draft PROMIS items. After developing our initial PRO item library and expanding it with input from patient focus groups, we next developed draft items. As the extant items varied in terms of phrasing styles, recall periods, response options, and literacy demands, we streamlined the items into a uniform style to create a harmonized item set using published PROMIS standards (14). We employed the following principles to create new items for the PROMIS GI symptom banks:

- Does not exceed a sixth grade reading level based on the validated "simple measure of gobbledygook" (SMOG) calculator (16).
- Minimizes ambiguity or cognitive difficulty.
- Avoids multi-barreled questions.
- Are as concise and simply worded as possible, attempting to use common English words and avoiding slang.
- Employ a 7-day recall period (standard PROMIS recall period (14)).
- Meets criteria for optimal translatability into non-English languages, as established by NIH PROMIS "translatability review" by the PROMIS linguist.

Next we created response scales for each item. For *bothersomeness* and *interference* of GI symptoms, we employed a five-point categorical response scale ranging from "not at all" to "very much," a preferred response scale for PROMIS (14). For frequency items we used the PROMIS five-level frequency scale (14). For bowel controllability we employed the PROMIS five-level capability scale ranging from "without any difficulty" to "unable to control" (14). For other items we created unique response sets that optimally suited the concept of interest, as necessary.

Patient cognitive debriefing for content validity. Following item development, we prepared a scripted interview to elicit patient feedback on the draft items. The script was based on guidance from PROMIS to evaluate respondent perceptions about language, comprehensibility, ambiguity, and relevance of item (see **Supplementary Appendix B** for moderator's guide) (14). The purpose of these interviews was to identify potentially problematic items and response scales, to help clarify and rewrite items that were not well understood, and to add additional items not already included in the bank. We developed our debriefing protocol to measure the following patient cognitions:

- *Comprehension*: What did the patient believe the question was trying to ask?
- *Memory retrieval process*: What strategy did the patient employ to retrieve information to answer the question?
- *Social desirability*: Was the patient motivated by social desirability in answering the question?
- *Response processing*: Did the patient's internal response metric for an item match the question's response options?

We used the retrospective verbal prompting technique to gauge these cognitions, following prior PROMIS work (14). After each draft item was completed, an interviewer posed scripted probes to elicit the patient's perceptions about the item and its response choices. We employed a standard set of probes developed and published by the PROMIS network (14). For example, following completion of an item, we asked: "In your own words, what do you think this question is asking?"

We obtained feedback from at least 10 patients from each patient group. Based on feedback and discussion, we created an updated set of items that included variations of the original items and additional items. Consistent with PROMIS standards, we then subjected the revised questionnaire to five additional patient interviews (14).

On the basis of these additional interviews and revisions, we crafted a final iteration of the items for subsequent testing. Finally, we classified each item on a matrix referring to the dimension of interest (e.g., intensity, frequency, difficulty, interference, predictability, bothersomeness) arranged in accordance with our previously described conceptual framework of GI symptoms (1,12). This process yielded our full PROMIS item set for subsequent psychometric evaluation, discussed below.

Phase 3: quantitative psychometric testing

In phase 3 of development, we sought to evaluate the psychometric properties of the PROMIS GI symptom scale by: (i) assessing the dimensionality of the scales and evaluating fit of item response theory (IRT) models in patients with different GI disorders and in representative members of the US GP; and (ii) evaluating the associations of the scales with legacy PRO instruments for GI illness and with patient-reported symptom severity. We tested the PROMIS GI Symptom scales in a diverse sample of GI patients and in a nationwide sample of the US GP for purposes of norming.

Selection of patients. We recruited participants from outpatient clinical practices and patients seeking care at university, community, and VA institutions. We invited patients seeking care at these outpatient clinics for an active GI symptom, including abdominal pain, bloating, nausea, vomiting, diarrhea, incontinence, constipation, dysphagia, or acid reflux. Our sample included patients with IBD seeking care at Cedars-Sinai Medical Center, a tertiary center in Los Angeles; patients with GI symptoms from SSc seeking care at rheumatology clinics at the University of Michigan; patients with functional GI disorders seeking care at a specialty clinic at

UCLA; and patients with diverse GI conditions seeking care at a general GI clinic at WLAVA. In addition, we partnered with the IFFGD (International Foundation for Functional Gastrointestinal Disorders) to survey a cohort of patients with diverse functional GI disorders enrolled in IFFGD mailing lists. The overall goal of this recruitment strategy was to enroll a widely diverse population of GI patients with active symptoms, ranging in demographics, disease type, and disease severity.

All patients were invited to complete the confidential online survey instrument, administered by Survey Monkey software (www.surveymonkey.com). Patients without Internet access could request paper surveys sent to their home, or completed in clinic, as needed. Patients were excluded from participation if they failed to provide informed consent or if they had cognitive impairment that would interfere with participation.

Selection of controls. In addition to GI patient recruitment, Cint (www.cint.com), a survey research firm, recruited a sample of individuals representative of the GP in terms of gender, ethnicity, race, and education level based on the 2010 census. Subjects were required to be 18 years of age and able to read English; there were no other exclusion criteria applied to the GP sample. Cint maintains panels with several million subjects across the United States. Cint maintained the PROMIS survey open until such time as the survey met all prespecified census-defined demographic requirements. This was completed within 3 weeks of opening the survey.

Measurements. In addition to the GI PROMIS Symptom items and demographic questions, we administered a wide range of concurrent legacy instruments that capture the biopsychosocial range of GI distress (2), including: (i) Visceral Sensitivity Index (17,18); (ii) PROMIS global health items (19); (iii) GI-specific global health item ("In the past 7 days, how would you rate your gastrointestinal condition? (excellent, very good, good, fair, or poor)); (iv) Gastrointestinal Symptom Rating Scale (20); and (v) EuroQOL health utility index (21). In addition to completing the common set of legacy instruments, patients completed relevant disease-targeted legacy instrument: IBS patients completed the IBS-QOL (Irritable Bowel Syndrome Quality of Life) (22,23), IBD patients completed the IBDQ (Inflammatory Bowel Disease Questionnaire) (24,25), and SSc patients completed the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract (GIT) 2.0 (26).

Psychometric analyses.

Overview of analyses: We first calculated descriptive statistics for demographic characteristics of the GP subjects and GI patients, including age, gender, race/ethnicity, education, marital status, and employment. We then followed PROMIS methodology to conduct quantitative psychometric analyses of the PROMIS items with the goal of developing symptom-specific scales based on IRT assumptions (13). Once these scales were created, we tested the construct validity of the resulting PROMIS scales against legacy instruments. In this report we present the cross-sectional psychometric analyses. Future reports will present longitudinal analyses including responsiveness to change and estimation of minimum clinically important differences for each scale.

IRT analyses: We first evaluated the extent to which items satisfied the IRT assumptions of monotonicity and unidimensionality. Monotonicity means that the probability of selecting a more favorable response option increases as the underlying health increases, and vice versa. Unidimensionality means the items in a scale measure a common underlying symptom domain. We evaluated dimensionality using confirmatory factor analytic methods. We fitted confirmatory categorical factor analytic models using MPLUS (Muthen & Muthen, Los Angeles, CA) in order to estimate polychoric correlations to adjust for ordinal rating scale data. We focused on practical fit indices such as the comparative fit index, as well as factor loadings and average absolute residual correlations to evaluate local dependence. We calibrated scales using the graded response model.

Reliability and information: We estimated internal consistency reliability and information at different points along the underlying scale for each PROMIS GI scale.

Construct validity: One method of establishing the validity of a PRO is to measure its relationship with other established legacy instruments. Thus, we hypothesized *a priori* that the PROMIS scales would significantly correlate with the five legacy instruments previously listed in the "Measures" section. We measured Pearson's correlation coefficients between each PROMIS GI symptom scale and each of the legacy instruments.

This study was approved by the institutional review boards of the West Los Angeles VA (PCC no. 0020), University of California at Los Angeles (IRB no. 11-003065), Cedars-Sinai Medical Center (PRO00027093), and the University of Michigan (HUM00052942), and was funded by grant NIH/NIAMS U01 AR057936A, the National Institutes of Health through the NIH Roadmap for Medical Research grant (AR052177).

RESULTS

Systematic literature review

The search strategy identified 15,697 titles, of which 183 met our final inclusion criteria There were 126 PRO instruments comprising over 2,300 GI symptom items, described in a previous publication (12). Item binning identified eight overarching symptom domains: (i) abdominal pain, (ii) gas/bloating, (iii) diarrhea, (iv) constipation, (v) bowel incontinence/soiling, (vi) heartburn/reflux, (vii) swallowing, and (viii) nausea/vomiting. We used these categories to guide our subsequent focus groups and item development.

Focus groups and cognitive interviews

Participants. **Table 1** shows demographic information of the 130 total participants in the qualitative research phase (102 in focus groups and 28 in cognitive interviews). The sample was demo-

Table 1. Descriptive characteristics of qualitative research sample

Characteristic	Values (*N*=130)
Mean age in years (range)	59 (24–86)
Gender	51% Female
Education	
High school graduate or less	12%
Some college	39%
College graduate	28%
One or more years after college	20%
Race/ethnicity	
White	69%
Black or African American	18%
Asian	5%
American Indian/Alaskan Native	2%
Other	1%
More than one race	1%
Hispanic/Latino	16%

graphically and clinically diverse. Of the 130 participants, there were 29%, 25%, 21%, and 25% with a functional GI disorder, IBD, SSc, and gastroesophageal reflux disease, respectively.

ATLAS.ti coding results. Participants in the focus groups spontaneously reported a diverse range of symptoms. Analysis of the transcripts yielded 42 unique codes grouped into the eight symptom domains. **Figure 1** shows the resulting conceptual framework resulting from ATLAS.ti coding of the symptom described by patients.

Qualitative item and scale development. Based on the literature search and focus groups, we developed candidate items within eight symptom domains. Overall, we found that the items were widely considered to be simple, understandable, and relevant in cognitive interviews. After iterative modification of the items, we developed 102 items contained within eight hypothesized domains, defined below based on qualitative item development:

Domain name: abdominal pain: Similar to previous work (15,27), we found that abdominal pain is multifaceted and can vary in location, intensity, and quality. Patients described how certain dimensions of pain drive illness severity more than others. The intensity, nature (sharp vs. dull), frequency, bothersomeness, and predictability (e.g., ability to tell in advance when a pain episode would occur) all contributed toward GI pain severity. In addition, patients indicated that involvement of more abdominal regions was related to higher pain severity. The items in the resulting PROMIS abdominal pain scale assess all dimensions of abdominal pain experienced over the past 7 days.Domain name: gas/bloating: The gas/bloating domain includes four facets: (i) bloating sensation (i.e., feeling pressure or fullness), (ii) bloating appearance (i.e., belly swollen or larger than usual size), (iii) flatulence (i.e., passing gas), and (iv) gurgling or rumbling. The first two facets reflect that bloating was described in terms of both its look and feel. "Flatulence" is a related but separate symptom that indicates passing gas (in contrast to gas retention with subsequent visible bloating). Flatulence was largely considered to be a discomfort symptom grouped within the bloating complex rather than as a defecation-related symptom, principally because flatulence most often occurs outside the context of bowel movements. The fourth facet is another related but separate symptom that refers to abdominal sounds. Gurgling or rumbling sounds were associated with gas and bloating. The gas/bloat domain items assess: (i) the frequency, sensation, appearance, predictability, and impact (bothersomeness and/or impact on daily activities) of gas/bloating during the past 7 days; (ii) the frequency and impact of flatulence during the same period; and (iii) the frequency of gurgling or rumbling during the same period.

Domain name: diarrhea: Diarrhea refers to loose, watery stools, urgency, and frequent bowel movements. The diarrhea items focus on capturing the frequency, form, bothersomeness, impact, controllability, and predictability of bowel urgency during the past 7 days.

Domain name: constipation: Constipation is the second defecation domain and encompasses the facets or cardinal subsymptoms of incomplete evacuation, straining, infrequent stools, and hard stools. Associated symptoms of rectal pain and need for manual maneuvers to facilitate stool evacuation are also assessed. The constipation domain items address the frequency, intensity, bothersomeness, and/or impact of all these facets of constipation during the past 7 days.

Domain name: bowel incontinence: This domain encompasses symptoms pertaining to a spectrum of bowel incontinence. Bowel incontinence was usually described as "having accidents" by most patients. This can be associated with bowel urgency or it can occur without the patient's awareness. In addition, however, some patients described stool leakage or "soiling" as a separate yet related symptom. Some patients described "passing gas" but subsequently finding out they also soiled their underwear, referred to as "gas incontinence." The bowel incontinence domain terms address frequency of these symptoms during the past 7 days.

Domain name: gastroesophageal reflux (GER): GER is the first of three domains associated with the foregut. The GER domain items assess four facets of patients' GER-related symptoms, including: (i) sensations associated (reflux, regurgitation) or unassociated (lump in the throat) with food intake; (ii) painful sensations (heartburn, chest pain, throat burn); and (iii) belching gas (burping)/hiccups. The GER items address the frequency, amount, bothersomeness, and/or impact of these symptoms during the past 7 days.

Domain name: nausea/vomiting: The nausea/vomiting domain encompasses a range of increasingly severe foregut symptoms

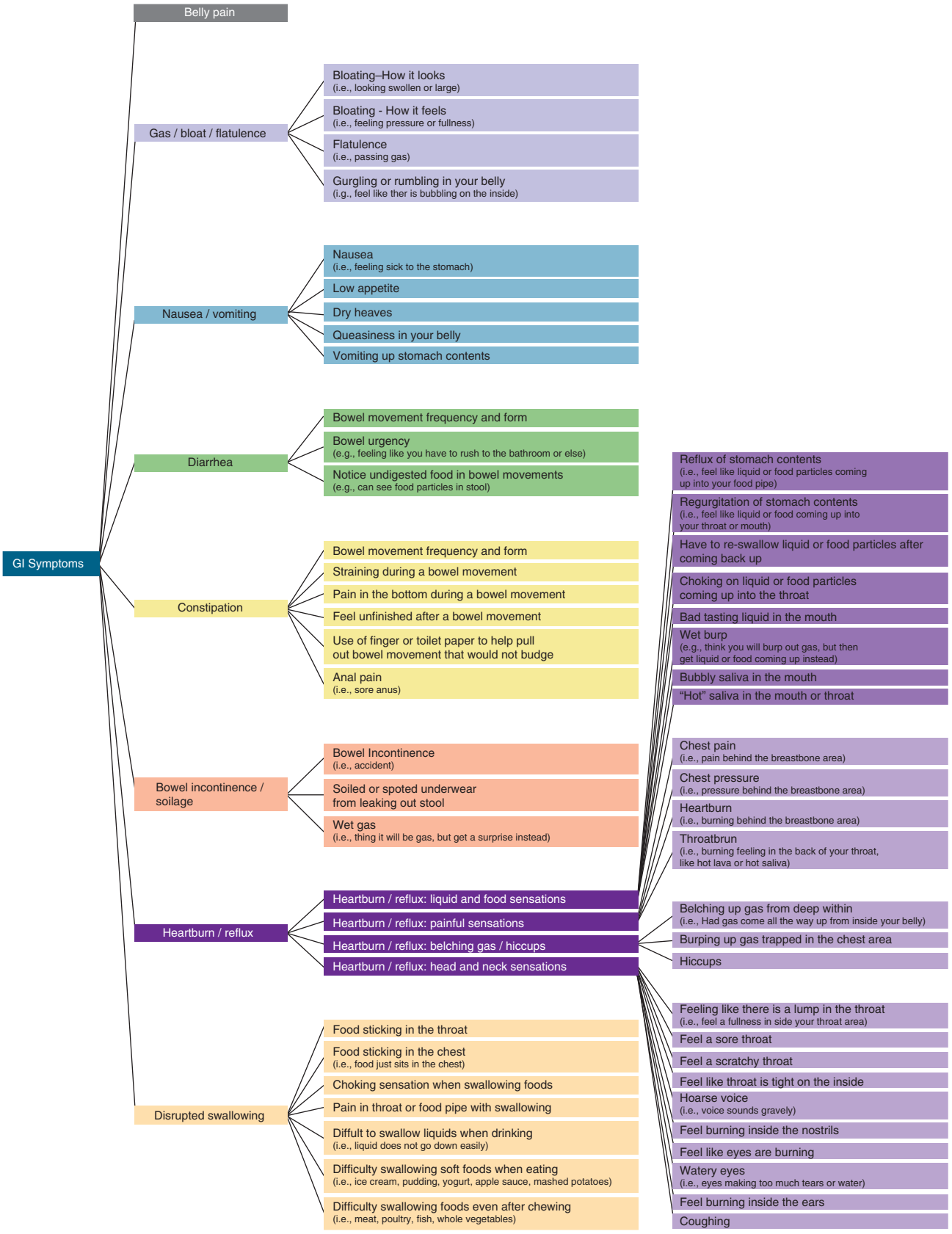

Figure 1. Patient-Reported Outcomes Measurement Information System (PROMIS) gastrointestinal (GI) Symptom Network.

Table 2. Descriptive characteristics of psychometric testing sample: GP vs. GI patients

Variable	GP (*n*=1177)	Patients (*n*=865)
Age	46 (s.d.=16)	48 (s.d.=16)
% Male*	43%	42%
% White*	72%	52%
% Black*	12%	17%
% Latino	12%	15%
% Asian*	3%	10%
% Other	2%	6%
% Less than HS	5%	2%
% HS grad*	33%	12%
% Some college	27%	29%
% College degree*	36%	58%
% Married	45%	44%
% Never married	33%	32%
% Widowed/divorced/ separated	22%	25%
% Employed	52%	49%
% Unemployed*	12%	8%
% Retired	15%	17%
% Disabled*	7%	14%
Self-reported GI disorders		
% IBS*	11%	40%
% GERD*	16%	33%
% IBD*	4%	28%
% Systemic sclerosis*	1%	18%
% Constipation*	19%	24%
% Other GI condition	47%	39%

GERD, gastroesophageal reflux disease; GI, gastrointestinal; GP, general population; HS grad, high school graduate; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome.

*$P<0.05$ comparing GP vs. patient groups.

Note that patients could endorse more than one GI condition. The most common "other" GI conditions were: intestinal surgery (*N*=72), symptomatic diverticular disease (*N*=63), dyspepsia (*N*=52), fecal incontinence (*N*=44), pancreatitis (*N*=25), celiac disease (*N*=15), peptic ulcer (*N*=15), and gastroparesis (*N*=11).

Table 3. PROMIS GI symptom scale characteristics

Scale	Number of items	Comparative fit index	Root mean square error of approximation
Belly pain	6	0.988	0.152
Gas/bloat	12	0.987	0.114
Diarrhea	5	0.966	0.154
Constipation	9	0.988	0.088
Bowel incontinence	4	0.999	0.080
Reflux	13	0.974	0.066
Nausea	4	0.992	0.068
Swallowing	7	0.966	0.154

GI, gastrointestinal; PROMIS, Patient-Reported Outcomes Measurement Information System.

that include "feeling sick to the stomach," decreased appetite, dry heaves, and finally vomiting up stomach contents. The nausea/vomiting domain items assess the frequency, severity, and/or predictability of these symptoms during the past 7 days.

Domain name: disrupted swallowing: Disrupted swallowing encompasses an array of symptoms described by patients ranging from pain to difficulty swallowing solids and/or liquids to food getting stuck in throat or chest when eating. The disrupted swallowing items assess the frequency of these swallowing-related symptoms during the past 7 days.

Refer to **Supplementary Appendix C** for the full set of PROMIS items. These will also be available online at the NIH Assessment Center (http://assessmentcenter.net/). In addition, we provide detailed scoring instructions and lookup tables in **Supplementary Appendix D**.

Psychometric evaluation

Patient characteristics and descriptive statistics. We recruited 865 patients to complete the online survey out of 2,217 invitations distributed among our partner clinics (39% response rate). Cint enrolled 1,177 GP subjects before closing the survey because of meeting enrollment criteria. **Table 2** presents the demographics characteristics of both samples. There was no significant difference in age or gender, but there were significant differences in race/ethnicity, education, marital status, and employment status. Of the GI patients, the most common diseases were IBS, gastroesophageal reflux disease, chronic constipation, IBD, and SSc. Notably, GI conditions were commonly reported in the US GP sample as well, demonstrating the high population prevalence of GI symptoms and related conditions.

IRT analyses. **Table 3** provides a summary of fit statistics for confirmatory factor analysis of calibrated PROMIS GI symptom scales. All the calibrated items had high fit indices supporting unidimensionality. The item properties from calibration are available in **Supplementary Appendix E**.

PROMIS GI symptom scale scoring. We calibrated each scale using the two-parameter IRT graded response model and scored on a T metric (the NIH PROMIS standard) with a mean of 50 and s.d. of 10 in the US GP. **Table 4** presents the mean scores among the GI patient population. With the exception of gastroesophageal reflux symptoms, the mean PROMIS scores were significantly higher in the patient population vs. GP. **Table 5** shows the correlations among the PROMIS GI Symptom Scales. **Supplementary Appendix D** demonstrates how to convert the scales into percentile scores, where each respondent is compared against the US GP on an easily interpreted percentile scale.

Scale reliability and information. Internal reliability was high for each of the scales, as follows: abdominal pain (0.87), gas/bloating (0.94), diarrhea (0.88), constipation (0.89), bowel incontinence (0.90), gastroesophageal reflux (0.88), nausea/vomiting (0.76), and disrupted swallowing (0.91).

Construct validity. **Table 6** provides evidence of construct validity for all eight PROMIS GI symptom scales compared with legacy instruments. Overall, the correlations between PROMIS GI symptom scales and the wide range of legacy instruments were statistically significant and in the anticipated direction.

Table 4. Average scores for general population and patients

Variable	General population (s.d.)	Patients (s.d.)
PROMIS gastroesophageal reflux[a]	50 (10)	51 (10)
PROMIS disrupted swallowing[a]	50 (10)	51 (10)
PROMIS diarrhea[a]	50 (10)	56 (11)
PROMIS incontinence[a]	50 (10)	53 (11)
PROMIS nausea/vomiting[a]	50 (10)	53 (10)
PROMIS constipation[a]	50 (10)	54 (10)
PROMIS belly pain[a]	50 (10)	57 (11)
PROMIS gas/bloat/flatulence[a]	50 (10)	57 (10)
PROMIS global physical[b]	50 (10)	45 (10)
PROMIS global mental[b]	50 (10)	47 (10)
EQ-5D[b]	0.77 (0.24)	0.69 (0.26)
VSI[c]	22 (21)	35 (21)

PROMIS, Patient-Reported Outcomes Measurement Information System; VSI, Visceral Sensitivity Index.
By design, all PROMIS scales are normed to a score of 50 and s.d. of 10 in the general population using a T-metric.
[a]Higher score denotes more gastrointestinal (GI) symptoms.
[b]Higher score denotes better health-related quality of life (HRQoL).
[c]Higher score denotes more GI-associated visceral sensitivity.

DISCUSSION

The eight NIH PROMIS GI symptom scales capture the breadth and depth of GI symptoms experienced by people with a wide range of digestive disorders. Unlike disease-targeted measures, which are designed for specific patient populations, the PROMIS GI symptom scales are system-targeted measures designed for anyone experiencing a GI symptom—whether patients or members of the population at large. This is an important distinction of PROMIS measures, because disease-targeted PROs are not useful across the population as a whole. PROMIS aims to support rigorously developed PROs that are applicable to all comers.

Similar to other PROMIS measures, the PROMIS GI scales are normed against GP distributions allowing for relative interpretation of symptom scores. As with clinical biomarkers, such as hemoglobin or creatinine levels, PROMIS scores are interpreted in relation to a background distribution of symptom experiences. For example, **Figure 2** shows sample results from a patient who completed the NIH PROMIS GI symptom scales using a computer administered patient–provider portal before a GI office visit (1). The "heat map" reports which of the eight symptoms the patient experienced over the past week, and records the symptom severity among the positively endorsed symptoms. Although the PROMIS scores are reported on a T metric, they can be easily converted to a percentile score against the US GP, as illustrated in **Figure 2**. We provide instructions in **Supplementary Appendix D** for how to calculate the PROMIS scores and convert them to percentile scores using lookup tables.

The PROMIS GI symptom scales are will become publicly available for download on the NIH PROMIS Assessment Center (http://assessmentcenter.net/). The Assessment Center provides score reports and T metric heat maps for users. Future functionality will yield age- and gender-normed scores. Even without Assessment Center, the instructions in **Supplementary Appendix D** allow for programming scores onto local systems as needed.

The PROMIS GI symptom scales can also be used for research. These scales offer the common-language benchmarks for symptoms across varied conditions. This provides a standardized outcome for epidemiological and clinical intervention trials. Future

Table 5. Correlations among PROMIS GI symptom scales

	Reflux	Swallow	Diarrhea	Incontinence	Nausea	Constipation	Pain	Gas/bloat
Reflux	1.00							
Swallow	0.74	1.00						
Diarrhea	0.44	0.39	1.00					
Incontinence	0.38	0.43	0.55	1.00				
Nausea/vomiting	0.62	0.58	0.46	0.40	1.00			
Constipation	0.46	0.43	0.42	0.29	0.46	1.00		
Pain	0.48	0.42	0.55	0.29	0.59	0.55	1.00	
Gas/bloat	0.50	0.43	0.47	0.30	0.52	0.55	0.70	1.00

GI, gastrointestinal; PROMIS, Patient-Reported Outcomes Measurement Information System.

Table 6. Correlations of GI scales with legacy measures

Scale	PROMIS Global Physical Health	VSI	EQ-5D	GSRS reflux	GSRS indigestion	GSRS belly pain	GSRS diarrhea	GRSR constipation	IBDQ	IBS-QOL	SSc-GIT
Reflux	−0.44	0.48	−38	0.68	0.55	0.57	0.36	0.45	−0.45	−0.25	0.48
Swallow	−0.43	0.43	−40	0.58	0.49	0.51	0.33	0.42	−0.36	−0.22	0.44
Diarrhea	−0.47	0.56	−36	0.37	0.53	0.31	0.80	0.39	−0.78	−0.50	0.67
Incontinence	−0.33	0.57	−32	0.32	0.38	0.34	0.53	0.28	−0.46	−0.37	0.57
Nausea/Vomiting	−0.44	0.53	−41	0.51	0.54	0.71	0.39	0.46	−0.56	−0.31	0.53
Constipation	−0.40	0.50	−37	0.40	0.53	0.50	0.36	0.77	−0.53	−0.32	0.28
Abdominal Pain	−0.51	0.66	−44	0.44	0.65	0.74	0.52	0.56	−0.70	−0.45	0.43
Gas/Bloat	−0.44	0.62	−39	0.45	0.76	0.64	0.45	0.60	−0.56	−0.53	0.59

GI, gastrointestinal; GSRS, Gastrointestinal Symptom Rating Scale; IBDQ, Inflammatory Bowel Disease Questionnaire; IBS-QOL, Irritable Bowel Syndrome-Quality of Life; PROMIS, Patient-Reported Outcomes Measurement Information System; SSc-GIT, Systemic Sclerosis-Gastrointestinal Tract; VSI, Visceral Sensitivity Index.
All correlation coefficients are significant at the $P<0.05$ level.

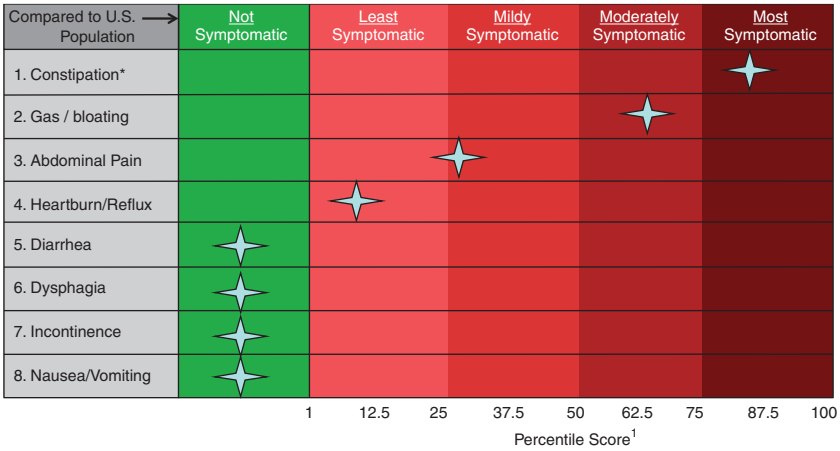

Figure 2. Sample "heat map" patient report of gastrointestinal (GI) Patient-Reported Outcomes Measurement Information System (PROMIS) scores. Patient scores are compared with the general US population benchmarks to add interpretability to the scores, similar to a lab test. For this use case, a provider can immediately detect that the patient reported many GI symptoms, but that constipation was the most severe and bothersome, falling within the top quartile of severity compared with the general population (GP). Gas and bloating were also elevated in this patient, falling in the third quartile of severity. In contrast, although the patient reported abdominal pain and heartburn/reflux symptoms, those scores were only in the first and second quartiles compared with people in the GP with similar symptoms. For instructions on how to convert PROMIS scores to percentile, see **Supplementary Appendix D**. *Patients "most bothersome" symptoms.

reports will present the longitudinal construct validity of the PROMIS GI symptom scales and minimally important difference estimates—additional attributes to assist with prospective intervention trials in gastroenterology.

The PROMIS GI symptom scales offer methodological and administrative advantages. Following the PROMIS methodology and constructed with oversight by the NIH PROMIS Steering Committee, the scales have been rigorously developed using modern psychometric techniques. This started with a grounded conceptual framework based on a systematic literature review and extensive patient focus groups. The participants ranged widely by demographics, GI disorders, and illness severity. The items were crafted to be understandable at a sixth grade level, and to be applicable to both patients and the GP at large. Support for the construct validity of the resulting scales was found using a diverse set of legacy instruments spanning from disease targeted (e.g., IBS-QOL, IBDQ, and SSc-GIT) to system targeted (e.g., Visceral Sensitivity Index and Gastrointestinal Symptom Rating Scale) to generic PROs (e.g., EuroQOL and PROMIS global health). Finally, unlike existing PROs in gastroenterology, the PROMIS GI symptom scales were also tested in the GP, thus offering a scale that is applicable to anyone with GI symptoms, regardless of whether they are seeking care for their symptoms.

As with any PRO development effort, the PROMIS GI symptom scales have limitations. Although we identified a wide range of patients representing the breadth and depth of typical GI symptoms, we did not include subjects from many GI conditions, such as GI malignancies or chronic liver diseases. Other common conditions, such as celiac sprue, had only small numbers of participants in this initial validation trial. The scales also do not measure signs like rec-

tal bleeding or weight loss. Future research is needed to evaluate the PROMIS GI symptom scales in other conditions and populations. In addition, the scales are currently designed for adult populations; we hope that future work will focus on using the PROMIS methodology in pediatric GI populations. The scales are further limited by their 7-day recall period; they are not currently suitable for momentary assessments, or for use as a daily diary. Future research should test retrofitted scales that can apply to shorter recall periods; this may be especially important for use of PROMIS in pharmaceutical trials. Finally, we did not validate the item bank against objective tests such as upper GI endoscopy, motility studies, or other diagnostics. Previous studies have shown that PROs complement the objective tests in clinical care and future research should assess the role of GI PROMIS in achieving this goal (28,29).

In conclusion, we developed the NIH PROMIS GI symptom scales—a publicly available set of valid and reliable PROs for use in people with GI symptoms. The eight scales can be used together or individually for clinical practice and clinical research in a disease-agnostic manner. The scales are broadly applicable across populations, GI symptoms, GI diseases, and demographics. Future work will report the longitudinal validity of the scales, including how they track with patient reports and physician illness assessments, and will evaluate how use of the scale affects clinical outcomes in diverse GI populations.

ACKNOWLEDGMENTS

We thank Sally Bolus for her skillful project management of the GI PROMIS scale development. We also thank William D. Chey, Douglas Drossman, and Jan Irvine for serving as members of the expert panel that reviewed the results of our literature search. We also thank Nancy and Bill Norton and the International Foundation for Functional Gastrointestinal Disorders (IFFGD) for partnering with us in recruiting patients for the study, Phil Tonkins, Jim Whitter, and Susana Sztein from the NIH, and David Cella for his leadership of the PROMIS Steering Committee. Finally, we dedicate this work to Vanessa Ameen, our Scientific Officer at NIH for the first part of this project and a strong advocate of PROMIS; Vanessa's passing did not allow her to see the completion of this work, but we know it is stronger because of her involvement.

CONFLICT OF INTEREST

Guarantor of the article: Brennan M.R. Spiegel, MD, MSHS, RFF, FACG, AGAF.

Specific author contributions: Obtained funding, study design, data collection, data analysis, and drafting the article: Brennan M.R. Spiegel; study design, data analysis, and editing the article: Ron D. Hays and Roger Bolus; data collection, and editing the article: Gil Y. Melmed and Lin Chang; data collection: Cynthia Whitman, Puja P. Khanna, Sylvia H. Paz and Tonya Hays; data analysis: Steve Reise; obtained funding, study design, data collection, data analysis and editing the article: Dinesh Khanna.

Financial support: This study was supported by NIH/NIAMS U01 AR057936A, the National Institutes of Health through the NIH Roadmap for Medical Research grant (AR052177). Puja P. Khanna was supported by Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training grant NIAMS 1 T32 AR053463 and ACR Research and Education Foundation Clinical Investigator Fellowship Award 2009-11. Dinesh Khanna was also supported by NIAMS K24 AR063120. Ron D. Hays was also supported by NIH/NIA grants P30-AG028748 and P30-AG021684, and NCMHD grant 2P20MD000182. Lin Chang was also supported by NIDDK P50 DK64539.

Potential competing interests: Brennan M.R. Spiegel has received grant support from Ironwood, Amgen, Shire Pharmaceuticals, and Theravance Pharmaceuticals, and served as a consultant to Ironwood, Forest, and Takeda North America. Dinesh Khanna has served as consultant and/or received grant support from Actelion, Astra-Zeneca, Bayer, BMS, DIGNA, Genentech, Gilead, InterMune, Merck, Roche, Takeda, Savient, and United Therapeutics. Ron D. Hays has served as a consultant to Amgen, Allergan, Pfizer, and the Critical Path Institute. Gil Y. Melmed has served as a consultant for Abbvie, Given Imaging, and Jannsen, is on the speaker's bureau for Prometheus and Abbott, and has received research support from Pfizer. Lin Chang has served as a consultant to Ironwood, Forest, Salix, Takeda North America, Purdue Pharma, and Entera Health, and has received grant support from Ironwood.

Disclaimer

The opinions and assertions contained herein are the sole views of the authors and are not to be construed as official or as reflecting the views of the Department of Veteran Affairs.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Patient-reported outcomes (PROs) capture the patients' illness experience in a structured format and may help providers and researchers understand symptoms from the patients' perspective.
- ✓ The National Institutes of Health (NIH) Patient-Reported Outcomes Measurement Information System (PROMIS) is toolbox of publicly available PROs (www.nihpromis.gov) that are highly efficient, computer-based, and short questionnaires that cover the breadth and depth of health and illness.
- ✓ Here we present the NIH PROMIS gastrointestinal (GI) symptom measures.

WHAT IS NEW HERE

- ✓ Psychometric analyses in 865 patients with diverse GI conditions and 1,177 participants from the general population found 8 major symptom complexes: gastroesophageal reflux, disrupted swallowing, diarrhea, bowel incontinence/soiling, nausea and vomiting, constipation, belly pain, and gas/bloating.
- ✓ Under the guidance of the NIH PROMIS consortium, we developed a scale for each GI symptom complex that correlates significantly with both generic and disease-targeted legacy instruments, and demonstrates evidence of reliability.
- ✓ The GI PROMIS scales can be used together or individually for clinical practice and clinical research in a disease-agnostic manner; they are broadly applicable across populations, GI symptoms, GI diseases, and demographics.

REFERENCES

1. Spiegel BM. Patient-reported outcomes in gastroenterology: clinical and research applications. J Neurogastroenterol Motil 2013;19:137–48.
2. Spiegel B, Khanna D, Bolus R *et al.* Understanding gastrointestinal distress: a framework for clinical practice. Am J Gastroenterol 2011;106:380–5.
3. Wagner EH, Austin BT, Davis C *et al.* Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20:64–78.
4. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74:511–44.
5. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. J Eval Clin Pract 2006;12:559–68.
6. Dobscha SK, Gerrity MS, Ward MF. Effectiveness of an intervention to improve primary care provider recognition of depression. Eff Clin Pract 2001;4:163–71.
7. Taenzer P, Bultz BD, Carlson LE *et al.* Impact of computerized quality of life screening on physician behaviour and patient satisfaction in lung cancer outpatients. Psychooncology 2000;9:203–13.
8. Velikova G, Booth L, Smith AB *et al.* Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004;22:714–24.
9. Brown RF, Butow PN, Dunn SM *et al.* Promoting patient participation and shortening cancer consultations: a randomised trial. Br J Cancer 2001;85:1273–9.
10. Cella D, Yount S, Rothrock N *et al.* The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45:S3–S11.
11. Burke LB, Kennedy DL, Miskala PH *et al.* The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. Clin Pharmacol Ther 2008;84:281–3.
12. Khanna P, Agarwal N, Khanna D *et al.* Development of an online library of patient reported outcome measures in gastroenterology: the GI-PRO database. Am J Gastroenterol 2014;109:234–48.
13. Reeve BB, Hays RD, Bjorner JB *et al.* Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;45:S22–31.
14. DeWalt DA, Rothrock N, Yount S *et al.* Evaluation of item candidates: the PROMIS qualitative item review. Med Care 2007;45:S12–21.
15. Spiegel BM, Bolus R, Agarwal N *et al.* Measuring symptoms in the irritable bowel syndrome: development of a framework for clinical trials. Aliment Pharmacol Ther 2010;32:1275–91.
16. McLaughlin G. SMOG grading: a new readability formula. J Reading 1969;12:639–46.
17. Labus JS, Bolus R, Chang L *et al.* The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale. Aliment Pharmacol Ther 2004;20:89–97.
18. Labus JS, Mayer EA, Chang L *et al.* The central role of gastrointestinal-specific anxiety in irritable bowel syndrome: further validation of the visceral sensitivity index. Psychosom Med 2007;69:89–98.
19. Hays RD, Bjorner JB, Revicki DA *et al.* Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009;18:873–80.
20. Revicki DA, Wood M, Wiklund I *et al.* Reliability and validity of the Gastrointestinal Symptom Rating Scale in patients with gastroesophageal reflux disease. Qual Life Res 1998;7:75–83.
21. The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.
22. Patrick DL, Drossman DA, Frederick IO *et al.* Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Dig Dis Sci 1998;43:400–11.
23. Drossman DA, Patrick DL, Whitehead WE *et al.* Further validation of the IBS-QOL: a disease-specific quality-of-life questionnaire. Am J Gastroenterol 2000;95:999–1007.
24. 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;91:1571–8.
25. Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. J Pediatr Gastroenterol Nutr 1999;28:S23–7.
26. Khanna D, Hays RD, Park GS *et al.* Development of a preliminary scleroderma gastrointestinal tract 1.0 quality of life instrument. Arthritis Rheum 2007;57:1280–6.
27. Spiegel BM, Bolus R, Harris LA *et al.* Characterizing abdominal pain in IBS: guidance for study inclusion criteria, outcome measurement and clinical practice. Aliment Pharmacol Ther 2010;32:1192–202.
28. Bae S, Allanore Y, Furst D *et al.* Associations between a scleroderma-specific gastrointestinal instrument and objective tests of upper gastrointestinal involvements in systemic sclerosis. Clin Exp Rheumatol 2013;31:57–63.
29. Khanna D, Nagaraja V, Gladue H *et al.* Measuring response in the gastrointestinal tract in systemic sclerosis. Curr Opin Rheumatol 2013;25:700–6.