

Systemic sclerosis (SSc) is a connective tissue disease characterized by autoimmunity, progressive vasculopathy, and fibrosis of the skin and internal organs¹. The most commonly affected internal organ in SSc is the gastrointestinal (GI) tract, which is involved in about 90% of patients. Any part of the GI tract from the esophagus to the anorectum may be involved, resulting in a variety of symptoms including gastroesophageal reflux disease (GERD), gastroparesis, small intestinal bacterial overgrowth, pseudo-obstruction, malabsorption, and fecal incontinence². The most common of these symptoms is GERD, which is caused by a weakened lower esophageal sphincter with or without loss of esophageal motility³. GERD affects 70–90% of patients^{2,3}, and symptoms include dysphagia, the sensation of acid reflux, or nausea. Chronic uncontrolled GERD may ultimately lead to strictures, erosive esophagitis, ulcerations, Barrett's esophagus, and/or adenocarcinoma^{4,5,6}. Though a subset of patients is asymptomatic, most patients experience symptoms, which is informative in guiding therapeutic intervention⁷.

The initial management of suspected GERD includes an empiric trial of acid-suppressive therapy⁸. If symptoms persist and endoscopy does not reveal evidence of GERD, esophageal function tests can be performed, including esophageal manometry and ambulatory reflux monitoring⁹. Objective measures are used to make a diagnosis of GERD, to determine whether existing therapy is effective in controlling acid and non-acid reflux, and to assess whether promotility agents may have a role in managing symptoms. Such objective testing includes pH monitoring with or without impedance (to measure reflux), and esophageal manometry with or without high resolution (to measure esophageal motility). While these tests are useful for clinical decision-making, they are invasive, costly, and at times not tolerated by patients with SSc¹⁰. As a result, patient-reported outcomes (PRO) have the potential to be more practical and cost-effective outcome measures for well-designed, randomized, placebo-controlled clinical studies and for guiding patient care for GERD management.

The University of California, Los Angeles (UCLA) Scleroderma Clinical Trials Consortium (SCTC) GI Tract Questionnaire (GIT) 2.0 and the US National Institutes of Health (NIH)-funded Patient-Reported Outcomes Measurement Information System (PROMIS) GI symptoms instrument have both been validated in SSc to assess patient-reported GI symptoms. The Quality of Life in Reflux and Dyspepsia (QOLRAD) assessed the effect of GERD symptoms on quality of life in different international cohorts, but has not yet assessed it in SSc^{11,12}. Our objectives were to determine reliability, construct validity, and sensitivity to change of the UCLA GIT 2.0 Reflux scale, PROMIS Reflux scale, and QOLRAD in participants with SSc and active GERD who were given a single new GERD-specific therapy to treat this symptom.

MATERIALS AND METHODS

Participants. Seven international SSc centers from the United States, Italy, Belgium, and Australia participated in this longitudinal observational cohort study: University of Michigan, Johns Hopkins University, University of Utah, University of Florence, Thomas Jefferson University, Ghent University, and Royal Adelaide Hospital. Participants were recruited during routine clinical visits. All participants met SSc American College of Rheumatology/European League Against Rheumatism 2013 criteria and had symptoms of active GERD, defined as having symptoms for at least 3 of the prior 7 days. In addition to the above criteria, all patients were required to meet at least one of the following additional criteria or have one of the following studies: (1) barium swallow showing spontaneous reflux; (2) pH impedance study; (3) abnormal 24-h pH monitoring; (4) upper GI endoscopy showing esophagitis or complications such as ulcers, stenosis, or Barrett's esophagus; or (5) diagnosis of GERD based on clinical symptoms. To maximize the likelihood of detecting a change in GI symptoms, we focused on recruiting participants who were initiated on a new treatment intervention or had a single change in their GI management (increase in pharmacologic or nonpharmacologic therapies). Such interventions included (1) initiation of lifestyle modifications (avoidance of aggravating foods/acidic foods, not eating < 4 h before bed, raising the head of the bed, avoidance of alcohol and smoking), (2) over-the-counter antacids, (3) initiation or increase of proton pump inhibitors; (4) initiation or increase of H₂-blocker, (5) other medications (sulfacate/carafate), (6) use of a combination of the above medications and prokinetic, or (7) endoscopic antireflux procedures. If the treating physician determined that the patient did not meet criteria for SSc or that the patient's GI disease was a complication of another condition, then the patient was excluded from the study. Patients could be recruited for only a single GI issue, though other complaints may have existed, and could have only a single intervention made at that clinic visit. PRO were recorded at baseline and again at 4 weeks of followup. This study complies with the Declaration of Helsinki, and the research protocol was approved by the Institutional Review Board (IRB00036572) at each center. Informed consent was obtained from all subjects.

Instruments. We assessed the overall severity of the underlying GI illness with an anchor questionnaire using a single global item ["In the past 7 days, how would you rate your gastrointestinal condition? (excellent, very good, good, fair, or poor)"]. UCLA SCTC GIT 2.0 instrument is a 34-item questionnaire that assesses 7 scales: reflux (8 items), distension/bloating (4 items), soilage (1 item), diarrhea (2 items), social functioning (6 items), emotional well-being (9 items), and constipation (4 items)^{13,14}. Each scale has a weighted subscore, and a total score is based on the sum of 6 domains (reflux, distension/bloating, fecal soilage, diarrhea, social functioning, emotional well-being). A 3-point categorical response scale (0–3) is used to assess all items except for items 15 and 31 in the diarrhea and constipation domains, respectively, where a score of 0 or 1 is provided. A higher score represents more severe disease. For the purpose of this analysis, only the Reflux scale was used, which collects information on reflux and regurgitation; however, the total score does encompass 6 domains.

The PROMIS Reflux scale GI Symptoms Scales instrument is a 60-item questionnaire that assesses 8 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/bloating/flatulence (12 items)¹⁵. This study used the GERD scale and did not use the computer adaptive training feature, so that all patients answered 20 items related to reflux and dysphagia. A 5-point categorical response scale was used to assess all items, and a higher score denotes more GI symptoms. This instrument does not record social functioning and emotional well-being.

The QOLRAD questionnaire is specific for reflux and dyspepsia and includes 25 items divided into 5 domains including emotional distress, sleep disturbance, food/drink problems, physical/social functioning, and vitality. The questions are rated on a 7-point Likert scale, with a lower value representing a more severe effect on daily functioning.

The Gastroesophageal Reflux Disease Questionnaire (GerdQ) is a self-administered 6-item questionnaire that is developed and validated in the diagnosis of GERD. The 6 domains include heartburn, regurgitation, upper abdominal pain, nausea, sleep interference, and medication, and they are scored on a 4-point Likert scale, with higher numbers representing more severe disease. The cutoff of ≥ 8 points has the highest specificity and sensitivity for a diagnosis of GERD¹⁶.

Anchors. The GI Anchor followup visit questionnaire was used to assess both overall improvement and improvement specific to GERD symptoms, and was used in a previous study¹⁷. All participants were asked 3 questions, which included the following: (1) In the past 7 days how would you rate your gastrointestinal condition? (2) Compared to your last visit, how is your overall gastrointestinal condition at this time? and (3) Compared to your last visit, how is your overall gastrointestinal condition for which you received treatment? For question 1, a 5-point response scale was provided, and response options included Excellent, Very Good, Good, Fair, and Poor. For questions 2 and 3, a 7-point response scale was provided and included the following potential responses: Completely better, Considerably better, Somewhat better, About the same, Somewhat worse, Considerably worse, and Completely worse. Participants who checked "Completely better," "Considerably better," or "Somewhat better" on Questions 2 and 3 were characterized as "Improved," and others were characterized as "Not Improved."

Statistical analysis. Descriptive statistics are presented as percentages for categorical variables. Student t tests were used to compare the means in normally distributed data. The Wilcoxon Mann-Whitney U test was used to compare the medians of non-normally distributed data. The percentages of respondents scoring the minimum (floor) and maximum (ceiling) possible scores were calculated to evaluate scale score distributions for PROMIS and legacy instruments. For easy interpretability, floor effect is presented as "worst" possible score and ceiling as "best" possible irrespective of the direction of the scale.

The internal consistency reliability was estimated using Cronbach's alpha coefficient, and a value ≥ 0.70 was considered satisfactory¹⁸. Construct validity was assessed by examining correlation coefficients at baseline between 3 PRO scales¹⁹. Cohen d effect sizes (ES) across the whole group, improved group, and not improved group were calculated to assess the sensitivity to change among important clinical subsets and interpreted as 0.20–0.49 as small magnitude, 0.50–0.79 as moderate magnitude, and ≥ 0.80 as large magnitude^{20,21}. All analyses were performed at the University of Michigan (VB) using R software version 3.4.2, and p values < 0.05 were considered statistically significant.

RESULTS

A total of 116 participants with SSc and active GERD completed baseline and 113 completed 4-week followup visits. Patient average (\pm SD) age was 53.8 (\pm 13.3) years and there was a mean disease duration of 12.0 (\pm 10.3) years (Table 1). Participants were more likely to be female (81%), and there was a similar distribution of participants with diffuse and limited cutaneous disease (42% and 49%, respectively). Mean body mass index (BMI) was 25.6 (\pm 5.2) kg/m². Antinuclear antibodies were present in 95.0% of participants. Anticentromere antibodies were reported in 40.4% of participants, anti-topoisomerase 1 (anti-Scl-70) antibodies in 21.8%, anti-RNA Pol-3 antibodies in 20.2%, anti-Ro antibodies in 12.2%, and anti-RNP in 10%. GI Anchors baseline visit scores represented patient self-reported assessment of their GI condition over the preceding 7 days. In response to this survey, 30.1% of participants rated their GI condition as poor, 40.7% rated their GI

Table 1. Baseline demographics and clinical characteristics.

Variables	N = 116
Age, yrs, mean \pm SD (range)	53.8 \pm 13.3 (23–82)
Female	81.0
Race	
White	91.4
Black	3.5
Other	5.1
Ethnicity	
Non-Hispanic	95.7
Hispanic	1.7
Disease duration (since first RP or non-RP), yrs, mean \pm SD; (range)	12.0 \pm 10.3 (109); (0–58)
Limited cutaneous disease	49.1
Diffuse cutaneous disease	42.2
Overlap	1.7
BMI, kg/m ² , mean \pm SD (range)	25.6 \pm 5.2 (16–40)
GI anchor self-reported baseline status (n = 113)	
Excellent	0.0
Very good	8.9
Good	20.4
Fair	40.7
Poor	30.1
Autoantibody-positive	
ANA, n = 113	95.0
Anti-Scl-70, n = 101	21.8
Anticentromere, n = 104	40.4
Anti-RNAP-3, n = 94	20.2
Anti-U1RNP, n = 100	10.0
Anti-Ro, n = 98	12.2
Negative for all above SSc-Ab, n = 113	2.7

Values are % unless otherwise specified. RP: Raynaud phenomenon; BMI: body mass index; GI: gastrointestinal; ANA: antinuclear antibody; RNAP-3: RNA polymerase antibodies-3; SSc-Ab: systemic sclerosis autoantibodies.

status as fair, 20.4% rated their GI status as good, 8.9% rated their GI status as very good, and none rated their GI status as excellent. Of the 63 patients who completed the GerdQ, 78% (49/63) scored above 8 points, which is highly sensitive and specific for GERD¹⁶.

At baseline, the mean (\pm SD) baseline UCLA GIT 2.0 Reflux score was 0.97 (0.63; n = 113) and the mean baseline PROMIS Reflux score was 53.3 (8.0, n = 65). GIT 2.0 Reflux score was suggestive of moderate GERD and the PROMIS Reflux score was 0.3 SD below mean US general population²². The mean QOLRAD scores (n = 110) for each scale were as follows: (1) emotional distress score 5.3 \pm 1.5 (n = 110); (2) sleep disturbance score 4.9 \pm 1.6; (3) food and drink problems score 4.8 \pm 1.5; (4) physical and social function score 5.8 \pm 1.2; and (5) vitality score 5.0 \pm 1.6. Cronbach's alpha coefficient, to assess internal consistent reliability, was adequate for all scales, ranging from 0.78 to 0.96. The percentages of participants having a minimum and maximum possible score at baseline on the UCLA GIT 2.0 Reflux scale, the PROMIS Reflux scale, and the 5 QOLRAD scales were then assessed, stratified by limited and diffuse cutaneous disease (Table 2). The percentage of participants

Table 2. Minimum and maximum scores and Cronbach's alpha coefficient for the UCLA GIT 2.0 (GERD), PROMIS (GERD), QOLRAD, and Reflux Disease Questionnaire at baseline.

Scale	Hypothetical Range for Each Scale	Min Score for This Cohort*	Max Score for This Cohort*	% with Min Score	% with Max Score	Cronbach's α^{\diamond}
UCLA GIT 2.0* Reflux, n = 113	0.0–3.0	0.0	2.6	0.9	0.0	0.78
PROMIS Reflux*, n = 65	30.3–87.6	36.4 [†]	68.8 [†]	0.0	0.0	0.82
QOLRAD**, n = 110						
Emotional distress	1.0–7.0	1.7	7.0	0.0	0.9	0.95
Sleep disturbance	1.0–7.0	1.0	7.0	0.9	0.9	0.96
FD problems	1.0–7.0	1.5	7.0	0.0	0.9	0.93
PS function	1.0–7.0	2.0	7.0	0.0	0.9	0.91
Vitality	1.0–7.0	1.3	7.0	0.0	0.9	0.93

* Higher score denotes more GI symptoms; ** Lower score denotes more GI symptoms; [†] Based on T score, so the min/max score for the cohort is different from the min/max score for the instrument. [◇] Cronbach's alpha reported to 2 decimal places as is standard reporting methodology. UCLA: University of California, Los Angeles; GIT: GI Tract; GERD: gastroesophageal reflux disease; PROMIS: Patient-Reported Outcome Measurement Information System; QOLRAD: Quality of Life in Reflux and Dyspepsia; FD: food and drink domain; PS: physical/social functioning.

with the minimum score on the UCLA GIT 2.0 Reflux scale was 0.88%. The percentage of participants with the minimum score on the PROMIS Reflux scale was 0%. On the QOLRAD, the percentage of participants with the minimum score was 0% percent on all scales, except for sleep disturbance, which was 0.91%. All instruments showed none to low floor and ceiling effect.

We assessed the correlations between the UCLA GIT 2.0 Reflux scale, the PROMIS Reflux scale, and the 5 QOLRAD scales (emotional, sleep, food and drink, physical/social, vitality; Table 3). Importantly, higher scores for the UCLA GIT 2.0 and for the PROMIS represent greater reflux disease, whereas higher scores for the QOLRAD represent better quality of life owing to reflux. The UCLA GIT 2.0 and PROMIS Reflux scales had a significant correlation coefficient at baseline (0.61, $p < 0.0001$). Of the 5 QOLRAD scales, the PROMIS Reflux scale correlated most strongly with the Food and Drink domain of QOLRAD (-0.66 ; $p < 0.01$), and UCLA GIT 2.0 Reflux scale correlated strongly with both the Food and Drink domain and the Sleep disturbance domains of the QOLRAD (-0.70 and -0.71 , respectively; $p < 0.01$).

Sensitivity to change was then assessed across the whole cohort ($n = 116$), and the ability to detect change over time was significant for the PROMIS Reflux scale, ($n = 65$, $p = 0.026$), and two of 5 QOLRAD scales ($n = 110$, emotional distress, $p = 0.034$; and vitality, $p = 0.047$; Supplementary Table 1 and Supplementary Figure 1, available with the online version of this article). The p values were of borderline statistical significance in Sleep Disturbance and Food and Drink domains of the QOLRAD ($n = 109$, $p = 0.055$ and $p = 0.08$, respectively), but not significant in the UCLA GIT 2.0 GERD scale ($n = 105$, $p = 0.11$), or Physical/Social Function scale of the QOLRAD ($p = 0.504$). The effect size for all the scales was < 0.20 , except for the PROMIS Reflux scale, which was 0.27.

As a smaller number of subjects completed PROMIS Reflux scale versus other instruments owing to nonavailability of translated PROMIS scale, we assessed sensitivity to change in participants who had completed both PROMIS Reflux scale and GIT 2.0. The effect size was similar for PROMIS and GIT 2.0 Reflux scales (0.28 and 0.24, respectively) for the whole group (Table 4). Among participants

Table 3. Correlation coefficients for patient-reported outcome surveys.

	UCLA GIT 2.0 Reflux	NIH PROMIS Reflux	QOLRAD Emotional Distress	QOLRAD Sleep Disturbance	QOLRAD FD Problems	QOLRAD PS Function	QOLRAD Vitality
UCLA GIT 2.0* Reflux	1.00	0.61	-0.63	0.71	-0.70	-0.61	-0.65
PROMIS* Reflux		1.00	-0.57	0.59	-0.66	-0.54	-0.61
QOLRAD**							
Emotional distress			1.00	0.83	0.86	0.84	0.85
Sleep disturbance				1.00	0.85	0.75	0.84
FD problems					1.00	0.78	0.85
PS function						1.00	0.86
Vitality							1.00

* Higher score denotes more GI symptoms; ** Lower score denotes more GI symptoms. All values are statistically significant ($p < 0.05$). GI: gastrointestinal; UCLA: University of California, Los Angeles; GIT: GI Tract; NIH: US National Institutes of Health; PROMIS: Patient-Reported Outcome Measurement Information System; QOLRAD: Quality of Life in Reflux and Dyspepsia; FD: food and drink domain; PS: physical/social functioning.

Table 4. Effect size across patients with both UCLA GIT 2.0, PROMIS Reflux, and QOLRAD scales.

Patient-reported Outcome Scales	N	Whole Group	p	N, (Q2 [†]), I/NI	Improved Group (Q2 [†])	p	Not Improved Group (Q2 [†])	p	N, (Q3), I/NI	Improved Group (Q3 [†])	p	Not Improved Group (Q3 [†])	p
UCLA GIT 2.0*													
Reflux	58	0.24	0.10	24/33	0.97	< 0.01	0.09	0.11	26/30	0.82	< 0.01	0.17	0.25
NIH PROMIS*													
Reflux	58	0.28	0.02	24/33	0.90	< 0.01	0.20	0.70	26/30	0.79	< 0.01	0.05	0.74
QOLRAD**													
Emotional distress	57	0.30	0.02	24/32	0.72	< 0.01	0.04	0.79	26/29	0.62	< 0.01	0.64	0.67
Sleep disturbance	57	0.30	0.01	24/32	0.93	< 0.01	0.10	0.24	26/29	0.86	< 0.01	0.13	0.14
FD problems	57	0.29	0.02	24/32	0.71	< 0.01	0.02	0.86	26/29	0.59	< 0.01	0.07	0.62
PS function	57	0.19	0.11	24/32	0.59	< 0.01	0.05	0.68	26/29	0.57	< 0.01	0.07	0.58
Vitality	57	0.23	0.07	24/32	0.67	< 0.01	0.06	0.66	26/29	0.58	< 0.01	0.04	0.79

* Higher score denotes more GI symptoms; ** Lower score denotes more GI symptoms. † Q2: Compared to your last visit, how is your overall GI condition at this time?; Q3: Compared to your last visit, how is your overall GI condition for which you received treatment? GI: gastrointestinal; UCLA: University of California, Los Angeles; GIT: GI Tract; NIH: US National Institutes of Health; PROMIS: Patient-Reported Outcome Measurement Information System; QOLRAD: Quality of Life in Reflux and Dyspepsia; FD: food and drink domain; PS: physical/social functioning.

who were characterized as “Improved” based on their self-reported assessment, the effect sizes were large for GIT 2.0 Reflux scale, moderate to large for the PROMIS Reflux scale, and moderate to large for QOLRAD, depending on the scale (range 0.573–0.928; Table 4, Figure 1, and Figure 2). For the participants who were characterized as “Not Improved,” the ES estimates were not significant.

DISCUSSION

GI dysmotility is the most frequent internal complication of SSc and has a major effect on morbidity. The esophagus is the most frequently involved region in the SSc GI tract, which can be affected by a weak lower esophageal sphincter, and/or significant dysmotility or aperistalsis. Each of these complications can lower the quality of life²³. The mechanisms underlying SSc esophageal dysmotility are poorly understood, and biomarkers of GI disease activity are not

defined, making it challenging to assess the effects of existing therapies. Though not all SSc patients with esophageal symptoms have GERD²⁴, and though some SSc patients may experience asymptomatic GERD²⁵, patient-reported outcomes such as the GerdQ are sensitive and specific in GERD diagnosis²⁶. Validated patient-reported outcome measures allow for a standardized assessment of important clinical response measures in SSc and may play a role for informing both clinical practice and trial design.

In an international prospective longitudinal cohort, we evaluated the ability of patient-reported GI outcome measures to detect clinically important changes in SSc after a single lifestyle modification or a GERD-specific change in medical therapy. All instruments were found to be reliable and have construct validity and sensitivity to change when assessing clinical response to therapeutic intervention in symptomatic SSc patients with GERD. In addition, there

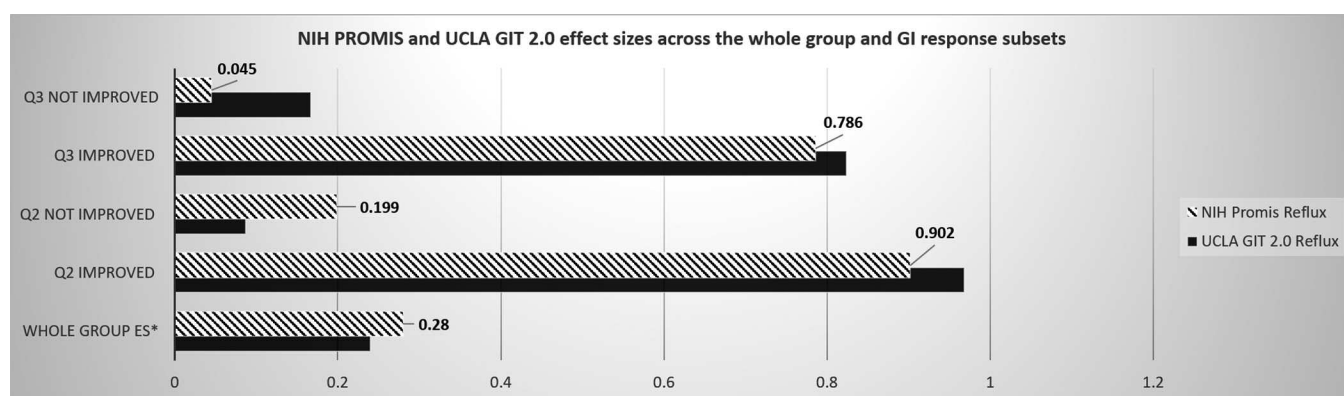


Figure 1. PROMIS and UCLA GIT 2.0 reflux scales effect sizes were calculated (1) for whole cohort, (2) by improved versus not improved on question 2 of the GI Anchor (i.e., Compared to your last visit, how is your overall gastrointestinal condition at this time?), and (3) by improved versus not improved on question 3 of the GI Anchor (i.e., Compared to your last visit, how is your overall gastrointestinal condition for which you received treatment?). Responsiveness indices: 0.20–0.49 = small magnitude, 0.50–0.79 = medium magnitude, and ≥ 0.80 = large magnitude. ES*: effect size; NIH: US National Institutes of Health; PROMIS: Patient-Reported Outcomes Measurement Information System; GI: gastrointestinal; UCLA GIT: University of California, Los Angeles GI Tract Questionnaire.

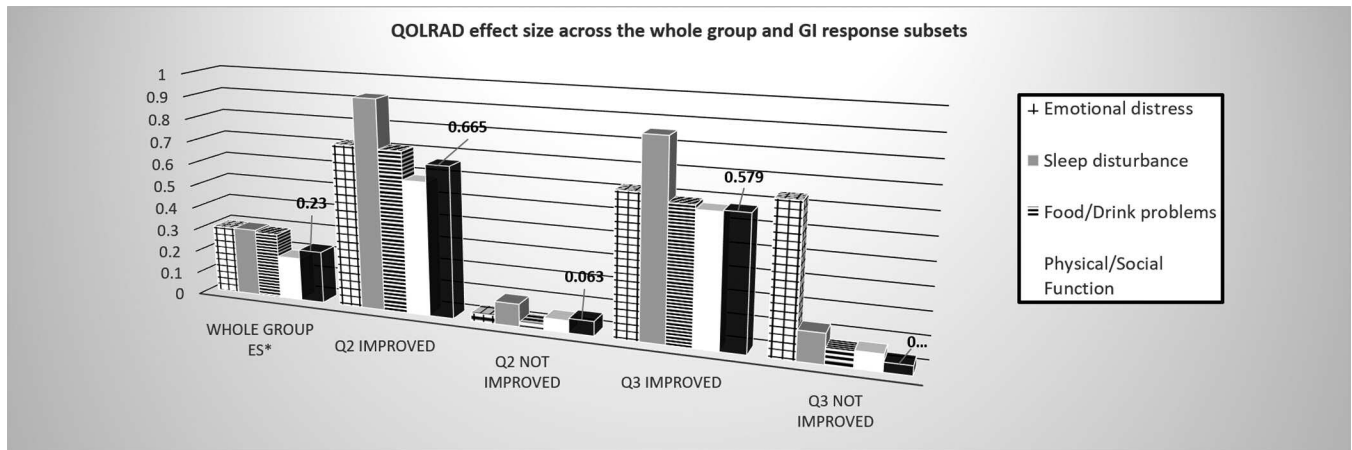


Figure 2. Quality of Life Reflux and Dyspepsia (QOLRAD) effect size across the cohort and among GI response subsets. ES was calculated (1) for whole cohort, (2) by improved versus not improved on question 2 of the GI Anchor (i.e., Compared to your last visit, how is your overall gastrointestinal condition at this time?), and (3) by improved versus not improved on question 3 of the GI Anchor (i.e., Compared to your last visit, how is your overall GI condition for which you received treatment?). Responsiveness indices: 0.20–0.49 = small magnitude, 0.50–0.79 = medium magnitude, and ≥ 0.80 = large magnitude. ES*: effect size; GI: gastrointestinal.

were minimal floor and ceiling effects of the instruments. The QOLRAD survey used in our study is an important complement to the other reflux-focused patient-reported outcome measures that we used, because it accounts for quality of life in several distinct domains, including emotional distress, sleep disturbance, food and drink, physical and social measures, and vitality. These measures are not assessed by the reflux domains of the UCLA GIT 2.0 and the NIH GI PROMIS, though they represent important external influences on GI symptoms and function.

The correlation coefficients for the UCLA GIT 2.0 Reflux scale demonstrated moderate correlation with the PROMIS Reflux scale (0.60) and the QOLRAD domains. Both the UCLA GIT 2.0 and the PROMIS changed in the same direction, with higher points representing more severe GI disease, while the QOLRAD changed inversely to these scales, with lower points representing more severe GI disease. In another study in SSc, the correlation coefficient between the PROMIS Reflux scale and UCLA GIT 2.0 Reflux scale was 0.77, supporting the current estimates.

For the overall group, the change in instruments' scores over time was minimal, consistent with data in the larger GI disorders¹⁷. During the development of the PROMIS GI Symptoms Scale¹⁵, only 23% "improved" in reflux symptoms on self-assessment when a similar design was used. In our current study, among participants who "improved," all instruments had moderate-to-large ES and were able to distinguish between participants who were "Improved" or "Not Improved," highlighting the importance of the selected instruments in future studies. In contrast, there was a minimal change or an increase in symptom severity among patients in the "Not Improved" group as determined by the GI Anchor. Scores from the 5 domains of the QOLRAD interestingly identified the vitality score,

physical/social function score, and sleep disturbance scores as the most prominent residual problems for the "Not Improved" group. Though some symptomatic improvement in the Food and Drink domain and in the Emotional Distress domain were noted among "Not Improved" participants, the ES and responsiveness were substantially lower compared to the "Improved" groups. This suggests that participants' assessments of being "improved" or "not improved" are strongly influenced by symptoms outside of the gut that are indirectly influenced by GI dysfunction, and not just GI dysfunction alone, and may support the importance of measuring emotional and physical functioning domains.

Our study had many strengths. To our knowledge, this is the first international study to assess the feasibility and reliability (internal consistency) of collection of validated patient-reported outcome surveys of SSc-related GERD in a prospective fashion. These data provide confidence that the application of these surveys in the clinical setting is feasible, even in the international community, when translations are available. Second, after changes in clinical management, these clinical instruments were sensitive to change after a clinical intervention, demonstrating their usefulness as longitudinal outcome measures. Finally, all instruments showed low floor and ceiling effect, thus minimally affecting sensitivity to change.

Weaknesses of our study include that the PROMIS instrument and GerdQ are not translated into the languages of all participating sites; therefore, data collection related to these specific surveys was limited to English-speaking countries. Thus we were not able to examine the results by subgroups of patients (i.e., limited versus diffuse cutaneous involvement, autoantibodies, or disease duration). We did not explore assessment with objective testing to confirm the diagnosis of GERD in each patient owing to limitations in

funding; however, we used an appropriate surrogate measure¹⁰. Previous data by Bae, *et al* showed that the GIT 2.0 Reflux scale had a high sensitivity for upper GI involvement, as assessed by upper endoscopy and manometry²⁷. As in any longitudinal observational cohort, there were missing data. However, our data provide robust estimates.

We prospectively evaluated a large international SSC cohort with symptomatic GERD to assess the sensitivity to change over time of validated PRO measures. We identified that these GI-specific PRO are sensitive to change in SSC. Incorporating such instruments into clinical trials as part of an outcome measure may provide a standardized approach to assess the symptomatic improvement of patients with GERD in this disease. Additionally, longitudinal data collection about the specifics on the effect on the instrument response are potential next steps for this large international cohort.

ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

REFERENCES

1. Denton CP, Khanna D. Systemic sclerosis. *Lancet* 2017; 390:1685-99.
2. Nagaraja V, McMahan ZH, Getzug T, Khanna D. Management of gastrointestinal involvement in scleroderma. *Curr Treatm Opt Rheumatol* 2015;1:82-105.
3. Gyger G, Baron M. Gastrointestinal manifestations of scleroderma: recent progress in evaluation, pathogenesis, and management. *Curr Rheumatol Rep* 2012;14:22-9.
4. Wipff J, Coriat R, Masciocchi M, Caramaschi P, Derk CT, Hachulla E, et al. Outcomes of Barrett's oesophagus related to systemic sclerosis: a 3-year EULAR Scleroderma Trials and Research prospective follow-up study. *Rheumatology* 2011;50:1440-4.
5. Ebert EC. Esophageal disease in scleroderma. *J Clin Gastroenterol* 2006;40:769-75.
6. Wipff J, Allanore Y, Soussi F, Terris B, Abitbol V, Raymond J, et al. Prevalence of Barrett's esophagus in systemic sclerosis. *Arthritis Rheum* 2005;52:2882-8.
7. Chunlertrith K, Noiprasit A, Foocharoen C, Mairiang P, Sukeepaisarnjaroen W, Sangchan A, et al. GERD questionnaire for diagnosis of gastroesophageal reflux disease in systemic sclerosis. *Clin Exp Rheumatol* 2014;32 Suppl 86:S-98-102.
8. Savarino E, Bredenoord AJ, Fox M, Pandolfino JE, Roman S, Gyawali CP; International Working Group for Disorders of Gastrointestinal Motility and Function. Expert consensus document: advances in the physiological assessment and diagnosis of GERD. *Nat Rev Gastroenterol Hepatol* 2017;14:665-76.
9. Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol* 2013;108:302-28.
10. DeVault KR, Castell DO, American College of Gastroenterology. Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *Am J Gastroenterol* 2005; 100:190-200.
11. Wiklund IK, Glise H. Quality of life in different gastrointestinal conditions. *Eur J Surg Suppl* 1998;582:56-61.
12. Wiklund IK, Junghard O, Grace E, Talley NJ, Kamm M, Veldhuyzen van Zanten S, et al. Quality of life in reflux and dyspepsia patients. Psychometric documentation of a new disease-specific questionnaire (QOLRAD). *Eur J Surg Suppl* 1998;583:41-9.
13. Khanna D, Hays RD, Maranian P, Seibold JR, Impens A, Mayes MD, et al. Reliability and validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. *Arthritis Rheum* 2009;61:1257-63.
14. Khanna D, Furst DE, Maranian P, Seibold JR, Impens A, Mayes MD, et al. Minimally important differences of the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. *J Rheumatol* 2011;38:1920-4.
15. Spiegel BM, Hays RD, Bolus R, Melmed GY, Chang L, Whitman C, et al. Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) gastrointestinal symptom scales. *Am J Gastroenterol* 2014;109:1804-14.
16. Jones R, Junghard O, Dent J, Vakil N, Halling K, Wernersson B, et al. Development of the GerDQ, a tool for the diagnosis and management of gastro-oesophageal reflux disease in primary care. *Aliment Pharmacol Ther* 2009;30:1030-8.
17. Khanna D, Hays RD, Shreiner AB, Melmed GY, Chang L, Khanna PP, et al. Responsiveness to change and minimally important differences of the Patient-Reported Outcomes Measurement Information System Gastrointestinal Symptoms Scales. *Dig Dis Sci* 2017;62:1186-92.
18. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
19. Wallace J. Confirmatory factor analysis of the cognitive failures questionnaire: evidence for dimensionality and construct validity. *Pers Individ Dif* 2004;37:307-24.
20. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al; Relaxin Study Group; Scleroderma Clinical Trials Consortium. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. *J Rheumatol* 2005;32:832-40.
21. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale: Lawrence Erlbaum Associates; 1988.
22. Khanna D, Nagaraja V, Gladue H, Chey W, Pimentel M, Frech T. Measuring response in the gastrointestinal tract in systemic sclerosis. *Curr Opin Rheumatol* 2013;25:700-6.
23. Kulig M, Leodolter A, Vieth M, Schulte E, Jaspersen D, Labenz J, et al. Quality of life in relation to symptoms in patients with gastro-oesophageal reflux disease — an analysis based on the ProGERD initiative. *Aliment Pharmacol Ther* 2003;18:767-76.
24. Stentoft P, Hendel L, Aggestrup S. Esophageal manometry and pH-probe monitoring in the evaluation of gastroesophageal reflux in patients with progressive systemic sclerosis. *Scand J Gastroenterol* 1987;22:499-504.
25. Lock G, Holstege A, Lang B, Scholmerich J. Gastrointestinal manifestations of progressive systemic sclerosis. *Am J Gastroenterol* 1997;92:763-71.
26. Arif T, Masood Q, Singh J, Hassan I. Assessment of esophageal involvement in systemic sclerosis and morphea (localized scleroderma) by clinical, endoscopic, manometric and pH metric features: a prospective comparative hospital based study. *BMC Gastroenterol* 2015;15:24.
27. Bae S, Allanore Y, Furst DE, Bodukam V, Coustet B, Morgaceva O, et al. Associations between a scleroderma-specific gastrointestinal instrument and objective tests of upper gastrointestinal involvements in systemic sclerosis. *Clin Exp Rheumatol* 2013;31 Suppl 76:57-63.