Reliability and Validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument

DINESH KHANNA,[1] RON D. HAYS,[2] PAUL MARANIAN,[1] JAMES R. SEIBOLD,[3] ANN IMPENS,[3] MAUREEN D. MAYES,[4] PHILIP J. CLEMENTS,[1] TERRI GETZUG,[1] NIHAL FATHI,[5] AMBER BECHTEL,[1] AND DANIEL E. FURST[1]

Objective. To refine the previously developed scleroderma (systemic sclerosis [SSc]) gastrointestinal tract (GIT) instrument (SSC-GIT 1.0).

Methods. We administered the SSC-GIT 1.0 and the Short Form 36 to 152 patients with SSc; 1 item was added to the SSC-GIT 1.0 to assess rectal incontinence. In addition, subjects completed a rating of the severity of their GIT involvement (from very mild to very severe). Evaluation of psychometric properties included internal consistency reliability, test–retest reliability (mean time interval 1.1 weeks), and multitrait scaling analysis.

Results. Study participants were mostly women (84%) and white (81%); 55% had diffuse SSc. Self-rated severity of GIT involvement ranged from no symptoms to very mild (39%), mild (21%), moderate (31%), and severe/very severe (9%). Of an initial 53 items in the SSC-GIT 1.0, 19 items were excluded, leaving a 34-item revised instrument (the University of California, Los Angeles Scleroderma Clinical Trial Consortium GIT 2.0 [UCLA SCTC GIT 2.0]). Analyses supported 7 multi-item scales: reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning. Test–retest reliability estimates were ≥0.68 and coefficient alphas were ≥0.67. Participants who rated their GIT disease as mild had lower scores on a 0–3 scale on all 7 scales. Symptom scales were also able to discriminate subjects with corresponding clinical GIT diagnoses. The Total GIT Score, developed by averaging 6 of 7 scales (excluding constipation), was reliable and provided greater discrimination between mild, moderate, and severe self-rated GIT involvement than individual scales.

Conclusion. This study provides support for the reliability and validity of the UCLA SCTC GIT 2.0, an improvement over the SSC-GIT 1.0, and supports a Total GIT Score in SSc patients with GIT.

INTRODUCTION

Gastrointestinal tract (GIT) involvement occurs in approximately 90% of patients with systemic sclerosis (SSc; scleroderma) (1,2), and has a negative impact on health-related quality of life (HRQOL) (3,4). Support for the reliability and validity of the Scleroderma Gastrointestinal Tract Involvement 1.0 (SSC-GIT 1.0) instrument was previously reported (5). However, the 52-item SSC-GIT 1.0 is fairly time consuming to administer, especially when administered in combination with other patient-reported measures. The goals of this study were to 1) refine the

SSC-GIT 1.0 and develop a shorter but still reliable and valid instrument, 2) differentiate reflux symptoms from the symptoms of distention/bloating, 3) add a scale to evaluate rectal incontinence because of its high prevalence in SSc (6), and 4) develop a composite score that captures the overall GIT burden association with SSc. We followed the Food and Drug Administration draft guidance on the development/modification of a patient-reported outcome measure (7) and developed a 34-item revised instrument (the University of California, Los Angeles Scleroderma Clinical Trial Consortium GIT 2.0 [UCLA SCTC GIT 2.0]).

MATERIALS AND METHODS

Development and evaluation of the SSC-GIT 1.0. We searched the PubMed electronic database for existing GIT questionnaires; convened an expert panel of rheumatologists, gastroenterologists, and psychometricians; pilot tested the SSC-GIT 1.0 (5) in focus groups comprised of patients with SSc; and field tested the SSC-GIT 1.0 in 88 patients with SSc and GIT involvement. Analyses of the SSC-GIT 1.0 supported 6 multi-item HRQOL scales: reflux/indigestion, diarrhea, constipation, pain, emotional well-being, and social functioning. The SSC-GIT 1.0 was found to be feasible, reliable, and valid.

Field testing of the UCLA SCTC GIT 2.0. Patients with SSc and GIT involvement were invited to participate at the following 3 scleroderma centers in the US: UCLA, Los Angeles, California; University of Michigan, Ann Arbor, Michigan; and University of Texas at Houston, Houston, Texas. The protocol was approved by the Institutional Review Board at each institution, and each subject signed a consent form prior to completing the questionnaires. In addition to completing the paper and pencil SSC-GIT 1.0 questionnaire (5), all of the patients provided sociodemographic information (age, sex, ethnicity, and level of education), completed the Short Form 36 version 2 (SF-36 v. 2) (8), a global self-rating of GIT severity: "In the past 1 week, how severe were your gastrointestinal (gut, GI) symptoms overall?" (none, very mild, mild, moderate, severe, or very severe), and a self-rating of upper and lower GIT severity: "In the past 1 week, how severe were your upper/lower gastrointestinal (gut, GI) symptoms?" (none, very mild, mild, moderate, severe, or very severe). Each physician performed a physical examination to determine the type of SSc (limited or diffuse) and provide their GIT diagnoses.

Patient-reported measures. The SF-36 v. 2 is a generic health status measure consisting of 36 items assessing 8 domains (8,9): physical functioning (10 items), bodily pain (2 items), role limitations due to physical health perceptions (4 items), general health perceptions (5 items), mental health (5 items), role limitations due to emotional problems (3 items), vitality (4 items), social functioning (2 items), and health transition (1 item). The 8 SF-36 scales are summarized into physical component summary (PCS) and mental component summary (MCS) scores. The summary scores are normalized to the US general population,

for whom the mean ± SD score is 50 ± 10. We used the acute (1-week) recall period version of the SF-36 v. 2 (8).

The SSC-GIT 1.0 is a 52-item instrument with 6 multi-item scales: reflux/indigestion, diarrhea, constipation, pain, emotional well-being, and social functioning (5). Each item is scored on a 0–100 possible range, where 100 represents better HRQOL. The SSC-GIT 1.0 has a 1-week recall period. In this study, we added 1 item to assess rectal incontinence. From the data derived from the SSC-GIT 1.0, we developed a 34-item revised instrument (the UCLA SCTC GIT 2.0) using the psychometric analysis as described below.

Psychometric analysis. The 53 items (52 items from the SSC-GIT 1.0 and 1 rectal incontinence item) were grouped into 8 hypothesized scales based on their content: reflux, distention/bloating, diarrhea, fecal soilage (to assess rectal incontinence), constipation, pain, emotional well-being, and social functioning. Since one of our objectives was to discriminate the reflux symptoms from the symptoms of distention/bloating, we divided the items in the original reflux/indigestion scale into the reflux scale and the distention/bloating scale (which is better wording than indigestion). Unlike the SSC-GIT 1.0, items were scored on a 0–3 possible range, with lower values indicating better HRQOL. A multitrait analysis was used to evaluate the extent to which each item correlated more strongly with its hypothesized scale than with other scales (10). Items identified as having either low item-to-hypothesized scale correlations ($r \leq 0.40$; n = 1), poor discrimination across scales (n = 5), or any subset of items having high correlations among themselves (n = 13) were excluded. Based on these analyses, 7 multi-item scales (reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning) were retained. Based on the multitrait analysis, dysphagia to solid food (question 1 of Supplementary Appendix A, available in the online version of this article at http://www3.interscience.wiley.com/journal/77005015/home) was retained under the reflux scale. In addition, because only 1 item of the pain scale was retained (question 34 of Supplementary Appendix A, available in the online version of this article at http://www3.interscience.wiley.com/journal/77005015/home) after multitrait analysis and pertained to the symptoms of constipation, the pain item was merged with the constipation scale and resulted in the final UCLA SCTC GIT 2.0 instrument. Version 2 of the GIT instrument contains 33 items from the SSC-GIT 1.0 and 1 item to assess fecal soilage.

Mean ± SD scores, ranges, and percentages of respondents scoring the minimum (floor) and maximum (ceiling) possible scores were calculated to evaluate scale score distributions.

Internal consistency reliability was estimated using Cronbach's alpha (11). For test–retest reliability, intraclass correlation coefficients (ICCs) over a mean interval of 1.1 weeks in 25 subjects were calculated using a 2-way random-effects model (12), previously described in detail (5). Four subjects each developed diarrhea and constipation during

test–retest visits, and ICCs were calculated including and excluding these patients.

We assessed the construct validity by exploring the association between the UCLA SCTC GIT 2.0 and other HRQOL measures. Correlations ≤0.29 were considered to be small, between 0.30 and 0.49 were moderate, and ≥0.50 were large (13). We hypothesized a priori moderate product-moment correlations between the social functioning scale of the UCLA SCTC GIT 2.0 and the social functioning scale of the SF-36, and between the emotional well-being scale of the UCLA SCTC GIT 2.0 and the role-emotional scale of the SF-36 (r ≥ 0.30). We also hypothesized that the upper GIT anchor would have larger correlations with upper GIT scales than lower GIT scales, and the lower GIT anchor would have larger correlations with lower GIT scales than upper GIT scales.

We examined the ability of the UCLA SCTC GIT 2.0 to differentiate among patients with mild, moderate, and severe GIT involvement. We also hypothesized that symptom-specific scale scores (e.g., the reflux scale) would be higher (worse HRQOL) in patients with a specific GIT clinical diagnosis (e.g., gastroesophageal reflux disease [GERD]). The self-rated severity of GIT (external anchor) was collapsed into 4 groups: no symptoms/very mild, mild, moderate, and severe/very severe. Tukey's post hoc adjustment was used to test for any significant differences in the analyses of variance.

Exploratory factor analysis was performed to evaluate the underlying structure of the 7 multi-item scales (14). Criteria used to select the most plausible model included components accounting for ≥5% of the variance and principal component eigenvalues >1 (15). Oblique promax rotation was performed to estimate factor correlations (rather than assume they were uncorrelated) (16). An oblique rotation often yields more interpretable factors with a simpler structure where factors are allowed to correlate rather than obtained with an orthogonal rotation where factors are uncorrelated.

All analyses were performed using Stata statistical software, version 9.2 (Stata, College Station, TX), and *P* values less than 0.05 were considered statistically significant.

RESULTS

Study population. The participants consisted of patients with SSc who were evaluated at 3 academic scleroderma centers. The majority of participants were women (84.1%) and white (81.4%), 55.3% had diffuse SSc, and their mean age was 50.9 years (Table 1). The mean ± SD SF-36 PCS score was 36.7 ± 9.3 and the mean ± SD SF-36 MCS score was 47.1 ± 12.5. Self-rated severity of GIT ranged from no symptoms to very mild (39%), mild (21%), moderate (31%), and severe/very severe (9%). The majority of patients had a diagnosis of GERD (91%), followed by small intestinal bowel bacterial overgrowth (11%), gastroparesis (11%), and diarrhea (11%). Patients with limited SSc had statistically higher (worse HRQOL) scores in the distention/bloating, constipation, social function, and emotional well-being scales compared with those with diffuse SSc (*P* < 0.05; data not shown).

Table 1. Baseline characteristics of the study participants (n = 152)*

Variables	Total sample
Age, mean ± SD years	50.9 ± 12.3
Women	128 (84.1)
Race†	
White	118 (81.4)
African American	11 (7.6)
Asian	9 (6.2)
Other	7 (4.8)
Hispanic	26 (17.1)
Education	
Less than or equal to high school graduate	40 (26.3)
Some college	44 (29.0)
College graduate	37 (24.3)
Graduate degree	31 (20.4)
Type of systemic sclerosis	
Limited	62 (40.7)
Diffuse	84 (55.3)
Overlap	6 (4.0)
Physician's GIT diagnoses, %	
Gastroesophageal reflux disease	91
Gastroparesis	11
Gastric antral venous ectasia	4
Small bowel bacterial overgrowth	11
Bowel pseudoobstruction	3
Diarrhea	11
Rectal incontinence	7
Constipation	5
HRQOL, mean ± SD	
SF-36 physical component summary	36.7 ± 9.3
SF-36 mental component summary	47.1 ± 12.5
Self-rated gastrointestinal severity	
No symptoms	29 (19.1)
Very mild	30 (19.7)
Mild	33 (21.7)
Moderate	47 (30.9)
Severe	9 (5.4)
Very severe	4 (2.6)

* Values are the number (percentage) unless otherwise indicated. GIT = gastrointestinal tract; HRQOL = health-related quality of life; SF-36 = Short Form 36.
† Data not available for 7 patients.

Evaluation of the UCLA SCTC GIT 2.0. The descriptive statistics and reliability estimates for the UCLA SCTC GIT 2.0 are shown in Table 2. Multitrait analyses provided support for 7 separate scales: reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning (see Supplementary Appendix A, available in the online version of this article at http://www3.interscience.wiley.com/journal/77005015/home). Mean ± SD scores ranged from 0.26 ± 0.51 (possible score 0–3) for the social functioning scale to 1.07 ± 0.82 for the distension/bloating scale. Test–retest reliability ICC estimates for the 7 scales over a mean interval of 1.1 weeks ranged from 0.68 (fecal soilage scale) to 0.89 (constipation scale), and the coefficient alpha was >0.70 for all scales except the constipation scale (α = 0.67). Ceiling effects ranged from 11% for the reflux scale to 80% for the fecal soilage scale.

Table 2. Descriptive statistics and internal consistency reliability of the UCLA SCTC GIT 2.0*

Scale	Sample size	No. of items	Mean ± SD score	Median score	Minimum score	Maximum score	% with floor effect	% with ceiling effect	Cronbach's α	ICC
Reflux	151	8	0.69 ± 0.53	0.63	0	2.65	0.0	11.3	0.74	0.85
Distention/bloating	151	4	1.07 ± 0.82	1.00	0	3.00	2.6	13.8	0.77	0.83†
Diarrhea	152	2	0.56 ± 0.67	0.50	0	2.00	7.2	50.0	0.74	0.84‡
Fecal soilage	151	1	0.30 ± 0.67	0.00	0	3.00	2.7	79.5	NA	0.68
Constipation	151	4	0.43 ± 0.50	0.25	0	2.25	0.0	39.7	0.67	0.89
Emotional well-being	151	9	0.49 ± 0.66	0.22	0	2.80	0.0	38.8	0.91	0.78
Social functioning	151	6	0.26 ± 0.51	0.00	0	0.00	3.0	60.9	0.88	0.88
Total GIT Score§	152	30	0.66 ± 0.46	0.59	0	2.10	0.0	4.0	0.71	0.81

* UCLA SCTC GIT 2.0 = University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract 2.0 instrument; ICC = intraclass correlation coefficient; NA = not applicable; GIT = gastrointestinal tract; HRQOL = health-related quality of life.
† The ICC was 0.53 when 4 patients who developed diarrhea and constipation during the test–retest period were included.
‡ The ICC was 0.55 when 4 patients who developed diarrhea and constipation during the test–retest period were included.
§ The sum of 6 of 7 scales (excludes the constipation scale). Each scale is scored from 0 (better HRQOL) to 3 (worse HRQOL).

The social functioning and emotional well-being scales of the UCLA SCTC GIT 2.0 correlated (r = 0.36, $P < 0.01$ for both) with the social functioning and emotional well-being scales of the SF-36, respectively. As hypothesized, the self-rated GIT severity was variably correlated with the 7 scales, ranging from 0.21 for the constipation scale to 0.50 for the distention/bloating scale. The reflux scale had a higher correlation with upper GIT self-rated severity, distention/bloating had a similar correlation with upper and lower self-rated severity, and the diarrhea, fecal soilage, and constipation scales had a higher correlation with lower GIT self-rated severity (Table 3).

The UCLA SCTC GIT 2.0 scales discriminated between the self-rated severity of GIT (Table 4). All 7 scales showed the lowest scores (least involvement) in people with very mild to mild disease and the highest scores (worst disease) in participants with severe to very severe disease.

In addition, symptom-specific scales were able to discriminate subjects with a corresponding clinical GIT diagnosis (Table 5). For example, the reflux scale was able to discriminate patients who had a clinical diagnosis of GERD versus those who did not ($P < 0.05$).

Exploratory factor analysis of the 7 scales suggested 2 underlying factors with eigenvalues >1, explaining >34% of the variation in the components (Table 6). Four of the 7

scales had noteworthy loadings on the first factor, and the second factor was defined primarily by the reflux, distension/bloating, and constipation scales. The estimated correlation between the factors was −0.55. As previously observed (5), the underlying factors did not represent anatomic upper and lower GIT.

Because there was a negative correlation between the diarrhea and constipation scales (−0.06) and a higher correlation between the diarrhea scale versus the other 5 remaining scales and self-rated severity of GIT when compared with the constipation scale, we created a Total GIT Score, excluding the constipation scale. The Total GIT Score is the average of the remaining 6 scales (average [reflux scale + distention/bloating scale + diarrhea scale + fecal soilage scale + emotional well-being scale + social functioning scale]) and is scored from 0 (better HRQOL) to 3 (worse HRQOL).

Total GIT Score. The mean ± SD Total GIT Score was 0.66 ± 0.46 with no floor effect and only a small ceiling effect (4%). The test–retest reliability ICC estimate was 0.81 and the coefficient alpha was 0.71. The Total GIT Score had large correlation coefficients with self-rated GIT severity (0.60) and upper (0.52) and lower (0.60) GIT severity. The Total GIT Score was able to discriminate among each of the 4 self-rated severity groups: no symptoms/very mild, mild, moderate, and severe/very severe (F = 31.3, $P < 0.001$). The F ratio was the largest for the Total GIT Score, indicating that it was the most sensitive measure to self-rated GIT disease (Table 4). However, symptom-specific scales were more discriminatory in separating subjects with a corresponding clinical GIT diagnosis compared with the Total GIT Score (Table 5).

DISCUSSION

GIT involvement has a major impact on the HRQOL of SSc patients (3,4). In this study, we refined our previously developed SSC-GIT 1.0. The results of this study provide support for the reliability and validity of the UCLA SCTC GIT 2.0. The UCLA SCTC GIT 2.0 captures SSc-related GIT

Table 3. Spearman's correlation coefficients between the 3 self-rated GIT severity scales versus the UCLA SCTC GIT 2.0 scales*

Scale	GIT severity	Upper GIT severity	Lower GIT severity
Reflux	0.47	0.56	0.31
Distention/bloating	0.50	0.48	0.47
Diarrhea	0.36	0.22	0.45
Fecal soilage	0.21	0.13†	0.29
Constipation	0.20	0.28	0.30
Social functioning	0.41	0.44	0.31
Emotional well-being	0.43	0.38	0.55
Total GIT Score	0.60	0.52	0.60

* See Table 2 for definitions.
† All other coefficients are significant at $P < 0.05$.

Table 4. Ability of the UCLA SCTC GIT 2.0 to differentiate between the severity of self-rated GIT involvement*

Scale	No symptoms/ very mild (n = 59)	Mild (n = 33)	Moderate (n = 47)	Severe/ very severe (n = 13)	F test	*P*
Reflux	0.41†	0.64‡	0.91§	1.23	17	< 0.001
Distention/bloating	0.59¶	1.15#	1.44	1.75	17	< 0.001
Diarrhea	0.30†	0.67#	0.65§	1.19	8.8	< 0.001
Fecal soilage	0.15**	0.27	0.36	0.77	3.3	0.02
Constipation	0.26	0.55	0.54	0.51	2.5	0.02
Emotional well-being	0.25†	0.33‡	0.61§	1.54	21	< 0.001
Social functioning	0.09†	0.14‡	0.41§	0.84	12	< 0.001
Total GIT Score	0.36¶	0.64‡	0.86§	1.31	31	< 0.001

* Values are the mean score unless otherwise indicated. Each scale is scored from 0 (better HRQOL) to 3 (worse HRQOL). See Table 2 for definitions.
† $P < 0.05$ between very mild and moderate disease; $P < 0.05$ between very mild and severe disease.
‡ $P < 0.05$ between mild and moderate disease; $P < 0.05$ between mild and severe disease.
§ $P < 0.05$ between moderate and severe disease.
¶ $P < 0.05$ between very mild and mild disease; $P < 0.05$ between very mild and moderate disease; $P < 0.05$ between very mild and severe disease.
\# $P < 0.05$ between mild and severe disease.
** $P < 0.05$ between very mild and severe disease.

activity and severity and can be used both in clinical trials and day-to-day care.

The reliability of the UCLA SCTC GIT 2.0, as assessed by test–retest and internal consistency, was found to be satisfactory (correlation coefficients ≥0.67 for both). The UCLA SCTC GIT 2.0 has 7 scales: reflux, distention/ bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning.

We made 5 refinements to our previous instrument, the SSC-GIT 1.0. First, we have decreased the number of items from 52 items to 34 items without sacrificing the reliability of the instrument. This should decrease patient burden, especially if the UCLA SCTC GIT 2.0 is administered in combination with other patient-reported measures.

Second, the SSC-GIT 1.0 had one reflux/indigestion scale. After the development of the SSC-GIT 1.0, it was realized that a combined reflux/indigestion scale might not be able to differentiate diagnoses of GERD, gastropare-sis, or bacterial overgrowth. Therefore, we decided to separate the scale into reflux and distention/bloating (which is more appropriate wording than indigestion). The distention/bloating scale is psychometrically sound and is able to differentiate between patients who had the diagnosis of bacterial overgrowth syndrome and gastroparesis and those who did not (Table 5). Third, we have added a 1-item fecal soilage scale to capture rectal incontinence, because up to one-third of SSc patients have soiling (6) and it has a negative impact on HRQOL (17). Fourth, we have developed a composite Total GIT Score to capture the overall burden of SSc-associated GIT. The Total GIT Score was found to be reliable and valid and more discriminatory in differentiating among self-rated GIT disease than individual scales. Last, we have simplified the scoring algorithm to improve its feasibility in clinical care; the instrument can be scored on a 0–3 scale and does not require conversion to a 0–100 scale.

Similar to the SSC-GIT 1.0, we noticed a high ceiling effect, ranging from 11.0% in the reflux scale to 80% in the fecal soilage scale with the UCLA SCTC GIT 2.0. This ceiling effect may be due to relatively mild to moderate disease recruited in our study and reflect variation in symptoms and anatomic GIT involvement from patient to patient (18,19). For example, only 13 patients (8%) rated their GIT disease as severe or very severe. However, the

Table 5. Comparison of mean scores of the UCLA SCTC GIT 2.0 scales in patients with and without a clinical GIT diagnosis*

Scale	Clinical diagnosis	With diagnosis, mean score	Without diagnosis, mean score
Reflux	GERD	0.72	0.38†
Distention/ bloating	Gastroparesis	1.63	1.00†
Distention/ bloating	Bacterial overgrowth	1.66	1.01†
Distention/ bloating	Pseudoobstruction	2.25	1.04†
Diarrhea	Diarrhea	1.00	0.51†
Fecal soilage	Rectal incontinence	0.50	0.28
Constipation	Constipation	0.81	0.41†

* Each scale is scored from 0 (better HRQOL) to 3 (worse HRQOL). GERD = gastroesophageal reflux disease; see Table 2 for additional definitions.
† $P < 0.05$.

Table 6. Oblique promax rotated 2-factor pattern matrix for the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract 2.0 instrument

Scale	Factor 1	Factor 2
Reflux	0.24	0.42
Distention/bloating	0.25	0.44
Diarrhea	0.74	−0.18
Fecal soilage	0.58	−0.11
Constipation	−0.30	0.58
Emotional well-being	0.71	0.11
Social functioning	0.58	0.29

Total GIT Score was associated with less ceiling effect (4%).

The 7 scales and Total GIT Score showed both convergent and divergent validity; the self-rated GIT severity correlated with the 7 scales and the Total GIT Score, and ranged from 0.21 for the constipation scale to 0.60 for the Total GIT Score. In addition, the 7 scales and the Total GIT Score were able to discriminate between the self-rated severities of GIT. All 7 scales and the Total GIT Score showed the lowest scores (indicating better GIT health) in people with very mild/mild disease and the highest scores (indicating worse GIT health) in participants with severe/ very severe disease. The Total GIT Score had the highest F statistics, suggesting that it has greater discriminatory ability than individual scales. We suggest that the Total GIT Score be reported to assess the overall burden/severity of GIT involvement. This can be used in both clinical practice and trials.

On the other hand, individual scales were able to discriminate between individual corresponding clinical diagnoses; scores for a specific scale (e.g., the reflux scale) were statistically lower in patients with a specific clinical diagnosis (e.g., GERD) compared with patients who did not have the diagnosis. Therefore, the scores of each scale can be used separately in clinical care and as outcome measures in randomized controlled trials as an activity index. As an outcome measure, we recommend presenting separate scores for each of the 7 scales, because it was more discriminatory for clinical GIT diagnoses and likely to change over time or with specific treatment.

Our study has significant strengths. First, this was a prospective 3-center study, with the primary objective being to refine the previously developed GIT instrument. Second, we showed that symptom-specific scales are able to discriminate subjects with corresponding clinical GIT diagnoses. These characteristics are sufficiently robust for both clinical trials and clinical care. Third, we showed that the Total GIT Score score is useful to assess overall burden/severity of GIT involvement.

Our study is not without limitations. We did not assess the UCLA SCTC GIT 2.0 versus radiologic test measures such as gastroesophageal endoscopy, esophageal manometry, and breath test because the objective of the current study was to refine the previously developed instrument and due to resource constraints. Also, we did not assess symptoms of xerostomia. Second, the Total GIT Score is an unweighted additive score rather than giving different weights to scales. Nevertheless, the Total GIT Score showed no floor and very low ceiling effects compared with individual scales, and a greater discriminatory ability than individual scales. One of our goals was to develop a measure that can be easily scored in clinical practice, so an unweighted composite score is preferable. Furthermore, other unweighted composite measures in rheumatology have performed well in observational and clinical trials (20–22).

In conclusion, we report the refinement of the previously developed SSC-GIT 1.0 in SSc. Current analysis is ongoing to examine its responsiveness to change and minimally important differences for the UCLA SCTC GIT 2.0.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Khanna had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study conception and design. Khanna, Seibold, Impens, Mayes, Clements, Getzug, Fathi, Bechtel, Furst.
Acquisition of data. Khanna, Seibold, Impens, Mayes, Clements, Getzug, Fathi, Bechtel, Furst.
Analysis and interpretation of data. Khanna, Hays, Maranian, Seibold, Impens, Getzug, Fathi, Furst.

REFERENCES

1. Sjogren RW. Gastrointestinal motility disorders in scleroderma [review]. Arthritis Rheum 1994;37:1265–82.
2. Lock G, Holstege A, Lang B, Scholmerich J. Gastrointestinal manifestations of progressive systemic sclerosis. Am J Gastroenterol 1997;92:763–71.
3. Nietert PJ, Mitchell HC, Bolster MB, Curran MY, Tilley BC, Silver RM. Correlates of depression, including overall and gastrointestinal functional status, among patients with systemic sclerosis. J Rheumatol 2005;32:51–7.
4. Gliddon AE, Dore CJ, Maddison PJ, and the Quins Trial Study Group. Influence of clinical features on the health status of patients with limited cutaneous systemic sclerosis. Arthritis Rheum 2006;55:473–9.
5. Khanna D, Hays RD, Park GS, Braun-Moscovici Y, Mayes MD, McNearney TA, et al. Development of a preliminary scleroderma gastrointestinal tract 1.0 quality of life instrument. Arthritis Rheum 2007;57:1280–6.
6. Trezza M, Krogh K, Egekvist H, Bjerring P, Laurberg S. Bowel problems in patients with systemic sclerosis. Scand J Gastroenterol 1999;34:409–13.
7. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures. Use in medical product development to support labeling claims. Rockville (MD): US Health and Human Services; 2006.
8. Ware J, Kosinski M, Dewey J. How to score version two of the SF-36 health survey. Lincoln (RI): QualityMetric; 2000.
9. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
10. Hadorn DC, Hays RD. Multitrait-multimethod analysis of health-related quality-of-life measures. Med Care 1991;29: 829–40.
11. Cronbach L. Coefficient α and the internal structure of tests. Psychometrica 1951;16:297–334.
12. Hays RD. Reliability and validity (including responsiveness). In: Fayers P, Hays RD, editors. Assessing quality of life in clinical trials. 2nd ed. New York: Oxford; 2005. p. 25–39.
13. Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum; 2003.
14. Joliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. Stat Methods Med Res 1992;1:69–95.
15. Shaw M, Talley NJ, Adlis S, Beebe T, Tomshine P, Healey M. Development of a digestive health status instrument: tests of scaling assumptions, structure and reliability in a primary care population. Aliment Pharmacol Ther 1998;12:1067–78.
16. Floyd F, Widaman K. Factor analysis in the development and

refinement of clinical assessment instruments. Psychol Assess 1995;7:286–99.

17. Kenefick NJ, Vaizey CJ, Nicholls RJ, Cohen R, Kamm MA. Sacral nerve stimulation for faecal incontinence due to systemic sclerosis. Gut 2002;51:881–3.

18. Sjogren RW. Gastrointestinal features of scleroderma. Curr Opin Rheumatol 1996;8:569–75.

19. Khanna D. Gastrointestinal involvement in systemic sclerosis. In: Font J, Ramos-Casals M, Rodes J, editors. Digestive involvement in systemic autoimmune diseases. Amsterdam: Elsevier; 2008. p. 51–69.

20. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.

21. Clements P, Lachenbruch P, Siebold J, White B, Weiner S, Martin R, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. J Rheumatol 1995;22:1281–5.

22. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23(5 Suppl 39):S100–8.