

# Longitudinal Evaluation of PROMIS-29 and FACIT-Dyspnea Short Forms in Systemic Sclerosis

Monique E. Hinchcliff, Jennifer L. Beaumont, Mary A. Carns, Sofia Podluszky, Krishna Thavarajah, John Varga, David Cella, and Rowland W. Chang

**ABSTRACT. Objective.** To assess the sensitivity of the Patient-Reported Outcomes Measurement Information System 29-item Health Profile (PROMIS-29) and the Functional Assessment of Chronic Illness Therapy-Dyspnea 10-item short form (FACIT-Dyspnea) for measuring change in health status and dyspnea in systemic sclerosis (SSc).

**Methods.** One hundred patients with SSc completed the PROMIS-29, FACIT-Dyspnea, and traditional instruments [Medical Research Council Dyspnea Score, St. George's Respiratory Questionnaire (SGRQ), Health Assessment Questionnaire-Disability Index (HAQ-DI), and Medical Outcomes Study Short Form-36 (SF-36)] at baseline and 1-year visits. PROMIS-29, FACIT-Dyspnea, and traditional instrument change scores were compared across composite modified Medsger Disease Severity and modified Rodnan Skin score (mRSS) change groups.

**Results.** Moderately high Spearman correlation coefficients were observed between FACIT-Dyspnea and SGRQ ( $r = 0.57$ ), FACIT-Dyspnea functional limitations and SF-36 physical component summary (PCS;  $r = 0.51$ ), PROMIS-29 physical functioning and HAQ-DI ( $r = 0.50$ ), and SF-36 PCS ( $r = 0.52$ ) change scores. In most validity comparisons, PROMIS-29, FACIT-Dyspnea, HAQ-DI, and SF-36 scores performed similarly. While PROMIS-29 covers more content areas than SF-36 (e.g., sleep), it may do so at the expense of responsiveness of its 4-item physical function scale as compared to the multiitem-derived SF-36 PCS. Statistically significant increases in SF-36 role physical ( $p = 0.01$ ) and physical component scale ( $p = 0.016$ ), but not PROMIS-29, were observed in patients with mRSS improvement.

**Conclusion.** PROMIS-29 and FACIT-Dyspnea are valid instruments to measure health status and dyspnea in patients with SSc. In physical function assessment, longer PROMIS short forms or computer adaptive testing should be considered to improve responsiveness to the effect of skin disease changes on physical function in patients with SSc. (First Release Nov 1 2014; J Rheumatol 2015;42:64–72; doi:10.3899/jrheum.140143)

## Key Indexing Terms:

PROMIS FACIT SEVERITY INDEX SYSTEMIC SCLEROSIS HEALTH STATUS

There are many impediments to conducting effective clinical trials and to making trial comparisons in scleroderma/systemic sclerosis (SSc). These include spontaneous improvement in some untreated patients with SSc, lack of rigorously validated indices of SSc disease severity and

activity, and the lack of an accepted set of patient-reported outcome (PRO) instruments that are used in clinical trials, among others. Several traditional PRO instruments have been validated or are commonly used in SSc clinical studies, including the Scleroderma Health Assessment

*From the Department of Medicine, Division of Rheumatology, Department of Medical Social Sciences, and the Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, and the Institute for Public Health and Medicine, Chicago, Illinois; Center for Lung Health, Henry Ford Hospital, Detroit, Michigan, USA.*

*Supported in part by the National Institutes of Health (NIH) K12 HD055884 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and K23 AR059763 (MH), by a research award from the Scleroderma Foundation and the Scleroderma Research Foundation (MH), Eleanor Wood Prince Grant Initiative from the Women's Board of Northwestern Memorial Hospital (KT), NIH-National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) P60 AR064464 (RWC), NIH-NIAMS R01 AR042309 (JV, MH), and by 5 U54 AR057951 (PROMIS Statistical Center, District of Columbia, USA).*

*M.E. Hinchcliff, MD, MS, Assistant Professor of Medicine; M.A. Carns, MS; S. Podluszky, BA, Department of Medicine, Division of Rheumatology, Northwestern University Feinberg School of Medicine, and the Institute for Public Health and Medicine; J.L. Beaumont, MS, Statistical*

*Analyst/Programmer, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, and the Institute for Public Health and Medicine; K. Thavarajah, MD, MS, Clinical Assistant Professor of Medicine, Center for Lung Health, Henry Ford Hospital; J. Varga, MD, Professor of Medicine, Department of Medicine, Division of Rheumatology, Northwestern University Feinberg School of Medicine; D. Cella, PhD, Professor of Medicine, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, and the Institute for Public Health and Medicine; R.W. Chang, MD, MPH, Professor of Medicine, Department of Medicine, Division of Rheumatology, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, and the Institute for Public Health and Medicine.*

*Address correspondence to Dr. M.E. Hinchcliff, Northwestern University Feinberg School of Medicine, Department of Medicine, Division of Rheumatology, 240 E. Huron St., Suite M300, Chicago, Illinois, USA 60611. E-mail: m-hinchcliff@northwestern.edu*

*Accepted for publication September 3, 2014.*

Questionnaire Disability Index (s-HAQ-DI), Medical Outcomes Study Short Form-36 (SF-36), St. George's Respiratory Questionnaire (SGRQ), and the Medical Research Council Dyspnea Scale (MRC-DS)<sup>1</sup>.

The s-HAQ includes 2 or 3 items for 8 activity domains, including dressing, grooming, arising, eating, walking, hygiene, reach, and grip, as well as 6 visual analog scales (pain, intestinal and breathing problems, Raynaud and digital ulcer interference in daily activities, and overall disease severity scale). Mean domain scores and a composite score are calculated (low scores favorable), but there is no standardized method for incorporating visual analog scores into the total score. The SF-36 requires a licensing fee and consists of 36 items assessing physical functioning, bodily pain, mental health, role limitations attributable to emotional health, vitality, and general health perceptions. Composite physical and mental component scores are calculated (high scores favorable). The SGRQ requires special permission before use and consists of 16 differentially weighted items that assess dyspnea (low score favorable). The freely available MRC-DS is a single 5-point item assessing dyspnea (low score favorable). Different scoring systems and lack of free use for some instruments complicate routine clinical or research use.

Beginning in 2004, the National Institutes of Health sponsored the Patient-Reported Outcomes Measurement Information System (PROMIS) to develop and validate item-response theory-based item banks to quantify physical, mental, and social health across patient populations<sup>2</sup>. In addition to item banks, short forms measuring general health status [PROMIS 29-Item General Health Form (PROMIS-29)] and dyspnea [Functional Assessment of Chronic Illness Therapy-Dyspnea (FACIT-Dyspnea)] were developed and validated in the general population and in patients with self-reported chronic obstructive pulmonary disease, respectively<sup>1,3,4,5</sup>. PROMIS-sponsored instruments are available free of charge, validated in many diseases, created in multiple languages, are readily administered electronically, can be administered as part of computerized adaptive tests, and use a uniform standardized scoring system that is simple to interpret<sup>6</sup>. We have demonstrated the construct validity for discriminative purposes of PROMIS-29 and FACIT-Dyspnea 10-item short form in SSc<sup>1</sup>. The goal of this project was to determine whether PROMIS-29 and FACIT-Dyspnea are responsive to change in SSc disease severity over 1 year, and thus useful in clinical trials and other longitudinal studies.

## MATERIALS AND METHODS

We studied the first 100 patients enrolled in the Northwestern Scleroderma Patient Registry who completed PRO instruments at baseline and at 1-year followup. Patients fulfilled the American College of Rheumatology (ACR) criteria for SSc or at least 3 out of 5 CREST criteria (calcinosis, Raynaud phenomenon, esophageal dysfunction, sclerodactyly, telangiectasia)<sup>7</sup>. Subsequent to the publication of the revised ACR/European League

Against Rheumatism (EULAR) SSc criteria, retrospective chart review was conducted to determine the number of subjects that fulfilled the new 2013 criteria<sup>8</sup>. Informed consent for study participation was obtained in accordance with the Northwestern University Institutional Review Board guidelines. The battery of PRO instruments administered at baseline (pen and paper) or 1 year later (by iPads) included traditional instruments (SGRQ, MRC-DS, S-HAQ-DI, and SF-36) and novel PROMIS-sponsored instruments (PROMIS-29 and FACIT-Dyspnea 10-item short form). These modes of administration (pen and paper, electronic) have been shown to be equivalent for PROMIS instruments<sup>9</sup>.

To evaluate PRO instruments' external validity for discriminative purposes, a composite modified 7-item Medsger SSc Disease Severity Score/severity score was calculated at baseline and 1 year as described<sup>1,10</sup>. Briefly, the composite severity score was calculated using data collected within 3 months of a clinic visit that included the modified Rodnan Skin score (mRSS); serum levels of brain natriuretic peptide (BNP), creatinine (crt), and hemoglobin (hgb); 2-dimensional echocardiography with tissue Doppler estimated right ventricular systolic pressure (est. RVSP); pulmonary function test (PFT); forced vital capacity in 1 s (FVC1); and DLCO percent predicted. Medsger Disease Severity Scale items that were not included in our modified score included hematocrit, change in weight, digital pitting scars/fingertip ulcerations/gangrene, finger to palm distance, proximal weakness, as well as esophageal and small bowel study information because those data were not included in our registry.

Demographic data, including self-reported race and smoking history, were collected on questionnaires administered to patients by paper or iPad. Anthropometric, laboratory (BNP, crt, hgb), mRSS, and PFT data were attained by querying the Northwestern Medicine Enterprise Data Warehouse (EDW) using an approach with proven validity<sup>11</sup>. The Northwestern Medicine EDW is an electronic clinical data repository of patient data stored in a variety of medical software systems at Northwestern University<sup>11</sup>. Any missing data were confirmed by manual chart review, and external results not available through the EDW were added. Echocardiographic data were obtained by manual chart review. Modified Medsger Disease Severity Scale variables were classified with scores of 0 (mild), 1 (moderate), or 2 (severe). Abnormal values were prospectively defined as previously described<sup>1</sup>. Individual item scores were summed to generate a composite severity score<sup>1</sup>.

To confirm the discriminative ability of PROMIS-29 and FACIT-Dyspnea for assessing differences in SSc disease severity, subjects were divided into categories based upon the severity score (0–1 = mild disease, 2–3 = moderate disease, 4+ = more severe disease), and descriptive statistics were generated for instrument scores at baseline and followup. PROMIS and FACIT instruments use a T score metric with a mean (SD) set to 50<sup>10</sup>. The PROMIS-29 reference population was the general population, while the FACIT-Dyspnea reference population consisted of patients with self-reported chronic obstructive pulmonary disease<sup>3,12</sup>. Instruments measure the amount of a domain present, thus high functional and low symptom domain scores are favorable.

To compare sensitivity to change over time between the new and the traditional instruments for the measurement of dyspnea and global health, Spearman correlation coefficients were calculated for change scores between baseline and followup. FACIT-Dyspnea was compared to the SGRQ and MRC-DS change scores, FACIT-Dyspnea functional limitation was compared to HAQ-DI change score, and PROMIS-29 physical functioning was compared to HAQ-DI and SF-36 physical component change scores. Definitions for correlation were established prospectively as follows:  $r < 0.3$  = low correlation,  $0.3 \leq r \leq 0.5$  = moderate correlation, and  $r > 0.5$  = high correlation<sup>1</sup>.

Three additional approaches were used to assess PROMIS-29 and FACIT-Dyspnea's responsiveness to change. In the first approach, patients were divided into worsened, unchanged, and improved severity score categories. Based upon this, PROMIS-29 and FACIT-Dyspnea descriptive statistics, as well as effect sizes (ES), were determined. In the second approach, patients were divided into worsened, unchanged, and improved

MRC-DS change groups, and FACIT-Dyspnea descriptive statistics as well as ES were determined. In the third approach, the sensitivity of PROMIS-29 to change in mRSS over 1 year was assessed. A  $\geq 5$ -point decrease in mRSS was defined as improvement based upon evidence that this is an important difference<sup>13</sup>. Subjects were classified as improved (mRSS decrease  $\geq 5$ ) versus worsened (mRSS increase  $\geq 5$ ) or stable (mRSS change 0–4), and descriptive statistics, as well as ES, were determined. Effect sizes  $> 0.2$  were deemed possibly clinically important<sup>14,15</sup>. ANOVA was used to determine the significance of between-group differences. McNemar’s test was used to compare pre- versus postclassification of abnormal versus normal values. All analyses were conducted using SAS 9.2 (SAS Institute Inc.).

RESULTS

One hundred patients with SSc who completed the traditional and new PRO instruments both at baseline and 1 year later were studied. Ninety-seven percent of the subjects fulfilled the revised ACR/EULAR SSc criteria. As shown in Table 1, 91% of the subjects were white women with longstanding SSc disease duration, median (range) of 4.5 (0–32) years, defined as the duration between onset of first

Table 1. Patient baseline characteristics. Values are median (range) unless otherwise specified.

Characteristics	Variables
Age, yrs	52 (27–71)
Female sex, %	90
White, %	90
Smoking history, %	
Never	67
Past	32
Current	1
SSc duration, interval from onset of Raynaud, yrs	5 (0–35)
SSc duration, interval from onset of first non-Raynaud, yrs	4.5 (0–32)
SSc subtype, lcSSc, %	53
Antitopoisomerase autoantibodies, positive, %	26
Anticentromere autoantibodies, positive, %	17
Anti-RNA polymerase III autoantibodies, positive, %	36
Primary ANA immunofluorescence pattern, positive, %	
Speckled	41
Nucleolar	20
Homogenous	22
Centromere	17
Modified Medsger Disease Severity Scale	
Composite Score	3.0 (0–9)
Brain natriuretic peptide, pg/ml, n = 92	54.4 (4.9–1143.7)
Creatinine, mg/dl, n = 98	0.8 (0.3–8.0)
FVC % predicted, n = 97	86 (22–121)
DLCO % predicted, n = 95	69 (5–142)
Est. RVSP % predicted, n = 82	31 (15–70)
mRSS, n = 100	7 (0–46)
Hemoglobin, mg/dl, n = 98	13 (7.2–17)

Four, 7, and 42 patients did not have ACA, Scl-70, and RNA polymerase III data, respectively. Five patients either did not have an ANA test performed (n = 3) or had a negative ANA (n = 2). SSc: systemic sclerosis; lcSSc: limited cutaneous systemic sclerosis; ANA: antinuclear antibody; FVC: forced vital capacity; est. RVSP: estimated right ventricular systolic pressure; mRSS: modified Rodnan Skin score; ACA: anticentromere antibodies.

non-Raynaud symptom and baseline PRO instrument administration. Just over half (53%) of patients met classification criteria for limited cutaneous SSc<sup>16</sup>. RNA polymerase III autoantibodies were the most prevalent SSc-specific autoantibody (21 out of 58; 36%) compared to anti-centromere (16 out of 96; 17%) and antitopoisomerase I (24 out of 93; 26%). A speckled antinuclear antibody was the most prevalent immunofluorescence pattern (39 out of 95; 41%).

The mean composite modified Medsger Severity Scale score was 3.2 (SD 2.1) at baseline and remained unchanged at 1-year followup. With the exception of an increase in the proportion of patients with low hgb, there were no statistically significant changes in the objectively measured disease variables that comprise the modified Medsger Severity Score between baseline and followup (Table 2). Similar percentages of patients had an abnormal BNP, crt, FVC, and DLCO percent predicted at baseline and followup. Creatinine significantly increased in 1 patient diagnosed with incident scleroderma renal crisis<sup>17</sup>. These data suggest that the majority of patients had stable cardiac, pulmonary, and renal disease over 1 year. There were 21 prevalent cases of anemia at baseline, and 14 incident cases of anemia at followup. Manual electronic medical record review for these 14 subjects demonstrated 5 cases of anemia because of chronic disease and 6 patients with iron deficiency anemia. No identifying cause was found in 3 subjects.

Contrary to our expectations, the percentage of patients with an abnormal estimated RVSP on echocardiography declined from baseline (n = 29, 35%) to 1 year (n = 12,

Table 2. Baseline and 1-year followup of modified Medsger SSc Severity Scale Score components. Values are % unless otherwise specified. P values are for paired t test or McNemar’s test for paired proportions, as appropriate.

Characteristics	Baseline	1-Yr Followup	p
Composite Medsger Disease Severity Scale, mean (SD)	3.2 (2.1)	3.2 (2.1)	NS
Brain natriuretic peptide, % $\geq 100$ pg/ml, n = 92, 55	25	22	0.206
Renal insufficiency, crt > 1.59 mg/dl, n = 98, 86	5	7	0.317
FVC % predicted, % < 79% predicted, n = 97, 83	40	43	0.617
DLCO % predicted, % < 79% predicted, n = 95, 82	79	83	0.366
Est. RVSP, % $\geq 35\%$ predicted, n = 82, 55	35	22	0.061
mRSS, % $\geq 15$ , n = 100, 97	25	18	0.262
Hemoglobin, % < 12 mg/dl, n = 98, 76	21	30	0.046

SSc: systemic sclerosis; crt: creatinine; FVC: forced vital capacity; est. RVSP: estimated right ventricular systolic pressure; mRSS: modified Rodnan Skin score; NS: not significant.



22%). Of the 12 subjects who had serial studies, 5 had persistent and 7 had newly diagnosed RVSP elevation. An etiology to account for increased RVSP was found in 4 out of 7 patients. One patient had pulmonary hypertension secondary to progressive interstitial lung disease. Another was diagnosed with pulmonary artery hypertension (mean pulmonary artery pressure at rest  $\geq 25$  mmHg with pulmonary capillary wedge pressure  $\leq 15$  mmHg on right-heart catheterization)<sup>18</sup>. One patient had worsened pulmonary venous hypertension secondary to diastolic heart failure, and another had worsening left ventricular filling pressures on echo.

To evaluate the ability of PRO instruments to detect cross-sectional differences in SSc disease severity, descriptive statistics for new and traditional instrument scores at baseline and followup were compared between patients stratified by modified Medsger Disease Severity Scale category (Table 3). Mean baseline and followup general and physical health PRO scores were significantly worse in patients with higher modified Medsger Disease Severity Scale scores: FACIT-Dyspnea ( $p \leq 0.001$ ); FACIT-Functional limitation ( $p \leq 0.001$ ); PROMIS physical function ( $p \leq 0.001$ ); PROMIS pain interference and social role (both  $p < 0.001$ ,  $p = 0.002$ ); HAQ-DI ( $p < 0.001$ ); SF-36 physical functioning, role physical, physical component summary, bodily pain, and general ( $p < 0.001$ ). Mean PROMIS Fatigue scores between groups were significant at baseline only ( $p \leq 0.001$ ). Mean PROMIS depression, anxiety, and sleep disturbance, and SF-36 emotional role, mental health, and mental component summary scores did not differ among the groups. These data indicate that PROMIS-29 and FACIT-Dyspnea are comparable to traditional instruments for discriminating between patients with SSc with different levels of disease severity as defined by a composite modified Medsger Scale.

The correlation between the new and traditional dyspnea instrument change scores in patients with SSc was assessed (Table 4). There was a high correlation between FACIT-Dyspnea and SGRQ total score ( $r = 0.57$ ,  $p < 0.001$ ) and a moderate correlation between FACIT-Dyspnea and MRC-DS ( $r = 0.32$ ,  $p < 0.05$ ). A moderate correlation was observed between FACIT-Functional limitations and SGRQ and HAQ-DI (both  $r = 0.35$ ,  $p < 0.001$ ), and MRC-DS ( $r = 0.32$ ,  $p < 0.05$ ). There was a strong negative correlation between FACIT-Functional limitations and SF-36 physical component score ( $-0.51$ ,  $p < 0.001$ ), which is in the expected direction because low FACIT functional limitation scores are comparable to high SF-36 physical component scores. These data suggest that dyspnea contributes to functional limitations in SSc, and that FACIT-Dyspnea change scores correlate with HAQ-DI and SF-36 change scores.

The correlation between new and traditional general health instrument 1-year change scores was assessed (Table

4). Strong negative and positive correlations, respectively, were observed between changes in PROMIS-29 physical functioning subscale and HAQ-DI ( $r = -0.50$ ) and SF-36 physical component scores ( $r = 0.52$ , both  $p < 0.001$ ; Table 4). Moderate negative correlations were found between changes in PROMIS-29 anxiety ( $r = -0.26$ ,  $p < 0.05$ ), depression ( $r = -0.36$ ,  $p < 0.001$ ), and fatigue ( $r = -0.42$ ,  $p < 0.001$ ) subscales, and SF-36 mental component summary. Negative correlation coefficients occurred in cases where high scores have different meaning (good vs bad) for different instruments. These results indicate that PROMIS-29 subscale change scores correlate well with similar traditional instrument scores and that PROMIS-29 is valid for measuring changes in overall health in patients with SSc.

Descriptive statistics for FACIT-Dyspnea, PROMIS-29, SF-36, and HAQ-DI change scores were determined for modified Medsger Scale change groups (improved, stable, worsened; data not shown). There were no significant between-group differences in any of the PRO change scores.

Descriptive statistics for FACIT-Dyspnea and change scores were determined for MRC-DS change groups (Figure 1). Patients with worsened MRC-DS change scores had higher FACIT-Dyspnea change scores [mean (SD) = 5.1 (6.2)] compared to patients whose MRC score remained stable or improved [ $-1.6$  (4.6),  $-3.0$  (6.4),  $p < 0.01$ ]. Effect sizes were greatest for the worsened, followed by improved, and then stable groups (0.83,  $-0.47$ ,  $-0.36$ ,  $p < 0.01$ ). These results indicate that FACIT-Dyspnea is a sensitive measure of 1-year change in dyspnea in patients with SSc.

Descriptive statistics for FACIT-Dyspnea and PROMIS-29 change scores were determined for mRSS change groups (Table 5). Patients with mRSS improvement demonstrated greater reductions in FACIT-Dyspnea and FACIT-Dyspnea functional limitations, and PROMIS-29 anxiety, fatigue, and sleep disturbance subscales than patients with worsened or stable skin scores, although these differences were not statistically significant. PROMIS physical function, depression, pain interference, and social role subscale scores remained stable in patients regardless of changes in skin score. These results suggest that FACIT-Dyspnea and FACIT-Dyspnea functional limitations, and PROMIS-29 anxiety, depression, fatigue, and sleep disturbance subscales may be more sensitive to mRSS changes over 1 year than PROMIS-Physical function, pain interference, and social role subscale scores. Statistically significant increases in SF-36 role physical ( $p = 0.01$ ) and physical component scale ( $p = 0.016$ ) were observed in patients with mRSS improvement.

## DISCUSSION

The paucity of validated outcomes that are responsive to change is an important barrier to therapeutic response assessment in SSc. To the best of our knowledge, ours is the

Table 3. PRO scores by modified Medsger SSc Severity Scale Score group.

Variables	Severity Score	n	Baseline Mean (SD)	p*	n	Followup Mean (SD)	p*
Dyspnea							
FACIT-Dyspnea	0–1	21	34.3 (6.7)	< 0.001	23	37.1 (8.9)	< 0.001
	2–3	41	41.6 (9.5)		37	37.9 (7.6)	
	4 +	35	44.7 (8.8)		39	45.2 (9.9)	
FACIT-Functional limitations	0–1	21	33.2 (5.9)	< 0.001	23	36.7 (8.3)	< 0.001
	2–3	41	42.1 (9.6)		37	38.6 (8.2)	
	4 +	38	46.8 (10.5)		40	45.5 (10.3)	
HAQ	0–1	21	0.23 (0.33)	< 0.001	23	0.46 (0.51)	< 0.001
	2–3	40	0.65 (0.65)		37	0.55 (0.61)	
	4 +	37	1.22 (0.74)		40	1.13 (0.68)	
Physical Functioning							
PROMIS physical functioning	0–1	21	52.7 (5.7)	< 0.001	23	51.0 (7.9)	< 0.001
	2–3	41	47.3 (8.6)		37	49.6 (8.2)	
	4 +	38	40.6 (7.3)		40	41.3 (7.6)	
SF-36 physical functioning	0–1	21	84.5 (18.1)	< 0.001	23	77.4 (23.6)	< 0.001
	2–3	41	64.4 (27.3)		37	71.4 (24.3)	
	4 +	38	41.6 (23.1)		40	45.4 (26.1)	
SF-36 role physical	0–1	21	83.3 (18.6)	< 0.001	23	78.3 (25.2)	< 0.001
	2–3	40	70.6 (26.6)		37	73.8 (28.6)	
	4 +	38	45.6 (33.0)		40	52.2 (29.9)	
SF-36 PCS	0–1	21	50.6 (8.2)	< 0.001	23	48.1 (8.8)	< 0.001
	2–3	40	42.1 (11.0)		36	45.8 (10.7)	
	4 +	38	33.1 (11.2)		38	34.5 (10.9)	
Mental Health							
PROMIS anxiety	0–1	21	46.6 (8.0)	0.087	23	50.0 (9.1)	0.782
	2–3	41	50.4 (9.2)		37	50.0 (9.0)	
	4 +	38	52.2 (9.6)		40	51.3 (9.0)	
PROMIS depression	0–1	21	46.9 (7.2)	0.061	23	48.4 (9.6)	0.699
	2–3	41	48.7 (8.8)		37	50.2 (7.6)	
	4 +	38	52.3 (9.8)		40	50.1 (9.2)	
SF-36 role emotional	0–1	21	91.7 (15.6)	0.070	23	82.2 (27.1)	0.428
	2–3	41	86.6 (21.0)		36	85.9 (17.2)	
	4 +	38	78.1 (27.3)		39	78.8 (25.6)	
SF-36 mental health	0–1	21	74.2 (19.9)	0.132	23	75.9 (21.9)	0.423
	2–3	41	74.8 (15.7)		37	73.8 (16.6)	
	4 +	38	66.6 (21.9)		40	69.5 (21.6)	
SF-36 MCS	0–1	21	50.1 (9.9)	0.551	23	49.5 (13.3)	0.902
	2–3	40	51.2 (9.4)		36	50.8 (8.5)	
	4 +	38	48.5 (13.0)		38	49.8 (12.5)	
Fatigue							
PROMIS fatigue	0–1	21	46.6 (6.2)	0.019	23	48.5 (11.4)	0.069
	2–3	41	51.8 (10.0)		37	49.1 (8.1)	
	4 +	38	53.9 (10.3)		40	53.3 (9.1)	
SF-36 vitality	0–1	21	58.8 (20.4)	0.019	23	59.2 (26.9)	0.003
	2–3	41	54.6 (23.1)		37	60.6 (19.0)	
	4 +	38	43.5 (21.4)		40	44.8 (19.9)	
Pain							
PROMIS pain interference	0–1	21	48.0 (7.3)	< 0.001	23	50.0 (9.7)	0.002
	2–3	41	52.8 (9.4)		37	51.5 (9.2)	
	4 +	38	58.6 (9.4)		39	57.3 (6.7)	
SF-36 bodily pain	0–1	21	81.9 (20.2)	< 0.001	23	73.8 (22.9)	< 0.001
	2–3	41	64.2 (24.0)		37	72.9 (21.9)	
	4 +	38	53.7 (23.9)		40	52.8 (19.6)	
Sleep							
PROMIS sleep disturbance	0–1	21	47.7 (9.4)	0.095	23	48.0 (8.2)	0.156
	2–3	41	52.1 (8.8)		37	51.2 (8.5)	
	4 +	38	52.7 (8.3)		39	52.1 (7.6)	
Social							
PROMIS social role	0–1	21	53.4 (9.6)	< 0.001	23	52.7 (11.3)	0.002
	2–3	41	49.9 (10.9)		37	51.7 (8.4)	
	4 +	38	42.5 (11.5)		39	45.1 (9.2)	
SF-36 social functioning	0–1	21	83.9 (19.8)	0.005	23	79.9 (31.9)	0.008
	2–3	41	79.0 (24.8)		37	83.1 (16.7)	
	4 +	38	64.1 (26.3)		39	66.3 (24.2)	
General							
SF-36 general health	0–1	21	63.4 (23.3)	< 0.001	23	62.6 (17.8)	< 0.001
	2–3	41	53.6 (23.6)		37	59.2 (24.9)	
	4 +	38	39.5 (21.3)		39	41.7 (18.2)	

\* ANOVA p value. Three patients were missing baseline, and 1 patient was missing FACIT-Dyspnea scores at followup. One patient did not complete a PROMIS-29 questionnaire at 1 year. PRO: patient-reported outcomes; SSc: systemic sclerosis; FACIT: Functional Assessment of Chronic Illness Therapy; HAQ: Health Assessment Questionnaire; PROMIS: Patient-Reported Outcomes Measurement Information System; SF-36: Medical Outcomes Study Short Form-36; PCS: physical component summary; MCS: mental component summary.

Table 4. Spearman correlation coefficients between newer and more traditional measure, 1-year change scores.

Characteristics	SGRQ Total Score	MRC-DS	HAQ-DI	SF-36 MCS	SF-36 PCS
FACIT-Dyspnea	0.57**	0.32*	0.14	0.02	−0.34**
FACIT-Functional limitations	0.35**	0.32*	0.35**	−0.08	−0.51**
PROMIS-29 subscales					
Physical functioning	−0.21*	−0.23*	−0.50**	0.11	0.52**
Anxiety	0.13	−0.03	0.00	−0.26*	0.07
Depression	0.07	−0.06	0.12	−0.36**	−0.05
Fatigue	0.37**	0.11	0.11	−0.42**	−0.32*
Pain interference	0.05	0.07	0.22*	−0.20	−0.45**
Sleep disturbance	0.10	0.11	0.08	−0.15	−0.09
Satisfaction with social role	−0.13	−0.08	−0.24*	0.14	0.29*

\*  $p < 0.05$ . \*\*  $p < 0.001$ . SGRQ: St. George's Respiratory Questionnaire; MRC-DS: Medical Research Council–Dyspnea Scale; HAQ-DI: Health Assessment Questionnaire–Disability Index; SF-36: Medical Outcomes Study Short Form-36; FACIT: Functional Assessment of Chronic Illness Therapy; PROMIS-29: Patient-Reported Outcomes Measurement Information System 29-item Health Profile; MCS: mental component summary; PCS: physical component summary.

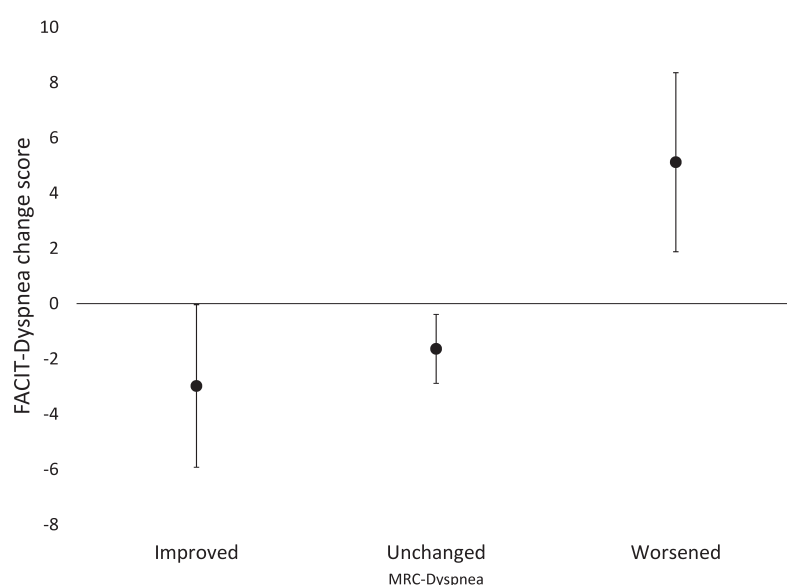


Figure 1. FACIT-Dyspnea scores by MRC-DS change groups. Mean values are expressed as T scores. FACIT-Dyspnea by MRC-DS change group could not be assessed for 13 subjects because of missing FACIT-Dyspnea or MRC-DS data. Three subjects were missing baseline FACIT-Dyspnea, 4 missing baseline MRC-DS, 5 missing followup MRC-DS, and 1 missing followup FACIT-Dyspnea score. FACIT: Functional Assessment of Chronic Illness Therapy; MRC-DS: Medical Research Council–Dyspnea Scale.

first study to assess the responsiveness to change of PROMIS-29 and FACIT-Dyspnea compared to traditional instruments in SSc.

Our SSc population was slightly more affected by their disease than in our previous report, as evidenced by a higher composite modified Medsger Disease Severity Index. Consecutive patients were approached to complete the battery of PRO instruments regardless of severity of illness. Studies have shown that most patients are willing to complete PRO instruments, suggesting that “favoring” the very sick by excluding them from health status research is unwarranted<sup>19,20</sup>.

Our results demonstrate that increased composite modified Medsger Disease Severity Scale scores were associated with statistically significant changes in new and traditional PRO instrument scores in the appropriate direction. Higher FACIT-Dyspnea, FACIT-Functional limitations, HAQ-DI, and PROMIS anxiety, depression, fatigue, pain interference, and sleep disturbance scores were observed as SSc disease severity increased. Similarly, lower SF-36 physical functioning, social role, physical component summary, role emotional mental health, mental component summary, vitality, bodily pain, social functioning, and general scores were observed in patients with higher disease

Table 5. Change in PRO scores by change in mRSS (improved = by at least 5 points); all patients. Mean values are expressed as T scores.

Variables	mRSS Change Group	Mean (SD)	Effect Size	p
Dyspnea				
FACIT-Dyspnea	Worsened/stable, n = 77	-0.5 (6.0)	-0.08	0.548
	Improved, n = 16	-1.5 (6.6)	-0.23	
FACIT-Functional limitations	Worsened/stable, n = 79	-0.8 (5.5)	-0.15	0.286
	Improved, n = 18	-2.4 (6.4)	-0.38	
HAQ	Worsened/stable, n = 79	0.00 (0.35)	0.00	0.101
	Improved, n = 16	-0.18 (0.58)	-0.31	
Physical Functioning				
PROMIS physical functioning	Worsened/stable, n = 79	0.7 (4.8)	0.16	0.866
	Improved, n = 18	1.0 (6.1)	0.16	
SF-36 physical functioning	Worsened/stable, n = 79	1.0 (14.8)	0.07	0.094
	Improved, n = 18	8.1 (20.7)	0.39	
SF-36 role physical	Worsened/stable, n = 78	-0.7 (23.9)	-0.03	0.010
	Improved, n = 18	16.3 (28.8)	0.57	
SF-36 PCS	Worsened/stable, n = 77	0.6 (6.2)	0.10	0.016
	Improved, n = 16	4.8 (6.8)	0.71	
Mental Health				
PROMIS anxiety	Worsened/stable, n = 79	0.5 (7.7)	0.06	0.239
	Improved, n = 18	-1.9 (7.3)	-0.26	
PROMIS depression	Worsened/stable, n = 79	0.2 (6.9)	0.03	0.666
	Improved, n = 18	-0.6 (8.2)	-0.07	
SF-36 mental health	Worsened/stable, n = 79	0.5 (17.0)	0.03	0.756
	Improved, n = 18	1.9 (20.9)	0.09	
SF-36 role emotional	Worsened/stable, n = 79	-2.4 (21.4)	-0.11	0.912
	Improved, n = 16	-3.1 (30.1)	-0.10	
SF-36 MCS	Worsened/stable, n = 77	0.0 (9.7)	0.00	0.769
	Improved, n = 16	-0.8 (13.6)	-0.06	
Fatigue				
PROMIS fatigue	Worsened/stable, n = 79	-0.7 (7.6)	-0.09	0.540
	Improved, n = 18	-1.9 (7.4)	-0.26	
SF-36 vitality	Worsened/stable, n = 79	2.8 (18.6)	0.15	0.966
	Improved, n = 18	3.0 (22.4)	0.13	
Pain				
PROMIS pain interference	Worsened/stable, n = 79	-0.4 (8.1)	-0.05	0.792
	Improved, n = 17	0.2 (7.1)	0.02	
SF-36 bodily pain	Worsened/stable, n = 79	1.2 (18.2)	0.07	0.713
	Improved, n = 18	-0.5 (17.5)	-0.03	
Sleep				
PROMIS sleep disturbance	Worsened/stable, n = 79	-0.1 (7.7)	-0.01	0.130
	Improved, n = 17	-3.2 (7.8)	-0.42	
Social				
PROMIS social role	Worsened/stable, n = 79	1.3 (7.9)	0.16	0.962
	Improved, n = 17	1.2 (7.8)	0.15	
SF-36 social functioning	Worsened/stable, n = 78	0.6 (19.3)	0.03	0.379
	Improved, n = 18	5.6 (28.5)	0.19	
General				
SF-36 general health	Worsened/stable, n = 78	1.6 (16.4)	0.10	0.755
	Improved, n = 18	3.0 (17.1)	0.18	

Three patients were missing followup mRSS. Three patients did not complete baseline, and 1 patient did not complete followup FACIT-dyspnea instrument. One patient did not complete PROMIS pain interference, sleep disturbance, and social role at followup. PRO: patient-reported outcomes; mRSS: modified Rodnan Skin score; FACIT: Functional Assessment of Chronic Illness Therapy; HAQ: Health Assessment Questionnaire; PROMIS: Patient-Reported Outcomes Measurement Information System; SF-36: Medical Outcomes Study Short Form-36; PCS: physical component summary; MCS: mental component summary.

severity. This is an expected finding because lower scores on all SF-36 domains represent worse outcomes. One important caveat regarding the composite modified Medsger Severity Scale that we used is that it is not sensitive to large changes in a single variable (i.e., a patient with a large change in only 1 variable will be considered stable). Another

interesting finding is that patients with SSc at both timepoints demonstrated FACIT-Dyspnea and FACIT-Functional limitations below 50. Thus, patients with SSc report less dyspnea than patients with chronic obstructive pulmonary disease (reference population).

There were statistically significant moderate and high

correlations between FACIT-Dyspnea and FACIT-Dyspnea functional limitations, and SF-36 physical component change scores, respectively. Further, significant correlations between PROMIS-29 physical functioning and SF-36 physical component change scores, and PROMIS-29 anxiety, depression, and fatigue subscale and SF-36 mental component change scores were observed. These data suggest that FACIT-Dyspnea, FACIT-Dyspnea functional limitations, and PROMIS-29 are sensitive to 1-year changes in dyspnea, anxiety, depression, and fatigue in patients with SSc as assessed by the correlation between change scores.

The responsiveness of HAQ-DI and SF-36 to mRSS changes in SSc has been demonstrated previously<sup>21,22</sup>. Patients with SSc who had a > 15% worsening of their skin score demonstrated worsening HAQ-DI scores, while patients with a > 15% improvement in skin score demonstrated improvement in HAQ-DI scores<sup>22</sup>. Similarly, patients with SSc with a  $\geq 30\%$  increase in skin score at 24 weeks during a clinical trial of the hormone relaxin demonstrated higher HAQ-DI scores and higher SF-36 physical component summary scores. We observed statistically significant differences between mean SF-36 role physical and physical component summary change scores, but not between mean HAQ-DI change scores in patients with worsening/stable or improved mRSS over 1 year. This is contrary to what was observed in the relaxin study where the HAQ-DI was more responsive to mRSS changes compared to the SF-36. Longer followup (1 year in our study compared to 24 weeks in the relaxin clinical trial) may account for the difference. There were no statistically significant differences in mean PROMIS-29 change scores in patients with worsening/stable or improved mRSS over 1 year. This is likely attributable to the fact that the SF-36 includes more than 10 items to assess physical function and role performance, all contributing to the physical component summary (PCS) score, while PROMIS-29 only includes 4 physical function items. If responsiveness in the area of physical function is desired from a PROMIS assessment, then longer physical function short forms such as the 10-item and 20-item versions should be considered, and could be used to supplement PROMIS-29. Similarly, PROMIS-43 or PROMIS-57 profiles, with 6-item and 8-item physical function short forms, respectively, may be preferable when the primary goal is a responsive physical function assessment. Of course, computerized adaptive testing (CAT) would likely provide the best measurement precision across the physical function continuum.

Improvement in mRSS over 1 year was associated with improved FACIT-Dyspnea, FACIT-Dyspnea functional limitations, and all PROMIS-29 subscales except pain interference, depression, and social role (these remained stable), but the results were not statistically significant. Less chest wall restriction and less work for activities of daily living in patients with lower skin scores may explain the association

between mRSS and dyspnea. The correlation between change in FACIT-Dyspnea and FACIT-Dyspnea functional limitations and change in forced vital capacity ( $-0.192$  and  $-0.042$ , respectively), and DLCO percent predicted ( $0.013$  and  $-0.130$ , respectively) were low. These data suggest that the correlation between change in FACIT-Dyspnea and FACIT-Dyspnea functional limitations, and change in mRSS, was not attributable to the improvement in SSc lung disease as assessed by change in FVC and DLCO percent predicted.

There were no significant differences in mean PRO change scores between modified Medsger Severity Scale change groups. This is likely attributable to the majority of patients having stable disease severity as defined by the modified Medsger Scale during 1 year of followup.

The major study limitation was the absence of patients who greatly improved or worsened as assessed by the modified Medsger Severity Scale. This led to nonstatistically significant associations between changes in the modified Medsger Severity scores and mRSS, and the PROMIS PRO measures. This could be addressed by larger sample sizes and/or longer duration of followup in future studies. Alternatively, additional items could be included in the modified Medsger Severity Scale; namely, tendon friction rub counts and novel echocardiogram variables that assess right ventricular size and function.

Study strengths include the wide breadth of clinical information available for the study cohort that enabled determination of etiologies for anemia, PFT changes, creatinine, and other clinical data, and the novel use of the modified Medsger Severity Scale administered at 2 timepoints as a surrogate of disease activity.

PROMIS-29 and FACIT-Dyspnea were comparable in performance to traditional questionnaires with the exception of assessing physical function change associated with improvement in mRSS. The SF-36 PCS was responsive to this change whereas the 4-item PROMIS physical function scale in the PROMIS-29 was not. With enhanced assessment of physical function, the PROMIS-29 and FACIT-Dyspnea may be useful in future SSc clinical research for measuring change in general health and dyspnea because they are available free of charge and easily administered on paper or electronically. The sensitivity of PROMIS instruments may be enhanced when they are administered as CAT. PROMIS-sponsored instruments are also validated in many diseases, including SSc as demonstrated here, and available in Spanish and a growing number of languages, including Portuguese, Mandarin, Dutch, Hebrew, and Hindi. Further, they use a uniform standardized scoring system that is simple to interpret.

The ultimate goal of the PROMIS initiative is to facilitate the use of PRO instruments in clinical practice, as well as in research. Our results demonstrate that PROMIS-29 and FACIT-Dyspnea may be useful for measuring changes in



SSc disease severity over 1 year. Future studies should be conducted to determine the responsiveness of PROMIS-43, -57, or -CAT to changes in SSc disease severity compared to the traditional instruments.

## ACKNOWLEDGMENT

We thank Andrew Varga for performing the manual chart review to determine the number of patients meeting new ACR/EULAR SSc criteria.

## REFERENCES

- Hinchcliff M, Beaumont JL, Thavarajah K, Varga J, Chung A, Podlasky S, et al. Validity of two new patient-reported outcome measures in systemic sclerosis: Patient-Reported Outcomes Measurement Information System 29-Item Health Profile and Functional Assessment of Chronic Illness Therapy-Dyspnea short form. *Arthritis Care Res* 2011;63:1620-8.
- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, 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.
- Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol* 2010;63:1179-94.
- Choi SW, Victorson DE, Yount S, Anton S, Cella D. Development of a conceptual framework and calibrated item banks to measure patient-reported dyspnea severity and related functional limitations. *Value Health* 2011;14:291-306.
- Yount SE, Choi SW, Victorson D, Ruo B, Cella D, Anton S, et al. Brief, valid measures of dyspnea and related functional limitations in chronic obstructive pulmonary disease (COPD). *Value Health* 2011;14:307-15.
- Khanna D, Maranian P, Rothrock N, Cella D, Gershon R, Khanna PP, et al. Feasibility and construct validity of PROMIS and "legacy" instruments in an academic scleroderma clinic. *Value Health* 2012;15:128-34.
- Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 1980;23:581-90.
- van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2013;65:2737-47.
- Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative. *Qual Life Res* 2014;23:217-27.
- Medsker TA Jr, Silman AJ, Steen VD, Black CM, Akesson A, Bacon PA, et al. A disease severity scale for systemic sclerosis: development and testing. *J Rheumatol* 1999;26:2159-67.
- Hinchcliff ME, Just E, Podlasky S, Varga J, Chang RW, Kibbe WA. Text data extraction for a prospective, research-focused data mart: implementation and validation. *BMC Med Inform Decis Mak* 2012;12:106.
- Victorson DE, Anton S, Hamilton A, Yount S, Cella D. A conceptual model of the experience of dyspnea and functional limitations in chronic obstructive pulmonary disease. *Value Health* 2009;12:1018-25.
- Amjadi S, Maranian P, Furst DE, Clements PJ, Wong WK, Postlethwaite AE, et al. Course of the modified Rodnan skin thickness score in systemic sclerosis clinical trials: analysis of three large multicenter, double-blind, randomized controlled trials. *Arthritis Rheum* 2009;60:2490-8.
- Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof* 2005;28:172-91.
- Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol* 2011;64:507-16.
- LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsker TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988; 15:202-5.
- Shanmugam VK, Steen VD. Renal disease in scleroderma: an update on evaluation, risk stratification, pathogenesis and management. *Curr Opin Rheumatol* 2012;24:669-76.
- McGoon M, Guterman D, Steen V, Barst R, McCrory DC, Fortin TA, et al. Screening, early detection, and diagnosis of pulmonary arterial hypertension: ACCP evidence-based clinical practice guidelines. *Chest* 2004;126:14S-34S.
- Krampe H, Bartels C, Victorson D, Enders CK, Beaumont J, Cella D, et al. The influence of personality factors on disease progression and health-related quality of life in people with ALS. *Amyotroph Lateral Scler* 2008;9:99-107.
- Yount S, Beaumont J, Rosenbloom S, Cella D, Patel J, Hensing T, et al. A brief symptom index for advanced lung cancer. *Clin Lung Cancer* 2012;13:14-23.
- Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. *J Rheumatol* 2005;32:832-40.
- Steen VD, Medsker TA Jr. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 1997;40:1984-91.