Published in final edited form as:

Arthritis Care Res (Hoboken). 2011 November ; 63(11): 1620–1628. doi:10.1002/acr.20591.

Validity of Two New Patient Reported Outcome Measures in Systemic Sclerosis: the PROMIS-29 Profile and the FACIT-Dyspnea

Monique Hinchcliff, MD MS¹, Jennifer L. Beaumont, MS², Krishna Thavarajah, MD MS³, John Varga, MD¹, Anh Chung, MS⁴, Sofia Podluszky, BA¹, Mary Carns, MS¹, Rowland W. Chang, MD, MPH^{1,4,5}, and David Cella, PhD²

¹Northwestern University Feinberg School of Medicine Division of Rheumatology

²Department of Medical Social Sciences

³Center for Lung Health, Henry Ford Hospital

⁴Department of Preventive Medicine

⁵Department of Physical Medicine and Rehabilitation

Abstract

Objective—Many patient-reported outcome (PRO) instruments used in systemic sclerosis (SSc) trials are limited by lack of validation, licensing fees, and complicated scoring systems. We assessed the construct validity for discriminative purposes of two new PRO instruments—the Patient Reported Outcomes Measurement Information System 29-item Health Profile (PROMIS-29) and the Functional Assessment of Chronic Illness Therapy-Dyspnea short form (FACIT-Dyspnea), measuring health status and dyspnea in SSc patients.

Methods—Seventy-three patients participated in a cross-sectional study at a tertiary SSc program. PROMIS-29, FACIT-Dyspnea, and legacy PRO instruments used in clinical trials (Medical Research Council [MRC] Dyspnea Score, St. George's Respiratory Questionnaire [SGRQ], Health Assessment Questionnaire-Disability Index [HAQ-DI] and Short-Form 36 [SF-36]) were administered. Composite severity scores using an adaptation of the Medsger Disease Severity Index were generated using clinical, diagnostic and laboratory information. PROMIS-29 and FACIT-Dyspnea scores were compared with legacy PRO measures and composite severity scores.

Results—The mean (range) patient age (84% women) was 51 years (22–72). The mean (range) SSc disease duration from the onset of the first non-Raynaud symptom was 7 years (0–45). Spearman correlation coefficients across FACIT-Dyspnea and PROMIS Physical Functioning scores with legacy PRO instruments were generally high (range=0.64–0.86); those between PROMIS and FACIT-Dyspnea with composite disease severity scores were more modest, but statistically significant (range=0.33–0.48, $p<0.01$).

Corresponding author: Monique Hinchcliff, MD MS, Northwestern University Feinberg School of Medicine, Division of Rheumatology, 240 E Huron Street, Suite M300, Chicago, IL 60611, Phone: 312-503-4844, Fax: 312-503-0994, m-hinchcliff@northwestern.edu.

Requests for Reprints: Monique Hinchcliff, MD MS, Northwestern University Feinberg School of Medicine, Division of Rheumatology, 240 E Huron Street, Suite M300, Chicago, IL 60611

Conflict of Interest: None of the authors have received any financial support or other benefits from commercial sources for the work reported on in the manuscript, and none have any other financial interests that could create a potential conflict of interest or the appearance of a conflict of interest with regard to the work.

Conclusion—PROMIS-29 and FACIT-Dyspnea are valid instruments to measure the health status of SSc patients. PROMIS-29 and FACIT-Dyspnea may be preferable to legacy instruments because they are freely available in multiple languages, and simple to administer, score and interpret.

Systemic sclerosis (SSc)/scleroderma is a rare, chronic connective tissue disease that causes skin and internal organ fibrosis, production of autoantibodies and deregulation of vascular homeostasis. Given the systemic nature of the disease, SSc can greatly impact quality of life(1). Accepted SSc clinical trial outcomes include assessment of functional, laboratory and radiographic disease markers, and patient reported outcomes (PRO). A PRO instrument is a patient-completed questionnaire that assesses a symptom or functional limitation.

A multitude of PRO instruments including the St. George's Respiratory Questionnaire (SGRQ)(2), Short-Form-36 (SF-36)(3), Health Assessment Questionnaire-Disability Index (HAQ-DI)(3, 4), Medical Research Council (MRC) Dyspnea Score, and others have been developed, validated in SSc (as indicated) and used to assess treatment response in SSc clinical trials. However, the lack of a standardized and uniformly scored set of PRO instruments to measure various SSc manifestations complicates between-study comparisons and interpretation of PRO results. In order to address this impediment to SSc and other disease-focused research, the National Institutes of Health (NIH) established the Patient Reported Outcomes Measurement Information System/PROMIS (www.nihpromis.org). The PROMIS is a network of NIH-funded research sites and coordinating centers working collaboratively to develop a series of dynamic tools to reliably and validly measure PROs(5).

In 2004, PROMIS investigators developed a plan to create a bank of PRO items and short forms that could be used by all investigators conducting patient-oriented research. Multidimensional scales including PROMIS-29 Health Profile (PROMIS-29), and symptom-specific instruments such as the Functional Assessment of Chronic Illness Therapy (FACIT)-Dyspnea short form were created in English and Spanish. PROMIS-29 and FACIT-Dyspnea were validated in the general US, and chronic obstructive pulmonary disease (COPD) populations respectively, and were standardized using a T-score metric with a mean set to 50 (standard deviation=10) to ensure simple scoring and interpretable results(6–8).

The purpose of the study was to assess the construct validity for discriminative purposes of the PROMIS-29 and FACIT-Dyspnea for the measurement of general health and dyspnea in SSc. We hypothesized that these new instruments to assess global health and dyspnea would perform as well as legacy instruments (SGRQ, MRC Dyspnea Score, HAQ-DI, SF-36) that have been used in SSc clinical trials. The benefit of the new instruments is that they are available in multiple languages (English and Spanish), free, and simple to administer, score and interpret.

Patients and Methods

All patients met American College of Rheumatology criteria for SSc or three out of five criteria for CREST (calcinosis cutis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly and telangiectasias)(9). Institutional Review Board consent was obtained from patients seen at a tertiary care scleroderma program to participate in the Northwestern Scleroderma Patient Registry and Biorepository. One optional component of the registry is the annual completion of PRO instruments. During regularly scheduled follow-up clinic visits, participants were given self-administered, paper-based, PRO instruments (PROMIS-29, FACIT-Dyspnea, SF-36, MRC, SGRQ and HAQ-DI) in the same order. Of

the 90 participants approached to complete the instruments between 12/2/2009 and 7/21/2010, 3 declined to participate, 7 did not complete the full battery, and 7 did not have a diagnosis of limited or diffuse SSc and were excluded from the analysis. Consequently, 73 subjects participated. One research assistant manually transcribed responses into a computerized database. A second research assistant reviewed a random sampling of 10% of entered PRO data to verify an error rate <0.5%.

All laboratory and diagnostic values were performed as clinically indicated. Laboratory values (brain natriuretic peptide, hemoglobin and creatinine) performed at one clinical laboratory obtained closest to the PRO instrument completion date were analyzed. Additional analyses using only laboratory data within 6 and 12 months of the PRO instrument completion date were also conducted. Pulmonary function test (PFT) parameters including forced vital capacity % predicted (FVC %) and diffusing capacity for carbon monoxide % predicted (DLCO %) were obtained by querying the Northwestern Medicine Enterprise Data Warehouse (EDW). The EDW is an electronic data repository that is updated nightly and contains all the clinical information collected across the Northwestern campus via a variety of clinical reporting tools including the electronic medical record. Estimated pulmonary artery systolic pressure on 2-D Doppler echocardiography performed for clinically indicated reasons was obtained by manual chart abstraction because echocardiography reports are less amenable to electronic data capture via the EDW. The modified Rodnan skin score (mRSS) was measured by an experienced scleroderma clinical specialist(10).

PROMIS and FACIT Patient-Reported Instruments

PROMIS investigators framed the overall PRO measurement model according to the classic World Health Organization definition of health as “a complete state of physical, mental and social health; not merely the absence of disease or infirmity”. Using confirmatory factor analysis to establish the unidimensionality of components of self-reported health, followed by item response theory models to calibrate items within banks, a hierarchical framework was populated around these three broad health components: Physical, Mental and Social. Each component was hypothesized, and confirmed, to have discrete, “bankable” symptoms and functional impacts (in the case of physical health), affective, behavioral and cognitive effects (in the case of mental health) and relationships and function (in the case of social health).

The PROMIS network developed item (question) banks and short forms in more than 20 health domains as well as a set of global health items and 29-, 43-, and 57-item profile measures (www.nihpromis.org)(8). To create a brief, practical-yet-inclusive short profile, a consensus-building process was used to identify 7 of these 20 domains to produce the PROMIS-29 that was used in this study(11, 12). The seven domains specifically relate to physical, mental and social health and cover the most relevant areas of self-reported health for the greatest majority of people with chronic illness: pain, fatigue, depression, anxiety, sleep, physical function and sexual function. The PROMIS-29 includes four items each from these seven core PROMIS domains as well as one 11-point rating scale for pain intensity. Norm-based scores have been calculated for each domain, such that a score of 50 represents the mean of the general population (standard deviation=10). High scores represent more of the domain being measured. Thus, on symptom-oriented domains of PROMIS-29 (anxiety, depression, fatigue, pain interference, and sleep disturbance), higher scores represent worse symptomatology. On the function-oriented domains (physical functioning and social role) higher scores represent better functioning.

The Functional Assessment of Chronic Illness Therapy (FACIT)-Dyspnea short form was developed using item response theory methodology and input from COPD experts to select

the most informative and relevant subset of items for measuring dyspnea in a COPD patient population(13). Item response theory or latent trait theory is based on the idea that the probability of a certain response to an item is a mathematical function of patient and item parameters. Inclusion of items that assess dyspnea across a spectrum of severity permits accurate measurement of dyspnea for each patient. FACIT-Dyspnea has been demonstrated to be an internally consistent, reliable, and valid assessment of dyspnea for men and women with self-reported and a formal diagnosis of COPD(6, 7).

Respondents are presented with 10 common tasks to determine whether they performed the task and the severity of dyspnea when completing these tasks over the past seven days and, separately, they are asked to rate the level of difficulty in completing these tasks due to dyspnea. Patients who report not having done the task are asked whether it was due to shortness of breath or simply because an opportunity to do the task did not occur in the past week. Two scores are created: dyspnea and dyspnea-related functional limitation. Higher scores represent worse dyspnea or functional limitation. As with PROMIS, scores are distributed on a T-metric (mean=50; SD=10); however the reference population is people with self-reported COPD.

Legacy Patient Reported Outcome Instruments

The SF-36 consists of 36 items assessing physical functioning, social functioning, role limitation due to physical health, bodily pain, mental health, role limitations due to emotional health, vitality, and general health perceptions. Scores were calculated according to the manual, including the calculation of Physical and Mental Component Scores(14). On all SF-36 scales, lower scores (<50) represent worse outcome. QualityMetric® requires a yearly SF-36 licensing agreement and fee on a per study basis.

The SGRQ was designed to measure health impairment in patients with asthma and COPD. It has been validated in patients with SSc(2). Symptoms and Activity scores are calculated using item weights provided in the instrument with a high score indicating poor quality of life(15). The total score is calculated from 16 items and their respective weights. Special permission is required before administering the SGRQ.

The MRC Dyspnea score is a freely-available instrument to measure breathlessness(16). The MRC Dyspnea score grades severity of breathlessness on a 5-point scale where 0 indicates no trouble with breathlessness and 4 indicates very severe breathlessness. The MRC Dyspnea score has been validated in patients with COPD and idiopathic pulmonary fibrosis, but not in SSc patients(17, 18).

The HAQ-DI was developed to assess disease severity in patients with rheumatoid arthritis and includes 2–3 questions for 8 activity domains (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities of daily living), and a pain visual analog scale (VAS)(19, 20). A mean score for each domain and a composite score are calculated using an ordinal scale from 0 (no impairment) to 3 (maximal functional impairment). The pain VAS is anchored at 0 (no pain/limitation) and 100 (very severe pain/limitation) with 1cm=0.2 points on a 0–3 ordinal scale, but is not factored into the overall disability index score.

The Medsger Disease Severity Scale is widely used in clinical trials. It assesses the impact of SSc upon 9 organ systems (general health, peripheral vascular, skin, joint/tendon, muscle, gastrointestinal tract, lungs, heart and kidneys)(21, 22). Each organ system is scored on a zero (no involvement) to four (severe involvement) scale although no composite score is calculated. The Medsger Disease Severity Score has not been validated in patients with SSc(23).

Study Design

We conducted a cross-sectional study using a cohort of well-described patients with SSc to assess the construct validity for discriminative purposes of two new PRO instruments. Construct validity according to COSMIN terminology is defined as “the degree to which the scores of a health related-PRO (HR-PRO) instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores or other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validity measures the construct to be measured”(24). We hypothesized that PROMIS-29 and FACIT-Dyspnea would demonstrate construct validity because these instruments were designed to cover as full a range of the latent variable being measured by each bank (e.g., pain, fatigue, etc.).

Data Analysis

To generate a composite SSc disease severity score, a modified Medsger Disease Severity Scale was calculated using available laboratory values (brain natriuretic peptide, hemoglobin and creatinine), physical exam findings (modified Rodnan skin score) and clinical markers of disease (PFT and echocardiography) in our cohort. Each variable was first classified into levels of severity and scores of 0, 1, or 2 assigned to each level:

- modified Rodnan skin score values of 0–6 were recoded to 0, 7–14 were coded as 1, and value 15 and higher were coded as 2
- Diffusing capacity for carbon monoxide percent predicted of 80% or greater was coded 0, 71–79% was coded 1 and less than or equal to 70% was coded 2
- Estimated right ventricular systolic pressure on 2-dimensional Doppler echocardiography values of less than or equal to 50mm Hg were coded 0 and values greater than 50mm Hg were coded 1
- Brain natriuretic peptide values of 60 pg/ml or less were coded as 0, 61–99 pg/ml were coded as 1, and values of 100 pg/ml or greater were coded 2
- Hemoglobin values of 12 mg/dl or greater were coded 0, 10.6–11.9 mg/dl were coded 1 and values of 10.5 or less were coded 2
- Creatinine levels of 1.59 mg/dl or less were coded as 0, 1.60–2.99 mg/dl were coded 1 and levels of 3.00 mg/dl or greater were coded 2.

Individual item scores were summed to create the composite SSc disease severity score. This composite score was used in the analyses described below.

We selected endpoints for the validation analyses that we anticipated would be most related to the legacy measures. Construct validity for the new instruments was assessed using Spearman correlation coefficients between legacy measures and FACIT-Dyspnea and PROMIS-29 scores, and by comparing mean FACIT-Dyspnea and PROMIS-29 scores across groups defined by MRC Dyspnea Score and by the composite measure of disease severity. Analysis of variance (ANOVA) was used to test the differences in scores between groups. Effect sizes (mean difference/pooled standard deviation) were calculated for the differences between adjacent groups to aid in interpretation of differences. All analyses were conducted using SAS 9.2 (SAS Institute Inc., Cary, NC).

Results

The battery of two new and four legacy PRO measures were administered to 73 patients with SSc. Descriptive statistics for patient characteristics and laboratory results performed closest to the date of PRO instrument completion are described in Table 1. There were no

significant differences in the analyses when all, or only laboratory data within 6 and 12 months of PRO instrument completion date, were used (data not shown). Over three-quarters of the participants were women ($n=61$, 83.6%) and 54.8% had limited disease subtype. Mean (range) disease duration was 7.2y (0–45) if defined as the time since the onset of the first non-Raynaud symptom and 8.0y (0–36) if defined as the interval since the appearance of Raynaud phenomenon. Only 3 patients (4.1%) had renal insufficiency.

Descriptive statistics for all PRO measures are listed in Table 2. Because the new PRO instruments are specifically designed to use a T-score metric, the mean of each domain is approximately 50 (range 40.5–55.1). This contrasts with the legacy PRO instruments that use various scoring systems as reflected in the wide distribution of means (range 0.83–78.2). Figure 1 depicts PROMIS-29 domain scores for the scleroderma cohort relative to the US population. For some domains such as physical functioning and satisfaction with social role participation a high score (>50) is favorable. Our SSc cohort reported poorer physical functioning (mean 46.66) and satisfaction with social role participation (mean 48.36) than the general population. For other domains such as fatigue, pain interference and sleep disturbance a high score indicates more decrements in health-related quality of life. Scleroderma patients reported more anxiety (mean 50.28), fatigue (mean 51.75), pain interference (mean 55.06) and sleep disturbances (mean 52.05) though the degree of impairment was less than the minimal clinically important difference (0.5 SD, or 5 points) for all domains except pain interference(25). Both FACIT-Dyspnea scores were approximately one full standard deviation better than the average for individuals with COPD (Table 2). Thus, our cohort of patients with scleroderma reported less dyspnea than patients with COPD.

To assess the construct validity of PROMIS-29 and FACIT-Dyspnea compared to the legacy instruments in measuring general health status and dyspnea, we calculated Spearman correlation coefficients to examine the correlation between PRO instrument scores that assessed similar domains. FACIT-Dyspnea and FACIT-Dyspnea Functional Limitations scores were compared to SGRQ, MRC Dyspnea score and HAQ-DI while PROMIS-29 Physical Functioning was compared to HAQ-DI and the SF-36 Physical Component Score (Table 3). We prospectively defined an $r \leq 0.3$ as a low correlation, $0.3 \leq r \leq 0.5$ as moderate correlation, and $r > 0.5$ as high correlation(26). Correlation coefficients ≥ 0.50 were considered good evidence for construct validity. As expected, FACIT-Dyspnea and FACIT-Dyspnea Functional Limitations scores strongly correlated with SGRQ, MRC Dyspnea score and HAQ-DI scores ($r=0.64$ – 0.82). There was a strong correlation between PROMIS-29 Physical Functioning and HAQ-DI and the SF-36 Physical Component scores ($r=0.82$, 0.86 respectively). Additionally there was a strong correlation between PROMIS-29 Anxiety and Depression and the SF-36 Mental Component Scores ($r=0.50$, 0.70 respectively). These data suggest that PROMIS-29 and FACIT-Dyspnea are good alternatives to existing measures of disease burden.

To further evaluate construct validity of the new versus legacy PRO instruments, we assessed the correlation between new and legacy instruments and the modified Medsger Disease Severity Index using Spearman correlation coefficients as above. PROMIS-29 Physical Functioning, Pain Interference and Satisfaction with Social Role Participation domains were moderately correlated with the SSc Disease Severity Index ($r=0.37$ – 0.48). A moderate degree of correlation was also seen between both FACIT-Dyspnea scores and the disease severity index ($r=0.33$, 0.43 , for dyspnea and functional limitations, respectively). These moderate correlations suggest that physician and patient-reported assessments of disease severity often differ.

As summarized in Table 4, the FACIT-Dyspnea scores and PROMIS-29 Physical Functioning domain showed differences in means on the basis of MRC Dyspnea groups with the poorest mean score associated with the most severe MRC Dyspnea score category of self-reported health. All effect sizes for the differences between adjacent groups were >0.30 . The FACIT-Dyspnea scores, PROMIS-29 Physical Functioning, and Satisfaction with Social Role Participation scores showed differences in means on the basis of groups defined by the SSc Disease Severity Index, with most effect sizes > 0.30 . The range of effect sizes for group differences was similar for the legacy measures (SF-36, SGRQ and HAQ-DI). These observations suggest that new and legacy PRO instruments are comparable to one another in terms of their ability to detect meaningful differences between clinically distinct groups of SSc patients.

Discussion

Using a cohort of SSc patients, we demonstrated the construct validity for discriminative purposes of two new PRO instruments, the PROMIS-29 health profile and FACIT-Dyspnea. PROMIS-29 and FACIT-Dyspnea will likely be more efficient and meaningful PRO instruments compared to legacy instruments in clinical SSc trials because they are freely available, simple to administer, score and interpret, available in multiple languages, and have been validated in many diseases, now including SSc.

We found moderate to high degrees of correlation between PROMIS-29 domains and FACIT-Dyspnea and legacy instruments. All PROMIS-29 domains were moderately to highly correlated with the SF-36 Physical Component Score, SGRQ and the HAQ-DI that have been validated in patients with SSc. Correlation between PROMIS-29 domains and MRC Dyspnea score was moderate. The FACIT-Dyspnea was highly correlated with the SGRQ total score, MRC Dyspnea score, HAQ-DI, and SF-36 Physical Component Score as expected. These results suggest PROMIS-29 and FACIT-Dyspnea are acceptable alternatives to the SGRQ, HAQ-DI, SF-36 and MRC Dyspnea score in clinical studies.

To further evaluate the construct validity of PROMIS-29 and FACIT-Dyspnea in SSc, we examined the correlation between these new PRO instruments and the modified Medsger Disease Severity Index. Some of the PROMIS-29 domains (physical functioning, pain interference, and satisfaction with social role participation) and the FACIT-Dyspnea were moderately correlated with the modified Medsger Disease Severity Index. We did not expect a high degree of correlation between some of the PROMIS-29 domains such as anxiety, depression, sleep disturbance and satisfaction with social role performance and the disease severity index.

The primary reason we did not observe a strong correlation between the severity scale and FACIT-Dyspnea may stem from the overall good health of our cohort. Our finding that the date of laboratory testing (creatinine, brain natriuretic peptide and hemoglobin) in relation to PRO instrument completion had no effect on the modified Medsger Disease Severity Index score suggests that our cohort was relatively well: patients with stable SSc infrequently undergo laboratory testing. Additionally, patients with severe skin disease (a high mRSS) might experience dyspnea secondary to a restrictive lung defect, and patients with low DLCO and hemoglobin, and high brain natriuretic peptide and estimated RVSP on echocardiography (secondary to interstitial lung disease/pulmonary hypertension or pulmonary arterial hypertension) would likely endorse dyspnea. Although five of the six items in the Index were directly or indirectly related to dyspnea, the mean severity index in our cohort was 2.79 (potential range of 0–11, observed range of 0–9). This suggests that patients were relatively healthy as assessed by our metric.

The FACIT-Dyspnea and PROMIS-29 were able to successfully differentiate between groups defined by MRC Dyspnea score category or the SSc Disease Severity Index, with most effect sizes >0.30 . Effect sizes >0.2 are likely to be clinically important(27, 28). This is an indication of the construct validity for discriminative purposes of these new instruments in patients with SSc.

One important advantage of using PROMIS instruments such as the PROMIS-29 Health Profile is the ability to compare the results in a study cohort to the general US population. SSc patients reported poorer physical functioning and satisfaction with social role participation compared with the general US population(25). Levels of anxiety and depression were similar while fatigue, pain interference and sleep disturbances were higher compared with the age-adjusted general US population(25). Increased sleep disturbances in our cohort compared to the general US population corroborates recently published study results that found sleep disturbances in many SSc patients that were associated with greater dyspnea, depressed mood, and severity of reflux symptoms(29).

Because lung disease (pulmonary artery hypertension and interstitial lung disease) is the leading cause of death in SSc patients and can cause dyspnea, we compared PROMIS domain scores in our SSc patients to COPD patients from a prior study(25). Our SSc cohort reported better fatigue, anxiety, depression, physical function and pain scores than COPD patients(25). FACIT-Dyspnea scores were substantially better (>1 SD which is two times the minimal clinically important difference) than those of COPD patients(6, 7). The ability to compare results of new PRO instruments across diseases is a major strength of the PROMIS items and short forms, and will likely provide meaningful insights into disease burden for a broad range of conditions.

Limitations of this study include the inability to calculate a complete Medsger Disease Severity Index for study subjects because of a lack of information collected from our registry participants (i.e. tendon friction rubs, joint contractures, digital ulcers). Also, as stated previously, our cohort was relatively healthy with a low mean score on the modified Medsger Disease Severity Index. Administration of PROMIS-29 and FACIT-Dyspnea to a cohort of SSc patients with low, moderate and high disease severity to further determine the sensitivity of PROMIS-29 and FACIT-Dyspnea in SSc is needed.

Future studies should also include longitudinal analyses to assess the sensitivity of PROMIS and FACIT as indices of SSc disease activity. An important future research need is to determine if PROMIS-29 and FACIT-Dyspnea predict outcomes in SSc (e.g. development and/or progression of lung and skin disease, death,) as they do for cancer. In fact, studies have shown that PROs predict clinical outcomes (e.g. death) better than physician assessments in patients with cancer(30, 31).

The results of our study demonstrate the construct validity for discriminative purposes of PROMIS-29 and FACIT-Dyspnea as measures of general health status and dyspnea in SSc patients. The potential advantages of PROMIS-29 and FACIT-Dyspnea include simple scoring procedures using a T-scale metric, the ability to compare results to a variety of other populations (US general public and other chronic diseases), availability in multiple languages (soon to include Chinese), and lack of licensing costs. PROMIS-29 and FACIT-Dyspnea are attractive alternatives to legacy PRO instruments to measure general health status and dyspnea in patients with SSc.

Significance and Innovation

- First study in systemic sclerosis to assess the construct validity for discriminative purposes of two, new patient reported outcome instruments: the

Patient Reported Outcomes Measurement Information System (PROMIS)-29 Health Profile and the Functional Assessment of Chronic Illness Therapy-Dyspnea (FACIT-Dyspnea).

- Study results indicate that PROMIS-29 and FACIT-Dyspnea have construct validity for discriminative purposes compared to legacy PRO instruments and a disease severity measure, and can be used to assess general health status and dyspnea in SSc patients.
- PROMIS-29 profile and FACIT-Dyspnea may be superior to legacy instruments such as the Short Form-36 and the Health Assessment Questionnaire-Disability Index.

Acknowledgments

Grant Support: MH (Eunice Kennedy Shriver National Institute of Child Health and Human Development [HD055884 to M.H.]; and the Arthritis Foundation Chicago Chapter Grant), KT (Eleanor Wood Prince Grant Initiative from the Women's Board of Northwestern Memorial Hospital), JV (Collaborative Development Initiative, Northwestern Memorial Hospital), RWC (P60 AR48098 [MCRC] and UL1RR025741 [CTSA]), DC (U5AR057951-01 [PROMIS Statistical Center] and U01AR052177-06S1 [PROMIS supplement])

References

1. Bassel M, Hudson M, Taillefer SS, Schieir O, Baron M, Thombs BD. Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian National Survey. *Rheumatology (Oxford)*. 2010
2. Beretta L, Santaniello A, Lemos A, Masciocchi M, Scorza R. Validity of the Saint George's Respiratory Questionnaire in the evaluation of the health-related quality of life in patients with interstitial lung disease secondary to systemic sclerosis. *Rheumatology (Oxford)*. 2007; 46(2):296–301. [PubMed: 16877463]
3. Rannou F, Poiraudou S, Berezne A, Baubet T, Le-Guern V, Cabane J, et al. Assessing disability and quality of life in systemic sclerosis: construct validities of the Cochin Hand Function Scale, Health Assessment Questionnaire (HAQ), Systemic Sclerosis HAQ, and Medical Outcomes Study 36-Item Short Form Health Survey. *Arthritis Rheum*. 2007; 57(1):94–102. [PubMed: 17266096]
4. Steen VD, Medsger 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(11):1984–1991. [PubMed: 9365087]
5. Cella D, Kallich J, McDermott A, Xu X. The longitudinal relationship of hemoglobin, fatigue and quality of life in anemic cancer patients: results from five randomized clinical trials. *Ann Oncol*. 2004; 15(6):979–986. [PubMed: 15151958]
6. 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(2):307–315. [PubMed: 21402298]
7. 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(2):291–306. [PubMed: 21402297]
8. 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(11):1179–1194. [PubMed: 20685078]
9. 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(5):581–590. [PubMed: 7378088]

10. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr. et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol.* 1988; 15(2):202–205. [PubMed: 3361530]
11. Cella D, Beaumont J, Liu Z, Phan A, Choi S, Yao J. Neuroendocrine tumor disease state and PROMIS health-related quality of life scores. *Journal of Clinical Oncology.* 2010; 28:7s.
12. Beaumont J, Liu Z, Choi S, Yao J, Phan A, Cella D. Relationship between neuroendocrine tumor-related symptoms and PROMIS health-related quality of life scores. *Journal of Clinical Oncology.* 2010; 28:7s.
13. 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(6):1018–1025. [PubMed: 19402847]
14. Ware, JE.; Kosinski, M.; Dewey, JE. How to Score Version 2 of the SF-36 Health Survey (standard and Acute Forms). Lincoln, RI: QualityMetric, Incorporated, 640 George Washington Highway; 02865.
15. Jones, P.; Forde, Y. London SGsUo, editor. St. George's Respiratory Questionnaire Manual, version 2.2. 2008.
16. Stenton C. The MRC breathlessness scale. *Occup Med (Lond).* 2008; 58(3):226–227. [PubMed: 18441368]
17. Manali ED, Stathopoulos GT, Kollintza A, Kalomenidis I, Emili JM, Sotiropoulou C, et al. The Medical Research Council chronic dyspnea score predicts the survival of patients with idiopathic pulmonary fibrosis. *Respir Med.* 2008; 102(4):586–592. [PubMed: 18162388]
18. Papiris SA, Daniil ZD, Malagari K, Kapotsis GE, Sotiropoulou C, Milic-Emili J, et al. The Medical Research Council dyspnea scale in the estimation of disease severity in idiopathic pulmonary fibrosis. *Respir Med.* 2005; 99(6):755–761. [PubMed: 15878493]
19. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum.* 1980; 23(2):137–145. [PubMed: 7362664]
20. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol.* 1982; 9(5):789–793. [PubMed: 7175852]
21. Medsger 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(10):2159–2167. [PubMed: 10529133]
22. Hudson M, Thoms BD, Steele R, Panopalis P, Newton E, Baron M. Health-related quality of life in systemic sclerosis: a systematic review. *Arthritis Rheum.* 2009; 61(8):1112–1120. [PubMed: 19644906]
23. Hudson M, Steele R, Taillefer S, Baron M. Quality of life in systemic sclerosis: psychometric properties of the World Health Organization Disability Assessment Schedule II. *Arthritis Rheum.* 2008; 59(2):270–278. [PubMed: 18240187]
24. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010; 63(7):737–745. [PubMed: 20494804]
25. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, Cella D. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol.* 2010; 63(11):1195–1204. [PubMed: 20688471]
26. Cohen, J. Statistical power analysis for the behavioral sciences. Second ed.. Hillsdale: Lawrence Erlbaum Associates, Inc.; 1988.
27. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof.* 2005; 28(2):172–191. [PubMed: 15851772]
28. 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. *Journal of clinical epidemiology.* 2011; 64(5):507–516. [PubMed: 21447427]

29. Frech T, Hays RD, Maranian P, Clements PJ, Furst DE, Khanna D. Prevalence and correlates of sleep disturbance in systemic sclerosis--results from the UCLA scleroderma quality of life study. *Rheumatology (Oxford)*. 2011
30. Montazeri A. Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008. *Health Qual Life Outcomes*. 2009; 7:102. [PubMed: 20030832]
31. Eton DT, Fairclough DL, Cella D, Yount SE, Bonomi P, Johnson DH. Early change in patient-reported health during lung cancer chemotherapy predicts clinical outcomes beyond those predicted by baseline report: results from Eastern Cooperative Oncology Group Study 5592. *J Clin Oncol*. 2003; 21(8):1536–1543. [PubMed: 12697878]

Figure 1. PROMIS-29 Profile Function and Symptom Domain Scores

A high functional domain score is favorable (left). A low symptom domain score is favorable (right).

Table 1

Demographics, Laboratory and Clinical Characteristics of Disease Severity in Patients with SSc.

Characteristics	N=73
Clinical Information	
Age (years), mean, SD, range	51.1, 10.9, 22–72
Sex (% female)	61 (83.6%)
Year of onset of Raynaud, mean, SD, range	8.0, 7.5, 0–36
Year of onset of first non-Raynaud, mean, SD, range	9.2, 8.0, 0–45
SSc disease subtype (% lcSSc)	40 (54.8%)
mRSS, mean (SD)	9.2 (8.0)
Anti-topoisomerase (% positive)	22 (31.4%)
Anticentromere (% positive)	14 (22.2%)
Brain natriuretic peptide (%)	11 (17.5%)
Hemoglobin, mean (SD)	12.5 (1.5)
Renal insufficiency (crt >1.59)	3 (4.1%)
Diagnostic Information	
FVC % predicted, mean (SD)	78.5 (17.3)
DLCO % predicted, mean (SD)	65.6 (20.5)
Est RVSP, mean (SD)	35.9 (11.1)
Modified Medsger Disease Severity Index, (mean, SD, range)	2.79, 1.97, 0–9

SSc=systemic sclerosis, lcSSc=limited cutaneous SSc, mRSS=modified Rodnan Skin score, crt=creatinine kinase, FVC=forced vital capacity, DLCO=diffusing capacity for carbon monoxide, est RVSP=estimated right ventricular systolic pressure on 2-dimensional Doppler echocardiography.

Table 2

Mean and Median for FACIT-Dyspnea, PROMIS-29, and Legacy PRO Instruments in Patients with SSc.

	N	Mean (SD)	Median (range)
New Instruments			
FACIT-Dyspnea	72	40.5 (10.2)	39.2 (27.7 – 67.4)
FACIT-Dyspnea Functional Limitations	71	41.1 (11.1)	40.3 (29.6 – 68.8)
PROMIS-29 Physical Functioning	73	46.7 (9.2)	45.3 (26.8 – 56.9)
PROMIS-29 Anxiety	73	50.3 (9.4)	49.0 (40.3 – 78.0)
PROMIS-29 Depression	73	49.4 (8.7)	48.9 (41.0 – 71.4)
PROMIS-29 Fatigue	73	51.8 (12.0)	51.0 (33.7 – 75.8)
PROMIS-29 Pain Interference	73	55.1 (10.4)	55.7 (41.6 – 75.6)
PROMIS-29 Sleep Disturbance	73	52.0 (9.7)	52.8 (32.0 – 73.3)
PROMIS-29 Satisfaction with Social Role Participation	73	48.4 (11.4)	46.3 (29.0 – 64.1)
Legacy Instruments			
SGRQ Symptom Score	73	25.0 (22.4)	19.9 (0 – 83.6)
SGRQ Activity Score	70	36.6 (33.2)	35.3 (0 – 98.9)
SGRQ Impacts Score	70	13.0 (17.2)	5.6 (0 – 78.1)
SGRQ Total Score	69	22.2 (21.1)	16.6 (0 – 82.6)
HAQ-Disability Index	72	0.83 (0.82)	0.63 (0 – 2.63)
SF-36 General Health	73	49.2 (24.2)	50 (0 – 95)
SF-36 Physical Functioning	73	59.6 (29.4)	60 (0 – 100)
SF-36 Role Limitations-Physical	72	63.2 (32.1)	68.8 (0 – 100)
SF-36 Bodily Pain	73	61.3 (27.5)	62 (0 – 100)
SF-36 Mental Health	73	71.5 (20.9)	75 (5 – 100)
SF-36 Role Limitations-Emotional	73	78.2 (27.6)	91.7 (0 – 100)
SF-36 Social Functioning	73	73.0 (28.2)	75 (0–100)
SF-36 Vitality	73	52.4 (8.0)	50 (31.2 – 68.8)
MRC 0 (%)		29 (44.6)	
MRC 1 (%)		24 (36.9)	
MRC 2–4 (%)		12 (18.5)	

FACIT= Functional Assessment of Chronic Illness Therapy, PROMIS=Patient Reported Outcomes Measurement Information System, SGRQ=St. George's Respiratory Questionnaire, HAQ-DI=Health Assessment Questionnaire-Disability Index, SF-36=Short-Form-36, MRC=Medical Research Council Dyspnea Score.

Table 3

Spearman Correlation Coefficients Comparing FACIT-Dyspnea and PROMIS-29 Scales with Legacy PRO Instruments for the Measurement of Dyspnea and Physical Function in Patients with SSC (with 95% Confidence Intervals).

	SGRQ Total Score	MRC	HAQ-DI	SF-36 Physical Component Score	SF-36 Mental Component Score
FACIT-Dyspnea	0.82 (0.72, 0.88)	0.74 (0.60, 0.83)	0.64 (0.47, 0.76)	--	--
FACIT-Dyspnea Functional Limitations	0.78 (0.67, 0.86)	0.68 (0.52, 0.79)	0.76 (0.63, 0.84)	--	--
PROMIS-29 Physical Functioning	--	--	-0.82 (-0.88, -0.72)	0.86 (0.78, 0.91)	--
PROMIS-29 Anxiety	--	--	--	--	-0.50 (-0.66, -0.30)
PROMIS-29 Depression	--	--	--	--	-0.70 (-0.80, -0.55)

All $p < 0.01$.

FACIT= Functional Assessment of Chronic Illness Therapy, PROMIS=Patient Reported Outcomes Measurement Information System, SGRQ=St. George's Respiratory Questionnaire, HAQ-DI=Health Assessment Questionnaire-Disability Index, SF-36=Short-Form-36, MRC=Medical Research Council Dyspnea Score.

Table 4

Mean Scores and Effect Sizes of FACIT-Dyspnea and PROMIS-29 and the Legacy PRO Instruments by the Medical Research Council Dyspnea Score

	MRC Dyspnea Score Group	Mean (SD)	Effect sizes	p-value *
FACIT-Dyspnea	MRC 0 (n=28)	33.1 (6.2)		
	MRC 1 (n=24)	42.9 (5.5)	1.42 1.60	< 0.001
	MRC 2-4 (n=12)	53.9 (9.9)		
FACIT-Dyspnea Functional Limitations	MRC 0 (n=27)	34.0 (8.6)		
	MRC 1 (n=24)	42.0 (6.6)	0.93 1.47	<0.001
	MRC 2-4 (n=12)	54.6 (11.1)		
PROMIS Physical Funct.	MRC 0 (n=29)	51.5 (8.4)		
	MRC 1 (n=24)	45.2 (5.7)	-0.89 -1.36	<0.001
	MRC 2-4 (n=12)	35.6 (5.1)		
SGRQ Total Score	MRC 0 (n=27)	6.2 (9.5)		
	MRC 1 (n=22)	28.5 (13.0)	1.69 1.64	<0.001
	MRC 2-4 (n=12)	50.1 (18.5)		
HAQ-DI	MRC 0 (n=29)	0.5 (0.7)		
	MRC 1 (n=23)	0.9 (0.7)	0.59 1.02	<0.001
	MRC 2-4 (n=12)	1.6 (0.7)		
SF-36 Physical Component Score	MRC 0 (n=28)	47.4 (11.1)		
	MRC 1 (n=24)	37.0 (8.2)	-1.07 -0.91	<0.001
	MRC 2-4 (n=12)	28.0 (8.7)		

* ANOVA p-value.

FACIT= Functional Assessment of Chronic Illness Therapy, PROMIS=Patient Reported Outcomes Measurement Information System, SGRQ=St. George's Respiratory Questionnaire, HAQ-DI=Health Assessment Questionnaire-Disability Index, SF-36=Short-Form-36, MRC=Medical Research Council Dyspnea Score.

Table 5

Mean Scores and Effect Sizes of FACIT-Dyspnea and PROMIS-29 and the Legacy PRO Instruments by Disease Severity Using the Modified Medsger Disease Severity Index Score.

	SSc Severity Score Group	Mean (SD)	Effect sizes	p-value *
FACIT-Dyspnea	0–1 (n=25)	35.0 (7.6)		
	2–3 (n=20)	44.1 (9.8)	0.94 –0.13	0.003
	4+ (n=27)	42.8 (10.8)		
FACIT-Dyspnea Functional Limitations	0–1 (n=24)	35.3 (7.7)		
	2–3 (n=19)	42.1 (10.7)	0.65 0.31	0.003
	4+ (n=28)	45.3 (11.9)		
PROMIS Physical Funct.	0–1 (n=25)	52.4 (6.7)		
	2–3 (n=20)	45.6 (8.9)	–0.81 –0.41	<0.001
	4+ (n=28)	42.2 (9.0)		
PROMIS Social Role	0–1 (n=25)	53.9 (10.3)		
	2–3 (n=20)	48.1 (12.9)	–0.53 –0.42	0.004
	4+ (n=28)	43.6 (9.2)		
SGRQ Total Score	0–1 (n=22)	10.0 (15.3)		
	2–3 (n=19)	27.8 (23.6)	0.89 0.00	0.004
	4+ (n=28)	27.9 (19.9)		
HAQ-DI	0–1 (n=25)	0.3 (0.5)		
	2–3 (n=19)	0.8 (0.7)	0.62 0.72	<0.001
	4+ (n=28)	1.3 (0.9)		
SF-36 Physical Component Score	0–1 (n=25)	48.6 (8.6)		
	2–3 (n=19)	38.3 (12.3)	–0.99 –0.36	<0.001
	4+ (n=28)	34.5 (10.2)		

* ANOVA p-value.

Effect size=mean difference/pooled standard deviation. FACIT=Functional Assessment of Chronic Illness Therapy, PROMIS=Patient Reported Outcomes Measurement Information System, SGRQ=St. George's Respiratory Questionnaire, HAQ-DI=Health Assessment Questionnaire-Disability Index, SF-36=Short-Form-36, MRC=Medical Research Council Dyspnea Score.