Reliability, construct validity and responsiveness to change of the PROMIS-29 in systemic sclerosis-associated interstitial lung disease

C.J. Fisher[1], R. Namas[1], D. Seelman[1], S. Jaafar[1], K. Homer[1], H. Wilhalme[2], A. Young[1], V. Nagaraja[1], E.S. White[3], E. Schiopu[1], K. Flaherty[3], D. Khanna[1]

[1]University of Michigan Scleroderma Program, Division of Rheumatology Department of Internal Medicine, University of Michigan, Ann Arbor, MI; [2]Statistics Core, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; [3]Division of Pulmonary Medicine and Critical Care, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA.

Caitlyn J. Fisher
Rajaie Namas, MD*
*Daniela Seelman**, MD*
Sara Jaafar, MD
Kate Homer
Holly Wilhalme, MS
Amber Young, MD
Vivek Nagaraja, MD
Eric S. White, MD, MS
Elena Schiopu, MD
Kevin Flaherty, MD, MS
Dinesh Khanna, MD, MS

**Current address: Medical Subspecialties Institute, Division of Rheumatology, Cleveland Clinic, Abu Dhabi, UAE.*

***Current address: Department of Internal Medicine, Mayor University, Santiago 8580745, Chile.*

Please address correspondence to:
Dr Dinesh Khanna,
Division of Rheumatology,
Department of Internal Medicine,
University of Michigan,
Ann Arbor 48109, Michigan, USA.
E-mail: khannad@med.umich.edu

Received on October 16, 2018; accepted in revised form on January 28, 2019.

Clin Exp Rheumatol 2019; 37 (Suppl. 119): S49-S56.

Key words: validation, PROMIS-29, scleroderma, interstitial lung disease, health-related quality of life (HRQOL)

Competing interests: see page S-55.

ABSTRACT

Objectives. *PROMIS-29 is a generic health-related quality of life instrument. Our objective was to assess the reliability, construct validity, and responsiveness to change of PROMIS-29 in systemic sclerosis-associated interstitial lung disease (SSc-ILD).*

Methods. *Seventy-three participants with SSc-ILD were administered patient reported outcomes (PROs) at baseline and follow-up visits which included PROMIS-29 and other measures of generic health, dyspnea, and cough instruments. We assessed internal consistency reliability using Cronbach's α, an alpha of ≥ 0.70 was considered satisfactory. We assessed the responsiveness to change using linear regression models.*

Results. *Mean age of participants was 51.9 years and mean disease duration was 7.9 years after first non-Raynaud's symptom. Of the 73 participants, 56.2% were classified as diffuse SSc and 26% limited SSc. The baseline (mean ± SD) FVC % predicted was 73.9±15.5 with a DLCO % predicted of 57.7±21.1; 95.9% had fibrotic NSIP pattern on HRCT. PROMIS-29 scores were 0.2 to 0.9 SD below the US population. Cronbach's α reliability was acceptable for all domains (ranged from 0.77 to 0.98). All scales showed statistically significant correlations with hypothesised PROMIS-29 domains (p≤0.05 for all comparisons). PROMIS-29 showed none-to-small discriminatory ability in comparison with physiologic measures (FVC and DLCO). There was no significant relationship between the change in FVC versus the change in PROMIS-29 measures over time.*

Conclusion. *PROMIS-29 has adequate reliability and construct validity for evaluation in SSc-ILD. It has moderate-to-large correlations with other PROs. The PROMIS-29 domains were not found to change over time in this cohort, likely due to stable nature of the observational cohort.*

Introduction

Systemic sclerosis (SSc) is a chronic multisystem autoimmune disease characterised by vascular and immune dysfunction, leading to skin and internal organ fibrosis (1). Systemic sclerosis-associated interstitial lung disease (SSc-ILD) is responsible for up to 30% of mortality in patients with SSc (2-4). ILD may be found on high-resolution computed tomography (HRCT) in approximately 90% of patients with SSc (5) and 40-75% will exhibit changes in pulmonary function tests (PFTs) (6) at the time of diagnosis. The prognosis of SSc-ILD is highly variable, depending on the HRCT pattern and histologic subtypes; however, the presence of SSc-ILD has a great impact on health-related quality of life, which can be assessed using patient reported outcome measures (PROs) (7).

Patient Reported Outcome Measures Information System (PROMIS-29) is part of a National Institutes of Health (NIH) roadmap designed to improve the reporting and quantification of PROs (8). PROMIS-29 is a generic health-related quality of life (HRQOL) measure designed to assess 8 distinct clinically important domains: physical function, anxiety, depression, ability to participate in social roles, sleep disturbances, pain interference, pain intensity, and fatigue. In a single-centre study PROMIS-29 domains were found to have construct validity against legacy instruments in SSc (9). In another SSc cohort, PROMIS-29 domains had a moderate-to-high degree of correla-

tions *vs*. FACIT-Dyspnea and legacy instruments (10, 11).

We administered the PROMIS-29 and a battery of other PROs (that capture generic HRQOL, dyspnea, and cough) in an observational cohort of SSc-ILD. We sought to assess internal consistency and test-retest reliability, construct validity, and responsiveness to change of the PROMIS-29 in SSc-ILD.

Methods

In an ongoing prospective longitudinal cohort of SSc-ILD at the University of Michigan (UM), 73 participants completed PROMIS-29 at baseline and follow-up visits and formed the cohort for current analysis. All eligible patients were approached in the Scleroderma and CTD-ILD (connective tissue diseases-associated interstitial lung disease) clinics and invited to participate in the SSc-ILD cohort. All study participants signed an IRB-approved informed consent form, after verifying that they were 18 years of age or older, met the 2013 ACR/EULAR classification criteria for SSc (12), and had presence of ILD on HRCT defined by the presence of bilateral, subpleural, lower lobe predominant distribution of either: (i) reticular and/or ground glass opacity, with or without traction bronchiectasis, or (ii) honeycombing with the absence of a pattern that is predominantly nodular, cystic, peribroncho-vascular/central or upper lung predominant, mosaic attenuation, or consolidation. Patients were excluded from the study if FEV_1/FVC was less than 0.70, suggestive of significant obstructive lung disease.

Participants were administered PROs endorsed by the OMERACT CTD-ILD Working Group in 2015 (13). These included the SF-36, the Saint George Respiratory Questionnaire (SGRQ), the Leicester Cough Questionnaire (LCQ), a global visual analogue scale (VAS) for overall severity by the patient, the Dyspnea 12, and the Modified Medical Research Counsel Dyspnea Scale (MMRCDS). We also captured the PFT and HRCT during the initial visit at UM and analysed SSc serologies. For participants who consented at subsequent visits for the cohort, the serologies and HRCT data were captured based on chart review; PFTs were performed at every visit in the clinic and were generally obtained on the same day as PROs.

Outcome measures

Patient-reported outcomes

– Generic measures

PROMIS-29 version 2: Patient Reported Outcome Measures Information System (PROMIS) profile instruments are a collection of short forms used to assess the complete state of physical, social and mental health. PROMIS Profile instruments contain items from seven PROMIS domains; depression, anxiety, physical function, pain interference, fatigue, sleep disturbance, and ability to participate in social roles and activities. The PROMIS-29 Version 2.0 used in this study is a 29-item survey that assesses each of the seven domains with four questions each. In addition, PROMIS-29 version 2.0 includes numeric rating scale (NRS) for pain intensity.

For this study, response pattern scores were calculated through the online Assessment Center Scoring Service (http://www.healthmeasures.net/images/promis/manuals/PROMIS_Profile_Scoring_Manual.pdf).

The PROMIS-29 is standardised using a T-score metric where 50 represents the mean of the US general population (standard deviation = 10.) For each domain, a higher T-score represents greater magnitude of the trait being measured. For depression, anxiety, fatigue, pain interference, pain intensity, and sleep disturbance, a higher score represents higher impact of disease. For physical function and social role, higher scores represent better functioning.

SF-36 version 2 is a generic health status measure consisting of 36 items assessing 8 domains, summarised into Physical Component Summary (PCS) and Mental Component Summary (MCS) scores. The SF-36 has been tested in the general population, and the norm score (50) can be used in comparison with other populations. The SF-36 has been established as the 'gold standard' PRO and has been validated in multiple diseases (14, 15). Higher scores indicate better health. We have incorporated the standard version of the SF-36 V2 that has been evaluated in SSc-ILD (16).

Cough-specific measure

Leicester Cough Questionnaire is a disease-specific HRQOL with 19 patient-derived items relating to chronic cough, and has been previously validated in patients with ILD. Participants are directed to assign a value from a 7-point Likert scale, from 1 to 7. The LCQ is separated into three distinct domains: physical, psychological, and social. Higher scores indicate better health (scores range from 3-21) (17).

Dyspnea-specific measures

St. George's Respiratory Questionnaire is a 50-item tool originally used to assess the HRQOL in patients with chronic obstructive pulmonary disease. Scores are calculated from three domains: symptoms, activities, and impact. Scores range from 0-100, with 100 indicating more severe limitations due to disease. We have recently validated this in SSc-ILD (7).

Dyspnea 12 is a 12-item instrument assessing impact of dyspnea. Scores range from 0-36, with higher scores indicating more severe disease impact. Dyspnea-12 has shown adequate reliability and validity in SSc-ILD (18).

Modified medical research council dyspnea scale (MMRCDS) is a dyspnea-specific scale which assigns a grade of dyspnea severity from 1 to 5 based on level of activity or ability, in order to assess the extent of disability due to breathlessness (19).

Global measure

Patient global assessment for disease severity (20) was assessed using a single-item 100-mm visual analogue scale (VAS), scored 0–100, with higher scores representing greater disease severity.

Physiologic measure

Pulmonary function tests (PFTs)

PFTs were performed according to the recommendations of the American Thoracic Society (ATS)[21]. Values for forced vital capacity (FVC), total lung

capacity (TLC), forced expiratory volume in 1 second (FEV_1), and diffusing capacity of the lung for carbon monoxide (DLCO) were determined, reported as percent predicted, and compared with predicted values.

Autoantibodies

Participants had serologies as part of clinical care, including the immunofluorescence anti-nuclear antibodies (IF-ANA), extractable nuclear antigens including: anti-Smith antibody, anti-centromere antibody, anti-topoisomerase-1 ab, Anti-Ro ab, anti-RNP ab, and anti-RNA polymerase III done using multi-bead ELISA and confirmed by immunodiffusion, as needed.

Statistical analysis

The internal consistency reliability was assessed using Cronbach's α, with a score of ≥0.70 considered satisfactory (22). Intra-class correlations were calculated using linear mixed effects models with a random subject effect to assess test-retest reliability for a 15-patient convenience sample with repeated PROMIS-29 scores within 30 days. Pearson's correlation was used to assess the strength of association between PROMIS-29 domains and other measures. We hypothesised large correlation coefficients (≥0.37) (23) between PROMIS-29 domains and patient global assessment VAS. In addition, we hypothesised that domains measuring similar constructs will have large correlation coefficients. These include the (i) PROMIS-29 Physical Function domain with the SF-36 Physical Function domain, the SF-36 Role Function domain, the LCQ Physical domain, SGRQ Impacts and Activity domains, the Dyspnea 12, and MMRC; (ii) PROMIS-29 ability to participate in Social Roles domain with the SF-36 Social domain, the LCQ Social domain, and SGRQ Impacts domain; (iii) PROMIS-29 Anxiety and Depression domains with the SF-36 Role Function (emotional) domain, the LCQ Psychological domain, the SGRQ Impacts domain; (iv) the Dyspnea 12, PROMIS-29 Fatigue domain with the SF-36 Fatigue domain, the SGRQ Activity domain, Dyspnea 12, and MMRC; and (v) PROMIS-29 Pain Interference domain with the SF-36 Pain domain, and the Sleep Disturbance section will correlate with the Dyspnea 12.

We also assessed the PROMIS-29 domains' potential to differentiate mild restrictive lung disease *vs*. moderate-to-severe restrictive lung disease using the following indicators for severity: (i) FVC ≥ 70% *vs*. <70%, (ii) TLC ≥ 70% *vs*. <70%, (iii) DLCO < 57% *vs*. ≥ 57% (where 57% is the median value), and (iv) patient global severity ≤46 *vs*. >46 (where 46 is the median value). We used Student's *t*-test for continuous variables and the Chi-square test for categorical variables.

Change in PROMIS-29 outcomes were calculated as the difference between PROMIS-29 domains at baseline and at 6-, 12-, and 18-months. A linear mixed effects model with a random subject effect and fixed effects for the change in FVC (from baseline to 6-, 12-, and 18-months), the month, and the interaction between change in FVC and month was used to determine if there was a relationship between change in FVC and change in PROMIS-29, and if that relationship changed over time. Models were adjusted by baseline PROMIS-29 scores. *p*-values <0.05 were considered statistically significant for all analyses and no adjustment was made for multiple testing.

Results

Baseline characteristics

Of 73 participants, 59 (80.8%) were female, and 63 (86.3%) participants reported their race as White. The mean age (mean ± SD) of participants was 51.9±11.8 years, with mean disease duration of 7.9±8.3 after first non-Raynaud's symptom (Table I); 41 participants had disease duration of ≤7 years. Thirty-two (41%) participants had cough several/most days of the week based on SGRQ. The majority of participants (56.2%) were classified as diffuse cutaneous SSc, 26% as limited cutaneous, 8.2% overlap, and 4.1% sine scleroderma. The mean FVC% was 73.9%, mean DLCO was 57.7%, and HRCT patterns for ILD were predominantly fibrotic NSIP (95.9%) with 4.1% showing UIP pattern (Table I).

Table I. Baseline characteristics of participants.

	(n=73)
Age (years), mean (SD)	51.9 (11.8)
Female Sex, n (%)	59 (80.8)
Race, n (%)	
White 63	(86.3)
Black 6	(8.2)
Other 4	(5.5)
Ethnicity, n (%)	
Hispanic	7 (9.6)
Non-Hispanic	65 (89.0)
Unknown	1 (4.0)
Type of Systemic Sclerosis, n (%)	
Diffuse Cutaneous	41 (56.2)
Limited Cutaneous	19 (26.0)
Sine Scleroderma	3 (4.1)
Overlap with Scleroderma	6 (8.2)
Disease duration (years), mean (SD)	
After First Non-Raynaud's Symptoms	7.9 (8.3)
After First Raynaud's Symptoms	10.9 (11.0)
After ILD Diagnosis	4.7 (7.4)
mRSS, mean (SD)	9.7 (10.2)
Autoantibodies, n (%)	
Anti-Nuclear Antibody (ANA)	62 (84.9)
Anti-Smith (Anti-Sm)	12 (16.4)
Anti-centromere	3 (4.1)
Anti-topoisimerase-1	21 (28.8)
Anti-RNA polymerase 3	11 (15.1)
Anti-Ro	12 (16.4)
Anti-U1 ribonucleoprotein (RNP)	10 (15.2)
ILD Pattern on HRCT, n (%)	
NSIP	95.9 (70.0)
UIP	4.1 (3.0)
PFT values, %predicted, mean (SD)	
FVC %	73.9 (15.5)
TLC %	39.7 (52.8)
DLCO %	57.7 (21.1)

mRSS: Modified Rodnan skin score
ILD: Interstitial lung disease
NSIP: Non-specific interstitial pneumonia
UIP: Usual interstitial pneumonia
PFT: Pulmonary function test
FVC: Forced vital capacity
TLC: Total lung capacity
DLCO: Diffusing capacity of the lungs for carbon monoxide
HRCT: High resolution computed tomography

Descriptive statistics of PROs

Scores for PROMIS-29 domains were 0.1±SD (Depression) to 0.9±SD (Physical function) below that of the US general population. SF-36 showed similar findings, with the average score for each domain below the US general population 0.2 SD (Mental health) to 1.5 SD (Physical function). The LCQ showed a total average score of 17.5, indicating a mild level of cough in participants. The total SGRQ question-

naire had an average total score of 32.6 (Table II).

Internal consistency reliability

Cronbach's α for all PROMIS-29 domains was ≥0.70, with a minimum value of 0.91 (anxiety domain) and a maximum value of 0.98 (Pain Interference domain). Cronbach's α for the other questionnaires, SF-36, SGRQ, and LCQ, also showed values ≥0.70, suggesting acceptable reliability in SSc-ILD (Table II).

Test-retest reliability

Intra-class correlations ranged from 0.58 (pain) to 0.89 (depression) (Table IV). Test-retest shows consistency over time for most domains, with 0.70 reliability for anxiety, depression, fatigue, and sleep disturbance.

Construct validity

1. Hypothesised correlation coefficients between PROMIS-29 domains and PROs

All PROMIS-29 domains met our a priori hypotheses of correlation coefficient of ≥0.37, except for LCQ Psychological, with a correlation coefficient of 0.05 with PROMIS-29 Depression domain, and -0.07 with PROMIS-29 Anxiety domain. LCQ Social had a coefficient of 0.20 with the PROMIS-29 Social domain. (Table III). We explored if the poor coefficient correlations between the LCQ domains may be due to lack of cough in our cohort. We further analysed the data in participants (41% of the cohort) who reported cough on several/most days of the week. In this subgroup, we again did not find larger correlation coefficients between hypothesised LCQ domains and PROMIS domains (data not shown). Other correlation coefficients are shown in Tables III and IV.

2. Discriminative validity

PROMIS-29 domains were unable to discriminate between restrictive lung disease (p≥0.05 for all comparisons). More participants with mild/moderate restrictive lung disease with an FVC ≥70% (n=45), and TLC ≥70% (n=70) were observed in our cohort compared to severe restrictive lung disease FVC ≤70% (n=28). Thirty-four participants had a DLCO ≤ Median, and 57 participants had DLCO ≥ Median (Table V). PROMIS-29 Physical Function (p=0.004), Pain Interference (p=0.015), and Pain Intensity (p=0.001) domains were able to discriminate patient global assessment VAS at p<0.05.

3. Responsiveness to change

For all PROMIS-29 outcomes, there was no significant relationship between the change in FVC% and the change in PROMIS-29 domains over time. Additionally, this relationship was not significantly different depending on study month (interaction p-value >0.05 for all measures).

The results of the linear regression of PROMIS measures based on groupings of FVC% and DLCO% measures also indicated there was no significant difference in the 6-month change of any PROMIS-29 between FVC (assessed at 5% change), DLCO (assessed at 10% change), or TLC (assessed at 5% change). This is probably explained by our small sample size and the short follow up period. In addition, we had a largely stable cohort over time, with

Table II. Baseline scores and internal consistency reliability for study participants.

Evaluation	Scores % (n=73)	Cronbach's Alpha	Test-Retest
PROMIS-29 (n=73), mean (SD)			
Physical function[2]	41.4 (8.1)	0.92	0.65
Social role[2]	45.9 (8.2)	0.95	0.65
Anxiety[1]	52.5 (9.6)	0.91	0.85
Depression[1]	51.2 (11.0)	0.96	0.89
Fatigue[1]	56.4 (10.4)	0.94	0.87
Pain Interference[1]	55.9 (11.0)	0.98	0.58
Sleep Disturbance[1]	52.9 (11.0)	0.92	0.8
Pain[1]	3.5 (2.7)	NA	ND
SF-36 (n=61), mean (SD)			
PF (Physical function)	35.2 (12.5)	0.93	ND
MH (Mental health)	47.9 (10.6)	0.82	ND
Role Function (Physical)	37.4 (12.4)	0.96	ND
Role Function (Emotional)	41.9 (13.5)	0.95	ND
Social	41.4 (11.6)	0.87	ND
Energy/fatigue	41.1 (11.7)	0.83	ND
Pain	43.4 (12.8)	0.93	ND
General health	36.3 (12.3)	0.89	ND
SF-36 PCS, Mean± SD	35.9 (12.7)	NA	ND
SF-36 MCS, Mean± SD	46.6 (11.3)	NA	ND
Leicester Cough Questionnaire (n=70), mean (SD)			
Physical	5.6 (1.0)	0.78	ND
Psychological	6.0 (1.0)	0.77	ND
Social	5.9 (1.1)	0.91	ND
LCQ total	17.5 (3.1)	0.95	ND
SGRQ (n=69), mean (SD)			
Symptom score	30.4 (20.8)	0.79	ND
Activity score	52.1 (27.3)	0.91	ND
Impact score	22.1 (17.8)	0.87	ND
Total score	32.6 (19.0)	0.94	ND
Dyspnea 12 (n=70), mean (SD)	8.2 (9.3)	0.97	ND
MMRCDS (n=62), mean (SD)	2.2 (0.9)	NA	ND
VAS patient global assessment (n=61), mean (SD)	48.6 (28.2)	NA	ND

*Large correlation coefficient (≥0.37).
PROMIS-29: Patient Reported Outcome Measures Information System, scored from 0-100.
Pain: Intensity scale from 0-10.
SF-36: Short Form 36, scored from 0-100.
LCQ: Leicester Cough Questionnaire, scored from 3-21.
SGRQ: Saint George Respiratory Questionnaire, scored from 0-100.
Dyspnea 12: Scored from 0-36.
MMRCDS: Modified Medical Research Counsel Dyspnea Scale, scored from 1-5.
VAS: Patient Global Assessment for Disease Severity, Visual Analogue Scale.

mean (SD) FVC% change in this cohort being 0.64% (5.38%), -1.28% (7.30%) and 1.16% (5.92%) at 6 months, 12 months, and 18 months respectively.

Discussion

SSc-ILD is one of leading causes of mortality in patients with SSc. Current management of SSc-ILD includes aggressive screening and treatment with immunosuppressive agents (24, 25). Although clinicians are focused on physiological and radiological measures to make a decision on severity of the underlying ILD, it is also important to capture how the patient *feels and functions* (26) as PROs can complement objective measures and aid in treatment decisions.

In this single-centre cohort of SSc-ILD, we demonstrated that the PROMIS-29 has acceptable reliability and construct validity in SSc-ILD. We were not able to show that PROMIS-29 domains are responsive to change in SSc-ILD, largely due to stable nature of the cohort. In addition, we confirmed the detrimental impact of SSc-ILD on HRQOL.

Our study supports the findings reported in previous studies that PROMIS-29 have construct validity in SSc (10, 11). Hinchcliff *et al*. (10) assessed PROMIS-29 in SSc, though without distinguishing participants with or without ILD. We assume that some of the participants enrolled in that study had ILD, given the mean FVC % predicated was 78.5 and DLCO % predicated was 65.6. The baseline demographics in Hinchcliff *et al*. and our study had similar mean age groups (51.9 years *vs*. 51.1 years) (current cohort *vs*. Hinchcliff), but more females enrolled in our study (81% *vs*. 61%), as well as participants with dcSSc (56.2% *vs*. 45.2%), similar disease onset from first non-Raynaud's symptom (7.9 *vs*. 7.2). Higher scores among PROMIS-29 domains were observed in our study when compared to Hinchcliff *et al*. in the following domains (current cohort *vs*. Hinchcliff): anxiety: 52.5 *vs*. 50.3; depression: 51.2 *vs*. 49.4; fatigue: 56.4 *vs*. 51.8; sleep disturbance: 52.9 *vs*. 52.0; and pain interference: 55.9 *vs*. 55.1. Lower scores compared to Hinchcliff *et al*. were found in the following domains: physical function: 41.4 *vs*. 46.7; social role: 45.9 *vs*. 48.4.

Table III. Correlation coefficients for hypothesised domains of PROMIS-29 *vs*. other PROs.

PROMIS-29 Domains	Hypothesised	Correlation coefficients
Physical function	SF-36 Physical Function	0.89*
	SF-36 Role Function (Physical)	0.67*
	LCQ Physical	0.36*
	SGRQ Impacts	-0.68*
	SGRQ Activity	-0.84*
	Dyspnea 12	-0.57*
	MMRC	-0.70*
	VAS	-0.56*
Social role	SF-36 Social	0.75*
	LCQ Social	0.20
	SGRQ Impact	-0.61*
	VAS	-0.40*
Anxiety	SF-36 Role Function (Emotional)	-0.74*
	LCQ Psychological	-0.07
	SGRQ Impacts	0.54*
	Dyspnea 12	0.55*
	VAS	0.34
Depression	SF-36 Role Function (Emotional)	-0.70*
	LCQ Psychological	0.05
	SGRQ Impacts	0.38*
	Dyspnea 12	0.39*
	VAS	0.34*
Fatigue	SF-36 Energy/Fatigue	-0.83*
	SGRQ Activity	0.56*
	Dyspnea 12	0.44*
	MMRC	0.50*
	VAS	0.43*
Pain interference	SF-36 Pain	-0.90*
	VAS	0.51*
Pain intensity	SF-36 Pain	-0.92*
	VAS	0.62*
Sleep disturbance	Dyspnea 12	0.40*
	VAS	0.39*

*Large correlation coefficient (≥0.37).
PROMIS-29: Patient reported outcome measures information system, scored from 0-100.
Pain: Intensity scale from 0-10.
SF-36: Short form 36, scored from 0-100.
LCQ: Leicester cough questionnaire, scored from 3-21.
SGRQ: Saint George respiratory questionnaire, scored from 0-100.
Dyspnea 12: Scored from 0-36.
MMRCDS: Modified medical research counsel dyspnea scale, scored from 1-5.
VAS Patient Global Assessment for Disease Severity, Visual Analogue Scale.

In this study, all PROMIS-29 domains had large correlation coefficients (>0.37) with hypothesised PROs, except for the LCQ domains, and provides construct validity for the PROMIS-29 domains in SSc-ILD. There was a lack of large hypothesised correlations with the cough-specific instrument, the LCQ. We explored if the relationship between LCQ domains and participants who reported cough on majority of days had larger coefficients (41% of the cohort) but that was not the case. It is likely that cough is not a major component of impact on generic HRQOL in SSc-ILD (27).

The baseline PROMIS-29 domains were unable to discriminate between mild *vs*. moderate-to-severe pulmonary physiology (FVC%, TLC%, or DLCO%). In addition, longitudinal assessment of FVC% in this cohort failed to show an association with PROMIS-29 domains. We assume that this observation may

Table IV. Correlation coefficients between PROMIS-29 and other HRQOL measures.

Correlation Analysis Qol	SF-36 PF Scales	SF-36 MH	SF-36 RE-P	SF-36 RF-E	SF-36 E/F	SF-36 Pain	SF-36 GH	SF-36 Social	SF-36 PCS	SF-36 MCS	Dysp. 12	LCQ PH	LCQ PS	LCQ SC	LCQ Total	MMRC DS	SGRQ SS	SGRQ AS	SGRQ IS	SGRQ Total	FVC % predicted	DLCO % predicted	VAS
PROMIS-29 Physical Function	0.89*	0.33*	0.67*	0.42*	0.56*	0.60*	0.77*	0.56*	0.85*	0.27*	-0.57*	0.36*	0.21	0.26*	0.29*	-0.70*	-0.49*	-0.84*	-0.68*	-0.79*	0.15	0.19	-0.56*
PROMIS-29 Social Role	0.70*	0.46*	0.74*	0.60*	0.71*	0.66*	0.69*	0.75*	0.73*	0.55*	-0.53*	0.27*	0.15	0.2	0.22	-0.49*	-0.34*	-0.69*	-0.61*	-0.67*	0.14	0.05	-0.40*
PROMIS-29 Anxiety	-0.53*	-0.74*	-0.43*	-0.75*	-0.53*	-0.59*	-0.59*	-0.65*	-0.42*	-0.78*	0.55*	-0.14	-0.07	-0.06	-0.09	0.37*	0.27*	0.56*	0.54*	0.56*	-0.06	0.01	0.34*
PROMIS-29 Depression	-0.55*	-0.7*	-0.41*	-0.77*	-0.43*	-0.61*	-0.56*	-0.56*	-0.42*	-0.71*	0.39*	0.02	0.05	0.05	0.04	0.27*	0.1	0.5*	0.38*	0.42*	-0.10	-0.09	0.34*
PROMIS-29 Fatigue	-0.64*	-0.54*	-0.73*	-0.65*	-0.83*	-0.64*	-0.64*	-0.70*	-0.67*	-0.64*	0.44*	-0.22	-0.12	-0.14	-0.16	0.40*	0.28*	0.56*	0.49*	0.54*	0.18	0.04	0.43*
PROMIS-29 Pain Interference	-0.60*	-0.41*	-0.60*	-0.57*	-0.52*	-0.9*	-0.60*	-0.68*	-0.71*	-0.47*	0.42*	-0.18	-0.11	-0.16	-0.16	0.33*	0.29*	0.52*	0.42*	0.49*	0.00	0.09	0.51*
PROMIS-29 Sleep Disturbance	-0.40*	-0.63*	-0.46*	-0.51*	-0.55*	-0.57*	-0.52*	-0.62*	-0.44*	-0.62*	0.40*	-0.32*	-0.23	-0.23	-0.27*	0.26*	0.37*	0.35*	0.44*	0.44*	-0.11	-0.09	0.39*

*Large Correlation Coefficient (≥0.37).
PROMIS-29: Patient Reported Outcome Measures Information System, scored from 0-100.
SF-36: Short Form 36, scored from 0-100.
LCQ: Leicester Cough Questionnaire, scored from 3-21.
SGRQ: Saint George Respiratory Questionnaire, scored from 0-100.
Dysp. 12: Dyspnea 12, Scored from 0-36.
MMRCDS: Modified Medical Research Counsel Dyspnea Scale, scored from 1-5.
VAS: Patient Global Assessment for Disease Severity, Visual Analogue Scale.

Table V. Relationship between the PROMIS-29 Domains *vs*. PFTs and Patient Global Assessment on the Visual Analogue Scale.

PROMIS-29 Scores	All (N=73)	Disease duration≤7 (N=41)	FVC category			TLC category			DLCO category			VAS Category		
			FVC≤70 (N=28)	FVC>70 (N=45)	*p*-value	TLC≤70 (N=32)	TLC>70 (N=16)	*p*-value	DLCO≤57 (N=34)	DLCO>57 (N=31)	*p*-value	VAS≤46 (N=31)	VAS>46 (N=30)	*p*-value
Physical function, mean (SD)	41.4 (8.1)	41.4 (7.9)	40.2 (7.5)	42.2 (8.4)	0.31	42.4 (8.1)	43.3 (7.9)	0.722	41.2 (7.8)	41.6 (8.9)	0.851	44.1 (8.8)	38.2 (6.6)	0.004*
Anxiety, mean (SD)	52.5 (9.6)	52.9 (9.7)	53.9 (10.0)	51.6 (9.4)	0.313	49.8 (9.0)	53.4 (9.8)	0.219	51.8 (10.0)	53.6 (9.0)	0.441	52.7 (8.7)	54.0 (10.1)	0.584
Depression, Mean (SD)	51.2 (11.0)	52.0 (11.4)	52.9 (11.9)	50.1 (10.4)	0.303	49.6 (9.8)	50.7 (13.8)	0.741	50.5 (11.3)	52.4 (11.0)	0.503	50.7 (11.2)	53.5 (11.3)	0.342
Fatigue, mean (SD)	56.4 (10.4)	56.5 (9.9)	54.6 (9.3)	57.6 (10.9)	0.225	54.9 (8.7)	57.0 (11.5)	0.48	55.8 (8.7)	57.8 (11.5)	0.425	55.7 (11.1)	59.1 (8.5)	0.177
Pain interference, mean (SD)	55.9 (11.0)	55.4 (11.8)	55.5 (12.2)	56.2 (10.3)	0.813	53.9 (9.8)	57.3 (10.8)	0.282	53.9 (10.1)	58.2 (11.3)	0.113	53.2 (10.4)	59.9 (10.4)	0.015*
Sleep, mean (SD)	52.9 (9.7)	52.7 (8.3)	54.7 (8.3)	51.8 (10.4)	0.215	53.0 (8.8)	51.0 (9.3)	0.481	54.0 (8.3)	52.7 (9.2)	0.546	51.3 (9.1)	54.7 (7.6)	0.122
Social, mean (SD)	45.9 (8.2)	45.6 (8.6)	44.6 (8.1)	46.8 (8.3)	0.275	47.6 (6.9)	45.2 (8.9)	0.32	46.3 (7.1)	45.2 (8.8)	0.579	46.3 (8.5)	44.2 (7.0)	0.299
Pain intensity (SD)	3.5 (2.7)	3.4 (2.7)	3.3 (3.0)	3.6 (2.6)	0.629	2.9 (2.7)	3.6 (2.4)	0.365	2.9 (2.7)	4.2 (2.6)	0.05*	2.4 (2.1)	4.7 (2.8)	0.001*

*Statistically significant (*p*-value ≤0.05).
PROMIS-29: Patient reported outcome measures information system, scored from 0-100.
SF-36: Short Form 36, scored from 0-100.
LCQ: Leicester Cough Questionnaire, scored from 3-21.
SGRQ: Saint George Respiratory Questionnaire, scored from 0-100.
Dyspnea 12: Scored from 0-36.
MMRCDS: Modified Medical Research Counsel Dyspnea Scale, scored from 1-5.
VAS: Patient Global Assessment for Disease Severity, Visual Analogue Scale.

be due to a small sample size or due to the lack of an association between PROMIS-29 domains and physiological measures. In the Scleroderma Lung Study-I, a randomised controlled trial of daily oral cyclophosphamide *vs*. placebo, SF-36 Mental Component Summary domain was able to differentiate between mild *vs*. moderate-to-severe restrictive FVC% and physical summary score was able to differentiate between DLCO% categories ≤57 and >57 (16). More participants in our current cohort had mild impaired lung mechanics (as this was not a controlled trial that enriched for more severe disease), reflecting less clinical symptoms such as cough and shortness of breath. This made discrimination between the 2 groups more difficult.

Test-retest reliability also showed stability of the cohort, except in the PROMIS-29 domains Pain Interference, Physical Function, and Social Roles, which may indicate that higher levels of pain, which can be variable, affects physical and social aspects of patient HRQOL. This relationship may be related to a lack of reliability of the domain in ILD, and should be evaluated further in future studies. Like PROMIS-29 domains, SF-36 domains were unable to differentiate between

mild *vs*. moderate-to-severe pulmonary physiology (data not shown) and may be related to lack of sensitivity of generic HRQOL for milder disease. PROMIS-29 was able to differentiate the global severity of the disease by VAS with statistically significant changes in Physical Function and Pain Interference domains ($p<0.05$).

Our study has many strengths. First, this study reports from a prospective longitudinal cohort carefully designed to assess overall impact of SSc-ILD based on OMERACT CTD-ILD working group (13). Second, we show construct validity with hypothesised PROs that were endorsed in a consensus-based exercise with pulmonologists, rheumatologists, and patient partners for SSc-ILD. Third, this is the first study to examine PROMIS-29 along with generic and symptom-specific instruments and physiologic measures in SSc-ILD. The broad set of variables allowed us to assess construct validity of PROMIS-29 domains along with other patient-reported instruments and physiological measures in SSc-ILD.

However, our study is not without limitations. We did not account for multiple testing in our cohort. Our data should be considered hypothesis-generating, and needs to be confirmed in another cohort. Our goal when calculating the correlation coefficients was to examine direction and magnitude, and not to test statistical significance. Due to the single-centre cohort design, our analysis is limited by a small sample size. PROMIS-29 should be tested in a larger cohort, in order to account for non-significant findings that may have been due to the small sample size. Although University of Michigan thoracic radiologists read all HRCTs, we did not quantify the degree of total lung involvement for this analysis. Since this is an observational trial and participants are administered PROs during their visit, some missing data is inevitable.

In conclusion, the PROMIS-29 was found to have acceptable internal consistency reliability and construct validity in participants with SSc-ILD. PROMIS-29 was not able to discriminate between physiologic measures and was not responsive to change, likely related to the stable nature of the overall cohort or insensitivity of a generic PRO to capture milder changes over time. Further work is needed in ongoing clinical trials in SSc-ILD where PROMIS-29 has been incorporated as an outcome measure to assess the discriminative ability of PROMIS-29 domains in differentiating categories of restrictive lung disease and change over time with treatment in an enriched cohort.

Competing interests

D. Khanna has received consultancies from Acceleron, Actelion, Bayer, Blade Therapeutics, BMS, Cytori, Galapagos, Genentech/Roche, GSK, Mitsubishi Tanabi, Sanofi-Aventis/Genzyme, UCB Pharma, and owns stocks from Eicos Sciences, Inc. All other authors have declared no competing interests.

References

1. ORLANDI M, BARSOTTI S, LEPRI G *et al*.: One year in review 2018: systemic sclerosis. *Clin Exp Rheumatol* 2018; 36 (Suppl. 113): S3-23.
2. STEEN VD, MEDSGER TA: Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis* 2007; 66: 940-4.
3. KHANNA D, TASHKIN DP, DENTON CP, LUBELL MW, VAZQUEZ-MATEO C, WAX S: Ongoing clinical trials and treatment options for patients with systemic sclerosis-associated interstitial lung disease. *Rheumatology* 2018.
4. KHANNA D, SEIBOLD J, GOLDIN J, TASHKIN DP, FURST DE, WELLS A: Interstitial lung disease points to consider for clinical trials in systemic sclerosis. *Rheumatology* 2017; 56 (Suppl. 5): v27-v32.
5. SCHURAWITZKI H, STIGLBAUER R, GRANINGER W *et al*.: Interstitial lung disease in progressive systemic sclerosis: high-resolution CT versus radiography. *Radiology* 1990; 176: 755-9.
6. STEEN VD, CONTE C, OWENS GR, MEDSGER TA, JR.: Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum* 1994; 37: 1283-9.
7. WALLACE B, KAFAJA S, FURST DE *et al*.: Reliability, validity and responsiveness to change of the Saint George's Respiratory Questionnaire in early diffuse cutaneous systemic sclerosis. *Rheumatology* 2015; 54: 1369-79.
8. KWAKKENBOS L, THOMBS BD, KHANNA D *et al*.: Performance of the Patient-Reported Outcomes Measurement Information System-29 in scleroderma: a Scleroderma Patient-centered Intervention Network Cohort Study. *Rheumatology* 2017; 56: 1302-11.
9. KHANNA D, MARANIAN P, ROTHROCK N *et al*.: Feasibility and construct validity of PROMIS and "legacy" instruments in an academic scleroderma clinic. *Value Health* 2012; 15: 128-34.
10. HINCHCLIFF M, BEAUMONT JL, THAVARAJAH K *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.
11. HINCHCLIFF ME, BEAUMONT JL, CARNS MA *et al*.: Longitudinal evaluation of PROMIS-29 and FACIT-dyspnea short forms in systemic sclerosis. *J Rheumatol* 2015; 42: 64-72.
12. VAN DEN HOOGEN F, KHANNA D, FRANSEN J *et al*.: 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann Rheum Dis* 2013; 72: 1747-55.
13. KHANNA D, MITTOO S, AGGARWAL R *et al*.: Connective Tissue Disease-associated Interstitial Lung Diseases (CTD-ILD) - Report from OMERACT CTD-ILD Working Group. *J Rheumatol* 2015; 42: 2168-71.
14. MCHORNEY CA, WARE JE, JR., RACZEK AE: The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; 31: 247-63.
15. 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.
16. KHANNA D, YAN X, TASHKIN DP *et al*.; SCLERODERMA LUNG STUDY G: Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: results from the scleroderma lung study. *Arthritis Rheum* 2007; 56: 1676-84.
17. NAGATA K, TOMII K, OTSUKA K *et al*.: Evaluation of the chronic obstructive pulmonary disease assessment test for measurement of health-related quality of life in patients with interstitial lung disease. *Respirology* 2012; 17: 506-12.
18. SWIGRIS JJ, YORKE J, SPRUNGER DB *et al*.: Assessing dyspnea and its impact on patients with connective tissue disease-related interstitial lung disease. *Respir Med* 2010; 104: 1350-5.
19. BESTALL JC, PAUL EA, GARROD R, GARNHAM R, JONES PW, WEDZICHA JA: Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999; 54: 581-6.
20. 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: 1984-91.
21. PELLEGRINO R, VIEGI G, BRUSASCO V *et al*.: Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948-68.
22. M M: Coefficient alpha: a basic introduction from the perspectives of classical test theory and structural equation modeling. *Struct Equation Model* 1995; 2: 73.
23. COHEN J: Statistical Power Analysis for the Behavioral Sciences. 1st ed: Academic Press; 1977. 490 p.

24. DENTON CP, KHANNA D: Systemic sclerosis. *Lancet* 2017; 390: 1685-99.
25. KHANNA D, SERRANO J, BERROCAL VJ *et al*.: A randomized controlled trial to evaluate an internet-based self-management program in systemic sclerosis. *Arthritis Care Res* (Hoboken) 2019; 71: 435-47
26. POWERS JH, 3RD, PATRICK DL, WALTON MK *et al*.: Clinician-reported outcome assessments of treatment benefit: Report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. *Value Health* 2017; 20: 2-14.
27. CHENG JZ, WILCOX PG, GLASPOLE I *et al*.: Cough is less common and less severe in systemic sclerosis-associated interstitial lung disease compared to other fibrotic interstitial lung diseases. *Respirology* 2017; 22: 1592-7.