

UC Irvine

UC Irvine Previously Published Works

Title

Reliability, validity and responsiveness to change of the Saint George's Respiratory Questionnaire in early diffuse cutaneous systemic sclerosis

Permalink

<https://escholarship.org/uc/item/6gs4b41b>

Journal

Rheumatology, 54(8)

ISSN

1462-0324

Authors

Wallace, Beth
Kafaja, Suzanne
Furst, Daniel E
et al.

Publication Date

2015-08-01

DOI

10.1093/rheumatology/keu456

Peer reviewed

Original article

Reliability, validity and responsiveness to change of the Saint George's Respiratory Questionnaire in early diffuse cutaneous systemic sclerosis

Beth Wallace^{1,*}, Suzanne Kafaja^{2,*}, Daniel E. Furst², Veronica J. Berrocal³, Peter A. Merkel⁴, James R. Seibold⁵, Maureen D. Mayes⁶ and Dinesh Khanna⁷

Abstract

Objective. Dyspnoea is a common, multifactorial source of functional impairment among patients with dcSSc. Our objective was to assess the reliability, construct validity and responsiveness to change of the Saint George's Respiratory Questionnaire (SGRQ) in patients with early dcSSc participating in a multi-centre prospective study.

Methods. At enrolment and 1 year, patients completed the SGRQ (a multi-item instrument with four scales: symptoms, activity, impact and total), a visual analogue scale (VAS) for breathing and the HAQ Disability Index (HAQ-DI) and underwent 6 min walk distance and pulmonary function tests, physician and patient global health assessments and high-resolution CT (HRCT). We assessed internal consistency reliability using Cronbach's α . For validity we examined the ability of the SGRQ to differentiate the presence vs absence of interstitial lung disease (ILD) on HRCT or restrictive lung disease and evaluated the 1 year responsiveness to change using pulmonary function tests and patient- and physician-reported anchors. Correlation coefficients of 0.24–0.36 were considered moderate and >0.37 was considered large.

Results. A total of 177 patients were evaluated. Reliability was satisfactory for all SGRQ scales (0.70–0.93). All scales showed large correlations with the VAS for breathing and diffusing capacity of the lung for carbon monoxide in the overall cohort and in the subgroup with ILD. Three of the four scales in the overall cohort and the total scale in the ILD subgroup showed moderate to large correlation with the HAQ-DI and the predicted forced vital capacity ($r = 0.33$ – 0.44). Each scale discriminated between the presence and absence of ILD and restrictive lung disease ($P \leq 0.0001$ – 0.03). At follow-up, all scales were responsive to change using different anchors.

Conclusion. The SGRQ has acceptable reliability, construct validity and responsiveness to change for use in a dcSSc population and differentiates between patients with and without ILD.

Key words: scleroderma and related disorders, respiratory, quality of life, patient attitude to health, autoimmune-inflammatory conditions, systemic sclerosis, dyspnoea, patient-reported outcomes.

Rheumatology key messages

- The Saint George's Respiratory Questionnaire (SGRQ) has acceptable reliability, validity and responsiveness to change in early dcSSc patients with and without interstitial lung disease.
- SGRQ scales were able to differentiate patients with and without interstitial lung disease.
- SGRQ can be used in early dcSSc or interstitial lung disease trials.

¹Division of Rheumatology, University of Michigan, Ann Arbor, MI, ²Division of Rheumatology, UCLA, Los Angeles, CA, ³Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, ⁴Department of Internal Medicine, University of Pennsylvania, Philadelphia, PA, ⁵Scleroderma Research Consultants, Litchfield, CT, ⁶Division of Rheumatology, University of Texas Health Science Center, Houston, TX and ⁷Division of Rheumatology, Department of Internal Medicine, University of Michigan Scleroderma Program, Ann Arbor, MI, USA

Submitted 28 April 2014; revised version accepted 30 September 2014

Correspondence to: Dinesh Khanna, University of Michigan Scleroderma Program, Division of Rheumatology, Department of Internal Medicine, 300 North Ingalls Street, Suite 7C27, Ann Arbor, MI 48109, USA. E-mail: khannad@med.umich.edu

*Beth Wallace and Suzanne Kafaja contributed equally to this study.

Introduction

SSc is an uncommon CTD that can be associated with vasculopathy, skin thickening and fibrosis of the cardiopulmonary, renal and gastrointestinal systems [1]. dcSSc, the form of SSc involving proximal skin thickening and often severe multi-organ dysfunction, is associated with significant mortality and a negative impact on health-related quality of life (HRQoL) [2].

Dyspnoea is common and associated with functional impairment in patients with SSc [3, 4]. It is usually multifactorial, with major contributors being SSc-related disease such as interstitial lung disease (ILD), pulmonary arterial hypertension and left heart failure, as well as effects of chronic illness such as anaemia and deconditioning. Cardiopulmonary disease is the primary cause of SSc-related mortality [5], and SSc-ILD is an independent predictor of poor prognosis and an important cause of death in this population [6, 7].

The Saint George's Respiratory Questionnaire (SGRQ) was developed to evaluate HRQoL in patients with obstructive lung disease [8, 9]. It has been evaluated in other respiratory diseases including ILD [10], bronchiectasis [11] and chronic pulmonary aspergillosis [12]. Two small cross-sectional studies in SSc-ILD showed preliminary evidence of validity in this population [13].

In the current analysis, our objective was to examine the reliability, construct validity and responsiveness to change of the SGRQ in patients with early dcSSc and in patients with SSc-ILD. We used data from the Combined Response Index in Systemic Sclerosis (CRISS) cohort to assess internal consistency reliability and to correlate the four SGRQ scales with other HRQoL measures pertinent to the SSc population and with objective measures of pulmonary disease severity [high-resolution CT (HRCT) and pulmonary function tests (PFTs)] and cardiopulmonary functional status. We examined the ability of the SGRQ to discriminate the presence/absence of both ILD on HRCT and restrictive lung disease, and examined its responsiveness to change relative to patient- and physician-reported anchors and physiological parameters.

Methods

Patients

The CRISS database was launched to develop a composite index in dcSSc [14–16]. Eligible patients met criteria for early dcSSc, with early defined as ≤ 5 years from the first non-RP disease manifestation to their enrolment and dcSSc defined as skin thickening both proximal and distal to the elbows and/or knees, with or without face and neck involvement. The CRISS cohort has been shown to be representative of patients in large multicentre dcSSc randomized controlled trials (RCTs) [17].

A total of 200 patients were enrolled at four US scleroderma centres and the standardized set of CRISS outcome measures, including the measures below, was gathered at enrolment and at 1 year [14, 15]. One hundred and seventy-seven patients completed the SGRQ at

baseline and were included in our analysis. Ethical approval was obtained from the institutional review board of each study centre prior to any data collection and informed written consent was obtained from all patients at the time of their enrolment. De-identified data were used in this study.

Outcome measures

Dyspnoea measures

SGRQ. The SGRQ is a self-administered questionnaire for assessing HRQoL in respiratory diseases [8, 9]. It contains 50 items distributed over three scales. The symptom scale (SYM) assesses respiratory symptom severity. The activity scale (ACT) examines activity impairment due to respiratory symptoms. The impact scale (IMPACT) evaluates the effects of respiratory symptoms on overall function and well-being. Responses on each scale are weighted and a subscore ranging from 0 (no impairment) to 100 (worst impairment) is derived for each. A total score (TOTAL), also ranging from 0 to 100, is the weighted average of these three subscores. We used version 2.3 of the SGRQ, evaluating average impairment over the previous 12 months.

VAS for breathing. The VAS for breathing allows patients to self-assess the degree to which dyspnoea impairs daily activities. The scale ranges from 0 to 150, with higher values denoting worse impairment [18, 19].

Global health and HRQoL measures

HAQ-DI. The HAQ Disability Index (HAQ-DI) is a 20-item self-administered questionnaire assessing functional disability in eight domains. Scores range from 0.0 (best) to 3.0 (worst) [20]. It is fully validated in SSc [21, 22].

SF-36. The 36-item Short Form Health Survey (SF-36) is a self-administered survey focusing on functional status and HRQoL. It generates eight scores in different scales, as well as a physical component summary (PCS) and a mental component summary (MCS) [23, 24]. It is scored on a *T* score metric with a US population mean of 50 (s.d. 10); a higher score denotes better HRQoL. It has been previously validated for use in SSc [25].

Transitional disease question. Physicians and patients used a modified Likert scale ranging from 1 (much better) to 5 (much worse) to rate the patient's change in overall condition over the past year. Ratings of 1 or 2 were categorized as improved and ratings of 3–5 were categorized as unimproved.

Cardiopulmonary functional assessment

6 min walk distance (6MWD). This is the distance a patient can walk at his/her usual pace in 6 min. It correlates with other indicators of pulmonary disease severity in SSc patients with cardiopulmonary involvement [26] but is confounded by deconditioning and musculoskeletal involvement. The 6MWD was conducted and measured in accordance with American Thoracic Society (ATS) guidelines [27].

Physiological and radiographic evaluation

PFTs. PFTs were performed according to the recommendations of the ATS [28]. Values for forced vital capacity (FVC), total lung capacity (TLC), forced expiratory volume in 1 s (FEV1) and diffusing capacity of the lung for carbon monoxide (DLCO) were determined and compared with predicted values [29, 30].

HRCT. Thoracic HRCT has largely replaced surgical lung biopsy as a non-invasive means of identifying and characterizing SSc-ILD [31] and grading its severity and impact on mortality [32, 33]. HRCT was carried out as part of clinical care and was interpreted by local radiologists at the centres where care was received. Lung involvement attributable to SSc (ground glass opacity, interstitial markings and/or traction bronchiectasis, defined as SSc-ILD) was categorized as present or absent on each scan.

Statistical analysis

Summary statistics were calculated for all demographic and clinical variables. For continuous variables we reported mean (s.d.). For discrete categorical or binary variables we reported the number (%) of subjects having a certain characteristic.

To compare the distribution of SGRQ scores in the general population [34] vs the CRISS cohort, we computed weighted averages for the SYM, ACT, IMPACT and TOTAL scores for each population. In each case, average scores in each SGRQ scale were obtained in six different age and gender strata (males 40–49 years, males 50–59 years, males 60–69 years, females 40–49 years, females 50–59 years, females 60–69 years) for both populations and multiplied by the percentage of subjects in each strata of the CRISS cohort.

To assess the internal consistency reliability of the SGRQ scales, we computed Cronbach's α [35] for subjects in the CRISS cohort. A Cronbach's α of ≥ 0.70 is acceptable [36]. We also assessed the number and percentage of subjects achieving minimum and maximum scores in each subscale.

To determine whether there was an association between SGRQ scores and pulmonary function, HRQoL and dyspnoea at baseline, we calculated Pearson correlations among all four SGRQ scales as well as between each SGRQ scale vs VAS for breathing, HAQ-DI, 6MWD, percentage TLC predicted (TLC%), percentage FVC predicted (FVC%) and percentage DLCO predicted (DLCO%). For each pair of variables we tested whether correlations were significantly different from zero using a *t* test. Pearson correlation coefficients were interpreted as proposed by Cohen: 0.0–0.10, no correlation; 0.10–0.23, small correlation; 0.24–0.36, moderate correlation; ≥ 0.37 , large correlation [37].

To investigate whether SGRQ scales can discriminate between subjects with and without SSc-ILD and/or restrictive lung disease, we compared baseline SGRQ scales scores in subjects with (i) FVC% $\geq 70\%$ vs $<70\%$, (ii) TLC% $\geq 70\%$ vs $<70\%$ [38], (iii) DLCO% $\geq 66\%$ vs $<66\%$ (where 66% is the median value) and

(iv) SSc-related lung involvement on HRCT vs no such involvement.

We assessed the responsiveness to change in SGRQ scales relative to the transitional disease question asked at follow-up (see Global health and HRQoL above) as well as the change in VAS breathing, relative change in FVC%, relative change in TLC% and relative change in DLCO% from baseline to follow-up. We identified three VAS breathing subgroups: VAS worsened, or subjects with follow-up VAS breathing scores >10 points above baseline; VAS unchanged, those with scores within 10 points of baseline; and VAS improved, those with scores >10 points lower than baseline. We identified three subgroups for TLC%, FVC% and DLCO%: improved, i.e. subjects with follow-up TLC% and/or FVC% values $\geq 5\%$ (or DLCO% values $\geq 15\%$) higher than baseline; worsened, i.e. subjects with follow-up TLC% and/or FVC% values $\geq 5\%$ (or DLCO% values $\geq 15\%$) lower than baseline; and unchanged, i.e. subjects with follow-up TLC% and/or FVC% values within $\leq 5\%$ (or DLCO% values $\leq 15\%$) of baseline. The cut-off of 15% change in DLCO was used based on ATS/European Respiratory Society (ATS/ERS) criteria for defining a favourable response to treatment [39, 40]. Exploratory analysis using the ATS/ERS cut-off of a 10% change in FVC was also done. Changes in PFT values are relative, i.e. a baseline FVC of 60% that decreased by 10% would be 54% at follow-up. For changed scores, *t* tests were performed to determine significant differences in SGRQ scores between subgroups. For each of these groups, the effect size (ES) was calculated by deriving the average change in each SGRQ scale score from baseline to follow-up and dividing it by the baseline s.d. Cohen's rule of thumb for interpreting ES is that a value of 0.20–0.49 represents a small change, 0.50–0.79 a medium change and ≥ 0.80 a large change [37].

Results

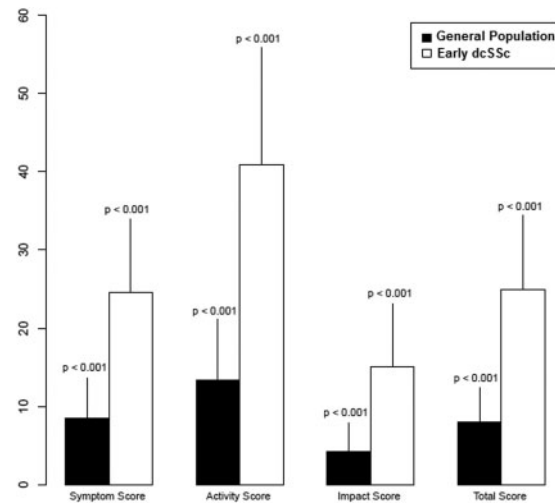
One hundred and seventy-seven participants completed the SGRQ at baseline. Of those, 132 (74.5%) were women, with a mean age of 50.5 years and mean disease duration 1.6 years (Table 1). The average modified Rodnan skin score was 20.6, denoting moderate to severe skin involvement. The mean PFT percentage predicted values were FVC% 83.1, DLCO% 67.3 and TLC% 87.5. Of patients who had HRCT, 73 (69.5%) had SSc-ILD (fibrosis, ground glass opacity, interstitial markings and/or traction bronchiectasis). The mean 6MWD was 489.3 feet and the mean HAQ-DI score was 1.1, denoting moderate functional disability. There were no statistical differences in SGRQ score between patients who completed the SGRQ and those who did not ($P=0.23$ – 0.40) or between patients with and without HRCT data [$P=0.12$ – 0.27 , except for the IMPACT score—mean 17.21 (s.d. 17.81) in 98 patients with HRCT, mean 12.99 (s.d. 14.54) in 73 patients without HRCT ($P=0.05$)]. These data are shown in tabular form.

Mean scores on the SYM, ACT, IMPACT and TOTAL scales of the SGRQ were 24.6 (S.D. 19.5), 40.9 (S.D. 31.1), 15.1 (S.D. 16.6) and 24.8 (S.D. 19.4), respectively.

TABLE 1 Baseline characteristics of CRISS cohort members who completed the Saint George's Respiratory Questionnaire

Demographics and disease severity	n	Value
Age, mean (s.d.), years	177	50.5 (11.7)
BMI, mean (s.d.), kg/m ²	158	25.18 (5.30)
Gender (female), %	177	74.5
Ethnicity, %		
Caucasian	142	80.23
African American	14	7.91
Asian	13	7.34
Native American	2	1.13
Pacific Islander	1	0.56
Mixed race	4	2.26
Other	1	0.56
MRSS, mean (s.d.)	177	20.6 (10.1)
Disease duration, mean (s.d.), years	171	1.6 (1.4)
Autoantibody status, %		
ANA negative	138	16.7
Centromere	97	11.3
Scl-70	135	26.7
Haemoglobin, mean (s.d.)	173	12.6 (1.6)
Dyspnoea		
SGRQ subscale scores for the overall cohort, mean (s.d.)		
SYM	177	24.6 (19.5)
ACT	174	40.9 (31.1)
IMPACT	171	15.1 (16.6)
TOTAL	168	24.8 (19.4)
SGRQ subscale scores for the SSc-ILD subgroup, mean (s.d.)		
SYM	73	30.3 (20.1)
ACT	71	48.1 (31.8)
IMPACT	71	19.8 (18.6)
TOTAL	69	30.2 (20.5)
VAS breathing, mean (s.d.)	175	20.4 (34.5)
PFT values (% predicted), mean (s.d.)		
FEV1%	164	85.9 (18.2)
FVC%	165	83.1 (18.3)
TLC%	131	87.5 (19.9)
DLCO%	165	67.3 (21.2)
HRCT, n (%) with SSc lung involvement	105	73 (69.5)
Functional status, HRQoL		
6MWD, mean (s.d.), m	54	489.3 (240.8)
HAQ-DI, mean (s.d.)	175	1.1 (0.8)
Physician global assessment (0-10 VAS), mean (s.d.)	157	4.26 (2.17)
Patient global assessment (0-10 VAS), mean (s.d.)	177	3.89 (2.66)
SF-36 PCS, mean (s.d.)	174	37.86 (12.77)
SF-36 MCS, mean (s.d.)	174	44.23 (6.06)

6MWD: 6 min walk distance; ACT: activity scale; DLCO: diffusing capacity of the lung for carbon monoxide; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; HAQ-DI: HAQ Disability Index; HRCT: high-resolution CT; HRQoL: health-related quality of life; ILD: interstitial lung disease; IMPACT: impact scale; MCS: mental component summary; MRSS: modified Rodnan skin score; PCS: physical component summary; PFT: pulmonary function test; SF-36: 36-item Short Form Health Survey; SGRQ: Saint George's Respiratory Questionnaire; SYM: symptom scale; TOTAL: total score; TLC: total lung capacity; VAS: visual analogue scale.

Fig. 1 Saint George's Respiratory Questionnaire (SGRQ) scale scores in the general population and in patients with early dcSSc

The vertical line over each bar represents the s.d. of the weighted average score. The mean SGRQ scores from the general population were obtained from Ferrer *et al.* [34].

After controlling for age and gender, scores on all scales at enrolment were statistically significantly higher (worse) than those obtained from a sample of the US general population (Table 1 and Fig. 1). Patients with SSc-ILD on HRCT had mean scores on the SYM, ACT, IMPACT and TOTAL scales of the SGRQ of 30.3 (s.d. 20.1), 48.1 (s.d. 31.8), 19.8 (s.d. 18.6) and 30.2 (s.d. 20.5), respectively.

Reliability

Cronbach's α for the SYM, ACT, IMPACT and TOTAL scales were satisfactory, ranging from 0.70 on the SYM scale to 0.93 on the ACT scale (supplementary Table S1, available at *Rheumatology* Online). The percentage of patients with the minimum possible score of 0 ranged from 4.8% on the TOTAL scale to 23.4% on the IMPACT scale. On the ACT scale 1.7% of patients ($n=3$) scored the maximum 100 points; there were no maximum scores on any other scales.

Correlations between SGRQ scales and other measures

Baseline SGRQ scales had large correlations among themselves ($r=0.52-0.93$) (Table 2). VAS breathing and DLCO% values showed large correlations with all scales ($r=0.46-0.79$) and FVC% values showed moderate to large negative correlations with all scales ($r=-0.25$ to -0.38). HAQ-DI had large correlations with ACT, IMPACT and TOTAL scores ($r=0.33-0.44$) and moderate correlations with VAS breathing. 6MWD had moderate correlations with IMPACT and TOTAL scores. Haemoglobin showed a large negative correlation with HAQ-DI and moderate positive correlation with DLCO. Patients

TABLE 2 Correlations between baseline Saint George's Respiratory Questionnaire scale scores and other measures

	SYM (n = 177)	ACT (n = 174)	IMPACT (n = 171)	TOTAL (n = 168)	VAS breathing (n = 175)	HAQ-DI (n = 177)	6MWD (n = 54)	FVC% (n = 165)	DLCO% (n = 165)	TLC% (n = 131)	Hgb (n = 173)
SYM	1	0.52* (<0.001)	0.68* (<0.001)	0.72* (<0.001)	0.52* (<0.001)	0.12 (0.12)	-0.22 (0.11)	-0.25* (0.001)	-0.46* (<0.001)	-0.13 (0.12)	-0.04 (0.63)
ACT		1	0.74* (<0.001)	0.92* (<0.001)	0.62* (<0.001)	0.44* (<0.001)	-0.21 (0.12)	-0.36* (<0.001)	-0.48* (<0.001)	-0.22* (0.01)	-0.20* (0.009)
IMPACT			1	0.93* (<0.001)	0.79* (<0.001)	0.33* (<0.001)	-0.33* (0.02)	-0.36* (<0.001)	-0.51* (<0.001)	-0.20* (0.02)	-0.14 (0.08)
TOTAL				1	0.75* (<0.001)	0.38* (<0.001)	-0.29* (0.04)	-0.38* (<0.001)	-0.55* (<0.001)	-0.23* (0.01)	-0.17* (0.03)
VAS breathing					1	0.25* (<0.001)	-0.19 (0.18)	-0.28* (<0.001)	-0.39* (<0.001)	-0.15 (0.09)	-0.08 (0.30)
HAQ-DI						1	-0.22 (0.11)	-0.31* (<0.001)	-0.26* (<0.001)	-0.17 (0.05)	-0.38* (<0.001)
6MWD							1	0.29* (0.04)	0.30* (0.03)	0.05 (0.76)	0.14 (0.32)
FVC%								1	0.67* (<0.001)	0.76* (<0.001)	0.16* (0.04)
DLCO%									1	0.55* (<0.001)	0.25* (0.001)
TLC%										1	0.09 (0.30)
Hgb											1

Values are given as correlation (*P*-value). **P* < 0.05. 6MWD: 6 min walk distance; ACT: activity scale; DLCO: diffusing capacity of the lung for carbon monoxide; FVC: forced vital capacity; HAQ-DI: HAQ Disability Index; Hgb: haemoglobin; IMPACT: impact scale; SYM: symptom scale; TLC: total lung capacity; TOTAL: total score; VAS: visual analogue scale.

with SSc-ILD on HRCT demonstrated similar correlation coefficients to the overall cohort with the exception of low correlation between 6MWD and TOTAL score in this subgroup (supplementary Table S2, available at *Rheumatology* Online).

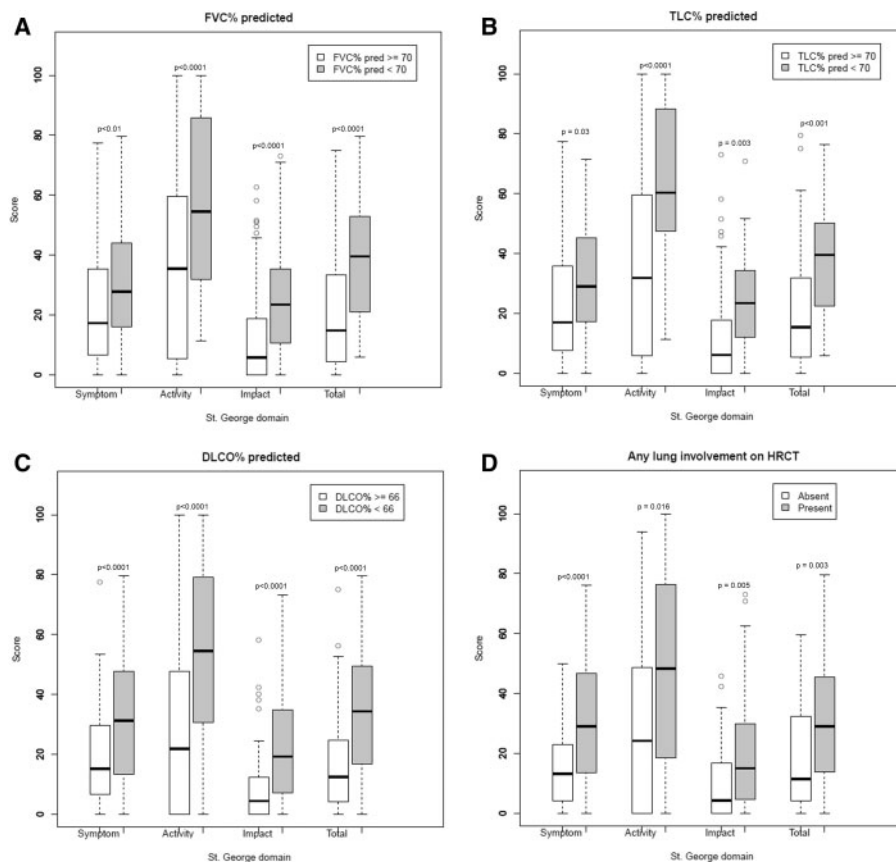
Discriminative validity of SGRQ scales for physiological and radiological parameters

Patients with low FVC% (*n* = 42), low TLC% (*n* = 23), low DLCO% (*n* = 92) and SSc-ILD on HRCT (*n* = 73) had statistically significantly higher scores on all SGRQ scales compared with patients with high FVC% (*n* = 135), high TLC% (*n* = 108), high DLCO% (*n* = 94) and no SSc-ILD (*n* = 104) (Fig. 2).

Responsiveness to change

Baseline and 1 year follow-up SGRQ scores were available for 150 participants (Tables 3 and 4). At 1 year, patients rated as improved by their physician had significantly lower SYM, ACT and TOTAL scores (*P* ≤ 0.02, ES: -0.27 to -0.21) and patients self-reporting improvement had significantly lower TOTAL scores relative to those categorized as unimproved (*P* ≤ 0.03, ES: -0.23) (Table 3). The subgroup with SSc-ILD on HRCT showed similar results, with significantly lower ACT scores in patients rated as improved by their physician (*P* = 0.03, ES = -0.38) and lower scores approaching significance on the other scales listed above (supplementary Table S3, available at *Rheumatology* Online).

Patients with improved VAS breathing values at follow-up had statistically significantly improved scores on all SGRQ scales compared with those with worsened values (*P* ≤ 0.01); ES was moderate to large (Table 4). These same patients had small improvements on all SGRQ scales relative to those with unchanged VAS breathing; changes in ACT, IMPACT and TOTAL scores were statistically significant (*P* ≤ 0.01). Patients with worsened follow-up VAS breathing values had worse scores on all SGRQ scales compared with those with unchanged values, with moderate ES in the SYM, IMPACT and TOTAL groups. Differences were significant for the TOTAL score (*P* < 0.05) and approached significance for the other scales (*P* = 0.07-0.09). Compared with those with worsened values at 1 year, patients with improved FVC% showed a moderate decrease in IMPACT and TOTAL scores (ES: -0.50, *P* = 0.02) when a cut-off of 5% change from baseline was used. When the ATS-recommended cut-off of 10% change was used, patients with worsened FVC% also showed a moderate increase in ACT, IMPACT and TOTAL scores (ES: -0.55 to -0.48) but the numbers of patients were small (*n* = 4-6) and *P*-values were not significant (*P* = 0.08-0.15). Patients with improved TLC% showed small to moderate significant decreases in all SGRQ scores (*P* ≤ 0.05) and patients with decreased TLC% showed small but significant increases in ACT and TOTAL scores compared with those with unchanged scores (*P* = 0.01). Patients with improved DLCO% showed a small but significant decrease in TOTAL score (*P* = 0.02). Patients with SSc-ILD on HRCT showed similar trends to the overall cohort, although ES

Fig. 2 Discriminative validity of baseline Saint George's Respiratory Questionnaire scales**(A)** FVC%; **(B)** TLC%; **(C)** DLCO%; **(D)** any lung involvement on high-resolution CT. DLCO: diffusing capacity of the lung for carbon monoxide; FVC: forced vital capacity; TLC: total lung capacity.

was larger and *P*-values did not reach significance in several cases (supplementary Table S4, available at *Rheumatology* Online). Similar to the overall analysis, when the ATS-recommended cut-off of a 10% change in FVC was used, patients with worsened FVC showed moderate to large increases in all four SGRQ scores (ES: -0.81 to -0.52), but the number was small ($n = 4$) and *P*-values were not significant ($P = 0.14$ – 0.50).

Moderate to large correlations were seen between changes in all SGRQ scales and changes in VAS breathing, FVC% and TLC% and between change in ACT, IMPACT and TOTAL scores and change in HAQ-DI. The largest correlations were seen between changes in VAS breathing and changes in all SGRQ scores ($r = 0.46$ – 0.61) (supplementary Table S5, available at *Rheumatology* Online). Similar trends were seen in the ILD population (supplementary Table S6, available at *Rheumatology* Online).

Discussion

Dyspnoea is a common and multifactorial source of functional impairment among patients with dcSSc [3, 4]. Dyspnoea is a complex construct and includes

symptoms, activity limitation and functional impairment [41]. Of the HRQoL instruments validated in SSc, the scleroderma HAQ (SHAQ)-associated VAS captures only a single dimension of dyspnoea (functional impairment). The SGRQ, like the Mahler dyspnoea index, measures symptoms and the impact of dyspnoea on multiple domains associated with dyspnoea. This makes it valuable as a potential outcome measure for dcSSc patients in general and those with SSc-ILD in particular.

Our data suggest that SGRQ is reliable and valid for use in an early dcSSc population. All four scales demonstrated good internal consistency [36], showed large correlations with other validated measures of functional capacity, breathing impairment, HRQoL and physiological impairment and could discriminate between patients with and without SSc-ILD by HRCT and/or restrictive lung disease. All SGRQ scales demonstrated responsiveness to change after 1 year in both the overall study population and the subgroup with SSc-ILD on HRCT.

The SGRQ was previously assessed in two small cross-sectional SSc studies. Beretta *et al.* [13] evaluated the SGRQ in 28 patients with SSc-related ILD. The study included patients with limited SSc or dcSSc with a long

TABLE 3 Responsiveness to change of the Saint George's Respiratory Questionnaire at 1 year compared with physician and patient disease assessments

	Overall score change	Improved, physician ^a	Unimproved, physician ¹	Improved, patient ¹	Unimproved, patient ¹
SYM					
Mean (s.d.)	-0.88 (19.38)	-5.27 (19.61)	4.53 (17.98)	-3.34 (17.22)	2.30 (22.07)
<i>n</i>	98	53	37	54	42
Effect size	-0.04	-0.27	0.22	-0.18	0.11
<i>P</i> -value	—	0.02	—	0.17	—
ACT					
Mean (s.d.)	-0.77 (24.36)	-7.95 (23.51)	7.95 (22.48)	-5.17 (23.37)	3.95 (25.03)
<i>n</i>	92	49	35	50	40
Effect size	-0.02	-0.25	0.23	-0.16	0.11
<i>P</i> -value	—	0.002	—	0.08	—
IMPACT					
Mean (s.d.)	-1.20 (15.65)	-3.58 (14.74)	0.56 (15.48)	-4.03 (11.91)	2.19 (22.07)
<i>n</i>	90	48	36	49	40
Effect size	-0.07	-0.21	0.03	-0.21	0.14
<i>P</i> -value	—	0.22	—	0.08	—
TOTAL					
Mean (s.d.)	-1.09 (16.52)	-5.35 (15.51)	3.56 (16.25)	-4.83 (12.79)	3.13 (19.53)
<i>n</i>	85	45	34	46	38
Effect size	-0.05	-0.26	0.17	-0.23	0.15
<i>P</i> -value	—	0.02	—	0.03	—

^aAt 1 year, physicians and patients rated the change in the patient's overall scleroderma severity on a scale of 1 (much better) to 5 (much worse). Scores of 1 and 2 were considered improved, scores 3–5 were considered unimproved. *P*-values for the reported group compared with the corresponding unimproved group. ACT: activity scale; IMPACT: impact scale; TOTAL: total score.

disease duration of 12.6 years (s.d. 7.3) and excluded patients with normal lung function (FVC% >80%), those with pulmonary hypertension on echocardiogram and those unable to perform the 6MWD due to non-respiratory limitations. Large correlations were observed between the 6MWD and ACT, IMPACT and TOTAL scores (−0.59 to −0.86, *P* < 0.01) and between FVC% and the ACT score (−0.47, *P* < 0.05). This study also demonstrated a relationship between disease severity on HRCT and ACT, IMPACT and TOTAL scores. Sözner *et al.* [42] evaluated 33 patients with limited SSc or dcSSc of any duration and ILD on HRCT. Inclusion and exclusion criteria were similar to those of Beretta *et al.* [13]. Only 13 patients had restrictive lung disease by PFTs. ACT scores had large inverse correlations with the 6MWD (−0.69, *P* < 0.0001) and DLCO% (−0.50, *P* = 0.005). Negative correlations were seen between all four scales and FEV1%, FVC% and 6MWD.

Similar correlations were also seen in our cohort except for smaller correlations between the SGRQ scales and 6MWD. The 6MWD is an imperfect outcome measure for use in SSc-related ILD. It does not correlate well with FVC [43] and is influenced by age, race, overall health status and both SSc-specific and non-specific co-morbidities [26, 44, 45]. It has only been partially validated by OMERACT for use as an endpoint in clinical trials of SSc, primarily because its poor specificity limits content validity [46, 47]. Furthermore, our population included

patients with early disease and concomitant small and large joint contractures, while both prior published studies included patients with longer disease duration and excluded patients with musculoskeletal limitations, a known confounder of 6MWD outcomes.

The current data also show a large correlation between the VAS for breathing scale and the SGRQ scales, particularly IMPACT and TOTAL. VAS breathing, which assesses overall functional impairment due to current breathing problems, was found to be a valid measure for assessing SSc-ILD in the Scleroderma Lung Study [48]. As the SGRQ assesses other domains associated with dyspnoea, such as respiratory symptom severity and physical activity impairment, it has greater content validity than VAS breathing in evaluating dyspnoea in SSc.

Our study adds valuable information to previously published studies. Our sample size was significantly larger and we assessed patients over 1 year. Our cohort had significantly higher SGRQ scores on all scales than previously published age- and gender-matched scores from the general population [34] and the subgroup with SSc-ILD had higher SGRQ scores on all scales than did patients without ILD, providing face validity in this cohort.

The SGRQ performed well both in a general dcSSc population and in the subgroup with HRCT-diagnosed SSc-ILD [49] and/or restrictive lung disease (defined as TLC% <70% of predicted value). The SGRQ has recently

TABLE 4 Responsiveness to change of the Saint George's Respiratory Questionnaire at 1 year compared with the VAS breathing, DLCO%, FVC% and TLC%

	VAS		VAS		VAS		DLCO%		DLCO%		FVC%		FVC%		TLC%		TLC%	
	Improved ^a	worsened ^a	unchanged ^a	improved ^b	worsened ^b	unchanged ^b	improved ^c	worsened ^c	unchanged ^c	improved ^d	worsened ^d	unchanged ^d	improved ^e	worsened ^e	unchanged ^e	improved ^f	worsened ^f	unchanged ^f
SYM																		
Mean (s.d.)	-0.9 (19.4)	-10.8 (22.4)	8.2 (23.4)	-1.6 (14.0)	-4.69 (18.3)	1.11 (18.3)	0.44 (19.8)	-7.5 (18.7)	3.2 (23.8)	0.9 (15.8)	-9.6 (19.4)	2.5 (17.2)	-2.9 (16.9)					
n	98	19	24	51	19	15	48	23	22	39	20	18	18					
Effect size	-0.04	-0.6	0.5	-0.09	-0.20	0.05	-0.03	-0.37	0.20	0.04	-0.45	0.14	-0.17					
P-value	—	0.01*	0.01**	0.11***	0.37*	0.79**	0.41***	0.10*	0.69**	0.08***	0.05*	0.35**	0.26**					
ACT																		
Mean (s.d.)	-0.8 (24.4)	-18.4 (26.7)	10.8 (25.5)	-0.3 (18.0)	-9.93 (25.4)	4.21 (16.7)	-1.28 (27.3)	-10.2 (16.8)	8.2 (26.8)	-3.7 (26.6)	-11.6 (20.1)	8.9 (10.4)	-10.2 (28.0)					
n	92	19	24	45	17	14	46	19	21	38	19	16	18					
Effect size	-0.02	-0.8	0.4	-0.0009	-0.27	0.13	-0.04	-0.30	0.24	-0.11	-0.32	0.29	-0.36					
P-value	—	<0.001*	<0.07**	0.01***	0.07*	0.37**	0.25***	0.01*	0.11**	0.27***	<0.001*	0.01**	0.86***					
IMPACT																		
Mean (s.d.)	-1.2 (15.7)	-13.6 (19.3)	7.04 (18.4)	-0.2 (8.9)	-7.64 (16.8)	1.22 (10.8)	-0.83 (15.8)	-9.7 (15.0)	6.0 (15.6)	-2.0 (13.0)	-8.3 (15.5)	4.8 (9.7)	-4.1 (16.3)					
n	90	17	23	47	18	13	45	22	21	35	19	18	16					
Effect size	-0.07	-0.7	0.5	-0.01	-0.39	0.06	-0.05	-0.50	0.35	-0.11	-0.45	0.34	-0.28					
P-value	—	0.002*	0.09**	0.01***	0.08*	0.59**	0.15***	0.002*	0.06**	0.05***	0.004*	0.07**	0.44***					
TOTAL																		
Mean (s.d.)	-1.1 (16.5)	-14.4 (18.2)	8.1 (18.5)	-0.4	-8.2 (14.1)	3.9 (12.2)	-0.83 (15.8)	-11.1 (13.5)	6.8 (19.0)	-2.1 (14.4)	-10.4 (13.7)	6.3 (9.1)	-7.1 (17.0)					
n	85	17	23	42	17	12	43	18	20	35	18	16	16					
Effect size	-0.05	-0.8	0.5	-0.02	-0.35	0.16	-0.05	-0.50	0.33	-0.10	-0.49	0.33	-0.41					
P-value	—	<0.001*	0.05**	0.007***	0.02*	0.27**	0.15***	0.002*	0.08**	0.03***	<0.001*	0.01**	0.54***					

^aVAS improved: score at 1 year is >10 points lower than baseline. VAS worsened: score at 1 year is >10 points higher than baseline. VAS unchanged: score at 1 year is within 10 points of the baseline score. ^bDLCO% improved: DLCO% at 1 year is $\geq 15\%$ higher than baseline. DLCO% worsened: DLCO% at 1 year is $\geq 15\%$ lower than baseline. DLCO% unchanged: DLCO% at 1 year is $\leq 15\%$ lower or $\leq 15\%$ higher than baseline. ^cFVC% improved: FVC% at 1 year is $\geq 5\%$ higher than baseline. FVC% worsened: FVC% at 1 year is $\geq 5\%$ lower than baseline. FVC% unchanged: FVC% at 1 year is $\leq 5\%$ lower or $\leq 5\%$ higher than baseline. ^dTLC improved: TLC% at 1 year is $\geq 5\%$ higher than baseline. TLC% worsened: TLC% at 1 year is $\geq 5\%$ lower than baseline. TLC% unchanged: TLC% at 1 year is $\leq 5\%$ lower or $\leq 5\%$ higher than baseline. P-values refer to *improved vs worsened; **worsened vs unchanged; ***improved vs unchanged. ACT: activity scale; DLCO: diffusing capacity of the lung for carbon monoxide; FVC: forced vital capacity; IMPACT: impact scale; SSYM: symptom scale; TLC: total lung capacity; TOTAL: total score; VAS: visual analogue scale.

been proposed as a valid outcome measure to assess SSc-ILD in an international Delphi exercise [50]. Our data support this, as patients with SSc-ILD on HRCT or restrictive lung disease on selected PFTs endorsed higher scores on all four SGRQ scales than those without these findings (Fig. 2). Furthermore, the correlations seen in the overall population among SGRQ scales and between each SGRQ scale and VAS breathing, 6MWD, HAQ-DI and PFTs were also seen in the subset of patients with SSc-ILD on HRCT (supplementary Table S2, available at *Rheumatology* Online).

All SGRQ scales were responsive to change over 1 year. Changes in all four scales tracked with changes in physician and patient assessments, changes in PFT values and scores on other validated instruments such as VAS breathing. Effect sizes were more robust (-0.6 to -0.8) when VAS breathing was used as an anchor compared with global assessments and change in FVC% or DLCO%. This is expected since both measures assess dyspnoea. Other anchors were associated with a smaller ES but were different from the no-change group (Table 4). Notably, a decline in FVC by 10% was associated with moderate change in IMPACT and TOTAL scores in the overall group and moderate to large changes in all four scores in the ILD group, but these results did not reach significance and the number of patients was small. Patients with improved physician assessment or worsened VAS breathing values at 1 year showed corresponding significant changes in SYM score, suggesting that it may assess subjective aspects of illness not captured well by physiological or functional assessments (Table 3). In the subset of patients with HRCT-defined ILD, similar trends were noticed.

This study has many strengths. It is the first multicentre longitudinal investigation of the SGRQ in SSc. The broad set of variables captured by the CRISS cohort allowed us to correlate SGRQ scale scores with other patient-reported measurements of dyspnoea (VAS breathing) and functional status (HAQ-DI), with objective measurements of functional capacity (6MWD) and with physiological and radiographic measurements of SSc-related pulmonary impairment (PFTs, HRCT). The longitudinal design also allowed us to demonstrate the SGRQ's responsiveness to change relative to the measures above.

The limitations of this study include lack of information about participants' treatment plans. However, since the current analysis focused on validation of the measure, it did not require data on therapy. The length of the instrument and complex computerized scoring algorithm may limit its utility in the clinical setting, but its use is feasible in the clinical trial setting as demonstrated by our cohort's high baseline completion rate. As this was an observational multicentre cohort study, missing data were inevitable. In addition, HRCT was interpreted by local radiologists and not scored using available semi-quantitative measures. Further studies are needed to carefully characterize the association between available HRCT scoring and dyspnoea instruments. In conclusion, the SGRQ was found to have acceptable reliability, construct

validity and responsiveness to change in dcSSc and differentiates between patients with and without ILD.

Acknowledgements

V.J.B. and D.K. were supported by grant NIH/NIAMS K24 AR063120-02. B.W. and S.K. wrote the first draft of the manuscript and were involved in data analysis. D.E.F., P.A.M., J.R.S., M.D.M. and D.K. were involved in study design, data acquisition and analysis, and manuscript drafting and editing. V.J.B. performed the statistical analysis. All authors critically revised the work and approved its publication. D.K. takes responsibility for integrity of the data and the accuracy of the data analysis.

Funding: This work was supported by the National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant numbers UO1 AR055057, K24 AR063120-02 to D.K. and V.B.).

Disclosure statement: P.A.M. has served as a consultant for Actelion, ChemoCentryx, GlaxoSmithKline and Sanofi and has received grant support from Actelion, Bristol-Myers Squibb, Celgene, ChemoCentryx, Genentech/Roche and GlaxoSmithKline. D.K. has served as a consultant and/or received grant support from Actelion, Bayer, Biogen-Idec, BMS, Cytosol, Genentech/Roche, InterMune, Lycera, Merck/EMD Serono and Sanofi-Aventis. J.R.S. has served as a paid consultant to Actelion, Aires, Bayer, Biogen Idec, Celgene, DART, Eicose, EMD Serono, InterMune, Sanofi-Aventis, Sigma Tau and United Therapeutics. D.E.F. has received grant/research support from AbbVie, Actelion, Amgen, BMS, Gilead, GSK, NIH, Novartis, Pfizer, Roche/Genentech and UCB; has served as a consultant for AbbVie, Actelion, Amgen, BMS, Gilead, GSK, Janssen, NIH, Novartis, Pfizer, Roche/Genentech and UCB and is on the speakers bureau [continuing medical education (CME) only] for AbbVie, Actelion and UCB. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

- Charles C, Clements P, Furst DE. Systemic sclerosis: hypothesis-driven treatment strategies. *Lancet* 2006;367: 1683–91.
- Thombs BD, Jewett LR, Assassi S *et al.* New directions for patient-centred care in scleroderma: the Scleroderma Patient-centred Intervention Network (SPIN). *Clin Exp Rheumatol* 2012;30(2 Suppl 71):S23–9.
- Bassel M, Hudson M, Taillefer SS *et al.* Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian National Survey. *Rheumatology* 2011;50:762–7.

- 4 Baron M, Sutton E, Hudson M *et al.* The relationship of dyspnoea to function and quality of life in systemic sclerosis. *Ann Rheum Dis* 2008;67:644–50.
- 5 Ioannidis JP, Vlachoyiannopoulos PG, Haidich AB *et al.* Mortality in systemic sclerosis: an international meta-analysis of individual patient data. *Am J Med* 2005;118: 2–10.
- 6 Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972–2002. *Ann Rheum Dis* 2007;66: 940–4.
- 7 Mathai SC, Hummers LK, Champion HC *et al.* Survival in pulmonary hypertension associated with the scleroderma spectrum of diseases: impact of interstitial lung disease. *Arthritis Rheum* 2009;60:569–77.
- 8 Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991;85(Suppl B): 25–31, discussion 3–7.
- 9 Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;145:1321–7.
- 10 Chang JA, Curtis JR, Patrick DL, Raghu G. Assessment of health-related quality of life in patients with interstitial lung disease. *Chest* 1999;116:1175–82.
- 11 Wilson CB, Jones PW, O'Leary CJ, Cole PJ, Wilson R. Validation of the St. George's Respiratory Questionnaire in bronchiectasis. *Am J Respir Crit Care Med* 1997; 156(2 Pt 1):536–41.
- 12 Al-Shair K, Atherton GT, Kennedy D *et al.* Validity and reliability of the St. George's Respiratory Questionnaire in assessing health status in patients with chronic pulmonary aspergillosis. *Chest* 2013;144:623–31.
- 13 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* 2007;46: 296–301.
- 14 Khanna D, Lovell DJ, Giannini E *et al.* Development of a provisional core set of response measures for clinical trials of systemic sclerosis. *Ann Rheum Dis* 2008;67: 703–9.
- 15 Khanna D, Distler O, Avouac J *et al.* Measures of response in clinical trials of systemic sclerosis: the Combined Response Index for Systemic Sclerosis (CRISS) and Outcome Measures in Pulmonary Arterial Hypertension related to Systemic Sclerosis (EPOSS). *J Rheumatol* 2009; 36:2356–61.
- 16 Wiese AB, Berrocal VJ, Furst DE *et al.* Correlates and responsiveness to change of measures of skin and musculoskeletal disease in early diffuse systemic sclerosis. *Arthritis Care Res* 2014;66:1731–9.
- 17 Gladue H, Furst DE, Berrocal V *et al.* Comparison of baseline characteristics of the Combined Response Index for Systemic Sclerosis (CRISS) cohort to patients enrolled in clinical trials of diffuse systemic sclerosis. American College of Rheumatology Annual Meeting. Washington, DC. November 12, 2012.
- 18 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.
- 19 Merkel PA, Herlyn K, Martin RW *et al.* Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. *Arthritis Rheum* 2002;46: 2410–20.
- 20 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23: 137–45.
- 21 Clements PJ, Wong WK, Hurwitz EL *et al.* Correlates of the disability index of the health assessment questionnaire: a measure of functional impairment in systemic sclerosis. *Arthritis Rheum* 1999;42:2372–80.
- 22 Khanna D, Clements PJ, Furst DE *et al.* Correlation of the degree of dyspnea with health-related quality of life, functional abilities, and diffusing capacity for carbon monoxide in patients with systemic sclerosis and active alveolitis: results from the Scleroderma Lung Study. *Arthritis Rheum* 2005;52:592–600.
- 23 Ware JE, Kosinski M, Keller SK. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA, USA: Health Institute, Tufts Medical Center, 1994.
- 24 Ware J. SF-36 health survey update. *Spine (Phila Pa 1976)* 2000;25:3130–9.
- 25 Khanna D, Furst DE, Clements PJ *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.
- 26 Villalba WO, Sampaio-Barros PD, Pereira MC *et al.* Six-minute walk test for the evaluation of pulmonary disease severity in scleroderma patients. *Chest* 2007;131: 217–22.
- 27 ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002;166:111–7.
- 28 Wanger J, Clausen JL, Coates A *et al.* Standardisation of the measurement of lung volumes. *Eur Respir J* 2005;26: 511–22.
- 29 Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011;139:52–9.
- 30 Pellegrino R, Viegi G, Brusasco V *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005;26: 948–68.
- 31 Wells AU, Hansell DM, Corrin B *et al.* High resolution computed tomography as a predictor of lung histology in systemic sclerosis. *Thorax* 1992;47:738–42.
- 32 Moore OA, Goh N, Corte T *et al.* Extent of disease on high-resolution computed tomography lung is a predictor of decline and mortality in systemic sclerosis-related interstitial lung disease. *Rheumatology* 2013;52: 155–60.
- 33 Goldin J, Elashoff R, Kim HJ *et al.* Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. *Chest* 2009;136:1333–40.
- 34 Ferrer M, Villasante C, Alonso J *et al.* Interpretation of quality of life scores from the St George's Respiratory Questionnaire. *Eur Respir J* 2002;19:405–13.

- 35 Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1992;1:73–5.
- 36 Miller M. Coefficient alpha: a basic introduction from the perspectives of classical test theory and structural equation modeling. *Struct Equation Model* 1995;2: 255–73.
- 37 Cohen J.. *Statistical power analysis for the behavioral sciences*. 2nd edn. Hillsdale, NJ, USA: Erlbaum, 1988.
- 38 American Thoracic Society. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991;144:1202–18.
- 39 American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS) and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* 2000;161(2 Pt 1):646–64.
- 40 Hanson D, Winterbauer RH, Kirtland SH, Wu R. Changes in pulmonary function test results after 1 year of therapy as predictors of survival in patients with idiopathic pulmonary fibrosis. *Chest* 1995;108:305–10.
- 41 Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest* 1984;85:751–8.
- 42 Celebi Sozener Z, Karabiyikoglu G, Duzgun N. Evaluation of the functional parameters in scleroderma cases with pulmonary involvement. *Tuberk Toraks* 2010;58: 235–41.
- 43 Matucci-Cerinic M, D'Angelo S, Denton CP, Vlachoyiannopoulos P, Silver R. Assessment of lung involvement. *Clin Exp Rheumatol* 2003;21(3 Suppl 29): S19–23.
- 44 de Oliveira NC, dos Santos Sabbag LM, Ueno LM *et al*. Reduced exercise capacity in systemic sclerosis patients without pulmonary involvement. *Scand J Rheumatol* 2007; 36:458–61.
- 45 Schoindre Y, Meune C, Dinh-Xuan AT *et al*. Lack of specificity of the 6-minute walk test as an outcome measure for patients with systemic sclerosis. *J Rheumatol* 2009;36: 1481–5.
- 46 Merkel PA, Clements PJ, Reveille JD *et al*. Current status of outcome measure development for clinical trials in systemic sclerosis. Report from OMERACT 6. *J Rheumatol* 2003;30:1630–47.
- 47 Avouac J, Kowal-Bielecka O, Pittrow D *et al*. Validation of the 6 min walk test according to the OMERACT filter: a systematic literature review by the EPOSS-OMERACT group. *Ann Rheum Dis* 2010;69:1360–3.
- 48 Tashkin DP, Elashoff R, Clements PJ *et al*. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006;354:2655–66.
- 49 Prosch H, Schaefer-Prokop CM, Eisenhuber E, Kienzl D, Herold CJ. CT protocols in interstitial lung diseases—a survey among members of the European Society of Thoracic Imaging and a review of the literature. *Eur Radiol* 2013;23:1553–63.
- 50 Saketkoo LA, Mittoo S, Huscher D *et al*. Connective tissue disease related interstitial lung diseases and idiopathic pulmonary fibrosis: provisional core sets of domains and instruments for use in clinical trials. *Thorax* 2014;69: 436–44.