BMJ Open Respiratory Research

Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis: insights from the INPULSIS trials

Jeffrey J Swigris,[1] Hilary Wilson,[2] Dirk Esser,[3] Craig S Conoscenti,[4] Wibke Stansen,[5] Nancy Kline Leidy,[2] Kevin K Brown[1]

[1]National Jewish Health, Denver, Colorado, USA
[2]Evidera Inc, Bethesda, Maryland, USA
[3]Boehringer Ingelheim International GmbH, Ingelheim am Rhein, Germany
[4]Boehringer Ingelheim Pharmaceuticals Inc, Ridgefield, Connecticut, USA
[5]Boehringer Ingelheim GmbH & Co KG, Ingelheim am Rhein, Germany

Correspondence to
Dr Jeffrey J Swigris;
swigrisj@njc.org

ABSTRACT

Introduction We evaluated the psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis (IPF) using data from the two INPULSIS trials.

Methods Data from 1061 patients treated with nintedanib or placebo were pooled. Internal consistency, test–retest reliability, construct validity, known-groups validity, responsiveness and responder thresholds were examined.

Results Cronbach's α was 0.93 for SGRQ total score and >0.75 for domain scores. In patients with stable disease based on change in forced vital capacity (FVC) ≤5% predicted or 'no change' on Patient's Global Impression of Change, intraclass correlation coefficients for the SGRQ total score were 0.72 or 0.76, respectively. Moderate to strong correlations were observed between SGRQ total and domain scores and the Cough and Sputum Assessment Questionnaire cough domains (−0.34 to −0.65), University of California San Diego Shortness of Breath Questionnaire (0.56 to 0.83) and EuroQol 5-Dimensional Quality of Life Questionnaire Visual Analogue Scale (−0.41 to −0.55); correlations with FVC % predicted were weak (−0.24 to −0.30). Longitudinal correlations between changes in SGRQ total score and these patient-reported outcomes over 52 weeks were moderate. Changes in SGRQ total, impact and activity scores were sensitive to detecting improvement or deterioration in FVC >10% predicted at week 52. Collectively, distribution-based and anchor-based approaches suggested using a change of 4–5 points in SGRQ total score as a starting point for responder analyses.

Conclusions The psychometric properties of the SGRQ support its use as a measure of health-related quality of life in patients with IPF.

Key messages

► We used pooled data from 1061 patients treated with nintedanib or placebo in the INPULSIS trials to evaluate the psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis (IPF).

► Based on measures of internal consistency, test-retest reliability, construct validity, known-groups validity, and sensitivity to detect change, our results support the use of the SGRQ as a measure of health-related quality of life in patients with IPF.

► Additional research is needed to test threshold estimates for deterioration and improvement in SGRQ total score.

Although developed for use in patients with chronic obstructive pulmonary disease (COPD) or asthma,[3] the St George's Respiratory Questionnaire (SGRQ) has also been used to measure HRQL in patients with IPF.[4] The SGRQ is a self-administered, 50-item questionnaire assessing three domains: symptoms, activity and impact. The symptoms domain addresses the frequency and severity of respiratory symptoms; the activity domain assesses activities that cause or are limited by breathlessness and the impact domain taps a range of aspects concerned with social functioning and the psychological impact of the disease.[3] Scores for each domain and the total score are weighted and range from 0 to 100, with higher scores indicating worse HRQL.

The SGRQ appears to possess acceptable internal consistency, construct validity and responsiveness in the assessment of HRQL in patients with IPF.[4 5] Additional research is needed to confirm the measurement

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a progressive fibrosing interstitial lung disease (ILD) characterised by exertional dyspnoea and cough.[1] As IPF progresses, health-related quality of life (HRQL) deteriorates, robbing patients of their emotional well-being and independence.[2]

properties of the SGRQ and develop estimates of clinically meaningful change in this patient population.

The results of the two, replicate, placebo-controlled, Phase III INPULSIS trials showed that nintedanib slowed disease progression by significantly reducing the annual rate of decline in forced vital capacity (FVC).[6] Change from baseline in SGRQ total score over 52 weeks was a key secondary endpoint in both trials. In INPULSIS-2, there was a significant difference in favour of nintedanib on this endpoint, whereas there was no significant difference between treatment groups in INPULSIS-1. In pooled data from both trials, the difference between the treatment groups in change from baseline in SGRQ total score at week 52 was −1.43 points (95% CI −3.09 to 0.23) (p=0.09) in favour of nintedanib.[6]

In this analysis, we used pooled data from the INPULSIS trials to evaluate the psychometric properties of the SGRQ in patients with IPF. We were particularly interested in assessing the reliability, validity and responder thresholds of the SGRQ in this patient population. To our knowledge, this represents the largest and most extensive analysis of the psychometric properties of the SGRQ in patients with IPF to date.

METHODS
Trial design
The INPULSIS trials were conducted at 205 sites in 24 countries. The design of these trials has been described.[6] In brief, patients ≥40 years of age, with IPF diagnosed within the previous 5 years, FVC ≥50% predicted and a diffusing capacity of the lungs for carbon monoxide (DLco) 30%–79% predicted were randomised 3:2 to receive nintedanib 150 mg twice daily or placebo for 52 weeks. Patients who prematurely discontinued trial medication were encouraged to attend all study visits as originally planned.[6]

Both trials were conducted in accordance with the principles of the Declaration of Helsinki and the Harmonised Tripartite Guideline for Good Clinical Practice from the International Conference on Harmonisation and were approved by local authorities. The clinical trial protocol was approved by an Independent Ethics Committee and/or Institutional Review Board at all the participating centres. All patients provided written informed consent prior to study entry.

Outcome measures
We present analyses based on measures of FVC and the following patient-reported outcomes at baseline and at week 52 of treatment: the SGRQ, the University of California San Diego Shortness of Breath Questionnaire (UCSD-SOBQ), the cough domains of the Cough and Sputum Assessment Questionnaire (CASA-Q) and the EuroQol 5-Dimensional Quality of Life Questionnaire Visual Analogue Scale (EQ-5D VAS). In addition, we used data from a single Patient's Global Impression of Change (PGI-C) measure taken at week 52.

Spirometry was conducted using sponsor-provided machines according to standardised criteria.[7] The 2008 version of the SGRQ with a recall of 4 weeks was used.[3] The UCSD-SOBQ assesses the severity and limitations of shortness of breath experienced during activities of daily living with varying levels of exertion; it is scored by summing responses across 24 items to form a total score ranging from 0 to 120, with higher scores indicating greater breathlessness.[8] The CASA-Q consists of four domains that assess cough and sputum, and their respective impacts, with response options on a 1–5 Likert scale. Domain scores range from 0 to 100, with lower scores indicating worse symptoms/impact.[9] The EQ-5D VAS is used to assess general health status on a scale of 0 to 100, with higher scores indicating better health.[10] The PGI-C was a single item asking patients to rate their overall status relative to baseline. Its response options were on a seven-point Likert scale ranging from 1 (very much better) to 7 (very much worse) (see online supplementary material).

Statistical analyses
All analyses were prespecified in a statistical analysis plan that was filed in the sponsor's archive before database lock of the INPULSIS trials, with the exception of the following: responder thresholds were analysed post hoc after a review of the initial analyses; stability/reliability analyses conducted at week 6, as well as those using change in FVC ≤2% predicted as a criterion for stability, were conducted to address reviewers' comments on this manuscript. All analyses were undertaken using pooled data from all patients treated with ≥1 dose of nintedanib or placebo in the INPULSIS trials.

Internal consistency
Cronbach's α coefficient was calculated for the SGRQ total and domain scores, UCSD-SOBQ and CASA-Q cough symptoms and impact domains at baseline. Values >0.7 were considered to indicate a homogeneous scale.[11]

Stability and test–retest reliability
Intraclass correlation coefficients (ICCs) and effect sizes over 52 weeks were calculated for SGRQ total and domain scores, UCSD-SOBQ, CASA-Q cough symptoms and impact domains and EQ-5D VAS in patients who were defined as clinically stable based on the following criteria: (1) a change in FVC ≤5% predicted from baseline in either direction,[12] (2) a change in FVC ≤2% predicted from baseline in either direction, (3) a change in UCSD-SOBQ score of <5 points from baseline in either direction and (4) a rating of 'no change' on the PGI-C at week 52. ICC and effect sizes over 6 weeks were calculated in patients who were defined as clinically stable based on a change in FVC ≤2% predicted from baseline in either direction and a change in UCSD-SOBQ score of <5 points in either direction. The ICC for instrument total/domain scores was estimated from random-effect

models with patient as random effect and by dividing the estimate for the between-patient covariance parameter by the sum of the estimates for the within and between-patient covariance parameter. ICC values >0.7 were considered to indicate acceptable test–retest reliability[11] (at week 6) or stability (at week 52), and an effect size of <0.20 was considered acceptable to support the stability of each score.

Construct validity

Construct validity was assessed cross-sectionally (at baseline) and longitudinally (as changes from baseline at week 52) by comparing SGRQ total and domain scores and FVC % predicted, UCSD-SOBQ score, CASA-Q cough domain scores and EQ-5D VAS. Spearman's correlation coefficients were considered to indicate moderate correlation if the value was between 0.30 and 0.60 and strong correlation if >0.60.[13]

Known-groups validity

Known-groups validity was assessed by evaluating differences in SGRQ total and domain scores at baseline between patients stratified by FVC % predicted (highest/lowest quartile) and use of supplemental oxygen (yes/no). Two-sided t-tests were used to investigate differences between groups.

Responsiveness

Changes from baseline in SGRQ total and domain scores in patients showing improvement or deterioration in health status at week 52 were assessed. Improvement or deterioration in health status was defined as an increase or decrease from baseline in FVC >10% predicted at week 52 and a PGI-C of very much or much better (score 1–2) or much or very much worse (score 6–7) at week 52. Two-sided t-tests were conducted to investigate differences between groups. An effect size of >0.20 supported the ability of a score to detect change.

Responder thresholds

Clinically meaningful change from baseline in SGRQ total score at week 52 was assessed using a triangulation method, including both anchor-based and distribution-based approaches.

In the anchor-based approach, analysis of variance and receiver operating characteristic (ROC) analyses of changes from baseline in SGRQ total and domain scores at week 52 relative to changes in FVC % predicted (≤2% (stable), deterioration/improvement >2% and ≤5%, >5% and ≤10%, >10%), UCSD-SOBQ score (<5 points (stable), deterioration/improvement ≥5 points) and PGI-C (very much or much better/worse, a little better/worse, no change) were assessed. In the distribution-based approach, three responder thresholds were estimated: 1 SE of the mean (SEM),[14 15] 0.2 SD (lower bound of estimate) and 0.5 SD (upper bound of estimate).[16]

Handling of missing data

All analyses were conducted using observed cases. If there were more than two, four or six missing items in the SGRQ symptoms, activity or impact domains, respectively, the domain was set to missing.

RESULTS

A total of 1061 patients were treated with at least one dose of nintedanib or placebo in the INPULSIS trials. The baseline characteristics of the patient population are presented in online supplementary table S1. Mean (SD) age was 66.8 (8.0) years, FVC was 79.6 (17.8) % predicted, DLco was 47.2 (13.5) % predicted and SGRQ total score was 39.5 (18.9).

Internal consistency

Cronbach's α coefficients were 0.93, 0.77, 0.86 and 0.86 for the SGRQ total, symptoms, activity and impact scores, respectively. Cronbach's α coefficients for the UCSD-SOBQ, CASA-Q cough symptoms and impact domains were 0.97, 0.84 and 0.93, respectively.

Stability over 6 weeks

In patients with stable disease defined as a change in FVC ≤2% predicted from baseline in either direction, SGRQ total score and domain scores met threshold criteria for stability based on both ICC (range 0.72–0.83) and effect size (range −0.060–0.030) (table 1). In patients with stable disease defined as a change in UCSD-SOBQ score of <5 points from baseline in either direction, threshold criteria for stability were met for the SGRQ total score and domain scores based on both ICC (0.76–0.91) and effect size (−0.078–0.015). Threshold criteria were also met for the UCSD-SOBQ and CASA-Q impacts and symptoms domains.

Stability over 52 weeks

In general, SGRQ total scores met threshold criteria for both ICC (0.72–0.87) and effect size (−0.003–0.181) regardless of the definition of stability over 52 weeks (change in FVC ≤2% predicted, change in FVC ≤5% predicted, change in UCSD-SOBQ score of <5 points, no change on PGI-C) (table 1). SGRQ symptoms and impact domain scores did not meet the ICC threshold criteria for stability. Threshold criteria for effect size were met for all SGRQ domain scores except for the activity score using PGI-C as an anchor. For the other instrument scores, only the UCSD-SOBQ met threshold criteria by all definitions over 52 weeks. The EQ-5D VAS did not meet threshold ICC criteria by any definition (0.49–0.68), but met threshold effect size criteria for all anchors except the PGI-C. Results for the CASA-Q domains were mixed.

Table 1 Stability of SGRQ total and domain scores in patients with stable disease

	Change in FVC ≤2% predicted at week 6 (n=389)		Change in UCSD-SOBQ score of <5 points at week 6 (n=351)		Change in FVC ≤2% predicted at week 52 (n=187)		Change in FVC ≤5% predicted at week 52 (n=419)		Change in UCSD-SOBQ score of <5 points at week 52 (n=194)		No change on PGI-C at week 52 (n=316)	
	ICC	Effect size	ICC	Effect size	ICC	Effect size	ICC	Effect size	ICC	Effect size	ICC	Effect size
SGRQ												
Total*	0.83 (0.80–0.86)	−0.029	0.91 (0.88–0.92)	−0.047	0.75 (0.68–0.81)	0.080	0.72 (0.67–0.77)	0.047	0.87 (0.83–0.90)	−0.003	0.76 (0.71–0.80)	0.181
Activity domain	0.81 (0.77–0.84)	0.030	0.88 (0.85–0.90)	−0.014	0.75 (0.68–0.81)	0.077	0.71 (0.66–0.76)	0.074	0.87 (0.83–0.90)	0.048	0.73 (0.68–0.78)	0.220
Impact domain	0.76 (0.72–0.80)	−0.035	0.87 (0.84–0.89)	−0.078	0.67 (0.59–0.75)	0.088	0.65 (0.58–0.70)	0.068	0.80 (0.75–0.85)	−0.011	0.68 (0.62–0.74)	0.166
Symptoms domain	0.72 (0.67–0.77)	−0.060	0.76 (0.71–0.80)	0.015	0.72 (0.64–0.78)	0.033	0.66 (0.60–0.71)	−0.071	0.68 (0.60–0.75)	−0.038	0.66 (0.59–0.72)	0.067
UCSD-SOBQ score†	0.83 (0.80–0.86)	0.044	N/A	N/A	0.75 (0.68–0.81)	0.166	0.72 (0.66–0.76)	0.167	N/A	N/A	0.75 (0.70–0.80)	0.235
CASA-Q												
Cough impacts domain	0.74 (0.70–0.79)	0.097	0.84 (0.81–0.87)	0.043	0.67 (0.58–0.74)	0.002	0.69 (0.63–0.74)	0.016	0.76 (0.70–0.82)	0.023	0.70 (0.64–0.75)	−0.099
Cough symptoms domain	0.70 (0.65–0.75)	0.055	0.75 (0.70–0.79)	0.058	0.60 (0.50–0.69)	−0.015	0.61 (0.54–0.67)	0.052	0.71 (0.64–0.77)	0.077	0.66 (0.59–0.72)	−0.028
EQ-5D VAS	N/C	N/C	N/C	N/C	0.60 (0.50–0.68)	−0.030	0.49 (0.42–0.56)	−0.074	0.68 (0.60–0.75)	−0.124	0.51 (0.42–0.58)	−0.204

*n=315 patients with no change on PGI-C.

†n=410 for patients with FVC ≤5% predicted group and n=309 for patients with no change on PGI-C.

N/C: EQ-5D VAS data were not collected at week 6.

CASA-Q, Cough and Sputum Assessment Questionnaire; EQ-5D VAS, EuroQol 5-Dimensional Quality of Life Questionnaire Visual Analogue Scale; FVC, forced vital capacity; ICC, intraclass correlation coefficient; PGI-C, Patient 's Global Impression of Change; SGRQ, St George's Respiratory Questionnaire; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

Table 2 Spearman's correlation coefficients (for cross-sectional and longitudinal correlations) between SGRQ total and domain scores and UCSD-SOBQ, CASA-Q cough domain scores, EQ-5D VAS scores and FVC % predicted

SGRQ score	UCSD-SOBQ score		CASA-Q cough symptoms domain score		CASA-Q cough impacts domain score		EQ-5D VAS		FVC % predicted	
	Baseline	Change from baseline at week 52	Baseline	Change from baseline at week 52	Baseline	Change from baseline at week 52	Baseline	Change from baseline at week 52	Baseline	Change from baseline at week 52
Total	0.83	0.61	−0.51	−0.43	−0.65	−0.59	−0.55	−0.41	−0.30	−0.34
Activity domain	0.82	0.57	−0.34	−0.29	−0.47	−0.41	−0.50	−0.34	−0.29	−0.30
Impact domain	0.76	0.53	−0.49	−0.38	−0.64	−0.55	−0.54	−0.38	−0.26	−0.28
Symptoms domain	0.56	0.35	−0.61	−0.45	−0.64	−0.49	−0.41	−0.25	−0.24	−0.23

P<0.0001 for all correlation coefficients.
CASA-Q, Cough and Sputum Assessment Questionnaire; EQ-5D VAS, EuroQol 5-Dimensional Quality of Life Questionnaire Visual Analogue Scale; FVC, forced vital capacity; SGRQ, St George's Respiratory Questionnaire; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

Validity
Construct validity
At baseline, moderate to strong correlations were observed between SGRQ total and domain scores and UCSD-SOBQ score, CASA-Q cough domain scores and EQ-5D VAS. Correlations between SGRQ total and domain scores and FVC % predicted were weak (table 2). Longitudinal correlations between changes in SGRQ total and domain scores and changes in UCSD-SOBQ score, CASA-Q cough domain scores and EQ-5D VAS at week 52 were generally moderate, whereas correlations with changes in FVC % predicted were generally weak but in the expected direction (table 2).

Known-groups validity
Patients in the lowest quartile of FVC % predicted at baseline had significantly higher mean SGRQ scores than patients in the highest quartile (figure 1 and online supplementary table S2). Patients using supplemental oxygen at baseline (n=91) had significantly higher mean SGRQ scores than patients who were not (n=952) (figure 1 and online supplementary table S2).

Responsiveness
Mean changes in SGRQ total score were −6.8 and 13.1 in patients with improvement and deterioration in FVC >10% predicted from baseline at week 52, respectively (between-group difference of 19.9 (95% CI 12.8 to 27.1); p<0.0001) (figure 2). Changes in SGRQ total, impact and activity scores were sensitive to detecting improvement or deterioration in FVC >10% predicted from baseline at week 52; effect sizes for every score were >0.20 (figure 2 and table 3). Changes in SGRQ symptoms score detected
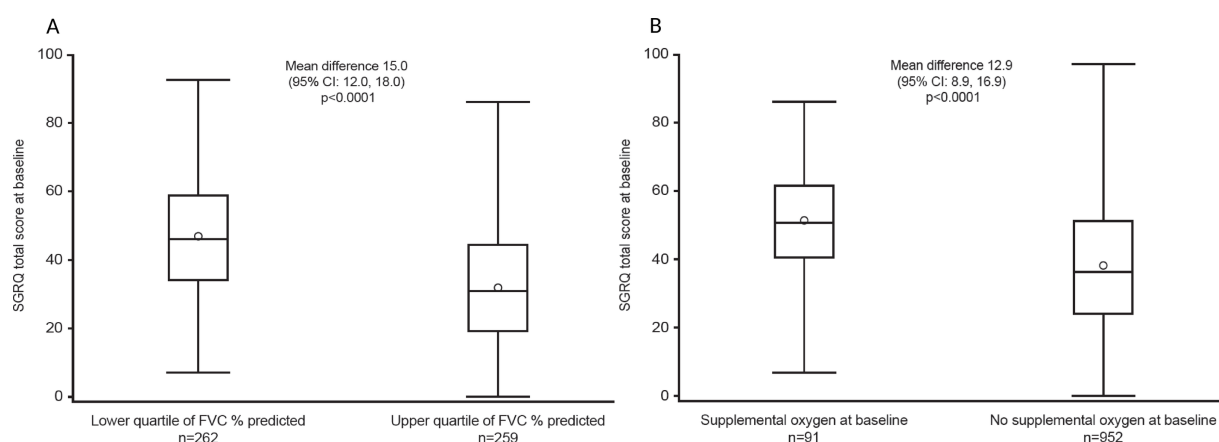

Figure 1 SGRQ total scores by (A) FVC % predicted at baseline and (B) use of supplemental oxygen at baseline. The circles denote the mean values, the midlines of the boxes indicate the median values and boundaries denote 25th and 75th percentiles; whiskers are the minimal and maximum values in the lower fence (1.5 interquartile range above 75th percentile) and upper fence (1.5 interquartile range below 25th percentile). FVC % predicted at baseline: lower quartile=65.92, upper quartile=90.62. FVC, forced vital capacity; SGRQ, St George's Respiratory Questionnaire.
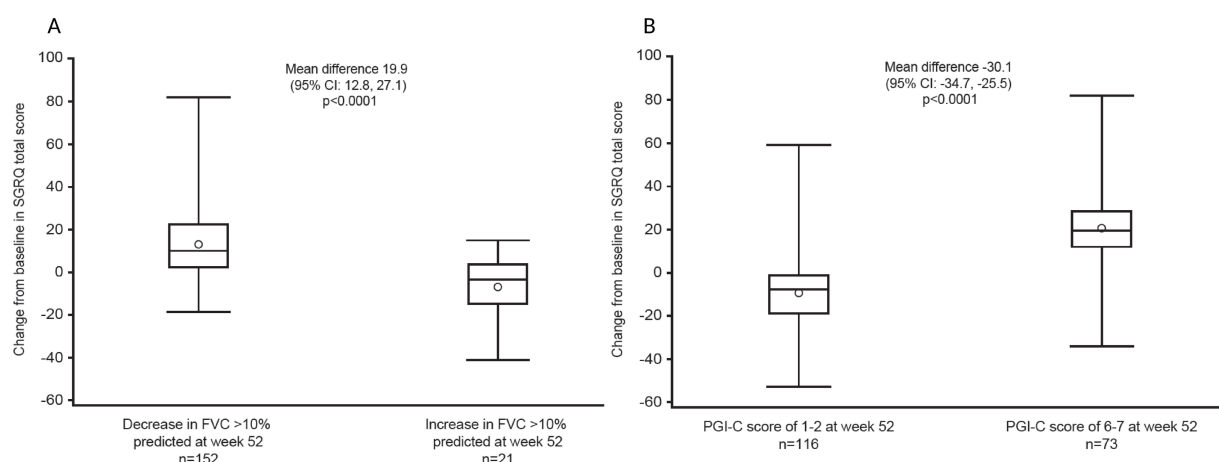
Figure 2 Changes from baseline in SGRQ total scores at week 52 by (A) changes in FVC >10% predicted at week 52 and (B) changes in PGI-C at week 52. The circles denote the mean values, the midlines of the boxes indicate the median values and boundaries denote 25th and 75th percentiles; whiskers are the minimal and maximum values in the lower fence (1.5 interquartile range above 75th percentile) and upper fence (1.5 interquartile range below 25th percentile). FVC, forced vital capacity; PGI-C, Patient 's Global Impression of Change; SGRQ, St George's Respiratory Questionnaire.

a deterioration (effect size of 0.38), but not an improvement (effect size of 0.14) in FVC >10% predicted from baseline at week 52 (table 3).

Mean changes in SGRQ total score were −9.7 and 20.4 in patients with improvement or deterioration according to PGI-C at week 52, respectively (between-group difference of −30.1 (95% CI −34.7 to −25.5); p<0.0001) (figure 2). Changes in SGRQ total and domain scores were sensitive to detecting improvement or deterioration according to PGI-C, with all effect sizes >0.20 (figure 2 and table 3).

Responder thresholds
Anchor-based approach
Mean (SD) changes in SGRQ total score over 52 weeks in patients who experienced a deterioration or improvement in FVC >5% and ≤10% predicted were 4.8 (15.4) and −4.4 (14.1), respectively (table 4). Smaller changes in FVC % predicted (>2–≤5%) did not correspond to significant changes in SGRQ total score over 52 weeks. ROC analyses were not used because the optimal thresholds for the SGRQ total score using FVC % predicted as a criterion variable had low specificity/sensitivity. Mean (SD) changes in SGRQ total score over 52 weeks in patients with deterioration or improvement in UCSD-SOBQ score of ≥5 points were 11.1 (15.1) and −8.1 (14.1), respectively (table 4). ROC analyses indicated a threshold of 3.3 for deterioration and −2.1 for improvement in SGRQ total score and had fair specificity and sensitivity. Patients who said they were a little worse on the PGI-C had a mean (SD) deterioration in SGRQ total score of 10.1 (12.4). Patients who said they were a little better had a mean (SD) improvement in SGRQ total score of 0.5 (13.9) (table 4). ROC analyses had low specificity and/or sensitivity.

Distribution-based approach
For distribution-based estimates of responder thresholds, 1 SEM was 10.74, 0.2 SD was 4.24 and 0.5 SD was 10.59.

DISCUSSION
A PRO measure should fulfil accepted criteria for reliability, validity and responsiveness to change. Its threshold for meaningful change is useful for interpreting the impact of interventions.[17] The SGRQ has been used previously to assess HRQL in patients with IPF, but this is the largest longitudinal dataset available that has been used to test its reliability, validity and meaningful change in this patient population.[18]

We analysed the psychometric properties of the SGRQ using pooled data from 1061 patients with IPF treated with nintedanib or placebo in the INPULSIS trials. Internal consistency was high for the SGRQ total and domain scores, confirming homogeneity of items within domains and for the entire instrument. The lower internal consistency of the symptoms domain is consistent with previous studies[5 19 20] and may be due to certain items within the symptoms domain (eg, those related to sputum and wheezing) lacking clinical relevance for many patients with IPF. In patients whose disease was clinically stable over 6 weeks, SGRQ total and domain scores all showed acceptable stability. In patients whose disease was clinically stable by all criterion variables over 52 weeks, the SGRQ total score had acceptable stability, but there was variability in the stability of domain scores, likely due to measurement error, which attenuated validity coefficients. The low ICC for the EQ-5D VAS is likely due to the generic nature of the instrument, which by definition is intended to capture all aspects of health in addition to IPF.

The content validity of the CASA-Q cough domains and UCSD-SOBQ has been demonstrated in patients with IPF, with patients reporting that the items in these questionnaires were relevant to their symptoms.[21] In our data, moderate to strong cross-sectional correlations were observed between SGRQ total and domain scores and the UCSD-SOBQ and CASA-Q cough domains, supporting

Table 3 Changes from baseline in SGRQ total and domain scores at week 52 and corresponding effect sizes by changes in FVC >10% predicted and PGI-C at week 52

Change from baseline in SGRQ score at week 52	Deterioration in FVC >10% predicted at week 52			Improvement in FVC >10% predicted at week 52					PGI-C score 1–2 (very much or much better) at week 52			PGI-C score 6–7 (much or very much worse) at week 52				
	n	Mean (SD)	Effect size	n	Mean (SD)	Effect size	Mean (SD) difference	P values	n	Mean (SD)	Effect size	n	Mean (SD)	Effect size	Mean (SD) difference	P-values
Total	152	13.1 (15.6)	0.64	21	−6.8 (15.2)	−0.35	19.9 (3.6)	<0.0001	116	−9.7 (14.9)	−0.55	73	20.4 (16.6)	1.20	−30.1 (2.3)	<0.0001
Activity domain	154	15.6 (19.0)	0.68	22	−4.4 (17.2)	−0.22	19.9 (4.3)	<0.0001	118	−7.5 (17.6)	−0.37	73	20.8 (20.8)	1.11	−28.3 (2.8)	<0.0001
Impact domain	153	13.2 (18.1)	0.58	21	−9.0 (18.2)	−0.41	22.1 (4.2)	<0.0001	116	−10.6 (18.3)	−0.54	73	21.4 (18.3)	1.12	−32.0 (2.7)	<0.0001
Symptoms domain	155	9.4 (20.5)	0.38	22	−3.2 (20.6)	−0.14	12.6 (4.7)	0.0076	118	−9.1 (18.7)	−0.43	74	17.8 (21.4)	0.92	−27.0 (2.9)	<0.0001

FVC, forced vital capacity; PGI-C, Patient 's Global Impression of Change; SGRQ, St George's Respiratory Questionnaire.

the SGRQ as an instrument capable of measuring aspects of disease relevant to patients with IPF. Longitudinal correlations, based on changes in these PROs over 52 weeks, were moderate and in the expected directions, thus further supporting validity. These data are consistent with recent findings from the Australian IPF registry[22] and the German INSIGHTS-IPF registry,[23] which demonstrated that dyspnoea and cough were strongly correlated with SGRQ total and domain scores at baseline.

SGRQ total and domain scores clearly distinguished patients in the highest and lowest quartiles of FVC % predicted at baseline and patients who were using or not using supplemental oxygen at baseline. Correlations between SGRQ scores and FVC % predicted were generally weak. This is consistent with the findings of previous studies in patients with IPF, including patient registries and the Phase II TOMORROW trial of nintedanib, in which correlations between the SGRQ total and domain scores and FVC % predicted at baseline ranged from −0.11 to −0.15.[5 22 23] Such weak correlations align with previously reported findings that factors other than lung function, such as patients' general health, comorbidities, mood disturbance, energy level and independence, may have an important impact on HRQL in patients with IPF.[22–26] However, despite the less-than-strong correlations, analyses showed that changes in SGRQ total, activity and impact scores were sensitive to detecting change in patients with >10% improvement or deterioration in FVC % predicted. This finding is consistent with pooled data from patients with IPF treated with bosentan or placebo in a Phase III study, in which changes in SGRQ scores were able to distinguish patients whose disease status declined, improved or remained stable based on changes in FVC % predicted, DLco % predicted and dyspnoea over a 6-month period.[20] In our analyses, changes in SGRQ scores were also sensitive to change in patients who reported improvement or deterioration based on the PGI-C.

A change in SGRQ total score of 4 points is generally accepted as clinically meaningful in patients with COPD.[3 27] Collectively, our analyses suggest that a change in SGRQ total score of approximately 4 to 11 points over 52 weeks may be clinically meaningful in patients with IPF, but it is not possible to draw a firm conclusion given the ambiguity of the ROC analyses. These preliminary results are, however, similar to findings from the placebo-controlled Phase III study of bosentan, which suggested that a change in SGRQ total score of 7 points over 6 months was meaningful.[20] We recognise that all the anchors we used have limitations. The UCSD-SOBQ was strongly correlated with the SGRQ, but data on its meaningfulness in patients with IPF are scarce. In a study conducted in 164 patients with various chronic lung diseases, a change in UCSD-SOBQ of 5 points was a reasonable estimate of clinically meaningful change.[28] In another analysis of data from 180 patients with IPF and advanced lung function impairment enrolled in a placebo-controlled trial of

Table 4 Changes from baseline in SGRQ total score at week 52 by categories of change in FVC % predicted, change in UCSD-SOBQ score and PGI-C at week 52

Anchor	n	Change from baseline in SGRQ total score at week 52 Mean (SD)	Comparison versus stable Adjusted mean (SE)	P values
Change in FVC % predicted				
Deterioration >10%	152	13.1 (15.6)	12.7 (1.6)	<0.001
Deterioration >5–≤10%	194	4.8 (15.4)	4.1 (1.5)	0.005
Deterioration >2–≤5%	141	0.7 (14.4)	0.0 (1.6)	0.998
Stable (≤2% change)	181	1.4 (13.7)	–	
Improvement >2–≤5%	78	−0.6 (13.9)	−1.1 (1.9)	0.566
Improvement >5–≤10%	56	−4.4 (14.1)	−5.0 (2.2)	0.022
Improvement >10%	21	−6.8 (15.2)	−6.1 (3.3)	0.065
Change in UCSD-SOBQ				
Deterioration ≥5	360	11.1 (15.0)	13.5 (1.2)	<0.001
Stable (change <5)	191	0.0 (9.8)	–	–
Improvement ≥5	192	−8.1 (14.1)	−4.9 (1.4)	<0.001
PGI-C				
Much worse/very much worse (6–7)	73	20.4 (16.6)	20.1 (1.7)	<0.001
A little worse (5)	156	10.1 (12.4)	8.2 (1.3)	<0.001
No change (4)	303	3.3 (12.9)	–	
A little better (3)	182	0.5 (13.9)	−2.0 (1.2)	0.098
Much better/very much better (1–2)	116	−9.7 (14.9)	12.1 (1.4)	<0.001

FVC, forced vital capacity; PGI-C, Patient's Global Impression of Change; SGRQ, St George's Respiratory Questionnaire; UCSD-SOBQ, University of California San Diego Shortness of Breath Questionnaire.

sildenafil, results suggested that a point estimate of 8 (range 5–11) represented a meaningful change in the UCSD-SOBQ.[29] We recognise the limitation of estimates based on the PGI-C, given the potential issues with recollection of change over a long-time period. Finally, although change in FVC is a well-established endpoint in trials of IPF,[30 31] the weak correlations with the SGRQ in our analysis suggest that it is not an ideal anchor to inform the meaningfulness of the SGRQ. Our finding that changes in FVC >5% predicted resulted in clinically meaningful changes in SGRQ scores is consistent with data showing a correlation between FVC changes of similar magnitude and outcomes such as hospitalisations and mortality,[12 32–35] but we acknowledge that a pooled analysis of data from the TOMORROW, INPULSIS, CAPACITY and ASCEND trials did not find a statistically significantly increased risk of death in patients who had a decline in FVC ≥5–<10 predicted compared with those who had a decline in FVC <5% predicted.[30] The present analysis supports using a change of 4–5 points as a starting point for responder threshold analyses of the SGRQ total score in patients with IPF. However, given the range of estimates in this analysis and an indication of greater thresholds in the Phase III study of bosentan,[20] we recommend conducting further sensitivity analyses.

This study had several strengths. First, with the exception of analyses of responder thresholds and stability/reliability analyses conducted at week 6 and using change in FVC ≤2% predicted as a criterion, all analyses were prespecified. Second, data were from two large prospective controlled trials. Third, the 52-week treatment duration provided a long follow-up period over which to assess the stability of scores. The main limitation of our analyses was the timing of the visits at which the PROs were assessed, which restricted certain validity tests. In particular, the PGI-C was only measured at week 52, and patients may have found it challenging to recall their health status over such a prolonged period. Another limitation was the exclusion of patients with severely impaired lung physiology (FVC <50% predicted).

At the start of recruitment into the INPULSIS trials in 2011, the SGRQ was considered the most suitable tool for use in this patient population. Subsequent data from clinical trials and patient registries support its utility in patients with IPF.[5 22 23] Additional tools have been developed specifically for use in patients with ILD (eg, the King's Brief Interstitial Lung Disease (K-BILD) questionnaire[36]) or IPF (eg, A Tool to Assess Quality of Life in IPF (ATAQ-IPF(-cA)) questionnaire[37 38] and Living with IPF (L-IPF) questionnaire[39]). The SGRQ has also been adapted for use with patients with IPF, the SGRQ-I,[19] but

analyses are needed to determine its performance characteristics. It is encouraging that these and other instruments are in development.[40] Future studies in patients with IPF should focus on evaluating these questionnaires and their measurement properties. Given its brevity, the K-BILD may be a particularly suitable instrument for use in clinical trials, where the need to collect comprehensive information needs to be balanced with the burden of filling out questionnaires. In such studies, the SGRQ may serve as a useful anchor and platform for development of more specific tools to collect patient-reported information.

In conclusion, the psychometric properties of the SGRQ observed in the INPULSIS trials support the use of this instrument as a measure of HRQL in patients with IPF. Additional research is needed to test the threshold estimates for deterioration and improvement in SGRQ total score over 52 weeks and over shorter intervals, including qualitative interviews with patients and clinicians regarding what they consider a meaningful outcome of treatment.

REFERENCES

1. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788–824.
2. Belkin A, Swigris JJ. Health-related quality of life in idiopathic pulmonary fibrosis: where are we now? *Curr Opin Pulm Med* 2013;19:474–9.
3. Jones PW, Quirk FH, Baveystock CM. The St George's respiratory questionnaire. *Respir Med* 1991;85(Suppl B):25–31. discussion 33–7.
4. Swigris JJ, Esser D, Conoscenti CS, et al. The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. *Health Qual Life Outcomes* 2014;12:124.
5. Swigris JJ, Esser D, Wilson H, et al. Psychometric properties of the St George's Respiratory Questionnaire in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2017;49:1601788.
6. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–82.
7. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005;26:319–38.
8. Eakin EG, Resnikoff PM, Prewitt LM, et al. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest* 1998;113:619–24.
9. Crawford B, Monz B, Hohlfeld J, et al. Development and validation of a cough and sputum assessment questionnaire. *Respir Med* 2008;102:1545–55.
10. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33:337–43.
11. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill, 1994.
12. du Bois RM, Weycker D, Albera C, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;184:1382–9.
13. Hinkle D, Wiersma W, Jurs S. *Applied Statistics for the behavioral sciences*. Boston: Houghton-Mifflin, 1988.
14. Wyrwich KW, Nienaber NA, Tierney WM, et al. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469–78.
15. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52:861–73.
16. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582–92.
17. Russell AM, Sprangers MA, Wibberley S, et al. The need for patient-centred clinical research in idiopathic pulmonary fibrosis. *BMC Med* 2015;13:240.
18. Olson AL, Brown KK, Swigris JJ. Understanding and optimizing health-related quality of life and physical functional capacity in idiopathic pulmonary fibrosis. *Patient Relat Outcome Meas* 2016;7:29.
19. Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010;65:921–6.
20. Swigris JJ, Brown KK, Behr J, et al. The SF-36 and SGRQ: Validity and first look at minimum important differences in IPF. *Respir Med* 2010;104:296–304.
21. Gries KS, Esser D, Wiklund I. Content validity of CASA-Q cough domains and UCSD-SOBQ for use in patients with Idiopathic Pulmonary Fibrosis. *Glob J Health Sci* 2013;5:131–41.
22. Glaspole IN, Chapman SA, Cooper WA, et al. Health-related quality of life in idiopathic pulmonary fibrosis: Data from the Australian IPF Registry. *Respirology* 2017;22:950–6.
23. Kreuter M, Swigris J, Pittrow D, et al. Health related quality of life in patients with idiopathic pulmonary fibrosis in clinical practice: insights-IPF registry. *Respir Res* 2017;18:139.
24. Swigris JJ, Kuschner WG, Jacobs SS, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;60:588–94.
25. Yount SE, Beaumont JL, Chen SY, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis. *Lung* 2016;194:227–34.
26. Furukawa T, Taniguchi H, Ando M, et al. The St. George's Respiratory Questionnaire as a prognostic factor in IPF. *Respir Res* 2017;18:18.
27. Spencer S, Calverley PM, Sherwood Burge P, et al. Health status deterioration in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2001;163:122–8.
28. Kupferberg DH, Kaplan RM, Slymen DJ, et al. Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. *J Cardiopulm Rehabil* 2005;25:370–7.
29. Swigris JJ, Han M, Vij R, et al. The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respir Med* 2012;106:1447–55.
30. Paterniti MO, Bi Y, Rekić D, et al. Acute exacerbation and decline in forced vital capacity are associated with increased

mortality in idiopathic pulmonary fibrosis. *Ann Am Thorac Soc* 2017;14:1395–402.

31. Karimi-Shah BA, Chowdhury BA. Forced vital capacity in idiopathic pulmonary fibrosis--FDA review of pirfenidone and nintedanib. *N Engl J Med* 2015;372:1189–91.

32. Richeldi L, Ryerson CJ, Lee JS, *et al*. Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax* 2012;67:407–11.

33. Zappala CJ, Latsi PI, Nicholson AG, *et al*. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J* 2010;35:830–6.

34. du Bois RM, Weycker D, Albera C, *et al*. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;184:459–66.

35. du Bois RM, Nathan SD, Richeldi L, *et al*. Idiopathic pulmonary fibrosis: lung function is a clinically meaningful endpoint for phase III trials. *Am J Respir Crit Care Med* 2012;186:712–5.

36. Patel AS, Siegert RJ, Brignall K, *et al*. The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012;67:804–10.

37. Swigris JJ, Wilson SR, Green KE, *et al*. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes* 2010;8:77.

38. Yorke J, Spencer LG, Duck A, *et al*. Cross-atlantic modification and validation of the A Tool to Assess Quality of Life in Idiopathic Pulmonary Fibrosis (ATAQ-IPF-cA). *BMJ Open Respir Res* 2014;1:e000024.

39. Graney B, Johnson N, Evans CJ, *et al*. Living with idiopathic pulmonary fibrosis (L-IPF): Developing a patient-reported symptom and impact questionnaire to assess health-related quality of life in IPF. *Am J Respir Crit Care Med* 2017;195:A5353.

40. Moor CC, Heukels P, Kool M, *et al*. Integrating patient perspectives into personalized medicine in idiopathic pulmonary fibrosis. *Front Med* 2017;4:226.