Respiratory Medicine (2012) 106, 1447–1455

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/rmed

The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis

Jeffrey J. Swigris [a,*,1], Meilan Han [b,1], Rekha Vij [d,1], Imre Noth [d,1], Eric L. Eisenstein [c,1], Kevin J. Anstrom [e,1], Kevin K. Brown [a,1], Diane Fairclough [f,1]

[a] *Autoimmune Lung Center and Interstitial Lung Disease Program, National Jewish Health, 1400 Jackson Street, Denver, CO 80206, USA*
[b] *Division of Pulmonary and Critical Care Medicine, University of Michigan Health System, Ann Arbor, MI, USA*
[c] *Duke Clinical Research Institute's Outcomes Research and Assessment Group, Duke University School of Medicine, Durham, NC, USA*
[d] *Interstitial Lung Disease Program, University of Chicago Medical Center, Chicago, IL, USA*
[e] *Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA*
[f] *Department of Biostatistics and Informatics, Colorado Health Outcomes Program, University of Colorado Denver, Anshutz Medical Center, Denver, CO, USA*

Received 14 February 2012; accepted 23 June 2012
Available online 15 July 2012

KEYWORDS
Iterstitial lung disease;
Idiopathic pulmonary fibrosis;
Dyspnea;
Validity

Summary
Background: Idiopathic pulmonary fibrosis (IPF) is a progressive interstitial lung disease that often causes disabling dyspnea. In IPF and other lung diseases, patient-reported outcomes (PROs)—questionnaires designed to gather information from the patient's perspective—can determine whether therapies affect dyspnea or other outcomes meaningful to patients. Before a PRO can be used confidently as an outcome measure in a longitudinal trial, studies must demonstrate the PRO's ability to capture change over time in the target population. Our goal in this study was to examine whether the UCSD Shortness of Breath Questionnaire does so in patients with IPF.
Methods: We used data from the Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF) to perform analyses that examined associations between UCSD scores and five external measures (anchors) at baseline and over time. Anchors included the Activity domain from St. George's Respiratory Questionnaire (SGRQ-A), the Physical Functioning domain from the SF-36 (SF36-PF), forced vital capacity (FVC), diffusing capacity of the lung

* Corresponding author. Tel.: +1 303 398 1621; fax: +1 303 398 1040.
E-mail address: swigrisj@njc.org (J.J. Swigris).
[1] For the IPFnet.

http://dx.doi.org/10.1016/j.rmed.2012.06.018

for carbon monoxide (DLCO), and distance walked during a timed walk test (6MWD). Linear regression models were used to examine relationships between UCSD scores and anchors over time.
Results: At baseline, UCSD scores were weakly correlated with percent predicted FVC (-0.21, $p = 0.005$) and percent predicted DLCO (-0.20, $p = 0.008$), moderately correlated with 6MWD (-0.39, $p < 0.0001$) and strongly correlated with SGRQ-A (0.79, $p < 0.0001$) and SF36-PF (-0.72, $p < 0.0001$). Change over time in UCSD scores was associated with change in FVC (estimate = 2.54, standard error [SE] = 1.23, $p = 0.04$), SGRQ-A (estimate = 7.94, SE = 1.11, $p < 0.0001$), SF36-PF (estimate = 6.00, SE = 1.13, $p < 0.0001$), and 6MWD (estimate = 4.23, SE = 1.18, $p = 0.0004$) but not DLCO (estimate = 0.33, SE = 1.33, $p = 0.80$).
Conclusions: These results support the validity of the UCSD to assess change in dyspnea over time in patients with IPF.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive, fibrosing interstitial lung disease that causes fatigue, decreased social participation, and most prominently, disabling dyspnea and shortened survival.[1] Given the often devastating effects of IPF and the absence of effective therapies to prolong life, interest has developed in identifying ways to improve health-related quality of life (HRQL) in IPF patients. Key goals include managing dyspnea, cough and fatigue, increasing physical activity and social participation, and easing the emotional burden of living with IPF. Most measures available to assess these outcomes fall in the category of patient-reported outcome measures (PROs).

Patient-reported outcome measures attempt to quantify a person's perceptions and are administered as questionnaires or surveys, completed by patients themselves, and target outcomes like health status, HRQL or symptoms (e.g., dyspnea). Before a PRO can be used confidently as a trial endpoint in a given population, studies need to be conducted to show the PRO performs as expected in that population. For example, prior to using a PRO in a longitudinal IPF trial, existing data should demonstrate its ability to accurately capture change over time in patients with IPF.

The University of California San Diego Shortness of Breath questionnaire (UCSD) is a PRO that has been used in longitudinal research studies, including therapeutic trials in IPF patients.[2,3] Despite its extensive use in IPF, the UCSD's basic psychometric properties, including its ability to capture change in dyspnea over time, have not been established. We conducted this study to determine the UCSD's validity as a PRO capable of assessing dyspnea longitudinally (i.e., its longitudinal construct validity) in IPF.

Methods

We used data from the Sildenafil Trial of Exercise Performance in Idiopathic Pulmonary Fibrosis (STEP-IPF) which was a placebo-controlled trial designed to examine the effects of sildenafil in an IPF population with DLCO < 35% predicted.[4] Data were collected at baseline, 6, 12, 18 and 24 weeks.

The UCSD

The UCSD is a 24-item dyspnea questionnaire that asks respondents to rate themselves from 0 ("Not at all") to 5 ("Maximally or unable to do because of breathlessness") in two areas: 1) how short of breath they are while performing various activities (21 items); and 2) how much shortness of breath, fear of hurting themselves by overexerting, and fear of shortness of breath limit them in their daily lives (3 items).[5] Scores range from 0 to 120, with higher scores indicating greater dyspnea.

Anchors

We selected five measures related to dyspnea to serve as external anchors and hypothesized they would be associated with UCSD scores. These measures included: (1) forced vital capacity (FVC), (2) diffusing capacity of the lung for carbon monoxide (DLCO), (3) distance walked during the 6-min walk test (6MWD), (4) the Activity domain from St. George's Respiratory Questionnaire (SGRQ-A), and (5) the Physical Functioning domain from the Medical Outcomes Study Short-form 36-item instrument (SF36-PF). We chose FVC and DLCO because they are measures used universally to describe IPF severity.[1] Each has been shown in prior cross-sectional studies to correlate with dyspnea[6,7]; both (especially FVC) have been used as primary outcome measures in IPF trials[2,8]; and both predict survival in patients with IPF.[9–11] The 6-min walk test (6MWT), and 6MWD in particular, is commonly used as a functional assessment in patients with IPF. The SGRQ-A assesses activities that either cause or are limited by dyspnea. Scores from the SGRQ-A correlate with dyspnea in cross-sectional studies,[12–14] and prior work supports the longitudinal validity of the SGRQ-A in IPF.[15] For the SGRQ-A, higher scores indicate greater impairment. Likewise, scores from the SF36-PF correlate with dyspnea[6,16] and possess longitudinal validity as a measure of physical functioning in patients with IPF.[15] In contrast to the SGRQ, lower scores connote greater impairment for the SF-36.

Statistical analysis

We performed analyses of baseline and longitudinal data. We used Spearman correlation to examine the relationship between baseline UCSD scores and values for each of the five anchors. We used analysis of variance (ANOVA) to compare mean UCSD change scores (from baseline to 12 weeks as well as from baseline to 24 weeks) across quartiles of change within each anchor over the respective time intervals. We followed the ANOVAs with p-value-adjusted, pair-wise (parametric and then non-parametric to assess the robustness of) comparisons of mean UCSD change scores between anchor quartiles while using the quartile of greatest decline for each anchor as the reference. We also tested for the presence of linear trends in UCSD change scores over successive quartiles of change within each anchor.

Next, we used simple linear regression to examine the association between change (from baseline to 12 weeks and from baseline to 24 weeks) in UCSD score and change in anchor values over the same timeframes. After interrogating the data set and omitting outliers with highly influential values, we re-ran these analyses. We then used mixed-effects models to further examine these associations longitudinally, across all study time points. Specifically, for each anchor, we generated a mixed-effects model with UCSD score as the outcome variable and the anchor and time as predictors. In these models, we dichotomized the anchor at a meaningful cut-point (stable/improved vs. worsened compared with baseline) and tested the null hypothesis that UCSD scores would not differ between the dichotomized subgroups within an anchor over the duration of the study. For FVC, we considered a decline in the raw value of 7% or greater as "worsened" and a decline of less than 7% as "stable/improved." Our selection of 7% stems from emerging data suggesting that small changes in FVC are meaningful and prognostically important.[17,18] For DLCO, we considered a decline in the raw value of 15% or greater as "worsened" and a decline of less than 15% as "stable/improved."[15] For 6MWD, for certain analyses, we considered a decline of 20% or greater as "worsened" and a lesser decline or improvement as "stable/improved," and for other analyses, we used the IPF-specific minimum important difference (MID) estimate of 30 m as a cut-point.[19–21] For the SGRQ-A and SF36-PF, we used as cut-points published estimates for their IPF-specific MIDs (5 and 3 points respectively).[15] In the mixed-effects models, we used an unstructured covariance matrix (type = un option in SAS PROC MIXED) to model the correlation structure of the repeated measure.

Next, as a visual representation of the relationship between UCSD and SGRQ-A or SF36-PF, we generated cumulative distribution function (CDF) plots for the UCSD with these anchors using the 24-week data. Finally, we generated MID estimates for the UCSD using distribution- and anchor-based approaches. For the distribution-based approach, we used the effect size, standardized response mean and standard error of measurement,[22] and for the anchor-based approach, we applied the within-patient anchor method using the SGRQ-A.[23] Institutional review board approval was not required to perform these analyses on de-identified, previously collected data. All statistical analyses were run using SAS, Version 9.2 (SAS, Inc.; Cary, NC).

Results

Most subjects were male, and the average duration of IPF was two years (Table 1). At baseline, mean UCSD scores were weakly correlated with percent predicted FVC (FVC%) and percent predicted DLCO (DLCO%), moderately correlated with 6MWD, and strongly correlated with SGRQ-A and SF36-PF scores (Table 2). At 12 and 24 weeks, UCSD change scores were associated with (and in the hypothesized direction of) change scores for each of the five anchors (Table 3). After omitting outliers, as directed by the DFBETA measure, the results remained the same, except for the FVC at 12 weeks (beta coefficient −2.4, SE 1.5, $p = 0.1$). From baseline to week 12, UCSD change scores increased linearly along quartiles of increasing impairment in SGRQ-A and SF36-PF ($p < 0.0001$ and $p = 0.0002$ for linear contrasts for the respective anchor). There were trends toward statistically significant linear increases in UCSD change scores along quartiles of change in 6MWD and quartiles of change in FVC over the same time period (Table 4). From baseline to week 24, UCSD change scores increased linearly along quartiles of increasing impairment in SGRQ-A and SF36-PF and along quartiles of change in 6MWD. UCSD change scores trended toward statistically significant linear increases along quartiles of change in FVC and DLCO over the same time period. The non-parametric analyses yielded the same results, except the comparison between the two extreme quartiles for 6MWD at 24 was not significant ($p = 0.1$).

Results from the longitudinal analysis are comparable: a significant change in UCSD scores was seen in individuals who experienced clinically significant changes in FVC,

Table 1 Baseline characteristics of STEP-IPF subjects.

Characteristic	Total subjects = 180
Age, yrs	69.0 (9.0)
Male, %	83
Time since diagnosis, yrs	2.0 (1.9)
Smoking status, %	Current 0 Former 76 Never 24
FVC, liters	2.3 (0.7)
FVC%	56.8 (14.2)
DLCO, ml/min/mmHg	7.8 (2.1)
DLCO%	26.3 (6.1)
6MWD, meters	265.0 (117.1)
UCSD SOB questionnaire	47.0 (21.4)
SGRQ activity domain	69.6 (17.6)
SF36 PF domain[a]	30.2 (8.4)

Data are percentages or mean (standard deviation); FVC = forced vital capacity; FVC% = percent of predicted FVC for gender, age, and height; DLCO = diffusing capacity of the lung for carbon monoxide; DLCO% = percent of predicted DLCO for gender, age and height; 6MWD = distance walked during 6-min walk test; UCSD = University of California San Diego; SGRQ = St. George's Respiratory Questionnaire; SF36 = Medical Outcomes Study Short-form 36-item Instrument; PF = physical functioning.

[a] Standardized scoring (for the US general adult population, this domain has a mean score of 50 with a standard deviation of 10).

Table 2 Mean UCSD scores per anchor quartile and correlations between baseline UCSD and anchor scores.

Anchor	*N*	UCSD	Correlation (*p*-value)
FVC%			−0.22 (0.003)
>68	43	43.0 (18.2)	
56−68	46	43.8 (22.8)	
46−56	45	48.3 (21.3)	
<46	45	52.7 (22.1)	
DLCO%			−0.20 (0.007)
>31	43	42.3 (21.4)	
27−31	42	45.3 (22.6)	
22−27	49	45.2 (18.7)	
<22	46	54.9 (21.4)	
6MWD, m			−0.39 (<0.0001)
>355	45	36.5 (17.9)	
263−355	45	42.9 (21.6)	
181−262	45	51.3 (20.8)	
<181	45	57.4 (19.6)	
SGRQ activity domain			0.80 (<0.0001)
<58	45	26.5 (13.8)	
58−71	44	38.5 (10.9)	
71−86	44	51.9 (13.3)	
>86	46	70.4 (16.3)	
SF36 PF domain[a]			−0.72 (<0.0001)
>36	51	28.5 (16.4)	
30−36	48	43.1 (11.6)	
23−30	46	55.2 (16.7)	
<23	34	69.0 (18.5)	

Data are percentages or mean (standard deviation); FVC = forced vital capacity; FVC% = percent of predicted FVC for gender, age, and height; DLCO = diffusing capacity of the lung for carbon monoxide; DLCO% = percent of predicted DLCO for gender, age and height; UCSD = University of California San Diego; SGRQ = St. George's Respiratory Questionnaire; SF36 = Medical Outcomes Study Short-form 36-item Instrument; PF = physical functioning.

[a] Standardized scoring (for the US general adult population, this domain has a mean score of 50 with a standard deviation of 10).

Table 3 Results of linear regression analyses showing association between change in UCSD and change in anchors (as continuous variables unless otherwise specified) over time.

Baseline to Week 12			Baseline to Week 24		
Anchor	Estimate (SE)	*p*	Anchor	Estimate (SE)	*p*
FVC: per 10% change from baseline in raw FVC *N* = 160	−3.96 (1.46)	0.007	FVC: per 10% change from baseline in raw FVC *N* = 138	−4.78 (1.58)	0.003
DLCO: per 10% change from baseline in raw DLCO *N* = 151	−1.19 (0.48)	0.02	DLCO: per 10% change from baseline in raw FVC *N* = 135	−1.15 (0.52)	0.04
6MWD: per 100m change from baseline 6MWD *N* = 164	−2.86 (1.36)	0.04	6MWD: per 100m change from baseline 6MWD *N* = 140	−2.81 (1.41)	0.04
SGRQ-A: per one point change from baseline in SGRQ-A score *N* = 162	0.62 (0.09)	<0.0001	SGRQ-A: per one point change from baseline in SGRQ-A score *N* = 139	0.56 (0.08)	<0.0001
SF36-PF[a]: per one point change from baseline in SF36-PF score *N* = 163	−0.75 (0.21)	0.0004	SF36-PF[a]: per one point change from baseline in SF36-PF score *N* = 140	−1.03 (0.18)	<0.0001

SE = standard error; FVC = forced vital capacity; DLCO = diffusing capacity of the lung for carbon monoxide corrected for hemoglobin; 6MWD = distance walked during the 6-min walk test; UCSD = University of California San Diego; SGRQ = St. George's Respiratory Questionnaire; SF36 = Medical Outcomes Study Short-form 36-item Instrument; PF = physical functioning.

[a] Standardized scoring (for the US general adult population, this domain has a mean score of 50 with a standard deviation of 10).

Table 4 Change from baseline to week 12 (left) or week 24 (right) in UCSD scores according to quartiles of change from baseline to week 12 or week 24 for each anchor.

Baseline to Week 12			
Quartiles of change in anchors	*N*	Change in UCSD	*p*
Raw FVC			0.17[b]
Increase by 2.2% or more	44	1.7 (16.8)	0.18
Increase by 2.1%—drop by 2.7%	45	2.4 (17.4)	0.23
Drop 2.8—7.0%	39	3.7 (14.0)	0.41
Drop by 7.1% or more	41	7.0 (21.2)	Reference
Raw DLCO			0.60[b]
Increase by 11.0% or more	42	4.8 (23.5)	0.42
Increase by 10.9%—drop by 0.6%	45	2.7 (11.5)	0.16
Drop by 0.7—10.9%	43	0.3 (14.1)	0.03
Drop by 11% or more	39	8.1 (19.6)	Reference
6MWD			0.08[b]
Increase by 12 m or more	44	1.5 (17.9)	0.02
Increase by 11—drop by 19 m	41	6.8 (18.5)	0.88
Drop by 20—82 m	42	2.4 (14.4)	0.24
Drop by 83 m or more	42	6.9 (18.4)	Reference
SGRQ Activity Score			<0.0001+
Drop by 6.0 points or more	55	−5.2 (15.4)	<0.0001
Drop 0.0—5.9 points	42	3.1 (13.3)	0.001
Increase 0.1—7.3 points	31	4.9 (17.8)	0.01
Increase by 7.4 points or more	41	14.4 (18.0)	Reference
SF36 PF Domain[a]			0.0002[b]
Increase by 2.6 points or more	36	−3.3 (12.8)	<0.0001
Increase by 2.5—drop by 1.2 points	46	1.7 (11.4)	0.003
Drop by 1.1—4.0 points	30	0.9 (15.3)	0.005
Drop by 4.1 points or more	52	11.8 (22.7)	Reference
Quartiles of change in anchors	*N*	Change in UCSD	*p*
Raw FVC			0.08[b]
Increase by 1.1% or more	38	5.4 (15.4)	0.20
Increase by 1.0%—drop by 4.9%	38	−1.2 (11.6)	0.003
Drop by 5.0—9.7%	35	5.3 (16.6)	0.20
Drop by 9.8% or more	35	10.3 (18.9)	Reference
Raw DLCO			0.16[b]
Increase by 8.0% or more	33	0.9 (13.4)	0.09
Increase by 7.9%—drop by 2.6%	35	6.4 (19.1)	0.78
Drop by 2.6—4.4%	37	4.2 (12.5)	0.40
Drop by 14.5% or more	36	7.4 (18.3)	Reference
6MWD			0.02[b]
Increase by 5.0 m or more	41	−0.9 (11.2)	0.03
Increase by 4.9—drop by 36 m	34	4.5 (12.3)	0.49
Drop by 36—114 m	35	8.8 (16.0)	0.65
Drop by 114 m or more	36	7.1 (21.7)	Reference
SGRQ Activity Score			<0.0001[b]
Drop by 6.2 points or more	32	−6.1 (12.6)	<0.0001
Drop by 6.2—increase by 2.3 points	36	4.3 (12.2)	0.003
Increase by 2.3—12.0 points	36	5.0 (17.2)	0.006
Increase by 12.1 points or more	37	14.5 (15.5)	Reference

(*continued on next page*)

Table 4 (*continued*)

Quartiles of change in anchors	*N*	Change in UCSD	*p*
SF36 PF domain[a]			<0.0001[b]
Increase by 2.2 points or more	34	−4.9 (13.8)	<0.0001
Increase by 2.1—drop by 2.0 points	41	2.8 (11.0)	0.0001
Drop by 2.1—6.2 points	31	3.2 (12.5)	0.0005
Drop by 6.3 points or more	40	15.6 (19.0)	Reference

Footnote Data are percentages or mean (standard deviation); FVC = forced vital capacity; DLCO = diffusing capacity of the lung for carbon monoxide; UCSD = University of California San Diego; SGRQ = St. George's Respiratory Questionnaire; SF36 = Medical Outcomes Study Short-form 36-item Instrument; PF = physical functioning.

[a] Standardized scoring (for the US general adult population, this domain has a mean score of 50 with a standard deviation of 10).

[b] For overall linear effect.

6MWD, SGRQ-A and SF36-PF but not DLCO over the duration of the study (Table 5). Graphs of the UCSD cumulative distribution function for the SGRQ-A and SF36-PF anchors are presented in Fig. 1. Among subjects with UCSD change scores showing any improvement ($\leq$0 on the *x*-axis of the figures), a greater proportion had improved SGRQ-A (Panel A) or SF36-PF (Panel B) than stable/worsened SGRQ-A or SF36-PF. The MID estimates for the UCSD ranged from 5 to 11, with a point estimate of 8 (Table 6).

Discussion

Using data from the STEP-IPF trial, we conducted several analyses whose results support the construct validity of the UCSD as an instrument capable of assessing and capturing change over time in dyspnea in patients with IPF.

Choice of anchors ideally implies that there is a gold standard for the construct of interest. This is rarely the case in the assessment of symptoms, and researchers must rely on either clinical or other patient-reported measures that are related to the symptom of interest to serve as anchors. In IPF, there are inextricable links between pulmonary physiology, functional capacity, HRQL and dyspnea: 1) as the impairment in pulmonary physiology worsens, dyspnea increases; 2) functional capacity (assessed as either the ability to complete tasks or the rate at which they are completed) depends greatly on the level of dyspnea; and 3) dyspnea is a potent driver of HRQL and health status in patients with IPF.[12,24] Hence, in the absence of a gold standard, we believe FVC, DLCO, 6MWD, SGRQ-A and SF36-PF were suitable anchors for this study.

Table 5 Results of mixed effects models showing association between change in UCSD and worsening in anchors over time.

Anchor	Estimate (SE)	*p*
FVC decline > 7%	2.54 (1.23)	0.04
DLCO decline > 15%	0.33 (1.34)	0.80
6MWD decline > 30 m	3.32 (1.24)	0.008
SGRQ Activity Score worsening ≥ 5 points	7.94 (1.13)	<0.0001
SF36 PF Domain[a] decline ≥ 3 points	6.00 (1.14)	<0.0001

SE = standard error; FVC = forced vital capacity; DLCO = diffusing capacity of the lung for carbon monoxide corrected for hemoglobin; 6MWD = distance walked during the 6-min walk test; UCSD = University of California San Diego; SGRQ = St. George's Respiratory Questionnaire; SF36 = Medical Outcomes Study Short-form 36-item Instrument; PF = physical functioning.

[a] Standardized scoring (for the US general adult population, this domain has a mean score of 50 with a standard deviation of 10).

From the patient's perspective, if a therapeutic intervention could improve only one thing in IPF, it would be dyspnea.[24] To assess whether therapies have beneficial (or adverse) effects on dyspnea, reliable, valid measurement instruments, sensitive to underlying change in dyspnea, are needed. Results from our analyses support the use of the UCSD as such an instrument. In the simple linear regression analyses, UCSD scores were significantly related to each of the five anchors. Despite a loss of power, in the ANOVA analyses, the UCSD was able to discriminate between subjects who remained stable (or improved) and those who declined according to the SGRQ-A and SF36-PF at 12 and 24 weeks. From the mixed-effects model, a decline from baseline in raw FVC of 7% or greater (at 6, 12, 18 or 24 weeks) or in 6MWD of 20% or greater (at 6, 12, 18 or 24 weeks) was associated with a 2.5 or 4.2 point increase respectively in UCSD score. Likewise, HRQL or health status worsening at any study time point, defined as change from baseline by more than the IPF-specific MID for the SGRQ-A (five points) or the SF36-PF (three points), was associated with a nearly 8-point or 6-point respective increase in UCSD score. The CDF graphs provide a different way to look at UCSD data: a far greater proportion of subjects with stable/improved UCSD scores had stable/improved, as opposed to worsened, anchor scores. These types of results—where the UCSD can track changes in IPF anchors—support the longitudinal validity of the UCSD as an instrument capable of assessing dyspnea over time in IPF patients.

The UCSD is one of several PROs developed to assess dyspnea. Before investigators can use any PRO confidently, they need to confirm data exist to support its validity for a specific purpose, and importantly, in the target population. A PRO has validity if its scores can be used to make accurate inferences about a patient in the target population. A mistake that has been perpetuated in the medical literature is that one cross-sectional, correlation study "validates" an instrument for use in a longitudinal trial.[25] Data from such studies may support the so-called "concurrent validity" of the instrument but are not sufficient to confirm its ability to assess change in longitudinal

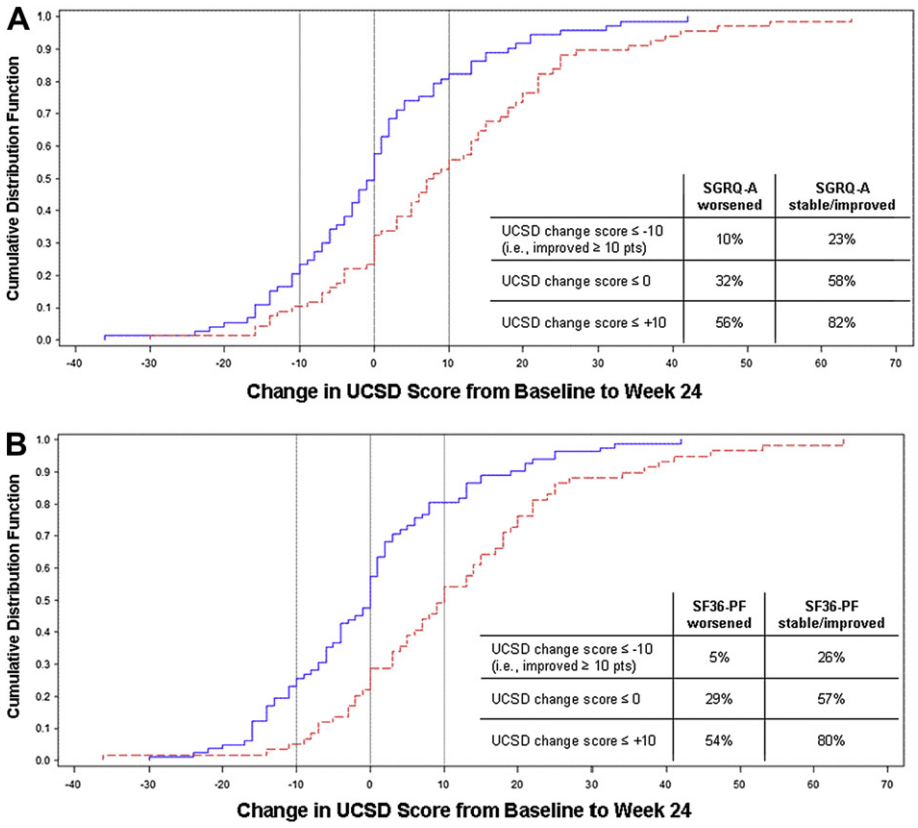

Figure 1 **Panel A.** Graph of cumulative distribution function for UCSD Change Scores Anchored on Dichotomized SGRQ-A Change Score at 24 weeks, *Footnote*: SGRQ-A = Activity domain from St. George's Respiratory Questionnaire; Solid line = SGRQ-A stable/improved; dashed line = SGRQ-A worsened, **Panel B.** Graph of cumulative distribution function for UCSD Change Scores Anchored on Dichotomized SF36-PF Change Score at 24 weeks, *Footnote*: SF36-PF = Physical Functioning domain from the SF-36 questionnaire; Solid line = SF36-PF stable/improved; dashed line = SF36-PF worsened.

studies.[26–28] One of the key (and anticipated) results from our study is this: patients with IPF who experience disease progression report greater dyspnea, as measured by the UCSD. For example, a decline in raw FVC of 7% or greater corresponds to a 2.5-point increase in UCSD score. Another important conclusion is that patients who report a decline in physical functioning will also develop greater dyspnea, as measured by the UCSD. Although these findings are intuitively obvious, validity is built only after performing analyses, such as these, that confirm the instrument performs as hypothesized.

It is possible that other dyspnea questionnaires would yield similar information under the same circumstances, but until their longitudinal construct validity is assessed, they cannot be used confidently in longitudinal IPF research. This is the first study to systematically assess the longitudinal construct validity of any dyspnea questionnaire in IPF. However, data from a recent study examining the psychometric properties of FVC in IPF further support our findings.[17] In that study, 24-week change in FVC was significantly correlated with 24-week change in UCSD score ($r = -0.25$, $p < 0.001$).

Recently, the FDA began formalizing recommendations for how a PRO might qualify (for the drug indication approval process) as a valid, reliable outcome measure whose scores have "interpretable meaning" and one whose utility need not be reconfirmed when used again in the same target population.[29] To our knowledge, there are no FDA-qualified PROs for IPF; further, we are unaware if the UCSD is being considered by the FDA for qualification in IPF. Data reported here would strongly support such a submission.

Table 6 Minimum important difference (MID) estimates for the UCSD.

0.5 ES	0.5 SRM	$^1/_3$ SD	1-SEM	SGRQa at 12 weeks	SGRQa at 24 weeks	MID
10.7	12 weeks 8.8	7.1	4.8	ΔUCSD = 3.1 + 0.6(ΔSGRQa) For 5–11pt Δ in SGRQa, UCSD Δ 6–10	ΔUCSD = 2.2 + 0.6(ΔSGRQa) 9For 5–11pt Δ in SGRQa, UCSD Δ 5–9	8 Range
	24 weeks 8.0					5–11

ES = effect size; SD = standard deviation; SEM = standard error of measurement; UCSD = University of California San Diego shortness of breath questionnaire; SGRQ = St. George's Respiratory Questionnaire; MID = minimum important difference.

This study has limitations. The first is that subjects in the STEP-IPF trial—as required for inclusion—had DLCOs $<$ 35% predicted. Thus, our results may not apply to IPF patients with milder disease. However, it seems likely that, regardless of disease severity at baseline, increasing IPF severity, or declining HRQL or health status, would lead to increasing dyspnea (and thus worse UCSD scores). Some of the analyses yielded results that were not statistically significant. This is likely due to the variability in measurement of the anchors; for example, DLCO has been recognized by other investigators as statistically noisy, and that limits the utility of DLCO as a reliable endpoint in IPF trials.[18] Additionally the exclusion of subjects with DLCO $>$ 35% predicted in STEP-IPF may have introduced floor effects for this anchor.

Furthermore, dyspnea is a very complex symptom driven by multiple inputs. Ventilatory restriction and impaired oxygen diffusion are two (of many) related, but somewhat independent drivers of dyspnea. This likely explains the only modest associations between FVC or DLCO and UCSD scores at baseline. The results of our study cannot be extended to other dyspnea questionnaires; similar studies would be required to assess their longitudinal construct validity in IPF. Despite our results supporting the UCSD in IPF, it is predominantly focused on how short of breath one gets with various activities. Thus, investigators should examine its item content to determine if the UCSD is right for their particular study. For example, of the UCSD's 24 items, only two address any aspect of the effect of dyspnea on mental well-being; if one were interested in determining the impact of a given treatment on mental health, a distinct mental health metric would be required. Ours are the first MID estimates for the UCSD in IPF. The MID should be calculated in other samples to increase confidence in our results. For now, we believe 8 is a reasonable point estimate to use for group-level analyses, and 11 should be used for individual responder analyses.

Conclusion

We used data from a recently completed IPF trial to examine the ability of the UCSD to assess dyspnea longitudinally. Our results support the use of the UCSD in IPF trials; however, it is important to recognize that the UCSD's primary focus is on dyspnea associated with physical activity.

Funding source

The project described was supported by grants from the Supported by grants from the NHLBI: U10HL080413 (data coordinating center), U10HL080274, U10HL080370, U10HL080371, U10HL080383, U10HL080411, U10HL080509, U10HL080510, U10HL080513, U10HL080543, U10HL080571, U10HL080685 (clinical centers). Dr. Swigris is supported in part by a Career Development Award from the NIH (K23 HL092227).

Conflict if interest

None declared.

Acknowledgments

In addition to the authors, the following investigators and clinical centers participated in the STEP-IPF Trial: **Cleveland Clinic**, Cleveland, OH: Jeffrey Chapman, MD, Mitch Olman, MD, Susan Lubell, RN. **Duke University Medical Center**, Durham, NC: Lake D. Morrison, MD, Mark P. Steele, MD, Terri Haram, RN. **Emory University**, Atlanta, GA: Jesse Roman, MD, Rafael Perez, MD, Tamra Perez, RN. **Mayo Clinic**, Rochester, MN: Jay H. Ryu, MD, James P. Utz, MD, Andrew H. Limper, MD, Craig E. Daniels, MD, Kathleen Meiras, Suson Walsh. **National Jewish Health**, Denver, CO: Kevin K. Brown, MD, Marvin Schwarz, MD, Carol Bair, CRT, CCRC, Dolly Kervitsky, CRT, CCRC. **Tulane University**, New Orleans, LA: Joseph A. Lasky, MD, Sandy Ditta, RN. **University of Alabama, Birmingham**: Joao de Andrade, MD, Victor J. Thannickal, MD, Mae Stewart. **University of California, Los Angeles**: David A. Zisman, MD, Joseph Lynch, III, MD, Eileen Calahan, RN, Paul Lopez, RN. **University of California, San Francisco**: Talmadge E. King, Jr., MD, Harold R. Collard, MD, Jeffrey A. Golden, MD, Paul J. Wolters, MD, Renee Jeffrey, RN. **University of Chicago**, Chicago, IL: Imre Noth, MD, D. Kyle Hogarth, MD, Nathan Sandbo, MD, Mary E. Strek, MD, Steven R. White, MD, Cathy Brown, RN, Irena Garic, Spring Maleckar. **University of Michigan, Ann Arbor**: Fernando J. Martinez, MD, Kevin R. Flaherty, MD, MeiLan Han, MD, Bethany Moore, MD, Galen B. Toews, MD, Debra Dahlgren, RN. **University of Washington, Seattle**: Ganesh Raghu, MD, Jennifer Hayes, RN, Margaret Snyder, RN. **Vanderbilt University**, Nashville, TN: James E. Loyd, MD, Lisa Lancaster, MD, William Lawson, MD, Rhonda Greer, Wendi Mason, NP. **Weill Medical College of Cornell University**, New York, NY: Robert J. Kaner, MD, Vanessa Monroy, RN, Mei Wang, RN. **Core Lab Chairs**: David Lynch, MD (Radiology) at National Jewish Health, Denver, CO, Thomas Colby, MD (Pathology) at Mayo Clinic, Scottsdale, AZ. **Data Coordinating Center**: Duke Clinical Research Institute, Duke University Medical Center, Durham, NC - Kevin J. Anstrom, PhD, Richard C. Becker, MD, Eric L. Eisenstein, DBA, Neil R. MacIntyre, MD, Lake D. Morrison, MD, James Rochon, PhD, Mark P. Steele, MD, John S. Sundy, MD, Linda Davidson-Ray, MA, Patricia Dignacco, BSMT, Rex Edwards, BA, Robert Anderson, BA, Rosemary Beci, BS, Sara Calvert, PharmD, Kristy Cain, MA, Tedryl Gentry-Bumpass, BSBE, David Hill, MBA, Marc Ingham, JD, Ethel Kagan, RN, BA, Jasvinder Kaur, RN, Clare Matti, MA, Jennifer McClelland, RCP, Allyn Meredith, MA, Taylor Nguyen, BS, Jean Pesarchick, RN, BSN, Rhonda S. Roberts, MSPH, Wanda Tate, BA, Tony Thomas, JD, Jenny Walker, MLS, Doug Whelan, BBA, Jane Winsor, BA, Qinghong Yang, MPH, Eric Yow, MS. **NHLBI Representatives**: Herbert Y. Reynolds, MD; Xin Tian, PhD; James Kiley, PhD

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. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011;**377**:1760—9.

3. Raghu G, Brown K, Bradford W, et al. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *New Engl J Med* 2004;**350**:125—33.
4. Zisman DA, Schwarz M, Anstrom KJ, et al. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 2010;**363**:620—8.
5. Eakin EG, Resnikoff PM, Prewitt LM, et al. In: . *Validation of a new dyspnea measure: the UCSD shortness of breath questionnaire*, vol. 113. San Diego. Chest: University of California; 1998. p. 619—24.
6. Martinez T, Pereira C, dos Santos M, et al. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with idiopathic pulmonary fibrosis. *Chest* 2000;**117**:1627—32.
7. Martinez J, Martinez T, Galhardo F, et al. Dyspnea scales as a measure of health-related quality of life in patients with idiopathic pulmonary fibrosis. *Med Sci Mon* 2002;**8**:CR405—10.
8. King Jr TE, Brown KK, Raghu G, et al. BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;**184**:92—9.
9. Collard HR, King Jr TE, Bartelson BB, et al. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;**168**:538—42.
10. Flaherty KR, Andrei AC, Murray S, et al. Idiopathic pulmonary fibrosis: prognostic value of changes in physiology and six-minute-walk test. *Am J Respir Crit Care Med* 2006;**174**:803—9.
11. Flaherty KR, Mumford JA, Murray S, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003;**168**:543—8.
12. Nishiyama O, Taniguchi H, Kondoh Y, et al. Health-related quality of life in patients with idiopathic pulmonary fibrosis. What is the main contributing factor? *Respir Med* 2005;**99**:408—14.
13. Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. *Chest* 2005; **127**:284—94.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. du Bois RM, Weycker D, Albera C, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;**183**:1231—7.
20. Holland AE, Hill CJ, Conron M, et al. Small changes in six-minute walk distance are important in diffuse parenchymal lung disease. *Respir Med* 2009.
21. Swigris JJ, Wamboldt FS, Behr J, et al. The 6-minute walk in idiopathic pulmonary fibrosis: longitudinal changes and minimum important difference. *Thorax* 2010;**65**:173—7.
22. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes* 2003;**1**:4.
23. Schunemann HJ, Griffith L, Jaeschke R, et al. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *J Clin Epidemiol* 2003;**56**:1170—6.
24. Swigris JJ, Stewart AL, Gould MK, et al. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005;**3**:61.
25. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000;**9**: 887—900.
26. Streiner D, Norman G. *Health measurement scales: a practical guide to their development and use*. 4th ed. New York: Oxford University Press; 2008.
27. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1992;**1**:73—5.
28. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care* 2000;**38**:II14—25.
29. Food and Drug Administration Center for Drug Evaluation and Research. *Guidance for industry: qualification process for drug development tools*. MD: Silver Spring; 2010.