Title

Reliability and Minimal Clinically Important Differences of FVC. Results from the Scleroderma Lung Studies (SLS-I and SLS-II)

Journal

American Journal of Respiratory and Critical Care Medicine, 197(5)

ISSN

1073-449X

Authors

Kafaja, Suzanne

Clements, Philip J

Wilhalme, Holly

et al.

Publication Date

2018-03-01

DOI

10.1164/rccm.201709-1845oc

Peer reviewed

ORIGINAL ARTICLE

Reliability and Minimal Clinically Important Differences of FVC
Results from the Scleroderma Lung Studies (SLS-I and SLS-II)

Suzanne Kafaja[1], Philip J. Clements[1], Holly Wilhalme[1], Chi-hong Tseng[1], Daniel E. Furst[1], Grace Hyun Kim[2], Jonathan Goldin[2], Elizabeth R. Volkmann[1], Michael D. Roth[1], Donald P. Tashkin[1], and Dinesh Khanna[3]

[1]Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, California; [2]Department of Radiological Sciences, University of California, Los Angeles, California; and [3]Department of Medicine, University of Michigan, Ann Arbor, Michigan

Abstract

Rationale: FVC percent predicted (FVC%) is the primary outcome measure in clinical trials of systemic sclerosis interstitial lung disease. For interpretation of change in the FVC% over time, it is important to define whether these changes are clinically meaningful.

Objectives:: To assess the reliability and the minimal clinically important differences (MCID) for FVC% in the Scleroderma Lung Study I and II (SLS-I and -II).

Methods: Using data from SLS-I and -II ($N = 300$), we evaluated the test-retest reliability for FVC% (screening vs. baseline) using intraclass correlation. MCID estimates at 12 months were calculated in the pooled cohort (SLS-I and -II) using two anchors: Transition Dyspnea Index (⩾change of 1.5 units for improvement and worsening, respectively) and the Medical Outcomes Short Form-36 Health Transition question ("Compared with one year ago, how would you rate your health in general now"?), where "somewhat better" or "somewhat worse" were defined as the MCID estimates. We next assessed the association of MCID estimates for improvement and worsening of FVC% with patient-reported outcomes (PROs) and computer-assisted quantitation of extent of fibrosis (QLF) and of total interstitial lung disease (QILD) on high-resolution computed tomography. Student's *t* test was used to compare the mean difference in outcomes between the MCID improvement/worsening and the "no change" group.

Measurements and Main Results: Reliability of FVC%, assessed at a mean of 34 days, intraclass correlation was 0.93 for the pooled cohort. The MCID estimates for the pooled cohort at 12 months for FVC% improvement ranged from 3.0% to 5.3% and for worsening from −3.0% to −3.3%. FVC% improvement by greater than or equal to MCID was associated with either statistically significant or numerical improvements in some PROs, QILD, and QLF, whereas FVC% worsening greater than or equal to MCID was associated with statistically significant or numerical worsening of PROs, QILD, and QLF.

Conclusions: FVC% has acceptable test-retest reliability, and we have provided the MCID estimates for FVC% in systemic sclerosis interstitial lung disease–based changes at 12 months from baseline in two clinical trials.

Clinical trial registered with www.clinicaltrials.gov (NCT00004563 for SLS-I and NCT00883129 for SLS-II).

Keywords: interstitial lung disease; systemic sclerosis; FVC%; minimal clinically important differences; patient-reported outcomes

(Received in original form September 11, 2017; accepted in final form November 2, 2017)

Lists of the members of SLS-I and SLS-II can be found before the beginning of the References.

D.K., H.W., and C.-h.T. were funded by K24 NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) AR063120 (D.K.) and NIH/NIAMS R01AR070470 (D.K.). SLS-I was supported by grants from the NIH (U01 HL60587, U01 HL60606, and R01 HL089758). SLS-II was supported by grants from the NHLBI/NIH (R01 HL089758 and R01 HL089901). Study drug (cyclophosphamide) was supplied by Bristol-Myers Squibb for use in SLS-I and the study drug (mycophenolate) and matching placebo were supplied at no charge through Drug Supply Grant #CEL539 from Hoffmann-La Roche/Genentech in SLS-II.

Author Contributions: Concept and design, S.K., P.J.C., D.E.F., J.G., M.D.R., D.P.T., and D.K. Analysis and interpretation, S.K., P.J.C., H.W., C.-h.T., D.E.F., G.H.K., J.G., E.R.V., M.D.R., D.P.T., and D.K. Drafted manuscript for important intellectual content, S.K., P.J.C., D.P.T., and D.K.

Correspondence and requests for reprints should be addressed to Dinesh Khanna, M.D., M.Sc., University of Michigan Scleroderma Program, Division of Rheumatology, Department of Internal Medicine, 300 North Ingalls Street, SPC 5422, Ann Arbor, MI 48109. E-mail: khannad@med.umich.edu.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 197, Iss 5, pp 644–652, Mar 1, 2018

Originally Published in Press as DOI: 10.1164/rccm.201709-1845OC on November 3, 2017
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: FVC percent predicted (FVC%) is the primary outcome measure in most clinical trials of systemic sclerosis interstitial lung disease. For interpretation of change in the FVC% within a group of subjects with systemic sclerosis interstitial lung disease over time or differences between two groups, it is important to define whether these changes are clinically meaningful.

What This Study Adds to the Field: Using the data from the Scleroderma Lung Study I and II, the minimal clinically important difference estimates for the FVC% improvement ranged from 3.0% to 5.3% and for worsening from −3.0% to −3.3%. These changes were associated with statistically significant or numerical changes in patient-reported outcome measures and high-resolution computed tomography changes.

Systemic sclerosis (SSc; scleroderma) is a multiorgan disease with a complex interplay among inflammation, fibrosis, and vasculopathy. Although organ involvement in SSc varies, lung involvement is one of the leading causes of morbidity and mortality (1). SSc-related interstitial lung disease (SSc-ILD), therefore, has received a prime clinical and therapeutic emphasis in SSc patient care and clinical research, and evaluating lung physiology has gained a central role in clinical trials of SSc. FVC percent predicted (FVC%) is the primary outcome measure in most clinical trials of SSc-ILD (2–4), including the Scleroderma Lung Study (SLS)-I and SLS-II (5, 6).

For interpretation of change in the FVC% within a group of subjects with SSc-ILD over time or differences between two groups, it is important to define whether these changes are clinically meaningful. The minimal clinically important difference (MCID) is defined as the smallest difference in a measure or instrument of interest that is considered to be "worthwhile or important" to the patient (7). For the clinician, MCID helps guide treatment. Although there are several methods for calculating MCID, the estimate of MCID using an external anchor method is often preferred over other methods (8). In this article, we analyzed data from two clinical trials in SSc-ILD (SLS-I and SLS-II) to assess the test-retest reliability of FVC% and to calculate the MCID estimates of FVC% in subjects with SSc-ILD.

Methods

Subjects

All subjects with any 12-month follow-up outcome data in SLS-I and -II were evaluated in this *post hoc* analysis. The study protocols for both SLS-I and -II were approved by the local institutional review boards, and written informed consent was obtained from all subjects. The trial designs for both SLS-I and -II have been published elsewhere (5, 6). Briefly, subjects meeting the 1980 SSc classification criteria were included if they had changes on high-resolution computed tomography (HRCT) consistent with SSc-ILD, including ground-glass opacity, and symptoms of breathlessness (grade 2 on the Functional Impairment domain of the Mahler Baseline Dyspnea Index). Subjects in SLS-I were randomized to 1 year of oral placebo or oral cyclophosphamide, with the primary endpoint being change from baseline in FVC% at 1 year, whereas subjects in SLS-II were randomized to 2 years of mycophenolate mofetil or 1 year of oral cyclophosphamide followed by 1 year of placebo. The primary endpoint for SLS-II was the course of the FVC% from baseline to 24 months using a joint model, which examined the repeated measurements of FVC%.

Methods and Procedures

Subjects' clinical data included age, sex, race, disease duration (from first non-Raynaud symptom attributable to SSc), skin subtype of SSc (diffuse or limited cutaneous), and the modified Rodnan Skin Score (mRSS). The patient-reported outcome measures (PROs) included the Mahler Transition Dyspnea Index, the Health Assessment Questionnaire disability index, and the Medical Outcomes Short Form-36 (SF-36) in SLS-I and -II (9). In addition, SLS-II PROs included the Leicester Cough Questionnaire and the St. George's Respiratory Questionnaire (SGRQ).

Dyspnea. Dyspnea was assessed using Baseline Dyspnea Index (BDI) and Transition Dyspnea Index (TDI) (10). TDI measured the change from baseline, and ranged from −3 to +3 for three domains, for a sum ranging between −9 and +9. Higher positive scores connote less dyspnea. Although BDI and TDI were assessed in SLS-I using a paper questionnaire, both indices were assessed in SLS-II using a self-administered, computer-generated format.

SGRQ. SGRQ is a self-administered questionnaire for assessing health-related quality of life (HRQoL) in respiratory diseases (11). It contains 50 items distributed over three scales. A total score, from 0 to 100, is the weighted average of these three subscores. SGRQ differentiates between SSc with and without interstitial lung disease (ILD) (12).

Health Assessment Questionnaire disability index. Health Assessment Questionnaire disability index is a 20-item questionnaire assessing functional disability in eight domains. Scores range from 0.0 (best) to 3.0 (worst) (13). It is validated in SSc (14).

SF-36 version 2. SF-36 version 2 is a self-administered survey focusing on generic HRQoL. It generates a physical component summary and a mental component summary (15), and is measured on a T-score metric with a U.S. population mean of 50 (SD, 10); a higher score denotes better HRQoL. It has been previously validated in SSc (16).

Leicester Cough Questionnaire. Leicester Cough Questionnaire is a self-administered questionnaire that assesses the impact of chronic cough on HRQoL (17). It has physical, social, and psychological domains, with a total score between 3 and 21. A higher score suggests less impairment. The Leicester Cough Questionnaire was administered in SLS-II (18).

Quantitative lung fibrosis and quantitative ILD on HRCT. On HRCT quantitative lung fibrosis (QLF) measures the extent of reticulation with architectural distortion, and quantitative interstitial lung disease (QILD) measures the total extent of ILD. They each range from 0 to 100%, where high scores represent an increasing amount of ILD/fibrosis. The score is based on a classification algorithm using texture features from a calibrated HRCT (19). For this analysis, we used the most severe zone/lobe for the QILD and QLF (20) and present the data for the two trials separately because SLS-I obtained follow-up HRCT at 12 months, whereas SLS-II obtained it at 24 months. Based on sequential data obtained

in SSc-ILD a change of 2% was considered measurement error for most severe zone/lobe for the QILD and QLF and based on changes greater than 2% (19). In our study population, we assessed the proportion of subjects who had improvement and worsening in QILD and QLF in the MCID groups.

Statistical Analysis

Summary statistics were calculated for all demographic and clinical variables. Continuous variables were reported as mean and SD; frequencies were reported for categorical variables.

Reliability. To assess the test-retest reliability of the FVC%, intraclass correlation (ICC) was measured for the screening and baseline FVC% in SLS-I and SLS-II. ICC was also assessed for the overall subjects in pooled data from both studies. An ICC of greater than or equal to 0.90 is considered excellent at the individual level (21). In addition, we assessed the variability within each subject by calculating the coefficient of variation (CV; ratio of SD in relation to the mean). Confidence intervals for the CV were estimated using a bootstrapping sampling procedure. A total of 1,000 sample datasets were generated from the original sample with replacement and the CV was estimate for each. The 95% confidence interval of the CV was calculated as the 2.5th and 97.5th percentiles of the distribution of bootstrapped CVs.

Anchors to Assess the MCID. In our current analysis, we considered three anchors: *1*) the health transition (HT) question from the SF-36, *2*) the TDI, and *3*) the SGRQ based on their relationship to FVC% in previous studies (22–24). These were selected given their relationship between dyspnea and FVC%, and the previous use of HT in assessing MCID in ILD (25). In addition, experts recommend multiple anchors to obtain robust estimates (8). The HT question asks whether the subject is better or worse at 1 year than at baseline. The "somewhat better" and the "somewhat worse" responses to this question were chosen as the anchors for calculating MCID (24, 25). For the TDI, we chose previously published MCID estimates of 1.5 units for improvement and worsening in SSc-ILD (24). For SGRQ, we used a cutoff of 4.0 and 5.0 units as the MCID, based on the literature pertaining to COPD and idiopathic pulmonary fibrosis (IPF), respectively (26). For all these analyses, we chose a 12-month period to assess MCID because the experts agree that trial design for SSc-ILD should be a minimum of 12 months (4).

We judged the appropriateness of the anchors by assessing Spearman correlations between the anchors (TDI and changes in HT) and changes in FVC%; a correlation coefficient of greater than or equal to 0.30 was considered acceptable and reflects moderate effect size when using the Cohen rules of thumb (8, 27, 28). We assessed the magnitude of the MCID estimates using the effect size (mean change in the FVC% divided by the SD at baseline) and interpreted based on Cohen criteria: 0.20–0.49 represents a small change, 0.50–0.79 a medium change, and greater than or equal to 0.80 a large change (28).

We also sought to determine whether MCID estimates for FVC% were associated with changes in PROs, mRSS, and HRCT findings. Student's *t* test was used to compare the mean difference in outcomes (PROs and HRCT) between the MCID improvement or worsening group and the "no change" group. *P* values less than 0.05 were considered statistically significant and no adjustment was made for multiple testing.

Table 1. Intraclass Coefficient and Coefficient of Variation of FVC Percent Predicted in Combined Group, SLS-I, and SLS-II

	Intraclass Coefficient	Coefficient of Variation (%)	95% CI
Combined group	0.93	4.8	4.2–5.3%
SLS I	0.90	5.8	5.0–6.5%
SLS II	0.97	3.1	2.8–3.6%

Definition of abbreviations: CI = confidence interval; SLS = Scleroderma Lung Study.

Results

Baseline Characteristics

A total of 300 subjects were enrolled in both studies (158 in SLS-I and 142 in SLS-II) (*see* Table E1 in the online supplement).

Table 2. Estimation of MCID in FVC Percent Predicted using Medical Outcomes Short Form-36 Health Transition Anchor by Combined Group, SLS-I, and SLS-II

		Unadjusted Analysis			Adjusted Analysis	
	N	Mean	95% CI	ES	Mean	95% CI
Combined Group						
Much better	48	4.45	2.40 to 6.51	0.63	5.02	2.97 to 7.08
Somewhat better	34	2.40	0.00 to 4.80	0.34	2.97	0.57 to 5.37
Same	66	−0.57	−2.09 to 0.95	−0.09	0	−1.52 to 1.52
Somewhat worse	32	−3.89	−6.77 to 1.01	−0.48	−3.32	−6.20 to −0.44
Much worse	3	−2.02	−11.47 to 7.43	−0.53	−1.45	−10.90 to 8.00
SLS-I						
Much better	15	1.29	−2.58 to 5.17	0.18	3.75	−0.12 to 7.63
Somewhat better	13	1.44	−4.18 to 7.07	0.15	3.9	−1.72 to 9.53
Same	32	−2.46	−4.51 to 0.41	−0.43	0	−2.05 to 2.05
Somewhat worse	17	−4.49	−8.3 to 0.69	−0.61	−2.03	−5.84 to 1.77
Much worse	0					
SLS-II						
Much better	33	5.89	3.50 to 8.59	0.87	4.69	2.30 to 7.09
Somewhat better	21	2.99	0.72 to 5.26	0.60	1.79	−0.48 to 4.06
Same	34	1.20	−0.96 to 3.37	0.19	0	−2.16 to 2.17
Somewhat worse	15	−3.20	−8.09 to 1.69	−0.36	−4.4	−9.29 to 0.49
Much worse	3	−2.02	−11.47 to 7.43	−0.53	−3.22	−12.67 to 6.23

Definition of abbreviations: CI = confidence interval; ES = effect size; MCID = minimal clinically important difference; SLS = Scleroderma Lung Study.
Negative sign denotes decline in FVC percent predicted.

Follow-up data were available for 110 subjects in SLS-I and 142 subjects in SLS-II. The mean (SD) age of the pooled cohort was 50.3 (11.3) years, mean (SD) disease duration was 2.9 (2.0) years, and the mean (SD) FVC% at baseline was 67.4% (10.8%); 59% of the subjects had diffuse cutaneous SSc in both trials (*see* Table E1). Although both studies were comparable in most baseline characteristics, SLS-II had a lower mean disease duration (2.6 vs. 3.2 yr; $P = 0.01$) and lesser baseline dyspnea, as assessed by the BDI (7.2 vs. 5.7; $P < 0.001$). Baseline characteristics (sex, diffuse cutaneous SSc, race, FVC%, and $D_{L_{CO}}$%) did not significantly differ between those subjects who were included in this analysis in pooled cohort and those who were excluded because of missing follow-up data, except that those who were excluded had shorter disease duration (2.2 yr vs. 3.0 yr; $P = 0.01$) and older age (53.8 yr vs. 49.6 yr; $P = 0.02$).

Reliability of FVC%

ICC reliability of FVC% was 0.90 for SLS-I, 0.97 for SLS-II, and 0.93 for the overall subjects in pooled data from both trials. The mean difference between the screening and baseline FVC% in both trials was 0.58 (SD, 4.51; 95% confidence interval, 0.06–1.09) with the mean (SD) number of days between two measurements being 34 (33) days (Table 1). The within-subject CV was 5.8% in SLS-I, 3.1% in SLS-II, and 4.8% in pooled data.

Correlation Coefficients to Assess for the Appropriateness of Anchors

For our MCID analysis, the correlation coefficients between the anchors (HT and TDI) and the change in FVC% over 12 months met the 0.30 threshold (for HT: −0.30 for SLS-I, −0.42 for SLS-II, and −0.39 for pooled data; $P < 0.01$ for all comparisons) (for TDI: 0.41 for SLS-I, 0.42 for SLS-II, and 0.43 for pooled data; $P < 0.01$ for all comparisons). SGRQ, which was only administered in the SLS-II, had a coefficient of −0.24 and was discarded as an anchor because it was less than *a priori* cutoff greater than or equal to 0.30.

MCID Estimates

We provide both unadjusted and adjusted (adjusted for the mean change in the group reporting no change) MCID estimates for improvement and worsening in FVC% (Table 2). For the HT, the mean MCID estimate for worsening (defined as the "somewhat worse" group) of FVC% was −3.89% in the unadjusted analysis and −3.32% in the adjusted analysis. The mean MCID estimate for improvement (defined as the "somewhat improved" group) was 2.40 in the unadjusted analysis and 2.97 in the adjusted analysis. These estimates were larger than those for the no-change group (Table 2). For SLS-I, the adjusted mean estimates for improvement and worsening were 3.9% and −2.03%, respectively, and for SLS-II, the adjusted mean estimates for improvement and worsening were 1.8% and −4.4%, respectively. The effect size for MCID estimates for the improved group was 0.34 (small effect size) in the combined group, 0.15 in SLS-I, and 0.60 in SLS-II, and the effect size for MCID estimates for the worsened group was 0.48 (small effect size) in the combined group, 0.61 in SLS-I, and 0.36 in SLS-II.

The TDI was grouped into three levels: *1*) no change corresponding to −1.5 less than TDI less than 1.5; *2*) better corresponding to TDI greater than or equal to 1.5; and *3*) worse corresponding to TDI less than or equal to 1.5. The mean MCID estimate for worsening of FVC% was −4.18% in the unadjusted analysis and −2.86% in the adjusted analysis. The mean MCID estimate for improvement was 4.02% in the unadjusted analysis and 5.34% in the adjusted analysis. These estimates were larger than those for the no-change group (Table 3). For SLS-I, the adjusted mean estimates for improvement and worsening were 6.9% and −2.5%, and for SLS-II the adjusted mean estimates for improvement and worsening were 2.6% and −2.7%.

The effect size for MCID estimates for the improved group was 0.62 (moderate effect size) in the combined group, 0.54 in SLS-I, and 0.53 in SLS-II. The effect size for MCID estimates for the worsened group was 0.50 (moderate effect size) in the combined group, 0.77 in SLS-I, and 0.08 in SLS-II.

Relationship between the MCID Estimates with PROs, Skin Score, and $D_{L_{CO}}$% Predicted

We explored whether the subjects who improved or worsened by greater than or equal to MCID over 12 months translated into parallel changes in PROs scores in SLS I-and SLS-II (Tables 4 and 5). Using the HT anchor, we used a 3.0% change as improvement, and a −3.3% change for worsening as FVC% MCID estimates. For improvement, statistically significant improvements were noted for the SF-36 physical component summary, TDI, and Health Assessment Questionnaire disability index compared with the no-change group ($P < 0.05$ for all comparisons). For subjects who worsened, statistical significances were noted with TDI, SGRQ, mRSS, and DLCO% predicted compared with the no-change group (Table 4).

For the MCID based on TDI as the anchor, we considered a change of 5.3% for improvement and −3.0% for worsening. For improvement, statistically significant improvement was only noted for SF-36 physical component summary compared with the no-

Table 3. Estimation of MCID in FVC Percent Predicted Using the Transition Dyspnea Index Anchor by Combined Group, SLS-I, and SLS-II

		Unadjusted Analysis			Adjusted Analysis	
	N	Mean	95% CI	ES	Mean	95% CI
Combined Group						
TDI ⩾ 1.5	65	4.02	2.41 to 5.63	0.62	5.34	3.73 to 6.95
−1.5 < TDI < 1.5	113	−1.32	−2.72 to 0.08	−0.18	0	−1.40 to 1.40
TDI ⩽ −1.5	45	−4.18	−6.69 to −1.68	−0.50	−2.86	−5.37 to −0.36
SLS-I						
TDI ⩾ 1.5	31	3.41	1.1 to 5.71	0.54	6.85	4.54 to 9.15
−1.5 < TDI < 1.5	69	−3.44	−5.26 to −1.62	−0.45	0	−1.82 to 1.82
TDI ⩽ −1.5	30	−5.93	−8.8 to −3.06	−0.77	−2.49	−5.36 to 0.38
SLS-II						
TDI ⩾ 1.5	34	4.58	2.24 to 6.93	0.53	2.57	0.23 to 4.92
−1.5 < TDI < 1.5	44	2.01	0.14 to 3.88	0.33	0	−1.87 to 1.87
TDI ⩽ −1.5	15	−0.69	−5.54 to 4.16	−0.08	−2.7	−7.55 to 2.15

Definition of abbreviations: CI = confidence interval; ES = effect size; MCID = minimal clinically important difference; SLS = Scleroderma Lung Study; TDI = transition dyspnea index; TDI ⩾ 1.5 = improved group; TDI ⩽ −1.5 = worsened group.
Negative sign denotes decline in FVC percent predicted.

Table 4. Change in Patient-Reported Outcome Measures, Skin Score, and D_{LCO}% by MCID (as Based on the Health Transition Question from the SF-36)

	Change in FVC% by >3.0% Improvement (MCID Improved Group)	3.3% Worsening in FVC% to <3.0% Improvement in FVC% (No Change Group)	Change in FVC% by >3.3% Worsening (MCID Worsened Group)
SF-36 PCS*			
N	72	73	38
Mean (SD) difference	4.32 (7.96)	0.51 (7.82)	−1.74 (8.48)
P value	<0.001	ref	0.45
SF-36 MCS*			
N	72	73	38
Mean (SD) difference	3.36 (10.50)	1.44 (10.25)	0.51 (10.42)
P value	0.34	ref	0.65
TDI*			
N	77	89	58
Mean (SD) difference	2.61 (3.70)	0.12 (3.10)	−1.36 (3.19)
P value	<0.001	ref	0.006
HAQ-DI†			
N	85	93	61
Mean (SD) difference	−0.13 (0.50)	0.03 (0.43)	0.11 (0.50)
P value	0.01	ref	0.31
SGRQ (Total)†			
N	51	41	15
Mean (SD) difference	−5.72 (14.90)	−2.36 (10.41)	6.05 (16.35)
P value	0.23	Ref	0.030
LCQ*			
N	51	38	15
Mean (SD) difference	−0.05 (2.84)	0.59 (3.20)	0.49 (3.25)
P value	0.32	ref	0.92
mRSS†			
N	84	94	61
Mean (SD) difference	−3.45 (6.16)	−3.32 (5.69)	−1.16 (5.72)
P value	0.88	ref	0.02
D_{LCO}%			
N	84	96	61
Mean (SD) difference	1.74 (10.26)	−1.06 (9.66)	−8.41 (9.95)
P value	0.06	ref	<0.001

Definition of abbreviations: FVC% = FVC percent predicted; HAQ-DI = Health Assessment Questionnaire disability index; LCQ = Leicester Cough Questionnaire; MCID = minimal clinically important difference; MCS = Mental Component Summary; mRSS = modified Rodnan Skin Score; PCS = Physical Component Summary; SF-36 = Medical Outcomes Short Form 36; SGRQ = St. George's Respiratory Questionnaire; TDI = transition dyspnea index.
P value is based on the comparison between the no-change group.
*Higher score denotes better health or higher D_{LCO}.
†Higher score denotes worse health or more severe mRSS.

change group. For subjects who worsened, statistically significant differences were noted with SGRQ, mRSS, and DLCO% predicted compared with the no-change group (Table 5).

Relationship between the MCID Estimates and HRCT Findings

We assessed the relationship of MCID estimates to the HRCT QILD and QLF separately in the SLS-I and SLS-II cohorts, because the follow-up HRCT was performed at the 12-month period in SLS-I and the 24-month period in SLS-II (Table 6). Mean MCID estimates for both improvement and worsening using each anchor were associated with statistically significant changes (compared with the no-change groups) for QLF in SLS-II. In SLS-I, however, MCID estimates for worsening were significantly associated with changes in QLF only using the HT anchor ($P = 0.044$), and nearly significantly associated using the TDI anchor ($P = 0.080$). In SLS-II, MCID estimates for improvement in FVC% were significantly associated with changes in QILD using the TDI anchor ($P = 0.02$), and the estimates for worsening were significantly associated with changes in QILD using both the HT and TDI anchors ($P = 0.005$ and $P = 0.0003$, respectively). When assessing improvement in the QLF and QILD as defined by more than −2% change, a larger proportion of participants had improvement among those who also showed improvement in FVC% in SLS-I and SLS-II (Table 6). Similarly, a larger proportion of subjects who had worsening in the FVC% also exhibited worsening in the QILF and QILD, as defined by greater than +2% change.

Discussion

The FVC% has traditionally served as the primary endpoint in fibrotic ILD clinical trials, because a decrement in FVC% is the physiologic hallmark of ILD that defines the disease (5, 29–32). In addition, FVC% is easily measured, is responsive to change, and has an acceptable measurement error (if performed using standardized methodology). Using two randomized controlled trials (RCTs), SLS-I and -II, we show that the FVC% has an acceptable

Table 5. Change in Patient-Reported Outcome Measures and Skin Score by MCID (as Based on the Transition Dyspnea Index)

	Change in FVC% by >5.3% Improvement (MCID Improved Group)	3.0% Worsening in FVC% to <5.3% Improvement in FVC% (No Change Group)	Change in FVC% by >3.0% Worsening (MCID Worsened Group)
SF-36 PCS*			
N	61	84	38
Mean (SD) difference	3.57 (7.76)	0.67 (8.39)	−1.74 (8.48)
P value	0.04	Ref	0.15
SF-36 MCS*			
N	61	84	38
Mean (SD) difference	2.76 (10.80)	2.13 (10.14)	0.51 (10.42)
P value	0.72	Ref	0.42
HAQ-DI†			
N	71	107	61
Mean (SD) difference	−0.11 (0.52)	0.00 (0.44)	0.11 (0.50)
P value	0.14	Ref	0.12
SGRQ (Total)†			
N	43	49	15
Mean (SD) difference	−4.4 (15.34)	−4.07 (11.00)	6.05 (16.35)
P value	0.91	Ref	0.008
LCQ*			
N	43	46	15
Mean (SD) difference	−0.37 (2.83)	0.78 (3.08)	0.49 (3.25)
P value	0.07	Ref	0.75
mRSS†			
N	71	107	61
Mean (SD) difference	−3.85 (6.42)	−3.07 (5.53)	−1.16 (5.72)
P value	0.4	Ref	0.03
D_{LCO}*			
N	70	110	61
Mean (SD) difference	1.77 (10.21)	−0.73 (9.81)	−8.41 (9.95)
P value	0.10	Ref	<0.001

Definition of abbreviations: FVC% = FVC percent predicted; HAQ-DI = Health Assessment Questionnaire disability index; LCQ = Leicester Cough Questionnaire; MCID = minimal clinically important difference; MCS = Mental Component Summary; mRSS = modified Rodnan Skin Score; PCS = Physical Component Summary; SF-36 = Medical Outcomes Short Form 36; SGRQ = St. George's Respiratory Questionnaire.
P value is based on the comparison between the no-change group.
*Higher score denotes better health or higher D_{LCO}.
†Higher score denotes worse health or more severe mRSS.

test-retest reliability in SSc-ILD, and the MCID estimates for improvement range between 3.3% and 5.3%, and for worsening are between −3.0% and −3.3%.

Reliability is defined as the extent to which the measure yields the same score when the outcome has not changed (33). Reliability is important for the evaluation of an outcome measure, and test-retest reliability is central to the evaluation of an outcome measure for clinical trials. The test-retest reliability of FVC measurements over a short interval (approximately 1 mo) in our pooled database was 0.93, and is in agreement with the analysis from large RCTs of IFN-γ1b in IPF, where the ICC was also 0.93 (25). In both the SLS-I and -II trials, spirometry was conducted according to the American Thoracic Society/European Respiratory Society standard protocol (34). In addition, emphasis was given to individual site training, centralized quality assurance monitoring, and the involvement of both pulmonologists and rheumatologists at each site, and the acceptability and repeatability of all spirometric tests was evaluated by a central quality control core. It was also encouraged that repeat testing be done on the same equipment and by the same tester whenever possible. This may also explain the acceptable within-subject CV of 4.8% in the pooled database between screening and baseline visits. CV is a measure of variability, and in healthy subjects, a within-subject week-to-week CV of 5.0–7.8% has been reported (35). In restrictive lung disease, the within-subject week-to-week CV has not been reported but is likely larger.

MCID estimates are an approximation and experts have suggested using multiple anchors to define a range for these estimates (8). Our data suggest that a change between 3.0% and 5.3% is the MCID for improvement, and a change of −3.0% to −3.3% is the MCID for worsening (after adjusting for the no-change group). In the unadjusted analysis, the mean changes in the improvement and worsening MCID groups were larger and in the right direction compared with the no-change group, giving us confidence in the data. When using the effect size, the MCID estimates for the improved and worsened groups were 0.34 and 0.48, respectively, for the SF-36 HT question and 0.62 and 0.50, respectively, for the TDI question. Previously published data have suggested that the MCID estimates range between 0.2 and 0.6 SD (8, 27), and our estimates are in line with these observations. The variability noted in the estimates based on the two anchors may be caused by the five-point response for HT (no-change, somewhat worse or better, and much worse or better),

Table 6. Change in the HRCT Findings by MCID Estimates

	MCID Estimates Using HT as an Anchor			MCID Estimates Using TDI as an Anchor		
	Change in FVC% by >3.0% Improvement (MCID Improved Group)	3.3% Worsening in FVC% to <3.0% Improvement in FVC% (No Change Group)	Change in FVC% by >3.3% Worsening (MCID Worsened Group)	Change in FVC% by >5.3% Improvement (MCID Improved Group)	3.0% Worsening in FVC% to <5.3% Improvement in FVC% (No Change Group)	Change in FVC% by >3.0% Worsening (MCID Worsened Group)
SLS-I						
N	15	33	32	10	37	33
QILD, mean (SD) difference	−8.05 (15.11)	−0.11 (14.52)	3.82 (13.70)	−11.85 (15.91)	0.15 (14.05)	3.35 (13.75)
P value	0.089	Ref	0.27	0.02	Ref	0.34
Improvement >2%, *n* (%)	11 (73.3)	12 (36.4)	11 (34.4)	8 (80)	14 (37.8)	12 (36.4)
Worsening >2%, *n* (%)	3 (20.0)	16 (48.5)	19 (59.4)	1 (10)	18 (48.7)	19 (57.6)
QLF, mean (SD) difference	−2.84 (14.60)	0.44 (17.63)	9.66 (18.67)	−6.73 (14.17)	1.38 (17.15)	9.01 (18.75)
P value	0.53	Ref	0.04	0.18	Ref	0.08
Improvement >2%, *n* (%)	7 (46.7)	10 (30.3)	6 (18.8)	6 (60)	10 (27.0)	7 (21.2)
Worsening >2%, *n* (%)	4 (26.7)	15 (45.5)	20 (62.5)	2 (20)	17 (46.0)	20 (60.6)
SLS-II						
N	56	25	16	39	40	18
QILD, mean (SD) difference	−4.96 (9.38)	−1.79 (10.53)	9.33 (13.18)	−6.79 (9.09)	−2.01 (9.05)	9.58 (13.29)
P value	0.18	Ref	0.005	0.022	Ref	0.0003
Improvement >2%, *n* (%)	35 (62.5)	11 (44.0)	3 (18.8)	29 (74.4)	16 (40.0)	4 (22.2)
Worsening >2%, *n* (%)	12 (21.4)	8 (32.0)	11 (68.8)	6 (15.4)	13 (32.5)	12 (66.7)
QLF, mean (SD) difference	−2.43 (9.58)	2.40 (6.63)	11.19 (13.20)	−3.96 (9.84)	1.80 (7.40)	10.27 (12.69)
P value	0.03	Ref	0.007	0.004	Ref	0.002
Improvement >2%, *n* (%)	24 (42.9)	3 (12.0)	2 (12.5)	18 (46.2)	9 (22.5)	2 (11.1)
Worsening >2%, *n* (%)	16 (28.6)	10 (40.0)	12 (75.0)	9 (23.1)	16 (40.0)	13 (72.2)

Definition of abbreviations: FVC% = FVC percent predicted; HRCT = high-resolution computed tomography; HT = health transition question in Medical Outcomes Short Form 36; MCID = minimal clinically important difference; QILD = quantitative interstitial lung disease; QLF = quantitative lung fibrosis; SLS = Scleroderma Lung Study; TDI = transition dyspnea index.
Computer-assisted QLF and QILD scores were derived from the most severe lobe for volumetric scans or zone for nonvolumetric scan. A negative score denotes improvement in HRCT scores over time.

whereas the TDI was grouped into three levels (no-change, better, and worse, based on the MCID of 1.5 for the TDI). Our data are in accord with the MCID estimates in two large RCTs in IPF with a range between 2% and 6% for worsening in the FVC% (25). The authors did not report an MCID for improvement, because stabilization of FVC% is currently the best case scenario in IPF (25).

One point to highlight is that the MCID estimates are calculated at a group level and should not be confused with change in a measure in an individual patient. At an individual level, a larger change (likely greater than the CV) is required to be considered a statistically significant change, and is influenced by both measurement error and normal biologic variability. In other words, a change of 3.0–5.3% is clinically important in a group of patients, but is within measurement variability for an individual patient. Pennock and colleagues (35) have suggested multiplying the CV by 1.65 to determine the limit by which change in FVC might represent a significant change.

Our MCID estimates were associated with statistically significant or numerical changes in the PROs, mRSS, and HRCT findings in the right directions, a finding that suggests that these MCID estimates translate into how a patient feels and functions (36, 37). For HRCT QLF scores, the associations with our MCID estimates were statistically significant for improvement in SLS-II, but not SLS-I. It is possible that this difference in the results for the two trials could be caused by the fact that in SLS-II, a larger number of subjects noted clinically meaningful improvement in FVC% compared with SLS-I. In addition, the magnitude of improvement in radiographic evidence of ILD was also greater in SLS-II than SLS-I (19, 38). Furthermore, the follow-up HRCT in SLS-II was performed at 24 months, in contrast to 12 months in SLS-I, possibly contributing to the differences observed for the statistical significance of the association between the MCID for FVC% and this radiographic measure between the two trials.

Our study has many strengths. First, we used prospective data from two large SSc-ILD RCTs (SLS-I and -II) to assess MCID estimates. Second, we provided MCID estimates for both improvement and worsening, because these can be different. Third, our study provided MCID estimates that correspond both with PROs and with HRCT changes over time, supporting the validity of these estimates. Finally, we used two anchors to assess the MCID estimates and our MCID for FVC% was reassuringly similar to those reported for IPF, thus providing further confidence in our results.

Our study is not without limitations. First, the analysis was *post hoc* rather than *a priori*. Second, SLS-I and -II were of different durations and the methodology of administering the questionnaire for one

of the anchors (TDI) also differed. Second, although the correlation between the HT and FVC% was above the proposed cutoff point of 0.30, the relationship between this general health anchor and the lung function outcome could have been influenced by changes in variables unrelated to the lung, including skin, gastrointestinal involvement, and other disease manifestations. Third, our CV of 4.8% was based on strict institution of the American Thoracic Society/European Respiratory Society guidelines for spirometry and training of the technicians. In real life, the CV is likely to be greater than 4.8%. Fourth, the participants came from two different trials with different immunosuppressive treatments, and placebo. The differing interventions influenced FVC%, and subclinical or mild pulmonary hypertension may have influenced PROs. Therefore, the MCID estimates should be considered preliminary and confirmed in an independent clinical trial. Fifth, although we had missing data, baseline differences were largely similar in those with complete versus missing data. Lastly, a longer follow-up period is required to assess the impact of MCID estimates on mortality.

In conclusion, our study demonstrates that FVC has an acceptable test-retest reliability and we propose MCID estimates derived from two large RCTs. These data can be used for interpretation of the results of ongoing clinical trials in SSc-ILD and for sample size estimation in future trials. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

The following people and institutions participated in the Scleroderma Lung Study I: University of Michigan, Ann Arbor: D. Khanna; University of California at Los Angeles (UCLA), Los Angeles: P. J. Clements, D. P. Tashkin, R. Elashoff, J. Goldin, M. Roth, D. Furst, K. Bulpitt, W.-L. J. Chung, S. Viasco, M. Sterz, L. Woolcock, X. Yan, J. Ho, S. Vasunilashorn, and I. da Costa; University of Medicine and Dentistry of New Jersey, New Brunswick: J. R. Seibold, D. J. Riley, J. K. Amorosa, V. M. Hsu, D. A. McCloskey, and J. E. Wilson; University of Illinois at Chicago, Chicago: J. Varga, D. Schraufnagel, A. Wilbur, D. Lapota, S. Arami, and P. Cole-Saffold; Boston University, Boston: R. Simms, A. Theodore, P. Clarke, J. Korn, K. Tobin, and M. Nuite; Medical University of South Carolina, Charleston: R. Silver, M. Bolster, C. Strange, S. Schabel, E. Smith, J. Arnold, K. Caldwell, and M. Bonner; The Johns Hopkins University School of Medicine, Baltimore: R. Wise, F. Wigley, B. White, L. Hummers, M. Bohlman, A. Polito, G. Leatherman, E. Forbes, and M. Daniel; Georgetown University, Washington: V. Steen, C. Read, C. Cooper, S. Wheaton, A. Carey, and A. Ortiz; University of Texas at Houston, Houston: M. Mayes, E. Parsley, S. Oldham, T. Filemon, S. Jordan, and M. Perry; University of California at San Francisco, San Francisco: K. Connolly, J. Golden, P. Wolters, R. Webb, J. Davis, C. Antolos, and C. Maynetto; University of Alabama at Birmingham, Birmingham: B. Fessler, M. Olman, C. Sanders, L. Heck, and T. Parkhill; University of Connecticut Health Center, Farmington: N. Rothfield, M. Metersky, R. Cobb, M. Aberles, F. Ingenito, and E. Breen; Wayne State University, Detroit: M. Mayes, K. Mubarak, J. L. Granda, J. Silva, Z. Injic, and R. Alexander; Virginia Mason Research Center, Seattle: D. Furst, S. Springmeyer, S. Kirkland, J. Molitor, R. Hinke, and A. Mondt; Data Safety and Monitoring Board: Harvard Medical School, Boston: T. Thompson; Veterans Affairs Medical Center, Brown University, Providence: S. Rounds; Cedars Sinai Medical Center–UCLA Health System, Los Angeles: M. Weinstein; Clinical Trials & Surveys, Baltimore: B. Thompson; Mortality and Morbidity Review Committee: University of California, Los Angeles: H. Paulus and S. Levy; The Johns Hopkins University, Baltimore: D. Martin.

The Scleroderma Lung Study II Research Group includes the authors and the following support staff: E. Kissin and F. Y. Cheong (Boston University); G. Marlis, J. Mason-Berry, P. Saffold, M. Rodriguez, L. Guzman, J. Brook, G. Ibrahim, and K. Largaespada (UCLA); C. Fridley, M. Zulmastashvili, A. Manu, and S. Moore (Georgetown University); L. Hummers and G. Leatherman (Johns Hopkins University); F. N. Hant and K. Gibson (Medical University of South Carolina); M. Morrison (National Jewish Health); H. Donnelly, C. Marlin, and J. Gangar (Northwestern University); D. A. McCloskey (Rutgers University); A. Eller, D. Leong, M. Lalosh, and J. Obata (UCSF); S. Arami and D. Franklin (University of Illinois); E. Schiopu, M. Benedict-Blue, V. Leone, and J. Shaw (University of Michigan); F. Tan, M. Perry, J. Anderson, and A. Saulino (University of Texas, Houston); P. Carey and M. Esplin (University of Utah); and P. Carlson (University of Minnesota).

References

1. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis* 2007;66:940–944.
2. Khanna D, Saggar R, Mayes MD, Abtin F, Clements PJ, Maranian P, *et al*. A one-year, phase I/IIa, open-label pilot trial of imatinib mesylate in the treatment of systemic sclerosis-associated active interstitial lung disease. *Arthritis Rheum* 2011;63:3540–3546.
3. Simeón-Aznar CP, Fonollosa-Plá V, Tolosa-Vilella C, Selva-O'Callaghan A, Solans-Laqué R, Vilardell-Tarrés M. Effect of mycophenolate sodium in scleroderma-related interstitial lung disease. *Clin Rheumatol* 2011;30:1393–1398.
4. Khanna D, Seibold J, Goldin J, Tashkin DP, Furst DE, Wells A. Interstitial lung disease points to consider for clinical trials in systemic sclerosis. *Rheumatology* 2017;56(Suppl 5):v27–v32.
5. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, *et al*.; Scleroderma Lung Study Research Group. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006;354:2655–2666.
6. Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, *et al*.; Scleroderma Lung Study II Investigators. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. *Lancet Respir Med* 2016;4:708–719.
7. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000;18:419–423.
8. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102–109.
9. Khanna D, Yan X, Tashkin DP, Furst DE, Elashoff R, Roth MD, *et al*.; Scleroderma Lung Study Group. Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: results from the scleroderma lung study. *Arthritis Rheum* 2007;56:1676–1684.
10. 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;85:751–758.
11. Jones PW, Quirk FH, Baveystock CM. The St. George's Respiratory Questionnaire. *Respir Med* 1991;85(Suppl B):25–31; discussion 3–7.
12. Wallace B, Kafaja S, Furst DE, Berrocal VJ, Merkel PA, Seibold JR *et al*. Reliability, validity and responsiveness to change of the Saint George's Respiratory Questionnaire in early diffuse cutaneous systemic sclerosis. *Rheumatology (Oxford)* 2015;54:1369–1379.
13. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–145.
14. Clements PJ, Wong WK, Hurwitz EL, Furst DE, Mayes M, White B, *et al*. Correlates of the disability index of the health assessment questionnaire: a measure of functional impairment in systemic sclerosis. *Arthritis Rheum* 1999;42:2372–2380.
15. Ware JE Jr. SF-36 health survey update. *Spine* 2000;25:3130–3139.
16. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, *et al*.; Relaxin Study Group; Scleroderma Clinical Trials Consortium. Responsiveness of the SF-36 and the Health Assessment

Questionnaire Disability Index in a systemic sclerosis clinical trial. *J Rheumatol* 2005;32:832–840.
17. Birring SS, Prudon B, Carr AJ, Singh SJ, Morgan MD, Pavord ID. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). *Thorax* 2003;58:339–343.
18. Tashkin DP, Volkmann ER, Tseng CH, Roth MD, Khanna D, Furst DE, *et al*. Improved cough and cough-specific quality of life in patients treated for scleroderma-related interstitial lung disease: results of Scleroderma Lung Study II. *Chest* 2017;151:813–820.
19. Kim HJ, Brown MS, Elashoff R, Li G, Gjertson DW, Lynch DA, *et al*. Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. *Eur Radiol* 2011;21: 2455–2465.
20. Kim HJ, Tashkin DP, Gjertson DW, Brown MS, Kleerup E, Chong S, *et al*. Transitions to different patterns of interstitial lung disease in scleroderma with and without treatment. *Ann Rheum Dis* 2016;75:1367–1371.
21. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428.
22. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, *et al*. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med* 2010;104:296–304.
23. Swigris JJ, Wamboldt FS, Behr J, du Bois RM, King TE, Raghu G, *et al*. The 6 minute walk in idiopathic pulmonary fibrosis: longitudinal changes and minimum important difference. *Thorax* 2010;65: 173–177.
24. Khanna D, Tseng CH, Furst DE, Clements PJ, Elashoff R, Roth M, *et al*.; for Scleroderma Lung Study Investigators. Minimally important differences in the Mahler's Transition Dyspnoea Index in a large randomized controlled trial: results from the Scleroderma Lung Study. *Rheumatology (Oxford)* 2009;48:1537–1540.
25. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *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;184:1382–1389.
26. Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD* 2005; 2:75–79.
27. Khanna D, Hays RD, Shreiner AB, Melmed GY, Chang L, Khanna PP, *et al*. Responsiveness to change and minimally important differences of the patient-reported outcomes measurement information system gastrointestinal symptoms scales. *Dig Dis Sci* 2017;62:1186–1192.
28. Cohen J. Statistical power analysis for the behavioral sciences. 1st ed. Lawrence Erlbaum Associates: Academic Press; 1977.
29. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, *et al*.; INPULSIS Trial Investigators. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071–2082.
30. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, *et al*.; ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2083–2092.
31. Hoyles RK, Ellis RW, Wellsbury J, Lees B, Newlands P, Goh NS, *et al*. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum* 2006;54:3962–3970.
32. Khanna D, Albera C, Fischer A, Khalidi N, Raghu G, Chung L, *et al*. An open-label, phase ii study of the safety and tolerability of pirfenidone in patients with scleroderma-associated interstitial lung disease: the LOTUSS Trial. *J Rheumatol* 2016;43:1672–1679.
33. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. *Qual Life Res* 1993; 2:441–449.
34. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, *et al*.; ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J* 2005;26:319–338.
35. Pennock BE, Rogers RM, McCaffree DR. Changes in measured spirometric indices. What is significant? *Chest* 1981;80:97–99.
36. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 2012;31:2973–2984.
37. Powers JH 3rd, Patrick DL, Walton MK, Marquis P, Cano S, Hobart J, *et al*. Clinician-reported outcome assessments of treatment benefit: report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. *Value Health* 2017;20:2–14.
38. Kim GH, Tashkin D, Lo P, Lu P, Brown M, Kleerup E, *et al*. Transitional voxel-wise changes in the interstitial lung disease on high resolution computed tomography using Scleroderma Lung Study-II [abstract]. *Am J Respir Crit Care Med* 2016;193:A1138.