




Treatment With Mycophenolate and Cyclophosphamide Leads to Clinically Meaningful Improvements in Patient-Reported Outcomes in Scleroderma Lung Disease: Results of Scleroderma Lung Study II

Elizabeth R. Volkman,¹  Donald P. Tashkin,¹ Holly LeClair,¹ Michael D. Roth,¹ Grace Kim,¹ 
Jonathan Goldin,¹ Philip J. Clements,¹ Daniel E. Furst,² and Dinesh Khanna³ 

Objective. Our objective was to determine if treatment with cyclophosphamide (CYC) and mycophenolate mofetil (MMF) improves patient-reported outcomes (PROs) among patients with systemic sclerosis-related interstitial lung disease (SSc-ILD).

Methods. This study examined PROs in patients with SSc-ILD (N = 142) who participated in the Scleroderma Lung Study II, a randomized controlled trial comparing MMF for 2 years with oral CYC for 1 year followed by 1 year of a placebo. Joint models were created to evaluate the course of PROs over 2 years. The difference in PRO scores from baseline to 24 months was measured, and the percentage of patients meeting the minimum clinically important difference (MCID) was calculated. Correlations between PROs and SSc-ILD disease severity measures were also examined.

Results. Treatment with CYC and MMF led to improvements in several PROs with no between-treatment differences. Scores for the Transitional Dyspnea Index (TDI) and St. George's Respiratory Questionnaire (SGRQ) improved significantly over 2 years, and 29%/24% and 28%/25% of participants in the CYC/MMF groups met or exceeded the MCID estimates for TDI and SGRQ, respectively. At baseline, the forced vital capacity (FVC) percentage predicted (FVC%-predicted) did not correlate with the Baseline Dyspnea Index or SGRQ. However, improvements in the FVC%-predicted were weakly associated with improvements in dyspnea (assessed by the TDI) and SGRQ scores.

Conclusion. Treatment with CYC and MMF improved overall health-related quality of life in patients with SSc-ILD. The relationship between PRO measures and the FVC was relatively weak, suggesting that PROs provide complementary information about treatment efficacy not captured by changes in the FVC alone in this patient population.

INTRODUCTION

Patient-reported outcomes (PROs) directly assess how patients feel and function from their own perspective. For complex systemic diseases, such as systemic sclerosis, PROs play a central role in providing insight into a patient's experience living with this disease (1–3). Specifically, PROs measure symptoms,

such as dyspnea, and health-related quality of life (HRQOL), and they may reveal treatment effects, such as pulmonary function, that are not captured by other measures.

In systemic sclerosis (SSc) and interstitial lung disease (ILD) (SSc-ILD) therapeutic research trials, PROs have included dyspnea indexes (4–6), cough questionnaires (7,8), and HRQOL surveys (4). Although other outcomes, such as the forced vital

The study drug (mycophenolate) and matching placebo were supplied at no charge through Drug Supply grant CEL539 from Hoffmann-La Roche/Genentech. Dr. Volkman's work was supported by the Rheumatology Research Foundation (Scientist Development Award) and the Scleroderma Foundation (Young Investigator Award). Dr. Tashkin work was supported by grants from the National Heart, Lung, and Blood Institute, NIH (grants R01-HL-089758 and R01-HL-089901, respectively). Dr. Khanna's work was supported by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grants K24-AR-063120 and R01-AR-07047).

¹Elizabeth R. Volkman, MD, MS, Donald P. Tashkin, MD, Holly LeClair, MS, Michael D. Roth, MD, Grace Kim, PhD, Jonathan Goldin, MD, PhD, Philip J. Clements, MD, MPH: University of California, Los Angeles; ²Daniel E. Furst, MD: University of California, Los Angeles, University of Washington,

Seattle, and University of Florence, Florence, Italy; ³Dinesh Khanna, MD, MS: University of Michigan Medical School, Ann Arbor.

Dr. Volkman is a consultant to Boehringer Ingelheim and Forbuis. Dr. Khanna is a consultant to Acceleron, Actelion, Amgen, Bayer, Bristol-Myers Squibb, Boehringer Ingelheim, CSL Behring, Corbus, Galapagos, Genentech/Roche, GlaxoSmithKline, Horizon, Merck, Mitsubishi Tanabe Pharma, Sanofi-Aventis, and United Therapeutics, and he has stock options in Eicos Sciences, Inc. No other disclosures relevant to this article were reported.

Address correspondence to Elizabeth R. Volkman, MD, MS, University of California, Los Angeles, Department of Medicine, Division of Rheumatology, 1000 Veteran Avenue, Suite 32-59, Los Angeles, CA 90095. E-mail: evolkman@mednet.ucla.edu.

Submitted for publication December 7, 2019; accepted in revised form February 14, 2020.

capacity (FVC) percentage predicted (FVC%-predicted), shed light on treatment-related changes in pulmonary physiology, only the PROs provide information about how a particular treatment affects a patient's experience with the disease. In some SSc-ILD trials, improvements in PRO scores have paralleled improvements observed in the FVC%-predicted (4), whereas in other trials, changes in PRO scores do not parallel changes in the FVC%-predicted (5). These conflicting results suggest that PROs may be measuring distinct treatment-related effects that are important to consider in caring for patients with SSc-ILD.

The present study sought to evaluate changes in PROs in the Scleroderma Lung Study (SLS) II, a randomized controlled trial comparing the safety and efficacy of mycophenolate mofetil (MMF) and oral cyclophosphamide (CYC) for treating SSc-ILD (9). The results of SLS II demonstrated that the FVC%-predicted improved significantly over 24 months in both treatment arms (average FVC%-predicted improvement: MMF 3.3%; CYC 3.0%) (9). The primary objective of this study was to investigate how specific PROs changed in response to therapy with CYC and MMF. A secondary goal was to determine whether changes in PROs correlate with changes in the FVC%-predicted in SLS II.

PATIENTS AND METHODS

Patient population. Participants of SLS II were all adults (18-75 years old) with SSc-ILD and either limited SSc (lcSSc) or diffuse SSc (dcSSc) (10) with active ILD, defined as the presence of both a restrictive to borderline restrictive ventilatory impairment (FVC%-predicted: less than 80%-85% but greater than or equal to 45%) and the presence of any ground glass opacity (GGO) (hazy opacity through which normal lung markings can be discerned) on high-resolution computed tomography (HRCT). Participants also had to have exertional dyspnea based on the Mahler Baseline Dyspnea Index (BDI) (11) and a disease duration of less than or equal to 7 years from the onset of the first non-Raynaud's symptom of SSc. Key exclusion criteria included pulmonary hypertension, clinically significant abnormalities on HRCT not attributable to SSc, smoking within the past 6 months, and evidence of significant airflow obstruction, defined as a ratio of the forced expired volume in 1 second to the FVC%-predicted of 65% or less. The study was approved by the Office of Human Research Protection Program at University of California, Los Angeles (UCLA) (Institutional Review Board [IRB] No. 11-002659-CR-00005) and by the IRBs of all 14 participating centers.

SLS II Study Design. Participants in SLS II were randomized to receive either oral CYC for 1 year followed by 1 year of a placebo or MMF for 2 years (9). The primary end point for the study was the course of the FVC%-predicted, measured every 3 months over 2 years. Thoracic HRCT imaging was obtained at baseline and 2 years, and a computer-aided design scoring system was employed to provide quantitative measures of different patterns of

ILD as previously described (12). The quantitative ILD (QILD) score was the sum of all abnormally classified scores, including scores for quantitative lung fibrosis (QLF) (linear reticular markings with architectural distortion), GGO, and honeycomb changes (clustered air-filled cysts with dense walls). Scores were calculated as a percentage of total counted voxels for both the whole lung (WL), which included both lungs, and the zone of maximal involvement (ZM). The complete details of the SLS II protocol appear in the supplementary web appendix accompanying the main SLS II publication (9).

PROs. The following PROs were examined in SLS II: Short Form 36 (SF-36) (13), the Health Assessment Questionnaire Disability Index (HAQ-DI) (14), the BDI (11), the Transitional Dyspnea Index (TDI) (15), the Leicester Cough Questionnaire (LCQ) (16), St. George's Respiratory Questionnaire (SGRQ) (17), and the UCLA Scleroderma Clinical Trials Consortium (SCTC) Gastrointestinal Tract 2.0 (GIT 2.0) (18). Each PRO aimed to address a unique aspect of the disease experience of SSc-ILD. For instance, the SF-36 measures HRQOL and consists of eight scales with both physical and mental components of HRQOL (13). For this analysis, we focused on the scores for the physical component summary (PCS) and mental component summary (MCS). The HAQ-DI measures functional ability in patients with musculoskeletal conditions (14) and has been studied extensively in patients with SSc (19,20).

The PROs targeting respiratory symptoms included the BDI, the TDI, the LCQ, and SGRQ. The BDI assesses patients' perception of their breathlessness at baseline based on three categories (functional impairment, magnitude of task, and magnitude of effort), whereas the TDI measures the change in dyspnea from baseline in each of these categories, the results of which are summed into a total score (11,15). The LCQ is a 19-item HRQOL measure of chronic cough and is highly responsive to change (16). It is a patient-derived questionnaire; therefore, it contains items, domains, and response scales that are clinically meaningful to the patient. SGRQ is a 50-item questionnaire that was originally designed to measure the impact of overall health, daily life, and perceived well-being in patients with obstructive airway disease (21). However, prior studies have demonstrated that it correlates well with other measures of disease activity in patients with SSc-ILD (17) and is responsive to change in patients with dcSSc (22).

The UCLA SCTC GIT 2.0 is an instrument that measures gastrointestinal tract involvement in SSc and contains 34 items from seven scales (reflux, distention/bloating, diarrhea, fecal soilage, constipation, emotional well-being, and social functioning) (18). It has been translated into several languages and has also been found to discriminate between patients with and without objective evidence of gastrointestinal tract involvement (23). A recent study from six international SSc centers demonstrated that the reflux scale is sensitive to change in patients with SSc and active gastroesophageal reflux disease (24).

The SF-36, HAQ-DI, LCQ and SGRQ scores were assessed at baseline and every 3 months during the study, whereas the TDI score was assessed at baseline and every 6 months during the study. The GIT 2.0 score was assessed at baseline and at 12 and 24 months.

Statistical analysis. Summary statistics were generated for the SF-36, HAQ-DI, BDI, LCQ, SGRQ, and GIT 2.0 scores at baseline. Between-group comparisons in baseline PRO scores were performed using the Student's *t* test.

The percentage of participants who met or exceeded the threshold for the minimum clinically important difference (MCID) for each of the PROs was analyzed using the χ^2 test. The MCID is the smallest improvement in the PRO score necessary for the patients to perceive an improvement that is meaningful to them (25). MCID estimates are captured at a group level. The MCID for each PRO is summarized in Table 1 (26–31).

We used Pearson's correlation coefficient to examine the relationship between baseline PRO scores and the extent of physiologic impairment (baseline FVC%-predicted and diffusing capacity for carbon monoxide [DLCO]), structural lung disease (QLF and QILD), and the modified Rodnan skin score (mRSS) as well as the relationship between the change in PRO scores and the change in FVC%-predicted, DLCO percentage predicted (DLCO%-predicted), mRSS, QLF, and QILD scores. We included all of the radiographic imaging scores because we wanted to understand whether changes in certain structural parameters correlated better with PROs (eg, improvement in the total lung versus improvement in the ZM). Pearson correlation coefficients were interpreted as proposed by Cohen (32): 0.1, small correlation; 0.3, medium correlation; and 0.5, large correlation. We did not correct for multiple hypothesis testing.

For the SGRQ, TDI, and HAQ-DI scores, an inferential joint model was used to examine the course of the PRO score over the course of the study. The joint model consisted of a mixed-effects model for longitudinal outcomes and a survival model to handle nonignorable missing data caused by study dropout, treatment failure, or death (ie, likely related to disease or treatment and therefore not random) (33). Fixed effects for the longitudinal portion of the joint model included treatment assignment, a time trend, the PRO at baseline, and a treatment group by time trend interaction. The time trend was modeled by linear splines

with knots at 12 and 21 months, except for the TDI, which only included a knot at 12 months because the TDI was only collected every 6 months.

Statistical analyses were performed using SAS version 9.4 (SAS Institute, Inc.), except for the joint modeling, which was conducted in R (R Foundation for Statistical Computing). *P* values less than 0.05 were considered statistically significant.

RESULTS

Main SLS II study findings. In SLS II, 73 patients with SSc-ILD were randomized to receive CYC, and 69 patients were randomized to receive MMF. Participants were predominantly women (74%) with an average age of 52 years, an average disease duration of 2.6 years, and a moderate degree of restriction on pulmonary function testing (9). Both treatment arms had similar scores for PROs at baseline (see Supplementary Table S1 for a complete list of baseline characteristics). Moreover, both treatment arms experienced a significant improvement in the course of the FVC%-predicted over 24 months, with an average absolute improvement of 3.0 and 3.3 in the CYC and MMF arms, respectively (9). In addition to improvement in the FVC%-predicted from baseline, treatment with MMF and CYC also led to significant improvements in the course of the mRSS as well as the QILD-WL score, with no between-treatment differences (9,34).

Treatment with CYC and MMF improves breathlessness and respiratory HRQOL. Supporting the results of the primary SLS II article, the present analysis found significant improvements in overall HRQOL in patients who participated in SLS II. Participants randomized to receive CYC and MMF experienced significant improvements in self-reported dyspnea and respiratory-related HRQOL based on the TDI and SGRQ, respectively. Figure 1 demonstrates the course of the TDI score in SLS II based on the joint model analysis. At 12, 18, and 24 months, participants in both the CYC and MMF arms experienced significant improvements in breathlessness relative to baseline, as measured by the TDI (Figure 1, Supplementary Table S2). There was no difference in the course of the TDI score between participants randomized to receive CYC versus those randomized to receive MMF (Supplementary Table S2).

Table 1. MCID scores for the PRO instruments examined in SLS II

Instrument	Description	MCID	Interpretation
SF-36 (26)	Measures health status	≥5	Increase in score indicates improvement
HAQ-DI (27)	Measures functional ability	≤−0.14	Decrease in score indicates improvement
TDI (29)	Measures dyspnea	≥1	Increase in score indicates improvement
LCQ (30)	Measures cough	≥1.5	Increase in score indicates improvement
SGRQ (31)	Measures health status and well-being	≤−4.0	Decrease in score indicates improvement
UCLA GIT 2.0. (total score) (28)	Measures gastrointestinal tract involvement	<−0.21	Decrease in score indicates improvement

Abbreviation: HAQ-DI, Health Assessment Questionnaire Disability Index; LCQ, Leicester Cough Questionnaire; MCID, minimum clinically important difference; PRO, patient-reported outcome; SF-36, Short Form 36; SGRQ, St. George's Respiratory Questionnaire; SLS, Scleroderma Lung Study; TDI, Transitional Dyspnea Index; UCLA GIT 2.0, University of California, Los Angeles Gastrointestinal Tract 2.0.

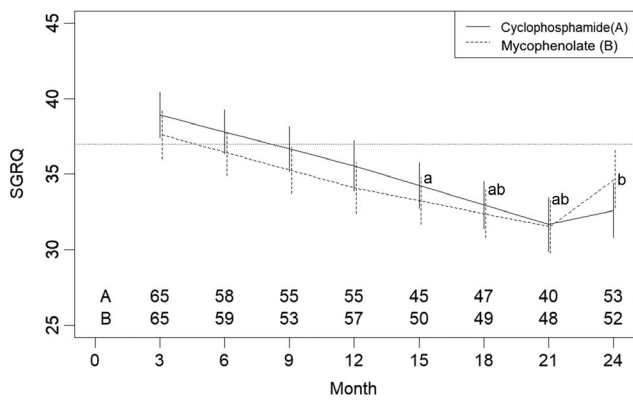


Figure 1. Course of the St. George’s Respiratory Questionnaire (SGRQ) score over 24 months by treatment arm. The solid line represents cyclophosphamide (CYC), and the dotted line represents mycophenolate mofetil (MMF). The horizontal line represents the mean baseline SGRQ score for the entire Scleroderma Lung Study II cohort. ^aA significant change from baseline within the MMF arm at 15, 18, and 21 months. ^bA significant change from baseline within the CYC arm at 18, 21, and 24 months. Please see Supplemental Table S3 for a complete summary of joint model results.

Similar to the TDI, the course of the SGRQ score improved over the course of SLS II based on the joint model analysis (Figure 2, Supplementary Table S3). At 15, 18, and 24 months, participants randomized to receive CYC experienced statistically significant improvements in the SGRQ score relative to baseline, whereas at 18 and 24 months, participants randomized to receive MMF experienced significant improvements in the SGRQ score relative to baseline. There was a slight increase in SGRQ scores (worsening) from 21 to 24 months, when participants who had prematurely stopped the study drug were invited to return for the final study visit. There was no difference in the course of the SGRQ score between participants randomized to receive CYC and those randomized to receive MMF (Supplementary Table S3).

The course of the HAQ-DI score also improved over the course of the study, although the change was not statistically significant within or between treatment arms (Supplementary Figure S1, Supplementary Table S4). For a summary of the change in scores for each PRO, please see Supplementary Table S5.

Proportion of participants treated with MMF and CYC whose PRO scores improved more than the MCID.

For each PRO, a number of participants in the CYC and MMF arms met or exceeded the MCID estimates at both 12 and 24 months (Table 2). There were no differences between treatment arms in the proportion of participants whose PRO scores improved more than the MCID at either time point. For SGRQ, 28% and 24% of participants randomized to receive CYC and MMF, respectively, met or exceeded the MCID for this outcome at 24 months. For the TDI, 29% and 24% of participants randomized to receive CYC and MMF, respectively, met or exceeded the MCID for this outcome at 24 months. For the HAQ-DI, 17% and 14% of participants randomized to receive CYC and MMF, respectively, met or exceeded the MCID for this outcome at 24 months.

omized to receive CYC and MMF, respectively, met or exceeded the MCID for this outcome at 24 months.

Overall generic health status also improved for a number of participants based on the results of the SF-36 (Table 2). In participants randomized to receive both CYC and MMF, 17% of participants met or exceeded the MCID for the SF-36 PCS at 24 months. For the SF-36 MCS, 20% and 21% of participants randomized to receive CYC and MMF, respectively, met or exceeded the MCID for this outcome at 24 months. A relatively smaller percentage of participants met or exceeded the MCID for the total GIT 2.0 score at 24 months (CYC: 13%; MMF: 10%) (Table 2).

Relationship between PRO measures and objective SSC-ILD disease severity measures.

At baseline, the FVC%-predicted did not correlate with any of the PRO measures, except for a weak correlation with the LCQ score (Table 3). There were medium correlations between the DLCO%-predicted and the LCQ and SGRQ scores (Table 3). The BDI, SGRQ, and LCQ scores all correlated with the extent of quantitative radiographic fibrosis and ILD, demonstrating that patients with more self-reported dyspnea at baseline had worse diffusing capacity and increased extent of radiographic fibrosis. The HAQ-DI score did not correlate with any of the objective measures of SSC-ILD disease severity, nor did the SF-36 MCS score, GIT 2.0 total score, or GIT 2.0 reflux score (Table 3).

The change in the FVC%-predicted significantly correlated with the change in the SF-36, HAQ-DI, TDI, and SGRQ scores at 24 months, indicating that patients who experienced an improvement in their FVC%-predicted experienced parallel improvements in their dyspnea and HRQOL (Table 4). Of all of the PRO

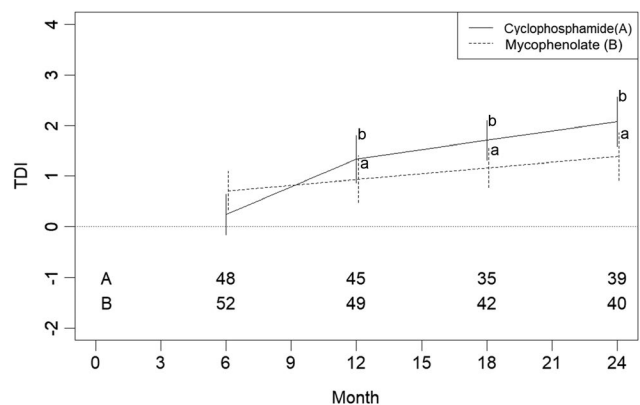


Figure 2. Course of the Transitional Dyspnea Index (TDI) score over 24 months by treatment arm. The solid line represents cyclophosphamide (CYC), and the dotted line represents mycophenolate mofetil (MMF). The horizontal line represents the mean Baseline Dyspnea Index score for the entire Scleroderma Lung Study II cohort. ^aA significant change from baseline within the MMF arm at 12, 18, and 24 months. ^bA significant change from baseline within the CYC arm at 12, 18, and 24 months. Please see Supplemental Table S2 for a complete summary of joint model results.

Table 2. Number of participants meeting the MCID scores for the PRO instruments examined in SLS II at 12 and 24 mo

PRO	n (%) at 12 Mo		<i>P</i> ^a	n (%) at 24 mo		<i>P</i> ^a
	CYC	MMF		CYC	MMF	
SF-36 PCS	19 (16.9)	17 (15.0)	0.55	18 (17.0)	18 (17.0)	1
SF-36 MCS	23 (20.4)	22 (19.5)	0.67	21 (19.8)	22 (20.8)	0.84
HAQ-DI	20 (17.7)	15 (13.3)	0.23	18 (17.0)	15 (14.2)	0.53
TDI	21 (21.9)	19 (19.8)	0.56	23 (29.1)	19 (24.1)	0.31
LCQ	17 (15.6)	11 (10.1)	0.17	16 (15.4)	13 (12.5)	0.51
SGRQ	31 (27.7)	28 (25.0)	0.44	29 (27.6)	25 (23.8)	0.5
UCLA GIT 2.0 (total score)	16 (14.3)	11 (9.8)	0.23	14 (13.3)	10 (9.52)	0.38

Abbreviation: CYC, cyclophosphamide; HAQ-DI, Health Assessment Questionnaire Disability Index; LCQ, Leicester Cough Questionnaire; MCID, minimum clinically important difference; MCS, mental component summary; MMF, mycophenolate mofetil; PCS, physical component summary; PRO, patient-reported outcome; SF-36, Short Form 36; SGRQ, St. George's Respiratory Questionnaire; SLS, Scleroderma Lung Study; TDI, Transitional Dyspnea Index; UCLA GIT 2.0, University of California, Los Angeles Gastrointestinal Tract 2.0.

^a*P* value for differences between treatment arms.

measures, the change in breathlessness (assessed by the TDI) correlated significantly with changes in the most objective measures of SSc-ILD severity (FVC%-predicted, DLCO%-predicted, QLF-WL, QILD-LM, and QILD-WL), whereas the change in the SGRQ score correlated only with the change in the FVC%-predicted, DLCO%-predicted, and QILD-LM score but not with changes in the other radiographic ILD and fibrosis scores. Similarly, the change in the HAQ-DI score only modestly correlated with the change in the FVC%-predicted and DLCO%-predicted (Table 4).

The relationship between cutaneous sclerosis and PROs in patients with SSc-ILD. The baseline mRSS correlated significantly with the baseline HAQ-DI and LCQ scores in all patients and in patients with dcSSc but not in patients with lcSSc (Table 3). At 24 months, 42% and 49% of patients randomized to the CYC and MMF arms, respectively, met or exceeded the MCID for the mRSS (defined as a decline of five points). An improvement in the mRSS from baseline to 24 months correlated significantly

with an improvement in the SF-36, HAQ-DI, and TDI scores in all patients (Table 4). Improvements in the mRSS correlated significantly with improvements in the SF-36 PCS, HAQ-DI, and TDI scores in patients with dcSSc, but only an improvement in the HAQ-DI score was associated with an improvement in the mRSS in patients with lcSSc (Table 4).

Of note, there were no significant correlations between the baseline mRSS and FVC%-predicted in all patients or in patients with dcSSc or lcSSc. There was a significant correlation between improvement in the mRSS and an improvement in the FVC%-predicted in all patients ($r = -0.24$) and in patients with dcSSc ($r = -0.28$) but not in patients with lcSSc.

DISCUSSION

Treatment with immunosuppression is typically the first-line therapeutic approach for patients with SSc-ILD (35). The results of this study affirm that this approach improves PRO measures, in addition to lung function and radiographic measures of ILD, in

Table 3. Baseline correlations between PROs and objective measures of SSc-ILD disease severity in SLS II^a

	FVC%	DLCO%	QLF-ZM	QLF-WL	QILD-ZM	QILD-WL	mRSS, All	mRSS, dcSSc	mRSS, lcSSc
SF-36 PCS	0.05	0.18*	-0.02	-0.07	-0.13	-0.16	-0.14	-0.03	0.18
SF-36 MCS	0.1	0.06	-0.08	-0.12	-0.13	-0.12	-0.01	0.00	-0.09
HAQ-DI	0.07	0.01	-0.10	-0.07	0.01	0.04	0.41*	0.30*	-0.01
BDI	0.09	0.22*	-0.18*	-0.19*	-0.25*	-0.25*	0.11	0.18	0.07
LCQ	0.20*	0.33*	-0.27*	-0.32*	-0.32*	-0.33*	0.26*	0.29*	0.08
SGRQ	-0.14	-0.30*	0.21*	0.29*	0.30*	0.33*	-0.11	-0.15	-0.07
UCLA GIT 2.0 (total score)	-0.08	-0.06	-0.09	0.03	-0.02	0.06	-0.04	-0.11	-0.05
UCLA GIT 2.0 (reflux)	-0.06	-0.09	-0.10	0.02	0.00	0.07	-0.04	-0.24*	0.08

Abbreviation: BDI, Baseline Dyspnea Index; dcSSc, diffuse systemic sclerosis; DLCO%, diffusing capacity for carbon monoxide; FVC%, forced vital capacity; HAQ-DI, Health Assessment Questionnaire Disability Index; LCQ, Leicester Cough Questionnaire; lcSSc, limited systemic sclerosis; MCS, mental component summary; mRSS, modified Rodnan skin score; PCS, physical component summary; QILD, quantitative interstitial lung disease; QLF, quantitative lung fibrosis; SF-36, Short Form 36; SGRQ, St. George's Respiratory Questionnaire; SLS, Scleroderma Lung Study; SSc-ILD, systemic sclerosis-related interstitial lung disease; UCLA GIT 2.0, University of California, Los Angeles Gastrointestinal Tract 2.0; WL, whole lung; ZM, zone of maximal involvement.

^aBold values denote medium correlation coefficients ($r \geq 0.3$).

* $P < 0.05$.

Table 4. Correlations between the change in PROs and the change in objective measures of SSc-ILD disease severity in SLS II from baseline to 24 mo^a

	FVC%	DLCO%	QLF-ZM	QLF-WL	QILD-ZM	QILD-WL	mRSS, All	mRSS, dcSSc	mRSS, lcSSc
SF-36 PCS	0.30*	0.29*	-0.14	-0.27*	-0.18	-0.20	-0.30*	-0.37*	-0.008
SF-36 MCS	0.08	0.15	0.08	-0.02	0.06	0.08	0.05	0.03	-0.04
HAQ-DI	-0.32*	-0.34*	0.02	0.04	0.07	0.00	0.36*	0.41*	0.31*
TDI	0.42*	0.28*	-0.14	-0.24*	-0.26*	-0.29*	-0.35*	-0.47*	0.05
LCQ	0.06	0.19	0.12	-0.04	-0.12	-0.15	-0.06	-0.09	0.04
SGRQ	-0.29*	-0.32*	0.03	0.14	0.21*	0.15	0.14	0.2	0.08
UCLA GIT 2.0 (total score)	-0.13	-0.12	-0.02	0.03	0.04	0.02	0.14	0.12	0.21
UCLA GIT 2.0 (reflux)	-0.05	-0.09	-0.06	0.07	0.10	0.17	0.11	0.07	0.24

Abbreviation: dcSSc, diffuse systemic sclerosis; DLCO%, diffusing capacity for carbon monoxide; FVC%, forced vital capacity; HAQ-DI, Health Assessment Questionnaire Disability Index; LCQ, Leicester Cough Questionnaire; lcSSc, limited systemic sclerosis; MCS, mental component summary; mRSS, modified Rodnan skin score; PCS, physical component summary; QILD, quantitative interstitial lung disease; QLF, quantitative lung fibrosis; SF-36, Short Form 36; SGRQ, St. George's Respiratory Questionnaire; SLS, Scleroderma Lung Study; SSc-ILD, systemic sclerosis-related interstitial lung disease; TDI, Transitional Dyspnea Index; UCLA GIT 2.0, University of California, Los Angeles Gastrointestinal Tract 2.0; WL, whole lung; ZM, zone of maximal involvement.

^aBold values denote at least medium correlation coefficients ($r \geq 0.3$).

* $P < 0.05$.

patients with this condition. Specifically, this study found that treatment with MMF and CYC led to significant and clinically meaningful improvements in self-reported dyspnea, health status, and physical function in patients with SSc-ILD.

Although the FVC is frequently used as the primary outcome measure in SSc-ILD clinical trials, changes in the FVC may not consistently translate into clinically meaningful improvements from a patient's perspective. For example, in the SENSICIS trial, patients with SSc-ILD treated with nintedanib had a lower rate of annual FVC decline than those receiving a placebo (treatment difference of 41 ml); however, there was no significant difference in PROs (based on SGRQ or the HAQ-DI) between treatment groups (5). In contrast, both SLS I (6) and our current analyses showed clinically meaningful improvements in PROs with treatment of SSc-ILD. Both SLS II (9) and SENSICIS (5) had similar baseline characteristics (FVC%-predicted, DLCO%-predicted, disease duration, mRSS, and SGRQ); however, there are several plausible explanations for these discrepancies. First, in SLS II (9), there was an overall trend for improvement in the FVC%-predicted, whereas in SENSICIS (5), there was an overall trend for a decline in the FVC%-predicted. Second, PROs capture the overall impact of an intervention on the whole person. Treatment with MMF and CYC favorably affects extrapulmonary manifestations of SSc, including cutaneous sclerosis, and this may in turn affect PROs (36). Our findings indicate that improvements in cutaneous sclerosis, particularly among patients with dcSSc, are associated with improvements in PROs (eg, HAQ-DI, TDI, and SF-36 PCS).

With the present results, we found no significant correlation at baseline between the FVC%-predicted and dyspnea (BDI or HRQOL), with the exception of a weak correlation with the LCQ. On the other hand, at baseline, the DLCO%-predicted and the quantitative extent of ILD and fibrosis scores significantly correlated with the BDI and SGRQ. These findings suggest that when evaluating disease severity of SSc-ILD, these additional

assessment measures may provide a more comprehensive understanding of a patient's experience with the disease and could help inform treatment decisions.

We did find that an improvement in the FVC%-predicted and DLCO%-predicted was significantly associated with an improvement in HRQOL (SF-36, HAQ-DI, and SGRQ) and dyspnea (TDI). These results illustrate that underlying changes in lung function and physiology may lead to meaningful changes in how a patient feels and functions. Consistent with this hypothesis, our prior analysis of SLS I and II demonstrated that clinically meaningful improvements in the FVC%-predicted were associated with improvements in the PCS, TDI, and HAQ-DI scores (37).

We also examined the course of the TDI and SGRQ scores using a joint model approach to adjust for nonignorable missing data and baseline disease severity and demonstrated significant improvements in these measures over the 2-year trial. For the TDI, there was a steady increase (improvement) in TDI scores during the trial in both treatment groups. Similarly, there was a steady decrease (improvement) in SGRQ scores during the trial in both treatment groups. Both the TDI and SGRQ scores continued to improve in year 2, even in patients randomized to receive 1 year of CYC followed by 1 year of a placebo. Of note, 10 participants in the CYC arm began treatment with potentially disease-modifying immunosuppressant therapy during year 2 of the study (azathioprine [$n = 2$], MMF [$n = 7$], and CYC [$n = 1$]) (38), raising the possibility that the improvements that occurred during this time period could have been influenced by this additional therapy. In SLS I, we also appreciated significant improvements in the course of the TDI score beyond the 12-month treatment period (persisted until 18 months), suggesting that the effects of CYC persist even after the treatment is stopped for at least 6 months (39).

We observed a slight worsening of SGRQ scores at month 24 compared with month 21 in both treatment groups. This was likely due to the fact that patients who withdrew from active treatment

during the study were encouraged to return for the final 24-month study visit; therefore, this SGRQ assessment included participants who were not on active therapy.

The improvements appreciated in the joint model analysis of the SGRQ and TDI scores mirrored the improvements we observed in the joint model of analysis of the FVC%-predicted in SLS II (9). The peak improvement in the FVC%-predicted in both treatment arms occurred at 21 months. In the present analysis, we also observed peak improvements in the SGRQ score at this time point (the TDI score was not obtained at 21 months). These findings further support the results of our correlation analyses, demonstrating that improvements in lung function are accompanied by parallel improvements in patients' perception of their breathlessness.

Changes in quantitative radiographic scores for lung fibrosis and ILD correlated poorly with changes in all of the PROs, with the exception of dyspnea as assessed by the TDI. These findings suggest that the relationship between the radiographic progression of ILD and how a patient feels and functions is likely influenced by other factors. These factors could include a patient's level of physical conditioning and the presence of comorbidities that limit mobility, such as arthritis. However, the participants in this trial did experience improvements in their overall health status (eg, SF-36 MCS, SF-36 PCS, and HAQ-DI). Taken together, these results may signify that treatment with CYC and MMF improves health outcomes in patients with SSc by also exerting beneficial effects on extrapulmonary features of SSc (eg, improvements in cutaneous sclerosis and arthritis).

The results of this study should be interpreted in the context of specific limitations. A substantial number of patients withdrew prematurely from active treatment during the study (CYC: 44%; MMF: 30%). Although the joint model analysis adjusts for nonignorable missing data from dropouts (33), the high attrition rate in this trial could lead to biased estimates. We are reassured, however, that the improvements we observed in all of the PROs were robust and sustained over the course of the trial. Another limitation is that although most of the PROs were assessed every 3 months, the TDI score was only assessed every 6 months. Additional data points for the TDI, especially earlier in the course of the trial, may have allowed us to further explore how breathlessness changes in response to treatment with CYC and MMF. Nonetheless, the joint model results clearly demonstrate an improvement in the course of the TDI score with these therapies. Finally, the MCID estimates applied in the article may be influenced by different aspects of the disease and other medical conditions. The data should be interpreted with this caveat in mind.

Strengths of this work include the scientific rigor of SLS II, a study that used the expertise of experienced SSc investigators at each study site and went to great lengths to ensure quality data collection and management. Another strength of this study is the use of a diverse array of PROs. Because SSc is a systemic disease that affects multiple organ systems, it is important to understand

how various PROs change in response to particular treatments. As new treatment options for SSc emerge, understanding how these various treatment options affect different aspects of a patient's overall health may help guide treatment selection and continuation.

In summary, treatment with oral CYC and MMF led to clinically meaningful improvements in overall health status, function, and breathlessness in patients with SSc-ILD, bearing in mind the limitations noted previously. Improvements in breathlessness paralleled improvements in lung function to a modest degree despite the finding that the baseline level of restrictive ventilatory impairment did not correlate with how a patient felt. Above all, the findings of this study demonstrate that a comprehensive evaluation combining pulmonary physiology, the radiographic extent of fibrosis, and PROs is essential to understanding the impact of treatment on progression of ILD in SSc.

ACKNOWLEDGMENTS

We thank the patients, investigators, and coordinators who participated in the SLS II. The following persons and institutions participated in SLS II: A. C. Theodore, R. W. Simms, E. Kissin, and F. Y. Cheong (University of Boston, Boston, MA); V. D. Steen, C. A. Read Jr., C. Fridley, and M. Zulmatashvili (Georgetown University, Washington, DC); R. A. Wise, F. M. Wigley, L. Hummers, and G. Leatherman (Johns Hopkins University, Baltimore, MD); R. M. Silver, C. Strange, F. N. Hant, J. Ham, K. Gibson, and D. Rosson (Medical University of South Carolina, Charleston, SC); D. P. Tashkin, R. M. Elashoff, M. D. Roth, P. J. Clements, D. Furst, E. R. Volkman, S. Kafaja, E. Kleerup, D. Elashoff, J. Goldin, E. Ariola, G. Marlis, J. Mason-Berry, P. Saffold, M. Rodriguez, L. Guzman, and J. Brook (University of California, Los Angeles, Los Angeles, CA); J. Golden, M. K. Connolly, A. Eller, D. Leong, M. Lalosh, and J. Obata (University of California, San Francisco, San Francisco, CA); S. Volkov, D. Schraufnagel, S. Arami, and D. Franklin (University of Illinois, Chicago, IL); J. Varga, J. Dematte, M. Hinchcliff, C. DeLuca, H. Donnelly, and C. Marlin (Northwestern University, Chicago, IL); D. J. Riley, V. M. Hsu, and D. A. McCloskey (University of Medicine and Dentistry of New Jersey, New Brunswick, NJ); K. Phillips, D. Khanna, F. J. Martinez, E. Schioppa, and J. Konkle (University of Michigan, Ann Arbor, MI); M. Mayes, B. Patel, S. Assassi, and F. Tan (University of Texas, Houston, TX); A. Fischer, J. Swigris, R. Meehan, K. Brown, T. Warren, and M. Morrison (National Jewish Health, Denver, CO); M. B. Scholand, T. Frecht, P. Carey, and M. Villegas (University of Utah, Salt Lake City, UT); and J. Molitor and P. Carlson (University of Minnesota, Minneapolis, MN).

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published.

Study concept and design. Volkman, Tashkin, Roth, Khanna.

Acquisition of data. Volkman, Tashkin, Roth, Kim, Goldin, Clements, Furst, Khanna.

Analysis and interpretation of data. Volkman, Tashkin, LeClair, Khanna.

REFERENCES

1. Khanna D, Hays RD, Furst DE. Functional disability and other health-related quality-of-life domains: points to consider for clinical trials in systemic sclerosis. *Rheumatology (Oxford)* 2017;56 Suppl 5:v17–22.
2. Lumetti F, Baron L, Alfieri C, Silva M, Serra V, Delsante G, et al. Quality of life and functional disability in patients with interstitial lung disease related to systemic sclerosis. *Acta Biomed* 2015;86:142–8.
3. Hudson M, Thombs BD, Steele R, Panopalis P, Newton E, Baron M, et al. Health-related quality of life in systemic sclerosis: a systematic review. *Arthritis Rheum* 2009;61:1112–20.
4. Khanna D, Yan X, Tashkin DP, Furst DE, Elashoff R, Roth MD, et al. Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: results from the scleroderma lung study. *Arthritis Rheum* 2007;56:1676–84.
5. Distler O, Highland KB, Gahlemann M, Azuma A, Fischer A, Mayes MD, et al. Nintedanib for systemic sclerosis-associated interstitial lung disease. *N Engl J Med* 2019;380:2518–28.
6. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006;354:2655–66.
7. Tashkin DP, Volkman 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;151:813–20.
8. Theodore AC, Tseng CH, Li N, Elashoff RM, Tashkin DP. Correlation of cough with disease activity and treatment with cyclophosphamide in scleroderma interstitial lung disease: findings from the Scleroderma Lung Study. *Chest* 2012;142:614–21.
9. Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomized controlled, double-blind, parallel group trial. *Lancet Resp Med* 2016;4:708–19.
10. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). *Arthritis Rheum* 1980;23:581–90.
11. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea: contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest* 1984;85:751–8.
12. Kim HJ, Tashkin DP, Clements PJ, Li G, Brown MS, Elashoff R, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol* 2010;28 Suppl 62:S26–35.
13. Saris-Baglama RN, Dewey CJ, Chisholm GB, Plumb E, King J, Kosinski MA, et al. QualityMetric health outcomes™ scoring software 4.0: installation guide. Lincoln (RI): QualityMetric Incorporated; 2010.
14. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
15. Mahler DA, Ward J, Fierro-Carrion G, Waterman LA, Lentine TF, Mejia-Alfaro R, et al. Development of self-administered versions of modified baseline and transition dyspnea indexes in COPD. *COPD* 2004;1:165–72.
16. 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;58:339–43.
17. Beretta L, Santaniello A, Lemos A, Masciocchi M, Scorza R. Validity of the Saint George's Respiratory Questionnaire in the evaluation of the health-related quality of life in patients with interstitial lung disease secondary to systemic sclerosis. *Rheumatology (Oxford)* 2007;46:296–301.
18. Khanna D, Hays RD, Maranian P, Seibold JR, Impens A, Mayes MD, et al. Reliability and validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. *Arthritis Rheum* 2009;61:1257–63.
19. 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;42:2372–80.
20. Khanna D, Furst DE, Clements PJ, Park GS, Hays RD, Yoon J, et al. Responsiveness of the SF-36 and the Health Assessment Questionnaire Disability Index in a systemic sclerosis clinical trial. *J Rheumatol* 2005;32:832–40.
21. Barr JT, Schumacher GE, Freeman S, LeMoine M, Bakst AW, Jones PW. American translation, modification, and validation of the St. George's Respiratory Questionnaire. *Clin Ther* 2000;22:1121–45.
22. 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;54:1369–79.
23. Bae S, Allanore Y, Furst DE, Bodukam V, Coustet B, Morgaceva O, et al. Associations between a scleroderma-specific gastrointestinal instrument and objective tests of upper gastrointestinal involvements in systemic sclerosis. *Clin Exp Rheumatol* 2013;31 Suppl 76:57–63.
24. McMahan ZH, Frech T, Berrocal V, Lim D, Bruni C, Matucci-Cerinic M, et al. Longitudinal assessment of patient-reported outcome measures in systemic sclerosis patients with gastroesophageal reflux disease: Scleroderma Clinical Trials Consortium. *J Rheumatol* 2019;46:78–84.
25. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15.
26. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;45:384–91.
27. Khanna D, Furst DE, Hays RD, Park GS, Wong WK, Seibold JR, et al. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. *Ann Rheum Dis* 2006;65:1325–9.
28. Khanna D, Furst DE, Maranian P, Seibold JR, Impens A, Mayes MD, et al. Minimally important differences of the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. *J Rheumatol* 2011;38:1920–4.
29. Witek TJ Jr, Mahler DA. Meaningful effect size and patterns of response of the transition dyspnea index. *J Clin Epidemiol* 2003;56:248–55.
30. Birring S, Muccino D, Bacci ED, Vernon MK, Nguyen AM. Defining minimal clinically important differences (MCID) on the Leicester Cough Questionnaire (LCQ): analyses of a phase 2 randomized controlled trial in chronic cough. *J Allergy Clin Immunol* 2019;143 Suppl:AB52.
31. Jones PW. St. George's Respiratory Questionnaire: MCID. *COPD* 2005;2:75–9.
32. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.

33. Li N, Elashoff RM, Li G, Tseng CH. Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Stat Med* 2012; 31:1707–21.
34. Goldin JG, Kim GH, Tseng CH, Volkmann E, Furst D, Clements P, et al. Longitudinal changes in quantitative interstitial lung disease on computed tomography after immunosuppression in the Scleroderma Lung Study II. *Ann Am Thorac Soc* 2018;15:1286–95.
35. Fernandez-Codina A, Walker KM, Pope JE, Scleroderma Algorithm Group. Treatment algorithms for systemic sclerosis according to experts. *Arthritis Rheumatol* 2018;70:1820–8.
36. Namas R, Tashkin DP, Furst DE, Wilhalme H, Tseng CH, Roth MD, et al. Efficacy of mycophenolate mofetil and oral cyclophosphamide on skin thickness: post hoc analyses from two randomized placebo-controlled trials. *Arthritis Care Res (Hoboken)* 2018;70:439–44.
37. Kafaja S, Clements PJ, Wilhalme H, Tseng CH, Furst DE, Kim GH, et al. Reliability and minimal clinically important differences of forced vital capacity: results from the Scleroderma Lung Studies (SLS-I and SLS-II). *Am J Resp Crit Care Med* 2018;197:644–52.
38. Volkmann ER, Tashkin DP, Sim M, Li N, Khanna D, Roth MD, et al. Cyclophosphamide for systemic sclerosis-related interstitial lung disease: a comparison of Scleroderma Lung Study I and II. *J Rheumatol* 2019;46:1316–25.
39. Tashkin DP, Elashoff R, Clements PJ, Roth MD, Furst DE, Silver RM, et al. Effects of 1-year treatment with cyclophosphamide on outcomes at 2 years in scleroderma lung disease. *Am J Respir Crit Care Med* 2007;176:1026–34.