



# Treatment of Scleroderma-Interstitial Lung Disease With Cyclophosphamide Is Associated With Less Progressive Fibrosis on Serial Thoracic High-Resolution CT Scan Than Placebo

## Findings From the Scleroderma Lung Study

Jonathan Goldin, MB ChB, PhD; Robert Elashoff, PhD; Hyun J. Kim, PhD; Xiaohong Yan, PhD; David Lynch, MD; Diane Strollo, MD; Michael D. Roth, MD, FCCP; Philip Clements, MD; Daniel E. Furst, MD; Dinesh Khanna, MD; Srainnapha Vasunilashorn, BS; Gang Li, PhD; and Donald P. Tashkin, MD, FCCP

**Background:** The Scleroderma Lung Study (SLS) demonstrated significant treatment-associated improvements in pulmonary function and symptoms when patients with scleroderma-related interstitial lung disease (SSc-ILD) were treated with a 1-year course of cyclophosphamide (CYC) in a randomized, double-blinded, placebo-controlled clinical trial. This study examined thoracic high-resolution CT (HRCT) scans obtained during the SLS for treatment-associated changes over time.

**Methods:** Ninety-eight of the 158 subjects (CYC group, 49 subjects; placebo group, 49 subjects) participating in the SLS underwent thoracic HRCT scans both at baseline and after 12 months of treatment, which were available for analysis. Two independent radiologists visually scored the baseline HRCT scans for the presence of ground-glass opacities (GGOs), fibrosis (FIB), and honeycomb cysts (HCs) on a scale of 0 to 4. The treatment effect at 12 months was assessed by a blinded comparison of baseline and follow-up scans for evidence of stability and improvement (not worse) or deterioration (worse).

**Results:** At the end of treatment, FIB was significantly worse in the placebo treatment group than in the CYC treatment group ( $p = 0.014$ ). Furthermore, differences in the 12-month change in FIB between the CYC and placebo groups correlated significantly with other outcome measures, including the 12-month changes in FVC ( $p < 0.05$ ), total lung capacity ( $p < 0.05$ ), and dyspnea ( $p < 0.001$ ) scores. However, no differences were noted between the two groups with respect to changes in either GGOs or HCs.

**Conclusions:** A 1-year course of treatment of SSc-ILD with CYC was associated with treatment-related changes in FIB scores on HRCT scans, which correlated with other measures of treatment response.

**Trial registration:** ClinicalTrials.gov Identifier: NCT00004563 (CHEST 2009; 136:1333–1340)

**Abbreviations:** CYC = cyclophosphamide; DLCO = diffusing capacity of the lung for carbon monoxide; FIB = fibrosis; GGO = ground-glass opacity; HAQ-DI = Health Assessment Questionnaire-Disability Index; HC = honeycomb cyst; HRCT = high-resolution CT; SF-36 = Medical Outcomes Study 36-item short-form; SLS = Scleroderma Lung Study; SSc-ILD = scleroderma-related interstitial lung disease; TDI = transitional dyspnea index; TLC = total lung capacity

Systemic sclerosis, scleroderma is a complex and life-threatening autoimmune disease associated with tissue fibrosis (FIB) and small-vessel vasculopathy that target the skin, lungs, heart, GI tract, peripheral circulation, and musculoskeletal system. Lung involvement has emerged as the leading

cause of morbidity and mortality<sup>1–3</sup> and has been the target of a number of clinical investigations.<sup>4</sup> Corticosteroids, D-penicillamine, relaxin, or an endothelin receptor antagonist have not been proven to be effective.<sup>4</sup> However, the results of two prospective randomized clinical trials<sup>5,6</sup> have suggested that

treatment with cyclophosphamide (CYC) can modify pulmonary outcomes. Findings from a recent retrospective, open-label study<sup>7</sup> of 6 months of treatment with IV CYC followed by 18 months of treatment with oral azathioprine showed comparable results.

Pulmonary function test results, particularly for FVC, have been studied<sup>5-10</sup> extensively as the primary measures of treatment efficacy. Although the response to a 1-year course of treatment with CYC in the Scleroderma Lung Study (SLS) was statistically significant, the magnitude of change in FVC percent predicted between patients treated with placebo and those treated with CYC averaged only 2.53% after 12 months of treatment<sup>5</sup> and 4.80% at 18 months (*ie*, 6 months after completing therapy).<sup>11</sup> These modest changes have raised questions about the clinical significance of CYC therapy.<sup>12</sup> Although high-resolution CT (HRCT) scanning has been used to characterize the nature and extent of scleroderma-related interstitial lung disease (SSc-ILD)<sup>13-17</sup> and even to predict outcomes,<sup>18-20</sup> there is limited experience<sup>21-24</sup> with its use as an outcome measure in therapeutic trials. HRCT findings<sup>16,17</sup> have suggested that SSc-ILD is associated with a nonspecific interstitial pneumonitis pattern more commonly than a usual interstitial pneumonitis pattern. Classically, nonspecific interstitial pneumonitis is characterized by finer FIB and a higher proportion of ground-glass opacity (GGO), and is associated with a better prognosis than is usual interstitial pneumonitis.<sup>16</sup>

In the SLS,<sup>5</sup> a greater amount of HRCT scan-detected FIB on the baseline scan was found to be predictive of a positive treatment response to CYC than placebo at 12 months. In the present study, we examined the difference in parenchymal abnormalities between baseline and follow-up HRCT scans

performed after patients completed 1 year of blinded treatment with either CYC or placebo. We hypothesized that treatment with CYC is associated with a lack of progression and improvement in HRCT evidence of parenchymal abnormalities, and that treatment-associated changes seen on HRCT scans would correlate with changes in pulmonary function and patient-reported outcome scores.

## MATERIALS AND METHODS

### *Patient Selection*

The SLS was a multicenter, randomized, double-blind comparison of a 1-year course of treatment with CYC ( $\leq 2$  mg/kg/d) vs treatment with placebo in 158 patients with dyspnea with active SSc-ILD as defined by evidence on either a HRCT scan (the presence of any GGO through which lung architecture could be seen) or BAL fluid testing ( $\geq 3.0\%$  neutrophils,  $\geq 2.0\%$  eosinophils, or both).<sup>5</sup> Pulmonary function was assessed every 3 months during the first year; the primary end point was FVC percent predicted at 12 months adjusted for baseline FVC percent predicted and the degree of FIB seen on the baseline HRCT scan. Secondary end points included the percent predicted for total lung capacity (TLC) and single-breath diffusing capacity of the lung for carbon monoxide (DLCO); dyspnea was assessed using the transitional dyspnea index (TDI),<sup>25</sup> and health-related quality of life was assessed using the Medical Outcomes Study 36-item short-form (SF-36) general health survey and the Health Assessment Questionnaire-Disability Index (HAQ-DI).<sup>26</sup>

### *HRCT Scanning*

HRCT scans were obtained at the 13 clinical sites with one of four CT platforms (Somatom Plus 4; Siemens; Erlangen, Germany; TCT-900S; Toshiba Medical Systems; Tokyo, Japan; or HiSpeed Advantage or HiSpeed Cti; GE Medical Systems; Milwaukee, WI) using a standardized protocol that was developed and monitored by a radiology core at the University of California (Los Angeles, CA), as previously described.<sup>16,17</sup> Patients were scanned with 1- or 2-mm collimation at 10-mm intervals from lung apices to bases during suspended end inspiration while in the prone position without IV contrast. Images were reconstructed with a high-spatial frequency algorithm. The SLS protocol included a baseline scan obtained during screening, and this was supplemented by funding from an ancillary study to obtain a second HRCT scan at the end of the 12-month treatment period.

### *Interpretation of HRCT Scans*

Two independent SLS core thoracic radiologists with 15 and 20 years of experience who were blinded to all patient information visually scored the HRCT scans. A specially designed training disk with characteristic HRCT scan features was used to standardize the scoring. At the completion of the independent scoring phase, scans that had been scored in a discordant manner (8 scans for FIB, 3 for GGOs, and 11 for honeycomb cysts [HCs]) were jointly reviewed with a third reader to obtain a final consensus decision. Each lung was divided into three zones (lung apex to aortic arch, aortic arch to inferior pulmonary veins, inferior pulmonary veins to lung bases), and the extent of parenchymal abnormality in each zone was scored on a scale of 0 to 4 (0 = absent; 1 = 1 to 25%; 2 = 26 to 50%; 3 = 51 to 75%;

Manuscript received January 14, 2009; revision accepted June 10, 2009.

**Affiliations:** From the Departments of Radiologic Sciences (Drs. Goldin and Kim, and Ms. Vasunilashorn), Biomathematics (Drs. Elashoff and Yan), and Medicine (Drs. Roth, Clements, Furst, Khanna, and Tashkin), David Geffen School of Medicine, and the Department of Biostatistics (Dr. Li), School of Public Health, University of California Los Angeles, Los Angeles, CA; the Department of Radiologic Sciences (Dr. Lynch), National Jewish Hospital, Denver, CO; and the University of Pittsburgh School of Medicine (Dr. Strollo), Pittsburgh, PA.

**Funding/Support:** This study was funded by research grants R01 HL072424 (Dr. Golding), 5U01 HL60587 (Dr. Tashkin), 5 P30 CA016042-35 (Dr. Li), P01AT003960 (Dr. Li), and 5U01 HL060606 (Dr. Elashoff) from the National Institutes of Health/National Heart, Lung, and Blood Institute.

**Correspondence to:** Donald P. Tashkin, MD, FCCP, Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, 10833 Le Conte Ave, Los Angeles, CA 90095; e-mail: dtashkin@mednet.ucla.edu

© 2009 American College of Chest Physicians. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians ([www.chestjournal.org/site/misc/reprints.xhtml](http://www.chestjournal.org/site/misc/reprints.xhtml)).

DOI: 10.1378/chest.09-0108

4 = 76 to 100%). The following HRCT scan findings were recorded: pure GGO (defined as increased lung attenuation in the absence of reticular interstitial thickening or architectural distortion); lung FIB (defined as reticular intralobular interstitial thickening, traction bronchioectasis, bronchioectasis, or any combination); and HCs (defined as clustered air-filled lung cysts with contiguous walls).

The comparison of baseline and 12-month HRCT scans for each patient was carried out side by side by randomly assigning one as scan A and one as scan B in order to keep readers blinded to which scan was the true baseline scan and which was the true 12-month follow-up scan. Scans A and B for each patient were then assessed for changes in each outcome measure (*ie*, FIB, GGO, or HC) for each lung zone in a dichotomized manner, with scan B read as either worse or not worse (*ie*, same or better) compared with scan A, assuming that scan A was always the true baseline scan (even though it could have been the 12-month scan). The same tie-breaker radiologist scored all cases for change between scans A and B, and the majority vote was recorded as final. Subsequently, the statistician reordered each pair of scans for each patient and readjusted the scoring (worse or not worse) accordingly.

### Pulmonary Function Tests

Pulmonary function tests (spirometry, lung volumes, and DLCO) were performed at baseline and every 3 months during the treatment period, using standardized methods, as previously described.<sup>5</sup>

### BAL

Right-middle-lobe BAL was performed prior to randomization at each center, as previously described,<sup>26</sup> to detect the presence or absence of neutrophilia ( $\geq 3\%$ ) and eosinophilia ( $\geq 2\%$ ).<sup>27</sup> The baseline HRCT scan was performed shortly before BAL or, if it was performed first, at least 4 weeks after BAL.

### Patient-Centered Measures

The cough index, dyspnea index, SF-36, HAQ-DI, and skin thickness (modified Rodnan skin score) were assessed at baseline and every 3 months during the treatment period, as previously reported.<sup>5</sup>

### Statistical Analysis

Baseline characteristics of patients assigned to the CYC and placebo arms of the SLS were compared by an unpaired *t* test or Fisher exact test. Differences in the HRCT scan measures (FIB, GGO, and HC) between the baseline and 12-month scans for each patient were recorded in a dichotomized manner as 0 for worse (if any region of the lung had a worse score) and 1 for not worse (if all regions of the lung were either stable or improved) based on the readers' visual side-by-side assessment of change. The changes in HRCT scan measures from baseline to the 12-month follow-up were then analyzed using logistic regression (analysis of covariance), with separate logistic covariance regressions performed for each variable (FIB, GGO, and HC). Factors included in the covariance model were treatment group assignment (CYC or placebo), baseline maximum HRCT FIB score (the maximum score over the six regions), and baseline FVC percent predicted. Odds ratio estimates for each covariate were calculated from the regression model.

Kendall  $\tau$  rank correlations also were determined between the binary FIB outcome variable (worse or not worse) and the

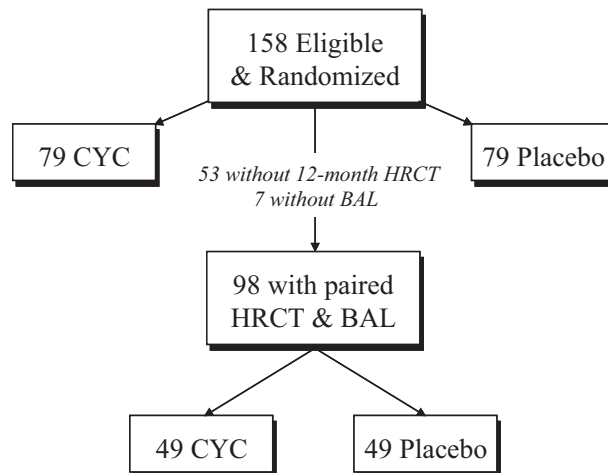


FIGURE 1. Patients with symptomatic SSc-ILD ( $n = 158$ ) were eligible and randomized into the SLS to receive a 12-month course of treatment with either CYC or placebo according to a randomized, double-blind, placebo-controlled study design. Of these, 98 patients met the current study criteria by having both paired baseline and 12-month HRCT scans available for analysis and BAL fluid cell counts. Forty-nine patients were randomized to the CYC arm, and 49 patients were randomized to the placebo arm.

12-month changes from baseline in physiologic measures (FVC, TLC, and DLCO, all as percent predicted). The McNemar test was used to determine the association between the binary FIB outcome and clinically relevant patient-centered outcome variables (*ie*, TDI, modified Rodnan skin score, HAQ-DI, and cough). The latter outcome measures were dichotomized using 0 for any response that was worse by comparing 12-month with baseline values, and 1 for an outcome that was not worse (*ie*, stable or improved). Agreement among the three radiologists with regard to scoring each radiographic feature (FIB, GGO, and HC) at baseline or during the comparison reading was assessed using the  $\kappa$  statistic.<sup>28</sup>

## RESULTS

Of the 158 patients who were eligible and randomized into the SLS, 98 patients (CYC group, 49 patients; placebo group, 49 patients) who had both baseline and 12-month HRCT scans and had completed a BAL procedure were analyzed (Fig 1). Table 1 shows the mean values (or percentage of patients) for clinically relevant baseline demographics and disease characteristics for these 98 patients as well as their worst recorded scores for HRCT scan measures of FIB, GGO, and HC (*ie*, the worst score among all lung zones). The CYC and placebo treatment groups were well balanced for all of these baseline features (Table 1), and this subset of 98 patients was comparable with the entire group of 158 randomized patients.<sup>5</sup>

The main focus of the current analysis was the visually identifiable change over time in HRCT findings of FIB, GGO, and HC using a dichotomous,

**Table 1—Baseline Patient Characteristics With Both Baseline and 12-Month HRCT Scan Readings by Treatment Assignment**

Characteristics	CYC Group (n = 49)	Placebo Group (n = 49)	p Value*
Age, yr	46.1	47.0	0.70
Female gender	38 (87.6)	35 (71.4)	0.49
Disease duration, yr	3.2	3.2	0.92
FVC, % predicted	68.3	69.8	0.52
FEV <sub>1</sub> /FVC ratio	82.3	82.9	0.74
TLC, % predicted	70.0	70.0	0.91
FRC, % predicted	76.0	71.9	0.29
DLCO, % predicted	46.0	47.9	0.50
BDI (0–12)	5.9	5.5	0.22
Cough SF-36 (0–100)	32 (68.1)	36 (73.5)	0.56
Physical component summary	32.9	34.6	0.45
Mental component summary	48.5	50.8	0.29
HAQ-DI (1–3)	0.95	0.73	0.11
Skin score (0–51)	15.5	14.2	0.58
HRCT scan score (range)			
FIB	2.0 (0–4)	1.96 (0–4)	0.89
GGO	0.67 (0–3)	0.65 (0–2)	0.61
HC	0.35 (0–3)	0.41 (0–2)	0.61

Values are given as the mean or No. (%), unless otherwise specified. BDI = baseline dyspnea index; FRC = functional residual capacity. \*Unpaired *t* test.

worse-or-not-worse scoring system as illustrated in Figure 2. At least moderate agreement was noted among the three readers with respect to changes in FIB ( $\kappa = 0.51$ ), but only fair vs poor agreement with respect to changes in GGO ( $\kappa = 0.36$ ) and HC ( $\kappa = 0.16$ ), respectively. Among patients in the CYC group, the FIB was worse at 12 months than at baseline in 14 patients and was not worse in 35 patients. In contrast, among patients in the placebo group, the FIB was worse at 12 months in 26 patients and was not worse in 23 patients (Table 2). FIB worsened in a significantly higher proportion of patients in the placebo group than in the CYC group ( $p = 0.014$  [ $\chi^2$  test]). On the other hand, no differential effect of treatment was noted on dichotomized changes in GGO or HC. When changes in GGO were expressed as worse, better, or no change, treatment with CYC was not found to result in any more frequent resolution of GGOs than treatment with placebo (Table 2).

Table 3 shows the results of the logistic regression for the FIB outcome with covariate adjustments as well as the odds ratios derived from the logistic regression for the binary changes in FIB. Treatment group (CYC vs placebo) significantly influenced the outcome after adjustment for the other variables ( $p = 0.012$ ) with an odds ratio of 3.26 (95% CI, 1.30 to 8.17%) for the treatment group. The baseline

FVC percent predicted was also a strong predictor of the FIB outcome ( $p < 0.01$ ).

Neither the inclusion of the percentage of neutrophils and the percentage of eosinophils from the BAL fluid nor the baseline extent of GGO (maximum score in any of the six lung regions) had any significant effect on these results. Similarly, neither of these variables was found to be a significant predictor of the FIB outcome. Separate logistic regression analyses in which the dependent variables were either GGO or HC also did not show a treatment effect with respect to these variables.

We assessed whether the dichotomized HRCT scan scores (worse vs not worse) for FIB correlated with the mean changes at 12 months from baseline for pulmonary function outcomes (percent predicted FVC, TLC, and DLCO) and the number of patients with worse or not worse scores for skin thickness, HAQ-DI, TDI, or cough. Table 4 lists Kendall  $\tau$  rank correlation coefficients or the McNemar test between these outcomes and the binary outcome for FIB (worse vs not worse). For the TDI, the total scores (sum of the functional, magnitude-of-task, and magnitude-of-effort domains) are shown. A change in TDI that was zero or positive indicated not worse, whereas a change in TDI that was negative indicated worse. The results show a statistically significant, albeit modest, correlation between changes in HRCT FIB and pulmonary function and skin thickness scores (correlation coefficients, approximately 0.2). Weak associations were noted between changes in FIB and HAQ-DI and cough (correlation coefficients, 0.071 vs 0.065, respectively). Changes over time in dyspnea (total TDI and the three domains) showed moderate and significant correlations with changes in FIB.

## DISCUSSION

SSc-ILD is a serious and therapeutically challenging manifestation of systemic sclerosis, scleroderma. Although a number of agents have been evaluated as treatments for SSc-ILD, only oral CYC has been proven to be effective in a randomized controlled study<sup>5</sup> demonstrating statistically significant placebo-adjusted improvements in FVC and TLC after 12 months of therapy. A small randomized controlled trial<sup>6</sup> using IV CYC showed similar results to those with oral CYC, reinforcing this finding. The current analysis provides for the first time a visual confirmation of this treatment efficacy by demonstrating less worsening of FIB on serial HRCT scans (baseline to 12 months) in patients treated with CYC than in those treated with placebo.

It has been argued that although the placebo-adjusted improvements in FVC and TLC in response

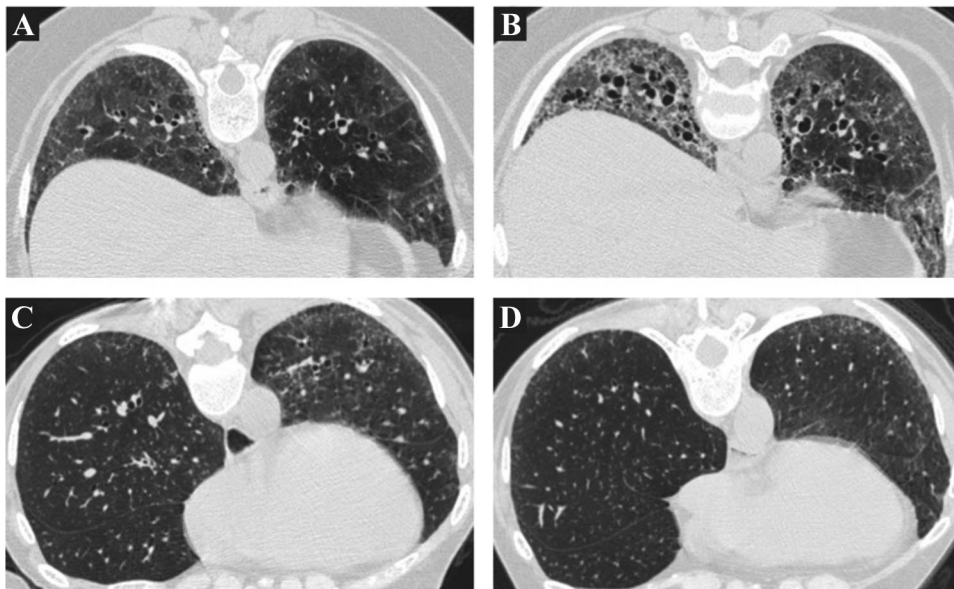


FIGURE 2. Representative paired images from a patient in the placebo arm whose FIB was visually scored as being worse over time (A, baseline; B, 12-month follow-up) and a patient in the CYC arm whose serial scans were scored as not worse (C, baseline; D, 12-month follow-up).

to treatment with CYC were statistically significant in the SLS,<sup>5</sup> they were small in magnitude.<sup>11</sup> However, concurrent improvements in dyspnea, skin changes, and several patient-reported outcomes have boosted confidence that the physiologic responses to CYC are real and beneficial, although these improvements still have to be weighed against the potential side effects of a 1-year course of treatment with oral CYC, including possible long-term consequences, such as bladder, hematologic, and skin cancers. Consistent with the improvements in FVC and TLC, the present study reveals a significantly more favorable course of HRCT scan-assessed FIB (*ie*, improvement or stability as opposed to worsening) in the CYC treatment group than in the placebo group ( $p = 0.014$  [ $\chi^2$  test]) but no difference in the course of other radiographic features (GGO or HC). Moreover, in separate analyses, treatment group assignment was a significant predictor of the change in FIB over the 12-month treatment period. In addition,

change in FIB over time was significantly correlated with physiologic measures of improvement (FVC and TLC) as well as with improvement in dyspnea and skin thickness.

These HRCT scan findings provided independent validation of the treatment effect produced by CYC and suggested that treatment-related differences in the 1-year course of measures of restriction (FVC and TLC) reflect structural changes in lung tissue. The HRCT scan findings, therefore, provide support for the interpretation that CYC slowed the progression of FIB compared with placebo over the 1-year treatment period.

At the same time, it is interesting that there was no observed treatment effect with respect to changes over time in GGO, including improvement, despite the fact that CYC is an immunosuppressive agent that might be expected to suppress inflammation. Thus, the absence of any beneficial effect of CYC on GGO argues against equating GGO with inflamma-

**Table 2—Number of Patients in Each Treatment Group With Worse (or Not Worse) FIB, GGO, or HC in at Least One Lung Zone on the 12-Month HRCT Scan Compared to the Baseline Scan**

Treatment Groups	Fibrosis		GGO		HC	
	Worse	Not Worse	Worse	Not Worse*	Worse	Not Worse
CYC	14	35	13	36	3	46
Placebo	26	23	16	33	3	46

$p = 0.014$  for FIB;  $p = 0.507$  for GGO;  $p = 1.000$  for HC ( $\chi^2$  test).

\*Of the 36 patients in the CYC group and 33 in the placebo group with not worse GGO, improvement in GGO was noted in 8 CYC patients and 10 placebo patients.

**Table 3—Impact of Covariates on FIB Outcome (Logistic Model)**

Covariates	Estimate	SE	p Value	Odds Ratio (95% CI)
Intercept	−4.05	1.67	0.015	
Group (CYC/placebo)	1.18	0.47	0.012	3.26 (1.30–8.17)
Max FIB score, baseline	−0.22	0.22	0.331	0.81 (0.52–1.25)
FVC % predicted, baseline	0.06	0.02	0.004	1.06 (1.02–1.10)

tion. These results are contrary to previously reported findings.<sup>21,22</sup> GGO is classically defined<sup>29</sup> as alveolar opacification through which vessel architecture is still visualized. In the setting of interstitial lung disease, GGO may reflect either alveolar filling by inflammatory cells or subresolution FIB. In patients with SSc-ILD in which nonspecific interstitial pneumonitis is the typical pathology, GGO has been correlated with alveolitis on BAL fluid findings in some studies,<sup>30</sup> and more extensive GGO seen on HRCT scans has been associated significantly with a lower DLCO, which is analogous to the association between BAL fluid evidence of alveolitis and a reduced DLCO.<sup>31</sup> In still other studies,<sup>26</sup> little correlation between BAL evidence for alveolitis and CT scan evidence for GGO has been shown. This uncertainty regarding the structural basis for GGO makes the absence of a change in GGO between the two treatment groups difficult to interpret. Interpretation of the lack of change in GGO is further hindered by the greater interreader variability in grading this HRCT scan finding ( $\kappa = 0.36$ ) compared with FIB ( $\kappa = 0.51$ ). The poorer interreader

agreement in scoring changes in GGO, especially in HC ( $\kappa = 0.16$ ), could be due to the more limited abnormalities in GGO and HC seen on the baseline HRCT scan (scores of 0.65 to 0.67 vs 0.35 to 0.41, respectively, on a scale of 0 to 3) compared with FIB (scores of 1.96 to 2.0, on a scale of 0 to 4) [Table 1].

There were several major limitations for the present study. First, the ancillary study needed to obtain an HRCT scan at the end of the 12-month treatment period was not available until well after the start of the SLS. Consequently, approximately one-third of the SLS patients did not have a follow-up HRCT scan, reducing the power to show a treatment effect. Second, in accordance with the intention-to-treat principle, some of the patients who underwent a 12-month HRCT scan had prematurely withdrawn from the double-blind treatment phase of the study. The failure of some patients in the CYC arm to receive a complete course of therapy could confound the results. However, these confounding effects would bias the results toward the null and could not have been responsible for the positive treatment effect on HRCT scan-assessed changes in FIB.

**Table 4—The 12-Month Change in Score Between FIB and FVC, TLC, DLCO, Skin Score, HAQ, Cough, and Dyspnea in 98 Patients**

Parameters	FIB		Kendall $\tau$ Rank Correlation	McNemar Test	p Value
	Worse	Not Worse			
Average change score	−4.35	−0.82	0.21		0.041
FVC <sub>12-0</sub> , % predicted	−4.87	−0.47	0.22		0.035
TLC <sub>12-0</sub> , % predicted	−6.38	−2.91	0.199		0.053
DLCO <sub>12-0</sub> , % predicted					
Number of patients					
Skin					
Worse	15	25		0.206	0.042
Not worse	11	47			
HAQ					
Worse	18	22		0.071	0.489
Not worse	22	36			
Cough					
Worse	9	31		0.065	0.52
Not worse	10	48			
Dyspnea					
Worse	21	19		0.355	< 0.001
Not worse	10	48			

FVC<sub>12-0</sub> = change in FVC between 0 (baseline) and 12 months; TLC<sub>12-0</sub> = change in TLC between 0 (baseline) and 12 months; DLCO<sub>12-0</sub> = change in DLCO between 0 (baseline) and 12 months.

Third, the scans were performed using axial incremental images every 10 mm (as was the convention at the time), resulting in image data sets at the two time points that were not always at exactly the same anatomic lung levels, potentially affecting the ability to infer changes due to sampling differences. Fourth, the side-by-side comparisons of the baseline and 12-month HRCT scans were scored visually, thus resulting in subject-to-interreader differences in interpretation. This interreader variability for changes over time in FIB was reflected by a  $\kappa$  statistic of 0.51 for agreement among trained readers. This only moderate degree of concordance could compromise the clinical utility of HRCT scans as an objective marker of treatment-related changes in FIB. Improvements in imaging technique (including full-lung volumetric acquisition) and technological advances in the development of software for computerized quantitative assessments of the distribution and extent of pulmonary FIB should help to resolve some of these difficulties.<sup>32</sup>

The assessment of FIB on the HRCT scans of patients with SSc-ILD has been shown to be a predictor of the progression of FIB in the absence of active treatment and of a favorable response to treatment with CYC compared with placebo.<sup>5</sup> It also has been shown to be a predictor of mortality in an observational study.<sup>20</sup> On the other hand, HRCT scanning has infrequently been used in a prospective, systematic manner as an outcome parameter for assessing therapeutic responses in patients with SSc-ILD.<sup>21,22</sup> The results of the present study suggest that the incorporation of follow-up HRCT scans in controlled clinical trials of interventions in patients with SSc-ILD (or potentially in other interstitial pneumonitides) might prove a usual adjunct to conventional physiologic and patient-centered end points, although cost and radiation exposure need to be considered.

#### ACKNOWLEDGMENTS

**Author contributions:** Drs. Tashkin, Goldin, Elashoff, Roth, Clements, Furst, and Khanna all contributed to the design, oversight, and conduct of the study and in the writing and editing of the manuscript. Drs. Kim and Yan, and Ms. Vasunilashorn all contributed to the processing and statistical analysis of the imaging data and the correlation of the imaging results with the physiologic and patient-centered outcome variables. Dr. Li contributed substantially to the statistical design and analysis of the results. Drs. Lynch and Strollo performed all of the interpretive comparisons of the baseline and 12-month HRCT scans and contributed to the review and editing of the manuscript.

**Financial/nonfinancial disclosures:** The authors have reported to the ACCP that no significant conflicts of interest exist with any companies/organizations whose products or services may be discussed in this article.

#### REFERENCES

- 1 Steen VD, Conte C, Owens GR, et al. Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum* 1994; 37:1283–1289
- 2 Steen VD, Medsger TA Jr. Severe organ involvement in systemic sclerosis with diffuse scleroderma. *Arthritis Rheum* 2000; 43:2437–2444
- 3 Medsger TA Jr. Classification, prognosis. In: Clements PJ, Furst DE, eds. *Systemic sclerosis*. Philadelphia, PA: Lippincott, Williams and Wilkins, 2004; 17–28
- 4 Latsi PI, Wells AU. Evaluation and management of alveolitis and interstitial lung disease in scleroderma. *Curr Opin Rheumatol* 2003; 15:748–755
- 5 Tashkin DP, Elashoff R, Clements PJ, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 2006; 354:2655–2666
- 6 Hoyles RK, Ellis RW, Wellsbury J, et al. Fibrosing Alveolitis in Scleroderma Trial (FAST): a multi-centre, prospective, randomized, double-blind, placebo-controlled trial of treatment with corticosteroids and intravenous cyclophosphamide followed by oral azathioprine. *Arthritis Rheum* 2006; 54:3962–3970
- 7 Berezne A, Ranque B, Valeyre D, et al. Therapeutic strategy combining intravenous cyclophosphamide followed by oral azathioprine to treat worsening interstitial lung disease associated with systemic sclerosis: a retrospective multicenter open-label study. *J Rheumatol* 2008; 35:1064–1072
- 8 Silver RM, Warrick JH, Kinsella MB, et al. Cyclophosphamide and low-dose prednisone therapy in patients with systemic sclerosis (scleroderma) with interstitial lung disease. *J Rheumatol* 1993; 20:838–844
- 9 Akesson A, Scheja A, Lundin A, et al. Improved pulmonary function in systemic sclerosis after treatment with cyclophosphamide. *Arthritis Rheum* 1994; 37:729–735
- 10 White B, Moore WC, Wigley FM, et al. Cyclophosphamide is associated with pulmonary function and survival benefit in patients with scleroderma and alveolitis. *Ann Intern Med* 2000; 132:947–954
- 11 Tashkin DP, Elashoff R, Clements PJ, 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; 176:1026–1034
- 12 Martinez FJ, McCune WJ. Cyclophosphamide for scleroderma lung disease [editorial]. *N Engl J Med* 2006; 354:2707–2709
- 13 Fujita J, Yoshinouchi T, Ohtsuki Y, et al. Non-specific interstitial pneumonia as pulmonary involvement of systemic sclerosis. *Ann Rheum Dis* 2001; 60:281–283
- 14 Kim DS, Yoo B, Lee JS, et al. The major histopathologic pattern of pulmonary fibrosis in scleroderma is nonspecific interstitial pneumonia. *Sarcoidosis Vasc Diffuse Lung Dis* 2002; 19:121–127
- 15 Bours D, Wells AU, Nicholson AG, et al. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am J Respir Crit Care Med* 2002; 165:1581–1586
- 16 Desai SR, Veeraraghavan S, Hansell DM, et al. CT features of lung disease in patients with systemic sclerosis: comparison with idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia. *Radiology* 2004; 232:560–567
- 17 Goldin JG, Lynch DA, Strollo DC, et al. High resolution CT findings in scleroderma-related lung diseases: findings from Scleroderma Lung Study. *Chest* 2008; 134:358–367
- 18 Wells AU, Rubens MB, du Bois RM, et al. Serial CT in fibrosing alveolitis: prognostic significance of the initial pattern. *AJR Am J Roentgenol* 1993; 161:1159–1165

- 19 Wells AU. High-resolution computed tomography and scleroderma lung disease. *Rheumatology (Oxford)* 2008; 47(suppl), v59–v61
- 20 Goh NSL, Desai SR, Veeraraghavan S, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med* 2008; 177:1248–1254
- 21 Giacomelli R, Valentini G, Salsano F, et al. Cyclophosphamide pulse regimen in the treatment of alveolitis in systemic sclerosis. *J Rheumatol* 2002; 29:731–736
- 22 Griffiths B, Miles S, Moss H, et al. Systemic sclerosis and interstitial lung disease: a pilot study using pulse intravenous methylprednisolone and cyclophosphamide to assess the effect on high resolution computed tomography scan and lung function. *J Rheumatol* 2002; 29:2371–2378
- 23 Pakas I, Ioannidis JPA, Malagari K, et al. Cyclophosphamide with low or high dose prednisolone for systemic sclerosis lung disease. *J Rheumatol* 2002; 29:298–304
- 24 Davas EM, Peppas C, Maragou M, et al. Intravenous cyclophosphamide pulse therapy for the treatment of lung disease associated with scleroderma. *Clin Rheumatol* 1999; 18:455–461
- 25 Mahler DA, Weinberg DH, Wells CK, et al. The measurement of dyspnea: contents, interobserver agreement, and physiologic correlates of two new clinical indexes. *Chest* 1984; 85:751–758
- 26 Strange C, Bolster MB, Roth MD, et al. Bronchoalveolar lavage and response to cyclophosphamide in scleroderma interstitial lung disease. *Am J Respir Crit Care Med* 2008; 177:91–98
- 27 BAL Cooperative Group Steering Committee. Bronchoalveolar lavage constituents in healthy individuals, idiopathic pulmonary fibrosis and selected comparison groups. *Am Rev Respir Dis* 1990; 141:S169–S202
- 28 Altman DG, ed. *Practical statistics for medical research*. London, UK: Chapman and Hall, 1991; 404
- 29 Webb WR, Muller NL, Naidich DP. *High-resolution CT of the lung*. Philadelphia, PA: Lippincott Williams & Wilkins, 2000
- 30 Remy-Jardin M, Remy J, Wallaert B, et al. Pulmonary involvement in progressive systemic sclerosis: sequential evaluation with CT, pulmonary function tests, and bronchoalveolar lavage. *Radiology* 1993; 188:499–506
- 31 Silver RM, Miller KS, Kinsella MB, et al. Evaluation and management of scleroderma lung disease using bronchoalveolar lavage. *Am J Med* 1990; 88:470–476
- 32 Kim HJ, Li G, Gjertson D, et al. Classification of parenchymal abnormality in scleroderma lung using a novel approach to denoise images collected via a multicenter study. *Acad Radiol* 2008; 15:1004–1016