

UCLA

UCLA Previously Published Works

Title

Early Radiographic Progression of Scleroderma: Lung Disease Predicts Long-term Mortality.

Permalink

<https://escholarship.org/uc/item/6sv3r48s>

Journal

Chest, 161(5)

ISSN

0012-3692

Authors

Volkman, Elizabeth R

Tashkin, Donald P

Roth, Michael D

et al.

Publication Date

2022-05-01

DOI

10.1016/j.chest.2021.11.033

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Early Radiographic Progression of Scleroderma Lung Disease Predicts Long-term Mortality



Elizabeth R. Volkman, MD; Donald P. Tashkin, MD; Michael D. Roth, MD; Jonathan Goldin, MD, PhD; and Grace H. J. Kim, PhD



BACKGROUND: Radiographic end points commonly are included in therapeutic trials for systemic sclerosis (SSc)-interstitial lung disease (ILD); however, the relationship between these outcomes and long-term mortality is unclear.

RESEARCH QUESTION: Do short-term changes in radiographic measures of ILD predict long-term survival in patients with SSc?

STUDY DESIGN AND METHODS: The Scleroderma Lung Study (SLS) I and II evaluated the safety and efficacy of cyclophosphamide (in SLS I and II) and mycophenolate mofetil (in SLS II) for the treatment of SSc-ILD. Changes in the extent of ILD over time were assessed on high-resolution CT scans of the chest by quantitative image analysis, an approach that applies a computer-based algorithm to assess changes in the radiographic extent of ILD objectively. Participants subsequently were followed for up to 12 years (SLS I) and 8 years (SLS II). Cox proportional hazards models determined whether the change in the quantitative radiographic extent of ILD predicted survival, adjusting for other known predictors of survival.

RESULTS: Among SLS I and II participants, 82 and 90 had follow-up imaging scans, respectively, and were included in the analysis. Participants in both trials who showed an increase in the total quantitative radiographic extent of ILD scores of $\geq 2\%$ at 12 months (SLS I) or 24 months (SLS II) experienced significantly worse long-term survival than those with change scores of $< 2\%$ ($P \leq .01$, log-rank test). In the multivariate Cox models, radiographic progression remained associated with worse long-term survival in SLS I ($P = .089$) and SLS II ($P = .014$).

INTERPRETATION: Data from two independent clinical trial cohorts with extensive long-term follow-up demonstrated that radiographic progression of ILD over 12 to 24 months, in both treatment and placebo arms, can predict increased risk for long-term mortality in patients with SSc. These findings suggest that radiographic end points may serve as surrogates for mortality in SSc-ILD.

CHEST 2022; 161(5):1310-1319

KEY WORDS: biomarkers; interstitial lung disease; mortality; systemic sclerosis

ABBREVIATIONS: DLCO = diffusing capacity for carbon monoxide; HR = hazard ratio; HRCT = high-resolution CT; ILD = interstitial lung disease; MMF = mycophenolate mofetil; QILD = quantitative interstitial lung disease; RCT = randomized controlled trial; SLS = Scleroderma Lung Study; SSc = systemic sclerosis

AFFILIATIONS: From the Department of Medicine (E. R. Volkman, D. P. Tashkin, and M. D. Roth), and the Department of Radiology (J. Goldin and Grace H. J. Kim), David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA.

This article was presented at the virtual European League Against Rheumatism Annual Congress, June 4, 2021.

FUNDING/SUPPORT: This study was funded by the National Heart, Lung, and Blood Institute, National Institutes of Health [Grants K23

HL150237 (E. R. V.), R01 HL089758 (D. P. T.), R01 HL089901 (see Acknowledgments), U01 HL 60587 (D. P. T.), and U01 HL 60606 (see Acknowledgments). Bristol-Myers Squibb supplied cyclophosphamide for use in Scleroderma Lung Study I and Hoffmann-LaRoche supplied mycophenolate mofetil for use in Scleroderma Lung Study II.

CORRESPONDENCE TO: Elizabeth R. Volkman, MD; email: evolkmann@mednet.ucla.edu

Copyright © 2021 The Author(s). Published by Elsevier Inc under license from the American College of Chest Physicians. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

DOI: <https://doi.org/10.1016/j.chest.2021.11.033>

Take-home Points

Study Question: Do short-term changes in radiographic measures of interstitial lung disease (ILD) predict long-term survival in patients with systemic sclerosis (SSc)?

Results: In patients with SSc-ILD enrolled in two large randomized controlled trials, an increase in the quantitative radiographic extent of ILD of $\geq 2\%$ over 1 to 2 years was associated with worse long-term survival.

Interpretation: Short-term changes in the radiographic extent of ILD in patients with SSc receiving treatment or placebo may serve as a proxy for long-term mortality. Future SSc-ILD clinical trials should consider including radiographic end points to assess treatment response.

Interstitial lung disease (ILD) is the leading cause of death in systemic sclerosis (SSc)^{1,2} and a major focus for therapeutic drug discovery in this field. Despite the burgeoning SSc-ILD therapeutic pipeline, no universally accepted end points exist that define an optimal treatment response. End points that directly measure how a patient with SSc-ILD feels and functions are lacking, and although mortality is an unequivocal end point, most SSc clinical trials are powered inadequately and are designed to use mortality as an end point.

The most commonly used surrogate end point in SSc-ILD clinical trials is FVC.³⁻⁵ Although studies have demonstrated that the course of FVC is related to survival in SSc-ILD,^{6,7} the reliability of this parameter as a direct measure of lung disease may be limited in SSc, where extrapulmonary manifestations can affect its measurement and interpretation substantially (eg, cutaneous sclerosis involving the chest wall, respiratory

muscle weakness).⁸ Variations in pulmonary function test protocols, along with patient and technician effort, can influence the reproducibility of the FVC in clinical practice and research further. In addition, studies have demonstrated that the FVC correlates poorly with the actual radiographic extent of ILD in SSc.⁹ Thus, an unmet need exists for the discovery of novel SSc-ILD study end points that are reliable and reproducible and can predict mortality outcomes consistently in patients with SSc.

Objectively quantifying the radiographic extent of ILD may represent a more direct assessment of parenchymal lung disease burden and potentially may obviate the impact of extrapulmonary disease, patient effort, and technical factors that can alter the measurement of FVC, as well as the intrareader and interreader variability in visual radiographic change assessment.¹⁰ Studies have demonstrated that increased radiographic extent of ILD at baseline predicts responsiveness to immunosuppressive therapy.^{11,12} Additional studies have found that the radiographic ILD end point is sensitive to change in patients with SSc-ILD undergoing treatment.¹³⁻¹⁶ However, no studies have evaluated whether a change in the radiographic extent of ILD predicts mortality in this population. To address this issue, the present study examined whether radiographic progression of ILD in patients with SSc receiving treatment predicts long-term mortality. Using data from the Scleroderma Lung Study (SLS) I³ and II,⁴ we hypothesized that those patients who experienced increased radiographic progression of ILD over the course of these trials would show worse long-term survival. The findings of this research may broaden our understanding of efficacy assessment in SSc-ILD research and ultimately may shape the design of future clinical trials for this often-fatal disease.

Study Design and Methods

Study Participants

All participants enrolled in SLS I³ ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01762449) Identifiers: NCT01762449 and NCT00004563) and SLS II⁴ ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT00883129) Identifier: NCT00883129) who underwent a follow-up, standardized, high-resolution CT (HRCT) scan of the chest were eligible to participate in this study. SLS I and II were randomized controlled trials (RCTs) that included an ethnically diverse population of both male and female patients with SSc-ILD followed up at multiple centers across the United States. Eligibility criteria for these trials were similar.^{3,4} The institutional review board of each site approved the primary studies and long-term follow-up. Informed consent was obtained from all participants.

SLS I and II Study Design

In SLS I, 158 participants were randomized to receive oral cyclophosphamide or placebo for 12 months and were followed up

for an additional 12 months off of therapy.³ In SLS II, 142 patients were randomized to receive mycophenolate mofetil (MMF) for 24 months or oral cyclophosphamide for 12 months followed by an additional 12 months of placebo.⁴

SLS I and II Assessment Measurements

The FVC (primary SLS I and II end point) was measured every 3 months during the 24-month study periods.^{3,4} HRCT thoracic imaging was obtained at baseline and at the conclusion of active treatment in both trials (ie, at 12 months in SLS I and at 24 months in SLS II). A Computer-Aided Design scoring system¹³⁻¹⁵ was used to calculate the quantitative ILD (QILD) score for the entire lung at baseline and follow-up (e-Fig 1 shows further details on the HRCT imaging protocol and scoring). The QILD score included the sum of all abnormally classified scores, including fibrosis (eg, reticular opacity with architectural distortion), ground-glass opacity (eg,

increased parenchymal attenuation), and honeycombing (eg, clustered air-filled cysts with dense walls).

The QILD score threshold to define radiographic ILD progression (QILD score of $\geq 2\%$) was determined first by performing a statistical analysis of the cohorts using a quantitative imaging biomarker algorithm proposed in our prior publication.¹⁷ This analysis examines both technical reproducibility and clinical reproducibility.¹⁷ Technical reproducibility was reported as 0.60% in the limited agreement (e-Fig 2). Subsequently, the factor of 2.77 ($= 1.96 \times \sqrt{2}$) was multiplied by the variability of the two HRCT scans, resulting in a threshold of 1.66%. This number then was rounded up to 2%.

Next, we considered clinically meaningful correlates of a $\geq 2\%$ increase in QILD, such as change in FVC percent predicted, patient-reported outcomes, and mortality. In the SLS I and II cohorts, an increase of QILD score of $\geq 2\%$ was associated with a clinically meaningful decline in FVC percent predicted.¹⁸ Specifically, most patients meeting the minimal clinically important difference estimates for FVC percent predicted worsening also experienced an increase in QILD score of $> 2\%$. The minimal clinically important difference criteria for FVC worsening (-3.0% to -3.3%) were derived based on an analysis of two valid patient-reported outcomes.¹⁸ As a final step, we tested the relationship between additional thresholds of QILD score worsening (eg, $\geq 1\%$, $\geq 3\%$, $\geq 4\%$, and $\geq 5\%$) and survival, and the $\geq 2\%$ increase in QILD score demonstrated the strongest association with survival in both of the SLS I and II cohorts.

Long-term Mortality

During the SLS I and II trials, mortality data were collected and causes of death were adjudicated by data safety and monitoring boards. After the 24-month trials, patients or their designated surrogates were contacted annually to assess morbidity and mortality outcomes. If the patient or previously designated contact person could not be reached, investigators contacted site investigators and searched publicly available death registries (eg, National Death Index and Social Security Death Index), as well as online obituaries. Survival status was ascertained for up to 12 and 8 years after the commencement of SLS I and II, respectively.

Results

Participant Characteristics

Among all of the SLS I ($n = 158$) and II ($N = 142$) participants, 82 and 90, respectively, underwent follow-up HRCT scans of the chest and were included in the present analysis. In SLS I, the follow-up HRCT scan was part of a substudy that was delayed in onset, limiting the ability to enroll patients who already had completed the 12-month follow-up visit. In SLS II, the primary reason not all patients had undergone an HRCT scan of the chest at 24 months was participant drop out, death, or both.

The baseline disease features and demographic characteristics of the SLS I and II participants who underwent follow-up HRCT scans were similar and closely reflected those of the overall study population (e-Table 1). Most patients were women with relatively

Statistical Analysis

All tests were 2-sided and were performed using SAS version 9.4 software (SAS Institute).

Baseline Characteristics: Summary statistics were generated for baseline characteristics from the two cohorts. Group comparisons were performed using two-sample t tests and χ^2 tests.

Primary Outcome: Survival: The primary outcome was all-cause mortality. The Kaplan-Meier estimate was used to generate survival curves, and the log-rank test was used to compare survival between participants who experienced progression of ILD ($\geq 2\%$ increase in QILD score) vs those who experienced stability or improvement of ILD ($< 2\%$ increase in QILD score). Cox proportional hazard models subsequently were developed. Using a comprehensive variable selection process described in our previous publication,⁷ the following variables were found to be associated significantly with mortality in univariate analyses in both the SLS I and II cohorts: age, modified Rodnan skin score, and baseline FVC percent predicted. Therefore, the aforementioned variables were included in the final Cox models. Treatment arm assignment was not associated significantly with mortality in SLS I or II⁷; however, given the potential interaction between treatment arm assignment and change in QILD, we created exploratory Cox models that also included treatment arm assignment. We also created exploratory Cox models that included variables associated with mortality based on expert knowledge (eg, sex, race, diffuse SSc subtype, and diffusing capacity for carbon monoxide [DLCO] % predicted).

Secondary Outcome: Course of FVC: Mixed-effects models were created to compare the course of the FVC percent predicted (measured every 3 months over 24 months) between patients who experienced progression of ILD ($\geq 2\%$ increase in QILD score) vs those who experienced stability or improvement of ILD ($< 2\%$ increase in QILD score). Covariates included the baseline FVC percent predicted, treatment arm assignment, and the change in QILD score.

early SSc (approximately 3 years from the onset of the first non-Raynaud symptom attributable to SSc) and a moderate degree of restriction on pulmonary function testing results.

Radiographic ILD Progression in SLS I and II

In SLS I, 34 participants (41%) experienced an increase in QILD score of $\geq 2\%$ for the entire lung at 12 months. Among these participants, 13 were randomized to cyclophosphamide and 21 were randomized to placebo. This difference between treatment arms is consistent with the statistically significant impact of cyclophosphamide on the change in QILD score at 12 months reported in our prior studies.¹² In SLS II, 28 participants (31%) experienced an increase in QILD score of $\geq 2\%$ for the entire lung at 24 months. Among these participants, 15 were randomized to cyclophosphamide and 13 were randomized to MMF. In contrast to the outcome in SLS I, in which half of the

participants received placebo, no significant difference in treatment effect was observed between the two active treatment arms (MMF and cyclophosphamide) in SLS II.¹⁴ No significant differences were found in the baseline characteristics of participants who experienced an increase in QILD score of $\geq 2\%$ and those who did not in either SLS I or II, with the exception of a lower baseline DLCO % predicted in SLS I participants who experienced an increase in QILD score of $\geq 2\%$ (Table 1).

SLS I participants who experienced an increase in QILD score of $\geq 2\%$ over 12 months also were more likely to experience a decline in FVC percent predicted over the course of the trial, whereas SLS I participants who experienced a change in QILD score of $< 2\%$ were more likely to experience stability or improvement in FVC percent predicted (Fig 1A). Similarly, in SLS II, those participants who experienced an increase in QILD score of $\geq 2\%$ over 24 months also were more likely to experience a decline in FVC percent predicted over the course of the trial, whereas those participants who experienced a change in QILD score of $< 2\%$ were more

likely to experience stability or improvement in FVC percent predicted (Fig 1B).

Long-term Survival in SLS I and II

In SLS I, 66 patients (42%) of the entire cohort had died within 12 years after the first patient was randomized (cyclophosphamide group, $n = 38$; placebo group, $n = 28$). Most deaths (65%) were the result of underlying SSc, and among the deaths resulting from SSc, 67% were the result of respiratory failure. In the SLS I participants with follow-up HRCT scans at 12 months, 28 deaths (34%) occurred during the long-term follow-up period (cyclophosphamide group, $n = 15$; placebo group, $n = 13$).

In SLS II, 30 patients (21%) of the entire cohort had died within 8 years after the first patient was randomized (cyclophosphamide group, $n = 16$; MMF group, $n = 14$). Most deaths (58%) were the result of underlying SSc, and among the deaths resulting from SSc, 50% were the result of respiratory failure. In the SLS II participants with follow-up HRCT scans at 24 months, 15 deaths

TABLE 1] Baseline Patient Characteristics of SLS I and II Participants Who Showed a Change in QILD Score of $\geq 2\%$ vs $< 2\%$ at 12 and 24 Months, Respectively

| Variable | SLS I | | | SLS II | | |
|---|---|--|----------------|---|--|----------------|
| | Change in QILD Score of $\geq 2\%$ ($n = 34$) | Change in QILD Score of $< 2\%$ ($n = 48$) | <i>P</i> Value | Change in QILD Score of $\geq 2\%$ ($n = 25$) | Change in QILD Score of $< 2\%$ ($n = 65$) | <i>P</i> Value |
| Age, y | 48.1 \pm 11.2 | 45.6 \pm 11.7 | .27 | 51.63 (9.74) | 51.36 (8.98) | .81 |
| Female sex | 21 (61.76) | 39(81.25) | .050 | 15 (60.00) | 51 (78.46) | .053 |
| SSc duration, mean (SD), y ^a | 2.18 (2.42) | 3.10 (3.47) | .031 | 2.50 (2.83) | 1.67 (3.00) | .59 |
| Diffuse cutaneous disease | 21 (61.76) | 27 (56.25) | .62 | 15 (60.00) | 38 (58.46) | .96 |
| Race ^b | | | | | | |
| White | 21 (61.76) | 34 (70.83) | .62 | 15 (60.00) | 44 (67.69) | .32 |
| Black | 6 (17.65) | 5 (10.42) | ... | 8 (32.00) | 14 (21.54) | ... |
| Asian | 2 (5.88) | 1 (2.08) | ... | 1 (4.00) | 5 (7.69) | ... |
| Other | 5 (14.71) | 7 (14.58) | ... | 1 (4.00) | 2 (3.08) | ... |
| Unknown | 0 (0.00) | 1 (2.08) | ... | 0 (0.00) | 0 (0.00) | ... |
| mRSS | 16.53 \pm 11.63 | 14.15 \pm 10.78 | .33 | 12.93 \pm 7.92 | 14.20 \pm 10.05 | .82 |
| History of prior smoking | 11 (33.33) | 17 (35.42) | .85 | 6 (25.00) | 19 (29.23) | .69 |
| FVC percent predicted | 66.67 \pm 10.85 | 71.16 \pm 11.68 | .21 | 67.69 \pm 7.78 | 65.91 \pm 9.00 | .93 |
| DLCO percent predicted ^b | 43.20 \pm 12.70 | 50.46 \pm 14.54 | .022 | 55.94 \pm 13.91 | 55.42 \pm 12.74 | .87 |
| QLF % whole lung | 8.23 \pm 5.98 | 10.14 \pm 10.78 | .35 | 7.77 \pm 5.72 | 7.80 \pm 7.01 | .97 |
| QILD score % whole lung | 31.57 \pm 11.99 | 35.62 \pm 17.04 | .24 | 26.61 \pm 8.84 | 27.05 \pm 13.91 | .86 |

Data are presented as No. (%), mean \pm SD, or median (interquartile range), unless otherwise indicated. DLCO = diffusing capacity for carbon monoxide; mRSS = modified Rodnan skin score; QILD = quantitative interstitial lung disease; QLF = quantitative lung fibrosis; SLS = Scleroderma Lung Study; SSc = systemic sclerosis.

^aFor disease duration, $n = 81$ and $n = 84$ in SLS I and II, respectively.

^bFor race, $n = 81$ in SLS I.

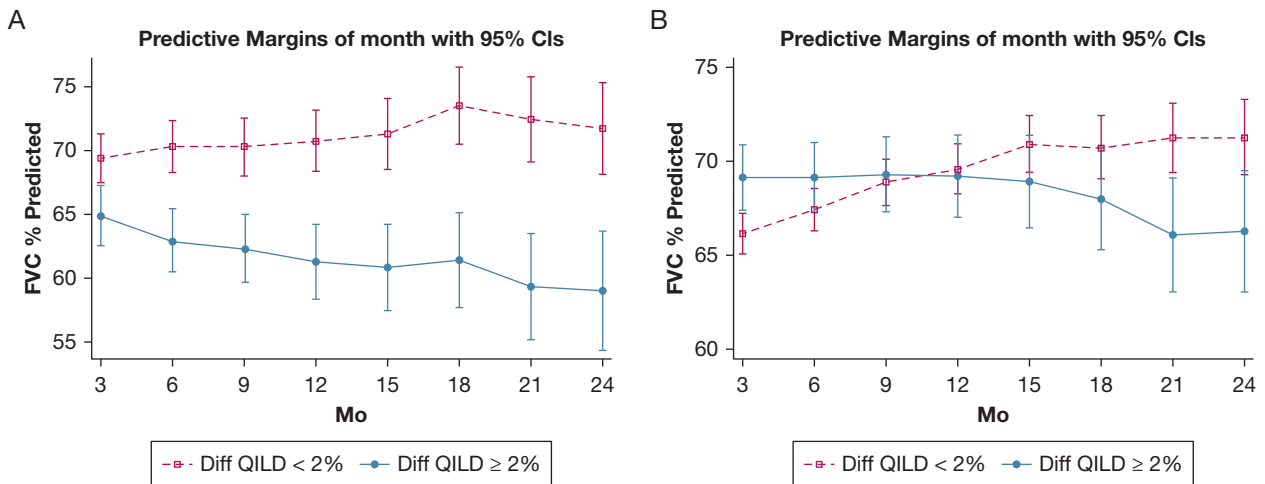


Figure 1 – A, B, Graphs showing the mixed-effects model of the course of FVC percent predicted over 24 months for participants with $\geq 2\%$ (red line) vs $< 2\%$ (blue line) increase in QILD score for SLS I (A) and SLS II (B). Covariates for both models included the baseline FVC percent predicted and treatment arm. QILD = quantitative interstitial lung disease; SLS = Scleroderma Lung Study.

(17%) occurred during the long-term follow-up period (cyclophosphamide group, n = 6; MMF group, n = 9).

Radiographic Progression of ILD Predicts Survival

During the 12-year long-term follow-up period, SLS I participants who experienced an increase in QILD score of $\geq 2\%$ for the entire lung at 1 year showed a significantly increased risk of death ($P = .01$, log-rank test) (Fig 2). During the 8-year long-term follow-up period, SLS II participants who experienced an increase

in QILD score of $\geq 2\%$ for the entire lung at 2 years showed a significantly increased risk of death ($P = 0.019$, log-rank test) (Fig 3).

After adjusting for age, modified Rodnan skin score, and baseline FVC percent predicted, a suggestive association was found between increase in QILD score of $\geq 2\%$ and mortality in SLS I participants (hazard ratio [HR], 1.98; $P = .089$) (Table 2). In addition, after adjusting for the aforementioned variables, an increase in QILD score of $\geq 2\%$ was associated with a significantly increased risk

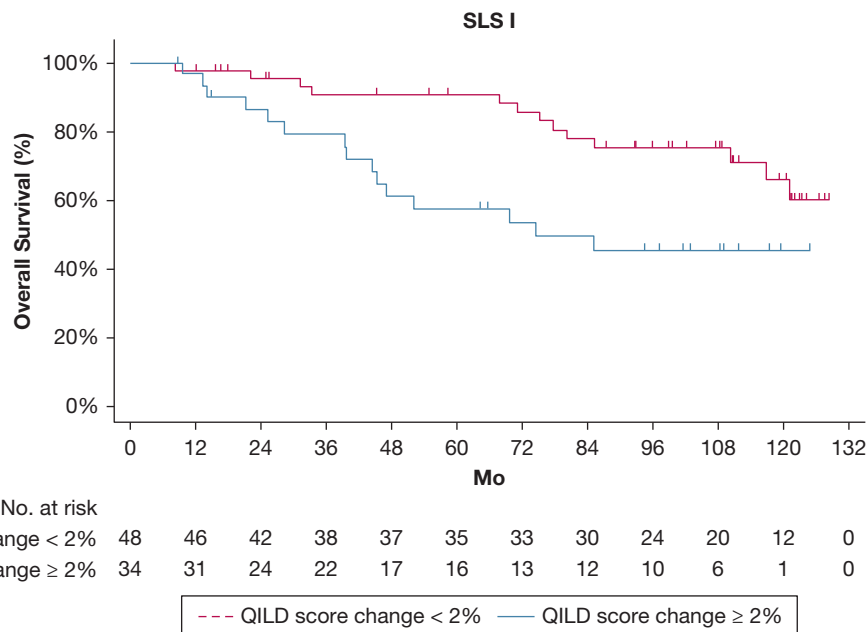


Figure 2 – Kaplan-Meier plot showing overall survival for participants with $\geq 2\%$ vs $< 2\%$ increase in QILD score for the entire lung from baseline to 12 months in SLS I. QILD = quantitative interstitial lung disease; SLS = Scleroderma Lung Study.

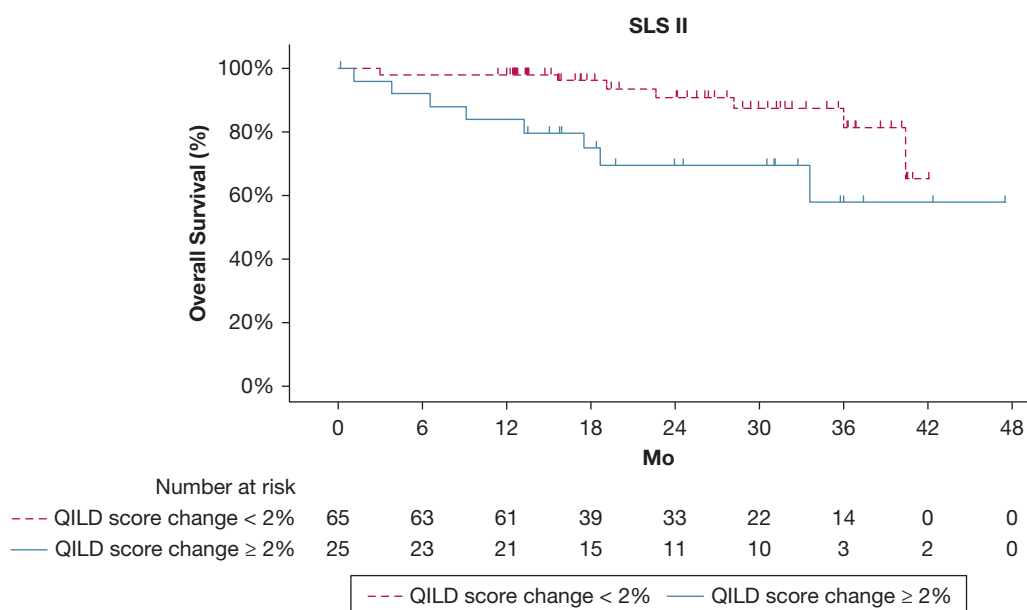


Figure 3 – Kaplan-Meier plot showing overall survival for participants with $\geq 2\%$ vs $< 2\%$ increase in QILD score for the entire lung from baseline to 24 months in SLS II. QILD = quantitative interstitial lung disease; SLS = Scleroderma Lung Study.

of mortality in SLS II participants (HR, 3.86; $P = .014$) (Table 3). After adding treatment arm assignment (e-Tables 2, 3), as well as the variables of sex, race, diffuse SSC subtype, and DLCO percent predicted (e-Tables 4, 5) to the aforementioned Cox models, an increase in QILD score of $\geq 2\%$ remained associated with mortality in both SLS I (suggestive association) and SLS II (significant association).

In an exploratory analysis, physiologic progression of ILD was substituted for the baseline FVC percent predicted as a covariate in the Cox model. Physiologic progression was defined according to the Outcome Measures in Rheumatology (OMERACT) criteria as FVC percent predicted decline of $\geq 10\%$ or FVC percent predicted decline of between 5% and 9% and DLCO percent predicted decline of $\geq 15\%$.^{6,19} In

SLS I, 15 patients met the OMERACT criteria for physiologic ILD progression at 12 months. After adjusting for age, modified Rodnan skin score, and physiologic progression of ILD at 12 months, an increase in QILD score of $\geq 2\%$ remained associated significantly with an increased risk of mortality (HR, 2.28; $P = .037$) (e-Table 6). In another exploratory analysis, we included the change in FVC percent predicted at 12 months (measured continuously) as a covariate, and this covariate was not associated significantly with mortality ($P = .97$) (e-Table 7).

In SLS II, too few patients ($n = 3$) met the OMERACT criteria for physiologic ILD progression at 24 months; therefore, a Cox model that included this covariate could

TABLE 2] Cox Proportional Hazards Model for Survival in SLS I (N = 82)

| Variable | Hazard Ratio | 95% CI | P Value | C Index |
|---------------------------------|--------------|-----------|---------|---------|
| Univariate analyses | | | | |
| QILD score change of $\geq 2\%$ | 2.61 | 1.23-5.57 | .013 | 0.637 |
| Continuous change in QILD score | 1.02 | 0.99-1.06 | .16 | 0.643 |
| Multivariate analyses | | | | |
| Age | 1.04 | 1.01-1.08 | .019 | 0.728 |
| mRSS | 1.04 | 1.00-1.07 | .028 | |
| Baseline FVC percent predicted | 0.97 | 0.93-1.01 | .098 | |
| QILD score change of $\geq 2\%$ | 1.98 | 0.90-4.40 | .089 | |

mRSS = modified Rodnan skin score; QILD = quantitative interstitial lung disease; SLS = Scleroderma Lung Study.

TABLE 3] Cox Proportional Hazards Model for Survival in SLS II (N = 90)^a

| Variable | Hazard Ratio | 95% CI | P Value | C Index |
|---------------------------------|--------------|------------|---------|---------|
| Univariate analyses | | | | |
| QILD score change of ≥ 2% | 3.17 | 1.15-8.76 | .026 | 0.688 |
| Continuous change in QILD score | 1.10 | 1.03-1.17 | .004 | 0.735 |
| Multivariate analyses | | | | |
| Age | 1.05 | 1.00-1.11 | .074 | 0.743 |
| mRSS | 1.00 | 0.94-1.07 | .614 | |
| Baseline FVC % predicted | 1.01 | 0.95-1.06 | .895 | |
| QILD score change of ≥ 2% | 3.86 | 1.31-11.27 | .014 | |

mRSS = modified Rodnan skin score; QILD = quantitative interstitial lung disease; SLS = Scleroderma Lung Study.

not be created. However, similar to SLS I, we also created a Cox model that included the change in FVC percent predicted at 24 months (measured continuously) as a covariate, and a suggestive association was found between change in FVC percent predicted with mortality (HR, 0.93; *P* = .06) (e-Table 8).

Discussion

Using data from two independent cohorts with extensive clinical characterization and follow-up, the present study demonstrated that radiographic progression of SSc-ILD (defined by a ≥ 2% increase in QILD score) over the course of 1 to 2 years is associated with an increased risk of long-term mortality. Even after adjusting for other factors known to affect survival in this patient population, worsening quantitative radiographic extent of ILD remained associated significantly with mortality in SLS II (suggestive association for SLS I).

Individualization of end points in RCTs for SSc-ILD is an evolving area of research with considerable implications for future drug discovery and development. Because no valid patient-reported outcomes currently exist for this disease state and an insufficient number of deaths occur over standard trial periods,³⁻⁵ an unmet need exists for establishing clinically meaningful and reliable surrogate end points for SSc-ILD RCTs.

The FVC has served as the primary end point for three of the largest SSc-ILD RCTs³⁻⁵; however, these trials all used different statistical methods for evaluating FVC change over time. For instance, in SLS I, a generalized estimating equation regression model was created that used imputation for missing data to compare outcomes in the FVC percent predicted over 1 year.³ In SLS II, a joint model that combined a mixed-effects model with a survival model to adjust for nonignorable missing data resulting from dropouts, treatment failures, and deaths

was used to compare the course of the FVC percent predicted over 2 years.⁴ In the Safety and Efficacy of Nintedanib in Systemic Sclerosis (SENSCIS) trial comparing nintedanib with placebo for SSc-ILD,⁵ the annual rate of decline in FVC (milliliters per year) over 1 year was assessed with a random-coefficient regression model that also used imputation for missing data. In the recently published phase 3 trial of tocilizumab for the treatment of diffuse cutaneous SSc, FVC was a key secondary end point, and that study evaluated the difference in distribution of change from baseline to week 48 in the FVC percent predicted.¹⁶ The diverse methodologies used to evaluate how various treatments affect lung function render it difficult to compare treatment effects across studies. In addition, studies have demonstrated considerable fluctuations in declines in the FVC percent predicted within individual patients receiving treatment for ILD.²⁰ Finally, lung function outcomes correlated poorly with various patient-reported outcomes in some trials,^{5,21,22} raising the question of whether lung function change truly is a clinically meaningful end point from the perspective of the patient.

The present study suggests that changes in the quantitative radiographic extent of ILD may serve as a proxy for long-term mortality in patients with SSc-ILD. Moreover, increased radiographic progression of ILD remained associated significantly with increased mortality in the multivariate analysis in SLS II, whereas physiologic progression on pulmonary function testing was not. These findings are consistent with studies of patients with other ILDs.²² For example, a recent study demonstrated that longitudinal changes in semiquantitative visual HRCT scan fibrosis scores of > 7% predicted lung transplantation-free survival in patients with idiopathic pulmonary fibrosis followed up for an average of 3 years.²³

Moreover, progression of ILD abnormalities in participants of the Framingham Heart Study based on visual assessment was associated with increased mortality.²⁴ However, the present study is the first study to demonstrate that increased quantitative radiographic progression of ILD beyond a prespecified threshold over the course of both 1- and 2 years was associated with an increased risk of death in patients with SSc.

In addition to predicting survival, a prior study from SLS II demonstrated that changes in the quantitative radiographic extent of ILD in the entire lung were associated significantly with patient-reported outcomes, such as dyspnea.²² Specifically, increased QILD score over 2 years was associated with worse scores on the transitional dyspnea index.²² Thus, radiographic progression of ILD may closely reflect actual changes in how a patient feels and functions.

In considering the results of the multivariate analyses, the HR for change in QILD score was higher for the SLS II survival model compared with the SLS I model. This finding could suggest that radiographic progression measured at 2 years is a better predictor of long-term mortality than that at 1 year, although to confirm this hypothesis, one would need to measure QILD score changes at these two time points in the same cohort. It is also notable that although substantially more patients randomized to placebo experienced radiographic progression compared with cyclophosphamide, the long-term mortality rates were similar between the SLS I study treatment arms, suggesting that 1 year of therapy is unlikely to lead to a sustained improvement in long-term survival.

Interestingly, the baseline characteristics of study participants who experienced radiographic progression of ILD and those who did not were fairly similar. Although not statistically significant, a greater proportion of men and Black participants experienced radiographic progression of ILD during these studies. These analyses may have been underpowered to detect significant differences, and larger studies are needed to determine whether sex and race affect radiographic progression of SSc-ILD.

Although quantitative imaging analysis can detect changes in the radiographic extent of ILD more sensitively than visual assessment,²⁵ a major limitation of this study is that automated methods for radiographic ILD scoring are not yet available widely in clinical

practice. However, this approach has been used in a number of SSc RCTs,^{3,4,16} demonstrating its feasibility as a study end point. Furthermore, insufficient inspiration, respiratory muscle strain, or both may curtail lung expansion during HRCT scan assessment. Furthermore, superimposed infection also may confound quantitative image assessment. Reassuringly, study technicians were trained to ensure adequate inspiration, and all images underwent quality control assessment to evaluate for the presence of infection before quantitative image analysis. Another potential shortcoming of this study is that radiographic assessment was performed at different time points in SLS I (1 year) and SLS II (2 years), which could introduce lead time bias. However, few deaths occurred during the active treatment periods ($n = 5$ and $n = 16$ for SLS I and II, respectively). Moreover, this also could be perceived as a strength because the findings suggest that radiographic changes at either of these time points have prognostic value.

Additional strengths of this study include the evaluation of two multicenter SSc-ILD cohorts who received standard treatment and follow-up during the radiographic assessment period. Moreover, although not all participants underwent a follow-up HRCT scan assessment, more than half of all patients did, and no difference was found in the baseline characteristics of participants without follow-up HRCT scan assessment and the entire study cohort. Finally, to our knowledge, this is the longest period during which participants in any RCT for SSc-ILD have been followed up for mortality outcomes.

Interpretation

In summary, measuring changes in the quantitative radiographic extent of ILD predicts long-term mortality in patients with SSc. Moreover, radiographic progression of ILD seemed to be a stronger predictor of mortality than longitudinal functional decline in two independent, multicenter cohorts. Radiographic end points may serve as more reliable and reproducible end points in SSc-ILD trials compared with FVC, and this is the first study to suggest that these end points are surrogates for mortality. Future studies are needed to determine whether measuring the change in QILD score at earlier time points (eg, 6 months) could be used to detect treatment effects and predict long-term outcomes in these patients.

Acknowledgments

Author contributions: E. R. V. takes full responsibility for the content of the manuscript, including the data and analysis. E. R. V. made substantial contributions to drafting the work. All authors made substantial contributions to the conception and design of the work; the acquisition, analysis, or interpretation of data for the work; revising the work critically for important intellectual content; final approval of the version to be published; and agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are investigated and resolved appropriately.

Financial/nonfinancial disclosures: The authors have reported to CHEST the following: E. R. V. reports consulting and speaker fees from Boehringer Ingelheim and grant support from Boehringer Ingelheim, Kadmon and Horizon Therapeutics. D. P. T. reports grant support from Genentech/Roche. M. D. R. reports grant support from Genentech/Roche. None declared (J. G., G. H. J. K.).

Role of sponsors: The study sponsors provided funding for study materials and data analysis. They had no role in the study design; collection, analysis, and the interpretation of the data; in the writing of the report; and in the decision to submit the paper for publication.

Other contributions: The data analyzed in the present study were collected under the National Heart, Lung, and Blood Institute, National Institutes of Health Grants R01 HL089901 and U01 HL 60606 to Robert M. Elashoff. The following people and institutions participated in the SLS I—*University of California at Los Angeles, Los Angeles:* P. J. Clements, D. P. Tashkin, R. Elashoff, J. Goldin, M. Roth, D. Furst, K. Bulpitt, D. Khanna, W.-L. J. Chung, S. Viasco, M. Sterz, L. Woolcock, X. Yan, J. Ho, S. Vasunilashorn, 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, J. E. Wilson; *University of Illinois Chicago, Chicago:* J. Varga, D. Schraufnagel, A. Wilbur, M. Lopata, S. Arami, P. Cole-Saffold; *Boston University, Boston:* R. Simms, A. Theodore, P. Clarke, J. Korn, K. Tobin, M. Nuite; *Medical University of South Carolina, Charleston:* R. Silver, M. Bolster, C. Strange, S. Schabel, E. Smith, J. Arnold, K. Caldwell, M. Bonner; *Johns Hopkins School of Medicine, Baltimore, MD:* R. Wise, F. Wigley, B. White, L. Hummers, M. Bohlman, A. Polito, G. Leatherman, E. Forbes, M. Daniel; *Georgetown University, Washington, DC:* V. Steen, C. Read, C. Cooper, S. Wheaton, A. Carey, A. Ortiz; *University of Texas at Houston, Houston:* M. Mayes, E. Parsley, S. Oldham, T. Filemon, S. Jordan, M. Perry; *University of California at San Francisco, San Francisco:* K. Connolly, J. Golden, P. Wolters, R. Webb, J. Davis, C. Antolos, C. Maynetto; *University of Alabama at Birmingham, Birmingham:* B. Fessler, M. Olman, C. Sanders, L. Heck, T. Parkhill;

University of Connecticut Health Center, Farmington: N. Rothfield, M. Metersky, R. Cobb, M. Aberles, F. Ingenito, E. Breen; *Wayne State University, Detroit, MI:* M. Mayes, K. Mubarak, J. L. Granda, J. Silva, Z. Injic, R. Alexander; *Virginia Mason Research Center, Seattle, WA:* D. Furst, S. Springmeyer, S. Kirkland, J. Molitor, R. Hinke, A. Mondt; *Data safety and monitoring board:* T. Thompson (Harvard Medical School, Boston), S. Rounds (Veterans Affairs Medical Center, Brown University, Providence, RI); M. Weinstein (Cedars Sinai—UCLA, Los Angeles), B. Thompson (Clinical Trials Surveys, Baltimore, MD); *Mortality and Morbidity Review Committee:* H. Paulus, S. Levy (UCLA, Los Angeles), D. Martin (Johns Hopkins University, Baltimore, MD). The following people and institutions participated in the SLS Study II—*University of Boston, Boston:* A. C. Theodore, R. W. Simms, E. Kissin, F. Y. Cheong; *Georgetown University, Washington, DC:* V. D. Steen, C. A. Read Jr., C. Fridley, M. Zulmatashvili; *Johns Hopkins University, Baltimore, MD:* R. A. Wise, F. M. Wigley, L. Hummers, G. Leatherman; *Medical University of South Carolina, Charleston:* R. M. Silver, C. Strange, F. N. Hant, J. Ham, K. Gibson, D. Rosson; *University of California, Los Angeles, Los Angeles:* D. P. Tashkin, R. M. Elashoff, M. D. Roth, P. J. Clements, D. Furst, S. Kafaja, E. Kleerup, D. Elashoff, J. Goldin, E. Ariola, G. Marlis, J. Mason-Berry, P. Saffold, M. Rodriguez, L. Guzman, J. Brook; *University of California, San Francisco, San Francisco:* J. Golden, M. K. Connolly, A. Eller, D. Leong, M. Lalosh, J. Obata; *University of Illinois, Chicago:* S. Volkov, D. Schraufnagel, S. Arami, D. Franklin; *Northwestern University, Chicago:* J. Varga, J. Dematte, M. Hinchliff, C. DeLuca, H. Donnelly, C. Marlin; *University of Medicine and Dentistry of New Jersey, New Brunswick:* D. J. Riley, V. M. Hsu, D. A. McCloskey; *University of Michigan, Ann Arbor:* K. Phillips, D. Khanna, F. J. Martinez, E. Schiopu, J. Konkle; *University of Texas, Houston:* M. Mayes, B. Patel, S. Assassi, F. Tan; *National Jewish Health, Denver:* A. Fischer, J. Swigris, R. Meehan, K. Brown, T. Warren, M. Morrison; *University of Utah, Salt Lake City:* M. B. Scholand, T. Frecht, P. Carey, M. Villegas; *University of Minnesota, Minneapolis:* J. Molitor, P. Carlson.

Additional information: The e-Appendix, e-Figures, and e-Tables are available online under “Supplementary Data.”

References

1. Tyndall AJ, Bannert B, Vonk M, et al. Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. *Ann Rheum Dis.* 2010;69:1809-1815.
2. Elhai M, Meune C, Boubaya M, et al. Mapping and predicting mortality from systemic sclerosis. *Ann Rheum Dis.* 2017;76:1897-1905.
3. Tashkin DP, Elashoff R, Clements PJ, et al. Cyclophosphamide versus placebo in

scleroderma lung disease. *N Engl J Med.* 2006;354:2655-2666.

4. Tashkin DP, Roth MD, Clements PJ, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease: Scleroderma Lung Study II (SLS II), a double-blind, parallel group, randomized controlled trial. *Lancet Respir Med.* 2016;4:708-719.
5. Distler O, Highland KB, Gahlemann M, et al. Nintedanib for systemic sclerosis-associated interstitial lung disease. *N Engl J Med.* 2019;380:2518-2528.
6. Goh NS, Hoyle RK, Denton CP, et al. Short-term pulmonary function trends are predictive of mortality in interstitial lung disease associated with systemic sclerosis. *Arthritis Rheum.* 2017;69:1670-1678.
7. Volkmann ER, Tashkin DP, Sim M, et al. Short-term progression of interstitial lung disease in systemic sclerosis predicts long-term survival in two independent clinical trial cohorts. *Ann Rheum Dis.* 2019;78:122-130.
8. Nawata T, Shirai Y, Suzuki M, Kuwana M. Chest wall muscle atrophy as a contributory factor for forced vital capacity decline in systemic sclerosis-interstitial lung disease. *Rheumatology (Oxford).* 2021;60:250-255.
9. Tashkin DP, Volkmann ER, Tseng C-H, et al. Relationship between quantitative radiographic assessments of interstitial lung disease and physiologic and clinical features of systemic sclerosis. *Ann Rheum Dis.* 2016;75:374-381.
10. Camiciottoli G, Orlandi I, Bartolucci M, et al. Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. *Chest.* 2007;131:672-681.
11. Khanna D, Nagaraja V, Tseng CH, et al. Predictors of lung function decline in scleroderma-related interstitial lung disease based on high-resolution computed tomography: implications for cohort enrichment in systemic sclerosis-associated interstitial lung disease trials. *Arthritis Res Ther.* 2015;17:372.
12. Roth MD, Tseng CH, Clements PJ, et al. Predicting treatment outcomes and responder subsets in scleroderma-related interstitial lung disease. *Arthritis Rheum.* 2011;63:2797-2808.
13. Kim HJ, Brown M, Elashoff F, 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 Radiology.* 2011;21:2455-2465.
14. Kim HJ, Tashkin DP, 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.
15. Goldin JG, Kim GH, Volkmann ER, et al. Longitudinal changes in quantitative lung disease on CT after immunosuppression in the Scleroderma Lung Study II. *Ann Am Thorac Soc.* 2018;15:1286-1295.

16. Khanna D, Lin CJF, Furst DE, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Respir Med*. 2020;8:963-974.
17. Obuchowski NA, Reeves AP, Huang EP, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. *Stat Methods Med Res*. 2015;24:68-106.
18. Kafaja S, Clements P, Wilhalme H, et al. Reliability and minimally clinically important differences of forced vital capacity: results from the Scleroderma Lung Studies (SLS-I and SLS-II). *Am J Respir Crit Care Med*. 2018;197:644-652.
19. Khanna D, Mittoo S, Aggarwal R, et al. Connective tissue disease-associated interstitial lung diseases (CTD-ILD)—report from OMERACT CTD-ILD Working Group. *J Rheumatol*. 2015;42:2168-2171.
20. Nathan SD, Albera C, Bradford WZ, et al. Effect of continued treatment with pirfenidone following clinically meaningful declines in forced vital capacity, analysis of data from three phase 3 trials in patients with idiopathic pulmonary fibrosis. *Thorax*. 2016;71:429-435.
21. Khanna D, Yan X, Tashkin DP, 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;56:1676-1684.
22. Volkman ER, Tashkin DP, LeClair H, et al. Treatment with mycophenolate mofetil and cyclophosphamide leads to clinically meaningful improvements in patient-reported outcomes in scleroderma lung disease: results of Scleroderma Lung Study II. *ACR Open Rheum*. 2020;2:362-370.
23. Taha N, D'Amato D, Hosein K, et al. Longitudinal functional changes with clinically significant radiographic progression in idiopathic pulmonary fibrosis: are we following the right parameters. *Respir Res*. 2020;21:119.
24. Araki T, Putman RK, Hatabu H, et al. Development and progression of interstitial lung abnormalities in the Framingham Heart Study. *Am J Respir Crit Care Med*. 2016;194:1514-1522.
25. Kim HG, Tashkin DP, Clements PJ, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol*. 2010;28:S26-S35.