



St. George's Respiratory Questionnaire Has Longitudinal Construct Validity in Lymphangioleiomyomatosis

Jeffrey J. Swigris, DO, MS; Hye-Seung Lee, PhD; Marsha Cohen, MD; Yoshikazu Inoue, MD; Joel Moss, MD, PhD, FCCP; Lianne G. Singer, MD, FCCP; Lisa R. Young, MD; and Francis X. McCormack, MD, FCCP; for the National Institutes of Health Rare Lung Diseases Consortium and the Multicenter International Lymphangioleiomyomatosis Efficacy and Safety of Sirolimus Trial Group

Background: Lymphangioleiomyomatosis (LAM) is an uncommon, progressive, cystic lung disease that causes shortness of breath, hypoxemia, and impaired health-related quality of life (HRQL). Whether St. George's Respiratory Questionnaire (SGRQ), a respiratory-specific HRQL instrument, captures longitudinal changes in HRQL in patients with LAM is unknown.

Methods: Using data from the Multicenter International Lymphangioleiomyomatosis Efficacy and Safety of Sirolimus trial, we performed analyses to examine associations between SGRQ scores and values for four external measures (anchors). Anchors included (1) FEV₁, (2) diffusing capacity of the lung for carbon monoxide, (3) distance walked during the 6-min walk test, and (4) serum vascular endothelial growth factor-D.

Results: SGRQ scores correlated with the majority of anchor values at baseline, 6 months, and 12 months. Results from longitudinal analyses demonstrated that SGRQ change scores tracked changes over time in values for each of the four anchors. At 12 months, subjects with the greatest improvement from baseline in FEV₁ experienced the greatest improvement in SGRQ scores (Symptoms domain, -13.4 ± 14.6 points; Activity domain, -6.46 ± 8.20 points; Impacts domain, -6.25 ± 12.8 points; SGRQ total, -7.53 ± 10.0 points). Plots of cumulative distribution functions further supported the longitudinal validity of the SGRQ in LAM.

Conclusions: In LAM, SGRQ scores are associated with variables used to assess LAM severity. The SGRQ is sensitive to change in LAM severity, particularly when change is defined by FEV₁, perhaps the most clinically relevant and prognostically important variable in LAM. The constellation of results here supports the validity of the SGRQ as capable of assessing longitudinal change in HRQL in LAM.

CHEST 2013; 143(6):1671-1678

Abbreviations: 6MWD = distance walked during 6-min walk test; CDF = cumulative distribution function; DLCO = diffusing capacity of the lung for carbon monoxide; LAM = lymphangioleiomyomatosis; SGRQ = St. George's Respiratory Questionnaire; VEGF-D = serum vascular endothelial growth factor D level

Lymphangioleiomyomatosis (LAM) is an uncommon, progressive lung disease that affects women and occurs either sporadically or in association with tuberous sclerosis complex.¹⁻³ In LAM, hallmark cystic destruction of the pulmonary parenchyma impairs lung function and induces debilitating dyspnea.^{4,5} In most patients, hypoxemia develops within a decade of symptom onset.⁶ Given the breathlessness, functional limitation, and need for supplemental oxygen that ultimately develop, it is not surprising that patients with LAM experience impaired health-related quality

of life (HRQL), particularly in domains that reflect respiratory symptoms or physical health and activities.⁷

St. George's Respiratory Questionnaire (SGRQ) is a respiratory-specific instrument that was designed to assess HRQL in patients with asthma or COPD.⁸ Despite the developer's initial intent—to develop a questionnaire for patients with either of those two conditions—the SGRQ has been used to assess HRQL in patients with other lung diseases, including women with LAM.⁹ In fact, investigators have observed that baseline scores from the SGRQ correlate with certain

baseline measures of pulmonary physiology, oxygenation, and functional capacity.^{7,9} The SGRQ has been shown to be sensitive to change when used in patients with various respiratory diseases,¹⁰⁻¹² but whether in patients with LAM it can track changes in HRQL that might occur as a result of disease progression or in response to a clinically beneficial (or harmful) medication has never been assessed. An HRQL instrument must possess this attribute to be considered useful as an outcome measure in longitudinal research. We conducted this study to examine the ability of the SGRQ to assess HRQL over time in patients with LAM (ie, to determine its longitudinal construct validity in this disease).

MATERIALS AND METHODS

We used data collected at baseline, 6 months, and 12 months in the Multicenter International Lymphangioleiomyomatosis Efficacy and Safety of Sirolimus (MILES) trial.¹³ The MILES trial was a two-stage trial—a 12-month randomized, double-blinded comparison of sirolimus vs placebo followed by a 12-month observation period—involving 89 patients with LAM who had a FEV₁ < 70% predicted. The primary end point was the difference between the groups in the rate of change in FEV₁. There were a number of secondary end points, including HRQL as assessed by the SGRQ. We assessed the association between SGRQ scores and certain variables hypothesized to be clinically meaningful measures of LAM severity; henceforth, we refer to those variables

as anchors. We hypothesized that changes in the anchors (ie, disease status or severity) would be associated with changes in HRQL and, thus, changes in SGRQ scores.

Saint George's Respiratory Questionnaire

The SGRQ is a self-administered, respiratory-specific questionnaire with three domains (Symptoms, Activity, and Impacts) and a total score designed to assess HRQL.⁸ The Symptoms domain, as its name implies, focuses on respiratory symptoms, including breathlessness, cough, and wheeze. The Activity domain probes for physical activities that either cause or are limited by dyspnea. The Impacts domain covers the effects of respiratory disease on several factors, including employment, social interactions, emotional well-being, and the sense of being in control. Scoring weights for the response options for each of the 50 items were derived using data from patients with asthma or COPD.¹⁴ Each domain score and the total score range from 0 to 100, and higher scores connote greater impairment.

Anchors

The four anchors we selected were: (1) FEV₁, (2) diffusing capacity of the lung for carbon monoxide (DLCO), (3) distance walked during the 6-min walk test (6MWD), and (4) serum vascular endothelial growth factor D level (VEGF-D). We chose FEV₁ and DLCO because each has been shown to be impaired in patients with LAM, and as such, they—particularly FEV₁—are measures used universally to characterize LAM severity.⁷ Both the FEV₁ and DLCO have been shown in prior cross-sectional studies to correlate with SGRQ scores.^{7,9} The 6-min walk test, and 6MWD in particular, is commonly used as a functional assessment in patients with respiratory diseases.¹⁵ We hypothesized that changes in 6MWD would reflect changes in overall physical functionality and that changes in physical functionality would lead to changes in subjects' perceptions of their HRQL. Serum VEGF-D level distinguishes LAM from other diseases, and here, changes in VEGF-D levels were hypothesized to track changes in LAM severity (eg, increases in VEGF-D would be associated with increased LAM severity) and, by association, HRQL.¹⁶

Statistical Analysis

We performed analyses to examine the association between SGRQ scores and anchor values cross-sectionally, as well as longitudinally, using the data collected at three different time points: baseline, 6 months, and 12 months. To enhance interpretability of results, serum VEGF-D values were log-transformed. In the first set of analyses, Spearman rank correlation was used to assess the relationship among variables cross-sectionally and longitudinally. Linear mixed-effects models were used to further examine these cross-sectional and longitudinal associations simultaneously. For each anchor, we generated four separate mixed-effects models (one model for each of the three SGRQ domains and one for the SGRQ total score). In each model, SGRQ score was the response variable and the anchor was a covariate. Each model included age at enrollment as a time-constant variable, the baseline anchor value (to allow examination of the cross-sectional association between SGRQ score and anchor), and the anchor change from baseline (to allow examination of the longitudinal association). For each model, to account for the within-subject correlation among the repeated measures in SGRQ score, anchor intercept and slope were incorporated as random effects with unstructured covariance. Next, we used general linear models to examine the association between SGRQ changes and quartiles of anchor change (defined as percent change from baseline at 6 or 12 months) after adjusting for the corresponding baseline SGRQ score. Finally, as a visual representation of the relationship between SGRQ change

Manuscript received January 19, 2012; revision accepted November 17, 2012.

Affiliations: From the Autoimmune Lung Center and Interstitial Lung Disease Program (Dr Swigris), National Jewish Health, Denver, CO; the Pediatrics Epidemiology Center (Dr Lee), University of South Florida College of Medicine, Tampa, FL; the Women's Health Research Institute (Dr Cohen), Women's College Hospital, Toronto, ON, Canada; the National Hospital Organization Kinki-Chou Chest Medical Center (Dr Inoue), Osaka, Japan; the National Institutes of Health (Dr Moss), Bethesda, MD; the Lung Transplantation Program (Dr Singer), University of Toronto; the Division of Allergy, Pulmonary, and Critical Care Medicine (Dr Young), Vanderbilt University School of Medicine, Nashville, TN; and the Department of Internal Medicine (Dr McCormack), Division of Pulmonary, Critical Care and Sleep Medicine, University of Cincinnati, Cincinnati, OH.

Funding/Support: Dr Swigris is supported in part by a Career Development Award from the National Institutes of Health (NIH) [Grant K23 HL092227]. Dr Moss was supported by the Intramural Research Program, NIH, National Heart, Lung and Blood Institute. The Multicenter International Lymphangioleiomyomatosis Efficacy and Safety of Sirolimus Trial was supported by the NIH Office of Rare Disease Research [Grant RR019498]; the US Food and Drug Administration [Grant FD003362]; the Japanese Ministry of Health, Labour, and Welfare; the Canadian Institutes of Health Research; Pfizer Pharmaceuticals; The LAM Foundation; the Tuberous Sclerosis Alliance; and Cincinnati Children's Hospital.

Correspondence to: Jeffrey J. Swigris, DO, MS, Associate Professor of Medicine, Autoimmune Lung Center and Interstitial Lung Disease Program, National Jewish Health, Southside Bldg, Office #G011, 1400 Jackson St, Denver, CO 80206; e-mail: swigrisj@njc.org
© 2013 American College of Chest Physicians. Reproduction of this article is prohibited without written permission from the American College of Chest Physicians. See online for more details.
DOI: 10.1378/chest.12-0161

scores and anchor change scores (again, defined as percent change from baseline), we generated cumulative distribution function (CDF) plots for each SGRQ score using data from subjects in the two extreme quartiles of anchor changes (greatest decline vs greatest improvement). Institutional review board approval was not needed for this study of deidentified, previously collected data. All statistical analyses were performed using SAS, Version 9.2 (SAS Institute Inc.), and $P < .05$ was considered to represent statistical significance for each analysis. Because our analyses were hypothesis driven, we did not adjust for multiple comparisons.

RESULTS

Baseline characteristics of the subjects are displayed in Table 1. On average, airflow limitation was moderately severe. At baseline, 6 months, and 12 months, there were significant correlations—most moderately strong—between various anchor values and SGRQ scores (Table 2). For each anchor, at each time point, the strongest correlations were with Activity domain scores (Table 2). Simple correlations between SGRQ change scores and anchor change scores are presented in Table 3.

The results from the mixed-effects models extend the results from the correlation analyses by yielding estimates for the cross-sectional relationship between SGRQ scores and anchors at baseline (Table 4, “Cross-sectional”) as well as how SGRQ scores were predicted to change in relation to changes over time in the anchors (Table 4, “Longitudinal”). At any time

point (ie, 6 or 12 months), improvements from baseline in any of the four anchors were predicted to generate improved (lower) SGRQ scores. For example, at 6 or 12 months, every 1% increase in FEV₁% was predicted to yield about a 0.5-point decrease (improved HRQL) in any SGRQ score (domain or total), and a 100-m increase in 6MWD was predicted to yield a 4-point decrease (improved HRQL) in the Symptoms, Activity, or SGRQ total score and a 5-point decrease in Impacts score.

Table 5 displays mean SGRQ change scores for subgroups stratified on quartiles of change in each anchor. After adjusting for baseline SGRQ score, there were significant differences between various SGRQ scores across quartiles of change in FEV₁, 6MWD, and log(VEGF-D). Although not all relationships were statistically significant, on balance, we observed that greater impairments (ie, decline in FEV₁, decline in DLCO, decline in 6MWD, or increase in VEGF-D) in the anchors were associated with greater impairments (increase) in SGRQ scores, and greater improvement in the anchors were associated with greater improvements (decrease) in SGRQ scores.

The CDF plots (Fig 1) provide a graphical representation of SGRQ change scores for subgroups in the extreme quartiles of change for each anchor. Consider an arbitrarily chosen improvement of at least 10 points in the Symptoms domain (ie, -10 or more extreme on the x-axis in Fig 1A): only 33% of subjects in the first quartile of change in FEV₁ (Q1: subjects with decline in raw FEV₁ of $> 11.1\%$ from baseline) compared with 61% of subjects in the fourth quartile of change in FEV₁ (Q4: subjects with an improvement in FEV₁ of $> 4.4\%$) had improvement of at least 10 points in the Symptoms domain from baseline to 12 months. For the Activity domain, 17% in Q1 vs 42% in Q4 had improvement of at least 10 points; for the Impacts domain, 17% in Q1 vs 32% in Q4 had improvement of at least 10 points; and for the SGRQ total, 17% in Q1 vs 28% in Q4 had improvement of at least 10 points.

DISCUSSION

Using data from the MILES trial, we performed several analyses whose results support the SGRQ as an instrument capable of assessing and responding to change over time in HRQL in patients with LAM. When investigators aim to generate data to support the validity of something (for an intended purpose), ideally, there is a gold standard against which it can be measured. Unfortunately, in the realm of HRQL, there is no gold standard. Thus, investigators must choose other variables, hypothesized—or known—to be related to HRQL to serve as anchors. In patients with LAM, pulmonary function tests (FEV₁ in particular)

Table 1—Characteristics of Study Sample

Variable	Baseline (n = 89)	6 mo (n = 83)	12 mo (n = 74)
Age, y	45.4 ± 10.6
Race, No. (%)			
White	59 (66)
Asian	27 (30)
Other	3 (3)
Oxygen therapy, No. (%)			
Intermittent use	52 (58)
Continuous use	28 (31)
FEV ₁			
L	1.37 ± 0.42	1.35 ± 0.42	1.33 ± 0.40
% Predicted	48.5 ± 13.8	48.5 ± 15.0	48.1 ± 14.1
DLCO			
mL/min/mm Hg	10.23 ± 4.61	9.95 ± 4.00	9.62 ± 3.95
% Predicted	43.4 ± 19.0	42.3 ± 16.3	41.2 ± 16.8
6MWD, m	403 ± 105	415 ± 119	425 ± 104
log(VEGF-D), pg/mL	7.22 ± 0.85	7.00 ± 0.82	6.90 ± 0.84
SGRQ scores			
Symptom domain	52.1 ± 18.4	49.8 ± 18.3	48.5 ± 19.2
Activity domain	65.5 ± 17.7	65.2 ± 18.3	64.5 ± 17.6
Impact domain	33.4 ± 17.4	34.5 ± 17.9	31.3 ± 15.0
Total	46.4 ± 15.2	46.6 ± 15.6	44.6 ± 13.7

Data are presented as mean ± SD unless otherwise noted. 6MWD = distance walked during 6-min walk test; DLCO = diffusing capacity of the lung for carbon monoxide; SGRQ = St. George's Respiratory Questionnaire; VEGF-D = serum vascular endothelial growth factor D level.

Table 2—Spearman Correlation Coefficients Between SGRQ Scores and Anchors at Baseline, 6 Months, and 12 Months

Domain	Symptom	Activity	Impact	Total
FEV ₁ %				
Baseline	−0.33 (.002)	−0.35 (.0007)	−0.26 (.01)	−0.33 (.001)
6 mo	−0.41 (.0001)	−0.48 (<.0001)	−0.42 (<.0001)	−0.49 (<.0001)
12 mo	−0.35 (.002)	−0.46 (<.0001)	−0.29 (.01)	−0.40 (.0005)
DLCO%				
Baseline	−0.21 (.048)	−0.43 (<.0001)	−0.20 (.06)	−0.33 (.002)
6 mo	−0.30 (.006)	−0.49 (<.0001)	−0.34 (.002)	−0.45 (<.0001)
12 mo	−0.17 (.14)	−0.54 (<.0001)	−0.33 (.005)	−0.41 (.0004)
6MWD				
Baseline	−0.14 (.19)	−0.36 (.0005)	−0.13 (.23)	−0.23 (.03)
6 mo	−0.28 (.01)	−0.45 (<.0001)	−0.37 (.0006)	−0.46 (<.0001)
12 mo	−0.20 (.08)	−0.49 (<.0001)	−0.37 (.001)	−0.41 (.0003)
Log(VEGF-D)				
Baseline	0.09 (.42)	0.23 (.03)	0.18 (.10)	0.23 (.03)
6 mo	0.21 (.07)	0.27 (.02)	0.10 (.37)	0.19 (.09)
12 mo	0.19 (.12)	0.23 (.06)	0.15 (.20)	0.21 (.08)

Values are Spearman correlation coefficients (*P* value). See Table 1 legend for expansion of abbreviations.

yield important prognostic information and are commonly used to assess disease status. And by extension, they yield some information about patients' well-being or HRQL. Thus, we hypothesized FEV₁ and DLCO as well as 6MWD (as a measure of functional status) would be related to HRQL in patients with LAM and, as such, in the absence of a gold standard for HRQL, would be useful as external anchors for our analyses. Serum VEGF-D has emerged as a diagnostic biomarker for LAM¹⁶; like FEV₁, DLCO, and 6MWD, we hypothesized VEGF-D levels would change in response to changes in clinically defined disease severity and might, therefore, be associated with HRQL.

The MILES trial revealed that targeting the mTOR (mammalian Target of Rapamycin) pathway is an effective approach to treating LAM, and MILES very likely paved the way for future trials of other mTOR signaling antagonists for LAM. The hope for such a trial is that any physiologic benefit of a drug would translate into improvements in survival and—arguably,

perhaps even more importantly—patient symptoms, functional capacity, and sense of well-being. Survival is easy to assess; symptoms and the more abstract constructs (eg, HRQL) are more challenging, but they can be measured, too. The key to doing so is using reliable, valid instruments that are sensitive to underlying change in the construct of interest. Results from the current study support the SGRQ as such an instrument. We identified a number of significant correlations between SGRQ scores and anchors at each of the study time points. The moderately strong correlations we observed (common to analyses like this), rather than being disappointing, are in fact reassuring. They show that the SGRQ performed as hypothesized, but, most importantly, they reveal that the SGRQ captures information about patients with LAM that FEV₁, DLCO, 6MWD, and serum VEGF-D levels do not.

Even more important for the process of building longitudinal validity than identifying significant

Table 3—Spearman Correlation Coefficients Between SGRQ Change Scores and Anchor Change Scores From Baseline to 6 and 12 Months

Domain	Symptom	Activity	Impact	Total
FEV ₁ %				
6 mo	−0.18 (.09)	−0.34 (.0016)	−0.37 (.0005)	−0.42 (<.0001)
12 mo	−0.16 (.18)	−0.16 (.1621)	−0.24 (.0378)	−0.27 (.0204)
DLCO%				
6 mo	−0.02 (.83)	0.07 (.51)	−0.08 (.46)	−0.05 (.65)
12 mo	−0.04 (.76)	−0.11 (.32)	−0.05 (.66)	−0.08 (.52)
6MWD				
6 mo	−0.07 (.51)	−0.17 (.1181)	−0.15 (.18)	−0.18 (.1128)
12 mo	−0.22 (.06)	−0.30 (.0086)	−0.22 (.06)	−0.35 (.0023)
Log(VEGF-D)				
6 mo	0.06 (.63)	0.09 (.45)	0.19 (.09)	0.20 (.09)
12 mo	0.10 (.39)	0.15 (.23)	0.09 (.46)	0.15 (.23)

Values are Spearman correlation coefficients (*P* value). See Table 1 legend for expansion of abbreviations.

Table 4—Mixed-Effects Model-Generated Parameter Estimates for SGRQ Change Scores Resulting From Change in Anchor Scores

Associations	Domain	Symptoms	Activity	Impact	Total
Cross-sectional	FEV ₁ , mL	−0.014 ± 0.004 (.002)	−0.016 ± 0.004 (<.0001)	−0.015 ± 0.004 (.0009)	−0.015 ± 0.004 (.0001)
	FEV ₁ %	−0.46 ± 0.13 (.0009)	−0.48 ± 0.11 (<.0001)	−0.39 ± 0.13 (.003)	−0.42 ± 0.11 (.0002)
	DLCO, mL/min/mm Hg	−1.11 ± 0.45 (.02)	−1.72 ± 0.40 (<.0001)	−0.84 ± 0.58 (.21)	−1.15 ± 0.38 (.005)
	DLCO %	−0.30 ± 0.11 (.01)	−0.42 ± 0.10 (<.0001)	−0.20 ± 0.11 (.08)	−0.29 ± 0.09 (.004)
	6MWD, m	−0.03 ± 0.02 (.12)	−0.07 ± 0.02 (.0002)	−0.04 ± 0.02 (.03)	−0.05 ± 0.02 (.004)
Longitudinal	Log(VEGF-D)	4.14 ± 2.16 (.06)	5.95 ± 2.22 (.009)	3.45 ± 2.17 (.12)	4.34 ± 1.95 (.03)
	FEV ₁ , mL	−0.018 ± 0.007 (.02)	−0.020 ± 0.005 (<.0001)	−0.021 ± 0.006 (.0007)	−0.020 ± 0.005 (<.0001)
		0.583	0.744	0.714	0.750
	FEV ₁ %	−0.49 ± 0.22 (.03)	−0.56 ± 0.14 (<.0001)	−0.61 ± 0.18 (.001)	−0.57 ± 0.14 (.0003)
		0.531	0.759	0.744	0.776
	DLCO, mL/min/mm Hg	−0.23 ± 0.75 (.76)	−0.42 ± 0.56 (.47)	−0.03 ± 0.82 (.97)	−0.28 ± 0.53 (.60)
		0.580	0.788	0.784	0.797
	DLCO %	−0.06 ± 0.18 (.73)	−0.10 ± 0.14 (.49)	−0.05 ± 0.15 (.76)	−0.09 ± 0.12 (.47)
		0.580	0.798	0.783	0.803
	6MWD, m	−0.05 ± 0.02 (.01)	−0.02 ± 0.02 (.15)	−0.05 ± 0.02 (.008)	−0.04 ± 0.01 (.003)
		0.612	0.789	0.716	0.776
	Log(VEGF-D)	4.85 ± 2.51 (.06)	1.63 ± 1.70 (.34)	3.86 ± 1.89 (.04)	3.27 ± 1.51 (.03)
		0.636	0.825	0.763	0.815

Values are mixed-effects model parameter estimate ± SE (*P* value). In the Longitudinal section of the table, the value below the *P* value is the lowest of the estimated within-patient correlations from the model. See Table 1 for expansion of abbreviations.

correlations between static anchor values and SGRQ scores, we found that SGRQ scores tracked changes in each of the four anchors over time. For example, from the mixed-effects analyses, a 200-mL increase in FEV₁, or an 8% increase in FEV₁%, was predicted to result in a 4-point decline in SGRQ total score (connoting an improvement in HRQL). Likewise, a 5% increase in DLCO% was predicted to result in a > 5-point decline in SGRQ Activity score. These

analyses also revealed that change in VEGF-D was significantly associated with change in SGRQ scores, but this was for very large (one log) changes in serum VEGF-D levels. Given what is known about serum VEGF-D, this was not too surprising to us: VEGF-D is a diagnostic marker, but its value as a biomarker of pulmonary disease activity (i.e., whether longitudinal changes in VEGF-D correlate with lung disease progression in LAM) is, at present, uncertain. Perhaps

Table 5—SGRQ Change Scores for Subjects Stratified Into Quartiles Based on Change From Baseline to 12 Months in Anchor Values

Anchor	Symptoms	<i>P</i> Value	Activity	<i>P</i> Value	Impacts	<i>P</i> Value	Total	<i>P</i> Value
FEV ₁		.01		.006		.16		.01
≤ −11.2% (Q1)	−0.12 ± 14.6	.017	−1.71 ± 8.76	.06	0.95 ± 12.1	.025	−0.13 ± 7.98	.006
−11.2% to −3.4% (Q2)	1.45 ± 13.1	.015	1.31 ± 9.84	.004	−0.34 ± 9.96	.17	0.34 ± 8.16	.01
−3.4% to 4.4% (Q3)	4.57 ± 21.4	.002	3.26 ± 8.43	.001	−2.30 ± 13.1	.24	0.66 ± 10.0	.007
> 4.4% (Q4)	−13.4 ± 14.6	Ref	−6.46 ± 8.20	Ref	−6.25 ± 12.8	Ref	−7.53 ± 10.0	Ref
DLCO		.6		.5		.6		.6
≤ −9.9% (Q1)	−3.74 ± 15.2	.736	0.65 ± 8.57	.418	−1.83 ± 13.1	.363	−1.29 ± 8.42	.388
−9.9% to −4.7% (Q2)	2.27 ± 18.5	.193	0.06 ± 8.28	.559	−2.06 ± 10.1	.582	−0.82 ± 8.58	.402
−4.7% to 4.7% (Q3)	−1.02 ± 15.9	.438	−3.14 ± 9.75	.522	−3.15 ± 11.7	.687	−2.92 ± 9.98	.789
> 4.7% (Q4)	−4.60 ± 20.1	Ref	−1.04 ± 11.1	Ref	−2.20 ± 13.9	Ref	−2.07 ± 11.8	Ref
6MWD		.7		.02		.17		.04
≤ −2.3% (Q1)	−1.03 ± 17.9	.400	2.87 ± 7.73	.05	1.81 ± 11.9	.233	2.02 ± 9.18	.06
−2.3% to 4.5% (Q2)	2.54 ± 22.7	.241	0.97 ± 6.53	.37	−0.75 ± 9.35	.662	0.34 ± 7.99	.36
4.5% to 17.1% (Q3)	−3.32 ± 14.5	.449	−5.41 ± 10.8	.26	−6.56 ± 14.8	.301	−5.88 ± 10.2	.33
> 17.1% (Q4)	−5.53 ± 11.9	Ref	−2.22 ± 10.7	Ref	−2.60 ± 11.4	Ref	−2.88 ± 9.41	Ref
Log(VEGF-D)		.04		.8		.5		.16
≤ −9.0% (Q1)	−7.54 ± 15.3	.7	−2.22 ± 8.43	.602	−5.97 ± 16.1	.452	−5.17 ± 11.6	.437
−9.0% to −3.7% (Q2)	−2.70 ± 18.4	.4	−2.48 ± 9.75	.859	−0.78 ± 11.6	.628	−1.51 ± 10.1	.440
−3.7% to 0.8% (Q3)	6.29 ± 17.9	.02	1.43 ± 9.65	.625	2.76 ± 11.5	.505	2.95 ± 9.03	.158
> 0.8% (Q4)	−4.88 ± 12.8	Ref	0.14 ± 10.8	Ref	−3.27 ± 8.49	Ref	−2.44 ± 5.86	Ref

Values are mean ± SD. Boldface numbers are *P* values for comparison across SGRQ means within anchor; other *P* values correspond to pairwise comparisons of SGRQ means with anchor using Q4 as the reference. See Table 1 legend for expansion of abbreviations.

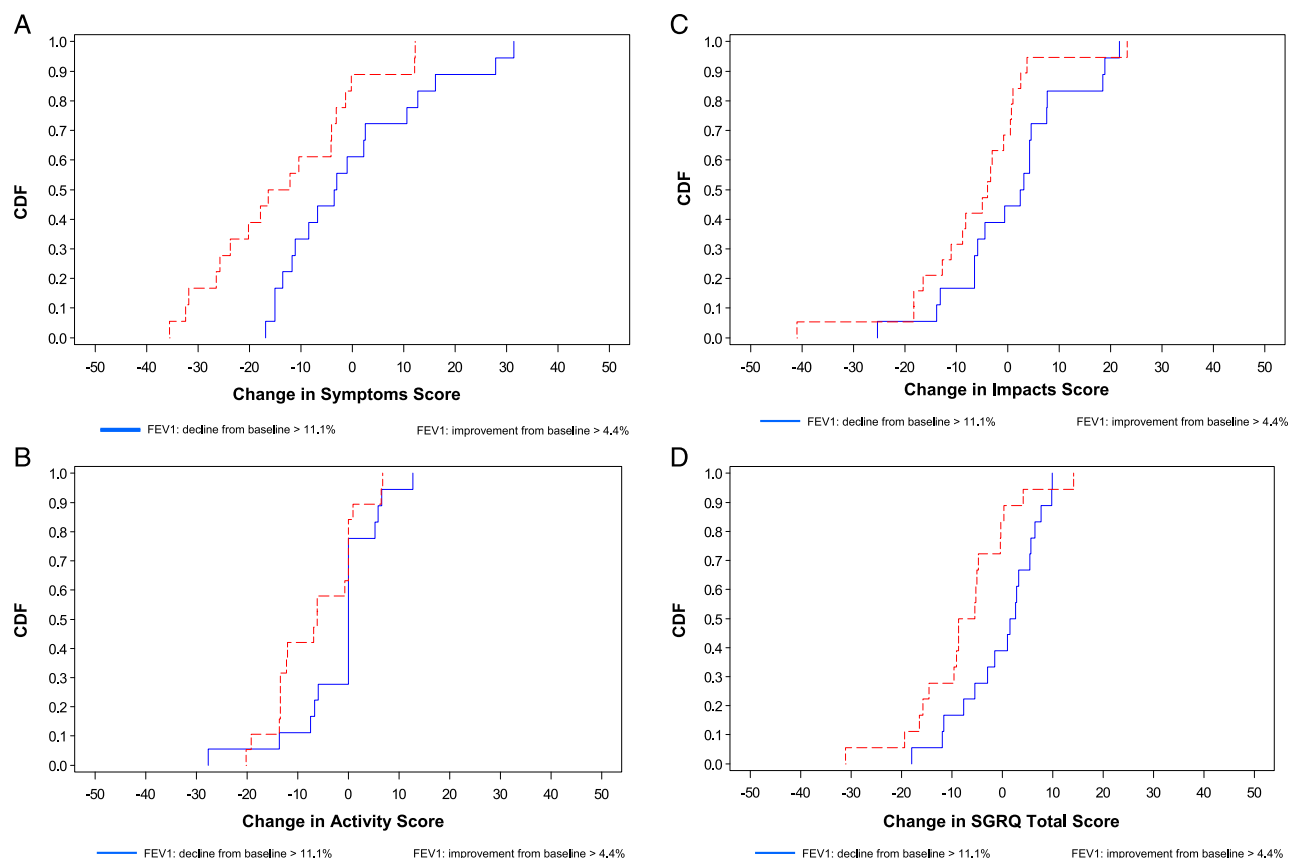


FIGURE 1. A, Plot of CDF for the two extreme quartiles of change from baseline in FEV₁ for SGRQ Symptoms domain. B, Plot of CDF for the two extreme quartiles of change from baseline in FEV₁ for SGRQ Activity domain. C, Plot of CDF for the two extreme quartiles of change from baseline in FEV₁ for SGRQ Impacts domain. D, Plot of CDF for the two extreme quartiles of change from baseline in FEV₁ for SGRQ total score. CDF = cumulative distribution function; SGRQ = St. George's Respiratory Questionnaire.

VEGF-D levels are exquisitely sensitive to LAM activity (at least lymphatic involvement) at the molecular level, but our clinical metrics (eg, FEV₁, DLCO, and 6MWD) are not: only extensive and prolonged lymphatic involvement (that drives up VEGF-D) translates into clinical worsening able to be captured by FEV₁, DLCO, or 6MWD. Clearly, more research into the role of VEGF-D as a biomarker of disease activity in patients with established LAM is warranted.

An effect of being sensitive to longitudinal changes in HRQL within a population is the ability to distinguish change over time between subgroups in that population. When we grouped subjects according to change over time in FEV₁, on balance, the SGRQ performed as hypothesized: subjects with the greatest improvement in FEV₁ experienced the greatest improvement in HRQL. For that particular set of analyses, as has been noted previously in similar analyses for another respiratory disease,¹² DLCO presents challenges as an anchor. First, DLCO changed very little (average change < 1 mL/mm Hg/min among all subjects) over the duration of MILES. Second, DLCO is a statistically noisy variable, and changes in DLCO may fluctuate for reasons unrelated to changes in

LAM severity. For example, DLCO is affected by changes in hemoglobin; DLCO was not adjusted for hemoglobin in MILES. Also, DLCO results are often affected by maldistribution of the inspired test gas mixture. This can occur when residual volume is elevated; in MILES, average residual volume was 141% of the predicted value. Finally, for DLCO and the other anchors, the loss of power induced by categorizing variables (as we did in this set of analyses) likely contributed to the inability to identify statistically significant differences between subgroups.

The SGRQ is just one of many instruments designed to assess HRQL and one of several respiratory-specific tools developed for this purpose. A mistake that has been perpetuated in the medical literature is that one cross-sectional correlation study—if statistically significant results emerge—“validates” an instrument for use in a longitudinal trial.¹⁷ In a handful of studies, investigators have generated data to support this so-called “concurrent validity” for the SGRQ (and other instruments) in LAM,^{7,9} but, to our knowledge, ours is the first to assess whether the SGRQ can be used confidently in LAM to assess change in longitudinal studies (eg, drug trials).^{18–20} The results of our

analyses are not surprising; however, these analyses must be done to confirm the SGRQ performs as hypothesized in LAM. This is the essence of building validity. The FDA has formalized recommendations for how instruments like the SGRQ might qualify as a valid, reliable outcome measure whose scores have “interpretable meaning” in a target population.²¹ To our knowledge, there have been no HRQL questionnaires submitted to the FDA for LAM, but data from this study could be useful in such a submission.

This study has limitations: the first is that subjects in the MILES trial may not be representative of the general LAM population. Given how uncommon LAM is, we doubt this is the case. However, subjects in MILES had moderately severe airflow limitation—more severe than the cohort enrolled in a nationwide registry⁷—so whether the SGRQ would perform equally well in a trial that included only subjects with milder LAM is uncertain. The results of our study cannot be extended to other HRQL questionnaires. It is possible that other HRQL questionnaires would perform similarly (or even better than the SGRQ) in LAM under the same circumstances. Until their longitudinal construct validity is assessed, those instruments cannot be used confidently in longitudinal LAM research. On the face of it, the SGRQ contains many items—those that ask about wheeze, cough, dyspnea, and physical activities—relevant to LAM patients.⁷ However, the SGRQ was not developed specifically for patients with LAM. Whether an instrument developed by specifically incorporating perspectives from patients with LAM would perform better than the SGRQ is unknown. Although not necessarily relevant to this study, in certain studies of patients with COPD, women’s scores are higher (greater impairment in HRQL) than men’s for the SGRQ.²² Some readers may not be familiar with mixed-effects models, but they are believed by most experts to be the models of choice when analyzing longitudinal data (including HRQL data from therapeutic trials).²³ These models parameterize the within-subject correlation that results from repeatedly measuring an outcome over time in the same individual, and they accommodate both incomplete data and time-varying covariates. In contrast, when simple correlation (or even linear regression) is used to assess the relationship between two variables, if data for either variable are missing for a subject, by necessity the subject is deleted from the analysis. Thus, mixed-effects models provide the most statistically efficient method for analyzing longitudinal data. Finally, certain analyses did not yield statistically significant results, but this was largely due to a loss of power resulting from categorization of continuous variables (eg, those in which the anchors were stratified into quartiles). The majority of our anal-

yses support the longitudinal validity of the SGRQ in LAM.

ACKNOWLEDGMENTS

Author contributions: Dr Swigris had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Dr Swigris: contributed to study conceptualization and planning, generating intellectual content for the manuscript, and critiquing and approving final content.

Dr Lee: contributed to study conceptualization and planning, analyzing data, generating intellectual content for the manuscript, and critiquing and approving final content.

Dr Cohen: contributed to generating intellectual content for the manuscript and critiquing and approving final content.

Dr Inoue: generating intellectual content for the manuscript and critiquing and approving final content.

Dr Moss: generating intellectual content for the manuscript and critiquing and approving final content.

Dr Singer: generating intellectual content for the manuscript and critiquing and approving final content.

Dr Young: generating intellectual content for the manuscript and critiquing and approving final content.

Dr McCormack: contributed to study conceptualization and planning, generating intellectual content for the manuscript, and critiquing and approving final content.

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

Role of sponsors: The sponsors had no role in the design of the study, the collection and analysis of the data, or in the preparation of the manuscript.

Other contributions: We thank Margaret Bevans, RN, PhD, Clinical Center, National Institutes of Health, for critical reading of the manuscript.

REFERENCES

1. Cabana H, Alexandre C, Agathos SN, Jones JP. Immobilization of laccase from the white rot fungus *Coriolopsis polyzona* and use of the immobilized biocatalyst for the continuous elimination of endocrine disrupting chemicals. *Bioresour Technol*. 2009;100(14):3447-3458.
2. Corrin B, Liebow AA, Friedman PJ. Pulmonary lymphangio-myomatosis. A review. *Am J Pathol*. 1975;79(2):348-382.
3. Taylor JR, Ryu J, Colby TV, Raffin TA. Lymphangioleiomyomatosis. Clinical course in 32 patients. *N Engl J Med*. 1990;323(18):1254-1260.
4. Johnson SR, Tattersfield AE. Decline in lung function in lymphangioleiomyomatosis: relation to menopause and progesterone treatment. *Am J Respir Crit Care Med*. 1999;160(2):628-633.
5. Taveira-DaSilva AM, Stylianou MP, Hedin CJ, Hathaway O, Moss J. Decline in lung function in patients with lymphangioleiomyomatosis treated with or without progesterone. *Chest*. 2004;126(6):1867-1874.
6. Johnson SR, Whale CI, Hubbard RB, Lewis SA, Tattersfield AE. Survival and disease progression in UK patients with lymphangioleiomyomatosis. *Thorax*. 2004;59(9):800-803.
7. Ryu JH, Moss J, Beck GJ, et al; NHLBI LAM Registry Group. The NHLBI lymphangioleiomyomatosis registry: characteristics of 230 patients at enrollment. *Am J Respir Crit Care Med*. 2006;173(1):105-111.
8. Jones PW, Quirk FH, Baveystock CM. The St George’s Respiratory Questionnaire. *Respir Med*. 1991;85(suppl B):25-31.
9. Xu KF, Wang L, Tian XL, et al. The St. George’s Respiratory Questionnaire in lymphangioleiomyomatosis. *Chin Med Sci J*. 2010;25(3):140-145.

10. Jones PW. Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *Eur Respir J*. 2002;19(3):398-404.
11. Spencer S, Calverley PM, Sherwood Burge P, Jones PW; ISOLDE Study Group. Inhaled Steroids in Obstructive Lung Disease. Health status deterioration in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2001;163(1):122-128.
12. Swigris JJ, Brown KK, Behr J, et al. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med*. 2010;104(2):296-304.
13. McCormack FX, Inoue Y, Moss J, et al; National Institutes of Health Rare Lung Diseases Consortium; MILES Trial Group. Efficacy and safety of sirolimus in lymphangioleiomyomatosis. *N Engl J Med*. 2011;364(17):1595-1606.
14. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis*. 1992;145(6):1321-1327.
15. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166(1):111-117.
16. Young LR, Vandyke R, Gulleman PM, et al. Serum vascular endothelial growth factor-D prospectively distinguishes lymphangioleiomyomatosis from other diseases. *Chest*. 2010;138(3):674-681.
17. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res*. 2000;9(8):887-900.
18. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 4th ed. New York, NY: Oxford University Press; 2008.
19. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res*. 1992;1(1):73-75.
20. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care*. 2000;38(suppl 9):II14-II25.
21. Food and Drug Administration Center for Drug Evaluation and Research. *Guidance for Industry: qualification process for drug development tools*. Silver Spring, MD: Office of Communications, Division of Drug Information, Center for Drug Evaluation and Research, Food and Drug Administration; 2010.
22. de Torres JP, Casanova C, Hernández C, et al. Gender associated differences in determinants of quality of life in patients with COPD: a case series study. *Health Qual Life Outcomes*. 2006;4:72.
23. Fairclough D. *Design and Analysis of Quality of Life Studies in Clinical Trials*. 2nd ed. New York, NY: CRC Press; 2010.