Performance of forced vital capacity and lung diffusion cut-points for associated radiographic interstitial lung disease in systemic sclerosis

Kimberly Showalter, MD[a], **Aileen Hoffmann, MS**[b], **Gerald Rouleau, MS**[c], **David Aaby, MS**[c], **Jungwha Lee, PhD MPH**[c,d], **Carrie Richardson, MD**[a], **Jane Dematte, MD MBA**[e], **Rishi Agrawal, MD**[f], **Rowland W. Chang, MD MPH**[b,c,d], and **Monique Hinchcliff, MD MS**[b,c,d]

[a]Northwestern University Feinberg School of Medicine, Department of Medicine, 240 E. Huron Street, Suite 1-200, Chicago, IL 60611

[b]Northwestern University Feinberg School of Medicine, Department of Medicine, Division of Rheumatology, 240 E. Huron Street, Suite M-300, Chicago, IL 60611

[c]Northwestern University Feinberg School of Medicine, Department of Preventive Medicine, 680 N. Lake Shore Drive, Suite 1400, Chicago, IL 60611

[d]Northwestern University Feinberg School of Medicine, Institute for Public Health and Medicine, 633 N St. Clair, 18th Floor, Chicago, IL 60611

[e]Northwestern University Feinberg School of Medicine, Department of Medicine, Division of Pulmonary and Critical Care Medicine, 240 E Huron, McGaw M-300

[f]Northwestern University Feinberg School of Medicine, Department of Radiology, 251 E. Huron Street, 4th Floor, Chicago, IL 60611

Abstract

Objective—Forced vital capacity (FVC) and carbon monoxide diffusion (DLCO) are used for systemic sclerosis-associated interstitial lung disease (SSc-ILD) screening. The study purpose was to determine the sensitivity, specificity, and negative predictive value (NPV) (proportion of true negative screening tests) of FVC and DLCO thresholds for SSc-ILD on chest high-resolution computed tomography (HRCT) scans.

Methods—Patients fulfilling American College of Rheumatology 2013 SSc criteria with a chest HRCT scan and pulmonary function tests (PFT) were studied. A thoracic radiologist quantified radiographic ILD. Optimal FVC and DLCO % predicted thresholds for ILD were identified using receiver operating characteristic curves. The FVC and DLCO combinations with greatest sensitivity and specificity were also determined. Sub-analysis was performed in patients with positive Scl-70 autoantibodies.

Results—265 patients were studied. Of 188 (71%) with radiographic ILD, 59 out of 188 (31%) had "normal" FVC (80% predicted), and 65 out of 151 (43%) had "normal" DLCO (60%

Corresponding Author: Monique Hinchcliff, MD, MS, Associate Professor of Medicine, Northwestern University Feinberg School of Medicine, Department of Medicine, Division of Rheumatology, McGaw Pavilion, 240 E Huron Street, Chicago, IL 60611, Phone: 312050308003, Fax: 312050300994, m-hinchcliff@northwestern.edu.

predicted). FVC <80% (sensitivity 0.69, specificity 0.74), and DLCO <62% (sensitivity 0.60, specificity 0.70) were optimal thresholds for radiographic SSc-ILD. All FVC and DLCO threshold combinations evaluated had NPV <0.70. The NPV for radiographic ILD for FVC <80% was lower in patients with positive Scl-70 autoantibody (NPV=0.05) compared to negative Scl-70 autoantibody (NPV=0.57).

Conclusions—Radiographic ILD is prevalent in SSc despite "normal" PFTs. No % predicted FVC or DLCO threshold combinations yielded high NPV for SSc-ILD screening. "Normal" FVC and DLCO in SSc patients, especially those with positive Scl-70 autoantibodies, should not obviate consideration of HRCT for ILD evaluation.

Introduction

Interstitial lung disease (ILD) is a leading cause of death in patients with systemic sclerosis (SSc) [1], and chest high-resolution computed tomography (HRCT) is the gold-standard diagnostic test [2]. Pulmonary function tests (PFTs) including forced vital capacity (FVC) and carbon monoxide diffusion (DLCO) % predicted are frequently used to screen for ILD to avoid HRCT-associated radiation. However, recent evidence shows that a significant number of SSc patients with FVC values within the normal range (80 % predicted) demonstrate radiographic ILD [3, 4]. We hypothesized that optimal individual FVC and DLCO % predicted thresholds, as well as optimal FVC and DLCO combinations, could be identified with acceptably high negative predictive value (NPV) for associated radiographic SSc-ILD. The prevalence of a disease in a population affects the NPV, and radiographic ILD is common in SSc patients [1, 3–5]. One HRCT study of 215 consecutive SSc patients found limited ILD (<10% of lung involvement on HRCT) in almost 50% of patients [1, 5]. However, Suliman et al. evaluated 102 consecutive patients with SSc at the University Hospital Zurich, and found that 40 out of 75 patients (53%) with an FVC >80% predicted had significant radiographic ILD (>20% lung involvement according to Goh et al. [3, 5]). Using data from SSc patients at our center, our aims were to (1) identify optimal individual FVC and DCLO % predicted cut-points for associated ILD on HRCT, and (2) to determine the sensitivity, specificity and negative predictive value (NPV) of varying individual and combined FVC and DLCO % predicted thresholds for associated ILD on HRCT.

Patients and Methods

The present study was approved by the Northwestern University Institutional Review Board (STU00066807). A waiver of informed consent was obtained for this retrospective study as all patients were consented for the Northwestern Scleroderma Patient Registry and permission to review electronic health records is included in that consent (STU00002669). Patients had sine, diffuse or limited cutaneous SSc, fulfilled American College of Rheumatology 2013 SSc criteria, and had a PFT within 12 months of a chest HRCT scan [6]. Patients who had pulmonary procedures or diagnoses that could potentially affect PFT results including lung transplant, lobectomy, or lung malignancy were excluded. Patients with comorbid PH-ILD were included in the analyses.

Clinical data (SSc subtype, smoking history (current, previous, or never), alcohol use (present or absent) and modified Rodnan skin score (mRSS)] were collected from

rheumatology clinic notes within one year of HRCT date. Systemic sclerosis disease duration was defined as duration between first non-Raynaud SSc symptom and HRCT scan date. Mean ± standard deviation (SD) or n (%) was reported for baseline characteristics.

Chest HRCT is defined by the use of images at 1-2 mm slice thickness reconstructed using specialized computer algorithms to increase the sharpness of lung parenchyma. Ground glass opacities (GGO) and fine reticulation on HRCT are characteristic of SSc-ILD [7]. Although lung biopsies are seldom performed for diagnosis and survival is independent of histology, non-specific interstitial pneumonia (NSIP) followed by usual interstitial pneumonia (UIP) are the most common histological patterns in patients with SSc [1]. An experienced thoracic radiologist (RA), blinded to clinical data, quantified ILD on HRCTs according to the method published by Kazerooni et al. and manually reviewed each HRCT to confirm lung changes consistent with ILD as described [8]. Briefly, each lobe was scored for GGO and fibrosis involvement using a modified Likert scale (0 = no disease; 1 = <5% involvement of the lobe, 2 = 6-25% involvement; 3 = 26-50% involvement; 4 = 51-75% involvement, and 5 = 76-100%). The total lung score was the sum of the GGO and fibrosis scores for each lobe. Interstitial lung disease was defined as present if the morphologic features of the disease matched a recognized pattern of disease (i.e., NSIP, UIP, etc.).

Pulmonary function test data obtained at our center or at an outside institution were analyzed. All available FVC values were studied and correlation with TLC was determined using Pearson's correlation coefficient. However, to optimize DLCO data quality, only PFTs with inspiratory vital capacity (IVC) and FVC within 0.85 of each other were included [9]. Patients with poor quality DLCO data as indicated by an IVC:FVC ratio less than 0.85 were excluded from the analyses of DLCO. However, all FVC data were analyzed. Percent predicted DLCO was corrected for hemoglobin values obtained from a complete blood count performed within 6 months of the PFT date. National Health and Nutrition Examination Survey III reference values were used [10].

Two independent receiver operating characteristic (ROC) curves were generated for radiographic SSc-ILD at varying FVC and DLCO % predicted values. For all analyses, R (version 3.4.0) ROCR package was used to determine the optimal cut-points, defined as the greatest combined sensitivity and specificity, for radiographic SSc-ILD. Various established PFT thresholds were evaluated. Sub-analysis was performed according to Scl-70 autoantibody profile, because SSc-ILD is highly prevalent in patients with anti-topoisomerase I (Scl-70) autoantibody positivity [1].

Results

Of 729 patients enrolled in the Northwestern Scleroderma Program, 404 patients had interpretable HRCT images for manual review. One hundred thirty-nine patients were excluded: 30 with an overlap autoimmune diagnosis or a confounding pulmonary procedure, and 109 with PFT greater than one year from HRCT date. The remaining 265 studies were analyzed. The DLCO values from 29 PFTs were excluded because the IVC to FVC ratio was 0.85 thereby rendering the DLCO potentially inaccurate [9]. Of the remaining 236 patients,

214 patients had hemoglobin assessed within six months of PFT for DLCO adjustment (Supplemental Figure 1) and were included for DLCO and combined threshold analyses.

The majority of patients were women (81.5%) and had diffuse cutaneous SSc (49%) (Table 1). Anti-topoisomerase I serum autoantibodies were present in 78 (30%). The mean SSc disease duration (time between first non-Raynaud SSc symptom to HRCT scan) was 6.4 (range 0-42) years with a median of 2 years. Radiographic ILD was present in 188 subjects (71%). There was a strong linear correlation between FVC and TLC (r=0.85), thus FVC was used in the remainder of the analyses. Fifty-nine out of 188 (31%) had normal FVC (80 % predicted). A total of 151 patients had available and accurate DLCO measurement (IVC:FVC within 85% of each other) and radiographic ILD, and 65 out of these 151 (43%) had normal DLCO (60 % predicted). Thirty-one out of 214 (14%) of these patients had both FVC and DLCO % predicted within normal range and radiographic ILD. To ensure that patients with poor PFTs were not systemically excluded from DLCO analysis, we compared the FVC between the group included vs. excluded for low IVC on the DLCO maneuver. Those with DLCO excluded due to low IVC:FVC ratio did not have a significantly lower FVC than those with acceptable DLCO maneuvers.

For associated ILD, the data-derived optimal FVC and DLCO cut-points were 80% predicted (sensitivity 0.69, specificity 0.73) for FVC, and 62% predicted (sensitivity 0.60, specificity 0.70) for DLCO (Figure 1, Table 2). The combination of DLCO <62% or FVC <80% predicted had a sensitivity and specificity for ILD of 0.80 and 0.56, respectively (Table 2). The sensitivity and specificity for various combinations of traditional and data-derived optimal FVC and DLCO cut-points for ILD are reported in Table 2. Using an ILD screening algorithm of meeting either FVC <80% or DLCO <62% predicted, 82 persons in our cohort (31%) would screen negative for ILD (NPV=0.53). A more liberal algorithm of either FVC <80% or DLCO <70% predicted would inappropriately classify 58 persons (22%) (NPV= 0.57), while an FVC <80% or DLCO <80% predicted would inappropriately classify 32 persons (12%) as lacking ILD (NPV=0.65) (Table 2).

Sub-analysis was performed based upon Scl-70 autoantibody status. The NPV for radiographic ILD for FVC <80% was lower in patients with positive Scl-70 autoantibody (NPV=0.05) compared to those negative (NPV=0.57) (Table 2).

Discussion

The study aim was to determine the sensitivity, specificity, and NPV of varying independent and combined FVC and DCLO % predicted thresholds for associated radiographic ILD in SSc. The clinical relevance of new FVC and DLCO threshold values is to inform the development of more rationale and data-driven ILD screening protocols that include HRCT in the most at-risk patients. Our ROC curve results show FVC=80% and DLCO=62% predicted are the optimal (maximized combined sensitivity and specificity) thresholds for SSc-ILD. Similar area under the curve (AUC) for FVC (AUC=0.74) and DLCO (AUC=0.71) indicate that FVC and DLCO are both "moderately accurate" tests for radiographic ILD in SSc patients, with 0.70-0.90 considered "moderately accurate", and >0.90 considered

"highly accurate" [11]. We found no individual or combined FVC and DLCO % predicted algorithm that had high NPV for SSc-ILD screening.

Our study population was predominantly white women reflecting SSc demographics and the patient population who receive care at our center, and thus our results may not be generalizable to men and racially diverse SSc populations. Only 407 out of 729 registry patients had available HRCT imaging and were included in the analysis. This likely enriched our population for patients with pulmonary disease. However, the 71% prevalence of radiographic ILD in our cohort is comparable to reported prevalence rates in other tertiary care cohorts (36-84%) [3–5]. The 30% prevalence of Scl-70 autoantibody positivity in our study cohort is slightly higher than previous reports (20%) [1]. Yet, ILD is known to be more common in SSc patients with Scl-70 antibodies rendering HRCT exam referral in this group more likely.

Individual and combined, traditional and data-derived, FVC and DLCO % predicted thresholds had relatively low sensitivity, specificity, and NPV values for associated radiographic ILD. Our data-derived optimal cut-points for radiographic ILD detection were identical (FVC 80% predicted) or similar to (DLCO 62 vs. 60% predicted), and did not out-perform traditional thresholds of "normal" that are based upon 95% confidence intervals from the NHANES III reference population. We examined alternative FVC and DLCO thresholds of 65%, 70% and 80% predicted and found low NPV for ILD.

Similar to prior study results, "normal" PFTs did not discriminate between SSc patients with and without radiographic SSc-ILD. Of our patients with radiographic ILD, defined using the quantitative Kazerooni method, 31% had "normal" FVC (80 % predicted), 43% had "normal" DLCO (60 % predicted), and 14% had both FVC 80 and DLCO 60 % predicted. Suliman et al. studied 102 SSc patients and defined radiographic ILD using the more simplistic Goh method that classifies ILD as mild (<20%), intermediate, or severe (>20%) [3, 5]. They found that 24 out of 102 patients (23%) had normal FVC (80% predicted) and radiographic ILD and 40 out of 75 (53%) patients with FVC 80% predicted had severe radiographic ILD [3]. Steele et al. evaluated various SSc-ILD screening algorithms and defined ILD as present if 1) crackles were noted on physical exam, 2) x-ray showed interstitial markings consistent with fibrosis, or 3) PFT parameters were met: FVC <70% and FEV1/FVC >70%; or FVC <80% and FEV1/FVC >70%. Both FVC <70 and <80 % predicted thresholds had low sensitivity for SSc-ILD (54% and 61%, respectively) [4].

At traditional thresholds (FVC <80% and DLCO <60%), we found FVC compared to DLCO to be more specific for associated SSc-ILD (72% vs. 69%). This is most likely due to SSc vascular disease, including concurrent pulmonary hypertension (PH), that also results in reduced DLCO [12]. In a multivariable analysis by Nihtyanova et al., low DLCO was considered a significant predictor variable in models for both PH and pulmonary fibrosis. In contrast, low FVC was only a predictor for pulmonary fibrosis and not PH [13]. Importantly, study results have shown that FVC is superior to DLCO for determination of risk of SSc-ILD related mortality [13]. However, our ROC curve results show that these tests perform similarly for ILD detection.

The association between SSc-ILD and positive Scl-70 autoantibodies is well recognized. The NPV for radiographic ILD of 5% (+Scl-70) vs. 57% (−Scl-70) for FVC <80 % predicted and of 10% (+Scl-70) vs. 48% (−Scl-70) for DLCO <60 % predicted in patients, demonstrates how crucial it is to maintain a high index of suspicion for ILD in patients with +Scl-70 antibodies.

Study limitations include analysis of data from participants recruited at one site that specializes in SSc care that limits generalizability. Also, DLCO was not reported for every patient and the absence of CBC within 6 months prevented correction for hemoglobin leading to exclusion of that patient from analyses. ILD was dichotomized as presence or absent that prevents determination of the impact of ILD severity varying PFT thresholds. Finally, we excluded patients with lung malignancy, transplant and lung resection and thus limited our evaluation of these patients who may have severe ILD. Study strengths include our large sample size (265 vs. 102 analyzed in Suliman et al. cohort), and our use of stringent criteria to identify high quality PFT (inclusion of DLCO with hemoglobin performed within 6 months and exclusion of DLCO values from PFTs with IVC: FVC ratio <0.85), to ensure DLCO quality [9]. The ILD scoring system also differed between studies as described above. We also add to the literature by examining a variety of cut-points to determine if new PFT thresholds exist that could be used to screen for SSc-ILD. Our study is limited by being a single center specializing in SSc care that may limit generalizability. As a cross-sectional analysis, we are also unable to report longitudinal change in PFTs. We also dichotomize ILD to presence or absent, therefore do not further elaborate on severity of ILD with regard to the varying PFT thresholds.

In conclusion, we show that significant radiographic ILD is present in SSc patients, especially in those with positive Scl-70 autoantibodies, despite a "normal" PFT. We conclude that clinicians should maintain a high index of suspicion for SSc-ILD despite normal PFT results, and should consider performing a screening chest HRCT exam. However, we recognize that the amount of radiation from one chest HRCT (~5-8 mSv) can exceed the equivalent dose of over 100 chest radiographs (~0.05 mSv) [14]. Additional studies that examine longitudinal change in PFT over time and the presence of radiographic ILD may help to shed light upon how clinicians may best monitor for SSc-ILD while avoiding unnecessary radiation exposure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Not applicable.

References

1. Wells AU. Interstitial lung disease in systemic sclerosis. Presse Med. 2014; 43:329–343.

2. Meyer K. Diagnosis and management of interstital lung disease. Transl Respir Med. 2014(2):4.

3. Suliman YA, Dobrota R, Huscher D, Nguyen-Kim TD, Maurer B, Jordan S, et al. Brief Report: Pulmonary Function Tests: High Rate of False-Negative Results in the Early Detection and Screening of Scleroderma-Related Interstitial Lung Disease. Arthritis Rheum. 2015; 67:3256–3261.

4. Steele R, Hudson M, Lo E, Baron M, Canadian Scleroderma Research G. Clinical decision rule to predict the presence of interstitial lung disease in systemic sclerosis. Arthritis Care Res. 2012; 64:519–524.

5. Goh NS, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. Am J Respir Crit Care Med. 2008; 177:1248–1254. [PubMed: 18369202]

6. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum. 2013; 65:2737–2747. [PubMed: 24122180]

7. Lammi MR, Baughman RP, Birring SS, Russell AM, Ryu JH, Scholand M, Distler O, LeSage D, Sarver C, Antoniou K, et al. Outcome Measures for Clinical Trials in Interstitial Lung Diseases. Curr Respir Med Rev. 2015; 11:163–174. [PubMed: 27019654]

8. Kazerooni EA, Martinez FJ, Flint A, Jamadar DA, Gross BH, Spizarny DL, et al. Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. AJR Am J Roentgenol. 1997; 169:977–983. [PubMed: 9308447]

9. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J. 2005; 26:720–735. [PubMed: 16204605]

10. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med. 1999; 159:179–187. [PubMed: 9872837]

11. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000; 45:23–41. [PubMed: 10802332]

12. Steen V. Predictors of end stage lung disease in systemic sclerosis. Ann Rheum Dis. 2003; 62:97–99. [PubMed: 12525376]

13. Nihtyanova SI, Schreiber BE, Ong VH, Rosenberg D, Moinzadeh P, Coghlan JG, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. Arthritis Rheumatol. 2014; 66:1625–1635. [PubMed: 24591477]

14. Kalra MK, Maher MM, Rizzo S, Kanarek D, Shepard JA. Radiation exposure from chest CT: issues and strategies. J Korean Med Sci. 2004; 19:159–166. [PubMed: 15082885]
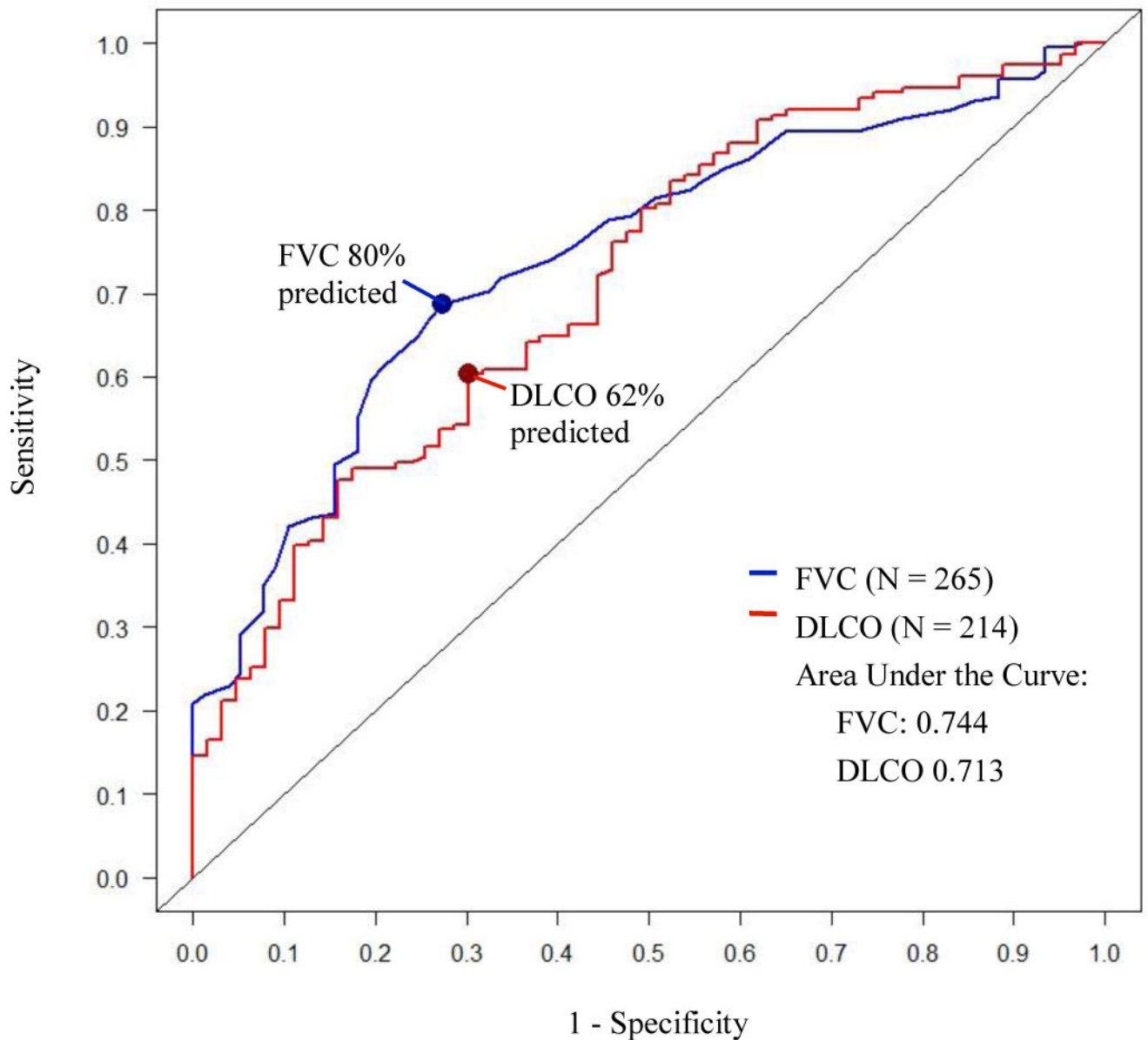
Figure 1. Receiver Operating Characteristic Curves
Receiver operating characteristic (ROC) curves for % predicted forced vital capacity (FVC) and diffusing capacity of the lung for carbon monoxide (DLCO) demonstrating the performance of varying FVC and DLCO % predicted cut-points for associated radiographic interstitial lung disease (ILD) in systemic sclerosis.

Table 1

Baseline study population characteristics (N = 265).

Characteristic	Mean ± SD or n (%)
Age at time of high-resolution computed tomography (HRCT), years	50 ± 12
Sex, woman	216 (82)
Body mass index, kg/m^2	29 ± 6
Ethnicity, white	192 (73)
Smoker, current or former	97 (37)
Alcohol, current	96 (46)
Proton pump inhibitor use, current	171 (65)
SSc disease subtype	
Diffuse cutaneous	129 (49)
Limited cutaneous	131 (49)
Sine scleroderma	5 (2)
Years since first non-Raynaud SSc symptom to HRCT date	6 ± 11
Systemic sclerosis specific antibodies, positive	
Anti-topoisomerase I (Scl-70)	78 (30)
Anticentromere (ACA)	48 (18)
Anti RNA polymerase III	59 (23)
Modified Rodnan skin score	13 (11)
Radiographic interstitial lung disease present	188 (71)
Forced vital capacity (FVC) % predicted	76 (18)
Diffusing capacity for carbon monoxide (DLCO) % predicted	60 (20)

Author Manuscript

Table 2

Performance of pulmonary function test thresholds for prevalent radiographic interstitial lung disease on chest high-resolution computed tomography images in patients with systemic sclerosis

	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Entire cohort				
FVC % predicted (n=265)				
<80 (conventional and optimal)	0.69	0.73	0.86	0.49
DLCO % predicted (n=214)				
<60 (conventional)	0.58	0.70	0.82	0.41
<62 (optimal)	0.60	0.70	0.83	0.42
<70 (alternative)	0.80	0.51	0.80	0.52
<80 (alternative)	0.92	0.32	0.77	0.63
Combination of PFT thresholds % predicted (n=214)				
FVC <80 **and** DLCO <60	0.46	0.81	0.85	0.38
FVC <80 **or** DLCO <60	0.79	0.57	0.82	0.53
FVC <80 **and** DLCO <62	0.49	0.81	0.86	0.40
FVC <80 **or** DLCO <62	0.80	0.56	0.81	0.53
FVC <80 **and** DLCO <65	0.53	0.78	0.85	0.41
FVC <80 **or** DLCO <65	0.82	0.46	0.78	0.52
FVC <80 **and** DLCO <70	0.61	0.76	0.86	0.45
FVC <80 **or** DLCO <70	0.87	0.43	0.78	0.57
FVC <80 **and** DLCO <80	0.66	0.73	0.85	0.47
FVC <80 **or** DLCO <80	0.94	0.27	0.76	0.65
Autoantibody Scl-70 positive patients				
FVC <80 (n=78)	0.75	0.05	0.98	0.05
DLCO <60 (n=62)	0.70	1.00	1.00	0.10
FVC <80 **or** DLCO <60 (n=62)	0.85	0.50	0.98	0.10
Autoantibody Scl-70 negative patients				
FVC <80 (n=183)	0.63	0.73	0.78	0.57
DLCO <60 (n=150)	0.50	0.68	0.70	0.48
FVC <80 **or** DLCO <60 (n=150)	0.74	0.57	0.72	0.60

PFT = pulmonary function test; FVC = forced vital capacity; DLCO = diffusing capacity of the lung for carbon monoxide. FVC 80% and DLCO 60% predicted represent traditional normal cut points based on 95% confidence interval in healthy population.