ORIGINAL RESEARCH

Prediction of progressive fibrosing interstitial lung disease in patients with systemic sclerosis: insight from the CRDC cohort study

Min Hui, Xinwang Duan, Jiaxin Zhou, Mengtao Li, Qian Wang, Jiuliang Zhao, Yong Hou, Dong Xu, Xiaofeng Zeng

ABSTRACT

Background This study aims to establish a reliable prediction model of progressive fibrosing interstitial lung disease (PF-ILD) in patients with systemic sclerosis (SSc)-ILD, to achieve early risk stratification and to help better in preventing disease progression.

Methods 304 SSc-ILD patients with no less than three pulmonary function tests within 6–24 months were included. We collected data at baseline and compared differences between SSC patients with and without PF-ILD. Least absolute shrinkage and selection operator (LASSO) regression and multivariable Cox regression were used to construct the prediction model, which were presented as nomogram and forest plot.

Results Among the 304 patients with SSc-ILD included, 92.1% were women, with a baseline average age of 46.7 years. Based on the 28 variables preselected by comparison between SSc patients without PF-ILD group (n=150) and patients with SSc PF-ILD group (n=154), a 9-variable prediction model was constructed, including age≥50 years (HR 1.8221, p<0.001), hyperlipidemia (HR 4.0516, p<0.001), smoking history (HR 3.8130, p<0.001), positive anti-Scl-70 antibodies or mycophenolate mofetil (HR 2.0487, p<0.001), decreased serum immunoglobulin A level (HR 2.3900, p=0.012), positive anti-Scl-70 antibody (HR 1.9753, p<0.001), arthritis (HR 2.0088, p<0.001), shortness of breath (HR 2.0467, p<0.001), decreased serum immunoglobulin A level (HR 1.8221, p<0.001), smoking history (HR 4.0516, p<0.001), and age≥50 years (HR 1.8221, p<0.001). The concordance index after enhanced bootstrap resampling adjustment was 0.874, while the optimism-corrected Brier score was 0.144 in internal validation.

Conclusion This study developed the first prediction model for PF-ILD in patients with SSc-ILD, and internal validation showed favourable accuracy and stability of the model.

INTRODUCTION

Interstial lung disease (ILD) has heterogeneous phenotypes stemming from diffused pulmonary parenchymal inflammation and fibrosis. Regardless of aetiology, ILDs with a progressive phenotype were collectively referred to as progressive fibrosing ILD (PF-ILD), which was characterised by rapidly aggravating dyspnoea, progressive deteriorating lung function, declining physical functional capacity, worsening health-related quality of life, adverse therapeutic response and often early mortality. PF-ILD accounting for 20%–30% in an ILD population without idiopathic pulmonary fibrosis (IPF). The PROGRESS Study in France showed that the prognosis of patients with PF-ILD was extremely poor, with a median survival time of only 3.7 years after rapid progression. However, early intervention of PF-ILD was often neglected in current clinical practice by waiting for lung function decline and/or extensive ILD involvement over preceding years. Therefore, early recognition and novel treatment concepts are required to prevent disease progression and avoid irreversible organ damages in patients with PF-ILD. The clinical features of patients with systemic...
sclerosis (SSc) who developed PF-ILD were still not fully elucidated, which may be limited by their small sample size, insufficient follow-up, patient selection biases and statistical instabilities.6 5 Although several retrospective studies have reported that male, old age, low baseline forced vital capacity (FVC) and carbon monoxide transfer factor and some imaging features (pathologically confirmed usual interstitial pneumonia (UIP) or typical features for UIP) were risk factors of PF-ILD,6 7 it still remained an unmet need to construct evidence-based prediction algorithms with significant clinical applicability to identify patients with SSc at high risk of PF-ILD. In this study, we first aimed to assess the disease course and patterns of SSc-ILD progression over a 5-year period and to further generated a prediction model of PF-ILD in patients with SSc-ILD to improve early recognition of this life-threatening disease.

MATERIALS AND METHODS

Study participants
The study cohort consisted of 3028 patients with SSc who prospectively registered in Chinese Rheumatism Data Center (CRDC) database from January 2008 to January 2022 by 127 participating centres nationwide. All patients (age ≥18 years) were diagnosed with SSc according to 2013 ACR/EULAR SSc classification criteria.8 Among those patients with SSc, 1411 were determined with presence of ILD by a rheumatologist and a respiratory physician based on symptoms, pulmonary function tests and ground glass opacification or scleroderma-related fibrosis on high-resolution CT (HRCT). Exclusion criteria were as follows: (1) overlapped with other connective tissue diseases; (2) pulmonary diseases caused by other aetiologies, such as ILD induced by drug, toxicant and occupational environmental exposure, chronic obstructive pulmonary disease, heart failure, acute pulmonary infection, malignancies and pulmonary embolism; and (3) combined with severe pulmonary hypertension, which was defined as pulmonary artery systolic pressure>60 mm Hg on echocardiography, mean pulmonary artery pressure (mPAP) ≥35 mm Hg or mPAP≥25 mm Hg with elevated right atrial pressure and/or cardiac index<2L/(min×m2) in right heart catheterisation. The study design was presented in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement,9 with the integral study process shown in figure 1.

Data collection
Medical records of patients with SSc were independently identified and reviewed by two CRDC investigators at each centre who were blinded to each other. The initial visit with record on SSc-ILD diagnosis was determined as baseline. Data were collected at baseline as follows: (1) overall information: demographic features (age, sex and body mass index), disease duration (recognised as the period from the first non-Raynaud phenomenon symptoms and baseline) and SSc subtypes according to the LeRoy and Medsger criteria 10; (2) medical history: comorbidities (eg, hypertension, diabetes mellitus,
hyperlipidemia with total cholesterol values ≥5.70 mmol/L and/or triglyceride values ≥1.70 mmol/L, as well as malignancy), family history of rheumatic diseases and smoking history (defined as smoking or used to smoke at least 0.1 pack years); (3) clinical manifestations: skin and microvascular lesions (e.g., Raynaud’s phenomenon, skin fibrosis with modified Rodnan skin score (mRSS) ≥1 and digital ulcers), musculoskeletal involvements (arthritis and myositis) and other visceral impairments (such as pulmonary arterial hypertension); (4) assessment tools for SSc: physical global assessment, mRSS and cardiopulmonary function assessment (verified by six-minute walk distance test, NYHA cardiac function classification and Borg index); (5) laboratory examinations: autoantibodies (e.g., antinuclear antibody, anti-Scl-70 antibody and antitcentromere antibody), complete blood cell count, inflammatory markers (C reactive protein and erythrocyte sedimentation rate), etc; and (6) treatments: glucocorticoids, immunosuppressant and antifibrotic agents.

Clinical outcomes

Registered patients were comprehensively followed up as planned every 3–12 months and 304 of them were further screened with fibrotic lung diseases on HRCT and no less than three available records of FVC at baseline and within the next 6–24 months. To reflect pattern of overall course, absolute changes in FVC% predicted were appraised annually (12±3 months) over a mean follow-up of 5 years, and patients were divided into five subgroups: major decline (annualised FVC decline of >20%); significant decline (annualised FVC decline of >10% and ≤20%); moderate decline (annualised FVC decline of 5%–10%); stable (annualised FVC decline of <5% or improvement of <5%); and improvement (annualised FVC improvement of ≥5%). According to the RELIEF definition by Behr et al,11 an absolute annualised FVC decline ≥5% predicted is proposed to define main outcome measure for progressive fibrosing ILD in this study.

Development of the prediction model

Preliminary screening for potential candidate variables was conducted based on the significant differences in univariable analysis (χ² test or Fisher’s exact test for categorical variables, Student’s t-test or Mann-Whitney U test for continuous variables, as well as univariate logistic regression), reports from previous literatures and expert opinions. 28 variables were preselected in accordance with their accessibility, stability and rationality from baseline data. Continuous candidate variables were converted into binary variables according to the appropriate range. Least absolute shrinkage and selection operator (LASSO) Cox model was further performed to filter out the most predictive variables to avoid overfitting. By increasing the penalty function (regularisation term λ) as constraint item, it can select variables more relevant to outcome and reduce the number of selected variables. Subsequently, the number and the rationality of the predictors included had been reconsidered before a Cox proportional hazards regression model with nine predictors was developed. The final prediction model was presented in the form of both nomogram and forest plot.

Internal validation

The optimal model was verified via internal cross-validation, in which enhanced bootstrap method was applied for repeated sampling (1000 times) to improve the efficiency. The discrimination performance of the model was evaluated by Harrell’s concordance index (C-index). Brier Score, also called cost function, is referred to as a set of calibration measures of prediction probability, which was used to assess the calibration performance of the model. The calibration curve was analysed by plotting the predicted nomogram and the actual probability of outcomes in patients. Better calibration of prediction model was concluded from lower Brier Score and calibration curve closer to diagonal line. The decision curve analysis (DCA) was then used to evaluate the clinical benefit of the prediction model. When the risk stratification determined by ROC curve is used as the threshold probability, the corresponding clinical net benefit value can be found in the DCA diagram to verify the clinical significance of the stratification threshold.

Risk stratification

The 5-year cumulative risks of SSc PF-ILD for an individual patient would be calculated using the following formula:

\[ P_{\text{at 5 years}} = 1 - S_0(t)^{\text{exp (prognostic index)}} \]

where \( S_0(t) \) is the 5-year average survival probabilities, and the prognostic index equals the sum of the products of the predictive factors and corresponding coefficients. Three risk groups were stratified with low, medium and high risk of SSc PF-ILD based on the cut-off thresholds at the 50th and 75th percentiles to develop clinician-friendly and patient-friendly stratification strategy.

Statistical analysis

Continuous variables were presented as the mean±SD or median (extreme ranges) for non-normal distribution, while categorical variables were reported as number (n) and frequency with percentage (%). Relevant differences were compared using a Student’s t-test (or Mann-Whitney test, as appropriate) for continuous variables and Pearson’s χ² test or Fisher’s exact test for categorical variables. Univariable and multivariable Cox regression analyses were proceeded to identify the predictive value of baseline variables. Statistical analyses, including data preprocessing (outlier deletion and missing value imputation), candidate variable screening, model establishment and calibration, model diagnosis, model presentation and risk probability calculation, were computed using R statistical software (V.4.1.2). A two-sided p value <0.05 was considered statistically significant difference.
RESULTS
Disease course and ILD patterns in patients with SSc-ILD
As demonstrated in online supplemental figure 1, the performances of progression rarely appeared in continuous 12-month periods. Most patients experienced at least one 12-month period of stable or improving FVC, though different severity of FVC decline (significant, major, or moderate) was highly probable to be inevitable. Most patients with SSc-ILD had a slow pattern of lung function decline with up to 52%–80% of the patients behaved as stable in each year over the 5-year period.

Clinical features and differential analysis of the patients
The 304 studied patients with SSc-ILD were overwhelmingly women (92.1%), with a baseline average age of 46.7±11.7 years, and 45.7% were classified as diffused cutaneous (dc) SSc (dcSSc) subtype. The median disease duration from first non-Raynaud SSc symptom to baseline was 41 months. The HRCT pattern categories included 132 cases of fibrotic non-specific interstitial pneumonitis, 108 cases of UIP, 25 cases of centrolobular fibrosis, 7 cases of chronic hypersensitivity pneumonitis and 14 cases of uncertain patterns. 154 patients were further identified as having SSc PF-ILD during follow-up. They were thus divided into SSc PF-ILD group (n=154) and SSc patients without PF-ILD group (n=150), and baseline data was afterwards compared between the two groups (table 1). Age≥50 years (51.3% vs 40.0%, HR 3.694, p<0.001), male sex (9.7% vs 6.0%, HR 1.993, p=0.012), combined hyperlipidemia (31.2% vs 8.7%, HR 3.564, p<0.001) and smoking history (18.2% vs 5.3%, HR 6.202, p<0.001) were significantly more prevalent in patients with SSc PF-ILD. Compared with SSc patients without PF-ILD, patients with SSc PF-ILD were more inclined to be diffused subtype (58.4% vs 32.7%, HR 2.363, p<0.001). Besides, patients with SSc PF-ILD were notably more complicated with arthritis (83.8% vs 24.7%, HR 3.868, p<0.001) and shortness of breath (84.4% vs 80.0%, HR 3.375, p<0.001). Anti-Scl-70 antibody positivity was higher in patients with SSc PF-ILD (53.2% vs 40.0%, HR 5.997, p<0.001) compared with SSc patients without PF-ILD. Decreased serum immunoglobulin A (IgA) level accounted for a significantly greater proportion in patients with SSc PF-ILD than SSc patients without PF-ILD (11.0% vs 5.3%, HR 2.665, p<0.001). Patients with SSc PF-ILD received markedly less cyclophosphamide (CYC) and/or mycophenolate mofetil (MMF) than SSc patients without PF-ILD (58.4% vs 68.7%, HR 0.168, p<0.001).

Development of prediction model
A total of 28 candidate variables were selected for further filtration, as is shown in online supplemental figure 2A,B. A combination of nine variables was identified in line with λ between kmin and λse. The final model consisted of the following significant indicators: age≥50 years (HR 1.8221, p=0.001), hyperlipidemia (HR 4.0516, p<0.001), smoking history (HR 3.8130, p<0.001), dcSSc subtype (HR 1.9753, p<0.001), arthritis (HR 2.0008, p<0.001), shortness of breath (HR 2.0487, p=0.012), decreased serum IgA level (HR 2.3900, p=0.002), positive anti-Scl-70 antibody (HR 1.9573, p=0.016) and usage of CYC/MMF (HR 0.4267, p<0.001). The 5-year cumulative risk of SSc PF-ILD for an individual patient with SSc-ILD thus can be calculated by the following formula:

\[
P_{\text{at} 5 \text{ years}} = 1 - 0.9863991^{\exp (\text{prognostic index})}
\]

The prognosis index=0.6000×age≥50 years+1.3991×hyperlipidemia+1.3384×smoking history+0.6807×dcSSc subtype+0.6936×arthritis+0.7172×shortness of breath+0.8713×decrease serum IgA level+0.6716×positive anti-Scl-70 antibody+0.8517×usage of CYC/MMF.

Model presentation and internal validation
The algorithm for the fitted model was graphically displayed by nomogram and forest plot as shown in figure 2. Primarily, the original C-index was 0.882 (95% CI 0.824 to 0.940), and the C-index after enhanced bootstrap resampling adjustment was 0.874 (95% CI 0.822 to 0.927), drawn from the ROC curve shown in figure 3. In addition online supplemental figure 3 demonstrated an apparent Brier Score of 0.142 (95% CI 0.108 to 0.176) and an optimism-corrected Brier Score of 0.144 (95% CI 0.114 to 0.174).

Risk stratification and model-based net benefit
The cumulative risk probability calculating equation enabled us to further divide the patients with SSc-ILD into three risk groups: low (<4.6%), moderate (4.6%–20.3%) and high (>20.3%) during the 5-year follow-up, accounting for 152, 76 and 76 individuals, respectively. For 50th and 75th percentiles cut-off thresholds of total points, the calculated sensitivities were 81.8% and 44.2%, respectively. The corresponding specificities were 82.6% and 94.7%. Positive predictive value and negative predictive value were also analysed, which were 82.9% and 81.6% for 50th percentile and 89.5% and 62.3% for 75th percentile. Moreover, the practical curve exhibited in online supplemental figure 4 indicated a positive net benefit for probability thresholds between 1% and 45%. Therefore, prophylactic screening for patients with SSc-ILD in both moderate-risk (4.6–20.3%) and high-risk (>20.3%) groups were recommended to be strengthened to facilitate early detection and intervention to reduce mortality.

DISCUSSION
PF-ILD was referred to as a subset of non-IPF fibrosing ILDs with progressive course even if conventional therapy had been used. As previously estimated, PF-ILD had a real-world prevalence ranging from 18% to 34% in patients without IPF-ILDs. The incidence rate of PF-ILD was approximately 45% in patients with CTD-ILD, while patients with SSc-ILD occupied the highest progression rate of 49%. This study also provides important insights into the disease course and patterns of
ILD progression in patients with SSc-ILD over a term of 5 years. 20.1% of the patients with SSc-ILD experienced various degrees of progression during the initial 12±3-month period, which was consistent with that derived from data in an EUSTAR Study.\textsuperscript{16} Patterns of FVC changes were frequently inconsistent between 12-month period, given that some patients who experienced an overall decline in FVC had periods of improvement, while some patients with overall FVC improvement were merited periods of worsened FVC. A few patients with SSc-ILD were revealed by a slow but cumulative declining of FVC, which might easily be neglected in clinical practice and

Table 1  Baseline demographic and clinical characteristics of the patients with SSc-ILD analysed for PF-ILD risk in univariate Cox proportional hazards regression analysis

<table>
<thead>
<tr>
<th></th>
<th>SSc PF-ILD (n=154)</th>
<th>SSc non-PF-ILD (n=150)</th>
<th>HR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td>48.5±11.9</td>
<td>44.9±11.3</td>
<td>1.023 (1.008 to 1.038)</td>
<td>0.003</td>
</tr>
<tr>
<td>Age≥50 years, n (%)</td>
<td>79 (51.3)</td>
<td>60 (40.0)</td>
<td>3.694 (2.579 to 5.290)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>15 (9.7)</td>
<td>9 (6.0)</td>
<td>1.993 (1.163 to 3.416)</td>
<td>0.012</td>
</tr>
<tr>
<td>Duration of SSc from diagnosis, months</td>
<td>37.9 (0.1, 417.6)</td>
<td>44.7 (0.3, 408.8)</td>
<td>0.999 (0.996 to 1.001)</td>
<td>0.314</td>
</tr>
<tr>
<td>Duration≥60 months, n (%)</td>
<td>39 (25.3)</td>
<td>27 (18.0)</td>
<td>1.007 (0.697 to 1.455)</td>
<td>0.970</td>
</tr>
<tr>
<td>BMI&gt;24 kg/m²</td>
<td>44/151 (29.1)</td>
<td>39/147 (26.5)</td>
<td>1.002 (0.972 to 1.075)</td>
<td>0.390</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>21/151 (13.9)</td>
<td>19/143 (13.3)</td>
<td>1.192 (0.583 to 2.437)</td>
<td>0.631</td>
</tr>
<tr>
<td>Hyperlipidemia, n (%)</td>
<td>48 (31.2)</td>
<td>13 (8.7)</td>
<td>3.564 (2.511 to 5.059)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Malignancy, n (%)</td>
<td>7 (4.5)</td>
<td>3 (2.0)</td>
<td>1.319 (0.617 to 2.818)</td>
<td>0.475</td>
</tr>
<tr>
<td>Family history of rheumatism, n (%)</td>
<td>30 (19.5)</td>
<td>29 (19.3)</td>
<td>1.057 (0.702 to 1.592)</td>
<td>0.790</td>
</tr>
<tr>
<td>Smoking history, n (%)</td>
<td>28 (18.2)</td>
<td>8 (5.3)</td>
<td>6.202 (4.033 to 9.538)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Allergic history, n (%)</td>
<td>13 (8.4)</td>
<td>9 (6.0)</td>
<td>1.509 (0.853 to 2.670)</td>
<td>0.157</td>
</tr>
<tr>
<td>Clinical features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DcSSc subtype, n (%)</td>
<td>90 (58.4)</td>
<td>49 (32.7)</td>
<td>2.363 (1.707 to 3.270)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Proximal skin sclerosis, n (%)</td>
<td>126 (81.8)</td>
<td>112 (74.7)</td>
<td>1.267 (0.840 to 1.911)</td>
<td>0.260</td>
</tr>
<tr>
<td>Digital ulcers, n (%)</td>
<td>44/139 (31.7)</td>
<td>36/130 (27.7)</td>
<td>1.079 (0.753 to 1.546)</td>
<td>0.677</td>
</tr>
<tr>
<td>Arthritis, n (%)</td>
<td>59 (38.3)</td>
<td>37 (24.7)</td>
<td>3.868 (2.716 to 5.510)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PAH, n (%)</td>
<td>23/137 (16.8)</td>
<td>24/123 (19.5)</td>
<td>0.659 (0.414 to 1.049)</td>
<td>0.079</td>
</tr>
<tr>
<td>GORD, n (%)</td>
<td>73 (47.4)</td>
<td>71 (47.3)</td>
<td>1.008 (0.732 to 1.387)</td>
<td>0.961</td>
</tr>
<tr>
<td>6MWD, min</td>
<td>471.2±83.6</td>
<td>480.0±90.4</td>
<td>1.000 (0.997 to 1.003)</td>
<td>0.880</td>
</tr>
<tr>
<td>6MWD&lt;500 m, n (%)</td>
<td>88 (57.1)</td>
<td>81 (54.0)</td>
<td>1.144 (0.830 to 1.577)</td>
<td>0.412</td>
</tr>
<tr>
<td>SOB, n (%)</td>
<td>130 (84.4)</td>
<td>120 (80.0)</td>
<td>3.375 (2.136 to 5.334)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ESR elevation, n (%)</td>
<td>68 (44.2)</td>
<td>61 (40.7)</td>
<td>1.067 (0.772 to 1.475)</td>
<td>0.693</td>
</tr>
<tr>
<td>Decreased serum IgA level, n (%)</td>
<td>17 (11.0)</td>
<td>8 (5.3)</td>
<td>2.665 (1.563 to 4.544)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Autoantibody</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANA, n (%)</td>
<td>142/152 (93.4)</td>
<td>142/149 (95.3)</td>
<td>0.793 (0.414 to 1.517)</td>
<td>0.484</td>
</tr>
<tr>
<td>Anti-Scl-70, n (%)</td>
<td>82 (53.2)</td>
<td>60 (40.0)</td>
<td>5.997 (3.869 to 9.296)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ACA, n (%)</td>
<td>16 (10.4)</td>
<td>22 (14.7)</td>
<td>0.985 (0.585 to 1.657)</td>
<td>0.954</td>
</tr>
<tr>
<td>Anti-RNP, n (%)</td>
<td>37/148 (25.0)</td>
<td>47/147 (32.0)</td>
<td>0.792 (0.543 to 1.154)</td>
<td>0.224</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucocorticoids≥15 mg/day, n (%)</td>
<td>78 (50.6)</td>
<td>81 (54.0)</td>
<td>0.843 (0.613 to 1.159)</td>
<td>0.294</td>
</tr>
<tr>
<td>CYC/MMF, n (%)</td>
<td>90 (58.4)</td>
<td>103 (68.7)</td>
<td>0.168 (0.114 to 0.248)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

ACA, anticientromere antibody; ANA, antinuclear antibody; BMI, body mass index; CYC, cyclophosphamide; dcSSc, diffused cutaneous systemic sclerosis; ESR, erythrocyte sedimentation rate; GORD, gastro-oesophageal reflux disease; IgA, immunoglobulin A; ILD, interstitial lung disease; MMF, mycophenolate mofetil; 6MWD, six-minute walk distance; PAH, pulmonary arterial hypertension; PF-ILD, progressive fibrosing interstitial lung disease; RNP, ribonucleic protein antibody; SOB, shortness of breath.
suffered a pitfall of delayed treatment under current strategy.

To our knowledge, this is the first prediction model that has been used for SSc PF-ILD. Clinicians can predict the 5-year risk of PF-ILD in patients with SSc-ILD by summing up the scores of each predictor and reading from the nomogram, which was feasible and user-friendly. The prediction model created in this study also possessed several strengths over existing predictors. First, a large-scale multicentre longitudinal study cohort evaluated through a multidisciplinary discussion fundamentally ensured the objectivity and reflected real-world practice. In addition, only variables with clinical plausibility, functionality and stability confirmed by expert opinions were on the cards to be finally included and their robustness was internally validated, while those with inter-centre differences, major errors, biases caused by missing data or abnormal distribution and herewith prominent changes over time were excluded. Besides, LASSO regression regularisation algorithm used for variable selection could effectively avoid overfitting. At last, the Cox regression analysis allowed us to take the time dimension for outcome event into account and realise the precise computing of incidence risk at a specific time point.

The prediction model established in this study included one demographic feature (age≥50 years), two terms of medical history (hyperlipidemia and smoking history), three SSc-related or ILD-related manifestations (dcSSc subtype, arthritis and shortness of breath), two serological features (decreased serum IgA and positive anti-Scl-70 antibody) and one therapeutic principle (CYC/MMF treatment), exhibiting good discrimination and calibration. Advanced age had been mentioned as an independent risk factor for PF-ILD in several previous literatures.14 16–20 Another registry study conducted in Canada also identified older age (HR 1.53, p<0.05) could predict the rapid deterioration of FVC.21 Moreover, age≥50 years was proved to be significantly associated with mortality in PROGRESS clinical cohort.2 Accordingly, ageing cells were observed with a stronger activation of profibrotic pathways in previous experimental studies. The probable mechanisms included telomeres dysfunction due to B cells dysregulation, higher endoplasmic reticulum stress and the age-related shift in fibroblasts populations leading to excessive wound healing.22–24 As Hambly et al recently mentioned, severe dyspnoea was associated with rapid progression of IPF,21 which was consistent with our findings. Given that shortness of breath, as a critical observable symptom, barely followed a consistent daily pattern, it was advised being
measured each day. Nevertheless, challenges introduced by patients' actions to release their symptoms, such as restricting their routines and receiving supportive treatment, should also be faced. A cohort study from EUSTAR confirmed that high baseline mRSS Score was clearly related to the rapid deterioration of FVC and its occurrence time,\(^6\) conforming to the potential predictive value of diffused subtype for the development of SSc PF-ILD.

Notably, we also confirmed hyperlipidemia as a strong predictive risk factor for PF-ILD in patients with SSc, in which underlying mechanisms are still intricate. It could be deduced as a substantial hit before endothelial dysfunction and repetitive injury of vasculature, which forms a feedback-forward loop circuit of fibroblast activation and excessive extracellular matrix, resulting in self-sustaining lung fibrosis.\(^25\) Consistent with existing literatures,\(^27\) smoking history was also incorporated into the equation, making it extra important to encourage prior smoking cessation in patients at risk. Arthritis has not been mentioned to be correlated with rapid progression of ILD in patients with SSc in previous studies, but evidence reviewed by Borthwick were provided with the activation of the interleukin-1 family, which plays a critical role in inflammation and fibrosis in multiple tissues including the lung.\(^28\) IgA was vital in the maintenance of respiratory tract microbiota homeostasis, while IgA deficiency would induce recurrent infections and lead to structural airway disease and ILD.\(^29\) Both previous researches and our results declared that anti-Scl-70 antibody was associated with a high risk of severe ILD, particularly in the early course of disease.\(^30-35\) Immuno-suppressive drugs are the cornerstone in the treatment of SSc-ILD, in which CTX and MMF were majorly recommended.\(^34\) Interestingly, in our study population, no significant relationship between the use of antifibrotic therapy and the occurrence of PF-ILD was found. It may be because by the time of the study, nintedanib has not been widely used in China, and there are also not much evidence for pirfenidone in SSc-ILD treatment.

Data adopted in this study was sufficient to prove that a considerable number of patients with SSc will develop PF-ILD during the disease course. Accordingly, high-risk patients should be identified early and receive more close monitoring, which requires the implementation of more accurate stratification criteria. In this study, by constructing the prediction model of PF-ILD in patients with SSc-ILD, the risk probability of PF-ILD at 5 years was fitted and all patients were divided into three groups (low risk, moderate risk and high risk), with an incidence risk of <4.6%, 4.6%–20.3% and >20.3%, respectively. Considering the decision analysis curve which demonstrated the clinical utility of the model, the probability thresholds between 1% and 45% enlightened a positive net benefit, giving a clear indication for rheumatologists to provide close monitoring for patients in moderate-risk and high-risk groups. For patients in these two groups, comprehensive following up every 3 months to half a year was highly recommended.

This study still has some limitations. Foremost, preservation of radiological imaging data from each centre was still deficient during the importing process, which is needed in the future for analysing relationship between the extent of lung fibrosis in HRCT and disease progression. Similarly, incomplete records of pulmonary function during follow-up may also hinder the judgement whether and when rapid progression occurs. Subsequently, the rarity of SSc and the strict inclusion criteria led to the limited study population, while no other qualified cohorts could be tracked. Thus, all cases were used to develop the model, resulting in a lack of external validation data set, which called for more consideration on the generalisability and transportability of the conclusion. Besides, given that some patients with SSc-ILD were observed with insidious onset and slow progression in the early stage with no obvious respiratory symptoms, the limited 5-year follow-up interval offered the possibility that some SSc-ILD patients with rapid progression in the later stage were missed.

CONCLUSIONS

In conclusion, this study developed the first prediction model for PF-ILD in patients with SSc-ILD based on the readily accessible data from CRDC database and internally validated its accuracy and stability. The established formula for individualised risk estimation of SSc PF-ILD within 5 years could help in identifying patients with the highest risk, who should receive enhanced managing strategies during follow-up. External validation will be needed in the future to realise its generalisation in different populations. Consensus and more specific algorithm are required to guide how it could be applied in clinical practice.

Author affiliations
1Department of internal medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
2Department of Rheumatology, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China
3Department of Rheumatology and Clinical Immunology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Acknowledgements We thank CRDC multi-center co-authors as above for assistance with cases collections.

Contributors MH, XD and JZhou: data curation and writing—original draft preparation; ML: visualisation and investigation; JZhao and QW: conceptualisation, methodology and software; YH: validation and writing—reviewing; and DX and XZ: writing—reviewing and editing. All authors worked on cases collections. DX is responsible for overall content as the guarantor.

Funding This study was supported by the National High-Level Hospital Clinical Research Funding (2022-PUMCH-B-013 and 2022-PUMCH-D-009).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by institutional review board of Peking Union Medical College Hospital (S-478). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.
REFERENCES


