RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

Validation of a semi-quantitative method to assess interstitial lung disease severity and progression in systemic sclerosis by standard and low-dose HRCT scans

Lucas Tschalèr ,[1] Suzana Jordan,[1] Trond Mogens Aaløkken,[2,3] Mike Becker ,[1] Cathrine Brunborg,[4] Cosimo Bruni ,[1] Christian Clarenbach,[5] Rucsandra Dobrota ,[1] Michael Thomas Durheim,[6] Muriel Elhai ,[7] Thomas Frauenfelder,[8] Håvard Fretheim,[3] Torhild Garen,[9] Oyvind Midtvedt,[9] Carina Mihai ,[1] Øyvind Molberg,[3,9] Oliver Distler ,[1] Anna-Maria Hoffmann-Vold [9,10]

To cite: Tschalèr L, Jordan S, Aaløkken TM, *et al*. Validation of a semi-quantitative method to assess interstitial lung disease severity and progression in systemic sclerosis by standard and low-dose HRCT scans. *RMD Open* 2025;**11**:e004938. doi:10.1136/rmdopen-2024-004938

OD and A-MH-V contributed equally.

Received 24 September 2024
Accepted 6 February 2025

For numbered affiliations see end of article.

Correspondence to
Professor Anna-Maria Hoffmann-Vold;
a.m.hoffmann-vold@medisin.uio.no

ABSTRACT

Background While the presence of distinct imaging abnormalities by high-resolution CT (HRCT) defines interstitial lung disease (ILD), there is a relative lack of validated methods to quantify these abnormalities in clinical practice, limiting ILD severity and progression assessments. We aimed to validate a semi-quantitative method for lung fibrosis assessment in patients with systemic sclerosis associated ILD (SSc-ILD) by standard and low-dose HRCT, considering lung structure and function as integral components of ILD evaluation.

Methods SSc patients from Oslo and Zurich with HRCT images, pulmonary function tests, including forced vital capacity (FVC), diffusing capacity for carbon monoxide (DLCO) and the 6-minute walk test with oxygen (O_2) desaturation were enrolled. We validated the semi-quantitative fibrosis extent method by HRCT using criteria for content and construct validity, discrimination, sensitivity to change and feasibility, as well as inter- and intra-rater variability.

Results 65 SSc patients from Zurich and 90 from Oslo were included. Significant correlations were observed between the extent of fibrosis on HRCT and FVC ($r=-0.517$, $p<0.001$), DLCO ($r=-0.400$, $p<0.001$) and O_2 desaturation ($r=-0.500$, $p<0.001$), indicating content, construct and criterion validity. Discrimination and sensitivity to change assessments showed moderate correlation with DLCO ($r=-0.377$, $p=0.003$) but not with FVC or O_2 desaturation. Inter- and intra-rater variability demonstrated excellent reliability ($\kappa=0.891$ and $\kappa=0.996$, respectively), with HRCT quantification averaging 9–15 min, indicating high feasibility.

Conclusion This study confirms that semi-quantitative fibrosis assessment of HRCT for SSc-ILD meets most validation criteria, supporting its use in clinical practice and showing additive value of structural to functional ILD assessment.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ High-resolution chest tomography (HRCT) is the preferred modality to identify interstitial lung disease in patients with systemic sclerosis (SSc-ILD). The lack of a validated method to quantify SSc-ILD makes it difficult in clinical practice to assess severity and progression of ILD.

WHAT THIS STUDY ADDS

⇒ This study validated a semi-quantitative method for ILD assessment by comparing HRCT findings with functional markers like lung function, exercise capacity and respiratory symptoms, using modified Outcome Measures in Rheumatology Clinical Trials criteria. We confirm that semi-quantitative fibrosis assessment of HRCT for SSc-ILD meets most validation criteria.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This semi-quantitative method for the assessment of pulmonary fibrosis in SSc patients has now been validated, is easily applicable in clinical practice and requires minimal training. The method can therefore be used in clinical practice to assess the severity of ILD on HRCT and monitor disease progression.

INTRODUCTION

Systemic sclerosis (SSc) is a severe multiorgan disease and population-level data indicate that >50% of patients develop interstitial lung disease (ILD).[1] At present, ILD is the main cause of death in patients with SSc.[2 3] While ILD is formally a histopathological entity, diagnosis is in clinical practice by imaging,

with high-resolution computed tomography (HRCT) as the preferred modality. Due to the invasive nature of lung biopsies and possible side effects, these are rarely conducted in clinical practice and radiological abnormalities consistent with ILD are considered the gold standard for ILD diagnosis.[4 5] Accordingly, in the first evidence-based consensus statements for the management of ILD in SSc, HRCT was included for screening and diagnosis.[6 7] There was also consensus for repeat HRCT scans as part of the monitoring strategy for SSc-associated ILD (SSc-ILD).[7] Further support for this strategy was given by a recent study showing that only a 2% increase in ILD from baseline HRCT to follow-up HRCT had an impact on long-term mortality.[8] Pulmonary function testing (PFT) remains a key tool in the assessment of ILD progression and is widely used in clinical practice and clinical trials.[7 9] However, its inherent limitations, such as biological variability and non-specificity for ILD, as well as its inability to detect structural changes in the lung, necessitate the use of complementary imaging techniques. Consequently, HRCT has gained prominence in evaluating ILD in SSc, offering more detailed visualisation of the lung parenchyma and enhancing the overall assessment of disease progression. In addition, with several novel treatment options approved and available for progressive ILD, there is increasing interest in disease monitoring to identify ILD patients that progress. This is reflected by the latest general ILD guidelines endorsed by the 2022 ATS/ERS/JRS/ALAT where progressive pulmonary fibrosis (PPF) was introduced as a new concept, defined by the presence of two or more of the following features across 12 months: (1) lung function decline, (2) respiratory symptoms worsening or (3) worsening lung fibrosis on HRCT.[5] The PPF concept shows that HRCT has become an integral part of ILD assessment, before and after treatment.[10–13] It is, however, noteworthy that the PPF description does not clearly define how 'lung fibrosis by HRCT' should be evaluated and scored, and it does not specifically define what is meant by 'lung fibrosis worsening'. These shortcomings underscore that there is a lack of standardised and validated tools for evaluation of imaging abnormalities, at baseline and during follow-up, in patients with ILD.

Accordingly, there is an unmet need for easily available and precise tools in clinical practice to determine the extent of pulmonary fibrosis on HRCT. To date, the visual scoring system by Goh *et al* is frequently applied in clinical practice. By this system, the extent of ILD is defined either as limited or extensive ILD.[14] The staging system is based on a combination of the estimated extent of lung fibrosis on HRCT with a threshold of 20% and in borderline cases combined with a forced vital capacity (FVC) threshold of 70%.[14] However, despite being simple and easily applicable in routine clinical practice, it is hampered by the lack of sensitivity to change. Aiming for better resolution and the ability to track ILD-related changes across time in individual patients, we developed an alternative, semi-quantitative method to determine the extent of changes attributable to lung fibrosis.[15] While this method appeared to detect changes in ILD extent from baseline HRCT to follow-up HRCT in individual patients, it has to date not been validated as a tool for ILD assessment.

A frequently discussed topic regarding serial HRCTs is the radiation exposure of standard high-dose HRCT. Therefore, the University Hospital Zurich (USZ) replaced the full-dose HRCT with a sequential, nine-slice, low-dose HRCT, which showed high accuracy and sensitivity in the detection of ILD in SSc patients.[16 17] This approach primarily benefits patients by considerably lowering radiation exposure, emphasising the advantage of low-dose HRCT in routine clinical use.[16] It is, however, not clear if it is feasible to apply semi-quantitative scoring of ILD extent on such low-dose HRCT images.

The aims of the current study were to validate the semi-quantitative ILD extent assessment as a tool for ILD assessment and evaluate the applicability of low-dose HRCT scans from SSc patients. Our validation compared changes observed on HRCT with alterations in functional markers such as lung function, exercise capacity assessed by the 6-minute walking test (6MWT) and respiratory symptoms representing patients' experience of respiratory health. This comparative analysis allowed us to assess the correlation between structural abnormalities detected on HRCT and functional impairments, thereby validating the utility of HRCT as a tool for ILD assessment and its impact on mortality.

METHODS

SSc patients from the Zurich and Oslo cohorts were included

We included SSc patients from the Oslo University Hospital (OUH) and University Hospital Zurich (USZ) cohorts who had to repeat HRCT images and PFTs from baseline and at least one follow-up examination with an HRCT scan and a concurrent PFT. Data were retrieved in Zurich from the local European Scleroderma Trials and Research database entries (EUSTAR), and in Oslo from the Norwegian systemic vasculitis and rheumatic disease registry (NOSVAR).[18] All patients fulfilled the 2013 American College of Rheumatology/European Alliance of Associations for Rheumatology (ACR/EULAR) criteria.[19]

Clinical data of the SSc patients, including demographics, disease duration, organ involvement and treatments were available. Immunosuppressive therapies included prednisone or equivalent ≥10 mg, cyclophosphamide, methotrexate, azathioprine, mycophenolate mofetil, tocilizumab and rituximab. Organ manifestations were defined as previously described.[9 15 20] All PFTs were carried out according to American Thoracic Society/European Respiratory Society guidelines.[21] Values of FVC and diffusing capacity for carbon monoxide (DLCO) were recorded as % predicted, with changes noted as changes of absolute FVC % predicted. The 6MWT with O_2 saturation at rest and after 6 min was recorded. A drop

in O_2 saturation to <94% after exercise was considered desaturation.[22] We also assessed the change in O_2 saturation, defined as the difference in O_2 saturation at baseline and follow-up. Respiratory symptoms were graded using the modified Medical Research Council definitions but assessed by the treating physician.[23]

Disease duration was defined as the time between the onset of the first non-Raynaud's symptom to baseline assessment. Baseline was defined as the first available HRCT and follow-up as the last available HRCT. The observation period was determined as the time between baseline and follow-up.

Assessment of the extent of lung fibrosis on HRCT on regular and low-dose HRCTs

The HRCT examinations were done using a 64-slice multidetector CT (Somatom Definition AS, Siemens Healthcare, Erlangen, Germany) or a 128-slice multidetector CT (Definition Flash Dual Source, Siemens Healthcare, Germany). The patients were scanned in a supine position and during inspiration. The slice thickness was <2 mm.

For the assessment of ILD on standard HRCTs, we evaluated 10 slices, equally distributed over the entire lung, with a beginning set at the apex and ending at the basal section of the lung as described.[15] In detail, for each slice, the area with reticular patterns and the area of the total lung were measured by manually tracing around the areas and then summed up to a total fibrotic volume and a total lung volume as previously described and shown in figure 1A,B.[15] The ratio of total fibrotic and total lung area defined the extent of lung fibrosis given in percentage. Reticular pattern changes defined as fine intralobular fibrosis without evident cysts, microcystic, macrocystic and honeycombing were defined as fibrosis. Superimposed ground-glass opacities were defined as being equivalent to fibrosis.[3 15] Fibrosis progression was defined as increasing the extent of lung fibrosis between baseline and follow-up.

The low-dose HRCT protocol has been described.[16 17] Briefly, the protocol is set up with the scanning of nine slices unevenly distributed over the lungs. The first slice is at the level of the manubrium sterni, the second at the carina and the third at the top of the lower lobe with an increment of 80 mm to each other. The final six slices were placed below the third slice with an increment of 15 mm, covering in detail the basal parts of the lungs. The slice thickness was 1 mm.

Quantification of fibrosis on the low-dose HRCT images was carried out with the nine available slices. We had access to concurrent standard HRCT and low-dose HRCT scans from 28 SSc patients, 14 of whom had pulmonary fibrosis. Using these data, we were able to define a conversion factor which compensated for the basal predominance of slices (and thereby an overestimation of fibrosis compared with the standard dose HRCT scans) in the low-dose HRCT. This enabled the direct comparison of fibrosis extent between standard and low-dose HRCT scans. The conversion factor was calculated with the following formula:

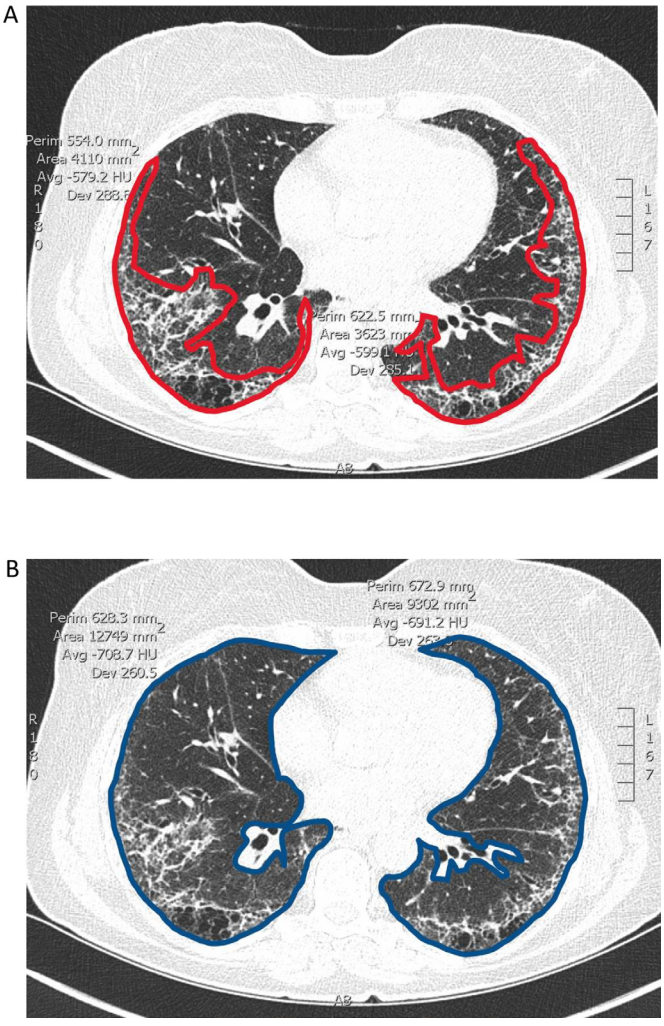

Figure 1 (A): Measurements of lung fibrosis on one HRCT slice made as an example. The area was measured by manually tracking the fibrotic area (marked in red). (B): Measurement of total lung volume on one HRCT slice made as an example. The area was measured by manually tracking the lung area (marked in blue).

On all low-dose HRCTs, this conversion factor was applied, and the same methodology for the extent of lung fibrosis assessment was used for both low and high-dose images. The quantification was carried out by two non-radiologists (one trained beginner and one with 9 years of experience, LT and AMHV) using PACS-Viewer (Impax, AGFA, Dübendorf, Switzerland). The HRCT scans from the Zurich and Oslo cohorts were fully quantified by one examiner and then blindly re-quantified by the second examiner for comparison and validation.

Validation of the semi-quantitative quantification of lung fibrosis on HRCT images

Structural changes assessed on HRCT, functional aspects of the lungs and respiratory symptoms reflect distinct but complementary dimensions of ILD capturing different aspects of respiratory health. For this reason, FVC, DLCO,

Table 1 The applied criteria for the assessment of validation and definitions and statistical analyses in this study

Criterion	Definition	In this study	Analysis
Truth			
Face validity (credibility)	Overall appropriateness of the measure, as assessed by investigators and clinicians	Appropriateness of lung fibrosis assessment on HRCT to evaluate SSc-ILD presence and severity	Agreement by investigators
Criterion validity and construct validity	The ability of the measure to match with the hypothesised expectations of the investigator compared with other indirect assessments The ability of the measure to predict all those components of health status relevant to the intervention being assessed	Ability of the structural severity of lung fibrosis (extent) to match functional assessments (PFT, 6MWT) and respiratory symptoms Ability of fibrosis score to predict mortality	Correlation between lung fibrosis score and PFT, 6MWT with O_2 desaturation and respiratory symptoms at baseline Association of fibrosis score and mortality
Discrimination			
Sensitivity to change	Change between two timepoints	Change of fibrosis score over time	Correlation of change in lung fibrosis score and change in FVC, DLCO and O_2 desaturation from baseline to follow-up Association of change in lung fibrosis score and mortality
Reliability (reproducibility)	Based on the evaluation of intraclass correlations	Intra- and inter-rater variability of lung fibrosis score	Intraclass correlation coefficient (ICC)
Feasibility			
Feasibility	The measure's ease of use, practicability and applicability	Time use of fibrosis score assessment Applicability for low-dose HRCT versus standard HRCT Applicability on different slides thicknesses	Minutes/assessment Assessment in 1 and 2 mm slice thickness HRCTs Comparison between Zurich and Oslo

The baseline refers to the timepoint when the first HRCT scan was made, while the follow-up corresponds to the timepoint of the last available HRCT scan.
DLCO, diffusing capacity for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution CT; ICC, intraclass correlation coefficient; 6MWT, 6-minute walking test; OUH, Oslo University Hospital; PFT, pulmonary function test; SSc-ILD, systemic sclerosis associated interstitial lung disease; USZ, University Hospital Zurich.

the 6MWT with O_2 desaturation and respiratory symptoms were used as functional and patients' experience, and the extent of lung fibrosis was semi-quantitatively assessed as structural aspects. By using functional markers and respiratory symptoms as comparators for our HRCT-based structural assessments, we aimed to validate the impact of semi-quantitative evaluation of ILD severity and progression on other parts of respiratory health and on mortality. For validation, we applied the proposed criteria from the Outcome Measures in Rheumatology Clinical Trials consensus group (table 1). We applied all these criteria to validate the impact of semi-quantitative quantification of lung fibrosis on HRCT images. Each criterion was rated valid (V), not valid (NV) or partially valid (PV) based on the statistical test results.[24 25] The criterion was considered valid if there was a correlation or if there was agreement among the authors. In contrast, the criterion was declared not valid if there was no correlation or disagreement among the authors. Partially valid was defined as differing results between the cohorts.

Statistical analyses and validation of the methodology

Statistical analyses were performed using STATA V.17 (StataCorp LLC, College Station, Texas, USA) and SPSS V.26 (IBM, Armonk, New York, USA). We did not conduct a formal power calculation. We randomly included patients with and without lung fibrosis from both cohorts, provided that HRCT images and lung function tests from the same time point were available. In Zurich, we also included patients who had low-dose HRCTs available. No specific patient groups were excluded from the study. Continuous variables were reported as mean (SD), and categorical variables as frequencies and percentages. Descriptive statistics were applied. Comparisons between groups were evaluated with the Pearson χ^2 test, Fisher's exact test and the Kruskal-Wallis t-test, and correlations

Table 2 Demographic and clinical characteristics of SSc patients in the Zurich and Oslo SSc cohorts

Characteristics	Zurich (n=65)	Oslo (n=90)	P value
Age, years (SD)	48 (16)	53 (15)	0.074
Disease duration at first available HRCT, years (SD)	9 (12)	5 (9)	0.014
Observation period, months (SD)	44 (27)	45 (31)	0.877
Female sex, n (%)	56 (86)	72 (80)	0.319
Ever smoker, n (%)	28 (43)	36 (40)	0.342
Diffuse cutaneous SSc, n (%)	21 (32)	22 (24)	0.281
Anti-centromere antibody, n (%)	26 (40)	46 (51)	0.133
Anti-topoisomerase I antibody, n (%)	16 (25)	10 (9)	0.026
Modified Rodnan Skin Score (SD)	8 (9)	10 (11)	0.259
Digital ulcers ever, n (%)	35 (54)	41 (46)	0.212
GORD, n (%)	43 (66)	40 (53)	0.007
Myopathy, n (%)	15 (23)	9 (10)	0.026
PH-ILD, n (%)	14 (22)	20 (22)	0.471
Deceased, n (%)	22 (34%)	39 (43%)	0.740
Immunosuppression, n (%)	30 (46)	34 (38)	0.223

Immunosuppressives included prednisone or equivalent ≥10mg, cyclophosphamide, methotrexate, azathioprine, mycophenolate mofetil and rituximab.
GORD, gastro-oesophageal reflux disease; HRCT, high-resolution CT; PH-ILD, pulmonary hypertension-interstitial lung disease; SSc, systemic sclerosis.

were measured by Pearson or Spearman coefficients as appropriate. For the correlation coefficient r, small effects were defined as 0.1, medium effects were 0.3 and large effects were 0.5.[26] The sensitivity to change was also assessed by receiver operating characteristic curves tested by area under the curve, where values >0.7 were considered acceptable. For inter- and intra-rater variability, we used the intraclass correlation coefficient. κ<0.4 was considered poor; values of 0.4–0.75 were considered moderate to good, and a κ of >0.75 represented excellent agreement.

RESULTS

Patient demographics

In the Zurich cohort, 65 SSc patients (86% females) were included with a mean age of 48 years (table 2). At baseline, 23/65 patients and at follow-up, 46/65 patients received a low-dose HRCT (table 3). In the Oslo cohort, 90 SSc patients were included with a mean age of 53 years, and 72/90 (80%) were females (table 2). All these 90 cases were examined by repeat standard HRCT scans. As the main characteristics of the Zurich and Oslo cohorts were similar, the two cohorts were combined for all further steps.

Validation of the quantification method

Truth was tested in several steps, with face validity first. The authors discussed the face validity of assessing lung fibrosis on HCRT. Consensus was reached on the validity of the proposed quantification method, as it directly assesses imaging indicators of tissue fibrosis, including reticular changes and associated ground glass opacities. Moreover, the extent of lung fibrosis serves as a surrogate marker for ILD severity. Finally, quantifying lung fibrosis was recommended as a valuable monitoring tool for identifying disease progression in ILD.

Next, we assessed construct validity. Based on available information, FVC%, DLCO%, respiratory symptoms, 6MWT and O_2 desaturation were considered as credible assessment measures for SSc-ILD. We also considered mortality as an important outcome for SSc-ILD (table 3). There was large construct and criterion validity in the Zurich cohort between the lung fibrosis score and FVC, O_2 desaturation and respiratory symptoms (table 4). In addition, a higher baseline lung fibrosis score was associated with mortality (14.6% (SD 13.4) in deceased patients compared with 3.8% (SD 5.7) in non-deceased; p<0.001). Consistent with the results from the Zurich cohort, there was a large correlation in the Oslo cohort between the lung fibrosis score and O_2 desaturation and a moderate correlation with FVC and DLCO (table 4). As in the Zurich cohort, a higher baseline extent of lung fibrosis was associated with mortality (13.5% (SD 20.9) in deceased compared with 6.1% (SD 13.05) in non-deceased patients; p=0.045).

In the next step, we determined discrimination and sensitivity to change between baseline and follow-up, which varied between 5 months and 120 months (table 3). The mean extent of lung fibrosis increased from 11.8% (SD 10.9) at baseline to 14.9% (SD 12.3) at follow-up, of which 17 (26%) had an increase of more than 2% in the Zurich cohort. In the Oslo cohort, the

Table 3 Extent of lung fibrosis on HRCT, pulmonary function parameters and respiratory symptoms at baseline and follow-up of SSc patients in the Zurich and Oslo SSc cohorts

	Zurich (n=65)		Oslo (n=90)	
	Baseline	Follow-up	Baseline	Follow-up
Time between assessments, years (SD)	3.7 (2.3)		3.7 (2.6)	
Any fibrosis, n (%)	41/65 (63)	42/65 (65)	51/90 (57)	50/90 (56)
Extent of lung fibrosis, % (SD)	7.4 (10.3)	9.6 (12.2)	9.7 (17.5)	11.7 (19.4)
Low dose HRCT, n/N (%)	23/65 (35)	46/65 (71)	n/a	n/a
Standard and low-dose HRCT, n (%)	23/28 (82)	5/28 (18)	n/a	n/a
FVC% (SD)	90 (21)	83 (23)	91 (23)	87 (22)
DLCO% (SD)	54 (13)	47 (17)	59 (21)	53 (18)
6 min walk test, m (SD)	458 (116)	428 (154)	457 (146)	477 (115)
O_2 saturation at rest, % (SD)	96 (2)	96 (4)	96 (2)	95 (5)
O_2 desaturation <94%, n/N (%)	7/62 (11)	9/58 (16)	3/16 (19)	4/16 (25)
Respiratory symptoms, >1, n/N (%)	42/57 (74)	n/a	56/85 (65)	n/a

The baseline refers to the timepoint when the first HRCT scan was performed, while the follow-up corresponds to the timepoint of the last available HRCT scan.
Respiratory symptoms were graded using the modified Medical Research Council definitions.
DLCO, diffusing capacity for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution CT; PFT, pulmonary function test; SSc, systemic sclerosis.

mean extent of fibrosis was 16.8% (SD 20.3) at baseline, increasing to 20.1% (SD 21.9) at follow-up; 20 (18%) had an increase of more than 2%. There was no correlation between these changes and the decline in FVC% and a moderate correlation of fibrosis progression assessed on HRCT with DLCO in the Zurich cohort (table 5). There was no correlation in the Oslo cohort (table 5). Correlations between 6MWD and respiratory were not made due to missing follow-up data. We found that increasing lung fibrosis on HRCT was associated with mortality compared with stable lung fibrosis in the Zurich cohort (14/30 (46.7%) compared with 8/35 (22.9%); p=0.043).

Next, we assessed the reliability by determining intra- and inter-rater variability of the lung fibrosis scores. The inter-rater variability analysis revealed excellent agreement between the two raters (κ=0.891). The semi-quantification of HRCT images was carried out twice in 31 randomly selected HRCTs. The difference between the two fibrosis scores was a mean of 1.37% (SD 1.63). Thus, the intra-rater variability was also excellent (κ=0.996).

Finally, we assessed the feasibility of using this semi-quantitative approach in clinical practice. We tested both the need for clinical experience and time use. The quantification was conducted by a non-radiologist without any experience in quantification of HRCTs. He was trained in two sessions of about 1 hour each and supervised the first five HRCTs. Both inter and intra-rater variability were excellent. The average time per HRCT quantification was between 9 and 15 min, depending on the image quality and fibrosis extent. Therefore, the feasibility was rated high (table 6).

DISCUSSION

According to the first evidence-based expert consensus for the management of SSc-ILD, assessment of ILD severity

Table 4 Construct and criterion validity assessed by correlation analyses between lung fibrosis score on HRCT and pulmonary function test, 6-minute walking test, O_2 desaturation and respiratory symptoms

	Zurich		Oslo	
Variables	Fibrosis at baseline (%)	P value	Fibrosis at baseline (%)	P value
FVC (%)	r=−0.538	<0.001	r=−0.406	<0.001
DLCO (%)	r=−0.400	<0.001	r=−0.346	0.004
6MWD (m)	r=−0.231	0.071	r=−0.105	0.349
O_2 desaturation (%)	r=−0.500	<0.001	r=−0.550	0.033
Respiratory symptoms	r=0.472	<0.001	r=0.290	0.007

The baseline refers to the timepoint when the first HRCT scan was performed.
Respiratory symptoms were graded using the modified Medical Research Council definitions.
DLCO, diffusing capacity for carbon monoxide; FVC, forced vital capacity; 6MWD, 6-minute walk distance; PFT, pulmonary function test.

Table 5 Sensitivity to change assessed with correlation analyses between lung fibrosis progression on high-resolution CT, pulmonary function and change of O_2 saturation

	Zurich			Oslo		
Variables	**Change of fibrosis (%)**	**P value**	**AUC**	**Change of fibrosis (%)**	**P value**	**AUC**
FVC change (%)	r=−0.115	0.383	0.56	r=−0.139	0.250	0.31
DLCO change (%)	r=−0.377	0.003	0.54	r=−0.039	0.758	0.51
Change of O_2 saturation	r=−0.155	0.268	0.65	n/a	n/a	n/a

AUC, area under the curve; DLCO, diffusing capacity for carbon monoxide; FVC, forced vital capacity; PFT, pulmonary function test.

assessed on HRCT and repeat HRCT were recommended for SSc-ILD monitoring.[7] Agreement was reached by the authors on the proposed quantification method, which directly measures imaging surrogates of the extent of tissue fibrosis as a marker of disease severity and serves as a monitoring tool for identifying ILD progression. Therefore, the authors agreed that face validity was considered valid.

The content and construct validity showed a moderate correlation between the assessed tools. Solely the 6MWD in both cohorts and respiratory symptoms in the Oslo cohort correlated mildly with the extent of fibrosis. The rather poor correlation between ILD detected on HRCT and lung function parameters was not surprising, as this had already been reported in other studies.[15 27] Importantly, normal lung function tests are frequently found in SSc patients despite the presence of ILD on HRCT, which makes the HRCT inevitable for screening and detection of ILD.[3 22 28] Overall content and construct validity was also rated as valid. The strong association between higher fibrosis scores and increased mortality rates observed in both cohorts confirms the criterion validity of this measure and highlights the necessity to assess ILD severity on HRCT at diagnosis.

The sensitivity to change analyses showed only a moderate correlation in the Zurich cohort between DLCO and the progression of lung fibrosis. Correlation of the change in extent of lung fibrosis and the change in FVC between baseline and follow-up showed no correlation. Therefore, sensitivity to change was rated as partially valid. This was again not unexpected, as similar results on the progression of lung fibrosis and decline in pulmonary function have been shown in previous studies.[15 29] For clinical practice, this highlights the necessity to monitor SSc-ILD patients using several modalities, such as lung function and serial HRCTs.[7] The mild to moderate correlations between the tools also support the role of multiple assessment tools for the evaluation of progressive SSc-ILD.[5 30] On the other hand, the association between increasing fibrosis on HRCT scans and mortality suggests a sensitivity to change in this measure, highlighting its potential utility in monitoring disease progression over time. These associations of fibrosis scores with mortality underscore the robustness and clinical relevance of baseline and repeat assessments of lung fibrosis scores as a prognostic marker in ILD, validating their use in clinical practice for risk assessment and management decisions.

In our study, we confirm that an HRCT without ILD at baseline indicates a good prognosis.[15 31] Of the 24/65 patients without ILD on baseline HRCT in the Zurich cohort, 3 (12.5%) developed ILD at follow-up after a mean of 3.7 (SD 2.3) years. In the Oslo cohort, 39/90 patients showed no fibrosis at baseline, and none developed fibrosis after a mean of 3.2 (SD 2.7) years. However,

Table 6 Overview of all Outcome Measures in Rheumatology Clinical Trials filter criteria rated as valid (green), not valid (red) or partially valid (orange)

Truth	Face validity (credibility) (V)	Agreement of the authors					
	Content validity and construct validity (V)	FVC	DLCO	6MWD	O_2 desaturation	Respiratory symptoms	Mortality
Discrimination	Sensitivity to change (PV)	FVC decline	DLCO decline		Change of O_2 saturation		Mortality
	Reliability (V)	Intra-rater variability	Inter-rater variability				
Feasibility	Feasibility (V)	Time use of lung fibrosis score assessment					

Respiratory symptoms were graded using the modified Medical Research Council definitions.
DLCO, diffusing capacity for carbon monoxide; FVC, forced vital capacity; 6MWD, 6-minute walk distance; PV, partially valid; V, valid.

it indicates also that repeat HRCTs are important in at-risk patients to identify new onset ILD as early as possible.

Important for clinical practice is that this methodology is not only valid and easily accessible but also feasible. We demonstrate here excellent intra- and inter-rater reliability, showing consistency, despite no prior experience of one of the raters and only short training sessions. This highlights that the assessment is easily applicable in everyday clinical practice. Nevertheless, the quantification will take 9–15 min per HRCT image, which especially in centres that care for many SSc patients can be time-consuming. Establishing a validated method to measure the extent of fibrosis, as demonstrated in our study, could also inform therapeutic decisions in the future. This improvement would enable more precise adjustments to treatment plans based on disease trajectories. This would need to be assessed in a follow-up study before implementation.

Our study is not without limitations. The most important limitation is the conversion factor between assessments from low and high-dose HRCTs which was developed only for this study. In addition, only 14 patients had available both low and high-dose HRCT assessments for comparison. Nevertheless, the conversion factor and the semi-quantitative assessment performed well in low-dose HRCTs suggesting that this method is also applicable in low-dose HRCTs. Another limitation of this study is the varying observation period and time between baseline and follow-up HRCTs, which may have led to different fibrosis progression and lung function values. We also acknowledge that the data originate from two single-centre cohorts and incorporating data from additional centres and expanding the dataset could have strengthened the study. As our study cohort is representative of the broader SSc-ILD population, where severe pulmonary involvement occurs less frequently, data should also be validated in patients with severe SSc-ILD. Larger, multicentre studies with geographically diverse populations are needed to confirm the generalisability of our findings to other populations. Additionally, while the study was supervised by experienced radiologists most but not all images were re-read by radiologists.

In conclusion, this study shows that most of the criteria were fulfilled, and we could therefore validate this easily applicable semi-quantitative quantification method of lung fibrosis assessment on HRCT in SSc patients not only in full but also low-dose HRCT. The methodology does not need any specific training and can be applied for assessment both at baseline for severity and at follow-up for progression assessment. The study underscores the significance of a comprehensive evaluation of lung involvement, encompassing structural, functional and respiratory symptom assessments. Each aspect offers unique insights into respiratory health, highlighting the importance of a multifaceted approach.

Author affiliations
[1]Department of Rheumatology, University Hospital Zurich, Zurich, Switzerland
[2]Department of Radiology, Oslo University Hospital, Oslo, Norway
[3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[4]Biostatistics, Epidemiology and Health Economics, Oslo University Hospital, Oslo, Norway
[5]Department of Pulmonology, University Hospital Zurich, Zurich, Switzerland
[6]Department of Respiratory Disease, Oslo University Hospital, Oslo, Norway
[7]University Hospital Zürich Center of Experimental Rheumatology, Zurich, Switzerland
[8]Department of Radiology, UniversitätsSpital Zürich, Zurich, Switzerland
[9]Department of Rheumatology, Oslo University Hospital, Oslo, Norway
[10]University Hospital Zurich, Zurich, Switzerland

X Cosimo Bruni @cosimobruni and Muriel Elhai @MurielElhai

Contributors LT, A-MH-V and OD conceived and planned the study. LT, A-MH-V and OD were involved in the further conceptualisation and methodology of the study. LT and A-MH-V accessed and verified the underlying data reported and wrote the initial draft of the manuscript. LT, SJ, A-MH-V, OD and ØMolberg were involved in the interpretation of results. LT, A-MH-V and CB performed statistical analysis. LT, SJ, TMA, MB, CBrunborg, CBruni, CC, RD, MTD, ME, TF, HF, GT, OMidtvedt, CM and A-MH-V collected study data. All authors contributed with intellectual, technical, material or administrative support. All authors critically revised the manuscript and approved the final submitted version. LT and A-MH-V had final responsibility for the decision to submit for publication. A-MH-V is the guarantor of this work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests LT, SJ and TMA report no conflicts. MB reports research funding, consultancy fees and/or speaker fees from GSK, Amgen, Novartis and Vifor. CB reports no conflicts. CC reports consulting fees and speaker fees from GSK, Novartis, Vifor, Boehringer, Astra Zeneca and Sanofi; speaker fees from Daiichi Sankyo; received support for attending meetings from Boehringer and Astra Zeneca. RD reports grants from the Iten-Kohaut Foundation; received support for attending meetings from Otsuka and Amgen; participated on an advisory board for Boehringer Ingelheim. MTD reports grants from Boehringer Ingelheim and Roche; received support for attending meetings and participated on an advisory board for Boehringer Ingelheim. ME received support for attending meetings from Janssen and Astra Zeneca. TF reports speaker fees from Bracco, Bayer and Boehringer; participated on an advisory board for AGFA; holds leadership or fiduciary roles with European Society of Thoracic Imaging and Swiss Society of Radiology. HF reports consulting fees from Bayer; speaker fees from Boehringer Ingelheim. GT and ØMolberg report no conflicts. CM reports speaker fees from MEDtalks Switzerland, Mepha, PlayToKnow AG, Medbase AG, Romanian Society of Rheumatology and Boehringer Ingelheim; support for attending meetings from Boehringer Ingelheim and Roche; participated on an advisory board for Boehringer Ingelheim and Janssen. OMidtvedt reports no conflicts. OD reports grants from Kymera, Mitsubishi Tanabe and Boehringer Ingelheim; consulting fees from 4P-Pharma, Abbvie, Acceleron, Alcimed, Altavant Sciences, Amgen, AnaMar, Arxx, Astra Zeneca, Blade Therapeutics, Bayer, Boehringer Ingelheim, Corbus Pharmaceuticals, CSL Behring, Galapagos, Glenmark, Horizon, Gossamer, Janssen, Kymera, Lupin, Medscape, Merck, Miltenyi Biotec, Mitsubishi Tanabe, Novartis, Orion, Prometheus Biosciences, Redxpharma, Roivant and Topadur; speaker fees from Bayer, Boehringer Ingelheim, Janssen and Medscape; has a patent issued for 'mir-29 for treatment of systemic sclerosis'; hold leadership or fiduciary roles with FOREUM foundation, ERS/EULAR Guidelines, EUSTAR, SCQM (Swiss Clinical Quality Management in Rheumatic Disease), SAMW (Swiss Academy of Medical Sciences) and Hartmann Müller Foundation; and reports other interest with Citus AG. A-MH-V reports grants from BI and Janssen; speakers fees from: Boehringer Ingelheim, Janssen, Medscape, Merck Sharp & Dohme, Novartis and Roche; consultancy fees of: ARXX, Boehringer Ingelheim, Genentech, Janssen, Medscape, Merck Sharp & Dohme, Pliant therapeutics, Roche and Werfen; and holds leadership or fiduciary roles with EULAR, and the Nordic PH vision group.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. The study was performed according to the Declaration of Helsinki and was approved by the local ethics committees. All patients had signed informed consent for EUSTAR or the general consent of the University Hospital Zurich for the use of their clinical data (EUSTAR BASEC-Nr. 2016-01515; BASEC-Nr. 2018-02165). In Oslo, patients had signed the NOSVAR consent form (approved by The Regional Committee of Health and Medical Research Ethics in South-East Norway, research protocol No. 2016/119). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. A de-identified patient data set will be made available to researchers upon reasonable request after the time of publication. To acquire data access a project plan must be submitted, and the research group has to approve the request. Please contact the corresponding author for such inquires. Data sharing will have to follow appropriate regulations.

ORCID iDs
Lucas Tschalèr http://orcid.org/0009-0007-5831-3674
Mike Becker http://orcid.org/0000-0001-9102-3088
Cosimo Bruni http://orcid.org/0000-0003-2813-2083
Rucsandra Dobrota http://orcid.org/0000-0001-9819-7574
Muriel Elhai http://orcid.org/0000-0001-8627-5758
Carina Mihai http://orcid.org/0000-0002-8627-8817
Oliver Distler http://orcid.org/0000-0002-0546-8310
Anna-Maria Hoffmann-Vold http://orcid.org/0000-0001-6467-7422

REFERENCES

1 Fretheim H, Halse A-K, Seip M, *et al*. Multidimensional tracking of phenotypes and organ involvement in a complete nationwide systemic sclerosis cohort. *Rheumatology (Oxford)* 2020;59:2920–9.
2 Elhai M, Meune C, Boubaya M, *et al*. Mapping and predicting mortality from systemic sclerosis. *Ann Rheum Dis* 2017;76:1897–905.
3 Hoffmann-Vold A-M, Fretheim H, Halse A-K, *et al*. Tracking Impact of Interstitial Lung Disease in Systemic Sclerosis in a Complete Nationwide Cohort. *Am J Respir Crit Care Med* 2019;200:1258–66.
4 Hatabu H, Hunninghake GM, Richeldi L, *et al*. Interstitial lung abnormalities detected incidentally on CT: a Position Paper from the Fleischner Society. *Lancet Respir Med* 2020;8:726–37.
5 Raghu G, Remy-Jardin M, Richeldi L, *et al*. Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2022;205:e18–47.
6 Denton CP, Goh NS, Humphries SM, *et al*. Extent of fibrosis and lung function decline in patients with systemic sclerosis and interstitial lung disease: data from the SENSCIS trial. *Rheumatology (Oxford)* 2023;62:1870–6.
7 Hoffmann-Vold A-M, Maher TM, Philpot EE, *et al*. The identification and management of interstitial lung disease in systemic sclerosis: evidence-based European consensus statements. *Lancet Rheumatol* 2020;2:e71–83.
8 Volkmann ER, Tashkin DP, Roth MD, *et al*. Early Radiographic Progression of Scleroderma: Lung Disease Predicts Long-term Mortality. *Chest* 2022;161:1310–9.
9 Hoffmann-Vold A-M, Allanore Y, Alves M, *et al*. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. *Ann Rheum Dis* 2021;80:219–27.
10 Distler O, Highland KB, Gahlemann M, *et al*. Nintedanib for Systemic Sclerosis-Associated Interstitial Lung Disease. *N Engl J Med* 2019;380:2518–28.
11 Hoffmann-Vold A-M, Distler O, Crestani B, *et al*. Recent advances in the management of systemic sclerosis-associated interstitial lung disease. *Curr Opin Pulm Med* 2022;28:441–7.
12 Khanna D, Lin CJF, Furst DE, *et al*. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Respir Med* 2020;8:963–74.
13 Maher TM, Tudor VA, Saunders P, *et al*. Rituximab versus intravenous cyclophosphamide in patients with connective tissue disease-associated interstitial lung disease in the UK (RECITAL): a double-blind, double-dummy, randomised, controlled, phase 2b trial. *Lancet Respir Med* 2023;11:45–54.
14 Goh NS, Desai SR, Veeraraghavan S. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med* 2008;177:1248–54.
15 Hoffmann-Vold A-M, Aaløkken TM, Lund MB, *et al*. Predictive value of serial high-resolution computed tomography analyses and concurrent lung function tests in systemic sclerosis. *Arthritis Rheumatol* 2015;67:2205–12.
16 Frauenfelder T, Winklehner A, Nguyen TDL, *et al*. Screening for interstitial lung disease in systemic sclerosis: performance of high-resolution CT with limited number of slices: a prospective study. *Ann Rheum Dis* 2014;73:2069–73.
17 Nguyen-Kim TDL, Maurer B, Suliman YA, *et al*. The impact of slice-reduced computed tomography on histogram-based densitometry assessment of lung fibrosis in patients with systemic sclerosis. *J Thorac Dis* 2018;10:2142–52.
18 Hoffmann-Vold A-M, Midtvedt Ø, Tennøe AH, *et al*. Cardiopulmonary Disease Development in Anti-RNA Polymerase III-positive Systemic Sclerosis: Comparative Analyses from an Unselected, Prospective Patient Cohort. *J Rheumatol* 2017;44:459–65.
19 van den Hoogen F, Khanna D, Fransen J, *et al*. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann Rheum Dis* 2013;72:1747–55.
20 Meier FMP, Frommer KW, Dinser R, *et al*. Update on the profile of the EUSTAR cohort: an analysis of the EULAR Scleroderma Trials and Research group database. *Ann Rheum Dis* 2012;71:1355–60.
21 Quanjer PH, Tammeling GJ, Cotes JE, *et al*. Lung volumes and forced ventilatory flows. *Eur Respir J* 1993;6 Suppl 16:5–40.
22 Suliman YA, Dobrota R, Huscher D, *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 Rheumatol* 2015;67:3256–61.
23 Rajala K, Lehto JT, Sutinen E, *et al*. mMRC dyspnoea scale indicates impaired quality of life and increased pain in patients with idiopathic pulmonary fibrosis. *ERJ Open Res* 2017;3.
24 Kowal-Bielecka O, Avouac J, Pittrow D, *et al*. Echocardiography as an outcome measure in scleroderma-related pulmonary arterial hypertension: a systematic literature analysis by the EPOSS group. *J Rheumatol* 2010;37:105–15.
25 Kowal-Bielecka O, Avouac J, Pittrow D, *et al*. Analysis of the validation status of quality of life and functional disability measures in pulmonary arterial hypertension related to systemic sclerosis: results of a systematic literature analysis by the Expert Panel on Outcomes Measures in Pulmonary Arterial Hypertension related to Systemic Sclerosis (EPOSS). *J Rheumatol* 2011;38:2419–27.
26 Ellis PD. *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press, 2010.
27 Ooi GC, Mok MY, Tsang KWT, *et al*. Interstitial lung disease in systemic sclerosis. *Acta Radiol* 2003;44:258–64.
28 Hoffmann-Vold AM, Molberg Ø. Detection, screening, and classification of interstitial lung disease in patients with systemic sclerosis. *Curr Opin Rheumatol* 2020;32:497–504.
29 Wu W, Jordan S, Becker MO, *et al*. Prediction of progression of interstitial lung disease in patients with systemic sclerosis: the SPAR model. *Ann Rheum Dis* 2018;77:1326–32.
30 Flaherty KR, Wells AU, Cottin V, *et al*. Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. *N Engl J Med* 2019;381:1718–27.
31 Launay D, Remy-Jardin M, Michon-Pasturel U, *et al*. High resolution computed tomography in fibrosing alveolitis associated with systemic sclerosis. *J Rheumatol* 2006;33:1789–801.