

Ultrasound in the Assessment of Interstitial Lung Disease in Systemic Sclerosis: A Systematic Literature Review by the OMERACT Ultrasound Group

Marwin Gutierrez^{ID}, Carina Soto-Fajardo^{ID}, Carlos Pineda^{ID}, Alfonso Alfaro-Rodriguez^{ID}, Lene Terslev, George Bruyn, Annamaria Iagnocco, Chiara Bertolazzi^{ID}, Maria Antonietta D'Agostino^{ID}, and Andrea Delle Sedie^{ID}

ABSTRACT. Objective. To provide an overview of the role of lung ultrasound (LUS) in the assessment of interstitial lung disease (ILD) in systemic sclerosis (SSc) and to discuss the state of validation supporting its clinical relevance and application in daily clinical practice.

Methods. Original articles published between January 1997 and October 2017 were included. To identify all available studies, a detailed search pertaining to the topic of review was conducted according to guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). A systematic search was performed in PubMed and EMBASE. The quality assessment of retrieved articles was performed according to the Oxford Center for Evidence-based Medicine. The methodological quality of the studies was assessed using the Cochrane Handbook for Systematic Reviews and the Quality Assessment of Diagnostic Accuracy Studies–2 tool.

Results. From 300 papers identified, 12 were included for the analysis. LUS passed the filter of face, content validity, and feasibility. However, there is insufficient evidence to support criterion validity, reliability, and sensitivity to change.

Conclusion. Despite a great deal of work supporting the potential role of LUS for the assessment of ILD-SSc, much remains to be done before validating its use as an outcome measure in ILD-SSc. (J Rheumatol First Release March 15 2020; doi:10.3899/jrheum.180940)

Key Indexing Terms:

SYSTEMIC SCLEROSIS

PULMONARY ULTRASOUND

FEASIBILITY

INTERSTITIAL LUNG DISEASE

VALIDITY

RELIABILITY

From the Division of Musculoskeletal and Rheumatic Disorders, and the Division of Neurosciences, Instituto Nacional de Rehabilitación, Mexico City; Doctorado en Ciencias Biológicas y de la Salud, Universidad Autónoma Metropolitana, Iztapalapa, México City, Mexico; Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; Department of Rheumatology, MC Groep, Lelystad, the Netherlands; Academic Rheumatology Center, Università degli Studi di Torino, Turin, Italy; Université Versailles Saint-Quentin en Yvelines, Paris, France; Rheumatology Unit, University of Pisa, Pisa, Italy.

M. Gutierrez, MD, PhD, Division of Musculoskeletal and Rheumatic Disorders, Instituto Nacional de Rehabilitación, and Doctorado en Ciencias Biológicas y de la Salud, Universidad Autónoma Metropolitana; C. Soto-Fajardo, MD, Division of Musculoskeletal and Rheumatic Disorders, Instituto Nacional de Rehabilitación; C. Pineda, MD, Division of Musculoskeletal and Rheumatic Disorders, Instituto Nacional de Rehabilitación; A. Alfaro-Rodriguez, MD, Division of Neurosciences, Instituto Nacional de Rehabilitación; L. Terslev, MD, PhD, Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Rigshospitalet; G. Bruyn, MD, PhD, Department of Rheumatology, MC Groep; A. Iagnocco, MD, PhD, Academic Rheumatology Center, Università degli Studi di Torino; C. Bertolazzi, MD, PhD, Division of Musculoskeletal and Rheumatic Disorders, Instituto Nacional de Rehabilitación; M.A. D'Agostino, MD, PhD, Université Versailles Saint-Quentin en Yvelines; A. Delle Sedie, MD, PhD, Rheumatology Unit, University of Pisa.

Address correspondence to Dr. M. Gutierrez, Division of Musculoskeletal and Rheumatic Disorders, Instituto Nacional de Rehabilitación, Luis

Guillermo Ibarra Ibarra, Calzada Mexico-Xochimilco 289, Colonia Arenal de Guadalupe, CP 143898, Mexico City, Mexico.
E-mail: dr.gmarwin@gmail.com

Accepted for publication June 18, 2019.

Interstitial lung disease (ILD) is a clinical manifestation affecting more than half of patients with systemic sclerosis (SSc)^{1,2}. It may be established within the first 4 years of the disease and is frequently subclinical^{3,4}. Although the severity of ILD varies considerably, it represents the leading cause of death in SSc^{5,6}. Thus, an increased awareness of this complication is a real need, because it may affect prognosis, quality of life, and response to treatment. In particular, a sensitive and accurate method is desirable to detect ILD in its early stages. Early detection of ILD in SSc may improve prognosis and lead to better treatment-related outcomes.

To evaluate the presence of ILD in patients with SSc, there are different available tools in addition to clinical evaluation, including pulmonary functional tests (PFT) and imaging methods.

It has been found that the clinical manifestations were not present in the initial stages of the ILD. Moreover, PFT could

be unspecific despite an established ILD⁷. In this context, imaging may play a key role in the accurate detection of ILD.

Chest radiography has been widely used as a first imaging approach to assess ILD, but its very low sensitivity in early stages limits its current use as an assessment tool for early changes. High-resolution computed tomography (HRCT) is sensitive, and is the most common imaging technique used in the assessment of ILD. It has demonstrated utility for diagnosis, disease activity, and therapy monitoring of ILD^{8,9}. Further, it can detect both early pulmonary changes and subclinical lung involvement⁸. However, it has limited routine use because of high costs and ionizing radiation, in spite of new-generation HRCT machines that have considerably reduced the radiation dose.

It has been proposed that lung ultrasound (LUS) may have a role for the assessment of ILD in patients with autoimmune rheumatic diseases^{10,11,12,13,14}. The LUS assessment of ILD is determined by the detection and quantification of B-lines, which consist of “comet tails” — artifacts fanning out from the lung surface — generated by the reflection of the LUS beam from thickened subpleural interlobar septa detectable between the lung intercostal spaces.

Despite the growing body of evidence supporting the utility of LUS in ILD, its validity, reliability, feasibility, and a standardized approach have not been thoroughly established. Several authors have developed and published different LUS methods to assess for ILD-SSc, but they are limited to the local clinical settings^{10,11,12,13}.

To validate the use of LUS as an outcome measurement instrument in the evaluation of patients with ILD in rheumatic diseases, an OMERACT (Outcome Measures in Rheumatology) LUS Subtask Force was formed.

The purpose of this paper from this task force is to provide an overview of the potential role of LUS in the assessment of ILD-SSc based on a systematic literature review and to discuss the current evidence and state of validation supporting its clinical relevance and application in daily clinical practice.

MATERIALS AND METHODS

Literature review criteria and search strategy. All relevant literature in the field of LUS for detection of ILD in SSc in the last 20 years has been reviewed. We included original articles about studies in humans published between January 1997 and October 2017. To identify all available studies, a detailed search pertaining to the topic of review was conducted according to PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines¹⁵.

A systematic search was performed in the electronic databases (PubMed and EMBASE), using the following search terms in all possible combinations: ultrasound, sonography, ultrasonography, interstitial lung disease, interstitial fibrosis, interstitial pulmonary fibrosis, pulmonary fibrosis, systemic sclerosis, and scleroderma. In addition, the reference lists of all retrieved articles were manually reviewed. In cases of missing data, study authors were contacted by e-mail to try to retrieve original data. Two independent authors analyzed each article and performed the data extraction independently. In case of disagreement, a third investigator was consulted.

Discrepancies were resolved by consensus. Titles, abstracts, and complete reports of the included articles were systematically evaluated.

Inclusion and exclusion criteria. Studies that have been performed using LUS in ILD-SSc were included in the present review. We excluded from this review the following nonanalytic types of publications: review articles, articles not published in English, case reports, letters to the editor, comments, editorials, non-human studies, or abstracts from scientific meetings. Retrieved papers were screened to avoid duplicates. Titles, abstracts, and full reports of articles identified were systematically screened regarding inclusion and exclusion criteria.

The quality assessment of retrieved articles was performed according to the Oxford Center for Evidence-based Medicine¹⁶. The methodological quality of the studies was assessed using the Cochrane Handbook for Systematic Reviews¹⁷ and the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies)¹⁸.

Data extraction. The following data were extracted using a template designed for this study: type and design of the study, number of patients, number of controls, comparative diagnostic methods, and aspects focused on the LUS measurements and technique, outcome domains, measures, content, criterion and construct validity, discrimination, and reliability.

RESULTS

About 300 publications were identified in PubMed and EMBASE databases between January 1997 and October 2017. From the 300 articles identified, after excluding the mentioned nonanalytic types of publications, 12 were finally included for further analysis (Figure 1).

Table 1 reports included studies, type of study, number of patients enrolled, methods of comparison, and variables analyzed (including LUS scoring systems used).

General characteristics of included studies. All 12 papers included were observational, cross-sectional, and/or descriptive studies^{10,11,12,13,14,19-25}.

No randomized controlled clinical trials or studies including a cohort followed prospectively or longitudinally to evaluate the progression of ILD were found. Three studies were performed using a control group and 11 studies (92%) used the HRCT as an imaging method comparator (Table 1). A total of 635 patients with SSc were recorded, with a median number of 36.5 patients per study (range 31.5–54.7). There were more women than men (82% vs 18%), with a median of 5.3 years of disease duration. The majority of the patients were white and in the sixth decade of life. In most of the studies, the subtype of SSc and the results of the respiratory tests were not mentioned. More details on the clinical characteristics of the patients included in the review are reported in Appendix 1.

The primary aim of all the studies was to determine the correlation between LUS and HRCT findings in detecting pulmonary fibrosis. In all the 12 included studies the LUS examination was performed by B-mode. No study reported the assessment by power Doppler technique.

Most of the articles (92%) included the B-lines as the main LUS finding for ILD, whereas a smaller number reported on pleural irregularities (Table 1). Several US B-lines scoring systems were reported: some were dichotomous (34%), others quantitative (16%) or semiquantitative (50%).

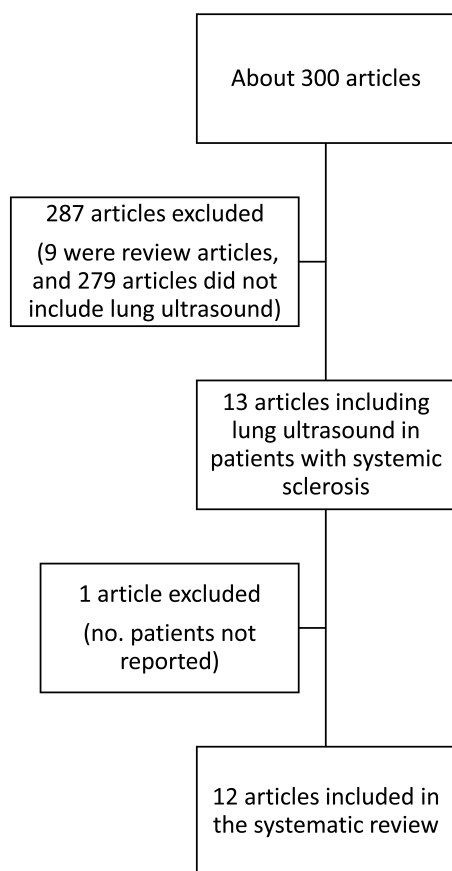


Figure 1. Flowchart of the review.

The US scanning protocol adopted by all the studies was based on the evaluation of intercostal spaces. The patient position was also similar in all studies. The patient was in the supine position for anterior and lateral scan and in a sitting position for posterior or dorsal scan (Table 2).

There was great variability in selecting the transducer for the US lung examination. Linear, convex, and cardiac transducers were used. A frequency of 3.5–5 MHz was generally used for the convex transducer, whereas the frequency varied from 8–11 MHz when the transducer was linear. Finally, only 3 studies reported that the sonographer was blinded to the patient's clinical data (Table 2).

Quality assessment of retrieved articles. All studies were classified as 2b level of evidence, according to the PRISMA guidelines for levels of evidence.

Ninety-two percent of the studies^{5,7,8,9,12-18} showed a low risk of bias; only one⁶ was judged as high risk of bias in the patient's selection section (Figure 2A). Regarding applicability, all the studies demonstrated low risk of bias (Figure 2B).

Criterion validity/construct validity. Because LUS was never tested against the external gold standard (lung histology) in any previous human study in SSc, it does not meet this aspect of validation. As an alternative, correlations

with other validated measures were searched, to estimate the concurrent and convergent validity as surrogates for criterion validity and as indicators of overall construct validity.

A total of 11 studies (92%) applied HRCT as the gold standard; in 7 of these studies (58%) the Warrick score was the HRCT score adopted for the correlation with LUS findings²⁶. Four out of 12 (33.3%) included also the PFT in addition to HRCT as a surrogate gold standard. Accuracy (sensitivity and specificity) data are reported in Table 3.

All studies demonstrated a positive correlation between LUS B-lines and HRCT in the assessment of ILD. However, these results were not confirmed by a multivariate analysis. **Discrimination.** Insufficient data were provided in the analyzed studies to assess the reliability and reproducibility of the LUS in ILD in patients with SSc. Only 3 studies (25%) performed intra- or interobserver reliability including κ coefficient. However, because these few tests indicated reproducibility, it was rated partially validated. None of the studies evaluated the sensitivity to change.

Moreover, no studies aimed to demonstrate the predictive validity regarding prognostic value (Table 4).

Feasibility. We found that 2 studies reported the time used to examine the lung by LUS, which may range from 6 to 31 min according to the severity of the ILD or the type of scanning technique (Table 4).

We could not find specific data on the day-to-day issues of feasibility, accessibility, or cost-effectiveness. Currently the number of intercostal spaces examined in the studies is highly variable, ranging from 10 to 72 per patient^{11,12,13,14}. Nevertheless, we found good evidence that LUS was available in medical centers, and the patient/physician acceptability was good.

DISCUSSION

This is the first systematic review addressing validity of LUS as an outcome measure in ILD-SSc, to our knowledge. Current evidence suggests that LUS passed the filter regarding face and content validity and feasibility. However, no validated or robust data allow full confirmation of criterion validity, reliability, and sensitivity to change (Table 3 and Table 4).

There have been interesting initiatives to promote new applications of US in rheumatology^{27,28,29}. Because of the increased competency and experience of the sonographers, and the availability of high-end equipment, preliminary data regarding the applications of US in lung disease are also available.

Overall the literature search showed encouraging results. However, some crucial points should be addressed before using LUS as a validated instrument for the assessment of ILD-SSc. First, no consensual definitions were used for the elementary lesions to evaluate during the examination. Second, we found a lack of information on the LUS proce-

Table 1. Included studies, type of study, number of patients enrolled, methods of comparison, and scoring systems used.

Reference	Type of Study	Patients	Comparison with Other Diagnostic Method	Name (domain) and US Definition	Variables	System of Measure
Gutierrez ¹⁴	O, P, C	36	HRCT	B-line, hyperechoic narrow-based reverberation type of artifact, spreading like a laser ray up to the edge of the screen		Semiquantitative score
Aghdashi ¹⁹	O, P, C	31	HRCT	B-line, hyperechoic narrow-based reverberation artifact generally not visible in normal lung parenchyma		Dichotomous (> 5 = positive results)
Barskova ²⁰	O, P, C	58	HRCT	B-line defined as an echogenic coherent wedge-shaped signal with a narrow origin in the near field of the image		Dichotomous (> 5 = positive results)
Buda ²¹	O, P, C, Co	52	HRCT	Am Line: subpleural, horizontal, and numerous reverberation artifacts, arising from pleural line and running to edge of screen, wide at the base and narrow at the top. Consolidations are hypogenic, usually wedge-shaped, rarely round or oval		Semiquantitative score
Gargani ¹¹	O, P, C	33	HRCT	Ultrasound lung comet sign was defined as an echogenic, coherent, wedge-shaped signal with a narrow origin in the near field of the image		Dichotomous (> 10 = positive results)
Gigante ²²	O, C	39	HRCT	B-line: discrete laserlike vertical hyperechoic reverberation artifact that arises from the pleural line extending to the bottom of the screen without fading, moving synchronously with lung sliding		Dichotomous (≥ 3 B-lines in at least 2 adjacent scanning sites or when a total of ≥ 5 B-lines were recorded = positive results)
Moazedi-Fuerst ²³	O, P, C, Co	25	NR	A-lines: repetitive horizontal reverberation artifacts that arise from the pleural line and are generated by subpleural air. B-lines: vertical artifacts arising from the pleural line and projecting the coexistence of elements with a major acoustic impedance gradient. Pleural irregularities: irregularities of the pleural line more than 2.8 mm		Semiquantitative score (B lines: 1–5 = score 1, > 5 = score 2; pleural irregularities: 1–5 areas = score 1, > 5 = score 2)
Pinal-Fernandez ²⁴	O, P, C	37	HRCT	B-line: a vertical hyperechoic artifact perpendicular to the pleural line, extending to the edge of the sonographic window. Pleural irregularity: loss of the normal hyperechoic linear pleural contour		Quantitative
Sperandeo ¹²	O, P, C	175	HRCT	Pleural thickening NR		Quantitative
Tardella ¹³	O, P, C	34	HRCT	B-lines: hyperechoic narrow-based reverberation type of artifact, spreading like a laser ray up to the edge of the screen		Semiquantitative (grade 0 or normal < 10 B-lines; grade 1 or mild = 11 to 20 B-lines; grade 2 or moderate = 21 to 50 B-lines, and grade 3 or marked > 50 B-lines)
Moazedi-Fuerst ¹⁰	O, P, C, Co	45	HRCT	Reverberation artifacts: repetitive horizontal artifacts that arise from the pleural line and are generated by subpleural air. B-lines/B-pattern: vertical artifacts arising from the pleural line. Pleural line: hyperechoic structure created by the parietal and visceral pleura. Thickening pleural: irregularities of the pleural line more than 3 mm observed in any scanned area		Semiquantitative (1–5 positive areas = comet-score of 1 and > 5 abnormal areas = comet-score of 2).
Mohammadi ²⁵	O, P, C	70	HRCT	B-line (NR)		Semiquantitative: 0 = normal (≤ 5 B-lines), 1 = mild (from 6 to 15 B-lines), 2 = moderate (from 16 to 30 B-lines), and 3 = severe (> 30 B-lines)

US: ultrasound; NR: not registered; O: observational; P: prospective; C: comparative; Co: control group; HRCT: high-resolution computed tomography.

dures of image acquisition. There is a crucial need to standardize the scanning technique and the approach for the LUS assessment of the lung as well as how many areas should be scanned (i.e., how many intercostal spaces should

be evaluated). Currently the number of intercostal spaces reported in the studies is highly variable, ranging from 10 to 72 per patient^{11,12,13,14}. Third, there is not a consensus on how to quantify the ILD by LUS — by a dichotomy approach or

Table 2. Technical aspect and characteristics of ultrasound machines.

Reference	US Mode	Transducer, and Technical Specifications	No. Sites Evaluated	Scanning Protocol and Position of the Patient	US Image Reader Blinded
Gutierrez ¹⁴	GS	Convex probe, MyLab 70 XVG (Esaote S.p.A.), 2–7 MHz	50 and 14	Standardized technique. Patient positions were supine or near-supine for the anterior chest scanning, while in a sitting position for the posterior chest scanning.	Yes
Aghdashi ¹⁹	GS	Linear probe, Siemens sonoline G-40 (Siemens), 7–10 MHz	10	Standardized technique. Patients were examined in supine position for assessment of anterior chest wall and in sitting position for assessment of posterior chest wall.	NR
Barskova ²⁰	GS	Cardiac sector transducer 2.5 cm in length (Mylab50, Esaote), 2.5–3.5 MHz	72	Standardized technique. Patients in the supine position for anterior and lateral scanning, and in the sitting position for dorsal scanning.	No
Buda ²¹	GS	Linear probe 8–11MHz, convex probe 3.5–5MHz, Logiq 7 system (GE Healthcare)	NR	Standardized technique. Patients remaining in the sitting and supine position.	NR
Gargani ¹¹	GS	Convex probe, cardiac sector transducer (2.5 cm long), Mylab25 (Esaote), 2.5–3.5 MHz	NR	Standardized technique. Patients in the supine or near-supine position for the anterior scanning, and in the sitting position for the dorsal scanning.	Yes
Gigante ²²	GS	Convex probe, Toshiba Ultrasound System, 2.5 to 3.5 MHz	NR	Standardized technique. Position of patients NR.	No
Moazedi-Fuerst ²³	GS	Convex probe 3.5 MHz, linear probe NR	NR	Standardized technique. The anterior pleural surface was investigated in a supine position while the lateral and posterior surfaces were scanned in a sitting position.	NR
Pinal-Fernandez ²⁴	GS	Linear probe. MyLabTwice system, 5 MHz	72	Standardized technique. Patients in supine position to record the anterior and anterolateral sonographic points and in sitting position for the posterior and posterolateral ones.	NR
Sperandeo ¹²	GS	Convex probe, 3.5–5 MHz	NR	Protocol and patient position NR.	NR
Tardella ¹³	GS	Convex probe, MyLab 70 XVG (Esaote Biomedica) equipped 2–7 MHz	50	Patients in the supine or near-supine position (with the arms elevated and hands clasped behind the neck) for anterior and lateral scanning, and in the sitting position (with the arms along the trunk) for posterior scanning.	Yes
Moazedi-Fuerst ¹⁰	GS	Convex probe and linear probe, 3.5 MHz	18	Standardized technique. The anterior pleural surface was investigated in a supine position while the lateral and posterior surface was scanned in a sitting position.	NR
Mohammadi ²⁵	GS	Linear probe, Medison Accuvix V20, 7–10 MHz	10	Standardized technique. Patients were examined in supine position for assessment of anterior chest wall and in sitting position for the posterior chest wall.	NR

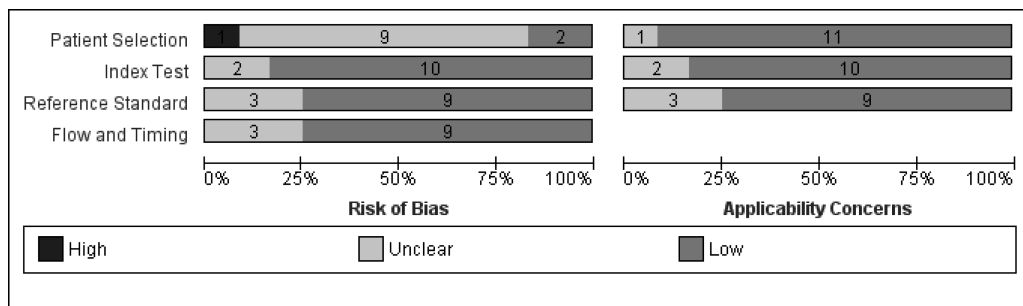
US: ultrasound; GS: greyscale; NR: not registered.

using quantitative or semiquantitative scoring systems. The problem is that there are different LUS B-lines scoring systems including different cutoffs to interpret the degree of ILD. Fourth, there was no agreement on the measurement to use (i.e., scoring systems), as well as the cutoff of normality. Fifth, there is no consensus regarding what the optimal US transducer is to use in the assessment of the lung. Although small surface probes with frequencies ranging between 3 and 3.5 MHz seemed suitable for this specific purpose, transducers with large surfaces and frequencies between 5 and 7.5 MHz were also used³⁰. Fifth, there were no studies including a cohort in which all newly ILD diagnosed by LUS were followed prospectively or longitudinally to see the longterm development. Finally, in general, the studies offered minimal to no information regarding how well LUS performs in the detection of early ILD. Only 1 study²⁰ was performed in very early SSc patients with mean of disease duration \pm SD of 1.9 ± 3.2 . The authors reported a sensitivity of 100% for the

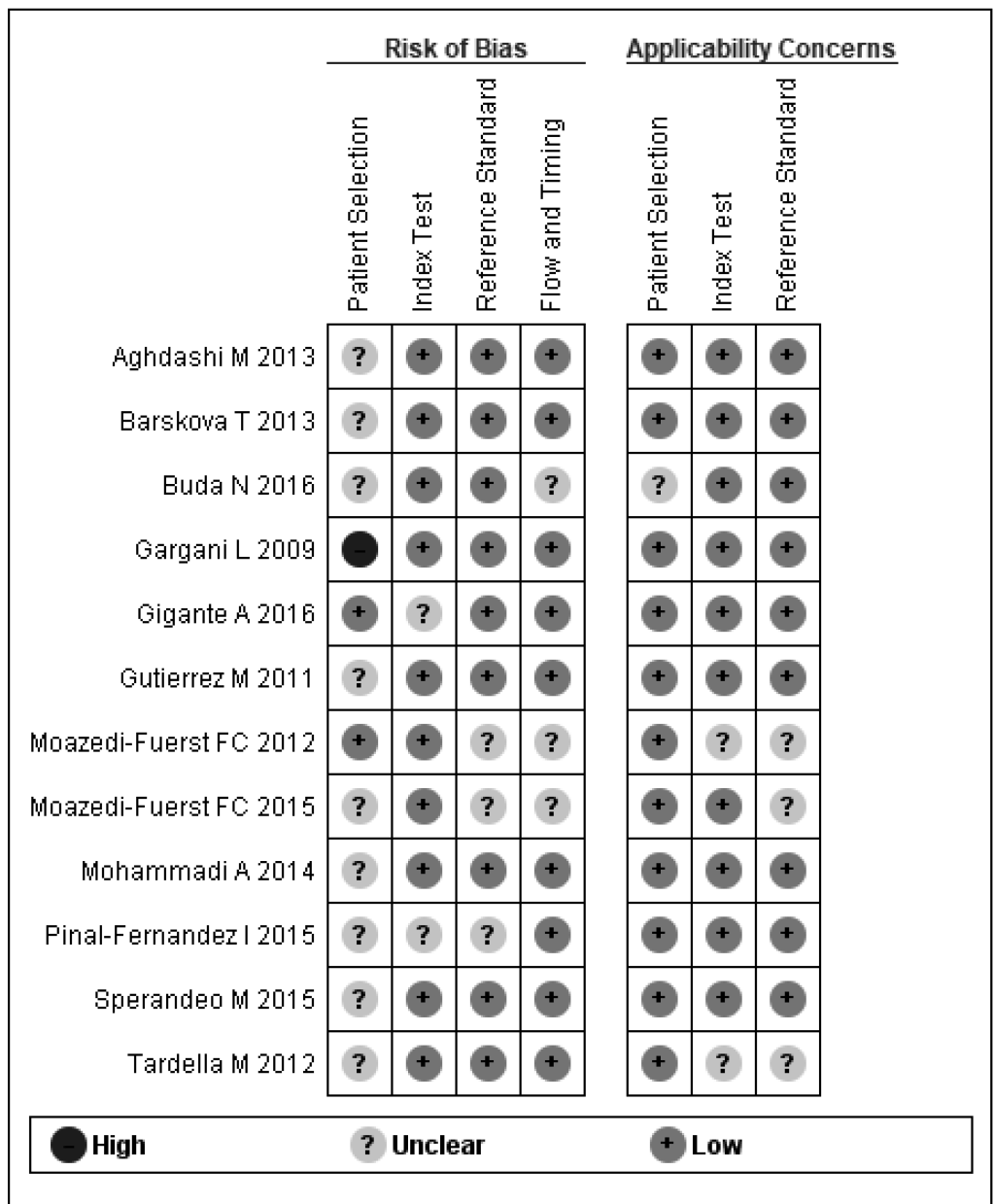
screening of ILD by LUS. These results may represent the basis to exploring the potential of LUS as a screening tool for the early detection of ILD-SSc. In this light, we recently conducted a study with the aim of determining diagnostic value of LUS in detecting subclinical ILD in 133 patients with SSc. We found that 40.6% of patients with SSc showed LUS signs of subclinical ILD in contrast to healthy controls (4.8%; $p = 0.0001$). Sensitivity and specificity of US in detecting ILD were 91.2% and 88.6%, respectively³¹.

This literature review revealed several aspects of LUS that need further validation (criterion/construct validity, reliability, and sensitivity to change), revealing a clear research agenda that needs to be addressed in the near future.

Definite validation of criterion validity of the LUS requires lung histology as a gold standard. To date there are no human studies using histology as the gold standard. However, previous studies performed in animal models showing a good correlation between number of B-lines and water level in pulmonary



A



B

Figure 2. Quality assessment of papers. A. Global risk of bias and applicability concerns. B. Risk of bias and applicability concerns for each paper.

Table 3. Validity of the studies included.

Reference	Construct	Validity	Criterion	Comparative Instrument	Sensitivity and Specificity, %	US Results
Gutierrez ¹⁴	Simplified assessment of the semiquantitative score [0 = normal (< 5 B-lines); 1 = mild (from 6 to 15 B-lines); 2 = moderate (from 16 to 30 B-lines); and 3 = marked (> 30 B-lines)]		Correlation between Warrick score and simplified assessment of the semiquantitative score	HRCT	NR	A positive correlation found between the US B-line assessment and Warrick score HRCT assessment in simplified method ($p = 0.0006$).
Aghdashi ¹⁹	Comet tails scoring system		Sensitivity, specificity, positive and negative predictive value of TTUS	HRCT	73.85 and 88.23	The sensitivity, specificity, positive and negative predictive values of TTUS were 73.58%, 88.23%, 95.12%, and 51.72%, respectively.
Barskova ²⁰	A scan was considered positive either when ≥ 3 B-lines were found in at least 2 adjacent scanning sites or when B-line total was > 5		Correlation between LUS and HRCT	HRCT	100 and 55	Individual patient analysis between LUS and HRCT showed a concordance between the 2 examinations of 83% in the overall population, with a sensitivity of 100%, negative predictive value of 100%, specificity of 55%, and positive predictive value of 78%.
Buda ²¹	Ultrasound Alveolar Index*. UFI: 3 to 35 points could be obtained. Mild pulmonary fibrosis: UFI is 3–14 points; moderate 15–20 points. Severe pulmonary fibrosis in LUS, 21–35 UFI points		Warrick score (FI): mild = FI < 8 points; moderate = FI 8–15 points; severe = FI ≥ 15 points.	HRCT	NR	Mild pulmonary fibrosis occurs in 24%, 12/52. Moderate in 38%, 20/52. Severe pulmonary fibrosis in 38%, 20/52.
Gargani ¹¹	ULC = absent ≤ 10 , present > 10		Warrick score	HRCT	NR	ULC were absent (< 10) in 16 patients and present (> 10) in 17. A significant positive linear correlation was found between echographic ULC score and Warrick score ($r = 0.72$; $p < 0.001$).
Gigante ²²	Scan positive either when > 3 B-lines present in at least 2 adjacent scanning sites or when a total of > 5 B-lines recorded		Warrick score	HRCT	NR	The mean number of B-lines is 29.1 ± 21.8 and the mean HRCT score is 9.5 ± 6.4 . A positive correlation exists between the number of B-lines and HRCT score ($r = 0.81$, $p = 0.0001$).
Moazedi-Fuerst ²³	Comet score system: 1 to 5 positive areas received a comet score of 1, and patients with > 5 abnormal areas had comet score of 2		NR	NR	NR	The median thickness of the pleural irregularities > 3.2 mm compared to the 1.3 mm in the volunteer group ($p < 0.001$). Nine patients with SSc (36%) had more than 2.8 mm of pleural thickness, which was declared as a cutoff. Pleural nodules sonographically observed in only 1 patient (4%) with SSc.
Pinal-Fernandez ²⁴	PI		Warrick score	HRCT	NR	PI detected in 28.9% (SD 20.2%) of US points. Mean Warrick score 16.1 (SD 8.6). The PI score correlated with the Warrick ($r = 0.63$, $p = 0.01$). The area under the ROC curve to detect ILD for the PI score 0.85 (95% CI 0.64–1); higher, but not significantly, than the B-line score (AUC = 0.65, 95% CI 0.32–0.98).
Sperandeo ¹²	Ultrasound pleural line thickness between 3.0 and 5.0 mm		HRCT reticular-nodular pattern	HRCT	80 and 99	Sensitivity 80.0% and specificity 99.0% for the HRCT reticular-nodular pattern.
Tardella ¹³	Semiquantitative score		Warrick score	HRCT	NR	A significant linear correlation was found between the US and the HRCT scores ($p < 0.001$; coefficient of rank correlation, $\kappa = 0.875$).
Moazedi-Fuerst ¹⁰	Semiquantitative score: comet score of 0 assigned to patients without positive areas; patients with 1–5 positive areas received a comet-score of 1, and patients with > 5 abnormal areas, comet score of 2		NR	HRCT	NR	Comet-tail artifacts/B-patterns present in all patients (100%) of the ILD group ($n = 20$) but in only 12% of patients with normal CT scans ($n = 25$, $p < 0.001$). Subpleural nodes observed in 55% ($n = 11$) of the ILD patients compared to 17% ($n = 4$) of patients without radiological signs of ILD ($p = 0.006$). Ninety-five percent of the ILD patients vs 5% of the non-ILD group showed pleural irregularities > 3 mm on thoracic US ($p < 0.001$). In healthy volunteers, B-lines observed in 3 patients (7%) and pleural noduli in 1 patient (2%). Intraarticular power Doppler signal gout 5/60 (8%), and CPPD 6/140 (4%) in knees.

Table 3. Continued.

Reference	Validity Construct	Criterion	Comparative Instrument	Sensitivity and Specificity, %	US Results
Mohammadi ²⁵	ULC assessment scored semiquantitatively as 0 = normal (≤ 5 B-lines), 1 = mild (6 to 15 B-lines), 2 = moderate (16 to 30 B-lines), and 3 = severe (> 30 B-lines)	Warrick score	HRCT	73.85 and 88.23	ULC assessment, compared to Warrick score, a significant positive correlation for severity of pulmonary involvement appreciation (Spearman's correlation coefficient = 0.695, $p < 0.001$; LR = 74.36, $p < 0.001$) was found. The sensitivity, specificity, positive and negative predictive values of TTUS were 73.58%, 88.23%, 95.12%, and 51.72%, respectively.

* Ultrasound Alveolar Index: assesses the degree of progression of the active changes in the lungs — from 2 to 4 points could be obtained. US: ultrasound; LUS: lung US; HRCT: high-resolution computed tomography; NR: not registered; ULC: US lung comets; ILD: interstitial lung disease; PI: pleural irregularity; UFI: Ultrasound Fibrosis Index; CPPD: calcium pyrophosphate deposition disease; TTUS: transthoracic lung US; SSc: systemic sclerosis; ROC: receiver-operating characteristic (curve); AUC: area under the curve; LR: likelihood ratio.

Table 4. Reliability, feasibility, and sensitivity to change.

Reference	Interobserver	Discrimination Intraobserver	Between-group Differences	Sensitivity to Change	Feasibility
Gutierrez ¹⁴	κ of comprehensive US semiquantitative assessment at parasternal, mid-clavicular, anterior axillary, mid-axillary, paravertebral, subscapular and posterior axillary level: 0.943, 0.846, 0.963, 0.932, 0.958, 0.969, and 0.980, respectively	2nd parasternal $\kappa = 0.864$, 4th mid-clavicular $\kappa = 0.881$, 4th anterior axillary $\kappa = 0.868$, 4th mid-axillary $\kappa = 0.845$, 8th paravertebral $\kappa = 0.894$, 8th subscapular $\kappa = 0.883$, 8th posterior axillary $\kappa = 0.862$	NR	NR	Estimated by comparing time spent regarding comprehensive assessment by independent samples t test. Significant difference found between comprehensive US B-lines assessment (mean $23.3 \pm SD 4.5$, range 16–31 min) and simplified US B-lines assessment (mean $8.6 \pm SD 1.4$, range 6–12 min, $p < 0.00001$).
Aghdashi ¹⁹	NR	NR	NR	NR	NR
Barskova ²⁰	NR	NR	NR	NR	Time needed for the scan and analysis always < 10 min
Buda ²¹	NR	NR	NR	NR	NR
Gargani ¹¹	NR	NR	NR	NR	NR
Gigante ²²	NR	NR	NR	NR	NR
Moazedi-Fuerst ²³	NR	NR	NR	NR	NR
Pinal-Fernandez ²⁴	NR	NR	NR	NR	NR
Sperandeo ¹²	NR	NR	NR	NR	NR
Tardella ¹³	Parasternal $\kappa = 0.943$, mid-clavicular $\kappa = 0.846$, anterior-axillary $\kappa = 0.963$, medial-axillary $\kappa = 0.932$, paravertebral $\kappa = 0.958$, subscapular $\kappa = 0.969$, posterior-axillary $\kappa = 0.980$	NR	NR	NR	NR
Moazedi-Fuerst ¹⁰	NR	NR	NR	NR	NR
Mohammadi ²⁵	Global κ value of agreement between 2 imaging methods 0.553 ($p < 0.001$).	Global κ values for intraobserver reliability of B-lines assessment 0.838	No	NR	NR

US: ultrasound; NR: not registered.

edema suggested that LUS could be a noninvasive and simple method to detect and quantify ILD in rheumatic disorders³².

Validation of reliability of the LUS in ILD-SSc requires

comparisons of repeated LUS assessments performed within a short time by the same investigator (intraobserver variability) and by 2 independent investigators (interobserver

variability) at the same time in patients with well-defined ILD-SSc.

To obtain more accurate and reliable information on the sensitivity and specificity, as well as the reproducibility of the lung US, additional studies are needed, which ideally must include a higher number of patients showing a full clinical spectrum of ILD-SSc. Additionally, the type of studies required to assess the validity of lung US regarding the sensitivity to change is longitudinal studies, including patients with ILD-SSc with and without treatment and parallel lung US and HRCT evaluations at different timepoints.

We are aware of limitations associated with the present review: the small number of articles found, and that the results described are based only on published studies in peer-reviewed journals and published in English. Another important limitation of our study is that many of the articles included had a small number of samples ($n < 40$), which decreases the external validity of the articles included. Finally, studies of LUS assessing other forms of ILD were not included, and including them would have extended the number of suitable papers and provided much information regarding the utility of LUS in other types of ILD.

Despite a great deal of work supporting the potential role of LUS for the assessment of ILD-SSc, too much remains to be done to validate its use as an outcome measure in ILD-SSc. In particular, future research should focus on validity of LUS in detecting ILD in the early stages, its accuracy in assessing the eventual response to therapy, the correct timing of LUS for diagnosis and followup, and its potential in monitoring the progression of ILD-SSc. Additionally, the research agenda should focus on promoting the development of consensus on definitions of elementary LUS lesions for ILD and on protocols of image acquisition as well as quantification of LUS findings for ILD.

REFERENCES

1. Michelfelder M, Becker M, Riedlinger A, Siegert E, Drömann D, Yu X, et al. Interstitial lung disease increases mortality in systemic sclerosis patients with pulmonary arterial hypertension without affecting hemodynamics and exercise capacity. *Clin Rheumatol* 2017;36:381-90.
2. Adler S, Huscher D, Siegert E, Allanore Y, Cziráj L, DelGallo F, et al. Systemic sclerosis associated interstitial lung disease — individualized immunosuppressive therapy and course of lung function: results of the EUSTAR group. *Arthritis Res Ther* 2018;20:17.
3. Salaffi F, Carotti M, Baldelli S, Bichi Secchi E, Manganelli P, Subiaco S, et al. Subclinical interstitial lung involvement in rheumatic diseases. Correlation of high resolution computerized tomography and functional and cytologic findings. *Radiol Med* 1999;97:33-41.
4. Manganelli P, Salaffi F, Pesci A. Clinical and subclinical alveolitis in connective tissue diseases assessed by bronchoalveolar lavage. *Semin Arthritis Rheum* 1997;26:740-54.
5. Maradit-Kremers H, Nicola PJ, Crowson CS, Ballman KV, Gabriel SE. Cardiovascular death in rheumatoid arthritis: a population-based study. *Arthritis Rheum* 2005;52:722-32.
6. Bernatsky S, Boivin JF, Joseph L, Manzi S, Ginzler E, Gladman DD, et al. Mortality in systemic lupus erythematosus. *Arthritis Rheum* 2006;54:2550-7.
7. Le Gouellec N, Duhamel A, Perez T, Hachulla AL, Sobanski V, Faivre JB, et al. Predictors of lung function test severity and outcome in systemic sclerosis-associated interstitial lung disease. *PLoS One* 2017;12:e0181692.
8. Molberg Ø, Hoffmann-Vold AM. Interstitial lung disease in systemic sclerosis: progress in screening and early diagnosis. *Curr Opin Rheumatol* 2016;28:613-18.
9. Bernstein EJ, Khanna D, Lederer DJ. Screening high resolution computed tomography of the chest to detect interstitial lung disease in systemic sclerosis: a global survey of rheumatologists. *Arthritis Rheumatol* 2018;70:971-2.
10. Moazedi-Fuerst FC, Kielhauser S, Brickmann K, Tripolt N, Meilinger M, Lufti A, et al. Sonographic assessment of interstitial lung disease in patients with rheumatoid arthritis, systemic sclerosis and systemic lupus erythematosus. *Clin Exp Rheumatol* 2015;33(4 Suppl 91):S87-91.
11. Gargani L, Doveri M, D'Errico L, Frassi F, Bazzichi ML, Delle Sedie A, et al. Ultrasound lung comets in systemic sclerosis: a chest sonography hallmark of pulmonary interstitial fibrosis. *Rheumatology* 2009;48:1382-7.
12. Sperandeo M, De Cata A, Molinaro F, Trovato FM, Catalano D, Simeone A, et al. Ultrasound signs of pulmonary fibrosis in systemic sclerosis as timely indicators for chest computed tomography. *Scand J Rheumatol* 2015;44:389-98.
13. Tardella M, Gutierrez M, Salaffi F, Carotti M, Ariani A, Bertolazzi C, et al. Ultrasound in the assessment of pulmonary fibrosis in connective tissue disorders: correlation with high-resolution computed tomography. *J Rheumatol* 2012;39:1641-7.
14. Gutierrez M, Salaffi F, Carotti M, Tardella M, Pineda C, Bertolazzi C, et al. Utility of a simplified ultrasound assessment to assess interstitial pulmonary fibrosis in connective tissue disorders—preliminary results. *Arthritis Res Ther* 2011;13:R134.
15. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
16. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al; OCEBM Table of Evidence Working Group. Oxford Centre for Evidence-Based Medicine 2011 levels of evidence. [Internet. Accessed October 8, 2019.] Available from: www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf
17. O'Connor D, Green S, Higgins JP. Defining the review question and developing criteria for including studies. In: *Cochrane Handbook for Systematic Reviews of Interventions - Cochrane Book Series*. London: The Cochrane Collaboration; 2008;81-94.
18. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
19. Aghdashi M, Broofeh B, Mohammadi A. Diagnostic performances of high resolution trans-thoracic lung ultrasonography in pulmonary alveoli-interstitial involvement of rheumatoid lung disease. *Int J Clin Exp Med* 2013;6:562-6.
20. Barskova T, Gargani L, Guiducci S, Randone SB, Bruni C, Carnesecchi G, et al. Lung ultrasound for the screening of interstitial lung disease in very early systemic sclerosis. *Ann Rheum Dis* 2013;72:390-5.
21. Buda N, Piskunowicz M, Porzezińska M, Kosiak W, Zdrojewski Z. Lung ultrasonography in the evaluation of interstitial lung disease in systemic connective tissue diseases: criteria and severity of pulmonary fibrosis – analysis of 52 patients. *Ultraschall Med* 2016;37:379-85.
22. Gigante A, Rossi Fanelli F, Lucci S, Barilaro G, Quarta S, Barbano

- B, et al. Lung ultrasound in systemic sclerosis: correlation with high-resolution computed tomography, pulmonary function tests and clinical variables of disease. *Intern Emerg Med* 2016; 11:213–17.
23. Moazedi-Fuerst FC, Zechner PM, Tripolt NJ, Kielhauser SM, Brickmann K, Scheidl S, et al. Pulmonary echography in systemic sclerosis. *Clin Rheumatol* 2012;31:1621–5.
 24. Pinal-Fernandez I, Pallisa-Núñez E, Selva-O'Callaghan A, Castella-Fierro E, Simeon-Aznar CP, Fonollosa-Pla V, et al. Pleural irregularity, a new ultrasound sign for the study of interstitial lung disease in systemic sclerosis and antisynthetase syndrome. *Clin Exp Rheumatol* 2015;33:136–41.
 25. Mohammadi A, Oshnoei S, Ghasemi-Rad M. Comparison of a new, modified lung ultrasonography technique with high-resolution CT in the diagnosis of the alveolo-interstitial syndrome of systemic scleroderma. *Med Ultrason* 2014;16:27–31.
 26. Warrick JH, Bhalla M, Schabel SI, Silver RM. High resolution computed tomography in early scleroderma lung disease. *J Rheumatol* 1991;18:1520–8.
 27. Gutierrez M, Wortsman X, Filippucci E, De Angelis R, Filosa G, Grassi W. High-frequency sonography in the evaluation of psoriasis: nail and skin involvement. *J Ultrasound Med* 2009;28:1569–74.
 28. Jousse-Joulin S, Nowak E, Cornec D, Brown J, Carr A, Carotti M, et al. Salivary gland ultrasound abnormalities in primary Sjögren's syndrome: consensual US-SG core items definition and reliability. *RMD Open* 2017;3:e000364.
 29. Naredo E, Möller I, Gutiérrez M, Bong DA, Cobo T, Corominas H, et al. Multi-examiner reliability of automated radio frequency-based ultrasound measurements of common carotid intima-media thickness in rheumatoid arthritis. *Rheumatology* 2011;50:1860–4.
 30. Delle Sedie A, Doveri M, Frassi F, Gargani L, D'Errico G, Pepe P, et al. Ultrasound lung comets in systemic sclerosis: a useful tool to detect lung interstitial fibrosis. *Clin Exp Rheumatol* 2010;28:S54.
 31. Gutierrez J, Gutierrez M, Almaguer K, Gonzalez F, Camargo K, Soto C, et al. OP0141 Ultrasound diagnostic and predictive value of interstitial lung disease in systemic sclerosis. Diagnostic and predictive value of ultrasound in the assessment of interstitial lung disease. *Ann Rheum Dis* 2018;77:121.
 32. Gargani L, Lionetti V, Di Cristofano C, Bevilacqua G, Recchia FA, Picano E. Early detection of acute lung injury uncoupled to hypoxemia in pigs using ultrasound lung comets. *Crit Care Med* 2007;35:2769–74.

APPENDIX 1. Clinical characteristics of the study populations included in the review.

Reference	Age, Yrs, Mean ± SD	Disease Duration, Mos, Mean ± SD	Sex	Race or Ethnicity	Country	SSc Subtype, n patients	FVC, %, Mean ± SD	FEV1/ FVC Ratio, %, Mean ± SD	DLCO%, Mean ± SD	Autoantibody Status
Gutierrez ¹⁴	57 ± 13	88 ± 83.1	NR	White	Italy	NR	NR	NR	NR	NR
Aghdashi ¹⁹	48.29 ± 9.7	68 ± 23.1	26 W, 5 M	Arab	Iran	NR	NR	NR	NR	NR
Barskova ²⁰	51 ± 14	70.8 ± 111.8	54 W, 4 M	White	Italy		77 ± 13	85 ± 10	63 ± 19	Scl-70 positivity 62%
Buda ²¹	50 ± 24	NR	20 W, 32 M	White	Poland	NR	NR	NR	NR	NR
Gargani ¹¹	53 ± 14	96 ± 84	29 W, 4 M	White	Italy	dcSSc 10, lcSSc 23	82 ± 19	NR	66 ± 22	Scl-70 positivity 27.2%
Gigante ²²	51 ± 15.2	102 ± 75.6	33 W, 6 M	White	Italy	NR	101.7 ± 18	NR	73.6 ± 16.5	Scl-70 positivity 29%, ACA positivity 25.6%
Moazedi- Fuerst ²³	51 ± 15	NR	23 W, 2 M	White	Austria	NR	NR	NR	NR	Scl-70 positivity 15%, ACA positivity 10%
Pinal- Fernandez ²⁴	51.5 ± 14.2	NR	24 W, 13 M	White	Spain	lcSSc 8, dcSSc 5, SSc sine scleroderma 3	NR	NR	NR	Scl-70 positivity 43.75%, ACA positivity 31.25%
Sperandeo ¹²	46.46 ± 15.33	90 ± 74.4	166 W, 9 M	White	Italy	lcSSc 38, dcSSc 137	NR	NR	NR	NR
Tardella ¹³	57.02 ± 12.96	85.58 ± 84.37	30 W, 2 M	White	Italy	NR	97.35 ± 26.37	NR	73.46 ± 21.03	NR
Moazedi- Fuerst ¹⁰	54	96	38 W, 7 M	White	Austria	NR	NR	NR	NR	NR
Mohammadi ²⁵	50.29 ± 9.7	88 ± 83.1	NR	Arab	Iran	NR	NR	NR	NR	NR

NR: not registered; FVC: forced vital capacity; FEV1/FVC: forced expiratory volume in 1 s/forced vital capacity; SSc: systemic sclerosis; ACA: anticentromere antibodies; lcSSc: limited cutaneous SSc; dcSSc: diffuse cutaneous SSc.