Title page

Title

Ultrasound in the assessment of interstitial lung disease in systemic sclerosis. A systematic literature review by the OMERACT Ultrasound Group.

Authors and affiliations
Marwin Gutierrez[1,2], Carina Soto-Fajardo[1] MD, Carlos Pineda[1], Alfonso Alfaro-Rodriguez[3], Lene Terslev[4], George Bruyn[5], Annamaria Iagnocco[6], Chiara Bertolazzi[1], Maria Antonietta D'Agostino[7] and Andrea delle Sedie[8]

[1]*Division of Musculoskeletal and Rheumatic Disorders. Instituto Nacional de Rehabilitacion, Mexico City, Mexico.*
[2] *Doctorado en Ciencias Biológicas y de la Salud, Universidad Autonoma Metropolitana, Iztapalapa, México City, México.*
[3] *Division of Neurosciencies. Instituto Nacional de Rehabilitacion, Mexico City, Mexico.*
[4] *Center for Rheumatology and Spine diseases, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.*
[5] *Department of Rheumatology, MC Groep, Lelystad, Netherlands*
[6] *Academic Rheumatology Center, Università degli Studi di Torino, Turin, Italy*
[7] *Université Versailles Saint-Quentin en Yvelines, Paris, France*
[8] *Rheumatology Unit, University of Pisa, Pisa, Italy.*

Academic degrees and ORCID of authors
Marwin Gutierrez MD, PhD. ORCID: https://orcid.org/0000-0002-7995-1110
Carina Soto-Fajardo MD. ORCID: http://orcid.org/0000-0001-6148-7789
Carlos Pineda MD. ORCID: http://orcid.org/0000-0003-0544-7461
Alfonso Alfaro-Rodriguez MD. ORCID: http://orcid.org/0000-0002-4599-1253
Lene Terslev MD, PhD.
George Bruyn MD, PhD.
Annamaria Iagnocco MD, PhD.
Chiara Bertolazzi MD, PhD. ORCID: https://orcid.org/0000-0002-4665-6596
Maria Antonietta D'Agostino MD, PhD. ORCID: http://orcid.org/0000-0002-5347-0060
Andrea delle Sedie MD, PhD. ORCID: https://orcid.org/0000-0002-7379-4732

Key indexing terms: systemic sclerosis, pulmonary ultrasound, interstitial lung disease, validity, reliability, feasibility.
Running title: ultrasound in systemic sclerosis
Word count: 4254

Correspondence to:

Marwin Gutierrez, M.D., PhD.
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
Tel. + (5255) 5999 1000 Ext. 19108-19109
E-mail: dr.gmarwin@gmail.com

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.

Conflict of interests

The authors would like to make the following statements with regard to their conflicts of interest/financial disclosures: **MG** has attended advisory board meetings, scientific consultancies and has obtained speaking fees for: AbbVie, Novartis, UCB, Esaote SpA, Janssen, Bristol-Myers Squibb, Merck Sharp & Dohme, Pfizer, Sanofi Aventis.
CS-F, CP, AA-R, LT, GB, AI, CB, MADA, ADS declare that they have no competing interests.

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 research pertaining to the topic of review was conducted according to PRISMA guidelines.
A systematic research was performed in PubMed and EMBASE. The Quality assessment of retrieved articles was performed according the Oxford Center for Evidence-based Medicine. The methodological quality of the studies was assessed using the Cochrane Handbook for Systematic Reviews and QUADAS-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. In conclusion, in spite of 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.

Introduction

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 diseases 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, which may affect prognosis, quality of life and response to treatment. In particular, a sensitive and accurate method is desirable in order 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 SSc patients there are different available tools in addition to clinical evaluation, including pulmonary functional tests (PFT) and imaging methods.

It was recently found that the clinical manifestations were not present in the initial stages of the ILD. Moreover, PFT could be unspecific in spite of an established ILD (7). In this scenario, imaging may play a key role in the accurate detection of ILD.

Chest X-ray has been widely used as first imaging approach to assess the ILD, but it's very low sensitivity in early stages limits the current use as assessment tool for early changes. High-resolution computer tomography (HRCT) is a sensitive and the most common imaging technique used in the assessment of ILD and has demonstrated utility for diagnosis, disease activity and therapy monitoring of ILD (8,9). Furthermore, it has shown ability to detect both early pulmonary changes and subclinical lung involvement (8). However, it has limited routine use due to high costs and ionizing radiation, in spite of new generation HRCT machines have reduced considerably the radiation dose.

Recently, it has been proposed that lung ultrasound (LUS) may have a role for the assessment of ILD in patients with autoimmune rheumatic diseases (10-14). The LUS assessment of ILD is determined by the detection and quantification of B-lines, which consist of "comet tails" – artefacts fanning out from the lung surface - generated by the reflection of the LUS beam from thickened sub-pleural interlobar septa detectable in-between the lung intercostal spaces.

Despite the growing body of evidence supporting the utility of LUS in ILD, validity, reliability, feasibility and 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 (10-13).

In order to validate the use of LUS as an outcome measurement instrument in the evaluation of patients with ILD in rheumatic diseases, an OMERACT - LUS Sub-Task 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.

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 concerning studies in humans, published between January 1997 and October 2017. To identify all available studies, a detailed

research pertaining to the topic of review was conducted according to PRISMA [Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines (15).

A systematic research 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 case of missing data, study authors were contacted by e-mail to try to retrieve original data. Two independent authors analysed 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 non-analytic 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 with regard to inclusion and exclusion criteria.

The Quality assessment of retrieved articles was performed according the Oxford Center for Evidence-based Medicine (16).

The methodological quality of the studies was assessed using the Cochrane Handbook for Systematic Reviews (17) and QUADAS-2 tool (18).

Data extraction

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

Results

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

Included studies, type of study, number of patients enrolled, methods of comparison, and variables analysed (including LUS scoring systems used) are reported in Table 1.

General characteristics of included studies

All 12 papers included were observational, cross-sectional and/or descriptive studies (10-14, 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 control group and 11 studies (92%) have used the HRCT as imaging method comparator (table 1). A total of 613 SSc patients were recorded, with a median of patient's number of 36.5 per study (range 31.5-54.7). There were more woman than men (82% vs 18%) with a median of 5.3 years of disease duration. The majority of the

patients were white 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 the supplementary table.

The primary aim of all studies was to determine the correlation between LUS and HRCT findings in detecting detect 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 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 scorings systems were reported: some of them were dichotomous (34%), other quantitative (16%) or semiquantitative (50%) scores.

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

There was a great variability in selecting the transducer for the US lung examination. Linear, convex, and cardiac transducers were indistinctly used. A frequency of 3.5-5 MHz was generally used for the convex transducer, whereas the frequency varies from 8-11MHz when the transducer was linear. Finally, only 4 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 guidelines for 'Levels' of evidence.

Ninety-two percent of the studies included (5, 7-9, 12-18) showed a low risk of bias; only one (6) was judged as high risk of bias in the patient's selection section (figure 2A). In terms of applicability all the studies demonstrated low risk of bias (figure 2B).

Criterion validity/construct validity

Since 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, correlation with other validated parameters 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 (92%) studies applied HRCT as gold standard; in 7 of these studies (58%) the Warrick score was the HRCT score adopted for the correlation with LUS findings [26].

Four out of 12 (42%) included also the PFT in addition to HRCT as 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 analysed studies to assess the reliability and reproducibility of the LUS in ILD in SSc patients. Only 3 studies (25%) performed intra or inter-observer reliability including kappa 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, in terms of prognostic value were conducted (table 4).

Feasibility

We found that 2 studies reported the time employed to explore the lung by LUS, which may range between 6 to 31 minutes according the severity of ILD or the type of scanning technique adopted (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 reported in the studies is highly variable ranging from 10 to 72 per patient (11-14). Nevertheless, we found good evidence that LUS was available in medical centres, 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. Current evidence suggests that LUS passed the filter in terms of 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 4).

In recent years there have been interesting initiatives to promote new applications of ultrasound in rheumatology (27-29). Due to the increased competency, experience of the sonographers, and to the availability of high-end equipment, preliminary data regarding the applications of ultrasonography in lung disease is also provided.

Overall the literature search showed encouraging results. However some crucial points should be addressed before using LUS as validated instrument for the assessment of ILD-SSc. First, no consensual definitions were used for defining the elementary lesions to evaluate during the examination. Second, we found a lack of information on the LUS procedures of images acquisition. There is a crucial need to standardized 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 (11-14). Third, there is not a consensus on how to quantify the ILD by LUS – by a dichotomy approach or using quantitative or semiquantitative scoring systems. The problem is that there are different LUS B-lines scorings including different cut-off to interpret the degree of ILD. Fourth, there was no agreement in the measurement to use (i.e. scoring systems), as well as the cut-off of normality. Fifth, there is no consensus regarding what the optimal ultrasound transducer is to use in the assessment of the lung. Although small surface probes with frequencies ranging between 3-3.5 MHz seemed suitable for this specific purpose, transducers with large surfaces and frequencies between 5-7.5 MHz were also used (30). Fifth, there were no studies including a cohort where all newly ILD diagnosed by LUS are followed prospectively or longitudinally to see the long-term development. Finally, In general, the studies presented offered minimal to no information regarding how well LUS performs in the detection of early ILD. Only one study (20) was performed in very early SSc patients with mean of disease duration ± SD of 1.9 ± 3.2. The authors reported a sensitivity of 100% for the screening of ILS 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. On this light, we recently conducted a study with the aim to

determine diagnostic value of LUS in in detecting subclinical ILD in 133 SSc patients. We reported that 40.6% of SSc patients showed LUS signs of subclinical ILD in contrast to healthy controls (4.8%) (p=0.0001). Sensitivity and specificity of US in detecting ILD was 91.2% and 88.6% respectively (31).

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 gold standard. To date there are no human studies using histology as gold standard. However, previous studies performed in animal models showing a good correlation between number of B-lines and water level in pulmonary oedema suggested that LUS could be a non-invasive, and simple method to detect and quantify ILD in rheumatic disorders (32).

Validation of reliability of the LUS in ILD-SSc requires comparisons of repeated LUS assessments performed within a short time period by the same investigator (intra-observer variability) and by two independent investigators (inter-observer variability) at the same time in patients with well-defined ILD-SSc.

To obtain a 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 with respect to the sensitivity to change should be ~~are~~ longitudinal studies including patients with ILD-SSc with and without treatment and parallel lung US and HRCT evaluations at different time points.

We are aware of limitations associated with the present review: the small number of articles found and the fact 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 small samples (n>40), which decreases the external validity of the articles included. Finally, studies of LUS assessing other forms of ILD were not included, which would have extended the number of suitable papers and provide a wide information regarding the utility of LUS in other types of ILD.

In conclusion, in spite of 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 researches should be focused on validity of LUS in detecting ILD in the early stages, its accuracy to assess the eventual response to the therapy, the correct timing of LUS for diagnosis and follow-up and its potential in monitoring the progression of ILD-SSc. Additionally, the research agenda should be focused in 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, Petersen F 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, Czirják L, DelGaldo 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;30;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, Crow- Son 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: 2550-57.
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;70:971-72.
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;48:1382-87.

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; 39:1641-47

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.P referred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097.

16. OCEBM Levels of Evidence Working Group, Durieux N, Pasleau F, Howick J. The Oxford 2011 Levels of Evidence. Group [Internet]. 2011;1(version):5653. Available from: http://www.cebm.net/index.aspxo=1025

17. O'Connor D, Green S, Higgins JP. Defining the Review Question and Developing Criteria for Including Studies. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. 2008. 81-94 p.

18. Whiting PF, Rutjes AWS, Westwood ME, Mallet S, Deeks JJ, Reitsma JB, et al. 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;6:562-66.

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;72:390–95.

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 Lungensonografie zur Bewertung von interstitiellen Lungenerkrankungen bei. Ultraschall der Medizin 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;31:1621–25.

24. Pinal-Fernandez I, Pallisa-Nuñez 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;18:1520–28.

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;50:1860-64.

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.

Authors' contributions.
MG and CS-F participated in the design of the review, the acquisition and interpretation of data, the drafting of the manuscript and gave final approval of the version of the paper to be published. CP, AA-R were involved in the selection of the articles to include in the review, made substantial contributions to the manuscript preparation and was involved in revising the manuscript for important intellectual content. LT, GB, AI, MADA, ADS participated in the review conception and gave substantial input to the data evaluation and manuscript preparation.
All authors read and approved the final version of manuscript.

Figure legends

Figure 1. Flowchart of the review

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

Reference	Type of study	No. of pts.	Comparison with other diagnostic method	Variables	
				Name (domain) and US definition Number of sites evaluated	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, hyper echoic narrow-based reverberation artifact that generally are not visible in normal lung parenchyma	Dichotomous (>5 = positive results)
Barskova (20)	O, P, C	58	HRCT	B-line was 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 artefacts, arising from pleural line and it is running to the edge of screen, wide at the base and narrow at the top. Consolidations are hipoechogenne, 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, T, 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 two adjacent scanning sites or when a total of ≥5 B-lines were recorded = positive results)
Moazedi-Fuerst (23)	O, P, C, Co	25	N/R	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 tickening (NR) NR	Quantitative

Tardella (13)	O, P, C	34	HRCT	B-lines: hyperechoic narrow-based reverberation type of artefact, 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 área	Semiquiantitative (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 Blines), 2 = moderate (from 16 to 30 B-lines), and 3 = severe (> 30 B-lines)

NR = not registered
O = observational, P = prospective; C= comparative, Co= control group
HRCT = High resolution computed tomography

Table 2. Technical aspect and characteristics of ultrasound machine adopted

Reference	US Mode	Transducer; Parameters and technical specifications	Number of 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., Genoa, Italy) 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 Siemense sonoline G-40 (Siemense, Germany) 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, Genoa, Italy) 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, WI)	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, Genoa, Italy)]. 2.5–3.5MHz	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 (2)
Gigante (22)	GS	Convex probe Toshiba's Ultrasound System (Tokyo, Japan) 2.5- to 3.5-MHz	NR	Standardized technique 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 (Esaote, Genoa) 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	NR NR	NR
Tardella (13)	GS	Convex probe MyLab 70 XVG (Esaote Biomedica, Genoa, Italy) equipped 2-7MHz	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 (2)
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 (Medison, South Korea) 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

	GS, grey scale; MHz, megahertz; Hz, Hertz; dB, decibels; NR, not referred; LON, longitudinal; TRV, transversal	

Table 3. Validity of the studies included.

Reference	Validity		Comparative instrument	Sensitivity and Specificity	US Results
	Construct	**Criterion**			
Gutierrez (14)	Simplified assessment 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 the semiquantitative score	HCRT	NR	A positive correlation was found between the US B lines 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	HCRT	73.85% and 88.23%	The sensitivity, specificity, positive and negative predictive value of TTUS was 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 two adjacent scanning sites or when a total of >5 B-lines	Correlation between LUS and HRCT	HCRT	100% and 55%	Individual patient analysis between LUS and HRCT showed a concordance between the two 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: assesses the degree of the progression of the active changes in the lungs, from 2 to 4 points could be obtained. Ultrasound Fibrosis Index, 3 to 35 points could be obtained. .Mild pulmonary fibrosis occurs when the UFI is 3 – 14 points; Moderate 15 – 20 points. Severe pulmonary fibrosis in LUS occurs when the patient has 21 – 35 UFI points	Warrick scale (Fibrosis Index): 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)	ULCs = absent ≤10, present >10	Warrick Scale	HRCT	NR	ULCs were absent (less than 10) in 16 patients and present (more than 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 was considered positive either when >3 B-lines were present in at least two adjacent scanning sites or when a total of >5 B-lines were recorded	Warrick Scale	HRCT	NR	The mean number of B-lines are 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: one to five positive areas received a comet score of 1, and patients with more than five abnormal areas got a comet score of 2	NR	NR	NR	The median thickness of the pleural irregularities was over 3.2 mm compared to the 1.3 mm in the volunteer group (p<0.001). Nine SSc patients (36 %) had more than 2.8 mm of pleural thickness, which was declared as a cutoff.

					Pleural nodules were sonographically observed in only one (4 %) patient with SSc.
Pinal-Fernandez (24)	PI	Warrick score	HRCT	NR	PI was detected in 28.9% (SD 20.2%) of US points. The mean Warrick score was 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 was 0.85 (95% CI 0.64–1), was higher, but not significantly, than the B-lines score (AUC=0.65, 95% CI 0.32–0.98).
Sperandeo (12)	Utrasound pleural line thickness between3.0 and 5.0 mm.	HRCT Reticular nodular pattern	HCRT	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, k= 0.875)
Moazedi-Fuerst (10)	Semiquantitative score: Comet-score of 0 was assigned to patients without positive areas, patients with 1–5 positive areas received a comet-score of 1 and patients with more than 5 abnormal areas got a comet-score of 2.	NR	HCRT	NR	Comet-tail artifacts/B-patterns were present in all patients (100%) of the ILD group (n=20) but only in 12% of the patients with normal CT-scans (n=25) ($p<0.001$). Subpleural nodes were observed in 55% (n=11) of the ILD patients compared to 17% (n=4) of the patients without radiological signs of ILD ($p=0.006$). Ninety-five percent of the ILD (n=19) patients *versus* 12, 5% (n=3) of the non-ILD group showed pleural irregularities >3 mm on thoracic ultrasound ($p<0.001$). In healthy volunteer B lines were observed in 3 patients (7%) and pleural noduli in one (2%) patient. Intraarticular PDS Gout 5/60 (8%) and CPPD 6/140 (4%) knees
Mohammadi (25)	ULCs assessment was scored semi-quantitatively as 0 = normal, (≤ 5 B-lines), 1 = mild (from 6 to 15 Blines), 2 = moderate (from 16 to 30 B-lines), and 3 = severe (> 30 B-lines)	Warrick score	HRCT	73.85% and 88.23%	ULCs assessment was compared to the 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 value of TTUS was 73.58%, 88.23%, 95.12% and 51.72% respectively

	HRCT = High resolution computed tomography NR = not registered ULCs = ultrasound lung comets ILD = interstitial lung disease PI = Pleural irregularity

Table 4. Realiability, feasibility and sensitivity to change.

Reference	Discrimination				Feasibility
	Interobserver	**Intraobserver**	**Between group Differences**	**Sensitivity to change**	
Gutierrez (14)	kappa of comprehensive US semi-quantitative assessment at para-sternal, mid-clavear, anterior axillary, mid-axillary, paravertebral, sub-scapular and posterior axillary level were: 0.943, 0.846, 0.963, 0.932, 0.958, 0.969 and 0.980 respectively	2nd para-sternal k= 0.864 4th mid-clavear k= 0.881 4th anterior axillary k= 0.868 4th mid-axillary k= 0.845 8th paravertebral k= 0.894 8th sub-scapular k=0.883 8th posterior axilary k= 0.862	NR	NR	It was estimated by comparing the time spent with respect to comprehensive assessment by the independent samples t-test. A significant difference between comprehensive US B-lines assessment (mean 23.3 ± SD 4.5, range 16 to 31 minutes) and simplified US B-lines assessment (mean 8.6 ± SD 1.4, range 6 to 12 minutes, P < 0.00001) was found
Aghdashi (19)	NR	NR	NR	NR	NR
Barskova (20)	NR	NR	NR	NR	The time needed for the scan and analysis was 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 k=0.943 Mid-clavicular k=0.846 Anterior-axillary k=0.963 Medial-axillary k=0.932 Paravertebral k=0.958 Subscapular k=0.969 Posterior-axillary k=0.980	NR	NR	NR	NR
Moazedi-Fuerst (10)	NR	NR	NR	NR	NR
Mohammadi (25)	The global kappa value of the agreement between two imaging methods was 0.553 (p<0.001).	The global kappa values for the intra-observer reliability of TTUS B-lines assessment was 0.838.	No	NR	NR

NR = Not registraded

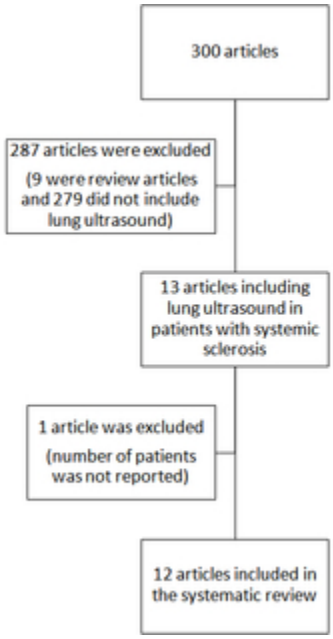

Figure 1

24x26mm (300 x 300 DPI)

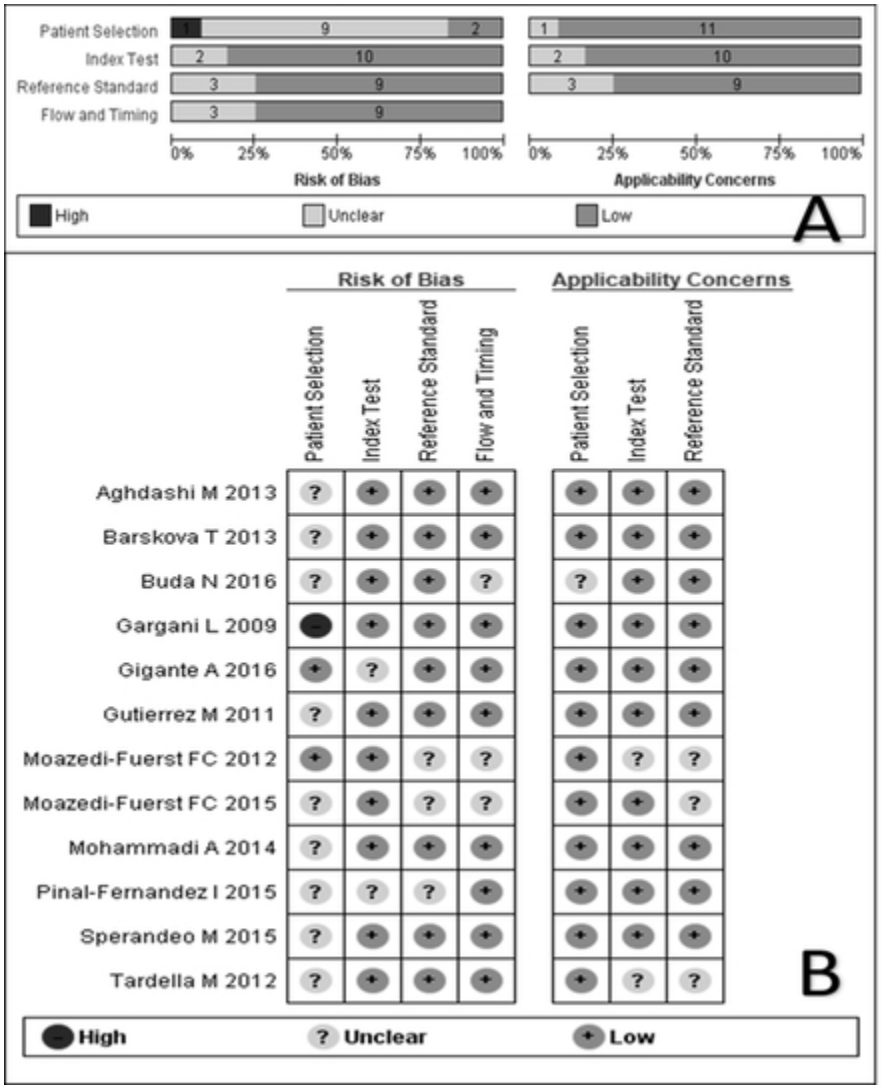

Figure 2

19x22mm (600 x 600 DPI)