Standard and pocket-size lung ultrasound devices can detect interstitial lung disease in rheumatoid arthritis patients

Chiara Cogliati1, Marco Antivalle2, Daniela Torzillo1, Simone Birocchi1, Alba Norsa3, Roberto Bianco3, Giorgio Costantino1, Maria Chiara Ditto2, Michele Battellino2, Pier Carlo Sarzi Puttini2 and Nicola Montano1

Abstract

Objectives. Interstitial lung disease (ILD) is a frequent extra-articular manifestation of RA associated with increased mortality. High-resolution CT (HRCT) is used for diagnosis and follow-up, but its accuracy is counterbalanced by high costs and radiological risk. In the presence of ILD, lung US (LUS) detects vertical artefacts called B-lines. The aims of the present study were to evaluate the accuracy of LUS in the diagnosis of ILD in RA and to validate the use of a pocket-size US device (PS-USD) as a screening tool.

Methods. LUS was performed with standard equipment by a trained physician through longitudinal scans following anatomical lines: 72 segments were considered (28 anteriorly and 44 posteriorly) and B-lines were counted in each segment. A B-line score >10 identified a positive examination (presence of ILD). A second LUS session for positive/negative judgment was performed by a short-trained physician using a PS-USD.

Results. Thirty-nine patients were studied. The sensitivity and specificity of standard LUS vs HRCT were 92% and 56%, respectively. The B-line score was significantly correlated with HRCT score (r = 0.806). A total of 29 patients were studied with a PS-USD. Sensitivity and specificity for PS-USD vs HRCT were 89% and 50%.

Conclusion. The sensitivity of LUS in the detection of ILD supports its use as a screening test for ILD in RA patients, even in the ambulatory setting with a PS-USD. The strong correlation between echographic and HRCT scores indicates LUS is a valid tool for grading and follow-up of ILD.

Key words: lung ultrasound, rheumatoid arthritis, interstitial lung disease, B-lines, pocket-size ultrasound device.

Introduction

Interstitial lung disease (ILD) is one of the most common extra-articular manifestations of RA [1]. According to recent evidence, lifetime risk of developing clinically relevant ILD in RA has been estimated to be 7.7%, with a median survival after diagnosis of only 2.6 years [2, 3]. ILD-associated mortality exceeds the combined number of deaths from all other extra-articular features of RA [4]. Furthermore, despite an overall decrease in RA-associated mortality rates, the prevalence of RA-ILD-related deaths seems to have increased in recent years [5, 6] and has not been reduced by the introduction of biologic treatments [7]. The severity of the prognosis might be related to a higher prevalence of the usual interstitial pneumonia pattern as compared with other CTDs [8] and justifies a screening strategy aimed at early diagnosis [9]. In almost half the patients, ILD is already present at disease onset or develops within 3 years [3], and subclinical, often progressive ILD, has been reported in 33% of unselected asymptomatic RA patients [10]. However, there are ample variations in the estimates of RA-ILD prevalence, depending on the population studied and...
the methods applied for the diagnosis. Currently high-resolution CT (HRCT) is the method of choice for the detection of ILD, with a diagnostic accuracy exceeding 94% [11], while lung biopsy is reserved for indeterminate cases [9]. In early RA, abnormalities consistent with ILD were found in 6%, 22% and 33% of patients by chest X-ray, pulmonary function tests (PFTs) and HRCT, respectively [12]. The use of HRCT as a screening method, however, is hampered by cost issues and by potential risks associated with the delivery of additional ionizing radiation to patients [13, 14]. PFTs, which are of fundamental prognostic importance [15–17], have been shown to correlate only weakly with HRCT findings [18, 19] and to be insensitive to limited involvement in early RA [20] and in asymptomatic patients [10]. Therefore the validation of a non-invasive, accurate, low-cost, easily performed and non-ionizing diagnostic tool could represent a major advance in the clinical management of RA-ILD. Lung US (LUS) pools all these characteristics. Recently, some studies have addressed the accuracy of lung LUS in the diagnosis of ILD in CTD. In SSc, LUS was shown to correlate well with HRCT score [21] and to be extremely sensitive in the diagnosis of early lung involvement [22]. Similarly, an excellent correlation between LUS and HRCT was reported in 34 patients including different CTDs [23]. So far, no published study has addressed the diagnostic accuracy of LUS in RA.

Moreover, in the last decade the use of US in clinical medicine has increased by the availability of hand-held devices with low cost and good performance profiles. These devices have been mostly used in the bedside or ambulatory setting to answer simple clinical questions [24–29].

The aim of the present study was to evaluate the accuracy of LUS in the diagnosis of ILD in RA patients, using HRCT as a reference standard, and to validate the use of a pocket-size US device (PS-USD) as a screening tool in the ambulatory setting.

Materials and methods

Patients fulfilled the 2010 ACR/European League Against Rheumatism (EULAR) criteria for RA [30]. In all patients, cigarette smoking (previous and/or current), 28-joint DAS (DAS28), disability (HAQ questionnaire) and drug exposure (prednisone, MTX and biologics) were assessed. All the patients were suspected or known to have ILD, with plans for an HRCT study for first diagnosis or follow-up.

Exclusion criteria were a diagnosis of pneumonia in the last month (possibly associated with an interstitial pattern at LUS), moderate or severe pleural effusion (determining atelectasis and thus affecting the correct visualization of pulmonary inferior lobes) and cardiac disease as a possible cause of heart failure (thus determining imbibition-related B-lines). To rule out cardiac disease, complete colour Doppler echocardiography was performed in each patient. Patients with moderate to severe mitral or aortic valvulopathy, left atrial or ventricular dilatation, ejection fraction <50%, diastolic dysfunction more than mild were excluded from the study. The study was approved by the institutional ethics board of L. Sacco Hospital, University of Milan (718/2011). Patients gave their written consent for the enrolment in the study.

Study protocol

Anthropometric and clinical data were collected for each patient. Resting capillary oxygen saturation was recorded while the patients were sitting on the day of the LUS study.

LUS study

LUS was performed by an expert physician (C.C. or D.T.), blinded with respect to clinical and HRCT data, using standard commercially available US equipment (Envisor, Philips, Andover, MA, USA) with a 5–2 MHz convex probe. Patients were studied in either the supine or sitting position for evaluation of the anterior and posterior chest walls, respectively. LUS was carried out through longitudinal scans following anatomical lines and B-lines were counted in each intercostal space. As previously described in the literature [21], 72 segments were considered (28 anteriorly and 44 posteriorly). B-lines were defined as vertical artefacts arising from the pleural line, deepening posteriorly and moving synchronously with respiration [31]. The total B-lines score was calculated as the sum of the B-lines counted in each segment. When there were five or more B-lines in a segment or B-lines were confluent, a value of 10 was attributed to that segment. A total B-lines score >10 identified a positive examination (presence of ILD). We previously reported that in our laboratory the r-value for interobserver variability for B-line quantification was 0.96 [32]. Moreover, a simplified method (eight-region technique) for interstitial pattern evaluation [33] was also applied: the anterolateral chest wall was divided into four areas on each side and the presence of at least three B-lines in the same scan identified a positive (B+) area. The presence of two or more positive regions bilaterally identified a positive examination (presence of ILD). Fig. 1 shows an example of a recorded US scan in two patients without and with ILD.

Finally, a short-trained physician, who underwent two sessions of 3 h each for recognition of B-lines, performed a US scan using a PS-USD (VSCAN, GE Healthcare, Fairfield, CT, USA) supporting a phased array transducer (1.7–3.8 MHz). He was unaware of the clinical, LUS and HRCT results. A positive/negative judgement was given when >10 or <10 B-lines, respectively, were identified on the anterior and posterior scans.

DLCO measurements

Single-breath diffusing capacity for carbon monoxide (DLCO) was measured using a VMAX229 (SensorMedics, Yorba Linda, CA, USA) according to the European Respiratory Society guidelines [34], while the patients were in the standing position. The breath-holding time was at least 10 s and the washout volume was 0.75 l. Results are expressed as a percentage of predicted values.

HRCT protocol

HRCT was performed on a spiral CT Light Speed QX/i GT GE-OC0 (General Electric Medical Systems, Milwaukee,
WI, USA) scanner with four rows of detectors and a rotation time of 1 s. All patients underwent an A-P and L-L scout in the supine or prone decubitus position, without administration of contrast agent. A total of 30–35 scans were then acquired during inspiratory apnoea, starting from the apex to the lung base. The acquisition parameters were as follows: sequential mode, 1.25 mm collimation, 10 mm interval, 240 mA and 140 kV tube voltage. A bone reconstruction with lung window was used. The duration of the CT acquisition was 30–35 s. The matrix was 512 × 512 and the absorbed dose was in the range of 5–6 mSv.

The radiologists gave a qualitative response (presence or absence of ILD). Moreover, a previously described method for grading pulmonary fibrosis (Warrick score) was used for a quantitative response. Briefly, a score between 1 and 5 is assigned for parenchymal alterations and a score between 1 and 3 is assigned for the number of pulmonary segments involved in the pathological process [35].

Statistical analysis

Data are presented as mean (s.d.). Relations between the HRCT and echographic score and between the HRCT and DLCO were evaluated by means of Spearman’s rank order correlation. Sensitivity and specificity were calculated for LUS performed by an expert physician with standard equipment and by a short-trained physician with PS-USD, with HRCT as the reference standard. The concordance (Cohen’s κ) between the expert and short-trained physician was calculated.

Results

Of the initial population considered for the study, one patient was excluded for the presence of moderate pleural effusion. Thus 39 patients were enrolled. The mean age, disease duration, smoking habits, disease activity, disability and drug exposure were not significantly different in the RA-ILD patients as compared with patients without lung involvement (Table 1). The mean oxygen saturation in the sitting position was 97% (s.d. 2%; range 91–100%). All patients underwent HRCT, with a positive finding (presence of ILD) in 13 patients. In these patients the mean Warrick score was 14 (s.d. 5), while a mean value of 3 (s.d. 2) was found in the HRCT-negative patients.

LUS was performed by an expert physician with standard equipment in all patients. When B-lines were counted, 25 of 39 patients showed a total >10 (presence of ILD).

The sensitivity and specificity were 92% (95% CI 78, 100) and 56% (95% CI 38, 75), respectively. When the simplified method [33] was applied, sensitivity was 69% (95% CI 44, 94) and specificity was 88% (95% CI 76, 100).

Moreover, using a receiver operating characteristics (ROC) curve, we calculated that 7 B-lines were needed to obtain 100% sensitivity; optimization for sensitivity and specificity was obtained using 17 B-lines as a cut-off, determining a sensitivity of 92% (95% CI 78, 100) and a specificity of 72% (95% CI 54, 90) (Fig. 2).

A significant correlation was found between the echographic score (total number of B-lines for each patient) and HRCT (Warrick) score, while a weak correlation was detected between the DLCO and Warrick score (Fig. 3).

In 29 of 39 patients an additional LUS using a PS-USD was performed. In Fig. 4, two scans (by PS-USD and standard equipment) are shown from the same patient. The sensitivity and specificity for PS-USD with respect to HRCT were 89% (95% CI 68, 100) and 50% (95% CI 28, 72). The κ coefficient with respect to standard US was 0.78.

Discussion

The major findings of this study are (i) LUS shows a very high sensitivity in recognizing ILD in RA patients, (ii) the sensitivity of LUS was high when scans were performed by an expert physician using standard US equipment as well as when a PS-USD was used by a short-trained operator and (iii) the number of B-lines (i.e. signs of
pulmonary fibrosis at LUS) showed a high correlation with the HRCT score.

RA is the most common inflammatory joint disease and ILD represents one of the most severe complications of RA, being responsible for a significant worsening in prognosis [4]. Physical examination and functional status questionnaires are generally used for screening pulmonary complications in ambulatory settings, while chest X-ray, spirometry and DLCO are periodically performed as first-line diagnostic tools. However, the greater sensitivity of HRCT accounts for its wide use in these patients, often repeated for follow-up even cases of a first negative result.

LUS is now recognized as an important diagnostic tool in many pathological conditions because of its easy availability, non-invasiveness, repeatability and a relatively rapid training curve. In the presence of lung fibrosis, the pulmonary US pattern is characterized by the presence of irregular pleural lines and frequently by small subpleural hypoechoic areas [36].

A recent paper reported a high sensitivity for LUS in early screening of sclerodermic patients [22], but no data are available regarding the accuracy of LUS in RA patients. In our study, LUS showed a high sensitivity in detecting ILD compared with HRCT, confirming the results obtained in sclerodermic patients.

Moreover, our study demonstrated that this sensitivity was maintained when US was performed by means of a PS-USD by a short-trained physician. Small, hand-held US devices, and more recently pocket size, have been introduced in recent years, and their use has been rapidly increasing in clinical practise, at the bedside of hospitalized patients and in the ambulatory setting. These devices have been demonstrated to improve diagnostic accuracy when added to physical examination [24, 25, 27, 28, 30] and have good reproducibility compared with standard US equipment, especially for

| TABLE 1 | Demographic and clinical variables in the population |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | All patients (n = 39) | RA-ILD (n = 13) | RA (n = 26) | P-value |
| Sex, n (%)     |                 |                 |               |         |
| Female         | 29 (74.4)       | 8 (61.5)        | 21 (80.8)     | 0.195   |
| Male           | 10 (25.6)       | 5 (38.5)        | 5 (19.2)      |         |
| Age, mean (S.D.), years | 64.87 (9.9)    | 65.31 (10.0)    | 64.65 (10.0)  | 0.607   |
| Disease duration, mean (S.D.), years | 11.21 (7.7)    | 12.46 (8.6)    | 10.56 (7.4)  | 0.543   |
| Smoker, n (%)  |                 |                 |               | 0.819   |
| No             | 22 (43.6)       | 7 (53.8)        | 15 (57.7)     |         |
| Yes            | 17 (56.4)       | 6 (46.2)        | 11 (42.3)     |         |
| RF, n (%)      |                 |                 |               | 0.039   |
| Negative       | 7 (17.9)        | 0 (0.0)         | 7 (26.9)      |         |
| Positive       | 32 (82.1)       | 13 (100)        | 19 (73.1)     |         |
| Anti-CCP, n (%)|                 |                 |               | 0.386   |
| Negative       | 10 (25.6)       | 2 (15.4)        | 8 (30.8)      |         |
| Positive       | 29 (74.4)       | 11 (84.6)       | 18 (69.2)     |         |
| HAQ, mean (S.D.) | 0.92 (0.7)    | 0.94 (0.7)      | 0.92 (0.6)   | 0.987   |
| DAS28, mean (S.D.) | 2.97 (0.9)   | 2.83 (0.7)      | 3.03 (1.0)   | 0.546   |
| Prednisone, n (%) |             |                 |               | 0.474   |
| No             | 1 (2.6)         | 0 (0.0)         | 1 (3.8)       |         |
| Yes            | 38 (97.4)       | 13 (100.0)      | 25 (96.2)     |         |
| Cumulative dose, mean (S.D.), g | 14.71 (23.8)  | 11.13 (11.8)    | 16.37 (27.7) | 0.545   |
| MTX, n (%)     |                 |                 |               | 1.000   |
| No             | 4 (10.3)        | 1 (7.7)         | 3 (11.5)      |         |
| Yes            | 35 (89.7)       | 12 (92.3)       | 23 (88.5)     |         |
| Cumulative dose, mean (S.D.), g | 3.81 (3.7)   | 3.16 (3.7)      | 4.08 (3.54) | 0.246   |
| Biologics, n (%) |             |                 |               |         |
| No             | 12 (30.8)       | 4 (30.8)        | 8 (30.8)      |         |
| Yes            | 27 (69.2)       | 9 (69.2)        | 18 (69.2)     |         |
| TNF-α, n (%)   |                 |                 |               | 0.357   |
| No             | 16 (41.0)       | 4 (30.8)        | 12 (46.2)     |         |
| Yes            | 23 (59.0)       | 9 (69.2)        | 14 (53.8)     |         |
| Others, n (%)  |                 |                 |               | 0.022   |
| No             | 17 (43.6)       | 9 (69.2)        | 8 (30.8)      |         |
| Yes            | 22 (56.4)       | 4 (30.8)        | 18 (69.2)     |         |

Patients are grouped according to the presence or absence of HRCT signs of interstitial involvement. For all considered variables, exposure refers to current and/or previous use. P-value refers to the comparison between the patients with and without ILD. Dichotomous and continuous variables were analysed by chi-square test and Mann-Whitney U test. ILD: interstitial lung disease; DAS28: 28-joint DAS.
cardiac evaluation [26, 37]. Regarding LUS, a previous paper [38] showed that the recognition of B-lines was reliable even when a hand-held device was used. In our sample we had just one false negative result for the PS-USD. The good correlation between the PS-USD and the standard machine, even when using different probes (convex vs cardiac), demonstrates that although the images can be slightly different, the clinical information does not change. Moreover, the high sensitivity of the PS-USD, even in the hands of a short-trained physician, indicates it is a valid tool for a first screening of ILD available to the rheumatologist in an ambulatory setting.

Fig. 2 Receiver operating characteristic (ROC) curve

ROC curve showing the accuracy of lung US in recognizing the presence of interstitial lung disease compared with high-resolution CT.

Fig. 3 Spearman’s correlation

Spearman’s correlation between Warrick score and the number of B-lines on lung US (left panel) and between Warrick score and diffusing capacity for carbon monoxide (DLCO) (right panel).

In the present study we used the cut-off value of 10 B-lines for considering the presence of ILD, as derived from literature [21]. Some discussion exists about the reproducibility of B-lines counts, these being artefacts and potentially depending upon the angle of the US beam [39]. In our experience, the reproducibility of the B-lines count is excellent for a maximum number of five B-lines in a single intercostal space. For this reason we assigned the arbitrary value of 10 when more than five vertical artefacts were observed. Moreover, in comparison with the absolute count of B-lines, a simplified method considering the count of positive areas in the anterolateral chest wall [33] provided a lower sensitivity. This suggests that even though the simplified method is the best one in an emergency setting for the differential diagnosis of dyspnoea, it is probably less suitable for the screening of pulmonary fibrosis, possibly because of the peculiar distribution of ILD in rheumatic diseases often starting from the basis of the lung. Data obtained with the ROC curve are similar to those reported in sclerodermic patients regarding the value of B-lines, allowing 100% sensitivity. Nevertheless, even considering the number of 17 B-lines indicated by the ROC curve for optimization of sensitivity and specificity, specificity remains lower than in previous papers [22]. This is probably explained by the fact that our study population was characterized by a relatively high mean age; in old patients, small pleural or parenchymal scars due to previous diseases are not unusual and may determine vertical artefacts, thus accounting for a lower specificity.

Finally, we found a high correlation between the number of B-lines and the HRCT score, while no correlation was present for DLCO values. These data strongly support the use of LUS as part of the instrumental grading of ILD and as well as long-term monitoring.

A possible limitation of our study is that we selected RA patients on the basis of the need for HRCT for clinical suspicion or follow-up in known ILD. Therefore the sensitivity that we assessed has to be tested in a population of asymptomatic RA patients.

In conclusion the results of our study support the use of LUS as a screening tool for ILD even in cases where only a low-cost PS-USD is available. In fact, from a practical
viewpoint, LUS could have avoided HRCT in 14 of the 39 patients. The high sensitivity of LUS, together with the possibility of using it in follow-up, may guide the clinician in deciding the timing of HRCT, thus avoiding periodic exposure to radiation.

Rheumatology key messages

- Lung US has a high sensitivity in detecting interstitial lung disease in RA patients and a strong correlation with CT.
- Pocket-size US devices provide a diagnostic yield similar to higher-level devices.
- Lung US may be used as a screening tool to detect and monitor interstitial lung disease.

Disclosure statement: The authors have declared no conflicts of interest.

References


