

## Visual versus automatic ultrasound scoring of lung B-lines: reliability and consistency between systems

Jorge Short<sup>1</sup>, Carlos Acebes<sup>2</sup>, Guido Rodriguez-de-Lema<sup>1</sup>, Giuliana Maria Concetta La Paglia<sup>3,4</sup>, Maria Pavón<sup>1</sup>, Olga Sánchez-Pernaute<sup>3</sup>, Julio Cesar Vazquez<sup>1</sup>, Fredeswinda Romero-Bueno<sup>3</sup>, Jesús Garrido<sup>5</sup>, Esperanza Naredo<sup>3</sup>

<sup>1</sup>Department of Emergency Medicine, Ultrasound Division. Hospital Universitario Fundación Jiménez Díaz, IIS Fundación Jiménez Díaz and Universidad Autónoma de Madrid. Madrid, Spain, <sup>2</sup>Rheumatology Unit, Hospital General de Villalba. Madrid, Spain, <sup>3</sup>Department of Rheumatology, Bone and Joint Research Unit. Hospital Universitario Fundación Jiménez Díaz, IIS Fundación Jiménez Díaz, and Universidad Autónoma de Madrid. Madrid, Spain, <sup>4</sup>Division of Rheumatology and Clinical Immunology, Humanitas Clinical and Research Center, Rozzano, Italy, <sup>5</sup>Department of Social Psychology and Methodology, Facultad de Psicología, Universidad Autónoma de Madrid, Spain

### Abstract

**Aims.** To evaluate the agreement between a visual and an automatic counting system of lung B-lines by ultrasound (US) as well as to test the inter- and intra-observer reliability of both systems in patients with lung diseases. **Material and methods.** We included four patients with different lung conditions. Four ultrasonographers expert in lung US blindly, independently and consecutively performed, in two rounds, a US B-mode assessment of 8 lung intercostal spaces of each patient. Each US assessment consisted of a visual and an automatic counting of B-lines. **Results.** Agreement between visual and automatic counting of B-lines was good to excellent [intraclass correlation coefficient (ICC) 0.79-0.84,  $p < 0.001$ ]. Intra-observer reliability was good to excellent [ICC 0.62-0.99,  $p < 0.001$ ] except for one investigator in whom it was close to moderate for the automatic system [ICC 0.49,  $p < 0.05$ ]. Inter-observer reliability was excellent for both systems in both rounds [ICC 0.86-0.90,  $p < 0.001$ ]. **Conclusions.** US automatic counting was consistent with US visual counting of lung B-lines, as performed by experts in the field. Both systems showed a high intra- and interobserver reliability.

**Keywords:** ultrasound; B lines; interstitial lung disease

### Introduction

Ultrasound (US) is a non-invasive, non-ionising, relatively inexpensive, safe and patient-friendly imaging modality, which is being increasingly used as a point-of-care tool to optimize diagnosis and therapeutic deci-

sions. Among all medical applications, US of the lung is being progressively implemented in clinical practice in different medical specialties, from Emergency medicine to Rheumatology. Since the late 60s lung US has been applied to detect pleural effusion and since the 90s the “comet-tail” artefact, described by Lichtenstein [1] and subsequently named “B-lines”, has been considered as a marker of alveolar-interstitial pathological changes which produce a disbalance between air, water and connective tissue in the subpleural lung tissue. B-lines are US signs of thickening of interlobular septa due to the presence of increased fluid or collagen tissue alteration (i.e. inflammation or fibrosis) [2,3]. Currently, there is a growing evidence supporting the diagnostic value of US-detected B-lines in a variety of lung conditions such as acute or chronic heart failure, chronic bronchitis or interstitial

Received 10.01.2019 Accepted 12.02.2019

Med Ultrason

2019, Vol. 21, No 1, 45-49

Corresponding author: Esperanza Naredo

Department of Rheumatology, Bone and Joint Research Unit, Hospital Universitario Fundación Jiménez Díaz and Autónoma University

Av. Reyes Católicos, 2. 28040 Madrid. Spain

Phone: +34 915 504 978

E-mail: enaredo@ser.es

ORCID: 0000-0003-0017-0096

lung involvement in connective tissue diseases (CTD) [4-18]. A recent meta-analysis [19] has shown a high diagnostic accuracy and a high concordance of lung US as compared to high resolution computed tomography, the reference imaging method, in detection and quantification, respectively, of interstitial lung disease (ILD) in patients with CTD. In addition, a high inter- and intra-observer reliability in detecting and scoring B-lines, mainly in patients with CTD, has been reported [2,10,12-14,17]

Traditionally, quantification of lung B-lines on US imaging has been done visually. This implies the need of high operator experience in this procedure since an accurate counting of B-lines is crucial to classify the patient lung as normal or pathological. In previous studies, either a number of B-lines  $\geq 3$  in two adjacent lung intercostal spaces [2,5,9,20,21] or a total number of B-lines  $\geq 5$  [2,5,9,20,21] or  $\geq 10$  [10] in all scanned intercostal spaces have been established as cut-offs between normality and abnormality. Recently launched software facilities allow an automatic real-time detection and quantification of lung B-lines, which potentially could make US assessment less operator-dependent. There are, however, no studies providing evidence of that.

The aim of this study was to evaluate the agreement between visual and automatic counting of B-lines as well as to test the inter- and intra-observer reliability of both systems in patients with lung diseases.

### Material and methods

We conducted an exercise to test the agreement between visual and automatic counting of B-lines by US and the intra- and interobserver reliability of both systems between 4 ultrasonographers, in 4 patients with lung pathology.

#### Patients

Four patients were included in the study: 1) a 77-year-old male diagnosed with severe chronic obstructive lung disease; 2) a 83-year old female with hypertension and acute heart failure; 3) a 68-year-old male diagnosed with systemic sclerosis with ILD; and 4) a 47-year-old female with antisyntetase syndrome with active lung involvement. They were recruited from the Emergency Unit (patients 1 and 2) or the outpatient Rheumatology Clinic (patients 3 and 4) by staff members of the respective departments, different from the investigators involved in the US assessment.

Each patient was randomly assigned to a US scanner where they remained during the US assessments of their lungs. The study was conducted in accordance with the World Medical Association Declaration of Helsinki, as revised in 2013 (Fortaleza) and was approved by the

local Ethics Committee of the Hospital Universitario Fundación Jiménez Díaz. Written informed consent was obtained from all patients before the study.

#### Ultrasonographers

The US investigators comprised 2 emergency specialists (JS, GRL) and 2 rheumatologists (EN, CA) with long experience in lung US. They were unaware of the diagnosis and clinical data of the patients.

#### US lung assessment

The US investigation was carried out using four commercially available real-time scanners, i.e. VENUE, General Electric Healthcare (Wauwatosa, WI, US), equipped with multifrequency convex transducers (1-5 MHz) and software for automatic B-lines counting (AUTO B lines tool). The B-mode settings of each US machine were optimised for lung assessment and standardized for the study as follows: frequency 4.5 MHz, gain -9 db, dynamic range 69 db, and persistence 0.0.

Four areas at each hemithorax, respectively located at the upper and lower front, and the upper and lower lateral chest, were selected for the US evaluation and marked beforehand to guide the ultrasonographers in probe positioning. Briefly, anterior regions were set at the third intercostal space (area 1) and fifth intercostal space (area 2) levels following the mid-clavicular line, while lateral regions were located at the fourth (area 3) and the sixth (area 4) intercostal spaces, at the level of the axillary midline [21,22].

The 4 ultrasonographers blindly, independently and consecutively performed a B-mode scanning of the 8 selected intercostal spaces of each patient, in two rounds in a blinded fashion. Each B-mode scanning consisted of the following: 1) first, after placing and stabilizing the probe at each intercostal space, the ultrasonographers visually counted the B-lines and recorded the B-line number in a standardised report sheet, and 2) after the visual counting, they activated the automatic counting software, which automatically detected the B-lines, selected the image with a higher B-line number from a 6-sec sequence and provided a number of B-lines from 0 to  $>5$  that they also recorded in a standardised report sheet. Each ultrasonographer was given a maximum of 10 mins to scan each patient and fill in the number of B-lines in the standardised report sheet. The ultrasonographers were assigned to the US machines in a different order for each round. An application specialist from the US company was near each machine to solve potential technical adjustment problems.

B-lines were defined as discrete laser-like vertical hyperechoic reverberation artifacts that arose from the pleural line and extended to the bottom of the screen without fading, while moving synchronously with lung sliding [21].



**Fig 1.** a) Ultrasound scan of an intercostal space showing no B-lines; b) Ultrasound scan of an intercostal space showing 4 B-lines; c) Automatic counting software showing 4 B-lines

**Statistical analysis**

Statistical analysis was performed using the IBM-SPSS 23 package. Quantitative variables were presented as mean, standard deviation (SD) and range. Intraclass correlation coefficients (ICC) were used to study agreement between the visual and automatic system of B-line counting in the two scanning rounds as well as intra-observer and interobserver reliability. Since the focus of the study was the agreement between systems, rater was considered as a fixed effect, and two-way mixed effect models (ICC 3,k) for consistence were used. ICC value <0.50 was considered poor, 0.50-0.60 moderate, 0.60-0.80 good and 0.80-1 excellent. A p value <0.05 was considered significant.

**Results**

A total of 256 visual counts and 247 automatic counts of B-lines were obtained. Nine automatic counts (4 in the first round, 5 in the second round) were missed, 8 of which by the same ultrasonographer, who could not produce images of sufficient quality in some intercostal spaces of a patient to allow automatic counting of B-lines.

**Table I.** Agreement between visual and automatic B-line counting

	n	ICC (95%CI)	p-value
All	247	0.794 (0.736-0.840)	<0.001
Round 1	124	0.746 (0.638-0.822)	<0.001
Round 2	123	0.840 (0.772-0.888)	<0.001

ICC – intraclass correlation coefficient; CI – confidence interval; n – number of B-line

The mean (SD; range) of B-lines was 2.23 (2.21; 0->5) for the visual counting system and 2.23 (2.45; 0->5) for the automatic counting system (p = 0.430). Figure 1 shows examples of B-mode images of intercostal spaces and an example of the automatic counting system of B-lines.

**Agreement between the visual and automatic system of B-line counting**

Table I shows the agreement between the visual and automatic counting of B-lines in all US assessments and in each round, which was good to excellent.

**Intra- and interobserver reliability of visual and automatic systems of B-line counting**

Table II displays the intra-observer reliability of each ultrasonographer for both, the visual and the automatic

**Table II.** Intra-observer reliability of visual and automatic B-line counting

Visual counting system			
Ultrasonographer	n	ICC (95%CI)	p-value
1	32	0.778 (0.545-0.892)	<0.001
2	32	0.986 (0.971-0.993)	<0.001
3	32	0.911 (0.818-0.957)	<0.001
4	32	0.688 (0.360-0.848)	0.001
Automatic counting system			
	n	ICC (95%CI)	p-value
1	28	0.485 (0.000-0.762)	0.045
2	32	0.826 (0.647-0.916)	<0.001
3	32	0.624 (0.230-0.816)	<0.001
4	31	0.702 (0.382-0.856)	0.001

ICC – intraclass correlation coefficient; CI – confidence interval; n – number of B-line

Table III. Interobserver reliability of visual and automatic B-line counting

Visual counting system			
	n	ICC (95%CI)	p-value
Round 1	32	0.878 (0.790-0.935)	<0.001
Round 2	32	0.902 (0.831-0.947)	<0.001
Automatic counting system			
	n	ICC (95%CI)	p-value
Round 1	28	0.866 (0.762-0.932)	<0.001
Round 2	27	0.856 (0.741-0.928)	<0.001

ICC – intraclass correlation coefficient; CI – confidence interval

system. The ICC were good to excellent except for one investigator on whom it was close to moderate for the automatic system.

Table III shows the interobserver reliability in each round for the visual and the automatic system, being the ICC excellent for both rounds.

### Discussions

In this study, we managed to show for the first time the high agreement of the automatic B lines counting system with the one performed visually by experts. Similarly, we observed that this measurement is as reliable as the visual system between experts.

Even though US assessment of lung involvement in CTD [10-18] is more comprehensive and involves a higher number of intercostal spaces than lung assessment in the Emergency setting, we chose a reduced intercostal space evaluation, such as the method recommended by Volpicelli et al [21] to make the exercise feasible and friendly for the patients.

A disadvantage of the automatic B line software used in this study is that it does not provide us with a number of B-lines >5 as the visual system does. However, this limitation has little relevance in clinical practice since the presence of >3 B-lines per intercostal space is considered pathological.

On the other hand, although one of the ultrasonographers had difficulties to get the correct image sequences for the automatic counting system in some intercostal spaces in one of the patients, the rest of the ultrasonographers did get accurate images in this patient. It remains to be demonstrated in future studies whether learning of the automatic system to get a reliable counting is, as it can be forecast, faster than the visual system and thus, if safe performance of lung US can be more widespread among the medical community.

In **conclusion**, our study showed that US automatic counting of lung B-lines is consistent with and as reliable as US visual counting performed by experts in the field.

### Acknowledgements

We would like to thank Luis Ortega, Fernando De la Torre and Carlos Salgado, from General Electric Healthcare, for providing the US machines and their technical support

### References

- Lichtenstein DA, Meziere G, Biderman P, Gepner A, Barré O. The comet-tail artifact. An ultrasound sign of alveolar-interstitial syndrome. *Am J Respir Crit Care Med* 1997;156:1640-1646.
- Jambrik Z, Monti C, Coppola V, Mottola G, Miniati M, Picano E. Usefulness of ultrasound comets as a nonradiologic sign of extravascular lung water. *Am J Cardiol* 2004;93:1265-1270.
- Agricola E, Bove T, Oppizzi M, et al. "Ultrasound comet-tail images": a marker of pulmonary edema: a comparative study with wedge pressure and extravascular lung water. *Chest* 2005;127:1690-1695.
- Soldati G. Sonographic findings in pulmonary diseases. *Radiol Med* 2006;111:507-515.
- Picano E, Frassi F, Agricola E, Gligorova S, Gargani L, Mottola G. Ultrasound lung comets: a clinically useful sign of extravascular lung water. *J Am Soc Echocardiogr* 2006;19:356-363.
- Frassi F, Gargani L, Gligorova S, Ciampi Q, Mottola G, Picano E. Clinical and ecocardiographic determinants of ultrasound lung comets. *Eur J Echocardiogr* 2007;8:474-479.
- Lichtenstein DA, Mezière GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. *Chest* 2008;134:117-125.
- Soldati G, Copetti R, Sher S. Sonographic interstitial syndrome: the sound of lung water. *J Ultrasound Med* 2009;28:163-174.
- Sperandeo M, Varriale A, Sperandeo G, et al. Transthoracic ultrasound in the evaluation of pulmonary fibrosis: our experience. *Ultrasound Med Biol* 2009;35:723-729.
- Gargani L, Doveri M, D'Errico L, et al. Ultrasound lung comets in systemic sclerosis: a chest sonography hallmark of pulmonary interstitial fibrosis. *Rheumatology (Oxford)* 2009;48:1382-1387.

11. Delle Sedie A, Doveri M, Frassi F, et al. Ultrasound lung comets in systemic sclerosis: a useful tool to detect lung interstitial fibrosis. *Clin Exp Rheumatol* 2010;28 (5 Suppl 62):S54.
12. Gutierrez M, Salaffi F, Carotti M, 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.
13. Tardella M, Gutierrez M, Salaffi F, et al. Ultrasound in the assessment of pulmonary fibrosis in connective tissue disorders: correlation with high-resolution computed tomography. *J Rheumatol* 2012;39:1641-1647.
14. Barskova T, Gargani L, Guiducci S, et al. Lung ultrasound for the screening of interstitial lung disease in very early systemic sclerosis. *Ann Rheum Dis* 2013;72:390-395.
15. Cogliati C, Antivalle M, Torzillo D, et al. Standard and pocket-size lung ultrasound devices can detect interstitial lung disease in rheumatoid arthritis patients. *Rheumatology (Oxford)* 2014;53:1497-1503.
16. Moazedi-Fuerst FC, Kielhauser SM, Scheidl S, et al. Ultrasound screening for interstitial lung disease in rheumatoid arthritis. *Clin Exp Rheumatol* 2014;32:199-203.
17. 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.
18. Moazedi-Fuerst F, Kielhauser S, Brickmann K, 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-S91.
19. Song G, Bae SC, Lee YH. Diagnostic accuracy of lung ultrasound for interstitial lung disease in patients with connective tissue diseases: a meta-analysis. *Clin Exp Rheumatol* 2016;34:11-16.
20. Gargani L. Lung ultrasound: a new tool for the cardiologist. *Cardiovasc Ultrasound* 2011;9:6.
21. Volpicelli G, Elbarbary M, Blaivas M, et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med* 2012;38:577-591.
22. Gargani L, Volpicelli G. How I do it: Lung ultrasound. *Cardiovasc Ultrasound* 2014;12:25.