

Sources of Variability in the Detection of B-Lines, Using Lung Ultrasound

Pivetta E¹, Baldassa F², Masellis S², Bovaro F³, Lupia E⁴, Maule MM⁵.

1 Cancer Epidemiology Unit and CRPT, Department of Medical Sciences, University of Turin, Turin, Italy; Division of Emergency Medicine, Department of Medical Sciences, University of Turin, Turin, Italy. Electronic address: emanuele.pivetta@gmail.com.

2 Division of Emergency Medicine, Department of Medical Sciences, University of Turin, Turin, Italy; School of Medicine, University of Turin, Turin, Italy.

3 Division of Emergency Medicine, Department of Medical Sciences, University of Turin, Turin, Italy; Residency Program in Emergency Medicine, University of Turin, Turin, Italy.

4 Division of Emergency Medicine, Department of Medical Sciences, University of Turin, Turin, Italy.

5 Cancer Epidemiology Unit and CRPT, Department of Medical Sciences, University of Turin, Turin, Italy.

Abstract

Lung ultrasound (LUS) is a largely employed diagnostic tool but an operational protocol for implementation has never been proposed. The lack of standardization clearly introduces variability in LUS results. We enrolled adult patients presenting for acute dyspnea with a clinical suspect of etiology related to heart failure. We calculated agreement among four providers in assessing B-lines. We varied probes, depth, evaluation time and scanning areas and we estimated the importance of each factors on B-lines assessment. Overall agreement among raters varied from a kappa of 0.70 to 0.81. The mean number of B-lines was 5.44 (95% confidence interval, CI, 4.1–6.8). This estimate did not suffer variation by the depth used (0.03, 95% CI –0.2–0.2, more B-lines, using 19 cm versus 10 cm). The use of a convex probe and expertise in LUS reduced the number of artifacts by 1.7 (95% CI 1.5–1.9) and 1.1 in comparison with a phased array probe and naive operators. Evaluation time increased estimates by 1.2 (95% CI 1–1.5) and 2.9 (95% CI 2.7–3.9) B-lines for 4” and 7” clips (reference was 2” clips). This study suggests that the probe, the evaluation time and the level of expertise might affect the results of quantitative assessment of B-lines.

Key Words

Lung ultrasound

Variability

Heart failure

Introduction

Lung ultrasound (LUS) was first described in 1961 (Dunn and Fry 1961) and its potential use as a point-of-care (POC) diagnostic tool has been known since 1997 (Lichtenstein et al. 1997). A number of studies (Barskova et al, 2013, Caiulo et al, 2011, Noble et al, 2009b, Xirouchaki et al, 2011) exploring LUS diagnostic accuracy in multiple clinical contexts and applications have been published.

However, until now a standardized operational protocol for the detection of B-lines with LUS has never been proposed. Since the study by Lichtenstein et al. (1997), where a micro-convex probe was used, all types of devices have been suggested. Moreover, various scanning depths and evaluation times are utilized in practice, and studies involving the detection of B-lines often fail to provide details of these technical specifications (Pivetta et al, 2015, Volpicelli et al, 2006).

This lack of technical standardization, in addition to the heterogeneity of patient characteristics and operators' level of expertise, affects data interpretation.

In 2012, an international experts' statement on POC LUS was published (Volpicelli et al. 2012) that summarized what is known about this topic. In this consensus statement, B-lines were defined as hyperechoic artifacts that arise from the *pleural* line and extend to the bottom of the screen without fading (*i.e.*, a synchronous movement with pleural sliding, “lung sliding”). Their nature was assumed to be related to the presence of transudate in the lung parenchyma. Recently, in animal models, it has been shown that B-lines are actually a manifestation of a density modification (Soldati et al, 2014, Soldati et al, 2016). Currently, the presence of three or more B-lines defines a positive zone; and two (or more) positive zones in each hemithorax indicate the presence of interstitial syndrome (ISy) (Volpicelli et al. 2012).

Precise recommendations on the type of probes, scan duration and depth needed to perform a diagnostic examination are not available.

Various approaches have been suggested, using both curvilinear and phased array probes, scanning depths from 10 to 18 cm, and examination intervals from 1 to 5 s or more.

We have assumed that all of these factors (direct acyclic graph, Fig. 1), as well as providers' expertise, might affect the reliability of POC LUS results as a valid addition to clinical findings.

The aim of the study was to identify and describe the various sources of variability in detecting B-lines in dyspneic acute patients and to assess their relative importance.

Materials and Methods

A prospective cohort was enrolled at Città della Salute e della Scienza di Torino University Hospital, Department of Emergency Medicine, Turin, Italy, between May and September 2016, after having been granted the approval of the institutional review board (project approval number 13132013). All patients signed an informed consent to participate in the study. After enrollment, data were anonymized. The study was conducted in accordance with the principles of the Declaration of Helsinki for clinical research involving human patients.

All adult patients (*i.e.*, age >18 y) admitted to the emergency department (ED), between May and September 2016, for acute dyspnea with a clinical suspicion of heart failure (Ponikowski et al. 2016) were considered eligible for the study. Exclusion criteria were diuretic assumption less than 1 h before admission and non-invasive ventilation treatment less than 15 min before LUS evaluation. We also excluded patients with a known history of lung fibrosis. We chose this study population to maximize the detection of B-lines unrelated to the acute condition.

Two operators (F.B. and E.P.) collected LUS videos for each patient, using a 4-zones-for-hemitorax scanning protocol, as suggested by the international recommendations for POC LUS (Volpicelli et al. 2012). Each patient was evaluated with a curvilinear and a phased array probe, at scanning depths of 10 and 19 cm, and a 7-s sonographic clip was recorded. After enrollment, two copies of each clip were made, which were then cut at 2 and 4 s from the beginning of the recording (which occurred randomly in the diagnostic process) to investigate the diagnostic value of various recording lengths.

We obtained a database in which each patient was examined with 2 probes, at 2 sonographic depths and for 3 interval lengths in 4 zones of each hemithorax, *i.e.*, we had a set of 96 clips for each of a total of 50 patients.

A total of 4 operators with different expertise in performing LUS (2 with previous expertise, *i.e.*, more than 40 examinations already performed at the time of the study (Goffi et al, 2013, Pivetta et al, 2015); and 2 naive operators with a specific basic training in LUS, *i.e.*, 90 min of theory and less than 10 POC examinations) blindly reviewed all 4800 videos and classified each zone qualitatively (positive and negative for the presence of at least 3 B-lines) and quantitatively (number of B-lines).

Descriptive data are shown as mean (\pm SD) or median (with interquartile range [IQR]) for continuous variables and as numbers and percentages for categorical variables.

Agreement among the 4 raters' qualitative evaluations was calculated, using the modified Cohen's kappa for multiple raters with 95% confidence interval (Fleiss 1971). Agreement on quantitative evaluations (the number of detected B-lines) was assessed, using the intra-class correlation 2-way coefficient (Bartko 1966), and multilevel linear regressions in which data are modeled with a 2-level structure with technical variables (probes, depth, evaluation time and scanning areas) as units at level 1, and raters' experience at level 2 (MacCallum et al. 1997).

Data were collected in a Microsoft Excel spreadsheet (Microsoft Inc., Redmond, WA, USA), and statistical analyses were conducted using Stata 13.0/SE (Stata Corp, College Station, TX, USA).

Results

Baseline characteristics of patients enrolled in the study are summarized in Table 1.

Table 1. Baseline characteristics of enrolled patients

	Age (median, IQR)	80.5 y (11)
M/F ratio		1.27
Symptoms at presentation to the ED (n, %)		
Dyspnea		43 (86%)
Tachycardia		3 (6%)
Chest pain		11 (22%)
Sent to the ED by general practitioner or EMS (n, %)		10 (20%)
History (n, %)		
Heart failure		19 (38%)
COPD		20 (40%)
Atrial flutter or fibrillation		24 (48%)
Hypertension		40 (80%)
Diabetes		14 (28%)
CAD (any type)		19 (38%)
Chronic renal failure		17 (34%)
Physical examination		
Rales		40 (80%)
Legs edema		34 (68%)
Fever (n=47)		2 (4.3%)
Home therapy		
Diuretics		32 (64%)
ED treatment		
NIV (any type)		3 (6%)
Diuretics		46 (92%)

Age (median, IQR)	80.5 y (11)
Nitrates	9 (18%)
Nt-proBNP >300pg/mL (n=43)	40 (93%)

CAD=coronary artery disease; COPD=chronic obstructive pulmonary disease; ED=emergency department; EMS=emergency medical services; F=female; IQR=interquartile range; M=male; NIV=non-invasive ventilation; Nt-proBNP=N-terminal pro b-type natriuretic peptide.

Overall agreement among all raters varied from a Cohen's kappa of 0.70 (lower lateral left zone) to 0.81 (for lower medial left zone—see Fig. 2 for details). It varied from 0.73 (upper medial left zone) to 0.83 (upper lateral left zone) between naive raters, and from 0.80 (lower medial left zone) to 0.63 (lower lateral left zone) between expert raters.

Table 2 presents the results of multilevel regression models. The unconditional (null) model predicted an overall mean number of B-lines of 5.44 (95% CI 4.1–6.8) with a between-subject SD of 4.9 (95% CI 4.1–6.0), a within-subject SD of 7.7 (7.6–7.7) B-lines and an intra-class correlation coefficient (ICC) of 0.29 (95% CI 0.2–0.4).

Table 2. Multilevel models for evaluation of number of B-lines

	Level 1 model	Level 2 model
Global number of B-lines	5.8 (95% CI 4.4–7.2)	6.3 (95% CI 4.9–7.8)
Duration of the clip (baseline: 2'')		
4'' clip	1.2 (95% CI 1–1.5)	1.2 (95% CI 1–1.5)
7'' clip	2.9 (95% CI 2.7–3.2)	2.9 (95% CI 2.7–3.2)
Type of probe (baseline: phased array)	–1.7 (95% CI –1.9 to –1.5)	–1.7 (95% CI –1.9 to –1.5)
Evaluation depth (baseline: 10cm)	0.03 (95% CI –0.2 – 0.2)	0.03 (95% CI –0.2 – 0.2)
Aggregated evaluation zones (baseline: upper medial zones)	1.35 upper lateral (95% CI 1.1–1.6)	1.35 upper lateral (95% CI 1.1–1.6)
	–0.5 lower medial (95% CI –0.8 to –0.2)	–0.5 lower medial (95% CI –0.8 to –0.2)
Expertise of raters (baseline: naive operators)	2.6 lower lateral (95% CI 2.3–2.9)	2.6 lower lateral (95% CI 2.3–2.9)
	—	–1.1 (95% CI –1.3 to –0.8)

CI=confidence interval.

Inclusion of level 1 covariates (evaluation depth, clip duration, type of probe and evaluation zones) changed the estimations as follows: the use of a convex probe reduced the number of B-lines by 1.7 (95% CI 1.5–1.9) compared with a phased array probe detection, whilst evaluation length increased B-lines by 1.2 (95% CI 1.0–1.5) and 2.9 (95% CI 2.7–3.2) for 4'' and 7'' clips, respectively, compared with a 2'' clip. The lateral zones, upper and lower, showed 1.35 (95% CI 1.1–1.6) and 2.6 (95% CI 2.3–2.9) more B-lines than the upper medial area. The lower medial area showed 0.5 (95% CI –0.8 to –0.2) fewer B-lines than the upper medial area.

Scan depth did not affect the estimates: a mean number of 0.03 more artifacts with 19 cm (95% CI –0.2 to 0.2) than 10-cm depth were observed.

Inclusion of these covariates into the model yielded a between-subject SD of 4.9 (95% CI 4.1–6.0), a within-subject SD 7.4 (95% CI 7.4–7.5) and an ICC of 0.30 (95% CI 0.2–0.4).

The inclusion of the raters' expertise in the model (level 2 covariate) yielded a reduction of 1.1 B-lines (95% CI 0.8–1.3).

Discussion

To our knowledge, this is the first study evaluating several sources of LUS variability, using both qualitative and quantitative approaches, including multilevel regression models, to assess their relative importance. We found that the operator's expertise, the type of probe and the clip duration affect the detection of B-lines in a cohort of dyspneic patients suspected of suffering from acute heart failure (HF) in the ED.

LUS has proven to be a useful tool for the evaluation and management of patients in several fields, including emergency medicine and critical care (Lichtenstein, 2015, Martindale et al, 2016, Noble et al, 2009a); however, no precise indications on how to perform this examination have been provided (Volpicelli et al. 2012). This suggested to us the rationale for studying the effect of a number of variables in determining the number of B-lines, independently of LUS diagnostic accuracy for HF.

Four raters, including both expert and naive operators, showed good agreement in classifying lung zones as positive or negative on the basis of the presence of B-lines. The lowest observed inter-rater agreement, measured through the kappa coefficient, was 0.70. This result indirectly confirmed the steep learning curve for the detection of B-lines already reported in the literature (Blehar et al, 2015, Pivetta et al, 2015, Volpicelli et al, 2012).

Raters showed good agreement in the qualitative assessment of B-lines also after stratifying by the other factors (type of probe, duration and depth of evaluation).

In 2010, Delle Sedie et al. addressed the issue of variability in the detection of B-lines, and, in 2014, Cogliati et al. presented a study on the possible variations in findings pulmonary artifacts, using a pocket and a high-end sonographic machine. Both studies were conducted in the rheumatologic setting.

In 2015, Platz et al. examined possible variations in B-line detection by using a pocket and a high-end device in the echocardiography laboratory and found small and non-statistically significant differences between the devices. To maximize the number of detectable B-lines, we designed our study in the ED, with patients highly suspected of suffering from acute HF; thus the large mean number of B-lines, varying between 5.4 and 6.3, was not surprising.

Similarly to the study by Platz et al. (2015), we found 1.2 and 2.9 more B-lines in 4" and 7" clips, respectively, compared with a 2" clip (Table 2). This difference suggests that the current criterion for the diagnosis of ISy (*i.e.*, at least 2 positive zones [defined by the presence of 3 or more B-lines] per hemithorax) is incomplete, because the number of detected B-lines tends to increase with the clip length. This could have a clinical impact, as longer evaluations might potentially lead to false-positive results and, *vice versa*, shorter evaluations might potentially lead to false-negative results.

We also assessed the effects of the type of probe, the evaluation depth and the raters' expertise.

The only factor that did not influence the number of B-lines was the depth of the evaluation (0.03 more B-lines, using 19-cm versus 10-cm depth).

A difference of about two B-lines was detected, using the phased array instead of the convex probe. The literature indicates no preference for one or the other, but our results suggest that the type of probe could considerably affect ISy detection.

Expertise also appears to have an effect: expert raters reported on average one B-line fewer than naive operators.

We also assessed possible variations among examination areas. Current expert statements (Volpicelli et al. 2012) suggest scanning 4 zones per hemithorax. Our finding that lateral areas of each hemithorax showed more B-lines than the medial ones is easily explained by the semi-supine position of patients.

Our study has some limitations. We examined 50 nonconsecutive patients, exclusively enrolled when at least 1 of the authors (F.B. or E.P.) was present in the ED. However, considering the investigators' presence was randomly allocated during the ED shifts, we can exclude the fact that a particular pattern of enrollment (such as night versus day, weekday versus weekend) may have affected our results.

The factors considered in our study (namely, depth, clip length, probe type, lung areas, level of expertise) are only some of the possible sources of variability. We did not have the opportunity to investigate other potential factors, such as those related to patient characteristics (*e.g.*, body mass index).

Conclusion

Our study has shown that the type of probe, the evaluation duration and the level of operator's expertise might affect the quantitative assessment of B-lines, using LUS. We believe that the current criteria for the diagnosis of ISy could be revised in the light of our findings. For example, studies could be appropriately designed to determine the procedures that maximize LUS diagnostic accuracy.

Acknowledgments

We are grateful to Elisabetta Segre and David Hill for editing the manuscript.

References

1. T. Barskova, L. Gargani, S. Guiducci, S.B. Randone, C. Bruni, G. Carnesecchi, M.L. Conforti, F. Porta, A. Pignone, D. Caramella, E. Picano, M.M. Cerinic **Lung ultrasound for the screening of interstitial lung disease in very early systemic sclerosis** *Ann Rheum Dis*, 72 (2013), pp. 390-395
2. J.J. Bartko **The intraclass correlation coefficient as a measure of reliability** *Psychol Rep*, 19 (1966), pp. 3-11
3. D.J. Blehar, B. Barton, R.J. Gaspari **Learning curves in emergency ultrasound education** *Acad Emerg Med*, 22 (2015), pp. 574-582
4. V.A. Caiulo, L. Gargani, S. Caiulo, A. Fisicaro, F. Moramarco, G. Latini, E. Picano **Lung ultrasound in bronchiolitis: Comparison with chest X-ray** *Eur J Pediatr*, 170 (2011), pp. 1427-1433
5. C. Cogliati, M. Antivalle, D. Torzillo, S. Birocchi, A. Norsa, R. Bianco, G. Costantino, M.C. Ditto, M. Battellino, P.C. Sarzi Puttini, N. Montano **Standard and pocket-size lung ultrasound devices can detect interstitial lung disease in rheumatoid arthritis patients** *Rheumatology (Oxford)*, 53 (2014), pp. 1497-1503
6. Delle Sedie, M. Doveri, F. Frassi, L. Gargani, G. D'Errico, P. Pepe, L. Bazzichi, L. Riente, D. Caramella, S. Bombardieri **Ultrasound lung comets in systemic sclerosis: A useful tool to detect lung interstitial fibrosis** *Clin Exp Rheumatol*, 28 (2010), p. S54

7. Dunn, W.J. Fry **Ultrasonic absorption and reflection by lung tissue** *Phys Med Biol*, 5 (1961), pp. 401-410
8. J.L. Fleiss **Measuring nominal scale agreement among many raters** *Psychol Bull*, 76 (1971), pp. 378-382
9. Goffi, E. Pivetta, E. Lupia, G. Porrino, M. Civita, E. Laurita, G. Griot, G. Casoli, G. Cibinel **Has lung ultrasound an impact on the management of patients with acute dyspnea in the emergency department?** *Crit Care*, 17 (2013), p. R180
10. D. Lichtenstein, G. Mézière, P. Biderman, A. Gepner, O. Barré **The Comet-tail artifact: An ultrasound sign of alveolar-interstitial syndrome** *Am J Respir Crit Care Med*, 156 (1997), pp. 1640-1646
11. D.A. Lichtenstein **BLUE-protocol and FALLS-protocol: Two applications of lung ultrasound in the critically ill** *Chest*, 147 (2015), pp. 1659-1670
12. R.C. MacCallum, C. Kim, W.B. Malarkey, J.K. Kiecolt-Glaser **Studying multivariate change using multilevel models and latent curve models** *Multivariate Behav Res*, 32 (1997), pp. 215-253
13. J.L. Martindale, A. Wakai, S.P. Collins, P.D. Levy, D. Diercks, B.C. Hiestand, G.J. Fermann, I. deSouza, R. Sinert **Diagnosing acute heart failure in the emergency department: A systematic review and meta-analysis** *Acad Emerg Med*, 23 (2016), pp. 223-242
14. V.E. Noble, L. Lamhaut, R. Capp, N. Bosson, A. Liteplo, J.-S. Marx, P. Carli **Evaluation of a thoracic ultrasound training module for the detection of pneumothorax and pulmonary edema by prehospital physician care providers** *BMC Med Educ*, 9 (2009), p. 3
15. V.E. Noble, A.F. Murray, R. Capp, M.H. Sylvia-Reardon, D.J.R. Steele, A. Liteplo **Ultrasound assessment for extravascular lung water in patients undergoing hemodialysis. Time course for resolution** *Chest*, 135 (2009), pp. 1433-1439
16. E. Pivetta, A. Goffi, E. Lupia, M. Tizzani, G. Porrino, E. Ferreri, G. Volpicelli, P. Balzaretto, A. Banderali, A. Iacobucci, S. Locatelli, G. Casoli, M.B. Stone, M.M. Maule, I. Baldi, F. Merletti, G.A. Cibinel, P. Baron, S. Battista, G. Buonafede, V. Busso, A. Conterno, P. Del Rizzo, P. Ferrera, P.F. Pecetto, C. Moiraghi, F. Morello, F. Steri, G. Ciccone, C. Calasso, M.A. Caserta, M. Civita, C. Condo, V. D'Alessandro, S. Del Colle, S. Ferrero, G. Griot, E. Laurita, A. Lazzerro, F. Lo Curto, M. Michelazzo, V. Nicosia, N. Palmari, A. Ricchiardi, A. Rolfo, R. Rostagno, F. Bar, E. Boero, M. Frascisco, I. Micossi, A. Mussa, V. Stefanone, R. Agricola, G. Cordero, F. Corradi, C. Runzo, A. Soragna, D. Sciuillo, D. Vercillo, A. Allione, N. Artana, F. Corsini, L. Dutto, G. Lauria, T. Morgillo, B. Tartaglino, D. Bergandi, I. Cassetta, C. Masera, M. Garrone, G. Ghiselli, L. Ausiello, L. Barutta, E. Bernardi, A. Bono, D. Forno, A. Lamorte, D. Lison, B. Lorenzati, E. Maggio, I. Masi, M. Maggiorotto, G. Novelli, F. Panero, M. Perotto, M. Ravazzoli, E. Saglio, F. Soardo, A. Tizzani, P. Tizzani, M. Tullio, M. Ulla, E. Romagnoli, SIMEU Group for Lung Ultrasound in the Emergency Department in Piedmont **Lung ultrasound-implemented diagnosis of acute decompensated heart failure in the ED: A SIMEU multicenter study** *Chest*, 148 (2015), pp. 202-210
17. E. Platz, E. Pivetta, A.A. Merz, J. Peck, J. Rivero, Cheng S. **Impact of device selection and clip duration on lung ultrasound assessment in patients with heart failure** *Am J Emerg Med*, 33 (2015), pp. 1552-1556
18. Ponikowski, A.A. Voors, S.D. Anker, H. Bueno, J.G.F. Cleland, A.J.S. Coats, V. Falk, J.R. González-Juanatey, V.-P. Harjola, E.A. Jankowska, M. Jessup, C. Linde, P. Nihoyannopoulos, J.T. Parissis, B. Pieske, J.P. Riley, G.M.C. Rosano, L.M. Ruilope, F. Ruschitzka, F.H. Rutten, P. van der Meer, Authors/Task Force Members **2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC** *Eur Heart J*, 37 (2016), pp. 2129-2200

19. G. Soldati, A. Smargiassi, R. Inchingolo, S. Sher, R. Nenna, S. Valente, C.D. Inchingolo, G.M. Corbo **Lung ultrasonography may provide an indirect estimation of lung porosity and airspace geometry** *Respiration*, 88 (2014), pp. 458-468
20. G. Soldati, M. Demi, R. Inchingolo, A. Smargiassi, L. Demi **On the physical basis of pulmonary sonographic interstitial syndrome** *J Ultrasound Med*, 35 (2016), pp. 2075-2086
21. G. Volpicelli, A. Mussa, G. Garofalo, L. Cardinale, G. Casoli, F. Perotto, C. Fava, M. Frascisco **Bedside lung ultrasound in the assessment of alveolar-interstitial syndrome** *Am J Emerg Med*, 24 (2006), pp. 689-696
22. G. Volpicelli, M. Elbarbary, M. Blaivas, D.A. Lichtenstein, G. Mahis, A.W. Kirkpatrick, L. Melniker, L. Gargani, V.E. Noble, G. Via, A. Dean, J.W. Tsung, G. Soldati, R. Copetti, B. Bouhemad, A. Reissig, E. Agricola, J.-J. Rouby, C. Arbelot, A. Liteplo, A. Sargsyan, F. Silva, R. Hoppmann, R. Breitzkreutz, A. Seibel, L. Neri, E. Storti, T. Petrovic, International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS) **International evidence-based recommendations for point-of-care lung ultrasound** *Intensive Care Med*, 38 (2012), pp. 577-591
23. N. Xirouchaki, E. Magkanas, K. Vaporidi, E. Kondili, M. Plataki, A. Patrianakos, E. Akoumianaki, D. Georgopoulos **Lung ultrasound in critically ill patients: Comparison with bedside chest radiography** *Intensive Care Med*, 37 (2011), pp. 1488-1493