Inter-observer reliability for different point-of-care lung ultrasound findings in mechanically ventilated critically ill COVID-19 patients

Šustić, Alan; Mirošević, Marko; Szuldrzynski, Konstanty; Marčun, Robert; Haznadar, Mehmed; Podbegar, Matej; Protić, Alen

LETTER TO THE EDITOR

Inter-observer reliability for different point-of-care lung ultrasound findings in mechanically ventilated critically ill COVID-19 patients

Alan Šustić[1,2] · Marko Mirošević[2] · Konstanty Szuldrzynski[3] · Robert Marčun[2,4,5] · Mehmed Haznadar[6] · Matej Podbegar[5,7] · Alen Protić[2]

To the Editor

Main clinical manifestation of the COVID-19 disease is viral pneumonia with varying degrees of severity from mild symptoms to severe acute respiratory distress syndrome (ARDS) requiring mechanical ventilation and intensive care treatment. Computed tomography scans of patients with COVID-19 ARDS typically show diffuse bilateral interstitial pulmonary infiltrates, with asymmetric, patchy lesions distributed mainly in the periphery of the lung. In the context of the pandemic and the strain on resources and utilization, lung ultrasound (LUS) has emerged as an alternative tool for diagnosing and monitoring COVID-19 patients. Typical LUS signs are heterogeneous B-lines clusters, an irregular or fragmented pleural line, pleural effusion, lung consolidation, and partial absence of lung sliding [1]. It was suggested that according to the LUS pattern, patients can be accurately divided into four groups; from pattern A which suggests a low probability of

COVID-19 disease to pattern D which indicates a high probability of disease [2]. However, use of LUS may be fraught with certain shortcomings, such as the considerable dependence on the observer's knowledge and skills, and dependence on the performance and settings of the ultrasound transducer. This operator-dependency is especially emphasized in the real-time point-of-care (POC) LUS examination in the ICU and, therefore, the comparability of LUS findings of different observers may be arguable in certain settings [3, 4].

In this pilot study, we analyzed the concordance between different observers in evaluating particular POC LUS findings in COVID-19 critically ill patients. We sent seventy-three different LUS video-clips (electronic supplementary material: five different examples of videos) for review by 10 observers from five European countries. The videos were recorded in three European countries by physicians who had no knowledge of the study; 22 clips were documented in Slovenia, 14 in Poland, and 37 in Croatia. All video-clips were recorded in mechanically ventilated, COVID positive adult patients with severe ARDS and were standardized as follows: all were recorded in the supine patient position, all lasted 10 s, and the LUS scans used were obtained in the medioclavicular line between second and fifth rib. The cross-sectional angle, choice of probe or frequency, as well as gain and depth of the ultrasound beam were left to the operator's choice. Observers who reviewed the video-clips were four from Croatia, two from Slovenia and Poland, and one from Serbia and Bosnia and Herzegovina. All ten observers were ICU physicians who self-identified as proficient LUS users (use on the regular basis for more than 1 year). However, the level of the education and years of experience in the application of (lung) ultrasound, as well as the type of the ultrasound machine and overall ICU environment differed significantly among the observers.

The observers were asked to view each video clip for not longer than 1 min and subsequently complete the accompanying questionnaire. The questionnaire, read by observers

✉ Alan Šustić
alan.sustic@uniri.hr

1 Department of Clinical Medical Science II, Faculty of Health Studies, University of Rijeka, Rijeka, Croatia

2 Department of Anesthesiology, Reanimatology, Intensive Care and Emergency Medicine, Faculty of Medicine, University of Rijeka, Br. Branchetta 20, 51 000 Rijeka, Croatia

3 Department of Anaesthesiology and Intensive Care, Central Clinical Hospital of the Ministry of Interior and Administration, Warsaw, Poland

4 University Clinic of Respiratory and Allergic Diseases Golnik, Golnik, Slovenia

5 Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

6 Cantonal Hospital Dr. "Safet Mujić", Mostar, Bosnia and Herzegovina

7 Department of Intensive Internal Medicine, General Hospital Celje, Celje, Slovenia

before the review, had 10 questions related to specific LUS findings according to Volppicelli et al. [2] with a simple yes and no answers, with the exception of the answer regarding to the number of B-lines, where observers had to enter the exact number of detected B-lines. At the end of the questionnaire, the examiner will have to provide a final grade on the probability of COVID-19, ranging from pattern A as a normal LUS finding or low probability for COVID-19, to pattern D representing a high probability of COVID 19 disease (Fig. 1). We evaluated inter-observer reliability for each question from the questionnaire using interclass correlation coefficients (ICC) according to Cicchetti (less than 0.40—poor; between 0.40 and 0.59—fair; between 0.60 and 0.74—good; between 0.75; and 1.00—excellent) [5].

The results of our study are somewhat unexpected and ICC for different LUS findings among observers (inter-rater reliability) is shown in Table 1. Comparing mutual agreement for each LUS finding between observers, we found a poor agreement for the following findings: pleural sliding, thickening of the pleural line with pleural line irregularity, observed A-lines, lobar consolidation with dynamic air bronchogram, tissue-like consolidation without bronchogram and peripheral lung consolidations. For the number of B-lines and the LUS finding of lung consolidation, we found an acceptable inter-rater agreement, while good concordance has been found in detecting the presence of B lines (B-pattern) and pleural effusion. Interrater reliability in the overall assessment of the likelihood of the presence of COVID 19

disease was also poor. Our analysis suggests that the general agreement between the different observers is unsatisfactory, especially for certain LUS findings. We found poor consistency of pleural-related findings, either pleural movement or pleural appearance. This can be explained by the fact that the recognition of pleural sliding in patients on lung protective ventilation with small tidal volume requires some experience and can be challenging, while the assessment of pleural thickening and irregularity is partly subjective. An acceptable interrater agreement on the presence of a lung

Table 1 Interclass correlation (ICC) for different LUS findings among observers (inter-rater reliability)

Variable		ICC (95% CI)	
1	Regular pleural sliding	0.330 (0.245–0.433)	Poor
2	Thickening of the pleural line	0.271 (0.194–0.369)	Poor
3	A-lines observed	0.230 (0.153–0.327)	Poor
4	B-lines pattern present	0.674 (0.595–0.753)	Good
5	No. of B lines	0.502 (0.366–0.611)	Fair
6	Lung consolidation	0.509 (0.384–0.582)	Fair
7	Lobar consolidation	0.247 (0.147–0.341)	Poor
8	Tissue like consolidation	0.302 (0.223–0.402)	Poor
9	Peripheral lung consolidation	0.156 (0.097–0.236)	Poor
10	Pleural effusion	0.693 (0.606–0.784)	Good
11	Probability of COVID-19 disease	0.198 (0.127–0.291)	Poor

LUS findings	Yes	No
1. Regular pleural sliding		
2. Thickening of the pleural line with pleural line irregularity		
3. A-lines observed		
4. B-lines pattern present		
Number of B lines		
5. Lung consolidation		
Lobar consolidation with dynamic air bronchogram		
Tissue like consolidation without bronchogram (obstructive atelectasis)		
Peripheral lung consolidations		
6. Pleural effusion		
Probability of COVID-19 disease according to LUS findings (A-D)		

Fig. 1 The questionnaire related to specific LUS findings

consolidation but a poor agreement on the type of consolidation present can also be similarly explained. Namely, the assessment of the presence of pulmonary consolidation is less demanding and less subjective than the assessment of whether it is an obstructive atelectasis, lobar or peripheral consolidation. As expected, good inter-observer comparability was found in the assessment of the B pattern and the presence of pleural effusion, as these LUS signs are likely to be the easiest to detect. Inter-observer reproducibility in quantifying the B line was acceptable and approximately equal to the one previously published in a similarly designed study [4]. Finally, a mutual agreement on the overall estimated likelihood of COVID-19 disease according to the LUS findings was correlated poorly between the observers.

Obviously, the most important drawback of the study is the lack of specific training related to non-standard (COVID-specific) findings of LUS (beyond B lines and pleural effusion) for observers. The results of our study suggest that the use of lung POCUS for this purpose probably requires special training. Also, an important limitation of the study is the lack of the standardization of machine settings including gain (overall or TGC), focal point, frequency, probe selection and depth during recording various video clips which can significantly affect the results of each observer [6].

Overall, the study points towards the likely discrepancy in the concordance of LUS findings between different observers from different ICUs, and further studies are needed to define the possible implications of these results.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10877-021-00726-9.

References

1. Volpicelli G, Gargani L. Sonographic signs and patterns of COVID-19 pneumonia. Ultrasound J. 2020;12:22. https://doi.org/10.1186/s13089-020-00171-w.
2. Volpicelli G, Lamorte A, Vilen T. What's new in lung ultrasound during the COVID-19 pandemic. Intensive Care Med. 2020;46:1445–8. https://doi.org/10.1007/s00134-020-06048-9.
3. Yamaga S, Ohshimo S, Shime N. Is lung ultrasonography really useful for diagnosing weaning-induced pulmonary oedema? Intensive Care Med. 2019;45:1329. https://doi.org/10.1007/s00134-019-05663-5.
4. Pičuljan A, Šustić M, Brumini G, Kuharić J, Šustić A. Reliability of B-line quantification by different-level observers and a software algorithm using point-of-care lung ultrasound. J Clin Monit Comput. 2020;34:1259–64. https://doi.org/10.1007/s10877-019-00440-7.
5. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6:284–90. https://doi.org/10.1037/1040-3590.6.4.284.
6. Schmickl CN, Menon AA, Dhokarh R, Seth B, Schembri F. Optimizing B-lines on lung ultrasound: an in-vitro to in-vivo pilot study with clinical implications. J Clin Monit Comput. 2020;2020(34):277–84. https://doi.org/10.1007/s10877-019-00321-z.