

# Interoperator Reliability of Lung Ultrasound during the COVID-19 Pandemic

## Inter-Operator-Reliabilität des Lungensultraschalls in der COVID-19-Pandemie

### Authors

Chiara De Molo<sup>1</sup>, Silvia Consolini<sup>2</sup>, Veronica Salvatore<sup>2</sup>, Alice Grignaschi<sup>2</sup>, Antonella Lanotte<sup>2</sup>, Livia Masi<sup>2</sup>, Fabrizio Giostra<sup>2</sup>, Carla Serra<sup>1</sup>

### Affiliations

- 1 Interventional, Diagnostic and Therapeutic Ultrasound Unit, Department of Medical and Surgical Sciences, Sant'Orsola Malpighi Hospital, Bologna, Italy
- 2 Emergency department, Azienda Ospedaliero-Universitaria di Bologna Policlinico Sant'Orsola-Malpighi, Bologna, Italy

### Key words

COVID-19, lung ultrasound, chest, LUS score, interoperator reliability

received 13.07.2020

accepted 10.03.2021

published online 15.04.2021

### Bibliography

Ultraschall in Med 2023; 44: 75–80

DOI 10.1055/a-1452-8379

ISSN 0172-4614

© 2021, Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

### Correspondence

Dr. Silvia Consolini

Pronto Soccorso, Azienda Ospedaliero-Universitaria di Bologna Policlinico Sant'Orsola-Malpighi, via Pietro Albertoni 15, 40138 Bologna, Italy  
Tel.: +39/51/2 14 47 20  
silvia.91@hotmail.it

### ABSTRACT

**Aim** Lung ultrasound (LUS) is a reliable, radiation-free, and bedside imaging technique used to assess several pulmonary diseases. Although COVID-19 is diagnosed with a nasopharyngeal swab, detection of pulmonary involvement is crucial for safe patient discharge. Computed tomography (CT) is currently the gold standard. To treat paucisymptomatic patients, we have implemented a “fast track” pathway in our emergency department, using LUS as a valid alternative. Minimal data is available in the literature about interobserver reliability and the level of expertise needed to perform a reliable examination. Our aim was to assess these.

**Materials and Methods** This was a single-center prospective study. We enrolled 96 patients. 12 lung areas were explored in each patient with a semiquantitative assessment of pulmonary aeration loss in order to obtain the LUS score. Scans were performed by two different operators, an expert and a novice, who were blinded to their colleague's results.

**Results** 96 patients were enrolled. The intraclass correlation coefficient (ICC) showed excellent agreement between the expert and the novice operator (ICC 0.975; 0.962–0.983); demographic features (age, sex, and chronic pulmonary disease) did not influence the reproducibility of the method. The ICC was 0.973 (0.950–0.986) in males, 0.976 (0.959–0.986) in females; 0.965 (0.940–0.980) in younger patients ( $\leq 46$  yrs), and 0.973 (0.952–0.985) in older ( $> 46$  yrs) patients. The ICC was 0.967 (0.882–0.991) in patients with pulmonary disease and 0.975 (0.962–0.984) in the other patients. The learning curve showed an increase in interobserver agreement.

**Conclusion** Our results confirm the feasibility and reproducibility of the method among operators with different levels of expertise, with a rapid learning curve.

### ZUSAMMENFASSUNG

**Ziel** Lungensultraschall (LUS) ist ein zuverlässiges, strahlungsfreies und bettseitiges Bildgebungsverfahren, das zur Beurteilung verschiedener Lungenerkrankungen eingesetzt wird. Obwohl die Diagnose von COVID-19 mittels Nasopharyngealabstrich erfolgt, ist der Nachweis einer pulmonalen Beteiligung für die sichere Entlassung des Patienten von entscheidender Bedeutung. Die Computertomografie (CT) ist derzeit der Goldstandard. Um paucisymptomatische Patienten zu behandeln, haben wir in unserer Notaufnahme einen „Fast Track“-Pfad implementiert, der LUS als valide Alternative nutzt. In der Literatur gibt es nur wenige Daten zur Inter-Observer-Reliabilität und den Grad der Expertise, die für eine zuverlässige Untersuchung erforderlich sind. Unser Ziel war es daher, diese zu bewerten.

**Material und Methoden** Es handelte sich um eine prospektive Single-Center-Studie. Wir haben 96 Patienten eingeschlossen. Bei jedem Patienten wurden 12 Lungenbereiche mit semiquantitativer Bewertung des pulmonalen Belüftungsverlustes untersucht, um den LUS-Score zu erhalten. Die

Scans wurden von 2 verschiedenen Anwendern, einem Experten und einem Neuling, durchgeführt, die gegenüber den Ergebnissen ihres Kollegen verblindet waren.

**Ergebnisse** 96 Patienten wurden in die Studie aufgenommen. Der Intraklassen-Korrelationskoeffizient (ICC) zeigte eine ausgezeichnete Übereinstimmung zwischen dem Experten und dem Neuanwender (ICC 0,975; 0,962–0,983). Demografische Merkmale (Alter, Geschlecht und chronische Lungenerkrankung) hatten keinen Einfluss auf die Reproduzierbarkeit der Methode. Der ICC betrug 0,973 (0,950–0,986) bei Männern;

0,976 (0,959–0,986) bei Frauen; 0,965 (0,940–0,980) bei jüngeren Patienten ( $\leq 46$  Jahre) und 0,973 (0,952–0,985) bei älteren Patienten ( $> 46$  Jahre). Der ICC betrug 0,967 (0,882–0,991) bei Patienten mit Lungenerkrankungen und 0,975 (0,962–0,984) bei den anderen Patienten. Die Lernkurve zeigte einen Anstieg der Inter-Observer-Übereinstimmung.

**Schlussfolgerung** Unsere Ergebnisse bestätigen die Durchführbarkeit und Reproduzierbarkeit der Methode bei Anwendern mit unterschiedlichem Fachwissen mit einer schnellen Lernkurve.

## Introduction

Early diagnosis and follow-up of pneumonia are essential in the management of patients with COVID-19 infection. To date, the current literature does not advocate lung ultrasound (LUS) for the diagnosis of COVID-19. However, its role in monitoring lung conditions is well known, especially in the critical care setting [1].

Chest radiography (CXR) is a less sensitive modality for the detection of COVID-19 lung disease compared to chest computed tomography (CT), with a reported baseline CXR sensitivity of 69% [2]. The most commonly reported CXR and CT findings of COVID-19 include lung consolidation and ground glass opacities. Ground glass densities observed on CT may often have a correlate that is extremely difficult to detect on CXR [3]. Furthermore, CT has been shown not only to be more sensitive than reverse transcriptase polymerase chain reaction (RT-PCR) with a sensitivity of 98% vs. 71% in the diagnosis of COVID-19, but also to correlate with disease progression and recovery [4–6].

Patients typically have bilateral multilobar ground-glass opacities (GGO) with a peripheral or posterior distribution, and lesions detected on CT usually progress, with the greatest severity of radiologic findings occurring around day 10 of symptom onset [7]. Despite its utility, CT is not readily available in many settings with limited resources. In addition, the need for disinfection of the CT scanners after they are used for suspected or confirmed COVID-19-positive patients results in delays in patient care. Moreover, the risk related to radiological exposure has to be considered, especially in view of these patients potentially needing repeated scans for clinical decision-making.

COVID-19 interstitial diffuse pneumonia typically involves the lung periphery, a feature that makes it particularly suitable for ultrasound investigation [8]. LUS examination allows rapid and reliable diagnosis of lung consolidation, pleural effusion, and interstitial–alveolar syndrome, and it is currently included in a consensus statement on core competencies in point-of-critical-care ultrasound [9].

Among the advantages of LUS, the lack of radiation, simplicity, and execution speed have to be highlighted. Compared to XCR, LUS also appears to be more sensitive for an early detection of interstitial syndrome. It can be performed at bedside, which is especially useful for sequential monitoring of patients in critical settings. Moreover, in experimental models of acute respiratory distress syndrome (ARDS), LUS has proved capable of detecting

lung lesions before the development of hypoxemia [10]. Furthermore, in COVID infection settings, it is important to stress that LUS involves and exposes a minimum number of health-care workers and medical devices to suspected or confirmed cases, thus avoiding contamination of infrastructures and nosocomial spreading of the virus. The main limitations of LUS lie in the areas of training, operator variability, and reliability. Ultrasound qualitative description is operator-dependent and based on artifacts due to the presence of air in the alveolar tissue, with only images that reach the lung surface and the pleura being detected. Moreover, the acoustic window may be limited by the rib cage. According to previous studies, LUS must be performed according to a systematic examination protocol [11–13].

In the setting of COVID-19 pneumonia, some typical patterns have been described in the literature [8–14] highlighting the features of pulmonary interstitial involvement:

- **B-lines** in COVID-19 pneumonia are visualized as clusters of B-lines, both in separate and coalescent forms, sometimes giving the appearance of a shining white lung with a bilateral specific patchy distribution that can be described as a “patchwork pattern”. Mixed with a normal A pattern nearby, different from cardiac pulmonary edema, which shows a homogeneous increase in B-lines with gravity
- **Pleural line** appears typically irregular or fragmented
- **Small peripheral consolidations** can be frequently visualized

Minimal data are available in the literature about interobserver reliability, need for training, and level of expertise needed in order to perform a trustworthy examination. Similarly, there is currently no agreement as to what constitutes an experienced LUS operator.

The reliability of counting B-lines has been previously evaluated comparing expert versus novice operator performance and expert performance versus the performance of a software algorithm, and  $\kappa$  was 0.66–0.80 (95% CI). Intra-rater reliability has also been evaluated, ranging from 0.82 to 0.95 using the ICC (Intraclass Correlation Coefficient) [15–17].

When considering the identification of pleural effusions, it has been shown that young doctors in training can be easily taught to reliably perform thoracic ultrasound to answer specific diagnostic questions and to guide safe intervention procedures [18].

Despite increasing interest in the technique, LUS training methods vary among centers and are not standardized. Based on clinical experience, it has been hypothesized that a short and easy-

to-implement training program based on 25 LUS determinations supervised by experts would be enough for trainees without expertise [19].

In order to provide early treatment for paucisymptomatic patients with COVID-19 infection, we have implemented a “fast track” pathway in our emergency department in collaboration with infectious diseases specialists. Together with the medical examination and oxygenation parameters test, bedside LUS was performed to confirm or exclude the presence of viral pneumonia.

The aim of this study was to evaluate the reproducibility of LUS in assessing the features of COVID-19 pneumonia among operators with different levels of expertise and to investigate any training effect.

## Materials and methods

### Study design, population, and setting

This study was a single-center prospective study. Patients with suspected COVID-19 who were referred to our “fast track” were enrolled to evaluate the ability of LUS to discriminate the presence of interstitial pneumonia in association with nasopharyngeal swab and clinical evaluation.

LUS was performed by two different operators, each performing repeat measurements on the same subject on the same day and blinded to their colleague’s results.

The “expert” operator was a clinician with more than 5 years of LUS experience, while the “novice” operator was a resident trainee in the emergency department with no expertise in LUS but at least 1 year of abdominal ultrasound expertise. The “novice” operator received a brief explanation and a dedicated training course of 60 minutes on how to perform the measurements. All operators were blinded to the results obtained by others.

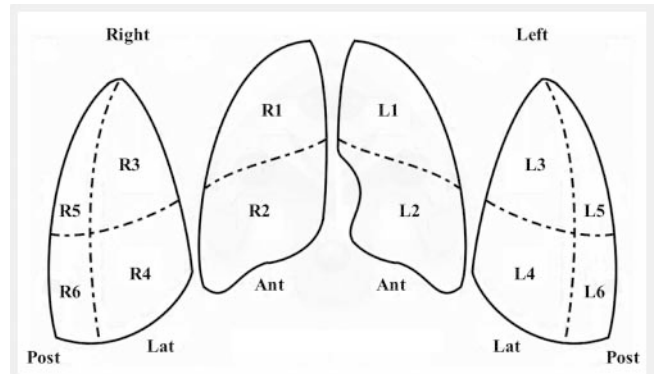
### Ultrasound imaging protocol

The US scans were all performed with Esaote My Lab 7. R

The convex probe (3–5 MHz) was used, with the widest acoustic window and maximum depth of 10 cm, with the focus on the pleural line. The exam was performed in a sitting, lateral, or supine position. As already reported in the literature, we have been using a systematic protocol of scanning that is rapid and practical[20]. In each patient 12 areas were explored once by the operator and reported in the LUS. The areas were registered as right and left lung areas (R and L) and divided using the anterior and posterior axillary lines, resulting in three areas per hemithorax (anterior, lateral, and posterior), with each of them split into superior and inferior. Each area was examined in the sagittal and axial views (► Fig. 1).

All lung areas were explored, and a score was defined for each one. The semiquantitative assessment of pulmonary aeration loss can vary between 0 and 36:

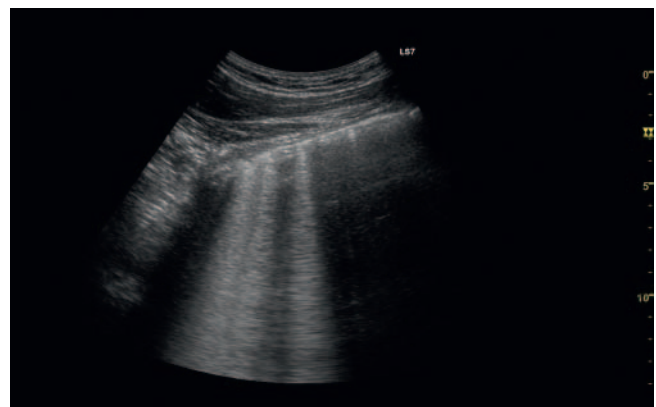
- 0: normal lung (A-lines) (► Fig. 2);
- 1: non-coalescent B-lines (B-lines occupying less than 50% of the intercostal space in the transverse plane) (► Fig. 3);
- 2: coalescent B-lines (B-lines occupying more than 50% of the intercostal space) (► Fig. 4);
- 3: consolidation > 1 cm (► Fig. 5).



► Fig. 1 Lung areas: both the right lung and the left lung are divided using the anterior and posterior axillary lines, resulting in three areas per hemithorax (anterior, lateral, and posterior), with each of them split into superior and inferior.



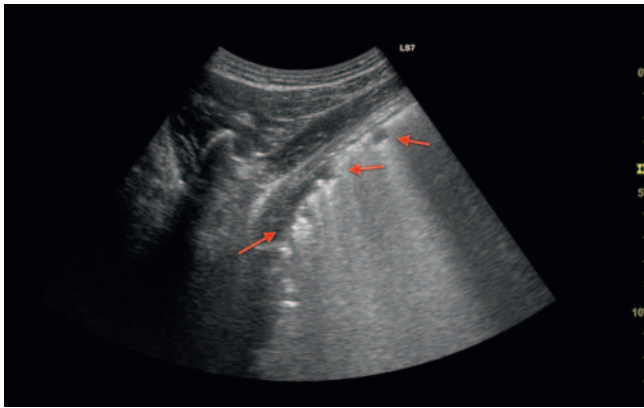
► Fig. 2 LUS score 0, normal A line pattern.



► Fig. 3 LUS score 1, non-coalescent B-lines (B-lines occupy less than 50% of the intercostal space in the transverse plane).

The sum of the 12 different lung areas represents the LUS score (LUSs).

For each operator, the whole examination lasted approximately 15 minutes.



► **Fig. 4** LUS score 2, coalescent B-lines (B-lines occupy more than 50 % of the intercostal space); the arrows show the pleural irregularity.



► **Fig. 5** LUS score 3, consolidation > 1 cm.

## Main outcome measurements

Interoperator variability was assessed by comparing the LUSs obtained on the same day by each operator. To establish if a training effect was present, the interoperator agreement in the last 5 days was compared with the agreements found in the first 5 days.

## Statistical analysis

Mean, median, standard deviation (SD), interquartile range (IQR), and frequencies were used as descriptive statistics.

Agreement between the two operators was expressed using intraclass correlation coefficients (ICCs) for single measurements and a 95% confidence interval (CI) for the ICC. Systematic differences were computed by means of the paired Student t-test. Bland-Altman plots were constructed to visualize agreement and the limits of agreement were evaluated together with their 95% CI (18). The repeatability coefficients were also computed. Statistical analysis was performed using SPSS (version 13.0 for Windows) and two-tailed P-values less than 0.05 were considered significant.

► **Table 1** Sample descriptive table of the population.

	all patients
<b>gender</b>	
male	39 (41 %)
female	57 (59 %)
<b>age</b>	
RANGE	15–88
median	46
≤ 46 yrs	50 (52 %)
> 46 yrs	46 (48 %)
<b>chronic pulmonary diseases</b>	
no	85
yes	11

## Results

96 patients were enrolled during the COVID-19 outbreak between April 17 and April 27, 2020. The median age was 46 years old (IQR 15–88); 39 males (41 %) and 57 females (59 %). Among the 96 patients, only 11 had known pulmonary disease: 1 case of COPD, 6 cases of asthma, and 4 other lung conditions.

No patients were excluded due to a suboptimal acoustic window.

The clinical characteristics of the patient population are described in ► **Table 1**.

The LUSs was 0 in 58 patients (60.4 %), ranged 1–3 in 24 patients (25.0 %), and ranged 4–14 in 13 patients (13.6 %) when performed by the expert operator. The LUSs for the novice operator was 0 in 56 patients (58.3 %), ranged 1–3 in 27 patients (28.1 %), and ranged 4–17 in 13 patients (13.6 %).

The ICC showed excellent agreement between the expert and the novice operator (ICC 0.975; 0.962–0.983) (► **Fig. 6**). As shown by the Bland-Altman analysis, both operators achieved a similar LUSs in the majority of patients, with only two cases showing a difference of 3 points. In particular, the novice operator overestimated the LUSs compared to the expert operator by a maximum of 3 points in only two cases, 17 instead of 14 and 8 instead of 5. A maximum underestimation of 2 points was reported by the novice operator (► **Fig. 7**).

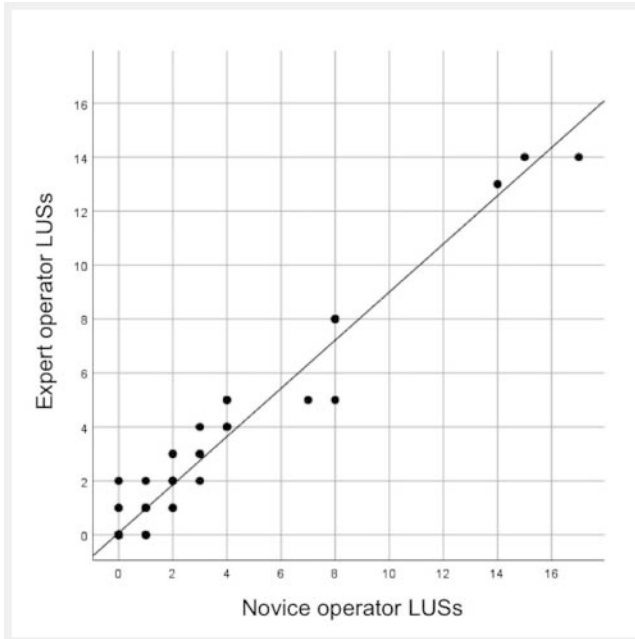
In 5 cases the novice operator reported a pathologic ultrasound with an LUSs of 1 compared to a negative score of the expert operator. Furthermore, the novice operator reported 3 cases as having a negative ultrasound, while the expert operator described an LUSs of 1 in two cases and an LUSs of 2 in one case.

Age, gender, and chronic pulmonary diseases did not influence the reported LUSs. In particular, the ICC was 0.973 (0.950–0.986) in males and 0.976 (0.959–0.986) in females.

The ICC in younger patients (≤ 46 yrs) was comparable with that seen in older patients (> 46 yrs): 0.965 (0.940–0.980) vs. 0.973 (0.952–0.985). When considering the influence of pulmo-

nary disease, the ICC of affected patients was 0.967 (0.882–0.991), comparable with the other patients with an ICC of 0.975 (0.962–0.984). In all analyses the difference was not statistically significant,  $p < 0.001$ .

In order to test a potential improvement in the novice operator learning curve, we divided the study population into two subpopulations: the first half (48 patients) and the second half (48 patients). The ICC was 0.971 (0.949–0.984) for the first subpopulation analysis and 0.981 (0.967–0.989) for the second subpopulation analysis,



► Fig. 6 Linear correlation between expert and novice.

showing a relative improvement in interobserver agreement (► Table 2).

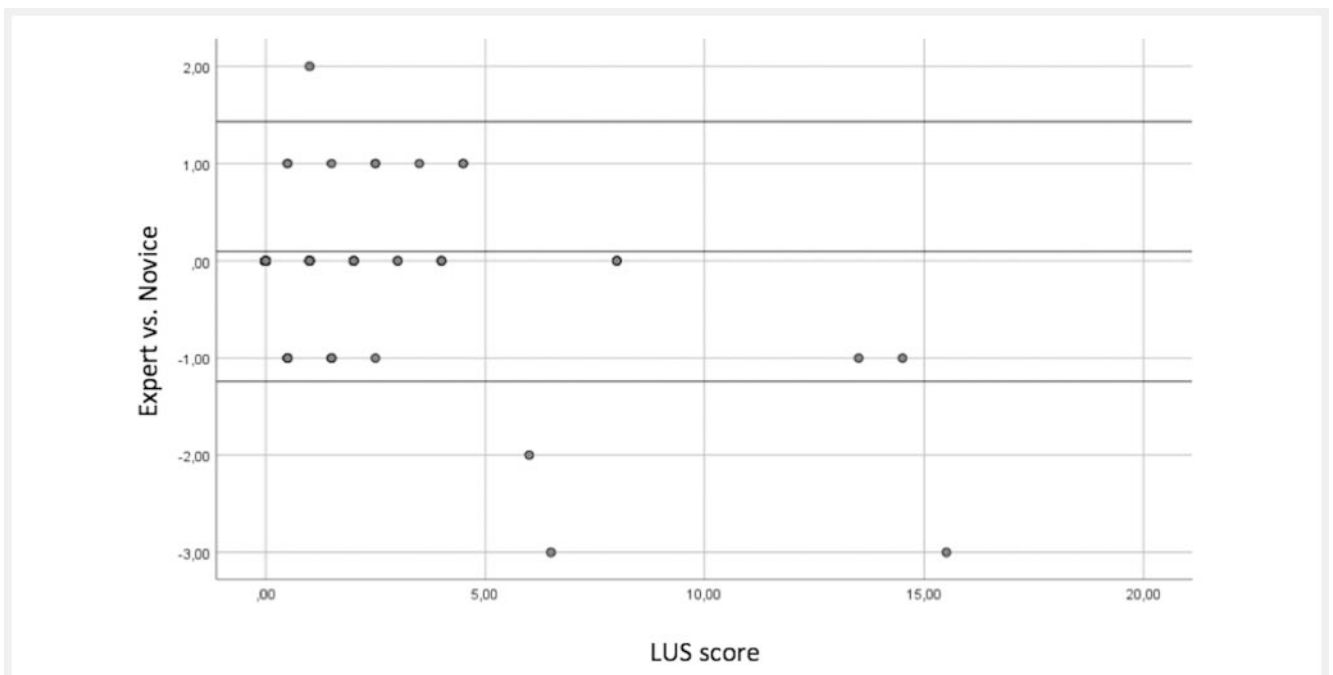
## Discussion

The main objective of this study was to assess the performance and reproducibility of LUS in patients suspected of having COVID-19 when conducted by operators with different levels of expertise.

Although the diagnosis of COVID-19 is made with a nasopharyngeal swab, detection of pulmonary involvement is essential in the context of COVID-19 infection in order to keep patients safe. So far, CT examination has been considered the gold standard in view of the high sensitivity and positive predictive values. However, it must be taken into consideration that this relates to the setting of a pandemic when there is a very high “a priori” probability of disease in the presence of respiratory symptoms [21].

Amidst the pandemic scenario characterized by limited technical resources, we consider ultrasound to be a valid alternative to CT, in particular for the assessment of paucisymptomatic patients. It is a safe and rapid method when used in conjunction with physical examination and arterial-blood gas analysis parameters. In particular, a negative LUS examination can allow the exclusion of pulmonary involvement with good accuracy [22].

Implementation of a standardized method is of utmost importance in order to use the technique widely. Our results confirm that this is feasible and reproducible even when performed by operators with different levels of expertise, including junior trainees with a brief LUS course with a very rapid learning curve. Furthermore, US has the advantage of being an easily portable system that could therefore also be used in out-of-hospital settings.



► Fig. 7 Bland-Altman plot representing the level of agreement between the two operators.



► **Table 2** Intra-class correlation between the two operators stratified according to different factors.

	ICC (95% CI) expert vs. novice	p-value
<b>all patients</b>	0.975 (0.962–0.983)	<0.001
<b>gender</b>		
male	0.973 (0.950–0.986)	<0.001
female	0.976 (0.959–0.986)	<0.001
<b>age</b>		
≤ 46 yrs	0.965 (0.940–0.980)	<0.001
> 46 yrs	0.973 (0.952–0.985)	<0.001
<b>chronic pulmonary diseases</b>		
no	0.967 (0.882–0.991)	<0.001
yes	0.975 (0.962–0.984)	<0.001
<b>learning curve</b>		
first period	0.971 (0.949–0.984)	<0.001
second period	0.981 (0.967–0.989)	<0.001

Demographic features did not influence the reproducibility of the method.

Considering the importance of giving a positive or negative imaging result, we saw a slight LUS discrepancy only in 8 patients. Greater discrepancies between the two operators have been registered in LUS > 5, but these can be considered not clinically relevant. Our study demonstrates that even a novice operator can easily exclude pulmonary involvement of the virus.

Limitations of the study are the lack of extremely high values of the LUSs in relation to the clinical features of our population and the lack of patients with possible confounding factors such as chronic pulmonary disease.

## Compliance with ethical standards

Ethical approval. The study protocol conforms to the ethical guidelines of the World Medical Association Declaration of Helsinki. All patients gave their consent, and the institutional review board of our hospital has been notified about the study.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

- [1] Wimalasena Y, Kocierz L, Strong D et al. Lung ultrasound: a useful tool in the assessment of the dyspnoeic patient in the emergency department. Fact or fiction? *Emerg Med J* 2018; 35 (4): 258–266. doi:10.1136/emered-2016-205937
- [2] Wong HYF, Lam HYS, Fong AH et al. Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. *Radiology* 2019; 201160: doi:10.1148/radiol.2020201160
- [3] Jacobi A, Chung M, Bernheim A et al. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clin Imaging* 2020; 64: 35–42. doi:10.1016/j.clinimag.2020.04.001
- [4] Fang Y, Zhang H, Xie J et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 2020; 200432: doi:10.1148/radiol.2020200432
- [5] Pan Y, Guan H, Zhou S et al. Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. *Eur Radiol* 2020. doi:10.1007/s00330-020-06731-x
- [6] Pan F, Ye T, Sun P et al. Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia. *Radiology* 2020; 200370: doi:10.1148/radiol.2020200370
- [7] Salehi S, Abedi A, Balakrishnan S et al. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. *Am J Roentgenol Am JRoentgenol* 2020; 1–7. doi:10.2214/AJR.20.23034
- [8] Volpicelli G, Gargani L. Sonographic signs and patterns of COVID-19 pneumonia. *Ultrasound J* 2020; 12: 22. doi:10.1186/s13089-020-00171-w
- [9] Mayo PH, Beaulieu Y, Doelken P et al. American College of Chest Physicians/La Société de Réanimation de Langue Française statement on competence in critical care ultrasonography. *Chest* 2009; 135 (4): 1050–1060. doi:10.1378/chest.08-2305
- [10] Soldati G, Smargiassi A, Inchingolo R et al. Is There a Role for Lung Ultrasound During the COVID-19 Pandemic? *J Ultrasound Med* 2020. doi:10.1002/jum.15284
- [11] Bouhemad B, Zhang M, Lu Q et al. Clinical review: Bedside lung ultrasound in critical care practice. *Crit Care* 2007; 11: 205
- [12] Bouhemad B, Mongodi S, Via G et al. Ultrasound for “lung monitoring” of ventilated patients. *Anesthesiology* 2015; 122: 437–447. doi:10.1097/ALN.0000000000000558
- [13] Dietrich CF, Mathis G, Cui XW et al. Ultrasound of the pleurae and lungs. *Ultrasound Med Biol* 2015; 41: 351–365. doi:10.1016/j.ultrasmedbio.2014.10.002
- [14] Volpicelli G, Lamorte A, Villén T. What's new in lung ultrasound during the COVID-19 pandemic. *Intensive Care Med* 2020. doi:10.1007/s00134-020-06048-9
- [15] Pičuljan A, Šuštić M, Brumini G et al. Reliability of B-line quantification by different-level observers and a software algorithm using point-of-care lung ultrasound. *J Clin Monit Comput* 2019. doi:10.1007/s10877-019-00440-7
- [16] Anderson KL, Fields JM, Panebianco NL et al. Inter-rater reliability of quantifying pleural B-lines using multiple counting methods. *J Ultrasound Med* 2013; 32: 115–120
- [17] Gullett J, Donnelly JP, Sinert R et al. Interobserver agreement in the evaluation of B-lines using bedside ultrasound. *J Crit Care* 2015; 30: 1395–1399. doi:10.1016/j.jcrr.2015.08.021
- [18] Moy RJ, Chapman AL, Bapusamy A. The Effectiveness of an Informal Teaching Programme for Junior Doctors Identifying Pleural Effusions Using Ultrasound at the Bedside. *JR Army Med Corps* 2010; 156: 233–235
- [19] Rouby JJ, Arbelot C, Gao Y et al. Training for lung ultrasound score measurement in critically ill patients. *Am J Respir Crit Care Med* 2018; 198: 398–401
- [20] Via G, Storti E, Gulati G et al. Lung ultrasound in the ICU: from diagnostic instrument to respiratory monitoring tool. *Minerva Anesthesiol* 2012; 78: 1282–1296. Epub 2012 Aug 3. PMID: 22858877
- [21] Piscaglia F, Stefanini F, Cantisani V et al. Benefits, Open questions and Challenges of the use of Ultrasound in the COVID-19 pandemic era. The views of a panel of worldwide international experts. *Ultraschall in Med* 2020. doi:10.1055/a-1149-9872
- [22] Tung-Chen Y, Marti de Gracia M, Diez-Tascon A et al. Correlation between chest computed tomography and lung ultrasonography in patients with coronavirus disease 2019 (covid-19). *Ultrasound Med Biol* 2020; 46: 2918–2926