Interoperator Reliability of Lung Ultrasound during the COVID-19 Pandemic

Inter-Operator-Reliabilität des Lungenultraschalls in der COVID-19-Pandemie

Authors
Chiara De Molo1, Silvia Consolini2, Veronica Salvatore2, Alice Grignaschi2, Antonella Lanotte2, Livia Masi2, Fabrizio Giostra2, Carla Serra1

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

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 (<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

Material und Methoden
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 0.97 (0.952–0.985) at baseline and 0.973 (0.952–0.985) at day 10 of symptom onset [7].

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-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 healthcare 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 k 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).

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

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

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

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

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**


