



HAL
open science

Comparison between Lung Ultrasonography Score in the Emergency Department and Clinical Outcomes of Patients with or Suspected COVID-19: An Observational Multicentric Study

Mehdi Benchoufi, Jérôme Bokobza, Anthony Chauvin, Elisabeth Dion, Marie-Laure Baranne, Fabien Levan, Maxime Gautier, Delphine Cantin, Thomas d'Humières, Cedric Gil-Jardine, et al.

► To cite this version:

Mehdi Benchoufi, Jérôme Bokobza, Anthony Chauvin, Elisabeth Dion, Marie-Laure Baranne, et al.. Comparison between Lung Ultrasonography Score in the Emergency Department and Clinical Outcomes of Patients with or Suspected COVID-19: An Observational Multicentric Study. *Journal of Ultrasound in Medicine*, 2023, 10.1002/jum.16329 . hal-04278973

HAL Id: hal-04278973

<https://hal.science/hal-04278973>

Submitted on 10 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

WE'VE GOT YOUR SOLUTION

KILLS TB IN 1 MINUTE

Tristel™ DUO DISINFECTANT

Fast-acting, cost-effective

Tristel DUO is designed to provide fast-acting and cost-effective cleaning and disinfection for instruments and non-critical medical devices including ultrasound transducers, holders, cables, keyboards, and ultrasound stations. With a maximum kill time of 2 minutes for bacteria and a minimum time of just 1 minute for tuberculosis and fungi, Tristel DUO really packs a punch. Tested and proven effective against HPV¹, Tristel DUO bridges the gap where low-level disinfection is insufficient and high-level disinfection is unnecessary.

ALCOHOL FREE • BLEACH FREE

The Power is in the Process

- Patented dispensing system generates the active ingredient – chlorine dioxide (ClO₂).
- ClO₂ destroys pathogens by separating electrons from microorganisms' vital structures resulting in molecular imbalance and microorganism death.
- Simply dispense, apply, and allow to dry. Use with any dry wipe.



Tristel DUO is manufactured and marketed in the United States by Parker Laboratories, Inc. as licensed by Tristel.

1. Meyers C, Milici J, Robison R. The ability of two chlorine dioxide chemistries to inactivate human papillomavirus-contaminated endocavitary ultrasound probes and nasendoscopes. J Med Virol. 2020 Aug;92(8):1298-1302. doi: 10.1002/jmv.25666. Epub 2020 Feb 4. PMID: 31919857; PMCID: PMC7497195.



Parker Laboratories, Inc.



The sound choice in patient care.™

973.276.9500

parkerlabs.com

Comparison between Lung Ultrasonography Score in the Emergency Department and Clinical Outcomes of Patients with or Suspected COVID-19

An Observational Multicentric Study

Mehdi Benchoufi, MD, PhD , Jérôme Bokobza, MD, Anthony Chauvin, MD, PhD, Elisabeth Dion, MD, PhD, Marie-Laure Baranne, MD, Fabien Levan, MD, Maxime Gautier, MD, Delphine Cantin, MD, Thomas d'Humières, MD, Cédric Gil-Jardiné, MD, Sylvain Benenati, MD, Mathieu Orbelin, MD, Mikaël Martinez, MD , Nathalie Pierre-Kahn, MD, Abdourahmane Diallo, PhD, Eric Vicaut, MD, PhD, Pierre Bourrier, MD

Received October 1, 2021, from the Center for Clinical Epidemiology, Hôtel-Dieu Hospital, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France (M.B., M.-L.B.); METHODS Team, Center for Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS-UMR 1153), Paris, France (M.B.); PICUS, Point of Care UltraSound Institute, Paris, France (M.B., J.B., M.-L.B., M.G.); Adult Emergency Department, Hôpital Cochin, Assistance Publique Hôpitaux de Paris, Paris, France (J.B., F.L., M.G.); Adult Emergency Department, Hôpital Lariboisière, Inserm U942 MASCOT, Université de Paris, Assistance Publique Hôpitaux de Paris, Paris, France (A.C.); Imaging Department Hôtel Dieu, Assistance Publique-Hôpitaux de Paris, Université de Paris, Paris, France (E.D., D.C., N.P.-K.); Centre de Recherche de l'Inflammation (CRI), INSERM U1149, Paris, France (E.D.); Physiology Department, Henri Mondor University Hospital, Créteil, France (T.d.); Adult Emergency Department SAMU-SMUR, Pellegrin Hospital, University Hospital Center, Bordeaux, France (C.G.-Jé.); Bordeaux Population Health, INSERM U1219, IETO Team, Bordeaux University, Bordeaux, France (C.G.-Jé.); Adult Emergency Department, Hospital Group South Ile-de-France, Melun, France (S.B.); Adult Emergency Department, New Civil Hospital, Strasbourg, France (M.O.); Adult Emergency Department, Forez Hospital Center, Montbrison, France (M.M.); Nord Emergency Network Ligérien Ardèche (REULIAN), Hospital Center Le Corbusier, Firminy, France (M.M.); Clinical Trial Unit Hospital, Lariboisière St-Louis AP-HP, Paris University, Paris, France (A.D., E.V.); and Imaging Department Saint-Louis Hospital, Assistance Publique des Hôpitaux de Paris (AP-HP), Paris, France (P.B.). Manuscript accepted for publication August 23, 2023.

†passed away.

Address correspondence to Mehdi Benchoufi, 1 Parvis Notre Dame, 75004, Paris, France.

E-mail: mehdi.benchoufi@aphp.fr

Abbreviations

AUC, area under the curve; Chest CT, chest computed tomography; GS, global score; ICU, intensive care unit; LUS, lung ultrasound; US, ultrasonography

doi:10.1002/jum.16329

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Objective—Chest CT is the reference test for assessing pulmonary injury in suspected or diagnosed COVID-19 with signs of clinical severity. This study aimed to evaluate the association of a lung ultrasonography score and unfavorable clinical evolution at 28 days.

Methods—The eChoVid is a multicentric study based on routinely collected data that was conducted in 8 emergency units in France; patients were included between March 19, 2020 and April 28, 2020 and underwent lung ultrasonography, a short clinical assessment by 2 emergency physicians blinded to each other's assessment, and chest CT. Lung ultrasonography consisted of scoring lesions from 0 to 3 in 8 chest zones, thus defining a global score (GS) of severity from 0 to 24. The primary outcome was the association of lung damage severity as assessed by the GS at day 0 and patient status at 28 days. Secondary outcomes were comparing the performance between GS and CT scan and the performance between a new trainee physician and an ultrasonography expert in scores.

Results—For the 328 patients analyzed, the GS showed good performance in predicting clinical worsening at 28 days (area under the receiver operating characteristic curve [AUC] 0.83, sensitivity 84.2%, specificity 76.4%). The GS showed good performance in predicting the CT severity assessment (AUC 0.84, sensitivity 77.2%, specificity 83.7%).

Conclusion—A lung ultrasonography GS is a simple tool that can be used in the emergency department to predict unfavorable assessment at 28 days in patients with COVID-19.

Key Words—COVID-19; lung ultrasound; lung ultrasound scoring

The clinical course of COVID-19 is marked by the predominance of respiratory symptoms and the occurrence of a major inflow of patients in hospitals. Significant clinical investigative resources, especially imaging tools, are now

being mobilized to assess and determine the best management of suspected or confirmed COVID-19.^{1,2} Learned radiology societies in France and the United States have considerably reduced the role of chest radiography for the initial assessment of patients with suspected or confirmed COVID-19.^{3,4} The increasing availability of handheld ultrasound devices⁵ and the non-irradiating nature of acoustic waves have led to considering ultrasonography (US) as a quick triage tool for suspected or diagnosed COVID-19.

Between March 2020 and August 2020, several studies documented the pulmonary US semiology of COVID-19.^{6,7} In July 2020, the WHO published a set of recommendations, mentioning lung US (LUS) as a tool of interest in the diagnosis.⁸ Some studies evaluated the interest of LUS for risk stratification of patients with diagnosed or suspected COVID-19.^{9–15} These studies evaluated the prognostic value of LUS with respect to severity outcomes such as respiratory support needs, intensive care unit (ICU) admission, and death, with lung scoring systems acquired on more than 10 chest zones. Most of these studies were mono-centric, with low sample sizes (in general $N < 100$, with the exception of¹⁶ with $N = 447$), rarely explored the clinical outcomes with sufficient follow-up (which matters when assessing death as an outcome), and led to more than 10 lung zone scores. Some studies showed an insufficient clinical prognostic value (again, mono-centric studies with low sample size).¹⁷ There is still a need to add to the scientific literature on the topic of LUS as a risk stratification tool in COVID-19, perhaps for inclusion in future meta-analyses.

In the eChoVid study, we use an 8 chest-zone scoring system. Buessler et al¹⁸ showed that to detect lung interstitial lesions in the context of acute heart failure, a 28-point LUS examination carried no specific benefits over a 6- or 8-point examination. Simplified lung scoring acquisition may be of interest when considering LUS as quick triage tool and also wide usage with its enhanced ease of use for non-expert operators.

Many studies showed a significant effect of a short training on the ability to perform LUS.¹⁹ The use of US, non-invasive, point of care, in real time, could facilitate the initial assessment of patients and management of patient flow.

We hypothesized that 1) a specific LUS score could be helpful in the emergency department for predicting the clinical evolution of COVID-19 at 28 days and 2) LUS could be safely performed by emergency physicians who were not experts in pulmonary US.

We performed a prospective study to evaluate 1) the association between a specific global score (GS) based on LUS for severity of COVID-19 and unfavorable clinical course at 28 days; 2) the performance of the GS and severity of lung damage according to chest CT scan; and 3) the performance, based on the GS, of a newly trained US operator and a US expert.

Methods

Study Design and Patient Selection

This was a multicentric, observational non-randomized study conducted in 8 emergency departments in France (Table 1). From March 19, 2020 to April 28, 2020, we enrolled patients referred to one of the recruitment centers because of initial clinical

Table 1. Investigation Centers and Material

Investigation Centers ^a		
Hospital	Location	
Cochin Hospital – APHP	Paris (75)	
Lariboisière Hospital – APHP	Paris (75)	
Henri Mondor Hospital – APHP	Créteil (94)	
Hôtel Dieu Hospital – APHP	Paris (75)	
Forez Hospital Center	Montbrison (42)	
Sud Ile-de-France Hospital Group	Melun (77)	
Strasbourg University Hospital	Strasbourg (67)	
Bordeaux University Hospital	Bordeaux (33)	
Material ^b		
Machine Name	Constructor	Probe Type
TE7	Mindray	Curved probe (2–5 MHz)
Ulight	Sonoscanner	Curved probe (2–5 MHz)
E2EXP	Sonoscape	Curved probe (3 MHz)
Spark	Philips	Curved probe (2–6 MHz)
VScan	GE	Linear probe (3.3–8.0 MHz)

^aThe multicentric, observational non-randomized study was conducted in 8 Emergency Departments located in France.

^bPhysicians used the ultrasonography available equipment with no specific requirement on machine performance.

suspicion of COVID-19 and with COVID-19 secondarily confirmed by RT-PCR. Exclusion criteria were patients for whom the LUS exploration could not be performed (ie, morbid obesity, extensive thoracic subcutaneous emphysema, absorbent subcutaneous infiltrations) or with any comorbidity that justified priority immediate intensive care.

Clinicians involved in the LUS assessment were emergency physician experts in pulmonary US (certification in US and more than 5 years' experience in US) or newly trained pulmonary US operators. All investigators had received the same curricula with regard to clinical US.

Intervention

After inclusion, each patient underwent both a clinical examination and LUS, each by a different emergency physician, with blinding to each other's findings. Emergency physicians who performed LUS were blinded to clinical, imaging, and biological patient data.

Each LUS evaluation consisted of a standardized exploration of 8 lung quadrants: right antero-superior, left antero-superior, right antero-inferior, left antero-inferior, right postero-superior, left postero-superior, right postero-inferior and left postero-superior.

For each quadrant, we defined 4 grades of severity (GS): *Grade 0*: up to a maximum of 3 observed B-lines; *Grade 1*: 4 to 8 B-lines, through the intercostal space at one of the pulmonary bases; *Grade 2*: B-lines in "curtain sign" (>8 B-lines) and/or diffusion of more than 4 B-lines in two-thirds of the pulmonary field; *Grade 3*: consolidation foci.

The US score used in our study for assessing lung condition was derived from the standard LUS score for 12 lung points.^{13,14} We simplified this score to 8 points on the upper and lower parts of the anterior and posterior regions of the left and right chest wall. Our GS, summing over 8 quadrants, ranged from 0 to 24 points. This simplification was deduced on the basis of the consensus in France with regard to LUS practices based on the Lichtenstein work,²⁰ with which all the investigators were trained.

For US evaluation, physicians used the available US equipment with no specific requirement for machine performance. All acquisitions were performed with curved probes; a handheld device (Vscan) was used in one of the sites.

Follow-up was performed on days 5, 15, and 28 to detect one of three outcomes: ICU admission, mechanical invasive ventilation, or death. If patients were not hospitalized, investigators contacted them by phone (or their trusted person in case of non-response).

Training of Trainees

Four participants never trained in pulmonary US completed a standardized training course: a 30-minute protocol of US theory with a review of pathology images from an image bank and practice on 5 patients with suspected COVID-19. Each lung evaluation performed by the trainee was repeated by an expert, with blinding to the trainee's results.

Data Collection and Data Sources

For each patient, the local investigator used a standardized computer form to collect patient demographic data, vital parameters at emergency department admission, details of the scoring for the 8 quadrants and the GS.

The results of the RT-PCR COVID-19 test and CT scan were collected. Each chest CT scan was reviewed by an expert radiologist, which led to a secondary analysis for confirmation of COVID-19 severity. The severity of lung damage on the CT scan was scored as normal (0%), minimal (0–10%), moderate (10–25%), and severe (>25%).²¹

Outcomes

The primary outcome was the association between GS at emergency department admission and clinical worsening, defined as the occurrence of at least one of ICU admission or mechanical invasive ventilation, or death within the 28-day follow-up.

Secondary outcomes were 1) a comparison of the performance of the GS and the severity of lung damage according to the chest CT scan and 2) the performance of a newly trained US operator and a US expert based on a comparison of the GS and score details of each quadrant in both groups.

Study Size and Statistical Analysis

A sample of 300 patients with documented level of severity would allow for estimating an area under the receiver operating characteristic curve (AUC) $\geq 85\%$ with precision $\pm 5\%$ or better. All quantitative data are presented with mean SD or median (Q1–Q3);

range) according to data distribution. Categorical data are summarized with number (%).

For the primary outcome, the association of clinical worsening and GS was evaluated by univariate logistic regression with the GS predicting clinical worsening. For each regression model, we calculated the AUC for the different values of GS.

The performance of the GS was evaluated by univariate logistic regression with the GS predicting CT disease quantification dichotomized as normal versus pathologic, normal or minimal versus moderate or severe, or normal, minimal or moderate versus severe. For each regression model, we calculated the AUC, Brier score, and Youden index for the different values of the scores. Model validation for calibration and discrimination ability involved bootstrap replications, and degrees of optimism were calculated for C statistics and the Brier score. For comparing GS and CT severity scores (normal or minimal vs moderate or severe, or normal, minimal, or moderate vs severe), we discarded patient data with collected CT status (pathological/normal) but missing CT severity score. For comparing the performance of a newly trained

US operator and a US expert, some patients were evaluated for GS by both an expert and a newly trained operator. We evaluated the agreement between them by calculating the weighted kappa²² of US severity grades for each quadrant and used the Bland and Altman method to evaluate agreement for LUS.

Statistical analysis involved using SAS 9.4 (SAS Institute Inc, Cary, NC) or R.4.1.0 $P < .05$ was considered statistically significant.

Ethics

According to French regulatory rules, the study was approved by the international review board (no. IORG0009855, no. CPP2020-03-035a/2020-A00768-31/20.03.24.70826). and no written informed consent was required from patients.

Results

A total of 409 patients with suspected or diagnosed COVID-19 between March 19, 2020 and April

Figure 1. Flow chart of included patients.

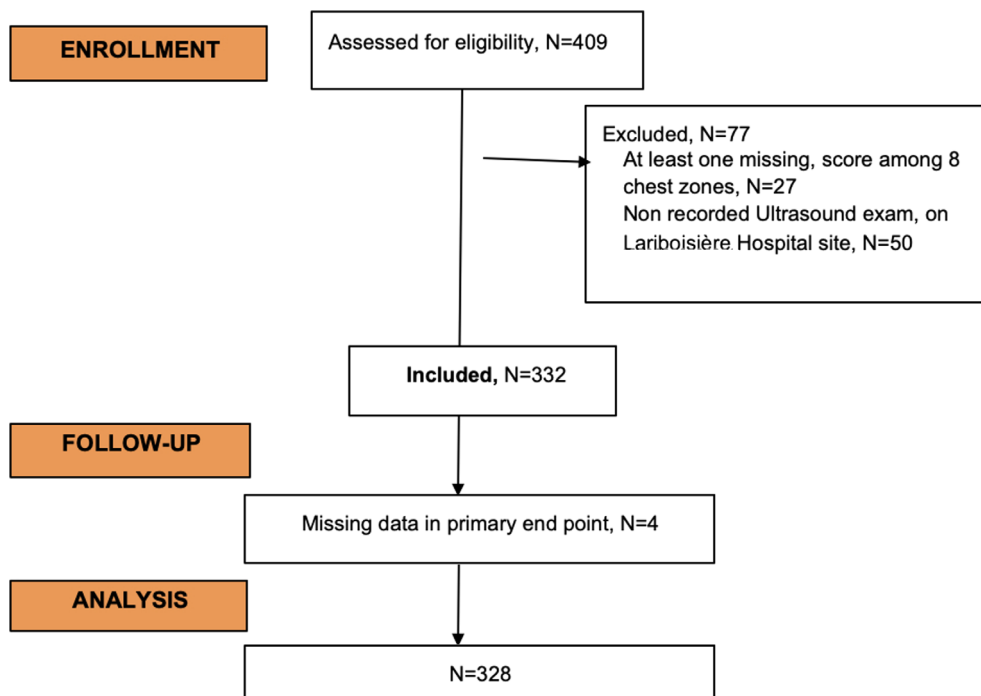


Table 2. Characteristics of Included Patients (n = 328)

	n/(miss.) (%)
<i>Demographic data</i>	
Mean age in years ± SD	54.2 ± 18.3
Male	159 (48.5%)
Active smokers	40/290 (13.8%)
<i>Medical background</i>	
Diabetes type 1	5/322 (1.6%)
Diabetes type 2	45/323 (13.9%)
Arterial hypertension	75/322 (23.3%)
Asthma	32/324 (9.9%)
Immunosuppressed status	12/322 (3.7%)
Chronic obstructive pulmonary disease	17/324 (5.2%)
Ischemic cardiopathy	15/324 (4.6%)
Active neoplasia	21/321 (6.5%)
Recent use of nonsteroidal anti-inflammatory drugs	3/314 (1.0%)
<i>Vital parameters at admission</i>	
Temperature (in °C)	37.3 ± 0.9
Mean heart rate ± SD	89.9 ± 17.5
Mean systolic blood pressure ± SD	136.3 ± 21.2
Mean diastolic blood pressure ± SD	80.9 ± 14.1
Mean respiratory rate ± SD	22.2 ± 7.2
<i>CT scan severity</i>	
Normal	14 (9.9%)
Minimal	35 (24.8%)
Moderate	46 (32.6%)
Severe	46 (32.6%)

28, 2020 were eligible. We included 328 patients in the analysis (Figure 1).

Descriptive Patient Data (Table 2)

Overall, 159/328 (48.5%) patients were men and the mean age was 54.2 ± 18.3 years (Table 2). One quarter of patients had arterial hypertension (n = 75/322, 23.3%) and 45/323 (13.9%) had type 2 diabetes mellitus. Twelve (3.7%) patients reported an immunosuppressed status. Four of 332 (1.20%) patients were lacking primary outcome data. All patients had a PCR testing within their hospital stay. Chest CT scan was performed for 141 (43%) patients; scans were pathological for 127 (90.0%).

Descriptive US Data

The two most frequently reported affected quadrants were the left and right postero-inferior quadrants, in 133 (40.5%) and 136 (41.5%) patients, respectively. These quadrants were frequently severely affected (grade = 3): 40 (12.2%) and 37 (11.3%) patients, respectively (Table 3). For the 328 patients, the mean

Table 3. Details of the Lung Ultrasound Score for Each Quadrant (n = 328)

Left Lung	N (%)	Right Lung	N (%)
<i>Antero-superior</i>		<i>Antero-superior</i>	
Grade 0	238 (72.6%)	Grade 0	221 (67.4%)
Grade 1	31 (9.5%)	Grade 1	39 (11.9%)
Grade 2	52 (15.9%)	Grade 2	50 (15.2%)
Grade 3	7 (2.1%)	Grade 3	18 (5.5%)
<i>Antero-inferior</i>		<i>Antero-inferior</i>	
Grade 0	217 (66.2%)	Grade 0	207 (63.1%)
Grade 1	37 (11.3%)	Grade 1	49 (14.9%)
Grade 2	45 (13.7%)	Grade 2	42 (12.8%)
Grade 3	29 (8.8%)	Grade 3	30 (9.1%)
<i>Postero-superior</i>		<i>Postero-superior</i>	
Grade 0	246 (75.0%)	Grade 0	236 (72.0%)
Grade 1	32 (9.8%)	Grade 1	38 (11.6%)
Grade 2	40 (12.2%)	Grade 2	43 (13.1%)
Grade 3	10 (3.0%)	Grade 3	11 (3.4%)
<i>Postero-inferior</i>		<i>Postero-inferior</i>	
Grade 0	195 (59.5%)	Grade 0	192 (58.5%)
Grade 1	40 (12.2%)	Grade 1	49 (14.9%)
Grade 2	53 (16.2%)	Grade 2	50 (15.2%)
Grade 3	40 (12.2%)	Grade 3	37 (11.3%)

GS score for all pulmonary quadrants (maximum achievable score: 24) was 4.9 ± 6.4.

Primary Analysis

Clinical status at day 28 was collected for 328 of 332 patients: 13 (4%) died, 16 (4.9%) underwent invasive mechanical ventilation, and 28 (8.5%) were referred to an ICU. In addition, 90 (27.4%) were hospitalized in a standard-care unit during the 28-day follow-up, and 181 (55.2%) returned home. Details of the GS for each status are in Table 4.

The association between clinical worsening and GS was optimal, with GS ≥6, corresponding to an AUC of 0.83 (Figure 2), sensitivity 84.2%, and specificity 76.4% (Youden Index: 60.6%). The optimism-corrected C-statistic was -0.0003 and the Brier Score was 0.12 with optimism-corrected C-statistic -0.0007. For patients with COVID-19, the association between clinical worsening and GS was optimal, with GS ≥6 (Table 5). Moreover, if we consider the clinical outcomes individually, the GS was optimal, with GS ≥6 for only the ICU admission outcome, ≥6 for the invasive mechanical ventilation outcome and ≥10 for only the death outcome (Table 6).

Table 4. Details of Lung Ultrasound Global Score for Each Status at 28 Days Follow-Up (n = 328)

Poor health status within 28 days							
	Total	Death	Invasive Mechanical Ventilation	Intensive Care Unit	Hospitalization	Back Home	
N	328	13	16	28	90	181	
Global Echography Score							
Mean ± SD		4.9 ± 6.4	10.7 ± 5.6	11.4 ± 5.4	10.6 ± 5.1	8.6 ± 7.1	1.2 ± 3.1
Median (Q1;Q3)		1.0 (0.0;9.0)	12.0 (10.0;14.0)	12.5 (7.0;16.0)	10.5 (7.0;14.5)	7.0 (3.0;15.0)	0.0 (0.0;1.0)
Min, Max		0, 23	0, 19	2, 18	0, 23	0, 22	0, 20

Figure 2. Area under the receiver operating characteristic (ROC) curve for the association between clinical worsening within the 28-day follow-up and the lung ultrasonography global score (GS).

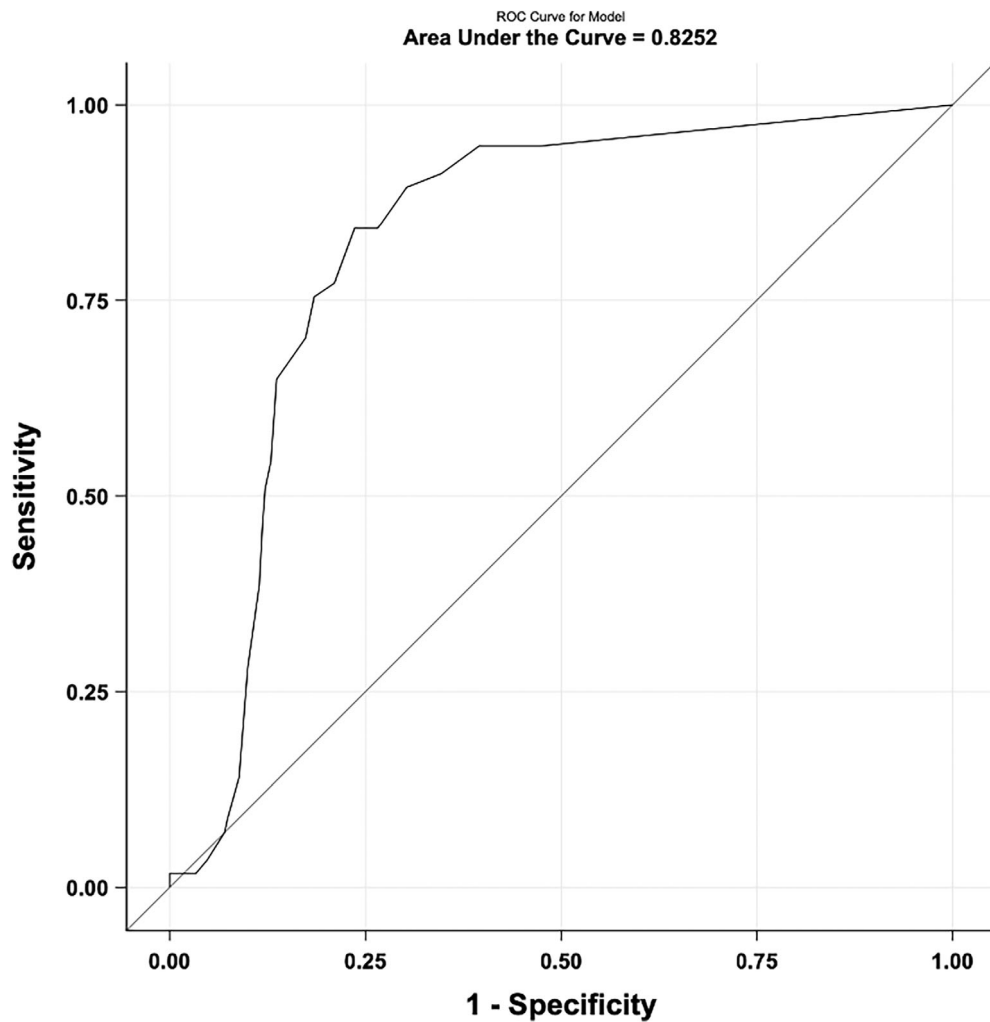


Table 5. Association Between Clinical Worsening Within the 28 Days Follow-Up Period and Lung Ultrasound Global Score for Patients Diagnosed With COVID-19

AUC (95% CI)	Score	Sensitivity	Specificity	False-Negative Rate	False-Positive Rate	Accuracy	Positive Predictive Value	Negative Predictive Value	False Omission Rate	False Discovery Rate	Youden Index
0.77 (0.69–0.84)	1	95.2%	40.9%	4.8%	59.1%	55.4%	37.0%	95.9%	4.1%	63.0%	36.1%
	2	95.2%	51.3%	4.8%	48.7%	63.1%	41.7%	96.7%	3.3%	58.3%	46.5%
	3	90.5%	54.8%	9.5%	45.2%	64.3%	42.2%	94.0%	6.0%	57.8%	45.3%
	4	88.1%	58.3%	11.9%	41.7%	66.2%	43.5%	93.1%	6.9%	56.5%	46.4%
	5	85.7%	63.5%	14.3%	36.5%	69.4%	46.2%	92.4%	7.6%	53.8%	49.2%
	6 ^a	85.7%	66.1%	14.3%	33.9%	71.3%	48.0%	92.7%	7.3%	52.0%	51.8%
	7	78.6%	70.4%	21.4%	29.6%	72.6%	49.3%	90.0%	10.0%	50.7%	49.0%
	8	78.6%	73.0%	21.4%	27.0%	74.5%	51.6%	90.3%	9.7%	48.4%	51.6%
	9	71.4%	73.9%	28.6%	26.1%	73.2%	50.0%	87.6%	12.4%	50.0%	45.3%
	10	64.3%	79.1%	35.7%	20.9%	75.2%	52.9%	85.8%	14.2%	47.1%	43.4%

^aOptimal cutoff point as the point maximizing the Youden index.

Secondary Analysis

The association between CT evaluation of disease (normal or minimal vs moderate or severe) and GS was optimal with GS ≥ 7 , corresponding to an AUC of 0.84 (95% confidence interval [CI]: 0.77–0.92), with sensitivity 77.2% and specificity 83.7% (maximal Youden index) (Table 7). The AUCs to predict clinical status (death or mechanical ventilation or ICU admission) for patients with full CT data were 0.71 (95% CI: 0.62–0.79) for the GS and 0.74 (95% CI: 0.65–0.82) for the CT severity score (Figure 3).

In total, 48 (14.6%) LUS evaluations were performed by newly trained operators. Bland and Altman plots revealed substantial agreement between expert and newly trained operators considering each quadrant individually (n = 48, 4 new trainees), with weighted kappa values of 0.62 to 0.81 (Table 8). Considering the above GS with maximal Youden index = 7 as a reasonable score to discriminate lung injury severity, we found good agreement between experts and newly trained operators to distinguish a GS ≥ 7 , with kappa values of 0.85 (95% CI: 0.69–1.00) and 0.85 (accelerated bias-corrected 95% CI: 0.60–0.96) (Table 9).

Discussion

In this multicentric, prospective, observational study, our GS was associated with clinical worsening for patients with suspected or diagnosed COVID-19 at 28 days after presentation to an emergency department. The AUC was 0.83 for comparing lung damage severity with the GS, summing the severity over 8 chest points, to the CT scan severity score. Another key finding, although very preliminary, is the concordance in GS scoring between an expert and newly trained US operator, with kappa values >0.85 .

With regard to the existing literature on LUS as a risk stratification tool, our study was performed with reasonable statistical power (N = 328), executed with both a standard and handheld US device, and led to a simplified score for lung scoring. Although this point is essentially not discussed in the studies we referenced, we took into account “true hospitalization” and ICU-like hospitalization (defined in our investigation centers as patients requiring over 5 L/min oxygeno-therapy) because in the event of

Table 6. Performance of the Lung Ultrasonography GS to Predict Each of the Severity Outcomes: Intensive Care Unit Hospitalization, Invasive Mechanical Intubation, and Death

Intensive Care Unit Hospitalization												
AUC (95% CI)	Score	Sensitivity	Specificity	False-Negative Rate	False-Positive Rate	Accuracy	Positive Predictive Value	Negative Predictive Value	False Omission Rate	False Discovery Rate	Youden Index	
0.75 (0.66–0.85)	1	89.5%	47.1%	10.5%	52.9%	49.5%	9.5%	98.6%	1.4%	90.5%	36.5%	
	2	89.5%	53.6%	10.5%	46.4%	55.7%	10.7%	98.8%	1.2%	89.3%	43.1%	
	3	89.5%	58.2%	10.5%	41.8%	60.0%	11.7%	98.9%	1.1%	88.3%	47.6%	
	4	89.5%	62.4%	10.5%	37.6%	64.0%	12.9%	99.0%	1.0%	87.1%	51.9%	
	5	84.2%	66.0%	15.8%	34.0%	67.1%	13.3%	98.5%	1.5%	86.7%	50.2%	
	6	84.2%	68.6%	15.8%	31.4%	69.5%	14.3%	98.6%	1.4%	85.7%	52.8%	
	7	73.7%	71.6%	26.3%	28.4%	71.7%	13.9%	97.8%	2.2%	86.1%	45.3%	
	8	68.4%	73.9%	31.6%	26.1%	73.5%	14.0%	97.4%	2.6%	86.0%	42.3%	
	9	63.2%	75.5%	36.8%	24.5%	74.8%	13.8%	97.1%	2.9%	86.2%	38.6%	
	10	57.9%	79.4%	42.1%	20.6%	78.2%	14.9%	96.8%	3.2%	85.1%	37.3%	
Invasive Mechanical Ventilation												
AUC (95% CI)	Score	Sensitivity	Specificity	False-Negative Rate	False-Positive Rate	Accuracy	Positive Predictive Value	Negative Predictive Value	False Omission Rate	False Discovery Rate	Youden Index	
0.79 (0.73–0.86)	1	100.0%	47.4%	0.0%	52.6%	50.2%	9.5%	100.0%	0.0%	90.5%	47.4%	
	2	100.0%	53.9%	0.0%	46.1%	56.3%	10.7%	100.0%	0.0%	89.3%	53.9%	
	3	88.2%	57.8%	11.8%	42.2%	59.4%	10.3%	98.9%	1.1%	89.7%	46.0%	
	4	88.2%	62.0%	11.8%	38.0%	63.4%	11.4%	99.0%	1.0%	88.6%	50.2%	
	5	88.2%	65.9%	11.8%	34.1%	67.1%	12.5%	99.0%	1.0%	87.5%	54.1%	
	6	88.2%	68.5%	11.8%	31.5%	69.5%	13.4%	99.1%	0.9%	86.6%	56.7%	
	7	76.5%	71.4%	23.5%	28.6%	71.7%	12.9%	98.2%	1.8%	87.1%	47.9%	
	8	70.6%	73.7%	29.4%	26.3%	73.5%	12.9%	97.8%	2.2%	87.1%	44.3%	
	9	64.7%	75.3%	35.3%	24.7%	74.8%	12.6%	97.5%	2.5%	87.4%	40.0%	
	10	58.8%	79.2%	41.2%	20.8%	78.2%	13.5%	97.2%	2.8%	86.5%	38.0%	
Death												
AUC (95% CI)	Score	Sensitivity	Specificity	False-Negative Rate	False-Positive Rate	Accuracy	Positive Predictive Value	Negative Predictive Value	False Omission Rate	False Discovery Rate	Youden Index	
0.75 (0.61–0.88)	1	84.6%	46.2%	15.4%	53.8%	47.7%	6.1%	98.6%	1.4%	93.9%	30.8%	
	2	84.6%	52.6%	15.4%	47.4%	53.8%	6.9%	98.8%	1.2%	93.1%	37.2%	
	3	84.6%	57.1%	15.4%	42.9%	58.2%	7.6%	98.9%	1.1%	92.4%	41.7%	
	4	84.6%	61.2%	15.4%	38.8%	62.2%	8.3%	99.0%	1.0%	91.7%	45.8%	
	5	84.6%	65.1%	15.4%	34.9%	65.8%	9.2%	99.0%	1.0%	90.8%	49.7%	
	6	84.6%	67.6%	15.4%	32.4%	68.3%	9.8%	99.1%	0.9%	90.2%	52.2%	
	7	84.6%	71.2%	15.4%	28.8%	71.7%	10.9%	99.1%	0.9%	89.1%	55.8%	

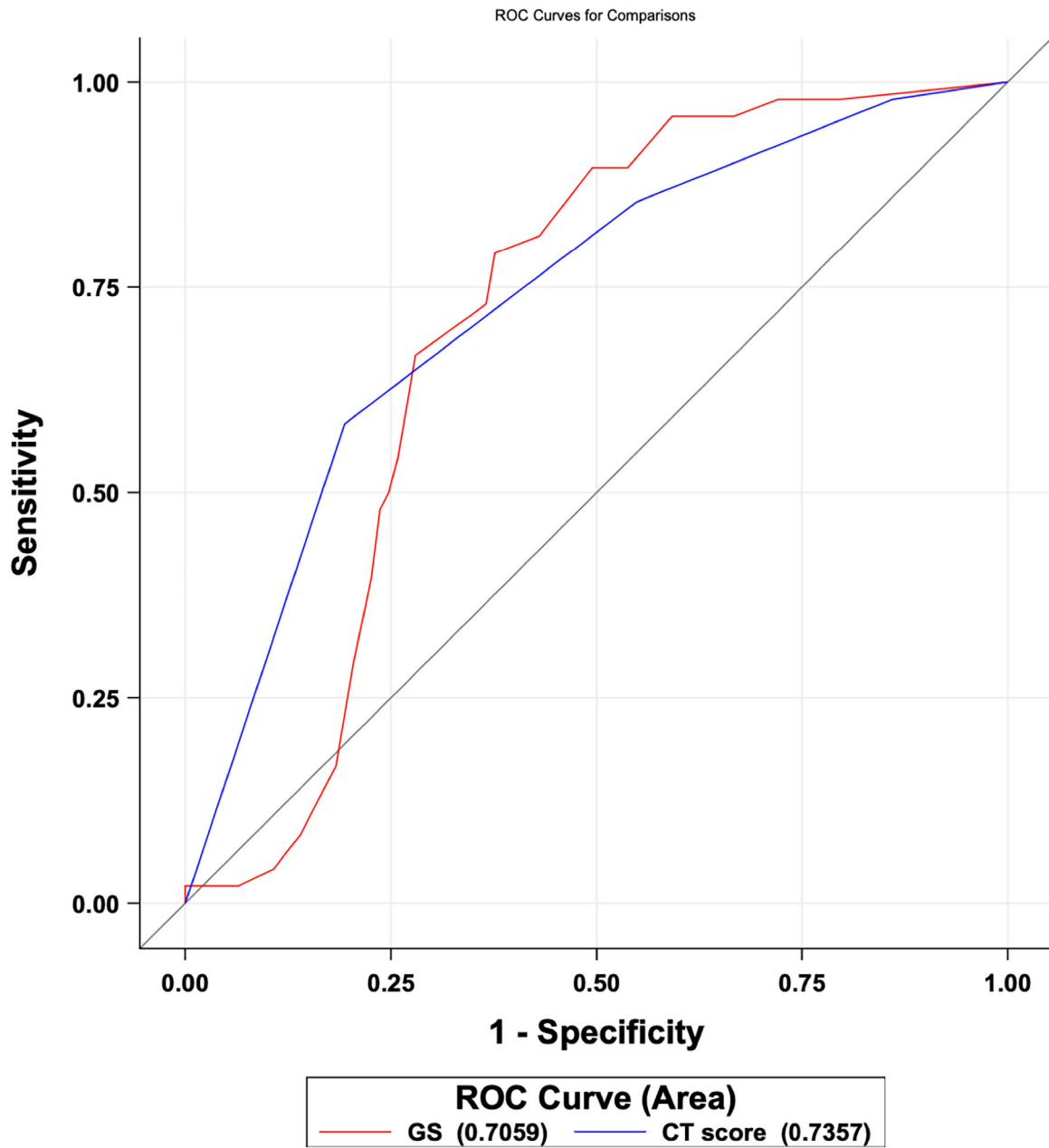
8	76.9%	73.4%	23.1%	26.6%	73.5%	10.8%	98.7%	1.3%	89.2%	50.3%
9	76.9%	75.3%	23.1%	24.7%	75.4%	11.5%	98.7%	1.3%	88.5%	52.2%
10	76.9%	79.5%	23.1%	20.5%	79.4%	13.5%	98.8%	1.2%	86.5%	56.4%
11	61.5%	81.4%	38.5%	18.6%	80.6%	12.1%	98.1%	1.9%	87.9%	42.9%
12	61.5%	82.7%	38.5%	17.3%	81.8%	12.9%	98.1%	1.9%	87.1%	44.2%

Table 7. Performances Lung Global Score GS to Predict CT Severity Score

AUC (95% CI)	Score	Sensitivity	Specificity	False-Negative Rate	False-Positive Rate	Accuracy	Positive Predictive Value	Negative Predictive Value	False Omission Rate	False Discovery Rate	Youden Index
0.84 (0.77–0.92)	1	96.7%	34.7%	3.3%	65.3%	75.2%	73.6%	85.0%	15.0%	26.4%	31.4%
	2	95.7%	46.9%	4.3%	53.1%	78.7%	77.2%	85.2%	14.8%	22.8%	42.6%
	3	93.5%	55.1%	6.5%	44.9%	80.1%	79.6%	81.8%	18.2%	20.4%	48.6%
	4	91.3%	65.3%	8.7%	34.7%	82.3%	83.2%	80.0%	20.0%	16.8%	56.6%
	5	84.8%	69.4%	15.2%	30.6%	79.4%	83.9%	70.8%	29.2%	16.1%	54.2%
	6	82.6%	73.5%	17.4%	26.5%	79.4%	85.4%	69.2%	30.8%	14.6%	56.1%
	7 ^a	77.2%	83.7%	22.8%	16.3%	79.4%	89.9%	66.1%	33.9%	10.1%	60.8%
	8	70.7%	83.7%	29.3%	16.3%	75.2%	89.0%	60.3%	39.7%	11.0%	54.3%
	9	66.3%	83.7%	33.7%	16.3%	72.3%	88.4%	56.9%	43.1%	11.6%	50.0%
	10	56.5%	87.8%	43.5%	12.2%	67.4%	89.7%	51.8%	48.2%	10.3%	44.3%

^aOptimal cutoff point as the point maximizing the Youden Index.

Figure 3. Statistics for AUCs: GS vs CT severity score.



ROC model	Area	Error type	de Wald confidence interval at 95%		D Somers	Gamma	Tau-a
GS	0.7059	0.0437	0.6202	0.7915	0.4117	0.4266	0.1862
CT score	0.7357	0.0418	0.6537	0.8177	0.4713	0.6292	0.2132

Table 8. Bland and Altman Plot for Agreement Between Experts and New Trainees on Each Chest Zone (n = 48, 4 New Trainees)

Cadrans	N	Weighted Kappa (Normal Approximation 95% CI)	Weighted Kappa (Percentile Meth. 95% CI)	Weighted Kappa (Bias-Corrected Percentile Meth. 95% CI)	Weighted Kappa (Accelerated Bias-Corrected [BCa] 95% CI)	P-Value
Antero-Superior R	48	0.81 (0.57;1.00)	0.81 (0.68;0.98)	0.81 (0.47;0.89)	0.81 (0.47;0.89)	<.0001
Antero-Inferior R	45	0.93 (0.87;1.00)	0.93 (0.84;0.98)	0.93 (0.84;0.98)	0.93 (0.83;0.98)	<.0001
Antero-Inferior R	47	0.78 (0.60;0.96)	0.78 (0.51;0.91)	0.78 (0.46;0.90)	0.78 (0.42;0.89)	<.0001
Postero-Inferior R	46	0.74 (0.59;0.90)	0.74 (0.56;0.88)	0.74 (0.56;0.87)	0.74 (0.55;0.87)	<.0001
Antero-Superior L	48	0.79 (0.58;0.99)	0.79 (0.52;0.94)	0.79 (0.52;0.94)	0.79 (0.48;0.93)	<.0001
Antero-Inferior L	45	0.71 (0.54;0.87)	0.71 (0.50;0.88)	0.71 (0.47;0.85)	0.71 (0.43;0.84)	<.0001
Postero-Superior L	47	0.76 (0.54;0.98)	0.76 (0.52;0.96)	0.76 (0.44;0.93)	0.76 (0.33;0.92)	<.0001
Postero-Inferior L	45	0.62 (0.33;0.91)	0.62 (0.31;0.90)	0.62 (0.32;0.90)	0.62 (0.28;0.87)	<.0001

lack of an ICU, patients may have been treated as ICU patients in a regular hospitalization department. eChoVid may serve as a supplementary input for meta-analysis (one meta-analysis²³ did not take into account many relevant studies) mitigating the low size, mono-centricity of many studies or eventually negative findings on the topic¹⁷ and helping the medical community define the place of LUS as a risk stratification tool.

Our study shows the interest of LUS in the first clinical assessment of patients with suspected COVID-19 and the triage of patients at high vs low risk of clinical worsening. Also, we merged the clinical outcomes of severity into one, which was relevant to measure the overall risk of clinical deterioration in a patient given a first US examination. In addition, at the time the study protocol was drafted (beginning of March 2020), hospitals in France were overwhelmed with COVID-19 patients, so we added days 5 and 15 as intermediate points to acquire clinical outcomes. Indeed, we thought that intermediary results regarding outcomes at days 5 and 15 may lead to a significant correlation between the US severity score and severity clinical outcomes and that these could be shared within the French medical authorities and radiology learned societies. However, this expected outcome was not observed in the intermediary analysis.

The good sensitivity of LUS is not surprising.^{12,24} Nevertheless, LUS findings must be interpreted with caution and should not lead to ruling out other causes of dyspnea (pulmonary embolism).²⁵ Without a blinding condition, the specificity of LUS may be improved with interpretation considering the clinical context. Also, LUS may be of interest for severity assessment of lung lesions; however and especially given the rapidity of PCR testing, there is no substantial argument to value LUS as an initial diagnostic tool.

LUS must not compete with chest CT, especially when the patient requires closer lung status assessment. Indeed, the GS may be difficult to determine for patients with causes that prevent GS interpretation (morbid obesity, sub-cutaneous emphysema, etc) or pre-existing conditions (emphysema or fibrosis), thus affecting the relevance of the operator’s interpretation. However, the GS score and CT severity score performed comparably for predicting the primary outcome (death or mechanical ventilation or ICU admission): AUCs were 0.72 (95% CI: 0.62–0.79) and 0.73 (95% CI: 0.65–0.82) for the GS and CT, respectively, and the difference between the two scores was non-significant (Mann–Whitney, *P* = .52).

We standardized the evaluation criteria for LUS and simplified according to Buessler et al.^{18,26} The GS for LUS severity may be improved. The score

Table 9. Agreement Between Experts and New Trainees for Discriminating GS ≥7 (N = 48 Pairs of Raters)

GS	N	Kappa (Normal Approximation 95% CI)	Kappa (Percentile Meth. 95% CI)	Kappa (Bias-Corrected Percentile Meth. 95% CI)	Kappa (Accelerated Bias-Corrected (BCa) 95% CI) (Better)	P-Value
GS ≥7	48	0.85 (0.69;1.00)	0.85 (0.67;1.00)	0.85 (0.65;1.00)	0.85 (0.60;0.96)	<.0001

carries information about the severity of lung injuries, but low scores do not sufficiently address their topographic distribution. Indeed, a high lung injury score in a sole quadrant and low scores affecting several quadrants may lead to the same GS. The median, average or maximum severity is too naive to provide specific or sufficient information on the extension of tissue injury throughout the lung.

To improve the US scoring, an ongoing study is comparing CT and LUS for each chest point. The study includes more refined statistics based on machine-learning techniques, exploring linear and non-linear effects (such as the non-linear jump in condition between GS at 0 and 1) or injury distribution patterns throughout chest zones. Although this approach suggests gains in specificity, a more complex score may not be easy to compute in practice.

In our study, although the agreement between experts and newly trained operators seems promising, training protocols may be improved and tested with a larger pool of newly trained operators. Some evidence suggests that short protocols may be sufficient, especially when the training focuses on specific medical issues.^{27–29} Our agreement findings are less significant when comparing the performance for chest zone by chest zone, especially because the assessment of B-lines may be variable for new trainees. This observation may be related to the fact that, together with the investigators of the study who performed the training sessions (both expert ultra-sonographers and experienced trainers), we set the training protocol as a quick practice of 5 scans. This training was grounded on experience and the fact that training focused on some specific signs and did not cover the wide range of US lung semiology. Though US training protocols are not well standardized,³⁰ our protocol may have benefited from following some practices documented in the literature, such as training with a higher number of scans (eg, 25 scans in Reference 31).

Also, we wanted to compare performance with regard to predicting clinical outcomes, and our results show that the global assessment was comparable between experts and new trainees. All images acquired and scored by new trainees were reviewed by two experts (both investigators of the study), and disagreements were resolved by discussion.

Our study has several limitations. First, we did not evaluate the treatment received by patients, which

could highly modify the prognosis and bias our primary outcome. Second, we used a clinical approach and included all patients with suspected COVID-19 and with radiology or RT-PCR confirmation during a period with a remarkably high incidence of COVID-19 in France. This situation could suggest recruitment bias. Our results must be confirmed in other studies including patients with all etiologies of febrile dyspnea.

Conclusion

The LUS severity GS could be used for assessing the severity of lung injuries in patients with suspected or diagnosed COVID-19 and for predicting clinical worsening at 28 days. LUS findings seemed to be consistent with chest CT findings. The assessment of training and learning curves need an enhanced protocol for further studies. The point-of-care nature of the examination, the accessibility of the device, the real-time interpretation, and the non-invasive technology may suggest LUS as a relevant screening tool for assessing lung injury severity.

References

1. Lythgoe MP, Middleton P. Ongoing clinical trials for the management of the COVID-19 pandemic. *Trends Pharmacol Sci* 2020; 41: 363–382.
2. Rubin GD, Ryerson CJ, Haramati LB, et al. The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner society. *Radiology* 2020; 296:172–180.
3. Actualisation Des Recommandations d'imagerie Thoracique Dans la Pneumonie COVID-19 [Internet]. SFR e-Bulletin; 2020. <https://ebulletin.radiologie.fr/actualites-covid-19/actualisation-recommandations-dimagerie-thoracique-pneumonie-covid-19>. Accessed April 7, 2021.
4. ACR Recommendations for the Use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection [Internet]. <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection>. Accessed April 8, 2021.

5. Baribeau Y, Sharkey A, Chaudhary O, et al. Handheld point-of-care ultrasound probes: the new generation of POCUS. *J Cardiothorac Vasc Anesth* 2020; 34:3139–3145.
6. Mojoli F, Bouhemad B, Mongodi S, et al. Lung ultrasound for critically ill patients. *Am J Respir Crit Care Med* 2019; 199:701–714.
7. Pivetta E, Goffi A, Tizzani M, et al. Lung ultrasonography for the diagnosis of SARS-CoV-2 pneumonia in the emergency department. *Ann Emerg Med* 2021; 77:385–394.
8. Akl EA, Blažić I, Yaacoub S, et al. Use of chest imaging in the diagnosis and management of COVID-19: a WHO rapid advice guide. *Radiology* 2021; 298:E63–E69.
9. Bonadia N, Carnicelli A, Piano A, et al. Lung ultrasound findings are associated with mortality and need for intensive care admission in COVID-19 patients evaluated in the emergency department. *Ultrasound Med Biol* 2020; 46:2927–2937.
10. Castelo J, Graziani D, Soriano JB, Izquierdo JL. Findings and prognostic value of lung ultrasound in COVID-19 pneumonia. *J Ultrasound Med* 2021; 40:1315–1324.
11. Yasukawa K, Minami T, Boulware DR, Shimada A, Fischer EA. Point-of-care lung ultrasound for COVID-19: findings and prognostic implications from 105 consecutive patients. *J Intensive Care Med* 2021; 36:334–342.
12. Ji L, Cao C, Gao Y, et al. Prognostic value of bedside lung ultrasound score in patients with COVID-19. *Crit Care* 2020; 24:1–12.
13. Brahier T, Meuwly JY, Pantet O, et al. Lung ultrasonography for risk stratification in patients with coronavirus disease 2019 (COVID-19): a prospective observational cohort study. *Clin Infect Dis* 2021; 73:e4189–e4196.
14. Le Bastard Q, Javaudin F, Perrier C, et al. Combination of lung ultrasound to standard evaluation in patients with suspected COVID-19 presenting to the emergency department: development of a composite score (EchoVID score) for prediction of day-2 respiratory failure. *Eur J Emerg Med* 2021; 28:149–150.
15. Lichter Y, Topilsky Y, Taieb P, et al. Lung ultrasound predicts clinical course and outcomes in COVID-19 patients. *Intensive Care Med* 2020; 46:1873–1883.
16. Mafort TT, Rufino R, da Costa CH, et al. One-month outcomes of patients with SARS-CoV-2 infection and their relationships with lung ultrasound signs. *Ultrasound J* 2021; 13:1–8.
17. Gil-Rodrigo A, Llorens P, Luque-Hernández MJ, Martínez-Buendía C, Ramos-Rincón JM. Lung ultrasound integration in assessment of patients with noncritical COVID-19. *J Ultrasound Med* 2021; 40:2203–2212.
18. Buessler A, Chouihed T, Duarte K, et al. Accuracy of several lung ultrasound methods for the diagnosis of acute heart failure in the ED: a multicenter prospective study. *Chest* 2020; 157:99–110.
19. Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med* 2020; 382:970–971.
20. Lichtenstein D. L'échographie pulmonaire en réanimation. 2022. <http://164.2.250.208/pdf/java07/EchoDL.pdf>.
21. La société d'Imagerie Thoracique propose un compte-rendu structuré de scanner thoracique pour les patients suspects de COVID-19 [Internet]. SFR e-Bulletin. <https://ebulletin.radiologie.fr/actualites-covid-19/societe-dimagerie-thoracique-propose-compte-rendu-structure-scanner-thoracique>. Accessed April 7, 2021.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174.
23. Song G, Qiao W, Wang X, Yu X. Association of lung ultrasound score with mortality and severity of COVID-19: a meta-analysis and trial sequential analysis. *Int J Infect Dis* 2021; 108:603–609.
24. Lichtenstein DA, Lascols N, Mezière G, et al. Ultrasound diagnosis of alveolar consolidation in the critically ill. *Intensive Care Med* 2004; 30:276–281.
25. Léonard-Lorant I, Delabranche X, Séverac F, et al. Acute pulmonary embolism in patients with COVID-19 at CT angiography and relationship to d-dimer levels. *Radiology* 2020; 296:E189–E191.
26. Lichtenstein DA, Mezière GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. *Chest* 2008; 134:117–125.
27. 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.
28. See KC, Ong V, Wong SH, et al. Lung ultrasound training: curriculum implementation and learning trajectory among respiratory therapists. *Intensive Care Med* 2016; 42:63–71.
29. 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.
30. Shah S, Santos N, Kisa R, et al. Efficacy of an ultrasound training program for nurse midwives to assess high-risk conditions at labor triage in rural Uganda. *PLoS One* 2020; 15:e0235269.
31. Pietersen PI, Madsen KR, Graumann O, Konge L, Nielsen BU, Laursen CB. Lung ultrasound training: a systematic review of published literature in clinical lung ultrasound training. *Crit Ultrasound J* 2018; 10:23.