Research Article

Lung Ultrasound using a Handheld Device to Diagnose COVID-19 in the Emergency Department

Abdulrahman M. Alfurahia, Abdussalam A. Alshehri, Hani M. Alshehri, Sami A. Alamri, Tariq S. Aleyyed, Khalid G. Alaufi, Mohammed J. Alsaadi

ARTICLE INFO

Article History
Received 09 August 2021
Accepted 12 August 2021

Keywords
COVID-19
SARS-CoV-2
point-of-care ultrasound
diagnostic test
sensitivity
specificity

ABSTRACT

The Coronavirus Disease 2019 (COVID-19) pandemic has critically struck healthcare systems and burdened emergency services. To date, there is no accurate and rapid point-of-care diagnostic test. This study aimed to investigate Lung Ultrasound (LUS) against Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) test in suspected COVID-19 patients presenting to the emergency department. In 20 eligible patients (mean age ± standard deviation, 49 ± 15 years), 12 had a positive RT-PCR test and underwent an LUS examination over 12 lung zones using a handheld ultrasound device. Each zone was semiquantitatively scored according to the Lung Ultrasound Scoring System (LUSS) from 0 to 3 based on the severity of findings (pleural line irregularity, B-lines, consolidations) and documented the presence of light beam artifacts. A second blinded reader scored the images to investigate interreader reproducibility. The LUSS score had a modest diagnostic performance at 66.6% [95% Confidence Interval (CI), 34.9–90.0%] sensitivity and 75.0% (95% CI, 34.9–96.8%) specificity. The light beam artifact was more prevalent and sensitive to COVID-19 patients with 81.8% (95% CI, 48.2–97.7%) sensitivity and 75.0% (95% CI, 34.9–96.8%) specificity. LUS had an almost perfect interreader reproducibility for LUSS (Kendall’s W = 0.961; 95% CI, 0.894–0.985) and light beam artifact (Cohen’s κ = 0.890; 95% CI, 0.683–1.00). Overall, LUS using handheld devices can offer a safe, reproducible, rapid, and feasible first-line tool for detecting COVID-19 patients in emergency departments. The light beam artifact was more sensitive and specific to COVID-19 patients and can be useful for effectively triaging suspected cases.

© 2021 Dr. Sulaiman Al Habib Medical Group. Publishing services by Atlantis Press International B.V.
This is an open access article distributed under the CC BY-NC 4.0 license (http://creativecommons.org/licenses/by-nc/4.0/).

1. INTRODUCTION

Numerous healthcare systems are facing the daily challenges of managing the Coronavirus Disease 2019 (COVID-19) pandemic. The virus spread worldwide, and close to 2.5 million lives were lost as of early 2021. Recent reports suggest that the actual COVID-19 attributable mortality is reflected by more than 20.5 million years of life lost globally [1]. Although most cases with COVID-19 develop a mild illness, approximately 14% develop severe disease requiring hospital admission, and 5% require intensive care unit support [2]. Despite the recent breakthroughs and distribution of vaccines, a recent survey showed that 89% of scientists believe the virus is “here to stay” for years because it has spread worldwide where global eradication is unlikely [3].

Emergency physicians are confronted with large cohorts of suspected cases amidst this pandemic. They are searching for feasible and accurate diagnostic instruments to identify suspected cases. The Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) test through a nasopharyngeal or oropharyngeal swab has been widely recognized as the reference test for diagnosing suspected cases. However, the test requires specialized laboratories; a luxury numerous countries are lacking [4]. It is also subject to delays, where results might take up to 8 h to be available. Additional critical disadvantages include the problematic false negative rates as well as logistic challenges [5]. Alternative radiologic tests such as chest X-ray and Computed Tomography (CT) have been used in adjunct to RT-PCR. These tests carry further disadvantages related to the capacity and use of ionizing radiation besides questionable diagnostic accuracy. The American College of Radiology (ACR) discourages such tests as the first line of investigation or screening for suspected COVID-19 [6]. They recommend a sparing use of CT to hospitalized or critical cases where they are clinically indicated.

In contrast, point-of-care Lung Ultrasound (LUS) is widely adopted in emergency departments. It can be safely performed at the bedside by the attending physician without patient transportation and risks of exposure to additional personnel. This real-time, noninvasive, and radiation-free tool is routinely used for detecting and monitoring pulmonary abnormalities. LUS has a high sensitivity (97%) and specificity (95%) for detecting pulmonary edema and more than 93% accuracy for pleural effusions [7].

Coronavirus disease 2019 predilection to the peripheral distribution of pathology and pleural involvement proposes a promising
role for LUS. Indeed, there is a growing research activity for LUS in COVID-19. Multiple studies have attempted to evaluate the role of LUS in diagnosis and management in emergency departments where it is hypothesized to be useful for triaging suspected COVID-19 patients by directing suspected cases to the isolation area or discharging them. Early signs of LUS in COVID-19 showed a presence of B-lines, consolidations, pleural thickening, and lack of pleural effusions in symptomatic cases [8,9]. Although the early findings of pulmonary signs are promising, most of the studies were of retrospective design and lacked a control group by investigating confirmed COVID-19 patients. There is also considerable disagreement in the interpretation of LUS findings. Manivel et al. [10] have recently proposed the Lung Ultrasound Scoring System (LUSS) as a semiquantitative criterion for diagnosing COVID-19 in the emergency department. However, its diagnostic performance and reproducibility have not been analyzed. Volpicelli et al. [11] recently reported a typical artifact coined as the “light beam” artifact that manifests as a broad band-shaped vertical artifact arising from a normal pleural line. Case studies observed this pattern invariably in most COVID-19-related pneumonia [12]. Others speculated that this artifact is specific to early COVID-19 because of the patchy pattern of the ground-glass opacity lesions present on CT [13]. Systematic evaluation of this artifact in more than case studies is lacking. Moreover, handheld devices’ performance in diagnosing COVID-19 cases is unknown despite their recommended use in emergency departments during the pandemic [14].

This study aims to investigate the diagnostic performance of LUS patterns in detecting COVID-19 using a handheld ultrasound device compared to RT-PCR in symptomatic patients presenting to the emergency department. Another study objective is to investigate the diagnostic accuracy of the recently reported light beam artifact.

2. MATERIALS AND METHODS

2.1. Study Design and Patient Population

The study was designed as a prospective case–control study at Prince Mohammed bin Abdulaziz Hospital in Riyadh, Saudi Arabia, from August 2020 to January 2021. Ethical approval was obtained (KACS registration number: H-01-R-012; NIH registration number: IRB00010471) from the institutional review board at King Fahad Medical City, the external governing ethical site. All patients provided written informed consent.

Eligible suspected and confirmed COVID-19 patients presenting to the emergency department were consecutively invited to participate using convenient sampling. The inclusion criteria were: (1) adult (>18 years); and (2) suspected COVID-19 presenting with at least one of the following symptoms: 1 = fever (>38°C), 2 = cough, 3 = shortness of breath (Medical Research Council dyspnea score ≥4), and 4 = desaturation (O₂ ≤ 94 in room air). Patients unable to provide consent by themselves or an authorized relative were excluded. An RT-PCR test was performed for all patients and considered the reference test to confirm the diagnosis of COVID-19. The latest swab results were collected, including any repeat swabs. A positive result categorized the patient as a positive COVID-19 case. Blood gas analysis through venous samples were performed including partial pressure of O₂ (pO₂), partial pressure of CO₂ (pCO₂), hemoglobin oxygen saturation (SO₂), bicarbonate (HCO₃⁻), and pH levels.

2.2. Ultrasound Imaging and Scoring

The patients were scanned using the Butterfly iQ (Butterfly Network Inc., Guilford, CT, USA) with a probe size of 185 × 56 × 35 mm² equipped with 9000 micromachined sensors. The lung preset with a default depth of 13 cm operating at frequency 1–5 MHz was used across all examinations. We followed the 12-zone protocol (Figure 1), which divides each lung into six zones (superior/inferior, anterior/lateral/posterior) [15]. The probe was placed longitudinally in each zone and a still B-mode image was recorded.

Scanning was performed by the attending Emergency Room (ER) physicians, all of whom are board-credited with prior knowledge and training in LUS. To standardize image interpretation, they were provided with case examples demonstrating the different ultrasound scores [10,16,17]. The LUSS from the COVID-19 Lung Ultrasound in Emergency (CLUE) protocol was utilized to score the images [10,16,17]. Each zone was scored from 0 to 3. The definition for each score is as follows: score 0 = smooth pleural line, A-lines, maximum of 2 B-lines; score 1 = irregular or thickened pleura, >2 B-lines; score 2 = subpleural consolidation (<1 cm thickness), confluent B-lines; and score 3 = large consolidation (>1 cm), air bronchograms. According to the LUS–CLUE protocol, the patients were classified based on the total score (0–36) as normal (score 0), mild (score 1–5), moderate (score 6–15), and severe (>15) [10]. Moreover, the patients were dichotomized, based on a previously reported threshold, as positive if they had a score >12 [18]. The images were also retrospectively evaluated for the presence of the light beam artifact, which is characterized as a thick and strong hyperechoic shining band of confluent B-lines artifact arising from a relatively smooth pleura and surrounded by spared areas of A-lines [11,13]. The physician had an aid to record the score for each zone while scanning.

To test for LUSS interreader reproducibility, an external reader with 10 years of experience in ultrasonography—who was blinded to the original scores and clinical information including RT-PCR results—retrospectively recorded the images. Finally, the examination duration was retrospectively calculated to evaluate the feasibility by calculating the time difference between the first and last images.

2.3. Statistical Analysis

Descriptive and inferential statistics were done using SPSS version 27 (IBM Corp., Armonk, NY, USA). Fisher’s exact test was used
to compare the controls against COVID-19 patients for dichotomous categorical variables. To compare the two groups in terms of continuous variables, an independent samples t-test was used or the alternative nonparametric Mann–Whitney U-test when the data were not normally distributed. The Receiver Operating Characteristic (ROC) curve was used to represent the overall diagnostic performance of the LUSS by computing the area under ROC curve. The 95% Confidence Intervals (CIs) were provided and computed using the exact binomial method. A p-value <0.05 was regarded as statistically significant. No missing or indeterminate data were encountered.

The interrater reproducibility of the ordinal LUSS of each image was analyzed using Kendall’s coefficient of concordance (W), which provides an agreement coefficient (W) between 0 and 1, where 1 indicates that all readers ranked the images perfectly in agreement with each other [19]. In addition, the interrater reproducibility on the total LUSS score for each patient was analyzed using Intraclass Correlation Coefficients (ICCs). The interreader reproducibility in detecting the light beam artifact was tested using Cohen’s kappa (κ). The reproducibility results were interpreted as follows: 0.00–0.20 = poor agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; and >0.80 = almost perfect agreement [20].

3. RESULTS

3.1. Patient Characteristics

A total of 20 patients participated in this study. The mean age ± standard deviation was 49 ± 15 (range 20–71) years, and three of the patients were female. A total of eight patients (40%) who had a negative RT-PCR test result were considered controls. The demographic and clinical characteristics of COVID-19 patients and controls are listed in Table 1. The complete patient characteristics including blood gas analysis results are included in Supplementary Table S1. Those in the control group were on average younger than those in the COVID-19 group by approximately 14 years (p = 0.044). The most frequent symptoms were cough (80%) and shortness of breath (75%) followed by tiredness/lethargy (50%), sore throat (45%), and chest pain (45%). Six patients (30%) presented with fever (>38.0°C).

3.2. Lung Ultrasound

All participants had a complete LUS scan. The median scanning duration was 10.5 minutes and ranged between 5 and 16 min. Pleural effusions were detected in one COVID-19 patient and in one control. The LUSS was almost double in COVID-19 patients, with a median score of 17 compared to 9.5 in controls (Table 1). However, this difference was not statistically significant (p = 0.17) because of the wide variation in the readings (interquartile range = 13). Indeed, the ROC curve (95% CI) for LUSS was modest at 0.693 (95% CI, 0.449–0.936; p = 0.153) indicating poor discrimination between the patients and controls (Figure 2). When categorizing patients >12 points as positive [18], the diagnostic performance was also modest at 66.6% (95% CI, 34.9–90.0%) sensitivity and 75.0% (95% CI, 34.9–96.8%) specificity. Higher LUSS scores of approximately 30% were noted in the inferior lung zones. As for the LUSS classifications, the Chi-square test of homogeneity showed that the proportions were not significantly different between the two groups (p = 0.11). Nevertheless, there was a particularly larger number of COVID-19 patients than controls in the severe classification (Figure 3).

The light beam artifact was detected in 11 patients, nine of whom had COVID-19 (Figure 4). The difference in proportions was statistically significant (p = 0.024). The results diagnostically achieved 81.8% (95% CI, 48.2–97.7%) sensitivity, 75.0% (95% CI, 34.9–96.8%) specificity, 82% (95% CI, 56.7–93.9%) positive predictive value, 75.0% (95% CI, 44.6–91.8%) negative predictive value, and an overall accuracy of 78.9% (95% CI, 54.4–93.9%).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>COVID-19 (n = 12)</th>
<th>Controls (n = 8)</th>
<th>P for difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>54.7 (12.6)</td>
<td>40.4 (14.6)</td>
<td>0.044</td>
</tr>
<tr>
<td>Sex (M:F)</td>
<td>9:3</td>
<td>8:0</td>
<td>–</td>
</tr>
<tr>
<td>Oxygen saturation in room air</td>
<td>86.5 (8.7%)</td>
<td>90.6 (9.9%)</td>
<td>0.36</td>
</tr>
<tr>
<td>Required oxygen therapy</td>
<td>9</td>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>Never smoked</td>
<td>11 (91%)</td>
<td>7 (87%)</td>
<td>0.65</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart disease</td>
<td>2</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Hypertension</td>
<td>4</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Diabetes</td>
<td>5</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Asthma</td>
<td>1</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>COPD</td>
<td>1</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fever (&gt;38°C)</td>
<td>5</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Cough</td>
<td>10</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>9</td>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>Sore throat</td>
<td>6</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Chest pain or pressure</td>
<td>6</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Loss of taste and/or smell</td>
<td>3</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Joint pain</td>
<td>5</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>Lethargy/tiredness</td>
<td>8</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Aches and pains</td>
<td>3</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Headache</td>
<td>6</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Lung ultrasound score</td>
<td>17 (18)</td>
<td>9.5 (12)</td>
<td>0.17</td>
</tr>
<tr>
<td>Presence of light beam artifact</td>
<td>9 (75%)</td>
<td>2 (25%)</td>
<td>0.024</td>
</tr>
</tbody>
</table>

*Reported as mean (standard deviation). †Reported as median (interquartile range). COPD, chronic obstructive pulmonary disease.
The light beam artifact (LBA) was originally hypothesized to reflect a characteristic pattern and resulted in superior diagnostic performance. The LUS image interpretation based on our findings. Nevertheless, further research on a larger sample size can elucidate the observed diagnostic performance discrepancy noticeable in the LUS literature can be explained by the varying definitions of LUS outcome measure (i.e., how to interpret the images). Some studies have merely used any presence of B-lines or consolidations indicative of COVID-19. Others developed an arbitrary LUS scoring system based on a 5-point Likert scale or a 6-point system. There is an apparent disagreement on the scoring of small consolidations, where sometimes they are scored as 1 whereas in other studies they are scored as 2. In contrast, other researchers deemed consolidations as signs against COVID-19 pneumonitis. The promising standardized LUS may, unfortunately, not be the called for solution to standardize LUS image interpretation based on our findings. Nevertheless, further research on a larger sample size can elucidate the observed high scores in the COVID-19 group. Recently, Volpicelli et al. developed a mutually exclusive classification system that categorizes the findings as low, intermediate, high, or alternative probability for COVID-19. Their system used a qualitative approach for classifying the patients, which yielded excellent accuracy. The light beam artifact was originally hypothesized to reflect a patchy area of COVID-19 pneumonia. It has been specifically described as "a shining band-form artifact spreading down from a large portion of a regular pleural line, often appearing and disappearing with an on–off effect in the context of a normal A-line lung pattern visible in the background". It is thought to represent an early sign for viral spread in the peripheral lung corresponding with the ground-glass opacifications detected on CT. However, few researchers argue the specificity of the light beam artifact to COVID-19, claiming that it is simply "a patchy area of COVID-19 (positive RT-PCR)."
white lung” [34]. In contrast, our data showed that it had a high diagnostic yield as a typical LUS feature of COVID-19 and can be reliably detected using a handheld ultrasound device. Indeed, our results correspond to Volpicelli et al.’s [33] findings that suggested its value as an independent predictor of RT-PCR positivity.

As with all ultrasound artifacts, the light beam artifact can be influenced by the probe orientation, machine settings, and probe type. We recommend using a convex profile probe operating at low frequency (<3 MHz) with the focus positioned at the pleural line for accurate interpretation. Notwithstanding, our results support the role of the light beam artifact as a more sensitive and specific pattern compared to conventional patterns in LUS. This study’s finding agrees with a few published case reports and observations [11,13,35]. Our study is the first to systematically analyze this artifact in an observational study on COVID-19. An international multicenter research is currently ongoing to investigate the specificity to particular COVID-19 phenotypes [11,35].

Few studies have evaluated the interreader agreement of LUS in COVID-19. A recent study by Kumar et al. [36] investigated the reproducibility of LUS qualitative findings. They reported a substantial interreader reproducibility of 0.79 (95% CI, 0.72–0.87) kappa value for normal LUS, single, and multiple B-lines but moderate (κ = 0.57; 95% CI, 0.50–0.64) and fair (κ = 0.23; 95% CI, 0.15–0.30) reproducibility for consolidations and pleural thickening, respectively. Other groups used the relatively similar intensive care unit LUS scoring system and reported a substantial interreader reproducibility (κ = 0.79) [22]. When investigating the reproducibility of the semiquantitative LUSS, this study demonstrated promising agreement between readers. This finding was also true for the characteristic light beam artifact. Despite the promising findings of such subjective LUS findings, the introduction of an automated computer-aided quantitative scoring technique might offer an objective and operator-independent assessment [37].

In terms of feasibility, Wolfshohl et al. [38] conducted a survey on emergency physicians which suggested the majority were not prepared to perform LUS in emergency settings because of lack of targeted training despite their awareness of its superior diagnostic performance. In contrast, Narinx et al. [26] reported a more positive experience in their study, which agrees with our experience, in which LUS was a useful, rapid, and safe tool in the ER. In our study, all physicians were trained and reported no scanning predicaments. The reported average scanning time of 10 min supports the feasibility of LUS in the emergency setting for COVID-19 cases.

The majority of published studies used cart-based ultrasound systems. Our study is the first to investigate the handheld Butterfly IQ device in an emergency setting. A single previous study retrospectively investigated this device in comparison to a high-end ultrasound system on hospitalized COVID-19 patients [39]. Bennett et al. [39] reported an excellent concordance coefficient (95% CI) of 0.989 (0.978–0.994). Such handheld devices are relatively cheaper and demonstrated excellent reproducibility. The most valuable feature is the small size factor, which suggests lower infection risks because of the relative ease of disinfection. They can also be encased in a cover and are not equipped with a cooling fan. Indeed, infection control is paramount in this pandemic, and several scientific societies have endorsed the use of handheld ultrasound devices for COVID-19 [14]. However, handheld devices offer limited to no adjustments to the technical sonographic settings such as focusing and frequency manipulation. They also come with a few trade-offs, such as lower spatial resolution and smaller screen size. Our reported lower diagnostic performance for the overall LUSS score might be confounded by the imaging ability of the handheld device. This explanation can be supported by the reported substantial interrater reproducibility when interpreting the images.

We support the use of LUS as a bedside assessment tool for suspected COVID-19 patients in emergency settings because of several advantages. It can be useful for prompt isolation of patients with suggestive COVID-19 findings, such as the light beam pattern and high LUSS scores. LUS can be carried out using rapid and simplified protocols. Dacrema et al. [40] adopted a simplified LUS protocol for COVID-19 diagnosis focusing on six posterior zones only, which yielded excellent sensitivity. Furthermore, we support the sole use of LUS and sparing chest X-ray and CT for selected and critical cases, in agreement with the Saudi COVID-19 safety guide for healthcare workers, the Multinational Consensus Statement from the Fleischner Society, and ACR [6,41].

This study had several limitations. The sample size was small for several reasons, including the steep decline in COVID-19 cases during the recruitment phase. The number of rejections to participate was not recorded. However, it was deemed relatively high (approximately 4:1), especially by female patients because of the examination’s revealing nature. The study did not correlate LUS with variables for arterial blood gas analysis or CT because of the data’s unavailability. The positive and negative predictive values we reported on the light beam artifact should be interpreted cautiously as they are highly dependent on the disease’s prevalence. They will unlikely represent the true predictive values in other communities or globally because of the fluctuating prevalence of COVID-19. More longitudinal research is warranted for investigating the prognostic role of LUS concerning mortality, ICU admission, and intubation as well as its value in patient management. This is merited with reports suggesting LUS can lead to overdiagnosis leading to higher hospital admissions, which may further overwhelm healthcare systems [42].

5. CONCLUSION

In conclusion, semiquantitative scores of LUS findings using the LUSS were relatively higher in COVID-19 patients than controls but had limited diagnostic performance for identifying suspected cases compared to RT-PCR as a reference test. In contrast, a recently recognized qualitative sign called the light beam artifact had superior sensitivity and specificity. Overall, the results suggest that LUS can be feasibly and reliably implemented in the emergency department while waiting for the RT-PCR confirmatory results. Owing to modest specificity, suggestive LUS outcomes with a positive light beam artifact should be confirmed with the RT-PCR test. Future research should investigate the temporal changes of LUS findings and study their prognostic value.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.
AUTHORS’ CONTRIBUTION

AMA and MJA contributed in study conceptualization. AMA, AAA, HMA, SAA, TSA and KGA contributed in study design. AAA, HMA and SAA contributed in data collection. AMA and MJA contributed in data interpretation and analysis. AMA and MJA contributed in manuscript writing and preparation. All authors reviewed the final draft.

FUNDING

This research has been funded under Dr Sulaiman Al Habib Medical Journal grant to support COVID-19 research. The funding body had no role or influence in the data interpretation or conclusions.

ACKNOWLEDGMENTS

The research team is grateful for Al Habib Medical Group for funding this project in support of local COVID-19 research. The authors are thankful to the emergency department residents (Dr. Abdullah Alshehri, Dr. Nader Bokhari, Dr. Mansour Aldhafeeri, Dr. Malek Alaaullah) and supporting staff as well as the research participants for their contribution to the research.

SUPPLEMENTARY MATERIAL

Supplementary data related to this article can be found at https://doi.org/10.2991/dsahmj.k.210823.001

REFERENCES


