Lung ultrasonography in pregnant women during the COVID-19 pandemic: an interobserver agreement study among obstetricians

Murat Yassa[1], Memis Ali Mutlu[1], Pinar Birol[1], Taha Yusuf Kuzan[2], Erkan Kalafat[3,4], Canberk Usta[1], Emre Yavuz[1], Ilkhan Keskin[1], Niyazi Tug[1]

Departments of [1]Obstetrics and Gynecology and [2]Radiology, Sehit Prof. Dr. Ilhan Varank Sancaktepe Training and Research Hospital, Istanbul; [3]Department of Statistics, Middle East Technical University, Faculty of Arts and Sciences, Ankara; [4]Department of Obstetrics and Gynecology, Faculty of Medicine, Ankara University, Ankara, Turkey

Purpose: This study investigated interobserver agreement in lung ultrasonography (LUS) in pregnant women performed by obstetricians with different levels of expertise, with confirmation by an expert radiologist.

Methods: This prospective study was conducted at a tertiary "Coronavirus Pandemic Hospital" in April 2020. Pregnant women suspected to have coronavirus disease 2019 (COVID-19) were included. Two blinded experienced obstetricians performed LUS on pregnant women separately and noted their scores for 14 lung zones. Following a theoretical and hands-on practical course, one experienced obstetrician, two novice obstetric residents, and an experienced radiologist blindly evaluated anonymized and randomized still images and videoclips retrospectively. Weighted Cohen's kappa and Krippendorff's alpha tests were used to assess the interobserver agreement.

Results: Fifty-two pregnant women were included, with confirmed COVID-19 diagnosis rate of 82.7%. In total, 336 eligible still images and 115 videoclips were included in the final analysis. The overall weighted Cohen's kappa values ranged from 0.706 to 0.912 for the 14 lung zones. There were only seven instances of major disagreement (>1 point) in the evaluation of 14 lung zones of 52 patients (n=728). The overall agreement between the radiologist and obstetricians for the still images (Krippendorff's α=0.856, 95% confidence interval [CI], 0.797 to 0.915) and videoclips (Krippendorff's α=0.785; 95% CI, 0.709 to 0.861) was good.

Conclusion: The interobserver agreement between obstetricians with different levels of experience on still images and videoclips of LUS was good. Following a brief theoretical course, obstetricians' performance of LUS in pregnant women and interpretation of pre-acquired LUS images can be considered consistent.

Keywords: COVID-19; Interobserver variability; Lung ultrasound; Pregnancy; SARS-CoV-2

Introduction

Considering its effectiveness and practicality, the clinical utilization of lung ultrasonography (LUS) has become widespread, particularly in emergency medicine, intensive care units, and pediatric settings [1–3]. LUS can be used to identify pathological changes in the peripheral lung tissue [4].

The utility of LUS was previously shown during the 2009 H1N1 influenza A pandemic and other influenza outbreaks [5,6]. Although the radiological diagnosis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is primarily based on chest computed tomography (CT) [7], a growing body of evidence suggests that LUS can be used for diagnosing coronavirus disease 2019 (COVID-19) [4,8–16]. LUS performed by an experienced clinician is helpful for triage and assessments of the severity of the disease [16].

LUS is a low-energy diagnostic tool without ionizing radiation that is a safe imaging modality for pregnant women [17]. LUS has been shown to be useful for the detection and monitoring of several respiratory diseases in pregnant women, and if performed by an experienced clinician, it aids in triage and assessments of disease severity [16,17]. It has recently been proposed that LUS should be performed by obstetricians during the COVID-19 pandemic given the exceptional circumstances and obstetricians' inherent familiarity with ultrasonography [12]. Our recent experience with a large case series established that the routine use of LUS after an obstetric ultrasound assessment can influence the clinical management of pregnant women with COVID-19 [18].

LUS is relatively easy to learn and less technically demanding than other sonographic examinations, since it is based on specific pattern recognition and does not require complex measurements [19]. However, it is operator-dependent and needs to be performed and interpreted by trained clinicians [20]. This has raised questions regarding the generalisability of LUS performed by obstetricians during the COVID-19 pandemic and the consistency of LUS among obstetricians, and the data in the literature on this issue are scarce. In this study, we investigated interobserver agreement in LUS in pregnant women performed by obstetricians with different levels of expertise, with confirmation of our findings by an expert radiologist.

Materials and Methods

Study Design, Setting, and Patients

This prospective, blinded, cross-sectional study was conducted at a single tertiary "Coronavirus Pandemic Hospital" in Istanbul, Turkey in April 2020. Women with a confirmed healthy pregnancy who were admitted to the antenatal service due to suspected COVID-19 were included. Reverse transcription polymerase chain reaction (RT-PCR) testing was performed from a nasal/throat swab sample to confirm the diagnosis of COVID-19. Women with chronic morbidities related to the respiratory system that could disrupt the accuracy and interpretation of LUS findings (e.g., tuberculosis or sarcoidosis) were excluded. All women underwent LUS following a routine fetal assessment. An image acquisition protocol was followed, as described in detail below. The study was approved by the Scientific Research Platform of the Health Ministry and the local institutional review board. Informed consent was obtained from all patients for the study and the anonymous use of their LUS findings.

LUS Protocol

According to our local protocol, all pregnant women admitted to the obstetrics unit systematically underwent routine LUS within 24 hours of their arrival at the hospital, as a first-line imaging technique. For the purposes of this study, the data were obtained in a two-step approach. First, two independent and experienced obstetricians who were blinded to patients' clinical severity and infection status performed LUS on pregnant women separately. Their blinding to patients' conditions was achieved with the help of an independent nurse who was working in the pandemic service. The same two obstetricians specifically scored 14 anatomical landmarks in real time while assessing the patient and noted the scores of their LUS findings separately. In order to minimize the exposure to the raters and the healthcare staff (chaperone nurses, other nurses working in the setting, etc.), the operators did not make special efforts to obtain images with the best quality for storage purposes. It was obligatory to save images taken from the posterior and lateral basal lobes (either normal or abnormal), as COVID-19 mostly effects the periphery of these regions, as well as any suspicious imaging findings. However, the remaining apparently normal images were stored at the discretion of the operator. This step aimed to establish interobserver agreement by dynamically evaluating the scores as a whole and separately for each lung zone.

Second, all still images and videoclips were stored digitally on an external computer for later analysis after anonymization and randomization with the help of an independent nurse who was working in the obstetrics service. One experienced specialist, one experienced radiologist, and two novice residents blindly evaluated and scored the still images and videoclips.

Sonographers and Raters

All sonographer and raters participating in the study underwent a single-day didactic course held by a radiologist with over 6 years of experience and by the first author, who is experienced in performing LUS in pregnant women [18]. The structured course

included a detailed presentation on how to perform and interpret LUS in pregnant women, and a brief hands-on practical course on 10 normal patients and five COVID-19 patients. The course was completed after observing alignment in didactic reference cases between the course attendees and the expert physician. Those reference cases were collected from the literature and from our previous case series [18].

Performing LUS

Following the recommended high-level protection rules [21], all lung images and videoclips were obtained with a dedicated machine (EA720, Esaote S.p.a., Italy; manufactured by Eizo Nanao Corp.) and a 1–8 MHz convex transducer with regular obstetric presets for patients with suspected or confirmed COVID-19 [21]. Sonographers were asked to obtain a representative image and videoclip from each zone. The thickness and irregularity of the pleura, sporadic/multiple B-lines, small and large consolidations, and pleural effusion were the sonographic elements that the sonographers focused upon. Images and videoclips were obtained while patients were sitting or in a 90° lateral position. A standardized 14-area scanning protocol [12,13] was applied (3 posterior, 2 lateral, and 2 anterior) for each patient and for 10 seconds in each area along the indicated lines. According to the International Standardization of the Use of Lung Ultrasound for Patients with COVID-19 protocol, the 14 anatomic landmarks were defined as follows [13]: (1) right basal on the paravertebral line above the curtain sign; (2) right middle on the paravertebral line at the inferior angle of the shoulder blade; (3) right upper on the paravertebral line at the spine of the shoulder blade; (4) left basal on the paravertebral line above the curtain sign; (5) left middle on the paravertebral line at the inferior angle of the shoulder blade; (6) left upper on the paravertebral line at the spine of the shoulder blade; (7) right basal on the midaxillary line below the internipple line; (8) right upper on the midaxillary line above the internipple line; (9) left basal on the midaxillary line below the internipple line; (10) left upper on the midaxillary line above the internipple line; (11) right basal on the midclavicular line below the internipple line; (12) right upper on the midclavicular line above the internipple line; (13) left basal on the midclavicular line below the internipple line; (14) left upper on the midclavicular line above the internipple line. Scanning from the intercostal space was preferred where applicable.

Interpretation of the Images

Each area was given a score between 0 and 3 according to the specific pattern [13]. A pattern with a continuous and regular pleura line, as well as horizontal artifacts referred to as A-lines, was scored as 0. A pattern with an indented pleural line and with sporadic vertical white areas below the point of discontinuity in the pleural line, referred to as sporadic B-lines, was scored as 1. A pattern with broken pleura, small consolidated areas below the discontinuity, and multiple vertical white areas that reached to the bottom of the field of view, referred to as multiple B-lines, was scored as 2. A pattern with severely broken pleura and a dense and largely extended white lung pattern, with or without larger consolidations, was scored as 3. At the end of the procedure, the highest score obtained for each area was noted (e.g., landmark 1, score 0; landmark 2, score 1; and so on) (Fig. 1). Images or videos that could not be scored were left empty and excluded from the analysis. The quality of the images after randomization and anonymization were evaluated by two independent experts, mainly taking the following features into consideration: the processability of the integrity of the pleural line throughout the image, an excessive increase in gain, the integrity of B-lines reaching to the bottom of the image, and the artifacts that can overlap physiological A-lines.

Statistical Analysis

The collected data were analyzed with SPSS version 22.0 (IBM Corp., Armonk, NY, USA) and R for Statistical Computing Software (version 4.0.2, R Foundation for Statistical Computing, Vienna, Austria). The normality of the demographic data was assessed with the Shapiro-Wilk test. Demographic data were summarised as the median and interquartile range for non-normally distributed data and as the mean and standard deviation for normally distributed data. The diagnostic performance of LUS including the sensitivity, specificity, accuracy and positive and negative predictive values was calculated by accepting LUS score of 1 or greater as abnormal. In the first analysis, Cohen's kappa with squared weights was used to assess agreement between two observers for 14 anatomical landmarks. In the second analysis, Krippendorff's alpha was used to assess overall agreement among four observers in the assessment of ultrasound images and video clips. Confidence intervals were calculated via bootstrapping (50,000 replications), and the Cohen's kappa results were interpreted as follows: values ≤0 as indicating no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement [22]. Krippendorff's alpha values above 0.80 were considered to indicate good agreement (0, no agreement; 1, perfect agreement) [23]. Data were visually demonstrated using plots depicting categorical differences between observers and the category as rated by one of the observers. The category differences between the observers, on a scale of −3 to 3, were plotted against the category as rated by the first observer. Bootstrapped limits of agreement were depicted on the graph and the proportion of overlapping points was reflected in the dot size [24]. A P-value of

<0.05 was considered to indicate a significant difference.

Results

In total, 52 pregnant women were included in the study. No patients were excluded. The positivity rate for COVID-19 was 82.7% (n=43). Age, parity, and gestational week were non-normally distributed, whereas body mass index was normally distributed. Demographic data are presented in Table 1. The majority of the patients (73.1%, 38 of 52) were symptomatic. The patients' clinical condition was mild in 53.8% (n=28), moderate in 15.4% (n=8), and severe in 3.8% (n=2) of the patients. A randomized and anonymized series of 350 still images and 115 videoclips were retrospectively evaluated

and scored. Fourteen images were excluded due to poor quality. A total of 336 still images and 115 videoclips were included in
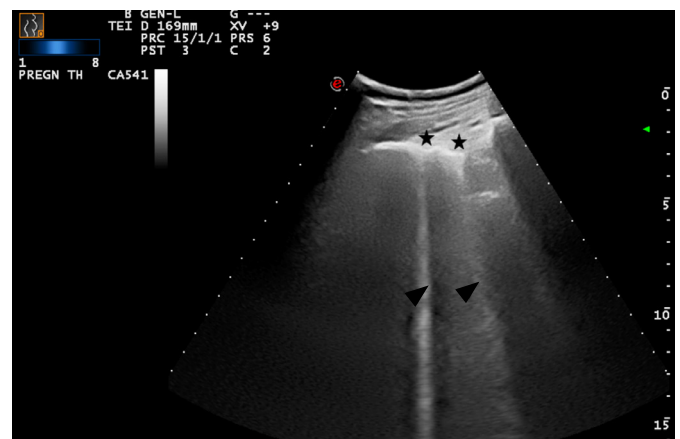
Table 1. Demographic features of the pregnant women suspected to have COVID-19 (n=52)

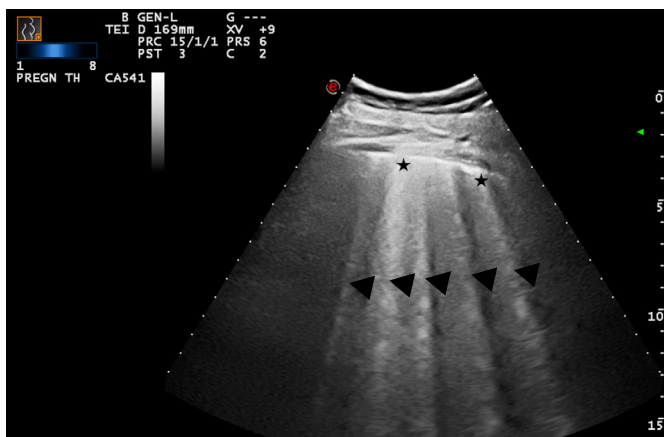| Variable	Median (IQR) or mean±SD	Minimum	Maximum
Age (yr)	28.0 (9.8)	19.0	40.0
Parity	2.0 (2.0)	0	6.0
Gestational week	27.0 (18.0)	4.0	40.0
BMI (kg/m^2)	27.0±4.6	17.3	40.8

COVID-19, coronavirus disease 2019; IQR, interquartile range; SD, standard deviation; BMI, body mass index.

A

B

C
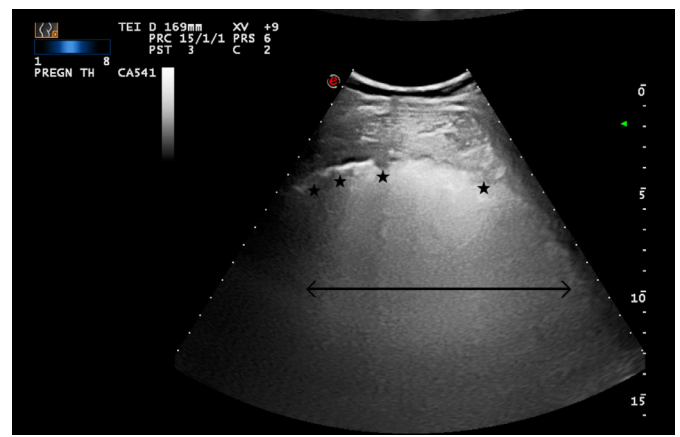
D

Fig. 1. Lung ultrasonography pattern.
A. This figure shows a normal lung ultrasonography (LUS) pattern with the convex transducer positioned longitudinally. The pleural line is continuous. Arrowheads indicate horizontal A-lines at regular intervals, classified as score 0. B. This figure shows an abnormal LUS pattern with the convex transducer positioned in the intercostal space. Stars indicate a discontinuity in the pleural line and the arrowheads indicate sporadic B-lines below the point of discontinuity, classified as score 1. C. Stars indicate a disrupted pleural line, and arrowheads indicate multiple B-lines reaching the bottom of the field of view, classified as score 2. D. Stars indicate a severely broken pleural line and associated subpleural consolidations. A double-headed arrow indicates a generalized white lung pattern, classified as score 3.

the final analysis (Fig. 2). In this cohort, the sensitivity, specificity, positive and negative predictive values, and the accuracy of LUS in detecting COVID-19 diagnosed with RT-PCR testing were as 72.1% (95% confidence interval [CI], 56.3% to 84.7%), 77.8% (95% CI, 40.0%. to 97.2%), 93.9% (95% CI, 81.8% to 98.2%), 36.8% (95% CI, 24.4% to 51.4%), and 73.1% (95% CI, 59.0% to 84.4%), respectively.

Scoring Agreement on All Anatomical Landmarks

The overall weighted Cohen's kappa values ranged from 0.706 to 0.912 (Table 2). All anatomic landmarks except 6 (left-sided, posterior, upper), 10 (left-sided, lateral, upper) and 11 (right-sided, anterior, lower) showed substantial to almost perfect agreement. Landmarks 6 and 10 showed moderate to perfect and landmark 11 showed moderate to substantial agreement among observers (Table 2).

In the first analysis, the proportions of LUS findings scored by the first obstetrician as 0, 1, 2 and 3 were 70.5% (n=513), 14.1% (n=103), 10% (n=73), and 5.4% (n=39), respectively. The proportions of LUS findings scored by the second obstetrician as 0, 1, 2, and 3 were 68.7% (n=500), 14.8% (n=108), 11.1% (n=81), and 5.3% (n=39), respectively. The bootstrapped limits of agreement plot of LUS score categories showed that >99.0% of observations were within the non-parametric limits of agreement (−1 to 1) and the observers had perfect agreement for the majority of observations (Fig. 3). There were only seven instances of major disagreement (>1 point) in the evaluation of 14 anatomical zones in 52 patients (n=728).

Interpretation Agreement for Scores of LUS Findings between Physicians with Different Expertise

The overall agreement among the radiologist, the specialist obstetrician, and the two obstetrics residents was good for both still images (α=0.856; 95% CI, 0.797 to 0.915) and videoclips (α=0.785; 95% CI, 0.709 to 0.861) (Table 3).

Table 2. Interobserver agreement between two obstetricians for lung ultrasonography using recorded images on 14 anatomical landmarks (n=52)

	Weighted Cohen's kappa[a]	95% Confidence interval[b]
L1	0.832	0.729−0.934
L2	0.821	0.697−0.945
L3	0.824	0.656−0.991
L4	0.808	0.683−0.933
L5	0.811	0.660−0.962
L6	0.720	0.493−0.947
L7	0.860	0.760−0.960
L8	0.837	0.733−0.941
L9	0.823	0.675−0.971
L10	0.713	0.489−0.938
L11	0.706	0.498−0.706
L12	0.838	0.694−0.982
L13	0.912	0.808−1.000
L14	0.868	0.703−1.000

[a]Cohen's kappa with squared weights. [b]Bootstrapped confidence intervals.

43 (82.7%) Women were tested positive for COVID-19 by RT-PCR

9 (17.3%) Women were tested negative for COVID-19 by RT-PCR

52 Pregnant women suspicious of or diagnosed with the COVID-19 infection

Lung ultrasound performed by two blinded obstetricians and findings were noted separately for 14 anatomical landmarks

728 Scoring were included to the first analysis

350 Images and 115 videoclips were collected

14 Images were excluded due to poor quality

336 Images, 115 videoclips were included to second analysis
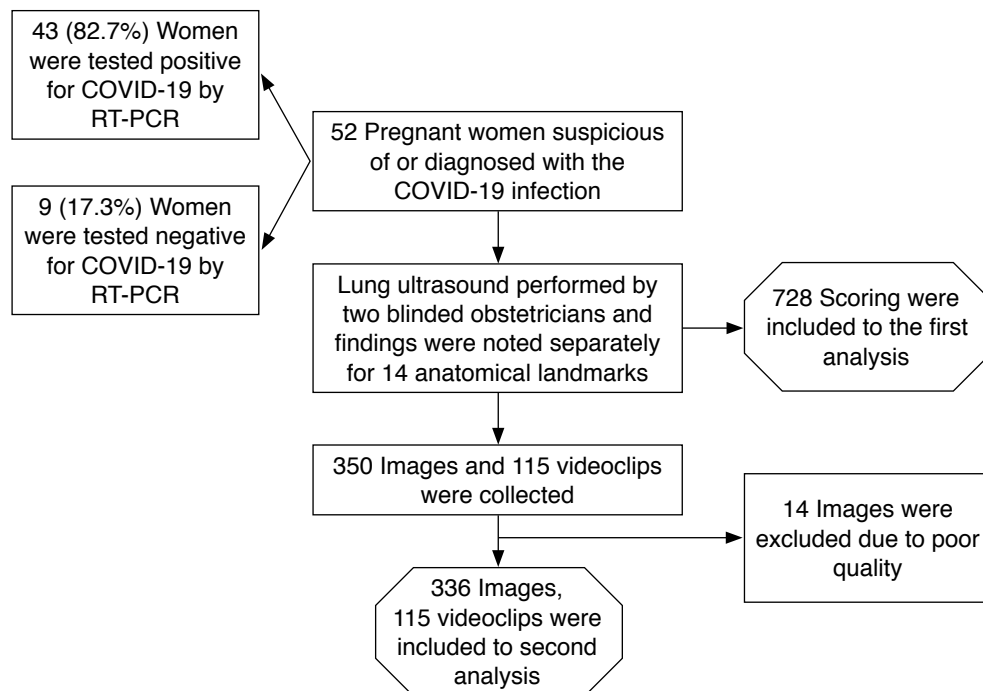
Fig. 2. Patient selection flowchart.

Good interobserver agreement for the interpretation of still images (Fig. 1) was found between the residents (0.831), between the obstetrician and the two obstetric residents (0.806 and 0.858, respectively), and between the obstetrician and the radiologist (0.780), whereas it was low to good between the radiologist and the obstetric residents (0.748 and 0.762, respectively) (Table 4).

The interobserver correlation matrix for the interpretation of videoclips (Video clips 1–3) was good between the residents (0.898), between the obstetrician and the two obstetric residents (0.896 and 0.873, respectively), between the obstetrician and the radiologist (0.832), and between the radiologist and the obstetric residents (0.782 and 0.801, respectively) (Table 4).
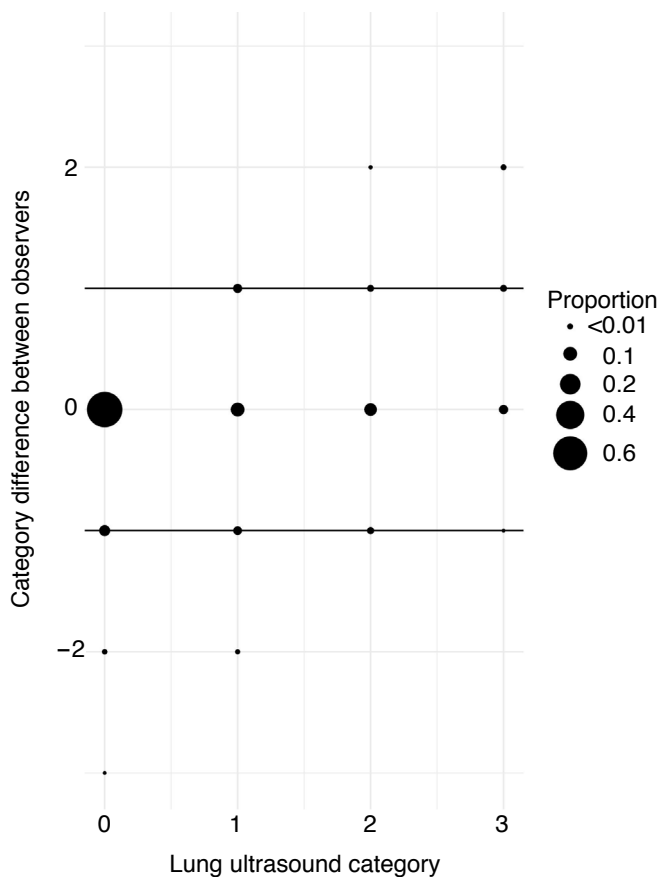

Fig. 3. Bootstrapped limits of agreement plot. The category differences between the observers on a scale of −3 to 3 was plotted against the category rated by one of the observers. In this bootstrapped limits of agreement plot, the category differences between the observers on a scale of −3 to 3 was plotted against the category rated by the first observer. Non-parametric limits of agreement are depicted on the graph and the proportion of overlapping points is reflected in the dot size. The bootstrapped limits of agreement plot of lung ultrasound score categories show that >99.0% of the observations were within the non-parametric limits of agreement (−1 to 1).

Table 3. Interobserver agreement between obstetricians and a radiologist for lung ultrasonography using either recorded images or videoclips

Comparison	Krippendorff's alpha	95% Confidence interval[a]
Images (n=336)		
Overall	0.856	0.797−0.915
Resident 1 vs. obstetrics consultant	0.913	0.847−0.978
Resident 2 vs. obstetrics consultant	0.830	0.742−0.918
Resident 1 vs. resident 2	0.892	0.838−0.946
Resident 1 vs. radiology consultant	0.852	0.776−0.929
Resident 2 vs. radiology consultant	0.831	0.740−0.923
Radiology consultant vs. obstetrics consultant	0.822	0.729−0.915
Videoclips (n=115)		
Overall	0.785	0.709−0.861
Resident 1 vs. obstetrics consultant	0.806	0.723−0.890
Resident 2 vs. obstetrics consultant	0.858	0.792−0.925
Resident 1 vs. resident 2	0.831	0.753−0.909
Resident 1 vs. radiology consultant	0.736	0.631−0.840
Resident 2 vs. radiology consultant	0.684	0.555−0.812
Radiology consultant vs. obstetrics consultant	0.780	0.690−0.902

[a]Bootstrapped confidence intervals.

Table 4. Inter-observer correlation matrix between obstetricians and radiologist for lung ultrasound with using either recorded images or videoclips

	Specialist obstetrician	Resident obstetrician 1	Resident obstetrician 2	Radiologist
Images (n=336)				
Specialist obstetrician	−	0.822	0.804	0.767
Resident obstetrician 1		−	0.921	0.762
Resident obstetrician 2			−	0.748
Radiologist				−
Videoclips (n=115)				
Specialist obstetrician	−	0.896	0.873	0.832
Resident obstetrician 1		−	0.898	0.782
Resident obstetrician 2			−	0.801
Radiologist				−

Discussion

This paper provides evidence that the agreement between obstetricians with different levels of experience in LUS in pregnant women was moderate to excellent in the prospective assessment of 14 anatomical zones. The agreement among obstetricians and a radiologist for LUS was retrospectively found to be good, using either recorded still images or videoclips. To the best of our knowledge, this study presents the first solid data in the literature regarding interobserver agreement between obstetricians when performing LUS in pregnant women.

The ongoing COVID-19 pandemic affects vulnerable populations more severely, including pregnant women. In a recent systematic review of 441 pregnant women, the need for oxygen support with a nasal cannula and serious morbidities, including mechanical ventilation, multi-organ dysfunction, and extracorporeal membrane oxygenation were observed in 24% and 11% of cases, respectively [25]. The clinical presentation of COVID-19 in women can vary widely across different populations and settings. Asymptomatic initial presentations are also common among pregnant women [25].

At this point, point-of-care or bedside LUS has clear benefits, particularly for use in pregnant women. These benefits include LUS being rapid, practical, cheap, and easy to learn and the fact that it provides dynamic real-time results [10]. Currently, chest CT is the most sensitive imaging method for detecting the respiratory involvement of COVID-19 infection, but CT and LUS have been found to be highly consistent in terms of diagnostic power [26]. LUS was proposed to have higher diagnostic accuracy than auscultation and chest X-rays [11].

Examinations involving ionizing radiation should only be used to diagnose COVID-19 pneumonia in pregnant women when it is necessary, and the dose should be kept as low as reasonably achievable [27]. *In utero* radiation exposure of a fetus can lead to prenatal death, malformations, growth restriction, mental retardation, childhood cancer, and maternal breast cancer, depending on the gestational age and the size of the mother [27,28]. Although the radiation dose of low-dose chest CT and limited repeats is considered to be acceptable and safe, the stochastic carcinogenic and mutagenic effects of ionizing radiation are known to have no dose threshold [29]. Therefore, clinicians should carefully weigh the risks against the potential benefits of imaging modalities that use ionizing radiation in pregnant women [27–29].

Although CT provides better diagnoses of deeper, central, and apical pulmonary lesions that do not extend to the pleural surface, LUS was found to be non-inferior to CT in detecting pleural and peri-pulmonary effusions, as well as smaller peri-pulmonary lesions [26,30]. With its non-ionizing nature, minimal contamination risk,

and repeatability for ongoing monitoring, LUS has advantages over CT and chest X-rays [26]. The routine use of LUS may also reduce the use of X-rays and promote selective CT scanning during the COVID-19 pandemic [11]. The refusal rates for chest CT may be higher among pregnant women with mild clinical conditions than among non-pregnant adults [31]. The use of LUS in the settings the COVID-19 pandemic may be beneficial in patients who do not give consent for chest CT and for follow-up in asymptomatic patients with initial negative chest CT findings [17].

Our findings are consistent with those of other LUS studies performed in pediatric settings. Gomond-Le Goff et al. [32] achieved good inter-rater agreement results independent of the raters' experience by retrospectively rating still images of LUS that had been performed in critically ill neonates. In another study similar to ours, four raters with different levels of experience retrospectively interpreted videoclips taken from 114 neonates with respiratory distress, and the physicians with different levels of experience were able to achieve good interobserver agreement [33]. Gullett et al. [34] evaluated the agreement among emergency physicians with different levels of experience with regard to findings in different lung zones. They found the best interobserver agreement in the anterior and superior thoracic zones. Similar to their findings, the interobserver agreement for anatomical landmarks 12 and 14 was good to excellent in our study. Gullett et al. [34] found moderate agreement between two emergency physician experts when they reinterpreted the videoclips retrospectively for B-lines. Since our interobserver agreement was good, we believe that this difference in the degree of agreement between our study and that of Gullett et al. can be explained by obstetricians' greater familiarity with ultrasonography compared to emergency physicians.

As new evidence emerges, future studies should focus on the correlation between LUS findings and the clinical severity of pregnant women infected with or suspected to have COVID-19, and on the routine use of LUS in labor wards [11,15]. Our findings of good interobserver agreement among obstetricians and an expert radiologist may pave the way for further research. The finding of good agreement among clinicians with different levels of experience will help respond to concerns that have been raised regarding LUS performed by obstetricians. Although we found good interobserver agreement between obstetricians with different levels of experience in performing LUS, performing and interpreting ultrasound is inherently user-dependent and inadequate training can therefore cause higher interobserver variability [10]. Vetrugno et al. [14] suggested that 25 supervised LUS examinations following a brief training should be adequate for performing LUS. Buonsenso et al. [35] proposed a specific single-day training program for obstetricians who are already well-trained in performing obstetric

ultrasonography, including a 2-hour theoretical course and structured tests to be applied before and after 10 ultrasound videoclips. We organized a specific structured course that consisted of a detailed theoretical component and a practical hands-on component with examinations of infected and healthy women. Later, the practical course was repeated with the examination of non-infected volunteer healthcare workers and by studying videoclips. Regular trainings consisting of a brief theoretical and practical course may explain the good interobserver agreement found in the current study.

One of the main limitations of this study is that it mainly focused on interobserver agreement, rather than the analysis of the diagnostic performance of LUS and the true reliability of LUS performed by obstetricians. The results of the current study can only confirm the interobserver agreement of obstetricians in the performance and interpretation of LUS in pregnant women, and the interobserver agreement of obstetricians with an experienced radiologist in the interpretation of pre-performed LUS images. However, this is an essential first step before conducting studies to establish the accuracy of LUS. Another limitation of this study is that it was aimed to assess interobserver, rather than intraobserver agreement, due to ethical considerations (i.e., in order to minimize the risk of transmission of COVID-19 to the observers and the healthcare staff). Secondly, the results regarding the accuracy of LUS in detecting SARS-CoV-2 infection compared with RT-PCR testing cannot be generalized, as the sample size was not adequate for that purpose and it was not the primary aim of this study. The design of this study may not be adapted for patients in critical conditions, as there were no intubated patients who underwent LUS. Lastly, the images of a significant part of the lung zones were not recorded and evaluated in the second analysis, although they were all scored in the first analysis. This may have caused selection bias, although there was no intentional exclusion of difficult cases.

One of the strengths of our study is the inclusion of videoclips for retrospective interpretation, since LUS is a dynamic and real-time examination modality. Furthermore, we produced interobserver agreement values for every lung zone, which may be useful for further research. Lastly, our cohort specifically represented the target population, since the positivity rate for COVID-19 was 82.7%. That may increase the generalizability of our findings.

In conclusion, the interobserver agreement in our findings from still images and videoclips that had been gathered from LUS was good between obstetricians with different levels of experience and a radiologist. The agreement did not vary across different lung zones, and it ranged from moderate to excellent. LUS performed on pregnant women by obstetricians after completing a brief theoretical and a practical course may be considered consistent.

ORCID: Murat Yassa: https://orcid.org/0000-0001-8661-1192; Memis Ali Mutlu: https://orcid.org/0000-0002-2756-0177; Pinar Birol: https://orcid.org/0000-0002-2194-0726; Taha Yusuf Kuzan: https://orcid.org/0000-0002-5420-8507; Erkan Kalafat: https://orcid.org/0000-0003-0658-138X; Canberk Usta: https://orcid.org/0000-0002-0825-2355; Emre Yavuz: https://orcid.org/0000-0003-3251-4906; Ilkhan Keskin: https://orcid.org/0000-0001-7839-5709; Niyazi Tug: https://orcid.org/0000-0001-7442-834X

Author Contributions

Conceptualization: Yassa M, Tug N. Data acquisition: Mutlu MA, Birol P, Kuzan TY, Yavuz E, Keskin I. Data analysis or interpretation: Kalafat E. Drafting of the manuscript: Yassa M, Kalafat E. Critical revision of the manuscript: Yassa M, Tug N. Approval of the final version of the manuscript: all authors.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Supplementary Material

Video clip 1. Normal lung ultrasonography pattern classified as score 0 with a continuous pleural line and horizontal A-lines at regular intervals due to the reverberation of the pleural line. The transducer is positioned in the intercostal space (https://doi.org/10.14366/usg.20084.v001).

Video clip 2. Discontinuity in the pleural line without significant consolidations. There are multiple discrete B-lines in some frames, but a large white area is also seen. Therefore, the lung ultrasonography findings are classified as score 3 according to the highest score obtained (https://doi.org/10.14366/usg.20084.v002).

Video clip 3. A severely broken pleural line, subpleural consolidations, and dense and largely extended white lung (https://doi.org/10.14366/usg.20084.v003).

References

1. Cibinel GA, Casoli G, Elia F, Padoan M, Pivetta E, Lupia E, et al. Diagnostic accuracy and reproducibility of pleural and lung ultrasound in discriminating cardiogenic causes of acute dyspnea in the emergency department. Intern Emerg Med 2012;7:65-70.
2. Enghard P, Rademacher S, Nee J, Hasper D, Engert U, Jorres A, et al. Simplified lung ultrasound protocol shows excellent prediction of extravascular lung water in ventilated intensive care patients. Crit Care 2015;19:36.
3. Pereda MA, Chavez MA, Hooper-Miele CC, Gilman RH, Steinhoff MC, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in children: a meta-analysis. Pediatrics 2015;135:714-

722.

4. Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, et al. Is there a role for lung ultrasound during the COVID-19 pandemic? J Ultrasound Med 2020;39:1459-1462.

5. Tsai NW, Ngai CW, Mok KL, Tsung JW. Lung ultrasound imaging in avian influenza A (H7N9) respiratory failure. Crit Ultrasound J 2014;6:6.

6. Tsung JW, Kessler DO, Shah VP. Prospective application of clinician-performed lung ultrasonography during the 2009 H1N1 influenza A pandemic: distinguishing viral from bacterial pneumonia. Crit Ultrasound J 2012;4:16.

7. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020;20:425-434.

8. Buonsenso D, Pata D, Chiaretti A. COVID-19 outbreak: less stethoscope, more ultrasound. Lancet Respir Med 2020;8:e27.

9. Buonsenso D, Raffaelli F, Tamburrini E, Biasucci DG, Salvi S, Smargiassi A, et al. Clinical role of lung ultrasound for diagnosis and monitoring of COVID-19 pneumonia in pregnant women. Ultrasound Obstet Gynecol 2020;56:106-109.

10. Kulkarni S, Down B, Jha S. Point-of-care lung ultrasound in intensive care during the COVID-19 pandemic. Clin Radiol 2020;75:710.

11. Antunez-Montes OY, Buonsenso D. Routine use of point-of-care lung ultrasound during the COVID-19 pandemic. Med Intensiva 2020 Apr 22 [Epub]. https://doi.org/10.1016/j.medin.2020.04.010.

12. Moro F, Buonsenso D, Moruzzi MC, Inchingolo R, Smargiassi A, Demi L, et al. How to perform lung ultrasound in pregnant women with suspected COVID-19. Ultrasound Obstet Gynecol 2020;55:593-598.

13. Soldati G, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, et al. Proposal for international standardization of the use of lung ultrasound for patients with COVID-19: a simple, quantitative, reproducible method. J Ultrasound Med 2020;39:1413-1419.

14. Vetrugno L, Bove T, Orso D, Barbariol F, Bassi F, Boero E, et al. Our Italian experience using lung ultrasound for identification, grading and serial follow-up of severity of lung involvement for management of patients with COVID-19. Echocardiography 2020;37:625-627.

15. Volpicelli G, Gargani L. Sonographic signs and patterns of COVID-19 pneumonia. Ultrasound J 2020;12:22.

16. Volpicelli G, Lamorte A, Villen T. What's new in lung ultrasound during the COVID-19 pandemic. Intensive Care Med 2020;46:1445-1448.

17. Inchingolo R, Smargiassi A, Mormile F, Marra R, De Carolis S, Lanzone A, et al. Look at the lung: can chest ultrasonography be useful in pregnancy? Multidiscip Respir Med 2014;9:32.

18. Yassa M, Birol P, Mutlu AM, Tekin AB, Sandal K, Tug N. Lung ultrasound can influence the clinical treatment of pregnant women

with COVID-19. J Ultrasound Med 2020 Jun 1 [Epub]. https://doi.org/10.1002/jum.15367.

19. Gargani L, Volpicelli G. How I do it: lung ultrasound. Cardiovasc Ultrasound 2014;12:25.

20. Correa M, Zimic M, Barrientos F, Barrientos R, Roman-Gonzalez A, Pajuelo MJ, et al. Automatic classification of pediatric pneumonia based on lung ultrasound pattern recognition. PLoS One 2018;13:e0206410.

21. World Federation for Ultrasound in Medicine and Biology Safety Committee, Abramowicz JS, Basseal JM. World Federation for Ultrasound in Medicine and biology position statement: how to perform a safe ultrasound examination and clean equipment in the context of COVID-19. Ultrasound Med Biol 2020;46:1821-1826.

22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

23. Krippendorff K. Content analysis: an introduction to its methodology. Beverly Hills, CA: Sage Publications, 2018.

24. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1:54-75.

25. Gajbhiye R, Modi D, Mahale S. Pregnancy outcomes, newborn complications and maternal-fetal transmission of SARS-CoV-2 in women with COVID-19: a systematic review of 441 cases. Preprint at medRxiv https://doi.org/10.1101/2020.04.11.20062356 (2020).

26. Sultan LR, Sehgal CM. A review of early experience in lung ultrasound in the diagnosis and management of COVID-19. Ultrasound Med Biol 2020;46:2530-2545.

27. McCollough CH, Schueler BA, Atwell TD, Braun NN, Regner DM, Brown DL, et al. Radiation exposure and pregnancy: when should we be concerned? Radiographics 2007;27:909-917.

28. Donders G, Bellen G, Byttebier G, Verguts L, Hinoul P, Walckiers R, et al. Individualized decreasing-dose maintenance fluconazole regimen for recurrent vulvovaginal candidiasis (ReCiDiF trial). Am J Obstet Gynecol 2008;199:613.

29. Eskandar O, Eckford SE, Watkinson T. Safety of diagnostic imaging in pregnancy. Part 1: X-ray, nuclear medicine investigations, computed tomography and contrast media. Obstetr Gynaecol 2010;12:71-78.

30. Huang Y, Wang S, Liu Y, Zhang Y, Zheng C, Zheng Y, et al. A preliminary study on the ultrasonic manifestations of peripulmonary lesions of non-critical novel coronavirus pneumonia (COVID-19). Preprint at http://dx.doi.org/10.2139/ssrn.3544750 (2020).

31. Power SP, Moloney F, Twomey M, James K, O'Connor OJ, Maher MM. Computed tomography and patient risk: Facts, perceptions and uncertainties. World J Radiol 2016;8:902-915.

32. Gomond-Le Goff C, Vivalda L, Foligno S, Loi B, Yousef N, De Luca D. Effect of different probes and expertise on the interpretation reliability of point-of-care lung ultrasound. Chest 2020;157:924-931.

33. Brusa G, Savoia M, Vergine M, Bon A, Copetti R, Cattarossi L. Neonatal lung sonography: interobserver agreement between physician interpreters with varying levels of experience. J Ultrasound Med 2015;34:1549-1554.

34. Gullett J, Donnelly JP, Sinert R, Hosek B, Fuller D, Hill H, et al. Interobserver agreement in the evaluation of B-lines using bedside ultrasound. J Crit Care 2015;30:1395-1399.

35. Buonsenso D, Moro F, Inchingolo R, Smargiassi A, Demi L, Soldati G, et al. Effectiveness of rapid lung ultrasound training program for gynecologists and obstetricians managing pregnant women with suspected COVID-19. Ultrasound Obstet Gynecol 2020;56:110-111.