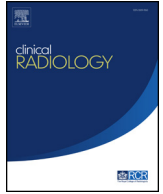




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



M-BLUE protocol for coronavirus disease-19 (COVID-19) patients: interobserver variability and correlation with disease severity

H. Xue^{a,†}, C. Li^{b,†}, L. Cui^a, C. Tian^c, S. Li^c, Z. Wang^b, C. Liu^a, Q. Ge^{b,*}

^a Department of Ultrasound, Peking University Third Hospital, Beijing, 1000191, China

^b Department of Critical Care Medicine, Peking University Third Hospital, Beijing, 1000191, China

^c Department of Emergency, Peking University Third Hospital, Beijing, 1000191, China

ARTICLE INFORMATION

Article history:

Received 21 October 2020

Accepted 4 February 2021

AIM: To retrospectively evaluate the interobserver variability of intensive care unit (ICU) practitioners and radiologists who used the M-BLUE (modified bedside lung ultrasound in emergency) protocol to assess coronavirus disease-19 (COVID-19) patients, and to determine the correlation between total M-BLUE protocol score and three different scoring systems reflecting disease severity.

MATERIALS AND METHODS: Institutional review board approval was obtained and informed consent was not required. Ninety-six lung ultrasonography (LUS) examinations were performed using the M-BLUE protocol in 79 consecutive COVID-19 patients. Two ICU practitioners and three radiologists reviewed video clips of the LUS of eight different regions in each lung retrospectively. Each observer, who was blind to the patient information, described each clip with M-BLUE terminology and assigned a corresponding score. Interobserver variability was assessed using intraclass correlation coefficient. Spearman's correlation coefficient analysis (R-value) was used to assess the correlation between the total score of the eight video clips and disease severity.

RESULTS: For different LUS signs, fair to good agreement was obtained (ICC = 0.601, 0.339, 0.334, and 0.557 for 0–3 points respectively). The overall interobserver variability was good for both the five different readers and consensus opinions (ICC = 0.618 and 0.607, respectively). There were good correlations between total LUS score and scores from three systems reflecting disease severity (R=0.394–0.660, $p<0.01$).

CONCLUSION: In conclusion, interobserver agreement for different signs and total scores in LUS is good and justifies its use in patients with COVID-19. The total scores of LUS are useful to indicate disease severity.

© 2021 The Royal College of Radiologists. Published by Elsevier Ltd. All rights reserved.

* Guarantor and correspondent: Q. Ge, Department of Critical Care Medicine, Peking University Third Hospital, Beijing, 1000191, China. Tel.: +8618101119839.

E-mail address: qingganggelin@126.com (Q. Ge).

† These authors are joint first authors.

Introduction

Since early 2020, the use of lung ultrasonography (LUS) in coronavirus disease-19 (COVID-19) has received much attention from both clinicians and radiologists as it has the advantage of identifying and classifying disease severity quickly and easily.^{1–6} Although none of the LUS features is pathognomonic for COVID-19, there has been much evidence to support its clinical value, especially in children and pregnant women.^{7–9} LUS is relatively easy to use, but in the hands of inexperienced operators, the accuracy and reproducibility might be reduced.¹⁰ In addition, interpretation of US images is dependent on the observer, which may not always provide reproducible results. The interoperator and interobserver reproducibility of LUS for its assessment of COVID-19 pulmonary involvement and disease severity should be validated before being widely used in clinical practice. Therefore, the present study was conducted with two purposes: to evaluate interobserver variability retrospectively between different intensive care unit (ICU) practitioners and radiologists who used the M-BLUE (modified bedside lung ultrasound in emergency) protocol to assess COVID-19 patients, and to determine the correlation between total M-BLUE protocol score and disease severity.

Materials and methods

Study population

This study was approved by the local ethics committee, which waived the need for a written informed consent. From 4 February 2020 to 29 March 2020, 79 consecutive patients with positive real-time polymerase chain reaction (RT-PCR) test results for COVID-19 were enrolled in this study. A total of 96 LUS were performed. All patients also underwent chest computed tomography (CT); however, no detailed correlation CT data were analysed as this was beyond the scope of the study.

Acquisition and analysis of LUS findings

The indication for LUS was not to screen or diagnose COVID-19, but to evaluate disease severity, and monitor disease progression/regression. In compliance with the highest level of personal protective equipment as per the World Federation for Ultrasound in Medicine and Biology (WFUMB),¹¹ all the bedside LUS examinations were performed by one ICU practitioner who had 5 years of experience in LUS, with a convex probe (M9 with C5-1s transducer, M7 with C5-2s transducer Mindray, Shenzhen, China). According to the M-BLUE protocol, eight areas (bilateral superior BLUE point, M point, PLAPS point, diaphragm point; Fig 1) were scanned per patient. Ten-second video clips instead of static images were saved to the hard disk for later analysis, as tiny alterations of LUS may not appear on every frame. A semi-quantitative scoring system was employed with the following rules: 0 point: A-lines or

less than two B-lines; 1 point: three or more separated B-lines; 2 points: confluent B-lines; 3 points: consolidation or atelectasis. Therefore, a total score for the eight regions of 0 is normal, and 24 would be the worst. The definitions of these LUS signs have been well described in previous studies.^{1,2,4,12,13} In the normal aerated lung, there is a thin, smooth hyperechoic line called the pleural line, and posterior horizontal echogenic lines called A lines (Fig 2a). Different from the horizontal A lines, B lines are vertical echogenic reverberation artefacts extending from the lung surface without attenuation (Fig 2b). Confluent B lines result in the “waterfall sign” (Fig 2c). B lines are caused by reverberation of the ultrasound beam between the slightly decreased alveolar air and increased interstitial fluids. Consolidation is visualised on LUS as a tissue-like hypoechoic region (Fig 2d), which reflects the process of highly reduced air and increased inflammatory cellular exudate.

Two ICU practitioners (with 5 and 3 years of experience of LUS) and three radiologists (with 8, 4, and 15 years of experience of LUS) reviewed the 768 video clips from 96 LUS examinations independently. To minimise bias in the scoring of the LUS video clips, readers were blinded to the clinical information during reading. After independent review, consensus scoring for each patient by the ICU practitioners and radiologists were obtained after group discussion.

Assessment of disease severity

The assessment of disease severity for each patient was based on three different scoring systems: APACHE II (acute physiology and chronic health evaluation II),^{14,15} CURB65 pneumonia severity score¹⁶ and qSOFA (quick sequential organ failure assessment).^{17,18} The three systems use point scores based upon values of age, previous health status, physiological measurements and laboratory-based prognostic markers to provide a general reflection of disease severity. A higher score (range 0–71 for APACHE II, range 0–5 for CURB65, and range 0–3 for qSOFA) indicates increased disease severity, and is closely correlated with the risk of poor prognosis.^{15,18,19} The time interval between LUS and assessment was <12 h.

Statistical analysis

Patient age and total scores for each patient rated by five readers were expressed as mean $\pm$ standard deviation, and all categorical variables were expressed as counts and percentages. Interobserver agreement for choosing LUS signs and total LUS score for each patient was analysed using the intraclass correlation coefficient (ICC). ICC was classified as poor (0–0.20), fair (0.20–0.40), good (0.40–0.75), or excellent (0.75).²⁰ For different LUS signs, data were pooled from all five readers to obtain overall percentages. Spearman's correlation coefficient analysis (R-value) was used to assess the correlation between the total score of eight video clips and disease severity. Correlation was considered high when the R-value was >0.6, as moderate when the R-value was between 0.4–0.6, or as slight when the R-value was <0.4. Two-sided $p < 0.05$ was considered statistically

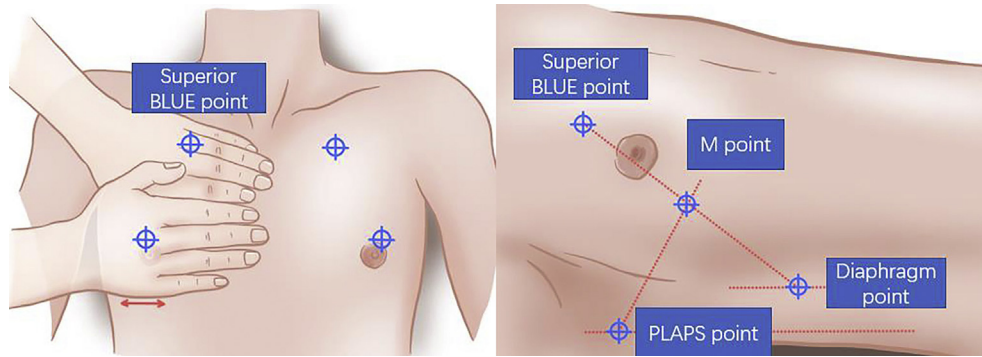


Figure 1 Standardised points used in the M-BLUE protocol. The patient was placed in the supine position during LUS examinations. Two hands (with approximately the patient's size) are applied as follows: little finger of the left hand just below the right clavicle, fingertips at middle line, and the right hand (excluding the thumb) just below the left hand. The superior BLUE point is at the middle of the left hand. The diaphragm point is built from the lung–liver or lung–spleen junction at mid-axillary line, while M point is at the midpoint between superior BLUE point and diaphragm point. PLAPS (posterolateral alveolar and/or pleural syndrome) point is the intersection of posterior axillary line and the vertical line from M point.

significant. Confidence intervals (CI) were reported at the 95% level. All statistics were calculated using SPSS software (version 25.0, IBM, New York, NY, USA).

Results

Participant characteristics

The study population comprised 79 consecutive patients (40 male and 39 female, mean age 61.1 ± 13.7 years, range

24–88 years). Two patients had four repeated LUS examinations, four patients had three LUS examinations, three patients had two LUS examinations, and the other 70 patients had one LUS examination. Therefore, a total of 96 LUS were performed.

Interobserver variability for LUS

The five readers assessing 768 video clips from 96 LUS produced a total of 3,840 “counts”. The range of total scores for each patient rated by five readers were: 7.8 ± 5.5 (0–20),

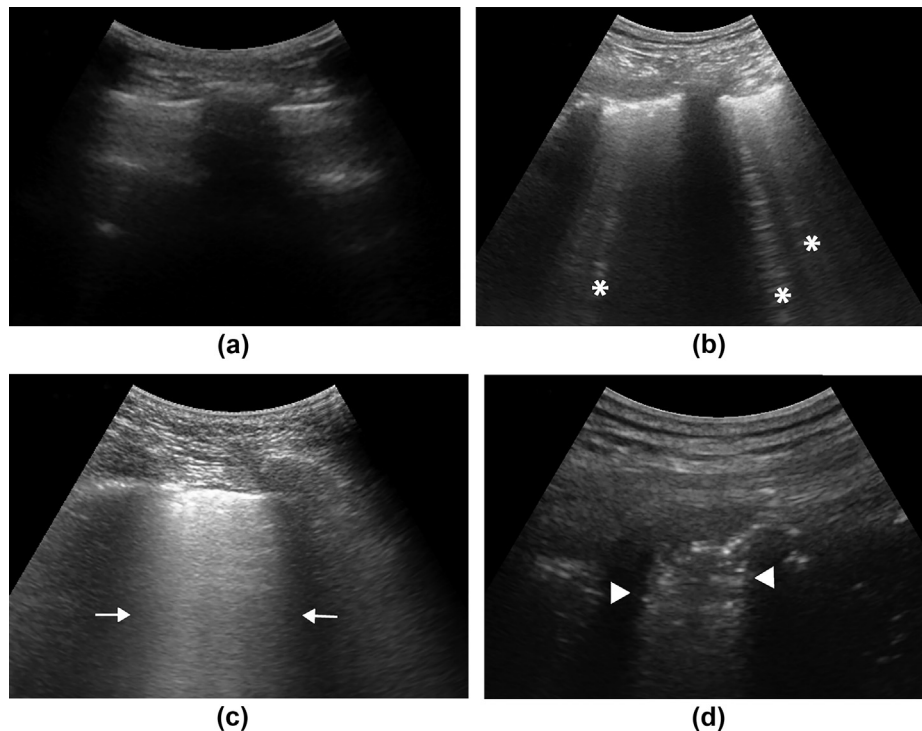


Figure 2 Images of different LUS scores in patients with COVID-19. (a) Left M point in a 24-year-old female patient. Zero points were given by five readers for normal LUS findings. (b) Right superior BLUE point in a 68-year-old male patient. One point was given for three of more separated B-lines (white asterisks). (c) A 62-year-old male patient. Confluent B-lines in left superior BLUE point yielded 2 points (between the white arrows). (d) Right diaphragm point in a 71-year-old female patient. Three points were given for hypoechoic subpleural consolidation (between white arrowheads).

Table 1

Percentage of lung ultrasonography (LUS) with each score and interobserver agreement.

LUS score	Counts and percentage	ICC value ^a
0 point	1,389 (36.2%)	0.601 (0.571–0.632)
1 point	783 (20.4%)	0.339 (0.305–0.375)
2 points	1,331 (34.7%)	0.334 (0.298–0.371)
3 points	337 (8.8%)	0.557 (0.525–0.589)
Overall	3,840 (100.0%)	0.618 (0.588–0.647)

^a Data in parentheses are 95% CIs.

10.8 ± 5.6 (0–20), 11.8 ± 6.0 (0–24), 12.4 ± 5.9 (0–21), 8.6 ± 7.0 (0–20); data in parentheses are ranges). **Table 1** shows the percentages of LUS examinations with each score and interobserver agreement. The overall percentages for different LUS scores were 36.2% for 0 points, 20.4% for 1 point, 34.7% for 2 points and 8.8% for 3 points, respectively. In describing different LUS signs, fair agreement was seen when 1 or 2 points were given (**Fig 2**; ICC, 0.339, 95% CI 0.305–0.375; ICC, 0.334, 95% CI 0.298–0.371) and good agreement was seen when LUS score was given as 0 points (ICC, 0.601, 95% CI 0.571–0.632) or 3 points (**Fig 2**; ICC, 0.557, 95% CI 0.525–0.589). The overall interobserver reliability for different LUS scores was good for both five different readers (ICC, 0.618, 95% CI 0.588–0.647) and consensus opinions among ICU practitioners and radiologists (ICC, 0.607, 95% CI 0.560–0.650).

The interobserver agreement of the total score for LUS was excellent for both the five different readers (ICC, 0.753, 95% CI 0.687–0.813) and the two groups (ICC, 0.753, 95% CI 0.649–0.827).

Correlation between LUS score and disease severity

Statistically significant correlation between total LUS score and three different systems reflecting disease severity was observed for all the five readers and group opinion ($p < 0.001$; for R-values, see **Table 2**). R-value between total LUS score and APACHE II was higher than two other scoring systems. Group opinions from ICU practitioners had slightly higher R-values than those from radiologists for all three systems. Interestingly, group discussion would not always yield higher R-values for both ICU practitioners and radiologists in all three systems.

Statistically significant correlation was also observed between the three scoring systems reflecting disease severity ($R = 0.818$ for APACHE II and CURB65; $R = 0.587$ for APACHE II and qSOFA; $R = 0.553$ for CURB65 and qSOFA).

Discussion

With the development and utilisation of LUS in the past decades, its application for triage and assessment of various lung diseases has been studied and promoted widely.^{1–4,10,12,18,21,22} COVID-19, with high contagiousness, rapid worldwide spread, and more severe clinical manifestations compared with common influenza, has resulted in worldwide healthcare crises. Although chest CT is the routine imaging method for early diagnosis and monitoring of the disease, LUS, with its advantages of repeatability, low cost, and point-of-care, may play a complimentary role in the work-up of COVID-19. Compared to chest radiography and CT, LUS does not require patients to be transported to rooms housing equipment, thus minimising the number of healthcare workers and medical devices exposed to COVID-19, which is important to avoid nosocomial outbreaks of the virus. In the setting of COVID-19, LUS can be used to detect not only signs of pulmonary involvement, but also disease progression or regression; however, the obvious disadvantage of LUS is operator dependency, and there is doubt regarding the interobserver variability, and whether total LUS scores could correlate with disease severity. These two questions remain to be investigated and clarified adequately.

The present results showed fair to good agreement in describing lesions on LUS, thus demonstrating the appropriateness of the terms chosen in LUS. The terminology was well accepted and familiar to both ICU practitioners and radiologists who perform LUS. Agreement for 0 point and 3 points on LUS was higher than 1 point and 2 points, suggesting easier decision making for normal and pulmonary consolidation or atelectasis, but more difficulty in distinguishing confluent B-lines. As the total score was the sum of eight video clips, different results for one or two video clips would not significantly affect the overall impression on LUS; thus, excellent agreement was achieved in the total score for both the five different readers and the two groups.

Several studies have demonstrated the usefulness of chest CT in evaluating the disease severity in patients with COVID-19.^{23,24} Similar to the observation on chest CT, the present study shows that total scores in LUS had good correlations with scores from APACHE II, CURB65, and qSOFA in all five different readers, with highest R-values in APACHE II. APACHE II is a well-accepted scoring system that provide accurate description of disease severity and prognosis for patients in ICU (15, 19). An increased LUS score indicated decreased lung aeration, and vice versa. The high correlation (R -value: 0.506–0.660) between total score of

Table 2

Spearman correlation coefficient analysis (R -value) for total lung ultrasonography (LUS) score and three different systems reflecting disease severity.

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	ICU ^a	Radiology ^a
APACHE II	0.660	0.611	0.506	0.547	0.621	0.652	0.587
CURB65	0.598	0.588	0.508	0.526	0.559	0.584	0.526
qSOFA	0.582	0.573	0.394	0.404	0.590	0.574	0.467

^a ICU and Radiology indicate consensus opinions for ICU practitioners and radiologists. Readers 1 and 2: ICU practitioners. Readers 3, 4 and 5: radiologists.

LUS and APACHE II justified the use of serial LUS in monitoring the effect of antiviral and supportive therapies. In a recent study by Zhao *et al.*,²⁵ similar LUS scores were used in diagnosing refractory respiratory failure (PaO₂/FiO₂ 100 mmHg or on extracorporeal membrane oxygenation) among 35 patients with COVID-19. In another recent study,²⁶ an inverse relationship between PaO₂/FiO₂, the aeration score, and the number of subpleural consolidations observed by 12-zone LUS was found. Compared to their study, dynamic video clips but not static images were evaluated, which was closer to clinical practice, and more comprehensive scoring systems reflecting disease severity were used as the reference standard in the present study. The high correlation with APACHE II, CURB65, and qSOFA guaranteed the use of LUS in guiding clinical decisions, as previously reported by Xirouchaki *et al.*,²⁷ and potentially reduce the need for chest radiography and CT. As APACHE II could predict the prognosis for patients in ICU, it is plausible to envision that LUS could also provide objective identification for patients with poor prognosis.

The present study had some limitations. First, all the LUS cases had been scanned and evaluated by Reader 1, who was in charge of the patients just 2 months prior to the study. Knowledge of their clinical information may have influenced the scanning and scoring of the LUS video clips. This may also explain why Reader 1 exhibited the highest R-value among all the readers. Second, whether different operators would affect the reproducibility of LUS on assessment of COVID-19 pulmonary involvement has not been assessed, because it is not ethical to expose two operators to the risk of becoming infected. Third, this study was based on the performance of experienced ICU practitioners and radiologists; therefore, there may be inconsistency in evaluating the LUS video clips with different level of expertise. Fourth, high-frequency linear probes were not used. LUS images with higher resolution may increase diagnostic confidence and reduce interobserver variability.

In conclusion, interobserver agreement for different signs and total scores using LUS is good and justifies its use in patients with COVID-19. Total LUS scores are useful to indicate disease severity, potentially reducing the need for chest radiography and CT, which would increase the efficiency of management of patients with COVID-19.

Conflict of interest

The authors declare no conflict of interest.

References

1. Xue H, Zhang Y, Cui L, *et al.* Lung ultrasonography in diagnosis and management of novel coronavirus (COVID-19) pneumonia: pearls and pitfalls. *Adv Ultrasound Diagn* 2020;**4**(2):57–9.
2. Zhang Y, Xue H, Wang M, *et al.* Lung ultrasound findings in patients with coronavirus disease (COVID-19). *AJR Am J Roentgenol* 2020;**216**(1):80–4.
3. Piscaglia F, Stefanini F, Cantisani V, *et al.* Benefits, open questions and challenges of the use of ultrasound in the COVID-19 pandemic era. The

- views of a panel of worldwide international experts. *Ultraschall Med* 2020;**41**(3):228–36.
4. Smith MJ, Hayward SA, Innes SM, *et al.* Point-of-care lung ultrasound in patients with COVID-19 — a narrative review. *Anaesthesia* 2020;**75**(8):1096–104.
5. Ghosh S, Deshwal H, Saeed MB, *et al.* Imaging algorithm for COVID-19: a practical approach. *Clin Imag* 2020;**72**:22–30.
6. Lichter Y, Topilsky Y, Taieb P, *et al.* Lung ultrasound predicts clinical course and outcomes in COVID-19 patients. *Intensive Care Med* 2020;**46**(10):1873–83.
7. Buonsenso D, Raffaelli F, Tamburrini E, *et al.* Clinical role of lung ultrasound for the diagnosis and monitoring of COVID-19 pneumonia in pregnant women. *Ultrasound Obstet Gynecol* 2020;**56**(1):106–9.
8. Denina M, Scolaro C, Silvestro E, *et al.* Lung ultrasound in children with COVID-19. *Pediatrics* 2020;**146**(1):e20201157.
9. Kalafat E, Yaprak E, Cinar G, *et al.* Lung ultrasound and computed tomographic findings in pregnant woman with COVID-19. *Ultrasound Obstet Gynecol* 2020;**55**(6):835–7.
10. 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**(3):398–401.
11. World Federation for Ultrasound in MBiology Safety C, Abramowicz JS, *et al.* 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**(7):1821–6.
12. Volpicelli G. Lung sonography. *J Ultrasound Med* 2013;**32**(1):165–71.
13. Peng QY, Wang XT, Zhang LN. Chinese Critical Care Ultrasound Study G. Findings of lung ultrasonography of novel corona virus pneumonia during the 2019-2020 epidemic. *Intensive Care Med* 2020;**46**(5):849–50.
14. Knaus WA, Draper EA, Wagner DP, *et al.* APACHE II: a severity of disease classification system. *Crit Care Med* 1985;**13**(10):818–29.
15. Gursel G, Demirtas SJR. Value of APACHE II, SOFA and CPIS scores in predicting prognosis in patients with ventilator-associated pneumonia. *Respiration* 2006;**73**(4):503–8.
16. Barlow G, Nathwani D, Davey P. The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia. *Thorax* 2007;**62**(3):253–9.
17. Marik PE, Taeb AM. SIRS, qSOFA and new sepsis definition. *J Thorac Dis* 2017;**9**(4):943–5.
18. Frost F, Bradley P, Tharmaratnam K, *et al.* The utility of established prognostic scores in COVID-19 hospital admissions: a multicentre prospective evaluation of CURB-65, NEWS2, and qSOFA. *BMJ Open Respir Res* 2020;**7**(1):e000729.
19. Naved SA, Siddiqui S, Khan FHJ, *et al.* APACHE-II score correlation with mortality and length of stay in an intensive care unit. *J Coll Physicians Surgeons Pakistan* 2011;**21**(1):4.
20. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;**19**(1):231–40.
21. Hankins A, Bang H, Walsh P. Point of care lung ultrasound is useful when screening for CoVid-19 in Emergency Department patients. *medRxiv* 2020.
22. Volpicelli G, Lamorte A, Villen T. What's new in lung ultrasound during the COVID-19 pandemic. *Intensive Care Med* 2020;**46**(7):1445–8.
23. Zhao W, Zhong Z, Xie X, *et al.* Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. *Am J Roentgenology* 2020;**214**(5):1072–7.
24. Yang R, Li X, Liu H, *et al.* Chest CT severity score: an imaging tool for assessing severe COVID-19. *Radiology* 2020;**2**(2):e200047.
25. Zhao L, Yu K, Zhao Q, *et al.* Lung ultrasound score in evaluating the severity of coronavirus disease 2019 (COVID-19) pneumonia. *Ultrasound Med Biol* 2020;**46**(11):2938–44.
26. Bitar ZI, Shamsah M, Maadarani O, *et al.* Lung ultrasound and sonographic subpleural consolidation in COVID-19 pneumonia correlate with disease severity. *Crit Care Res Pract* 2021.
27. Xirouchaki N, Kondili E, Priniakakis G, *et al.* Impact of lung ultrasound on clinical decision making in critically ill patients. *Intens Care Med* 2014;**40**(1):57–65.