Interobserver Agreement in Semi-Quantitative Scale-Based Interpretation of Chest Radiographs in COVID-19 Patients

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1,2 **Bartosz Mruk**
ACD 1,2 **Jerzy Walecki**
BCF 2 **Piotr Gustaw Wasilewski**
CD 1 **Łukasz Paluch**
ABDFG 1,2 **Katarzyna Sklinda**

1 Department of Radiology, Medical Centre for Postgraduate Education, Warsaw, Poland
2 Department of Diagnostic Radiology, Central Clinical Hospital of The Ministry of The Interior in Warsaw, Warsaw, Poland

Corresponding Author: Bartosz Mruk, e-mail: bartosz.mruk@icloud.com
Source of support: Departmental sources

Background: The chest X-ray is the most available imaging modality enabling semi-quantitative evaluation of pulmonary involvement. Parametric evaluation of chest radiographs in patients with SARS-CoV-2 infection is crucial for triage and therapeutic management. The CXR Score (Brixia Score), SARI CXR Severity Scoring System, and Radiographic Assessment of Lung Edema (RALE), proposed to evaluate SARS-CoV-2 infiltration of the lungs, were analyzed for interobserver agreement.

Material/Methods: This study analyzed 200 chest X-rays from 200 consecutive patients with confirmed SARS-CoV-2 infection, hospitalized at the Central Clinical Hospital of the Ministry of the Interior and Administration in Warsaw. Radiographs were evaluated by 2 radiologists according to 3 scales: SARI, RALE, and CXR Score.

Results: The overall interobserver agreement for SARI ratings was good (κ=0.755; 95% CI, 0.817-0.694), for RALE scale assessments it was very good (κ=0.818; 95% CI, 0.844-0.793), and for CXR scale assessments it was very good (κ=0.844; 95% CI, 0.846-0.841). A moderate correlation was found between the radiological image assessed using each of the scales and the clinical condition of the patient in MEWS (Modified Early Warning Score) (r=0.425-0.591).

Conclusions: The analyzed scales are characterized by good or very good interobserver agreement of assessments of the extent of pulmonary infiltration. Since the CXR Score showed the strongest correlation with the clinical condition of the patient as expressed using the MEWS scale, it is the preferred scale for chest radiograph assessment of patients with COVID-19 in the light of data provided.

Keywords: COVID-19 • Diagnostic Imaging • Radiography

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/931277

📄 2274   ▦ 7   📊 3   📖 20

Background

Besides computed tomography scans, chest radiographs (CXR) are the primary method for the assessment of the extent of pulmonary lesions in the course of SARS-CoV-2 infection [1-10]. Despite its lower sensitivity in the detection of pulmonary lesions compared to chest CT, radiography is the preferred diagnostic modality in multiple sites owing to its availability [3,10,11]. Toussie et al demonstrated the usefulness of chest radiographs acquired at a hospital emergency department as predictors of hospitalization and intubation of patients with COVID-19 [1]. Previous work involving patients examined during the acute respiratory syndrome (SARS) coronavirus outbreak in 2003 as well as patients with other pneumonias confirmed the relationship between the extent of pulmonary infiltrates and prognosis [12-14].

To determine the appropriate clinical management and respiratory support for COVID-19 patients, it is essential to quantitatively assess the extent of pulmonary infiltrates. There is no standardized and acknowledged scale that would be considered a criterion standard for reporting and interpretation of chest X-ray results in COVID-19 patients. At least 3 different scales have been described in the literature to evaluate chest radiographs of patients with COVID-19. The SARI CXR Severity Scoring System and RALE Classification have been proposed prior to the outbreak of COVID-19 and the CXR Score was designed specifically for evaluation of patients with confirmed SARS-CoV-2 infection [11,15,16].

The SARI CXR Severity Scoring System was proposed in the pre-COVID era, with an aim to simplify the clinical grading of CXR reports from inpatients with confirmed acute respiratory infection into 5 severity categories [15]. The CXR findings were categorized as: 1 – normal; 2 – hyperinflation and/or patchy atelectasis and/or bronchial wall thickening; 3 – focal consolidation; 4 – multifocal consolidation; and 5 - diffuse alveolar changes (**Figure 1**). Soon Ho Yoon et al used this scoring system to quantify the pulmonary involvement in patients with COVID-19 [4].

The Radiographic Assessment of Lung Edema (RALE) score as proposed by Warren et al was simplified by Wong et al and used in the assessment of COVID-19 patients [10,16]. This scale assessed each lung individually. The score of 0 to 4 points was assigned based on the extent of involvement, ie, ground-glass opacity or consolidation (0 – no involvement; 1 – less than 25%; 2-25% to 50%; 3-50% to 75%; 4 – more than 75% involvement), with the overall score being the total of points from both lungs (**Figure 2**).

To date, the CXR Score (Brixia Score) is the only available method for CXR assessment that has been designed specifically for patients with confirmed COVID-19 [11]. This CXR scoring system, as proposed by Andrea Borghesi and Roberto Maroldi, is comprised of 2 steps of imaging analysis [11]. The first step is to divide each lung as seen in frontal chest projection (posteroanterior – PA or anteroposterior AP view) into 3 zones designated with letters A, B, and C for the right lung and D, E, and F for the left lung. The letters divide lungs into 3 levels: the upper level (A and D) above the inferior wall of the aortic arch, the middle level (B and E) below the inferior wall of the aortic arch and above the inferior wall of the right inferior pulmonary vein (the hilar structures), and the lower level (C and F) below the inferior wall of the right inferior pulmonary vein (the lung bases) (**Figure 3**).

The purpose of this study was to analyze the interobserver agreement of chest radiographs obtained from patients with COVID-19 as assessed using the 3 scales described above by the same group of 2 independent radiologists as well as to establish correlations between the radiological image and the clinical condition of the patient as expressed using the Modified Early Warning Score (MEWS), which includes measurements of systolic blood pressure, heart rate, respiratory rate, body temperature, and level of consciousness (**Table 1**) [17].

Material and Methods

A total of 200 chest X-ray examinations collected from 200 consecutive patients hospitalized due to SARS-CoV-2 infection at the Central Clinical Hospital of the Ministry of the Interior and Administration in Warsaw were analyzed retrospectively in the study. Each patient admitted to the hospital had to have a positive PCR test result confirmed twice. All the patients' data were fully anonymized before they were accessed. Within the analyzed group there were 109 men and 91 women. The mean age was 62.6 (range 19-90 years old).

The study was approved by the Bioethics Committee of the Central Clinical Hospital of the Ministry of the Interior and Administration in Warsaw.

Radiographs were acquired using 2 Siemens Multix Pro stationary units and 1 Shimadzu Mobile Dart Evolution MX8 portable device, using a standardized technique (80 kV, 10 mAs, 180-cm film-focus distance for posteroanterior; 80 kV, 10 mAs, 100-cm film-focus distance for anteroposterior). There were 128 posteroanterior and 72 anteroposterior radiographs.

CXRs were independently assessed by 2 radiologists with 7 years of experience (B.M.) and 16 years of experience (K.S). Radiologists were aware of the positive results of RT-PCR tests for the presence of SARS-CoV-2 while having no access to the results of other laboratory tests, clinical data, and previous imaging scans. CXRs were interpreted using diagnostic workstations running OsiriX MD v.8.0.2 software.
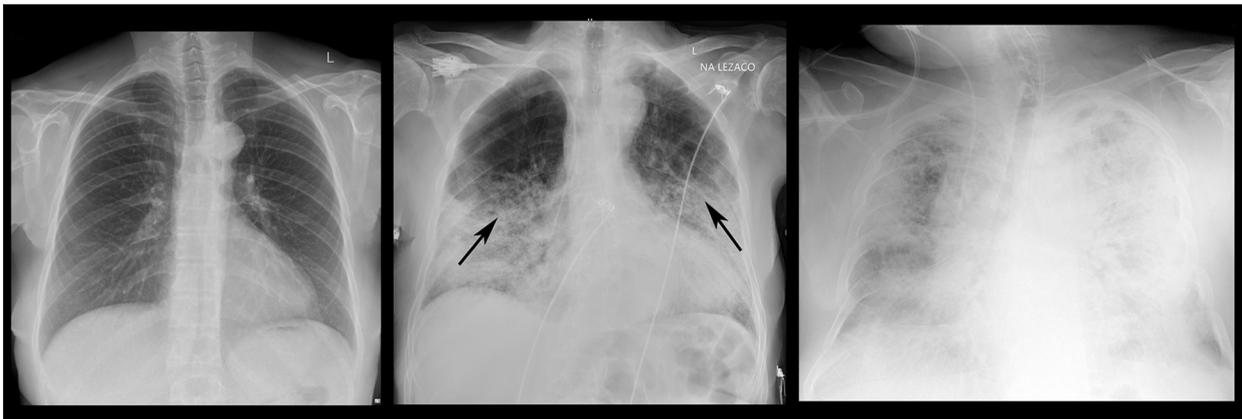
Figure 1. Chest X-ray images of 3 COVID-19-positive patients with different intensity of lung involvement assessed with SARI scoring system. In the **left** picture both lungs present no radiological signs of parenchymal involvement and were assessed as 1 with SARI scoring system; in the **middle** picture multifocal consolidations can be spotted and the image was assessed as 4 with SARI scoring system; in the **right** image nearly entire parenchyma of both lungs present diffuse alveolar changes and the image was assessed as 5 with SARI scoring system.
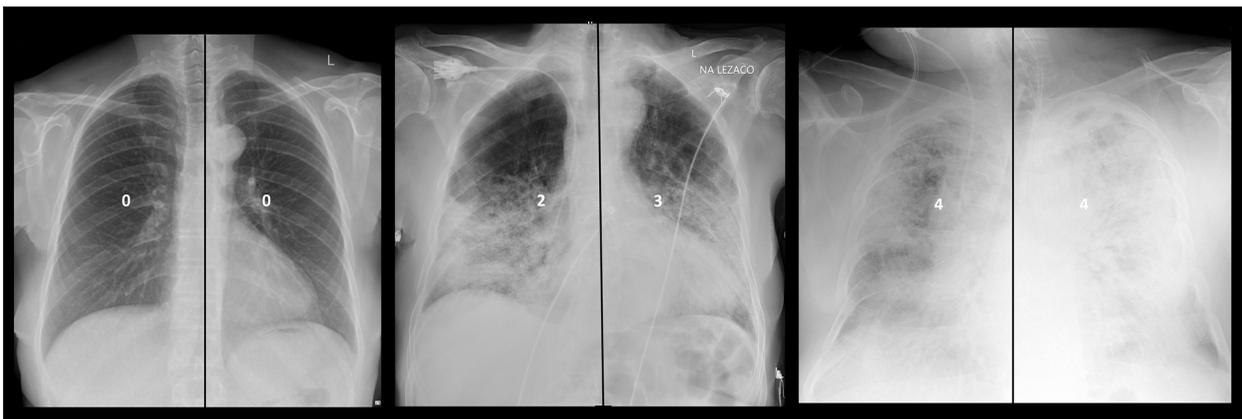

Figure 2. Chest X-ray images of 3 COVID-19-positive patients with different intensity of lung involvement assessed with RALE classification. In the **left** picture both lungs present no involvement and the overall score was assessed as 0; in the **middle** picture the right lung involvement is assessed as 25-50% and the left lung as 50-75%, the overall RALE score was assessed as 5; in the **right** image lungs are involved in nearly 100%, the overall RALE score was assessed as 8.
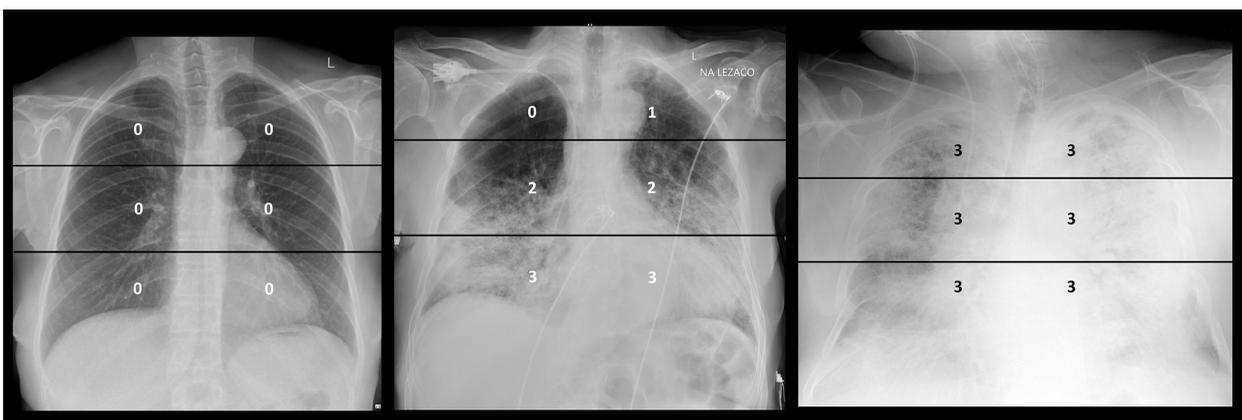

Figure 3. Chest X-ray images of 3 COVID-19-positive patients with different intensity of lung involvement assessed with CXR scoring system. In the **left** picture the image of lungs was assessed by CXR Sore at 0 points; in the **middle** picture the image of lungs was assessed by CXR Sore at 11 points; in the **right** picture the image of lungs was assessed by CXR Score at 18 points.

Table 1. The Modified Early Warning Score (MEWS).

Score	3	2	1	0	1	2	3
Systolic blood pressure (mmHg)	<70	71-80	81-100	101-199		≥200	
Heart rate (beats per minute)		<40	41-50	51-100	101-110	11-129	≥130
Respiratory rate (breaths per minute)		<9		9-14	15-20	21-29	≥30
Temperature (°C)		<35		35-38,4		>38,5	
Level of consciousness				Alert	Reacting to voice	Reacting to pain	Unresponsive

Radiographs were evaluated according to 3 scales: SARI in the range of 1-5 points; RALE in the range of 1-4 points for each of the 2 lungs (range 1-8 for both lungs); and CXR Score in the range of 1-3 points for each of the 6 anatomical regions of the lungs (range 1-18 for both lungs).

All patients whose images were included in the analysis had their clinical condition assessed using MEWS scale (on the day of the CXR). For the purposes of statistical analyses, patients were divided into 3 groups: Group A (MEWS score 0-1; 96 patients), Group B (MEWS score 2-3; 53 patients), and Group C (MEWS score ≥4; 51 patients).

To assess the interobserver agreement of CXR interpretation between 2 radiologists, Cohen's κ was calculated. Since the results were presented on ordinal scales, weighted Cohen's κ was used for the interobserver agreement analysis. The weights were selected using the Fleiss-Cohen method [18]. The inter-class correlation coefficient (ICC) was also calculated for the CXR scale. The weighted κ values were interpreted according to McHugh, while ICCs were interpreted according to Koo and Li [19,20]. Agreement was defined as moderate (κ >0.4-0.6), good (κ >0.6-0.8) and very good (κ >0.8-1.0). Spearman's linear correlation coefficient was used to analyze the correlation between the extent of inflammatory lesions and the clinical condition of the patient. The correlation coefficient was defined as low (r=0-0.3), moderate (r=0.3-0.5), strong (r=0.5-0.7), or very strong (r=0.7-1).

For the SARI scale, a general population and a group-by-group interobserver agreement analysis were performed depending on the type of exam (PA vs AP) and the patient's clinical condition as expressed using MEWS on the day of the exam: Group A (MEWS 0-1), Group B (MEWS 2-3), and Group C (MEWS ≥4).

For the RALE scale, a general population, the left and the right lung and a group-by-group interobserver agreement analysis were performed depending on the type of exam (PA vs AP) and the patient's clinical condition as expressed using MEWS on the day of the exam: Group A (MEWS 0-1), Group B (MEWS 2-3), and Group C (MEWS ≥4).

For the CXR Score scale, interobserver agreement analysis was performed for a general population, for 6 individual anatomical lung regions and a group-by-group analysis depending on the type of exam (PA vs AP) and the patient's clinical condition as expressed using MEWS on the day of the exam: Group A (MEWS 0-1), Group B (MEWS 2-3), and Group C (MEWS ≥4).

Results

SARI Scale

The overall interobserver agreement of SARI ratings was good (κ=0.755; 95% CI, 0.817-0.694). With regard to the group-by-group analyses carried out in patients with different MEWS scores, the highest interobserver agreement was observed in patients with mild disease (MEWS 0-1 points): κ=0.791; 95% CI, 0.835-0.746. The lowest interobserver agreement was observed in the group of patients with MEWS in the range of 2-3 points (κ=0.574; 95% CI, 0.849-0.349). In the group of patients with the most severe clinical course (MEWS ≥4), the kappa value was 0.681 (95% CI, 0.828-0.533). Significant differences were noted in the interobserver agreement of the radiographic assessments depending on the type of examinations. The interobserver agreement of the assessments of AP radiographs was lower (κ=0.624; 95% CI, 0.874-0.475) than the assessments of PA examinations (κ=0.819; 95% CI, 0.892-0.789) (**Table 2**).

RALE Scale

The overall interobserver agreement of RALE scale assessments was very good (κ=0.818; 95% CI, 0.844-0.793). With regard to the group-by-group analyses carried out in patients with different MEWS ratings, the highest interobserver agreement was

Table 2. Analysis of the interobserver agreement of SARI assessments of radiographs. The table presents the weighted κ values for the overall population as well as for individual exam types (PA vs AP) and the patient's clinical condition as expressed using MEWS: Group A (MEWS 0-1), Group B (MEWS 2-3), and Group C (MEWS ≥4).

		SARI					
		Exam type		Modified Early Warning Score			Total
		AP X-ray	PA X-ray	0-1	2-3	≥4	
κ		0.624	0.819	0.791	0.574	0.681	0.755
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.874	0.892	0.835	0.849	0.828	0.817
	LL	0.475	0.789	0.746	0.349	0.533	0.694

Table 3. Analysis of the interobserver agreement of RALE assessments of radiographs. The table presents the weighted κ values for the overall population as well as for individual exam types (PA vs AP) and the patient's clinical condition as expressed using MEWS: Group A (MEWS 0-1), Group B (MEWS 2-3), and Group C (MEWS ≥4).

		SARI					
		Exam type		Modified Early Warning Score			Total
		AP X-ray	PA X-ray	0-1	2-3	≥4	
κ		0.796	0.825	0.840	0.799	0.807	0.818
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.812	0.841	0.846	0.822	0.849	0.844
	LL	0.778	0.783	0.833	0.758	0.765	0.793

Table 4. Analysis of the interobserver agreement of RALE assessments of radiographs for the right and the left lung.

		RALE	
		LL	RL
κ		0.801	0.837
p		<0.001	<0.001
95% CI	UL	0.817	0.848
	LL	0.785	0.826

observed in patients with mild disease (MEWS 0-1 pt): (κ=0.840; 95% CI, 0.846-0.833). The lowest interobserver agreement was observed in the group of patients with MEWS score in the range of 2-3 points (κ=0.799; 95% CI, 0.822-0.758). In the group of patients with the most severe clinical course (MEWS ≥4), the kappa value was (κ=0.807; 95% CI, 0.849-0.865). The interobserver agreement of the assessments of AP radiographs was lower (κ=0.796; 95% CI, 0.812-0.778) than the assessments of PA examinations (κ=0.825; 95% CI, 0.841-0.783). The κ values were similar for both lungs and were indicative of nearly perfect interobserver agreement (**Tables 3, 4**).

CXR Scale

The overall interobserver agreement of CXR scale assessments was very good (κ=0.844; 95% CI, 0.846-0.841). With regard to the group-by-group analyses carried out in patients with different MEWS ratings, the highest interobserver agreement was observed in patients with mild disease (MEWS 0-1): κ=0.846 (95% CI, 0.849-0.843). The worst interobserver agreement was observed for patients with the most severe clinical course (MEWS ≥4): κ=0.724; 95% CI, 0.792-0.676. In the group of patients with MEWS of 2-3, the kappa weighted value was 0.747; 95% CI, 0.8-0.695. The interobserver agreement of the

Table 5. Analysis of the interobserver agreement of CXR assessments of radiographs. The table presents the weighted κ and ICC values for the overall population as well as for individual exam types (PA vs AP) and the patient's clinical condition as expressed using MEWS scale: Group A (MEWS 0-1), Group B (MEWS 2-3), and Group C (MEWS ≥4).

		CXR					
		Exam type		Modified Early Warning Score			Total
		AP X-ray	PA X-ray	0-1	2-3	≥4	
κ		0.796	0.846	0.846	0.747	0.724	0.844
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.817	0.849	0.849	0.8	0.792	0.846
	LL	0.775	0.844	0.843	0.695	0.676	0.841
ICC (3,k)		0.844	0.848	0.848	0.846	0.824	0.847
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.847	0.899	0.849	0.849	0.839	0.848
	LL	0.839	0.846	0.847	0.741	0.785	0.816

Table 6. Analysis of the interobserver agreement of CXR assessments of radiographs within 6 anatomical lung regions.

		CXR					
		LLF	MLF	ULF	LRF	MRF	URF
κ		0.809	0.823	0.817	0.828	0.817	0.828
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.836	0.847	0.849	0.874	0.844	0.849
	LL	0.783	0.799	0.786	0.806	0.792	0.806
ICC (3,k)		0.829	0.836	0.834	0.839	0.834	0.839
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.836	0.830	0.839	0.843	0.839	0.843
	LL	0.819	0.841	0.826	0.834	0.826	0.833

LLF – lower left field; MLF – middle left field; ULF – upper left field; LRF – lower right field; MRF – middle right field; URF – upper right field.

assessments of AP radiographs was lower (κ=0.796; 95% CI, 0.817-0.775) than the agreement of the assessments of PA examinations (κ=0.846; 95% CI, 0.849-0.844) (**Tables 5, 6**).

Correlation Between the Radiological Image and the Clinical Condition of the Patient as Expressed Using MEWS

There was a moderate correlation between the clinical condition of the patient as expressed using MEWS and the radiological image as assessed using each of the scales (r=0.425-0.591) (**Table 7**). According to both radiologists, the strongest correlation was observed for the CXR scale (r=0.577 and 0.591) and the weakest correlation was observed for the RALE scale (r= 0.425 and 0.462).

Discussion

The analysis confirmed good and very good interobserver agreement of assessments for CXRs evaluated using each of the 3 scales. Scores obtained using CXR Score scales are comparable to these presented by Borghesi et al (κ=0.82; 95% CI, 0.79-0.86) [11].

Although no validation of the SARI and RALE scales was performed in a COVID-19 patient group, the agreement of the 2 radiologists of the scale as assessed on the basis of pulmonary infiltrates of other etiology is within the range of κ=0.75-83 for SARI and ICC=0.93 for RALE scale [15,16].

Lower interobserver agreement was observed for AP radiographs as compared to PA radiographs for each scale,

Table 7. Analysis of the correlation between the radiological image as assessed in individual scales and clinical condition as expressed using MEWS scale.

		Radiologist 1			Radiologist 2		
		SARI	RALE	CXR	SARI	RALE	CXR
Spearman's ρ		0.534	0.425	0.577	0.541	0.462	0.591
p		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
95% CI	UL	0.661	0.574	0.698	0.668	0.604	0.711
	LL	0.377	0.249	0.417	0.414	0.292	0.452

suggesting the relationship between the reported results and the quality of the scan.

In the anatomical context, somewhat lower interobserver agreement was observed for SARI scale assessments of the left lung as compared to the right lung. Similarly, in the case of the CXR Score scale, the lowest interobserver agreement was observed for the lower left lung field.

These findings may suggest a conclusion that evaluation of regions where other structures cover the parenchyma of lungs (such as heart) can be more subjective. This affects the overall scoring of an assessing radiologist, and their evaluation may be biased.

In each of the analyzed scales, the best interobserver agreement was observed in patients in mild clinical condition (MEWS of 0-1). Lower agreement was observed both in patients with the moderate severity of symptoms (MEWS of 2-3) and in patients in severe condition (MEWS ≥4). Moderate correlation (r=0.425-0.591) was identified in the study between the score obtained in each of the analyzed scales and the clinical condition of the patient as expressed using MEWS.

The strongest correlation with the patient's clinical condition was shown for the 18-point CXR Score scale (r=0.577 and 0.591).

The present study is limited by a relatively small number of patients (200 cases) and radiologists assessing the scans. However, kappa values comparable to those presented in other studies on patients with COVID-19 suggest that these factors had no effect on the obtained results.

In our opinion, parametric evaluation of chest radiographs in patients with SARS-CoV-2 infection is crucial for patient triage and therapeutic decision making.

Further validation is required with regard to quantitative analysis of chest radiographs and their predictive value in the context of the clinical course of the disease.

Parameterization of radiological images can also provide a useful tool for the development of computer-aided diagnosis and AI artificial intelligence systems.

Conclusions

The analyzed scales are characterized by good or very good interobserver agreement of assessments of the extent of pulmonary lesions being made by independent, experienced radiologists.

The lowest interobserver agreement was observed for the SARI scale, while the results for the RALE and the CXR Score scales were similar, with overlapping CIs. Since the CXR Score showed the strongest correlation with the clinical condition of the patient as expressed using the MEWS scale, it is the preferred scale for chest radiograph assessment of patients with COVID-19 in the light of data provided.

Conflict of Interest

None.

Declaration of Figures Authenticity

All figures submitted have been created by the authors, who confirm that the images are original with no duplication and have not been previously published in whole or in part.

References:

1. Toussie D, Voutsinas N, Finkelstein M, et al. Clinical and chest radiography features determine patient outcomes in young and middle age adults with COVID-19. Radiology. 2020;297(1):E197-206

2. Bernheim A, Mei X, Huang M, et al. Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection. Radiology. 2020;295(3):200463

3. Zu ZY, Jiang MD, Xu PP, et al. Coronavirus disease 2019 (COVID-19): A perspective from China. Radiology. 2020;296(2):E15-25

4. Yoon SH, Lee KH, Kim JY, et al. Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): Analysis of nine patients treated in Korea. Korean J Radiol. 2020;21(4):494-500

5. Li Y, Xia L. Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management. Am J Roentgenol. 2020;214(6):1280-86

6. Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. Radiology. 2020;296(2):E115-17

7. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis. 2020;20(4):425-34

8. Pan F, Ye T, Sun P, et al. Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19). Radiology. 2020;295(3):715-21

9. Jędrusik P, Gaciong Z, Sklinda K, et al. Diagnostic role of chest computed tomography in coronavirus disease 2019. Pol Arch Intern Med. 2020;130(6):520-28

10. Wong HYF, Lam HYS, Fong AH, et al. Frequency and distribution of chest radiographic findings in COVID-19 positive patients. Radiology. 2019;296(2):E72-78

11. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: Experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiol Med. 2020;125(5):509-13

12. Chau TN, Lee PO, Choi KW, et al. Value of initial chest radiographs for predicting clinical outcomes in patients with severe acute respiratory syndrome. Am J Med. 2004;117(4):249-54

13. Hui DS, Wong KT, Antonio GE, et al. Severe acute respiratory syndrome: Correlation between clinical outcome and radiologic features. Radiology. 2004;233(2):579-85

14. Antonio GE, Wong KT, Tsui EL, et al. Chest radiograph scores as potential prognostic indicators in severe acute respiratory syndrome (SARS). Am J Roentgenol. 2005;184(3):734-41

15. Taylor E, Haven K, Reed P, et al. A chest radiograph scoring system in patients with severe acute respiratory infection: A validation study. BMC Med Imaging. 2015;15:61

16. Warren MA, Zhao Z, Koyama T, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax. 2018;73(9):840-46

17. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM. 2001;94(10):521-26

18. Fleiss JL. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement. 1973;33(3):613-19

19. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82

20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-63 [Erratum in: J Chiropr Med. 2017;16(4): 346]