The chest radiographic scoring system in initial diagnosis of COVID-19: Is a radiologist needed?

Marta Rorat1,A–F, Tomasz Jurek1,C–F, Krzysztof Simon2,C,E,F, Maciej Guziński3,B–F

1 Department of Forensic Medicine, Wroclaw Medical University, Poland
2 Department of Infectious Diseases and Hepatology, Wroclaw Medical University, Poland
3 Department of General Radiology, Interventional Radiology and Neuroradiology, Wroclaw Medical University, Poland

A — research concept and design; B — collection and/or assembly of data; C — data analysis and interpretation;
D — writing the article; E — critical revision of the article; F — final approval of the article

Address for correspondence
Tomasz Jurek
E-mail: tomasz.jurek@umed.wroc.pl

Funding sources
This work was carried out according to subject register in Simple system SUB.A120.19.036 and SUB.A120.21.030 and supported by statutory subvention granted by the Ministry of Science and Higher Education (Poland).

Conflict of interest
None declared

Received on December 22, 2020
Reviewed on May 4, 2021
Accepted on July 2, 2021
Published online on July 30, 2021

Cite as

Abstract

Background. Lung imaging, next to a polymerase chain reaction (PCR) test, is a key diagnostic tool in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. The degree of abnormalities correlates with clinical outcome. Imaging of the lungs using chest radiography (CXR) at the peak of a pandemic is considered a basic diagnostic tool at the triage stage. The CXR images are less characteristic than computed tomography (CT) and should be interpreted with a combination of clinical findings.

Objectives. Comparison of the usefulness of 2 CXR severity scores to evaluate the presence/severity of inflammation in the course of COVID-19 and the possibility of a non-radiologist to interpret the image independently.

Materials and methods. Retrospective analysis of the medical records of 152 consecutive patients (aged 19—96, 73 men), infected with SARS-CoV-2, confirmed using real-time PCR (RT-PCR). Five-point and twelve-point CXR severity scoring systems were used (independently by a radiologist and a referring physician) to assess the severity of inflammation.

Results. In 77 of 152 cases, the CXR revealed features of inflammation. Bilateral abnormalities were found in 48/77 (62.3%) cases. Statistically, the lower lobes were involved more often than the upper ones (p < 0.001) and the left lobe more often than the right one (p < 0.001). The intensity of the abnormalities using both scales correlated with the persistence of symptoms (p = 0.0133 and p = 0.0403). A positive and statistically significant correlation was found between both scales and dyspnea, decreased oxygen saturation, elevated C-reactive protein (CRP), ferritin, D-dimer, lactate dehydrogenase, and alanine aminotransferase activity. The interobserver agreement analysis did not show a statistically significant difference in the CXR severity score using the five-point (B = 0.8345, kappa = 0.82; p = 0.1480) or the twelve-point scale (B = 0.8219, kappa = 0.77; p = 0.0502).

Conclusions. The CXR severity score is a useful tool to assess the inflammation in the initial diagnosis of coronavirus disease 2019 (COVID-19). Quantifying lung abnormalities accurately may be performed by a referring physician. Both CXR severity scales correlate well with clinical parameters.

Key words: chest X-ray, imaging, viral infection, COVID-19, lung opacities
Background

Lung imaging, next to PCR testing, is a key diagnostic tool in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. So far, computed tomography (CT) has been a widely validated method in coronavirus disease 2019 (COVID-19). In CT tests, inflammatory lesions can be detected in symptomatic and asymptomatic patients, which significantly increases the sensitivity of this diagnostic method.1 The abnormalities present in patients with COVID-19 are diverse, depend on the severity of infection and the duration of symptoms, and undergo dynamic changes.2–4 Lung abnormalities appear the earliest, within the first days of infection, but the most severe abnormalities appear about 10 days after the onset of initial symptoms.3 The lack of visible changes on a CT or the presence of atypical changes does not exclude SARS-CoV-2 infection.5 Although assessment based on a CT test is characterized by high sensitivity and substantially expands the knowledge on the severity of inflammation, in many medical centers, access to this test is restricted, used mainly for epidemiological and logistic reasons. This is usually the case at the peak of the pandemic, in field hospitals and in countries where health services are insufficient. In such situations, the use of a portable chest X-ray (CXR) is sufficient in the triage stage as this method is fast and carries a low risk of cross-infection. This procedure is consistent with recommendations from the American College of Radiology (ACR), the British Society of Thoracic Imaging, and the Polish Agency for Health Technology Assessment and Tariff System.6–8

As the literature-based data suggests, the CXR image is less characteristic and requires more careful interpretation.9 Test sensitivity is estimated to be 68.1–89% and the specificity is 60.6%.10–12 Data concerning CXR itself are still limited. There are no standards for evaluating a CXR in patients with COVID-19, and reports on the frequency of occurrence and the type of changes are still scarce. The reports published so far indicate that the lesions found most commonly in the early phase of COVID-19 are ground-glass opacities, reticular alterations, and consolidations that gradually increase as the illness progresses.13 Abnormalities on a CXR are found less frequently than with CT, which is why the CXR image should be interpreted with a combination of clinical and laboratory findings.7

Objectives

This study aimed to explore the correlation with laboratory results and to compare the usefulness of 2 scoring scales to assess the presence and severity of inflammation in the course of COVID-19 using CXR, as well as to evaluate the possibility of a non-radiologist (referring physician) interpreting the presence and degree of inflammation in the lungs independently.

Materials and methods

A retrospective analysis of X-ray images of 152/165 consecutive patients (48% male, average age 56.6 ±16.8 years) infected with COVID-19 (confirmed using real-time polymerase chain reaction – RT-PCR), in which an anteroposterior chest X-ray was performed on admission. The patients were hospitalized in the period between March 6 and April 16, 2020. Patients who had undergone thoracic surgery in the 2 weeks beforehand (7 patients), those with active tuberculosis (2 patients), disseminated cancer of the lungs (3 patients), or who underwent an incorrectly performed CXR (1 patient) were excluded from the analysis. Another study based on the same group of medical records using only a five-point severity scoring system to assess the correlation of CXR with patients’ health conditions was performed.9

Table 1 presents the baseline characteristics of study patients.

Chest X-rays were interpreted and described independently by 2 doctors – an experienced radiologist and a referring physician (infectious diseases specialist who diagnoses and treats patients infected with SARS-CoV-2 daily).

Table 1. Baseline characteristics of study patients (n = 152)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age [years]</td>
<td>19–96, average 56.6; SD = 16.8 women average 58.1; SD = 16.5 men average 55.1; SD = 16.8</td>
</tr>
<tr>
<td>Sex</td>
<td>M 73; F 79</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
</tr>
<tr>
<td>cardiovascular diseases (including hypertension)</td>
<td>83 (54.6%)</td>
</tr>
<tr>
<td>pulmonary diseases</td>
<td>66 (43.4%)</td>
</tr>
<tr>
<td>malignant neoplasm</td>
<td>14 (9.2%)</td>
</tr>
<tr>
<td>obesity</td>
<td>19 (12.5%)</td>
</tr>
<tr>
<td>diabetes</td>
<td>22/83 (26.5%)</td>
</tr>
<tr>
<td>other (autoimmune diseases, liver cirrhosis)</td>
<td>19 (12.5%)</td>
</tr>
<tr>
<td></td>
<td>8 (5.3%)</td>
</tr>
<tr>
<td>Duration of hospitalization [days] (mean)</td>
<td>1–69 (14)</td>
</tr>
<tr>
<td>Deaths, n (%)</td>
<td>16 (10.5%)</td>
</tr>
</tbody>
</table>

SD – standard deviation.
The infectious diseases specialist had a short training session on using the 2 scoring systems based on validated radiological images. The 1st one (five-point scale) is a chest X-ray severity scoring system proposed by Taylor et al. CXR findings were categorized as: 1 – normal; 2 – patchy atelectasis and/or hyperinflation and/or bronchial wall thickening; 3 – focal consolidation, no more than 1 lobe; 4 – multifocal and bilateral consolidation; and 5 – diffuse alveolar changes.14 The 2nd scoring system (twelve-point scale) is our modification of a system proposed by Borghesi et al.15 The assessment of the severity of abnormalities was performed in 4 quadrants, similar to the system used in the radiographic assessment of lung edema (RALE), using a scoring system with 1–3 points for each of the 4 quadrants based on the percentage of the quadrant with opacification: 1 – normal, 2 – lesions <50% of the pulmonary field and 3 – lesions involving ≥50% of the pulmonary field. All CXRs were performed with the use of a portable device in an isolated room.

The results from the use of the 2 scales by a radiologist and non-radiologists were compared. The twelve-point scale results were contrasted with those from the five-point scale.9 The correlation between clinical parameters: the presence of comorbidities, dyspnea and cough, saturation and laboratory test results (morphology, capillary blood gas test, C-reactive protein – CRP, lactate dehydrogenase – LDH, serum alanine aminotransferase – ALT activity, D–dimer, and ferritin level) was analyzed.

### Statistical analyses

Given the ordinal nature of the scores compared, we used non-parametric statistics when comparing levels between groups. To compare the scores between the 2 groups, we used the Mann–Whitney test. The correlation between scores and quantitative variables was assessed using Spearman’s rank correlation coefficient. Assessment of the presence of inflammation among various positions was performed using Pearson’s χ² test of independence. The optimal cut-off point for CXR scores to predict death was performed by maximizing the Youden’s index in a receiver operating characteristic (ROC) curve analysis. All tests were considered significant when the p-value was lower than 0.05. Calculations were performed using the R statistical program for Windows (v. 4.0; https://www.r-project.org/).16 B-statistics and kappa statistics were used to quantify the agreement between the 2 observers. The following R packages were used:


The rest of the tests were performed using built-in tests.

### Results

In the research group, the severity of inflammation in CXR images was assessed using a five-point scale (Table 2) and a twelve-point scale (Table 3). Among 77 patients with features of pneumonia detected on CXR, bilateral changes were found in 48/77 (62.3%), peripheral (±central) opacities in 44/77 (57.1%), heart

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Number of points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Assessment of a radiologist</td>
<td>number of patients (%)</td>
</tr>
<tr>
<td>Assessment of a referring physician</td>
<td>number of patients (%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Number of points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Assessment of a radiologist</td>
<td>number of patients (%)</td>
</tr>
<tr>
<td>Assessment of a referring physician</td>
<td>number of patients (%)</td>
</tr>
</tbody>
</table>
enlargement in 22/77 (28.6%), pleural effusion in 6/77 (7.8%), and emphysema in 1 case as a complication of vena cava catheterization. We observed a statistically significant presence of inflammation in the lower than the upper lobes (48.03% compared to 15.3%; p < 0.001; Pearson’s χ² test) and in the left lobes rather than the right ones (44.74% compared to 34.87%; p < 0.001; Pearson’s χ² test). The frequency of pleural effusions (mean 2.27 ±1.39 compared to 2.67 ±1.86, p = 0.565; Mann–Whitney test, Shapiro–Wilk W statistics 381) was not statistically significant in contrast to heart enlargement (mean 3.42 ±0.85 compared to 3.82 ±0.79, p = 0.0477; Mann–Whitney test, W statistics 442) and peripheral opacities (mean 3.06 ±0.89 compared to 3.85 ±0.67, p = 0.0001; Mann–Whitney test, W statistics 373). Patients reported to the hospital with symptoms having evolved over 1–23 days (6.5 days on average). The severity of abnormalities measured using both scales remained in relation with the duration of symptoms (five-point scale: r = 0.21, p = 0.0133; twelve-point scale r = 0.17, p = 0.0403; Spearman’s rank correlation coefficient). Patients in which symptoms persisted for more than 7 days had a significantly higher score using both scales (five-point scale: <7 days mean 1.99 ±1.32, ≥7 days 2.68 ±1.45, p = 0.04; twelve-point scale: <7 days mean 4.93 ±1.54, ≥7 days 5.57 ±1.78, p = 0.009; Spearman’s rank correlation coefficient).

The interobserver agreement analysis did not show a statistically significant difference in CXR assessment using the five-point scale (B = 0.8345, kappa = 0.82; p = 0.148) or the twelve-point scale (B = 0.8219, kappa = 0.77; p = 0.0502). High compliance of assessments between the radiologist and referring physician was observed (Tables 2,3). An almost perfect interobserver agreement and substantial agreement were detected. The first researcher (radiologist) obtained lower results on average using the twelve-point scale.

We demonstrated, like in the other papers where the five-point scale had been used,⁷ the correlation between scores using the twelve-point scale – presence/severity of inflammation changes in CXR – and age (r = 0.45, p < 0.001; Spearman’s rank correlation coefficient), obesity (mean 4.92 ±1.38 compared to 6.77 ±2.05, p < 0.001; Mann–Whitney test, W statistics 290), cardiovascular system diseases (mean 5.77 ±1.19 compared to 6.70 ±1.71, p = 0.0222; Mann–Whitney test, W statistics 494), increased level of CRP > 6 mg/L (r = 0.61, p < 0.001), D-dimers >500 ng/mL (r = 0.52, p < 0.001; Spearman’s rank correlation coefficient), ferritin level >291 ng/mL (r = 0.49, p = 0.0081; Spearman’s rank correlation coefficient), ALT activity >48 IU/L (r = 0.57, p < 0.001; Spearman’s rank correlation coefficient), LDH > 246 U/L (r = 0.69, p < 0.001; Spearman’s rank correlation coefficient), decreased O₂ saturation <94% (mean 4.60 ±1.09 compared to 6.49 ±1.86, p < 0.001; Mann–Whitney test, W statistics 927), and finally, dyspnea (mean 4.69 ±1.23 compared to 5.63 ±1.82, p = 0.0002; Mann–Whitney test, W statistics 1916). We did not observe statistical significance between the presence of fever or cough and other concomitant diseases. On both scales, the higher scores occurred in patients with longer durations of hospitalization (five-point scale: r = 0.32, p = 0.001; twelve-point scale: r = 0.29, p = 0.0036; Spearman’s rank correlation coefficient) and the deceased (five-point scale: mean 2.15 ±1.34 compared to 3.44 ±1.46, p = 0.0005; twelve-point scale: mean 4.98 ±1.40 compared to 6.94 ±2.32, p = 0.0003; Mann–Whitney test, W statistics 534). We did not find any significant differences between the scales within this range. The ROC analysis on death prediction was used to establish the cut-off point for the five-point scale at 2 points and the twelve-point scale at 7 points.

The above data are also presented in Fig. 1, 2.

---

Fig. 1. A. April 10, a 53-year-old man, 4th day after the onset of symptoms: cough, fever, oxygen saturation 97%; chest X-ray (CXR): unilateral, focal, peripheral patchy consolidation – ground-glass opacity in the right lung; five-point scale score: 3; twelve-point scale score: 5 (1+2+1+1). B. April 17, CXR: peripheral multifocal diffuse patchy consolidations in the right and left lung; five-point scale score: 4; twelve-point scale score: 8 (2+3+1+2)
Discussion

According to the literature, the overall rate of a positive CXR in COVID-19 is between 43.4% and 94.4%,\textsuperscript{10–13,17} and is higher in patients with a long-lasting course of the disease.\textsuperscript{13} Our research paper revealed the presence of alterations in CXR, regardless of their type, in 50.7% of patients at the time they reported to the hospital. A lower number of positive results compared to Italian studies may be due to the fact that younger patients, often in a better general state of health, attended the hospital at an earlier stage of the disease and with a milder course thereof.

It is well known that community bacterial pneumonia is usually unilateral and involves 1 lobe. However, in infections with SARS-CoV-2, lung opacities are typically multifocal, bilateral and peripheral.\textsuperscript{10,12,13,17,18} In our study, alterations were bilateral in 62.3% of patients – in other papers, they were reported in 50–73.3% of cases,\textsuperscript{10,12,13,17,18} which, as is well known, depends on the persistence of the disease. Our research shows a statistically significant occurrence of heart enlargement compared to patients with no inflammation, which corresponds to reports by Cozzi et al.\textsuperscript{10} Peripheral involvement took place in over ½ of the cases, which is consistent with other reports.\textsuperscript{10,13,18} Just like other researchers, we detected that the involvement of the lower fields was more predominant.\textsuperscript{10,13,19,20} There was also a predominance of the left side over the right side in contrast to research by Vancheri et al., where neither side presented prevalence,\textsuperscript{13} or research by Cozzi et al. and Toussie et al., where the right side was predominant (58% and 42%, respectively).\textsuperscript{10,20} From a clinical point of view, this feature does not seem to be significant. Pleural effusions were present in <10% of those infected, similar to the research by Wong et al. and Vancheri et al.,\textsuperscript{12,13} whereas higher percentages were noted by Cozzi et al. – 16.7% and Ippolito et al. – 12.2%.\textsuperscript{10,18}

We considered a variety of CXR results in patients with COVID-19 at the moment of admission to the hospital – ranging from reticular alterations, more or less intensified ground-glass opacities co-occurring with reticular alterations or alone, to single or multiple, sometimes massive consolidations. In some patients, the changes were restricted to 1 lobe and in others, they were disseminated. According to the literature, the picture of changes depends, among others, on the phase of the disease. Consolidations are detected less frequently than other lesions, especially in the initial phase of the illness and they tend to increase over time.\textsuperscript{10,13,18} However, there is no research on the correlation between the type of changes with the clinical picture. In our opinion, the assessment of the severity of abnormalities is more important than the analysis of an individual alteration occurrence. The amount of involved lung tissue has a direct influence on lung impairment and clinical status. Therefore, we believe it is vital for a referring physician to perform a fast initial scoring of the severity of abnormalities using CXR. As our paper shows, the results obtained are highly consistent with an evaluation by a radiologist. In situations where a radiologist is unavailable to provide a quick evaluation, the interpretation of the severity of inflammatory abnormalities by a referring doctor serves as a valuable diagnostic and prognosticative guideline.

Scoring to assess the severity of inflammation in the lungs does not require the use of calculators. It is comprehensive and easy for any physician to do. We did not detect superiority in any of the scales. Both scales were equally correlated with many clinical parameters. The five-point scale is easy to use and to interpret, it informs about the progression of the illness, it has mainly qualitative and, to a lesser extent, quantitative features; it is also less subjective. In turn, the twelve-point scale, which is similar to the RALE score (Radiographic Assessment of Lung Edema) used by other

![Fig 2. A. April 15, a 54-year-old woman; 4th day after the onset of symptoms; cough, fever, tachypnea, dyspnea, oxygen saturation 86%; chest X-ray (CXR): bilateral consolidation of lower lobes; five-point scale score: 4; twelve-point scale score: 8 (1+1+2+2); B. April 29, CXR: progression of consolidations and patchy opacities; five-point scale score: 4; twelve-point scale score: 10 (2+3+2+3)](image-url)
researchers, is used for a semi-quantitative evaluation and in this range, better presents the severity of inflammatory changes. However, the aggregated number of points does not reflect the distribution of abnormalities in given quadrants as an 8/12 point evaluation might show involvement of 2 (3+3+1+1), 3 (3+2+2+1) or 4 quadrants (2+2+2+2).

In our opinion, scoring according to one of the scales is clinically more useful than a description on its own and should be an essential component of a structured reporting strategy. As no scale was deemed superior, it is more legitimate to use an easier tool in the form of a simpler and clearer five-point scale. The more complex the scale, the more uncertain the assessment. The authors’ experience in treating COVID-19 patients allows them to highlight the importance of interpreting the image and not just the description. Our paper proves that the five-point and twelve-point scales for CXR scoring in COVID-19 patients can be used by a referring physician with the risk of error not exceeding 10%. This is vital in situations when an urgent decision about subsequent treatment for a patient is required (to send them home, to admit them to a hospital or to perform further diagnostics with more tests, including imaging tests). A RALE score was used in research by Wong et al. and Cozzi et al.10,12 It is marginally more complex as it requires assessment of consolidations on a scale of 0–4 and density on a scale of 0–3 with the values then being multiplied by each other. The assessment is carried out in 4 quadrants and the sum thereof gives the final result.19 The greater complexity of the scale the less useful it is in emergency situations. In our research, the five-point scale produced a marginally greater consistency between the results of a radiologist and a referring physician.

The use of these scales also has prognostic importance as shown in the research by Tousie et al. In this research, a zonal scale was used to reveal that CXR severity scores represent an independent prognostic indicator of outcomes in COVID-19 patients.20 In this case, the lungs were divided into 6 zones and it was demonstrated that if opacities were present in at least 2 lung zones, the patient was more likely to require hospitalization, but if the changes were present in at least 3 lung zones, they required intubation.

Limitations

Our study has several limitations. Firstly, it is a retrospective research study. Secondly, the time between the onset of symptoms and reporting to the hospital was highly variable. The number of patients with a positive CXR result is too small to divide the patients into groups based on the duration of symptoms in order to perform a comparative analysis between the groups. The lower quality of portable X-ray images should also be taken into account.

Conclusions

The CXR severity score is a useful tool to assess the severity of inflammatory changes in the initial diagnosis of COVID-19. At the peak of a pandemic, when the system is overwhelmed, quantifying lung abnormalities accurately can be performed by a referring physician with a substantial agreement in respect to radiological evaluation. In such a situation, the function of a radiologist should be to conduct training for referring clinicians as well as being helpful in cases of uncertain images. Simple and complex CXR severity scales correlate well with clinical parameters thus the less complex five-point scale should be recommended as an essential component of a structured reporting strategy. The presence of inflammatory changes in CXR, even non-severe ones, is an independent factor of worse prognosis in COVID-19.

ORCID iDs

Marta Rorat https://orcid.org/0000-0002-5318-4945
Tomasz Jurek https://orcid.org/0000-0003-0110-0276
Krzysztof Simon https://orcid.org/0000-0002-8040-0412
Maciej Guziński https://orcid.org/0000-0002-9781-2114

References


