



## Research article

## Chest x-ray in the COVID-19 pandemic: Radiologists' real-world reader performance



Andrea Cozzi<sup>a,1</sup>, Simone Schiaffino<sup>b,1,\*</sup>, Francesco Arpaia<sup>c</sup>, Gianmarco Della Pepa<sup>c</sup>, Stefania Tritella<sup>b</sup>, Pietro Bertolotti<sup>b</sup>, Laura Menicagli<sup>b</sup>, Cristian Giuseppe Monaco<sup>b</sup>, Luca Alessandro Carbonaro<sup>b</sup>, Riccardo Spairani<sup>b</sup>, Bijan Babaei Paskeh<sup>b</sup>, Francesco Sardanelli<sup>a,b</sup>

<sup>a</sup> Department of Biomedical Sciences for Health, Università degli Studi di Milano, Via Luigi Mangiagalli 31, 20133 Milano, Italy

<sup>b</sup> Unit of Radiology, IRCCS Policlinico San Donato, Via Rodolfo Morandi 30, 20097 San Donato Milanese, Italy

<sup>c</sup> Postgraduate School in Radiodiagnostics, Università degli Studi di Milano, Via Festa del Perdono 7, 20122 Milano, Italy

## ARTICLE INFO

## Keywords:

Severe acute respiratory syndrome coronavirus 2  
COVID-19  
Pneumonia, viral  
Radiography, thoracic

## ABSTRACT

**Purpose:** To report real-world diagnostic performance of chest x-ray (CXR) readings during the COVID-19 pandemic.

**Methods:** In this retrospective observational study we enrolled all patients presenting to the emergency department of a Milan-based university hospital from February 24th to April 8th 2020 who underwent nasopharyngeal swab for reverse transcriptase-polymerase chain reaction (RT-PCR) and anteroposterior bedside CXR within 12 h. A composite reference standard combining RT-PCR results with phone-call-based anamnesis was obtained. Radiologists were grouped by CXR reading experience (Group-1, >10 years; Group-2, <10 years), diagnostic performance indexes were calculated for each radiologist and for the two groups.

**Results:** Group-1 read 435 CXRs (77.0 % disease prevalence): sensitivity was 89.0 %, specificity 66.0 %, accuracy 83.7 %. Group-2 read 100 CXRs (73.0 % prevalence): sensitivity was 89.0 %, specificity 40.7 %, accuracy 76.0 %. During the first half of the outbreak (195 CXRs, 66.7 % disease prevalence), overall sensitivity was 80.8 %, specificity 67.7 %, accuracy 76.4 %, Group-1 sensitivity being similar to Group-2 (80.6 % versus 81.5 %, respectively) but higher specificity (74.0 % versus 46.7 %) and accuracy (78.4 % versus 69.0 %). During the second half (340 CXRs, 81.8 % prevalence), overall sensitivity increased to 92.8 %, specificity dropped to 53.2 %, accuracy increased to 85.6 %, this pattern mirrored in both groups, with decreased specificity (Group-1, 58.0 %; Group-2, 33.3 %) but increased sensitivity (92.7 % and 93.5 %) and accuracy (86.5 % and 81.0 %, respectively).  
**Conclusions:** Real-world CXR diagnostic performance during the COVID-19 pandemic showed overall high sensitivity with higher specificity for more experienced radiologists. The increase in accuracy over time strengthens CXR role as a first line examination in suspected COVID-19 patients.

## 1. Introduction

Since the start of the COVID-19 pandemic, international recommendations [1,2] have repeatedly stated that the diagnosis of SARS-CoV-2 infection should primarily rely on viral testing rather than on chest imaging.

This endorsed reference standard, *i.e.* reverse transcriptase-polymerase chain reaction (RT-PCR) on nasal or throat swabs, has become essential in the triage and monitoring phases of patients with

suspected SARS-CoV-2 infection [3], but is encumbered by a sensitivity oscillating between 38 % and 89 % [4–6]. Moreover, during the pandemic peak, RT-PCR response times became often incompatible with appropriate triaging and management of the high number of suspect COVID-19 cases simultaneously presenting to emergency departments [7–9], forcing the incorporation of imaging in the diagnostic pathway to compensate the aforementioned shortcomings of RT-PCR [2,10,11]. While the use of chest CT – even as a triaging test – was almost ubiquitous [11–13], both initial reports from China and a recent

\* Corresponding author.

E-mail addresses: [andrea.cozzi1@unimi.it](mailto:andrea.cozzi1@unimi.it) (A. Cozzi), [simone.schiaffino@grupposandonato.it](mailto:simone.schiaffino@grupposandonato.it) (S. Schiaffino), [francesco.arpaia@unimi.it](mailto:francesco.arpaia@unimi.it) (F. Arpaia), [gianmarco.dellapepa@unimi.it](mailto:gianmarco.dellapepa@unimi.it) (G. Della Pepa), [stefania.tritella@grupposandonato.it](mailto:stefania.tritella@grupposandonato.it) (S. Tritella), [pietro.bertolotti@grupposandonato.it](mailto:pietro.bertolotti@grupposandonato.it) (P. Bertolotti), [laura.menicagli@grupposandonato.it](mailto:laura.menicagli@grupposandonato.it) (L. Menicagli), [cristiangiuseppe.monaco@grupposandonato.it](mailto:cristiangiuseppe.monaco@grupposandonato.it) (C.G. Monaco), [luca.carbonaro@grupposandonato.it](mailto:luca.carbonaro@grupposandonato.it) (L.A. Carbonaro), [riccardo.spairani@grupposandonato.it](mailto:riccardo.spairani@grupposandonato.it) (R. Spairani), [paskehbijan.babaei@grupposandonato.it](mailto:paskehbijan.babaei@grupposandonato.it) (B. Babaei Paskeh), [francesco.sardanelli@unimi.it](mailto:francesco.sardanelli@unimi.it) (F. Sardanelli).

<sup>1</sup> Andrea Cozzi and Simone Schiaffino contributed equally to this work.

meta-analysis highlighted its low specificity [14]. Therefore, an ever-growing number of institutions have come to prefer chest x-ray (CXR), also taking into account that CXR can be performed with portable equipment in isolation rooms [15] or even in external settings [16]. Such choice also minimizes potential contact between patients and operators, as well as other patients [15–18]. This has been the case in our hospital, located less than 25 miles from the first pandemic hotspot in Lombardy, Italy.

Apart from small-scale case series [19,20], at the time of writing three major retrospective studies have evaluated the diagnostic performance of CXR performed as a triaging test on emergency department admission [21–23]. The two largest by far are a retrospective review by a single radiologist of 518 CXRs acquired during the first phase of the pandemic peak (from March 1st to March 15th) – with a resulting overall sensitivity of 57 % [22] – and a study coming from our group and performed on 535 patients [23]. In our analysis we instead considered the dichotomized reports of all radiologists on duty during a larger period (*i.e.*, from February 24th to April 8th, 2020), obtaining an overall 89.0 % sensitivity and 60.6 % specificity, using a composite reference standard (RT-PCR supplemented by anamnestic data and patient follow-up, as well as by RT-PCR repetition in negative cases). We aim now to further analyse the radiologists' real-world performance in CXR reading during the COVID-19 pandemic, distinguishing them according to their CXR reading experience.

## 2. Materials and methods

This retrospective observational study was approved by the local Ethics Committee and performed between February 24th and April 8th, 2020, at IRCCS Policlinico San Donato (San Donato Milanese, Italy), a university hospital mainly focusing on cardiovascular diseases but promptly converted to a primarily COVID-19-dedicated hospital during the pandemic peak.

We included in this study all patients presenting to our emergency department for suspected SARS-CoV-2 infection who underwent both a nasopharyngeal swab for RT-PCR and an anteroposterior bedside CXR within 12 h from admission. At our hospital, CXRs are reported by the on-duty radiologist within about 60–90 min if performed during the day shift (07:00 am – 08:00 pm), and at the beginning of the following working day if performed during the night shift (08:00 pm – 07:00 am). Considering the delay in the availability of RT-PCR results, caused by the high number of patients incessantly presenting to the emergency department during the pandemic peak in our region, all CXRs in the study period were reported by radiologists forcedly blinded to RT-PCR results.

For the purposes of this study, as previously described [23], we then built a composite reference standard to improve RT-PCR sensitivity, by combining RT-PCR results with phone-call-based complete anamnesis in RT-PCR-negative patients who had not repeated the swab during hospitalization. Considering the rather unspecific nature of CXR findings in patients with COVID-19 pneumonia, a radiologist with 5 years of experience in CXR interpretation (S.S.) reviewed all routine CXR reports – being blinded for the original radiologists' signatures – in order to classify them dichotomously as positive or negative for COVID-19. The absence of pulmonary abnormalities on a CXR determined its classification as a negative one, while the presence of interstitial infiltrates – associated or not with alveolar infiltrates – with predominantly bilateral and basal distribution on a CXR implied its classification as a positive examination [1,2,11]. Conversely, CXR findings unrelated to COVID-19, such as lobar alveolar infiltrates (typically associated with bacterial pneumonia) pleural effusion, pneumothorax, were considered as non-COVID-19-related finding for the purpose of this dichotomization.

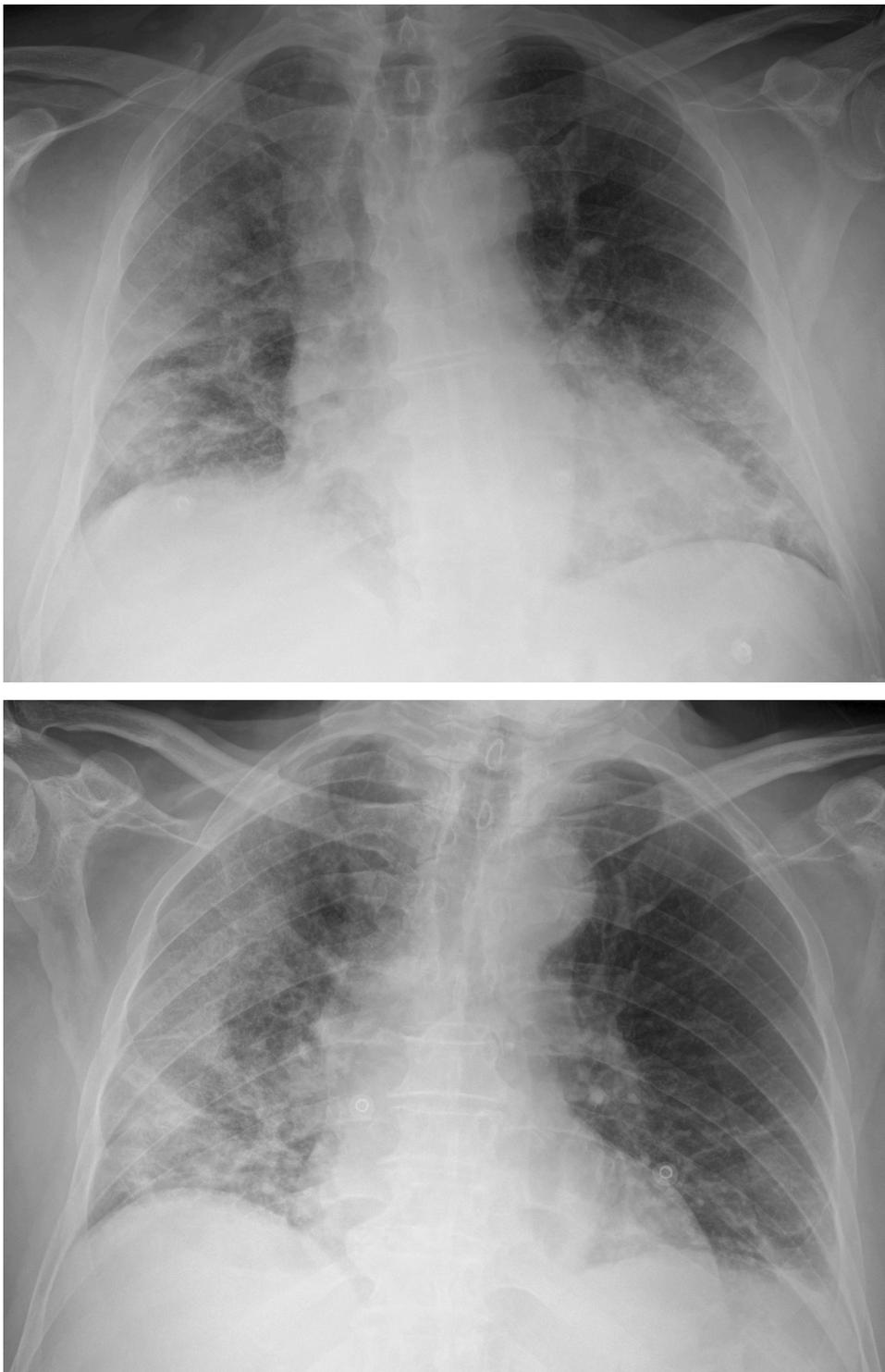
We grouped the seven radiologists from our department by their CXR reading experience: Group 1 included 4 radiologists (R1, R2, R3, and R4) with 10 or more years of experience in CXR reading; Group 2 included 3 (R5, R6, and R7) radiologists with less than 10 years of experience in CXR reading. All radiologists were board-certified: if a resident was in charge of drafting a first version of the report, the report was always

checked by a board-certified radiologist and the final version was signed by the same board-certified radiologist. Only one of the seven radiologists (in Group 1) has a particular dedication to breast imaging but practices at least half of his time as a general radiologist. Overall and patient-sex-specific diagnostic performance indexes were calculated for each radiologist and for the two groups over the 6-week timeframe and according to the first and second half of all CXRs read by each radiologist. Data are presented as sensitivity, specificity, positive predictive value, negative predictive value, accuracy, positive likelihood ratio, negative likelihood ratio, and their 95 % confidence intervals (CI). Statistical analyses were performed using Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA, USA).

## 3. Results

In the six-week study period, R1 read 180 CXRs, with a 79 % disease prevalence, R2 read 147 CXRs with a 70 % disease prevalence, R3 read 65 CXRs with an 80 % disease prevalence, and R4 read 43 CXRs with an 88 % disease prevalence. Overall, readers from Group 1 read 435 CXRs with a 77.0 % disease prevalence, obtaining an 89.0 % sensitivity (95 % CI 85.2 %–91.9 %), a 66.0 % specificity (95 % CI 56.3 %–74.5 %), an 83.7 % accuracy (95 % CI 79.9 %–86.9 %), an 89.8 % positive predictive value (95 % CI 86.0 %–92.6 %), a 64.1 % negative predictive value (95 % CI 54.5 %–72.7 %), a 2.62 positive likelihood ratio (95 % CI 1.99–3.45), and a 0.17 negative likelihood ratio (95 % CI 0.12–0.23). In Group 2, R5 read 59 CXRs with a 78 % disease prevalence, R6 read 27 CXRs with a 70 % disease prevalence, R7 read 14 CXRs with a 57 % disease prevalence; overall, readers from Group 2 read 100 CXRs with a 73.0 % disease prevalence, obtaining an 89.0 % sensitivity (95 % CI 79.8 %–94.3 %), a 40.7 % specificity (95 % CI 24.5 %–61.0 %), a 76.0 % accuracy (95 % CI 66.8 %–83.3 %), an 80.2 % positive predictive value (95 % CI 70.3 %–87.5 %), a 57.9 % negative predictive value (95 % CI 36.3 %–76.9 %), a 1.50 positive likelihood ratio (95 % CI 1.09–2.08), and a 0.27 negative likelihood ratio (95 % CI 0.12–0.60). Fig. 1 shows an example of a true positive and of a false positive case both for Group 1 and Group 2, Table 1 details overall performance indexes of all readers, and Table 2 shows the results of readers performance evaluation according to patients subgroups and different timeframes (*i.e.* the first and second three-week periods).

Considering the first half and the second half of all CXRs read by each radiologist, we observed an increase in disease prevalence for 5 out of 7 readers: disease prevalence in the CXR subset read by R1 increased from 77 % to 81 %, from 64 % to 77 % for R2, from 86 % to 90 % for R4, from 70 % to 86 % for R5, from 64 % to 77 % for R6, while decreasing from 85 % to 75 % for R3 and from 71 % to 43 % for R7. Group 1 readers attained an 87.2 % sensitivity (95 % CI 81.2 %–91.5 %), a 71.4 % specificity (95 % CI 58.5 %–81.6 %), an 83.2 % accuracy (95 % CI 77.7 %–87.5 %), an 89.9 % positive predictive value (95 % CI 84.3 %–93.7 %), a 65.6 % negative predictive value (95 % CI 53.0 %–76.3 %), a 3.05 positive likelihood ratio (95 % CI 2.01–4.64), and a 0.18 negative likelihood ratio (95 % CI 0.12–0.28) in the first half of all their reported CXRs, while in the second half they reached a 90.6 % sensitivity (95 % CI 85.3 %–94.2 %), a 59.1 % specificity (95 % CI 44.4 %–72.3 %), a 84.2 % accuracy (95 % CI 78.7 %–88.5 %), an 89.6 % positive predictive value (95 % CI 84.2 %–93.3 %), a 61.9 % negative predictive value (95 % CI 46.8 %–75.0 %), a 2.22 positive likelihood ratio (95 % CI 1.55–3.17), and a 0.16 negative likelihood ratio (95 % CI 0.09–0.27). Conversely, Group 2 readers had an 82.9 % sensitivity (95 % CI 67.3 %–91.9 %), a 43.8 % specificity (95 % CI 23.1 %–66.8 %), a 70.6 % accuracy (95 % CI 57.0 %–81.3 %), a 76.3 % positive predictive value (95 % CI 60.8 %–87.0 %), a 53.8 % negative predictive value (95 % CI 29.1 %–76.8 %), a 1.47 positive likelihood ratio (95 % CI 0.93–2.33), and a 0.39 negative likelihood ratio (95 % CI 0.16–0.98) in the first half of all their reported CXRs, while in the second half they showed a 94.7 % sensitivity (95 % CI 82.7 %–98.5 %), a 36.4 % specificity (95 % CI 15.2 %–64.6 %), a 81.6 % accuracy (95 % CI 68.6 %–90.0 %), an 83.7 % positive predictive value (95 % CI 70.0 %–91.9 %), a 66.7 % negative predictive value (95 % CI 30.0 %–90.3 %), a 1.49 positive likelihood ratio (95 % CI 0.95–2.34), and a



**Fig. 1.** Bedside chest x-rays of suspected COVID-19 patients as classified by the radiologist on duty and subsequently dichotomized for this study's purpose. Chest x-rays in afterwards-confirmed COVID-19 correctly (true positives, panel A and B) and incorrectly interpreted (false positives, panel C and D).

Panel A: mild bilateral patchy areas of ill-defined lung opacities with peripheral and lower zone distribution in an anteroposterior chest x-ray, correctly interpreted as positives by a radiologist with more than 10 years of experience

Panel B: diffuse and ill-defined lung opacities with interstitial thickening, mainly involving the right lung in an anteroposterior chest x-ray, correctly interpreted as positives by a radiologist with less than 10 years of experience.

Panel C: non-specific bronchial wall-thickening in a posteroanterior chest x-ray of an asthmatic patient, incorrectly interpreted as a finding suggestive for COVID-19 by a radiologist with more than 10 years of experience.

Panel D: bilateral interstitial thickening in an anteroposterior chest x-ray, incorrectly interpreted as a finding suggestive for COVID-19 by a radiologist with less than 10 years of experience.

0.14 negative likelihood ratio (95 % CI 0.03–0.69). Table 3 details performance indexes both overall and for each reader in the first and second half of their CXR subset, sensitivity, specificity, and accuracy being also plotted in Figs. 2, 3 and 4, respectively.

#### 4. Discussion

The role of CXR in COVID-19 imaging could be paramount in settings with temporarily- or permanently-limited RT-PCR availability, as anticipated by Murphy et al. [24], who also warned against potential low diagnostic performance of CXR when reported by non-dedicated chest

radiologists. Real-world data from this study, albeit conducted in a high-prevalence region and during a SARS-CoV-2 pandemic peak, seem to provide a better scenario, in which radiologists with less than 10 years of experience matched the 89.0 % sensitivity attained by radiologists with more than 10 years of experience, with similar disease prevalence in the CXR subsets read by each group (73 % versus 77 %, respectively). A non-negligible cost for Group 2 to attain such a sensitivity was a consistently lower specificity (41 %, 95 % CI 25 %–59 %) – a value similar to the pooled specificity reported for chest CT by a meta-analysis of 3 studies from non-high-epidemic areas and 2 studies from high-epidemic areas (37 %, 95 % CI 26 %–50 %) [14] – while Group 1 showed a smaller difference

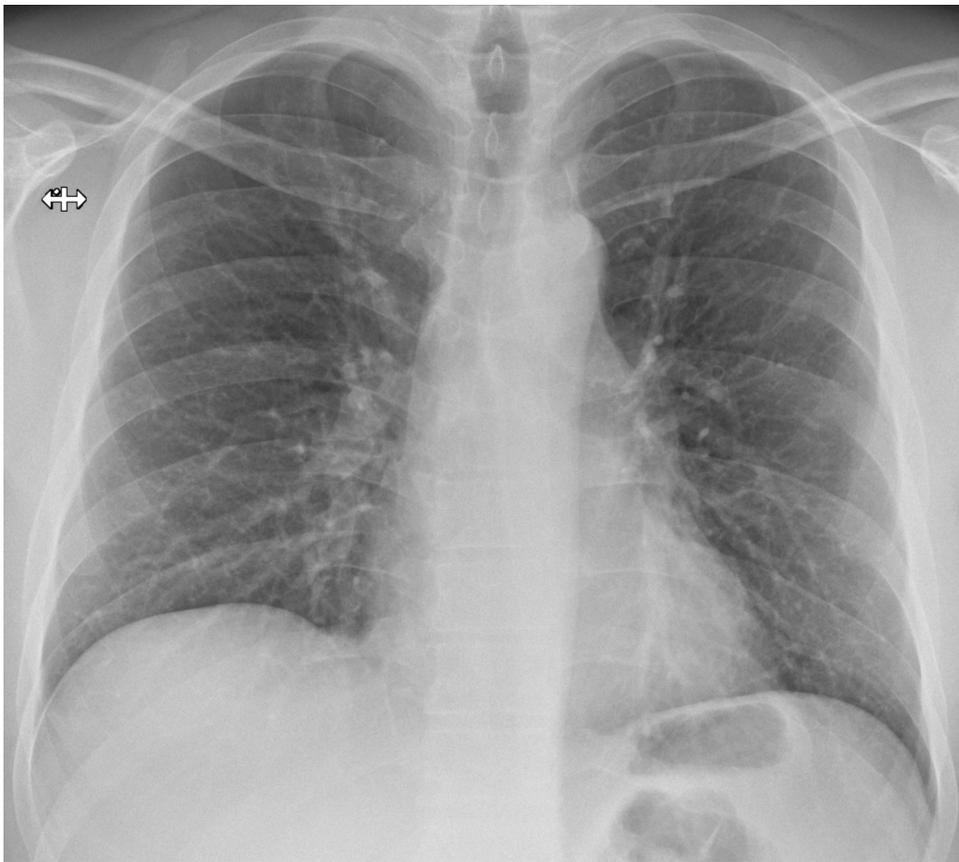


Fig. 1. (continued).



between sensitivity and specificity, with a constantly higher accuracy (Table 2). Such pattern was also observed comparing different timepoints or the total number of CXRs read by each radiologist: between the first and second half of the six-week study period overall accuracy increased from 76 % to 86 %, with corresponding increases both in Group 1 and Group 2; between the first and second half of CXRs read by each reader, overall accuracy increased from 81 % to 84 %, again with corresponding increases in

both groups, albeit more pronounced in the less experienced Group 2 (1% difference for Group 1, 11 % difference for Group 2). This trend was most likely driven in both groups by an adaptation to the escalation of examined cases (from 195 in the first three weeks to 340 in the following three), with an increase in sensitivity and accuracy mirrored by a specificity decrease. Of note, we can observe how in both groups there was a comparable number of readers exhibiting an inverse tendency towards a decrease in

**Table 1**  
Diagnostic performance indexes for chest x-ray reading for each radiologist and for the two experience-tiered groups.

	Disease prevalence	TP	TN	FP	FN	Sensitivity (95 % CI)	Specificity (95 % CI)	PPV (95 % CI)	NPV (95 % CI)	Accuracy (95 % CI)	
<b>Overall</b>	<b>77.0 %</b>	<b>298</b>	<b>66</b>	<b>34</b>	<b>37</b>	<b>89.0 %</b> (85.1 %–91.9 %)	<b>66.0 %</b> (56.3 %–74.5 %)	<b>89.8 %</b> (86.0 %–92.6 %)	<b>64.1 %</b> (54.5 %–72.7 %)	<b>83.7 %</b> (79.9 %–86.9 %)	
<b>Reader 1</b>	<b>78.9 %</b>	<b>131</b>	<b>26</b>	<b>12</b>	<b>11</b>	<b>92.3 %</b> (86.7 %–95.6 %)	<b>68.4 %</b> (52.5 %–80.9 %)	<b>91.6 %</b> (85.9 %–95.1 %)	<b>70.3 %</b> (54.2 %–82.5 %)	<b>87.2 %</b> (81.6 %–91.3 %)	
<b>Group 1</b>	<b>Reader 2</b>	<b>70.1 %</b>	<b>86</b>	<b>30</b>	<b>14</b>	<b>17</b>	<b>83.5 %</b> (75.1 %–89.4 %)	<b>68.2 %</b> (53.4 %–80.0 %)	<b>86.0 %</b> (77.9 %–91.5 %)	<b>63.8 %</b> (49.5 %–76.0 %)	<b>78.9 %</b> (71.6 %–84.7 %)
<b>Reader 3</b>	<b>80.0 %</b>	<b>47</b>	<b>5</b>	<b>8</b>	<b>5</b>	<b>90.4 %</b> (79.4 %–95.8 %)	<b>38.5 %</b> (17.7 %–64.5 %)	<b>85.5 %</b> (73.8 %–92.4 %)	<b>50.0 %</b> (23.7 %–76.3 %)	<b>80.0 %</b> (68.7 %–87.9 %)	
<b>Reader 4</b>	<b>88.4 %</b>	<b>34</b>	<b>5</b>	<b>0</b>	<b>4</b>	<b>89.5 %</b> (75.9 %–95.8 %)	<b>100.0 %</b> (56.6 %–100.0 %)	<b>100.0 %</b> (89.8 %–100.0 %)	<b>55.6 %</b> (26.7 %–81.1 %)	<b>90.7 %</b> (78.4 %–96.3 %)	
<b>Overall</b>	<b>73.0 %</b>	<b>65</b>	<b>11</b>	<b>16</b>	<b>8</b>	<b>89.0 %</b> (79.8 %–94.3 %)	<b>40.7 %</b> (24.5 %–59.3 %)	<b>80.2 %</b> (70.3 %–87.5 %)	<b>57.9 %</b> (36.3 %–76.9 %)	<b>76.0 %</b> (66.8 %–83.3 %)	
<b>Group 2</b>	<b>Reader 5</b>	<b>78.0 %</b>	<b>40</b>	<b>5</b>	<b>8</b>	<b>6</b>	<b>87.0 %</b> (74.3 %–93.9 %)	<b>38.5 %</b> (17.7 %–64.5 %)	<b>83.3 %</b> (70.4 %–91.3 %)	<b>45.5 %</b> (21.3 %–72.0 %)	<b>76.3 %</b> (64.0 %–85.3 %)
<b>Reader 6</b>	<b>70.4 %</b>	<b>19</b>	<b>3</b>	<b>5</b>	<b>0</b>	<b>100.0 %</b> (83.2 %–100.0 %)	<b>37.5 %</b> (13.7 %–69.4 %)	<b>79.2 %</b> (59.5 %–90.8 %)	<b>100.0 %</b> (43.9 %–100.0 %)	<b>81.5 %</b> (63.3 %–91.8 %)	
<b>Reader 7</b>	<b>57.1 %</b>	<b>6</b>	<b>3</b>	<b>3</b>	<b>2</b>	<b>75.0 %</b> (40.9 %–92.9 %)	<b>50.0 %</b> (18.8 %–81.2 %)	<b>66.7 %</b> (35.4 %–87.9 %)	<b>60.0 %</b> (23.1 %–88.2 %)	<b>64.3 %</b> (38.8 %–83.7 %)	

TP = true positives, TN = true negatives, FP = false positives, FN = false negatives, CI = confidence interval, PPV = positive predictive value, NPV = negative predictive value.

accuracy (Fig. 1) and sensitivity (Fig. 2), reinforced by a decrease in specificity in all but one less-experienced reader (Fig. 3).

Limitations of this study include its retrospective and monocentric nature, the fact that each radiologist read a different subset of images, and the imbalance in the number of CXRs read by Group 1 and Group 2, with the lesser-experienced Group 2 reading 18.6 % of all CXRs. However, the closely proportionate disease prevalence between the two groups substantiates the comparability of subsequent findings and seems to suggest a more pronounced influence of overall radiological experience on the diagnostic performance of each group. Such an hypothesis should be verified with a conventional multi-reader study, to ascertain if these differences

in diagnostic performance are also influenced by the number of COVID-19-positive CXRs read by each radiologist, or indeed result from a combination of these factors. However, we should also consider that any multi-reader study performed after a pandemic outbreak would not reproduce the condition of the first outbreak, when the new disease first spread in a country. Other than a conventional multi-reader study, further evaluations of real-world diagnostic performance should also target the potential impact on diagnostic performance of various types of subspecialty radiological training and of centre-specific contingencies, such as presence and employment of residents, different radiologists workloads, and disparities in CXR reporting conducted during day or night shifts. In addition, the

**Table 2**  
Diagnostic performance indexes for chest x-ray reading between radiologists' groups according to different patients' characteristics and timeframe.

	Disease prevalence	TP	TN	FP	FN	SN	SP	PPV	NPV	ACC	LR+	LR–
<b>All patients</b>												
<b>Overall</b>	<b>76.3 %</b>	<b>363</b>	<b>77</b>	<b>50</b>	<b>45</b>	<b>89.0 %</b>	<b>60.6 %</b>	<b>87.9 %</b>	<b>63.1 %</b>	<b>82.2 %</b>	<b>2.3</b>	<b>0.2</b>
G1 (Exp. > 10y)	77.0 %	298	66	34	37	89.0 %	66.0 %	89.8 %	64.1 %	83.7 %	2.6	0.2
G2 (Exp. < 10y)	73.0 %	65	11	16	8	89.0 %	40.7 %	80.2 %	57.9 %	76.0 %	1.5	0.3
<b>Male patients</b>												
<b>Overall</b>	<b>79.7 %</b>	<b>243</b>	<b>42</b>	<b>27</b>	<b>28</b>	<b>89.7 %</b>	<b>60.9 %</b>	<b>90.0 %</b>	<b>60.0 %</b>	<b>83.8 %</b>	<b>2.3</b>	<b>0.2</b>
G1 (Exp. > 10y)	81.2 %	202	35	17	22	90.2 %	67.3 %	92.2 %	61.4 %	85.9 %	2.8	0.1
G2 (Exp. < 10y)	73.4 %	41	7	10	6	87.2 %	41.2 %	80.4 %	53.8 %	75.0 %	1.5	0.3
<b>Female patients</b>												
<b>Overall</b>	<b>70.3 %</b>	<b>120</b>	<b>35</b>	<b>23</b>	<b>17</b>	<b>87.6 %</b>	<b>60.3 %</b>	<b>83.9 %</b>	<b>67.3 %</b>	<b>79.5 %</b>	<b>2.2</b>	<b>0.2</b>
G1 (Exp. > 10y)	69.8 %	96	31	17	15	86.5 %	64.6 %	85.0 %	67.4 %	79.9 %	2.4	0.2
G2 (Exp. < 10y)	72.2 %	24	4	6	2	92.3 %	40.0 %	80.0 %	66.7 %	77.8 %	1.5	0.2
<b>First three weeks (24/02–15/03)</b>												
<b>Overall</b>	<b>66.7 %</b>	<b>105</b>	<b>44</b>	<b>21</b>	<b>25</b>	<b>80.8 %</b>	<b>67.7 %</b>	<b>83.3 %</b>	<b>63.8 %</b>	<b>76.4 %</b>	<b>2.5</b>	<b>0.3</b>
G1 (Exp. > 10y)	67.3 %	83	37	13	20	80.6 %	74.0 %	86.5 %	64.9 %	78.4 %	3.1	0.3
G2 (Exp. < 10y)	64.3 %	22	7	8	5	81.5 %	46.7 %	73.3 %	58.3 %	69.0 %	1.5	0.4
<b>Second three weeks (16/03–08/04)</b>												
<b>Overall</b>	<b>81.8 %</b>	<b>258</b>	<b>33</b>	<b>29</b>	<b>20</b>	<b>92.8 %</b>	<b>53.2 %</b>	<b>89.9 %</b>	<b>62.3 %</b>	<b>85.6 %</b>	<b>2.0</b>	<b>0.1</b>
G1 (Exp. > 10y)	82.3 %	215	29	21	17	92.7 %	58.0 %	91.1 %	63.0 %	86.5 %	2.2	0.1
G2 (Exp. < 10y)	79.3 %	43	4	8	3	93.5 %	33.3 %	84.3 %	57.1 %	81.0 %	1.4	0.2

TP = true positives, TN = true negatives, FP = false positives, FN = false negatives, SN = sensitivity, SP = specificity, PPV = positive predictive value, NPV = negative predictive value, ACC = accuracy, LR+ = positive likelihood ratio, LR– = negative likelihood ratio, G1 = Radiologists' group 1, G2 = Radiologists' group 2, Exp. = experience.

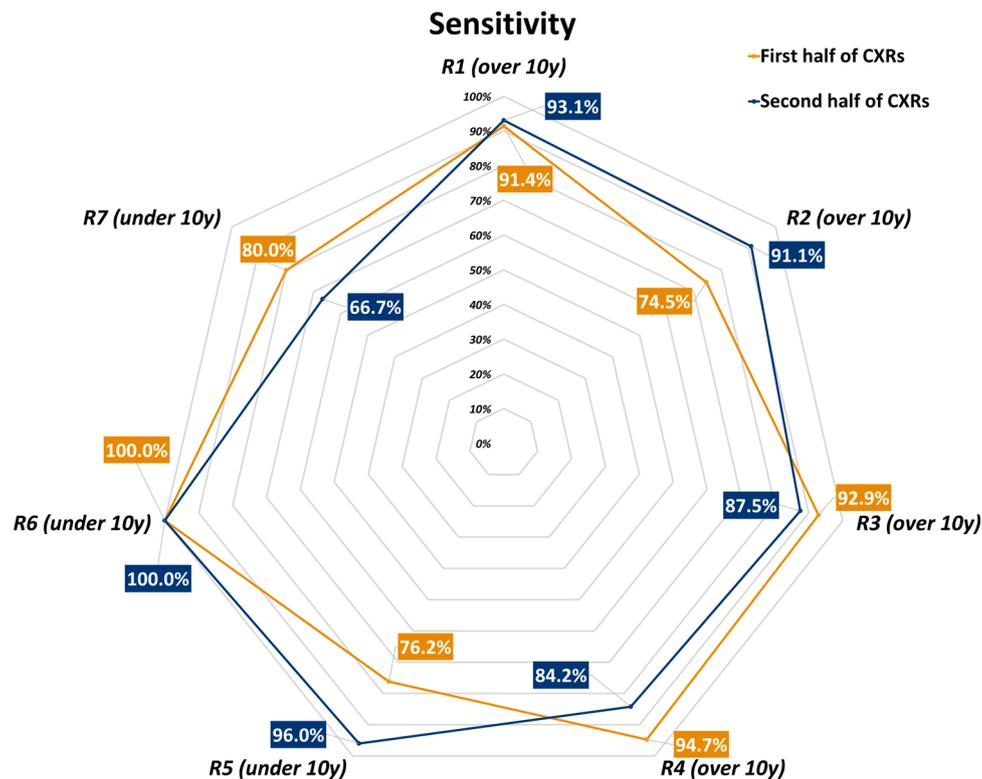
**Table 3**

Different diagnostic performance indexes for chest x-ray reading between the first and second half of interpreted chest x-rays for each reader and both radiologists' groups.

Group	Reader	Disease prevalence	TP	TN	FP	FN	SN	SP	PPV	NPV	ACC
<i>First half of chest x-rays for each reader</i>											
Overall		73.4 %	172	47	25	27	86.4 %	65.3 %	87.3 %	63.5 %	80.8 %
G1 (Exp. > 10y)	Reader 1	76.9 %	64	14	7	6	91.4 %	66.7 %	90.1 %	70.0 %	85.7 %
	Reader 2	63.5 %	35	20	7	12	74.5 %	74.1 %	83.3 %	62.5 %	74.3 %
	Reader 3	84.8 %	26	3	2	2	92.9 %	60.0 %	92.9 %	60.0 %	87.9 %
	Reader 4	86.4 %	18	3	0	1	94.7 %	100.0 %	100.0 %	75.0 %	95.5 %
G2 (Exp. < 10y)	Reader 5	70.0 %	16	4	5	5	76.2 %	44.4 %	76.2 %	44.4 %	66.7 %
	Reader 6	64.3 %	9	2	3	0	100.0 %	40.0 %	75.0 %	100.0 %	78.6 %
	Reader 7	71.4 %	4	1	1	1	80.0 %	50.0 %	80.0 %	50.0 %	71.4 %
<i>Second half of chest x-rays for each reader</i>											
Overall		79.2 %	191	30	25	18	91.4 %	54.5 %	88.4 %	62.5 %	83.7 %
G1 (Exp. > 10y)	Reader 1	80.9 %	67	12	5	5	93.1 %	70.6 %	93.1 %	70.6 %	88.8 %
	Reader 2	76.7 %	51	10	7	5	91.1 %	58.8 %	87.9 %	66.7 %	83.6 %
	Reader 3	75.0 %	21	2	6	3	87.5 %	25.0 %	77.8 %	40.0 %	71.9 %
	Reader 4	90.5 %	16	2	0	3	84.2 %	100.0 %	100.0 %	40.0 %	85.7 %
G2 (Exp. < 10y)	Reader 5	86.2 %	24	1	3	1	96.0 %	25.0 %	88.9 %	50.0 %	86.2 %
	Reader 6	76.9 %	10	1	2	0	100.0 %	33.3 %	83.3 %	100.0 %	84.6 %
	Reader 7	42.9 %	2	2	2	1	66.7 %	50.0 %	50.0 %	66.7 %	57.1 %

TP = true positives, TN = true negatives, FP = false positives, FN = false negatives, SN = sensitivity, SP = specificity, PPV = positive predictive value, NPV = negative predictive value, ACC = accuracy, G1 = Radiologists' group 1, G2 = Radiologists' group 2, Exp. = experience.

result herein reported should be considered in light of the pandemic peak – another disease (such as pleural effusion, pneumothorax, bacterial pneu-



**Fig. 2.** Sensitivity of each radiologist in each one's first and second half of interpreted chest x-rays.

with very high disease prevalence – and could be not reproducible in low prevalence settings [25,26]. Being this a real-world data study, our results rely on a practical dichotomization of CXR reports: their potential generalizability must be therefore very carefully considered, especially when, in case of suspected COVID-19, we have a non-typical CXR for SARS-CoV-2 pneumonia. Clinical translation of our findings would still result in at least two different scenarios, also taking into account the unspecific nature of CXR findings in COVID-19 pneumonia and other viral pneumonias. First, when a patient displays suspicious symptoms for COVID-19 that can however be justified by alternative pathological CXR findings pointing to

monia), the management of the patient would remain the one that would have normally been followed in the detected condition. Otherwise, if in a general situation of increased patient influx to emergency departments a patient presents with suspicious symptoms for COVID-19 but no suggestive CXR findings or other findings that can justify a COVID-19 diagnosis, the use of chest CT could be considered [2,14]. However, taking into account the suboptimal diagnostic performance of chest CT – in particular the potentially low specificity and positive predictive value [14] – if the patient's clinical conditions are stable and it is therefore possible to wait for RT-PCR confirmation of SARS-CoV-2 infection, preventive isolation would

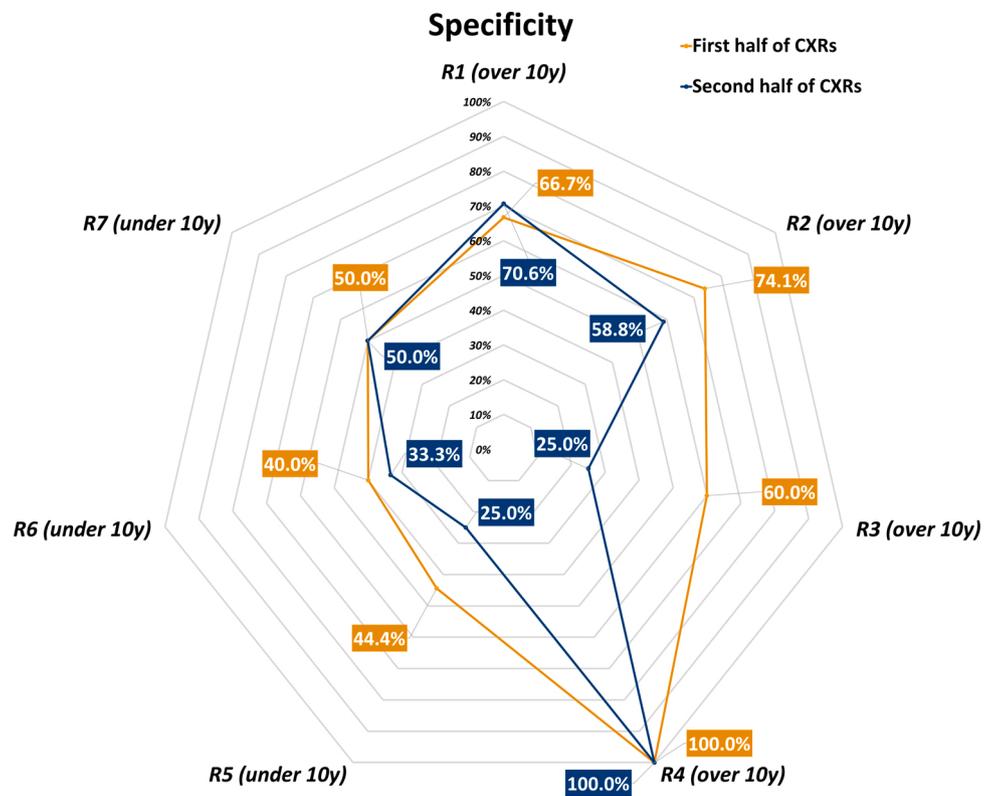


Fig. 3. Specificity of each radiologist in each one's first and second half of interpreted chest x-rays.

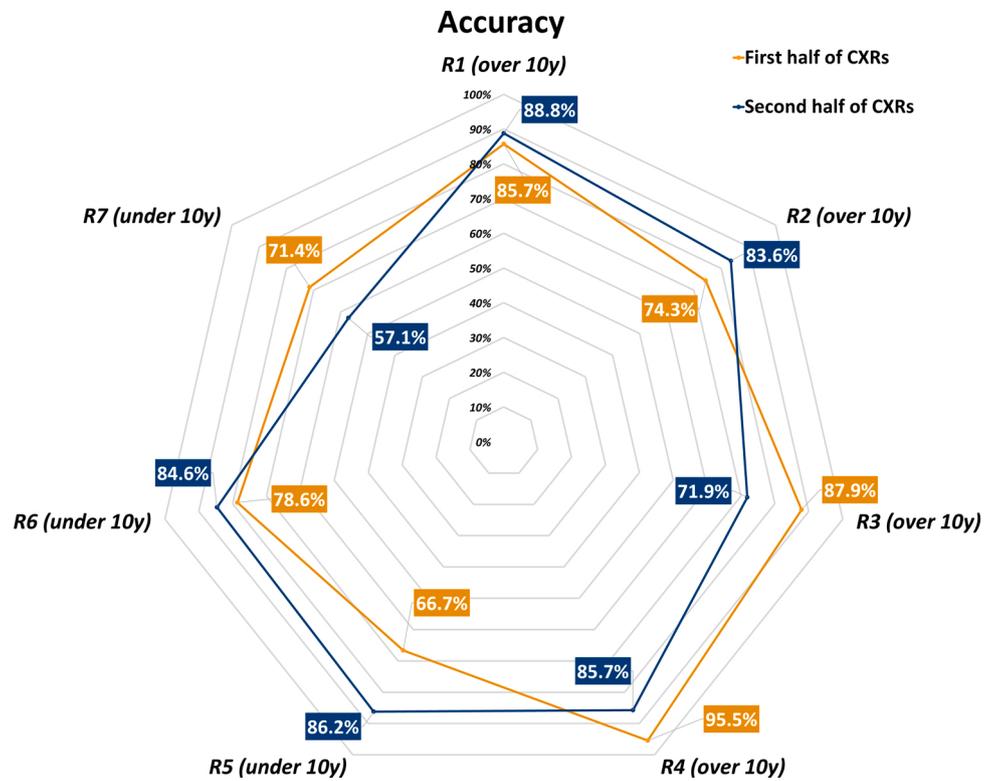


Fig. 4. Accuracy of each radiologist in each one's first and second half of interpreted chest x-rays.

remain the safest approach.

To summarize, the real-world diagnostic performance of CXR during the COVID-19 pandemic peak reached a relatively well-balanced overall accuracy (76%–86%), with an 89% sensitivity and a higher specificity for the more experienced radiologists (66%), lower for the less experienced radiologists (41%). Such data play in favour of the use of CXR as first line examination when chest imaging is required to aid the triage process of suspected COVID-19 patients during a pandemic peak.

### CRediT authorship contribution statement

**Andrea Cozzi:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Simone Schiaffino:** Conceptualization, Methodology, Investigation, Data curation, Validation, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Francesco Arpaia:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Gianmarco Della Pepa:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Stefania Tritella:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Pietro Bertolotti:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Laura Menicagli:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Cristian Giuseppe Monaco:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Luca Alessandro Carbonaro:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Riccardo Spairani:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Bijan Babaei Paskeh:** Investigation, Data curation, Writing - original draft, Writing - review & editing. **Francesco Sardanelli:** Conceptualization, Methodology, Validation, Resources, Funding acquisition, Supervision, Project administration, Writing - review & editing.

### Declaration of Competing Interest

A. Cozzi, F. Arpaia, G. Della Pepa, S. Tritella, P. Bertolotti, L. Menicagli, C.G. Monaco, L.A. Carbonaro, R. Spairani, and B. Babaei Paskeh, all declare that they have no conflict of interest and that they have nothing to disclose.

S. Schiaffino declares to have received travel support from Bracco Imaging and to be member of speakers' bureau for General Electric Healthcare.

F. Sardanelli declares to have received grants from or to be member of speakers' bureau/advisory board for Bayer Healthcare, Bracco, and General Electric Healthcare.

### Acknowledgements

This study was partially supported by Ricerca Corrente funding from Italian Ministry of Health to IRCCS Policlinico San Donato

### References

- [1] G.D. Rubin, C.J. Ryerson, L.B. Haramati, N. Sverzellati, J.P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I.B.K. Martin, D.J. Anderson, C. Kong, T. Altes, A. Bush, S.R. Desai, O. Goldin, J.M. Goo, M. Humbert, Y. Inoue, H.-U. Kauczor, F. Luo, P.J. Mazzone, M. Prokop, M. Remy-Jardin, L. Richeldi, C.M. Schaefer-Prokop, N. Tomiyama, A.U. Wells, A.N. Leung, The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society, *Radiology* 296 (2020) 172–180, <https://doi.org/10.1148/radiol.2020201365>.
- [2] E.A. Akl, I. Blazic, S. Yaacoub, G. Frijia, R. Chou, J.A. Appiah, M. Fatehi, N. Flor, E. Hitti, H. Jafri, Z.-Y. Jin, H.U. Kauczor, M. Kawooya, E.A. Kazerouni, J.P. Ko, R. Mahfouz, V. Muglia, R. Nyabanda, M. Sanchez, P.B. Shete, M. Ulla, C. Zheng, E. Van Deventer, M. Del R. Perez, use of chest imaging in the diagnosis and management of COVID-19: a WHO rapid advice guide, *Radiology* (2020), <https://doi.org/10.1148/radiol.2020203173>.
- [3] M. Dramé, M.T. Teguo, E. Proye, F. Hequet, M. Hentzien, L. Kanagaratnam, L. Godaert, Should RT-PCR be considered a gold standard in the diagnosis of Covid-19? *J. Med. Virol.* (2020) <https://doi.org/10.1002/jmv.25996>.
- [4] R. Liu, H. Han, F. Liu, Z. Lv, K. Wu, Y. Liu, Y. Feng, C. Zhu, Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020, *Clin. Chim. Acta* 505 (2020) 172–175, <https://doi.org/10.1016/j.cca.2020.03.009>.
- [5] C. Long, H. Xu, Q. Shen, X. Zhang, B. Fan, C. Wang, B. Zeng, Z. Li, X. Li, H. Li, Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *Eur. J. Radiol.* 126 (2020) 108961, <https://doi.org/10.1016/j.ejrad.2020.108961>.
- [6] M.J. Loeffelholz, Y.-W. Tang, Laboratory diagnosis of emerging human coronavirus infections – the state of the art, *Emerg. Microbes Infect.* 9 (2020) 747–756, <https://doi.org/10.1080/22221751.2020.1745095>.
- [7] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K.W.K. Chu, T. Bleicker, S. Brünink, J. Schneider, M.L. Schmidt, D.G.J.C.G. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.-L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M.P.G. P. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, *Eurosurveillance*. 25 (2020) 2000045, <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
- [8] G.M. Bwire, M.V. Majigo, B.J. Njiro, A. Mawazo, Detection profile of SARS-CoV-2 using RT-PCR in different types of clinical specimens: a systematic review and meta-analysis, *J. Med. Virol.* (2020), <https://doi.org/10.1002/jmv.26349>.
- [9] A. Tahamtan, A. Ardebili, Real-time RT-PCR in COVID-19 detection: issues affecting the results, *Expert Rev. Mol. Diagn.* 20 (2020) 453–454, <https://doi.org/10.1080/14737159.2020.1757437>.
- [10] M. Mossa-Basha, C.C. Meltzer, D.C. Kim, M.J. Tuite, K.P. Kolli, B.S. Tan, Radiology department preparedness for COVID-19: radiology scientific expert review panel, *Radiology* 296 (2020) E106–E112, <https://doi.org/10.1148/radiol.2020200988>.
- [11] N. Sverzellati, G. Milanese, F. Milone, M. Balbi, R.E. Ledda, M. Silva, Integrated radiologic algorithm for COVID-19 pandemic, *J. Thorac. Imaging* 35 (2020) 228–233, <https://doi.org/10.1097/RTI.0000000000000516>.
- [12] Q. Yang, Q. Liu, H. Xu, H. Lu, S. Liu, H. Li, Imaging of coronavirus disease 2019: a Chinese expert consensus statement, *Eur. J. Radiol.* 127 (2020) 109008, <https://doi.org/10.1016/j.ejrad.2020.109008>.
- [13] Z. Sun, N. Zhang, Y. Li, X. Xu, A systematic review of chest imaging findings in COVID-19, *Quant. Imaging Med. Surg.* 10 (2020) 1058–1079, <https://doi.org/10.21037/qims-20-564>.
- [14] H. Kim, H. Hong, S.H. Yoon, Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: a meta-analysis, *Radiology* 296 (2020) E145–E155, <https://doi.org/10.1148/radiol.2020201343>.
- [15] Y. Zhao, C. Xiang, S. Wang, C. Peng, Q. Zou, J. Hu, Radiology department strategies to protect radiologic technologists against COVID-19: experience from Wuhan, *Eur. J. Radiol.* 127 (2020) 108996, <https://doi.org/10.1016/j.ejrad.2020.108996>.
- [16] M. Zanardo, S. Schiaffino, F. Sardanelli, Bringing radiology to patient's home using mobile equipment: A weapon to fight COVID-19 pandemic, *Clin. Imaging* 68 (2020) 99–101, <https://doi.org/10.1016/j.clinimag.2020.06.031>.
- [17] S. Kooraki, M. Hosseiny, L. Myers, A. Gholamrezaezhad, Coronavirus (COVID-19) outbreak: what the department of radiology should know, *J. Am. Coll. Radiol.* 17 (2020) 447–451, <https://doi.org/10.1016/j.jacr.2020.02.008>.
- [18] N. Flor, R. Dore, F. Sardanelli, On the Role of Chest Radiography and CT in the Coronavirus Disease (COVID-19) Pandemic, *Am. J. Roentgenol* (2020), <https://doi.org/10.2214/AJR.20.23411>.
- [19] H.Y.F. Wong, H.Y.S. Lam, A.H.-T. Fong, S.T. Leung, T.W.-Y. Chin, C.S.Y. Lo, M.M.-S. Lui, J.C.Y. Lee, K.W.-H. Chiu, T.W.-H. Chung, E.Y.P. Lee, E.Y.F. Wan, I.F. N. Hung, T.P.W. Lam, M.D. Kuo, M.-Y. Ng, Frequency and distribution of chest radiographic findings in patients positive for COVID-19, *Radiology* 296 (2020) E72–E78, <https://doi.org/10.1148/radiol.2020201160>.
- [20] A. Pakray, D. Walker, A. Figacz, S. Kilanowski, C. Rhodes, S. Doshi, M. Coffey, Imaging evaluation of COVID-19 in the emergency department, *Emerg. Radiol.* (2020), <https://doi.org/10.1007/s10140-020-01787-0>.
- [21] H.W. Kim, K.M. Capaccione, G. Li, L. Luk, R.S. Widemon, O. Rahman, V. Beylergil, R. Mitchell, B.M. D'Souza, J.S. Leb, S. Dumeer, S. Bentley-Hibbert, M. Liu, S. Jambawalikar, J.H.M. Austin, M. Salvatore, The role of initial chest X-ray in triaging patients with suspected COVID-19 during the pandemic, *Emerg. Radiol.* (2020), <https://doi.org/10.1007/s10140-020-01808-y>.
- [22] D. Ippolito, A. Pecorelli, C. Maino, C. Capodaglio, I. Mariani, T. Giandola, D. Gandola, I. Bianco, M. Ragusi, C. Talei Franzesi, R. Corso, S. Sironi, Diagnostic impact of bedside chest X-ray features of 2019 novel coronavirus in the routine admission at the emergency department: case series from Lombardy region, *Eur. J. Radiol.* 129 (2020) 109092, <https://doi.org/10.1016/j.ejrad.2020.109092>.
- [23] S. Schiaffino, S. Tritella, A. Cozzi, S. Carriero, L. Blandi, L. Ferraris, F. Sardanelli, Diagnostic Performance of Chest X-Ray for COVID-19 Pneumonia During the SARS-CoV-2 Pandemic in Lombardy, Italy, *J. Thorac. Imaging* 35 (2020) W105–W106, <https://doi.org/10.1097/RTI.0000000000000533>.
- [24] K. Murphy, H. Smits, A.J.G. Knoop, M.B.J.M. Korst, T. Samson, E.T. Scholten, S. Schalekamp, C.M. Schaefer-Prokop, R.H.H.M. Philippen, A. Meijers, J. Melendez, B. Van Ginneken, M. Rutten, COVID-19 on the chest radiograph: a multi-reader evaluation of an AI system, *Radiology* 296 (2020) E166–E172, <https://doi.org/10.1148/radiol.2020201874>.
- [25] F. Sardanelli, G. Di Leo, Assessing the Value of Diagnostic Tests in the Coronavirus Disease 2019 Pandemic, *Radiology* 296 (2020) E193–E194, <https://doi.org/10.1148/radiol.2020201845>.
- [26] M.M.G. Leeflang, A.W.S. Rutjes, J.B. Reitsma, L. Hooft, P.M.M. Bossuyt, Variation of a test's sensitivity and specificity with disease prevalence, *Can. Med. Assoc. J.* 185 (2013) E537–E544, <https://doi.org/10.1503/cmaj.121286>.