

## **SUPPLEMENTARY MATERIAL**

### **CXR reconstructed from Average Intensity Projection-CT imaging**

The average intensity projection (AIP) algorithm was applied to the coronal reformatted HRCT scans on the local Picture Archiving and Communication System (PACS) workstation (suite Estensa, Esaote, Genova, Italy) by a senior chest radiologist (N.S.) with 16 years of experience. The slab thickness was adjusted to the individual chest size, notably accounting for the maximum AP torso diameter at the level of posterior costophrenic sulci (range 18-34 cm) (see video clip). The reconstructed (r-) CXR images were stored on the same PACS workstation for the study observers evaluation.

In order to evaluate the diagnostic value of the CXR-like imaging reconstructed by the AIP technique (r-CXR), a chest radiologist (M.S., with 9 years of experience) blindly compared 42 CXRs of subjects with rt-PCR confirmed COVID-19 pneumonia incidentally identified in the diagnostic work-up for other clinical indications with the r-CXRs reconstructed from the HRCT scans (CXR and CT were acquired on the same day according to variable clinical inquiries). First, the observer read r-CXR (the surrogate experimental tool) for the presence of individual abnormalities such as consolidation, nodules, ground glass opacity, reticular abnormalities and pleural effusion. Second, the observer compared the r-CXR and CXR side-by-side and recorded adjunct signs that were visible on CXR (the standard of reference in use for clinical practice). This process was structured on a case-by-case basis to avoid the intra-observer bias potentially generated by an independent scoring of the experimental tool and the standard of reference. The consistency of detection for each individual finding was given as the ratio between its detection on r-CXR and that on standard CXR, as follows: consolidation (10/12), reticular opacities (5/7), ground glass (7/8). No pleural effusion was observed.

The observer was also asked to provide her/his visual impression on the quality of r-CXR images, using the standard CXR as standard of reference. As compared to the standard CXR, the r-CXR was classified as follows: very similar (31/42), slightly different (7/42), different (2/42), remarkably different (2/42). r-CXR were all considered very similar to corresponding standard CXR (Fig. 1s, and 2s).

### **Computed tomography technique**

Non-contrast HRCT was performed with either a 128-slice scanner (SOMATOM Definition Edge, Siemens Healthineers, Erlangen, Germany) or a 16-slice mobile scanner on truck (SOMATOM Emotion, Siemens Healthineers, Erlangen, Germany). HRCT images were acquired with the patient in the supine position during end-inspiration breath-hold, without intravenous contrast material. The acquisition parameters were set at 110-120 kVp, 80 reference mAs, pitch 0.9–1.2, and collimation 0.625–1.0 mm. Reconstruction parameters for lung images: slice thickness 1.0 mm, increment 0.7-1.0 mm, sharp reconstruction algorithm (BI57 or B70s, respectively), lung window (width, 1600 HU; level, –600 HU). Reconstruction parameters for mediastinal images: slice thickness 2.0 mm, increment 1.5 mm, medium reconstruction algorithm (Br36 or B31s, respectively), mediastinal window (width, 400 HU; level, 30 HU). Sinogram affirmed iterative reconstruction (SAFIRE) strength 3 on Definition Edge scanner, filtered back projection (FBP) on Emotion scanner.

### **Visual assessment of the HRCT scans**

The extent of combined GGO and consolidation was visually scored at the nearest 5% on the whole lungs. The distribution was described as follows: a) axial distribution: predominantly peripheral (within the outer third of the lung), predominantly central, or mixed; b) cranio-caudal distribution: predominantly upper (above the carina), middle (between the carina and the right inferior pulmonary vein) or lower (below the inferior pulmonary vein) (1); c) bilateral or unilateral involvement; d) lobar involvement was accounted over 6 lobes (lingula considered as a single lobe). Description of the pattern was also tabulated into the various HRCT categories of our local COVID-19 protocol (2). These categories aimed to define disease severity by encompassing both morphology and extent of parenchymal findings, as follows: 1) non-COVID-19 findings, 2) findings indeterminate for COVID-19, either because of differential or overlapping disease, 3) typical pattern COVID-19, including different combinations of GGO and consolidations (from fluffy consolidation to organized pneumonia and their overall extent (Table 1).

The interobserver variability was tested against a second radiologist observer (M.S., with 9-year experience in imaging of interstitial lung disease).

### **Study observer pulmonologist characteristics**

The study observers ranged in age from 45 to 55 years, and their clinical experience ranged in age from 12 to 20 years. Three of the observers were pulmonologists with long-time (15 to 20 years) expertise in interstitial lung diseases. Two observers (E.A.R. and P.S.) work in hospitals using a CXR-based protocol triage (Royal Brompton Hospital from United Kingdom, and Azienda Ospedaliera di Padova from Italy, respectively), while one of them (F.B.) works in a hospital using a CT-based protocol triage (Klinik für Pneumologie from Essen, Germany). The other two study observers included one anesthesiologist (A.V.) and one emergency physician (I.C) who both work in a CT-based protocol triage at the University Hospital of Parma (Italy).

### **Derivation cohort characteristics**

Among the 200 patients with oxygen saturation  $\leq 95\%$ , fever (79.5%) and cough (56.5%) were the most frequent symptoms (dyspnea 44.5%). A total of 89/200 (44.5%) subjects had moderate pulmonary involvement (oxygen saturation 95-93%), 74 (37%) patients had severe pulmonary involvement (oxygen saturation 88-92%), 37 (18.5%) patients had critical pulmonary involvement (oxygen saturation  $<88$ ). Cardiovascular disease (56.5%), and chronic pulmonary disease (17.5%) were the most frequent comorbidities. 69 patients (34.5%) died after a mean time of 10 days.

Among the 100 patients with oxygen saturation 96-98%, fever (72%) and cough (58%) were most frequent symptoms (dyspnea 14%). Cardiovascular disease (43%), and chronic pulmonary disease (8%) were the most frequent comorbidities. 12 patients (12%) died after a mean time of 12 days.

### **Validation cohort characteristics**

In the 73 patients with oxygen saturation  $\leq 95\%$ , fever (94%) and dyspnea (62%) were the most common symptoms (cough 48%). The majority of the patients (34/73, 46%) showed severe pulmonary involvement (oxygen saturation 88-92%); critical pulmonary involvement (oxygen saturation  $<88\%$ ) was observed in 21 (29%) patients, while moderate pulmonary involvement (oxygen saturation 95-93%) in 18 (25%) cases. The most frequent comorbidities were cardiovascular disease (68.5%), diabetes (18%), and cancer (18%). Death occurred in 29 (40%) patients in a mean time of 9 days.

Among 31 patients with oxygen saturation >95%, fever (87.5%) and dyspnea (62.5%) were most frequent symptoms (cough 25%). The most common comorbidity was cardiovascular disease (54%), followed by diabetes (21%), cancer (17%), and kidney disease (17%). During a mean time of 12 days, death occurred in 8 (26%) patients.

r-CXRs were obtained from corresponding HRCT scans using the same method described for the study cohort, with a different PACS system tool (Syngo.plaza, Siemens, Erlangen, Germany).

### **Interobserver variability**

There was good ( $k_w = 0.74$ ; 95% CI: 0.67-0.81) interobserver agreement for the r-CXR diagnostic categories, and moderate ( $k_w = 0.49$ ; 95% CI: 0.40-0.57) agreement between r-CXR and HRCT diagnostic categories. Following the conversion of the imaging extent scores into a three point scale (for the prognostic scoring systems) - <20%; 20-50%; >50% - in subjects with either indeterminate or typical appearance on r-CXR, the interobserver variability was good for r-CXR (0.77; 95% CI 0.71-0.82) and excellent for HRCT ( $K_w = 0.85$ ; 95% CI: 0.80-0.90).

### **r-CXR vs. HRCT**

Diagnostic categories for COVID-19 pneumonia for r-CXR were 85/300 (28.3%) normal, 7 (2.3%) alternative diagnosis, 40 (13.3%) indeterminate, and 168 (56%) typical. Diagnostic categories for HRCT were 40 (13.3%) normal, 35 (11.7%) alternative diagnosis, 41 (13.7%) indeterminate, and 184 (61.3%) typical. The extent of disease showed moderate correlation between r-CXR and HRCT ( $R^2 = 0.23$ ,  $p < 0.0005$ ).

### **Survival analyses in the derivation cohort**

Mortality did not differ significantly, between patients with imaging appearances typical for COVID-19 pneumonia and those with indeterminate appearance, a finding applying equally to r-CXR and HRCT evaluation (further data are reported in the section 'survival analyses in the derivation cohort' of the supplementary material). In order to simulate clinical practice, survival analyses were

evaluated in patients with r-CXR findings compatible with COVID-19 pneumonia (e.g. either indeterminate or typical appearance for COVID-19 pneumonia).

In a logistic regression model, the four point semi-categorical r-CXR grade (e.g. normal, alternative diagnosis, indeterminate, typical) was linked to mortality (OR 1.45; 95% CI 1.14, 1.84;  $p < 0.005$ ). With compression of the r-CXR grade, comparing patients without CXR evidence of COVID (normal and alternative diagnosis) to those with r-CXR findings compatible with COVID (indeterminate and typical), COVID compatible features on r-CXR were associated with mortality (OR 3.09; 95% CI 1.53, 6.22;  $p < 0.005$ ). This finding was robust (OR 2.93; 95% CI 1.39, 6.16;  $p < 0.005$ ) after adjustment for age, which was also associated with mortality (OR 1.07; 95% CI 1.04, 1.09;  $p < 0.0005$ ).

In a logistic regression model, the four point semi-categorical HRCT grade was linked to mortality (OR 1.46; 95% CI 1.10, 1.94;  $p < 0.01$ ). With compression of the HRCT grade, comparing patients without HRCT evidence of COVID (normal and alternative diagnosis) to those with HRCT findings compatible with COVID (indeterminate and typical), COVID compatible features on HRCT were associated with mortality (OR 2.90; 95% CI 1.36, 6.17;  $p < 0.01$ ). This finding was robust (OR 4.00; 95% CI 1.73, 9.27;  $p < 0.001$ ) after adjustment for age, which was also associated with mortality (OR 1.07; 95% CI 1.05, 1.10;  $p < 0.0005$ ).

## Discussion

The lack of information from the observers about the reasons for requesting HRCT represents a study limitation that only became obvious at the analysis stage. The most likely explanation for the variable request of HRCT scans is the variability in the degree of mismatch between CXR findings and clinical data that would trigger the need for the most informative imaging diagnostic test to support challenging clinical decisions. However, the limitations of CXR with respect to CT are still largely unknown in the evaluation of suspected COVID-19 pneumonia. In a study evaluating the frequency and distribution of CXR findings in COVID-19 positive patients, only one out of four (25%) subjects had false negative CXR as compared to CT (3). In our cohort, the proportion of false negative r-CXRs was more than half of negative r-CXRs. Such discrepant findings may be related to several factors such as the larger study cohort, the high proportion of subjects with dominant ground-glass opacification on HRCT in this group of subjects with false-negative r-CXRs (86.7%), and potential diagnostic limitations of the r-CXR technique. Indeed, the higher frequency of abnormalities (mostly typical for COVID-19 pneumonia) of variable extent on HRCT reports would explain the higher frequency of hospitalizations among the observers clinical decisions at the HRCT-based protocol round.

The five point age/imaging scales provided very similar prognostic information for r-CXR and HRCT evaluation, which was reproduced in the validation cohort. These findings are also consistent with prior studies showing that the visual score of disease extent on CXR is independently predictive of outcome (e.g intubation, or mortality) (4-6). The advantage of this simplified approach is that other variables that had prognostic value on unadjusted analysis did not add materially to the accuracy of prognostic evaluation, quantified using ROC values. In particular, the comorbidity score was associated with mortality when examined in isolation, as observed in other cohorts, but was no longer significant when age and disease extent were taken into account. This likely reflects the association between comorbidities and age, as the comorbidity score remained significant when age was excluded from the multivariable analysis.

Given the impossibility of running a two arm-randomized-controlled trial between CXR and CT in the COVID-19 setting, we sought to retrospectively compare the two modalities by artificially reconstructing a bidimensional image from coronal HRCT scans that was very similar to a standard bedside CXR. The AIP – a post-processing algorithm available on most CT workstations and dicom viewers – easily and rapidly allowed for r-CXR images that showed individual COVID-related

findings very similar to those observed on standard CXR. Yet, the r-CXR may be less accurate of standard upright CXR that is obtained for subjects with suspicious COVID-19 pneumonia in the pandemic setting. Nevertheless, the good interobserver agreement for the interpretation of r-CXR, the levels of sensitivity and specificity in keeping with standard CXR, as well as the significant prognostic value of both diagnostic categories and disease extent on r-CXR further substantiate the utility of this surrogate tool. Furthermore, any potential limitations related to this post-processing technique was mitigated by the fact that the study triagers were informed that r-CXR reports were identical to those achievable from standard CXR.

## References

1. Zhou S, Wang Y, Zhu T, Xia L. CT Features of Coronavirus Disease 2019 (COVID-19) Pneumonia in 62 Patients in Wuhan, China. *AJR Am J Roentgenol*. 2020:1-8.
2. Sverzellati N, Milanese G, Milone F, Balbi M, Ledda RE, Silva M. Integrated Radiologic Algorithm for COVID-19 Pandemic. *Journal of thoracic imaging*. 2020.
3. Wong HYF, Lam HYS, Fong AH, Leung ST, Chin TW, Lo CSY, et al. Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. *Radiology*. 2019:201160.
4. Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, et al. Clinical and Chest Radiography Features Determine Patient Outcomes In Young and Middle Age Adults with COVID-19. *Radiology*. 2020:201754.
5. 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.
6. Schalekamp S, Huisman M, van Dijk RA, Boomsma MF, Freire Jorge PJ, de Boer WS, et al. Model-based Prediction of Critical Illness in Hospitalized Patients with COVID-19. *Radiology*. 2020:202723.

**Supplementary Table S1. Kendall's coefficient of concordance for management choices of the five readers for the overall study population, and for subgroups based on prespecified oxygen saturation levels.**

	<b>r-CXR</b>	<i>p-values</i>	<b>HRCT</b>	<i>p-values</i>
Overall study population	0.365	<0.001	0.654	<0.001
Oxygen saturation 96 – 98%	0.375	<0.001	0.607	<0.001
Oxygen saturation 93 – 98%	0.367	<0.001	0.611	<0.001
Oxygen saturation ≤95%	0.278	<0.001	0.655	<0.001
Oxygen saturation ≤92%	0.236	0.093	0.692	<0.001

**Supplementary Table S2. Inter-observer agreement for the whole study population based on r-CXR-round and HRCT-round (weighted kappa coefficients).**

		<b>r-CXR</b>				
	<b>Reader</b>	<b>R<sub>1</sub></b>	<b>R<sub>2</sub></b>	<b>R<sub>3</sub></b>	<b>R<sub>4</sub></b>	<b>R<sub>5</sub></b>
<b>HRCT</b>	<b>R<sub>1</sub></b>		0.23 (0.15 - 0.31)	0.23 (0.15 - 0.31)	0.28 (0.20-0.37)	0.17 (0.11 - 0.24)
	<b>R<sub>2</sub></b>	0.31 (0.23 - 0.40)		0.34 (0.25 - 0.42)	0.51 (0.42 - 0.60)	0.31 (0.23 - 0.39)
	<b>R<sub>3</sub></b>	0.44 (0.36 - 0.52)	0.51 (0.44 - 0.59)		0.27 (0.18 - 0.35)	0.19 (0.12 - 0.26)
	<b>R<sub>4</sub></b>	0.33 (0.24 - 0.41)	0.75 (0.68 - 0.81)	0.57 (0.49 - 0.65)		0.13 (0.05 - 0.21)
	<b>R<sub>5</sub></b>	0.23 (0.16 - 0.30)	0.59 (0.52 - 0.67)	0.54 (0.46 - 0.61)	0.57 (0.49 - 0.65)	



**Supplementary Table S3. Intra-observer agreement between management according to r-CXR and HRCT.**  
**Subjects for whom chest CT was recommended during r-CXR-round were excluded on a reader basis.**

	$k_w$	$n$
<b>R<sub>1</sub></b>	0.59 (0.50-0.69)	235
<b>R<sub>2</sub></b>	0.71 (0.62-0.80)	227
<b>R<sub>3</sub></b>	0.37 (0.27-0.47)	276
<b>R<sub>4</sub></b>	0.53 (0.42-0.65)	233
<b>R<sub>5</sub></b>	0.54 (0.44-0.64)	162

**Supplementary Table S4. Clinical characteristics, presentation vitals and laboratory results of subjects potentially managed with CXR only and those for whom at least one reader requested HRCT.**

	r-CXR-only (76)	Further work-up by HRCT (224)	p-value
<b>Age (mean, SD)</b>	65.2 ( $\pm$ 18)	67.4 ( $\pm$ 15)	0.40
<b>Gender</b>	47 M (62%); 29 F (38%)	141 M (63%); 83 F (37%)	0.86
<b>Contact with COVID-19 cases</b>			
- Yes	41 (54%)	31 (14%)	< 0.001
- No	6 (8%)	43 (19%)	
- Not reported	29 (38%)	150 (67%)	
<b>History of Fever</b>	54 (71.1%)	177 (79%)	0.82
- Days with fever (median, IQR)	5 (3 - 7)	7 (3 - 8)	0.28
<b>History of Cough</b>	43 (56.6%)	128 (57%)	0.36
- Days with cough (median, IQR)	6 (3 - 8)	5 (3 - 7)	0.89
<b>History of Dyspnoea</b>	25 (33.9%)	84 (38%)	0.20
- Days with dyspnoea (median, IQR)	5 (2.5 - 7)	3 (3 - 7)	0.59
<b>Other symptoms</b>	22 (28.9%)	76 (34%)	
<b>Hypertension</b>	27 (35.5%)	87 (39%)	0.61
<b>Diabetes</b>	8 (10.5%)	34 (15)	0.31
<b>Obesity</b>	2 (0.3%)	17 (8)	0.12
<b>Comorbidities</b>			
- Cancer	6 (7.9%)	27 (12)	0.32
- Respiratory disease	10 (13.1%)	34 (15%)	0.67
- Kidney disease	2 (2.6%)	10 (4%)	0.48
- Liver disease	0 (0%)	1 (0.4%)	0.56
<b>Temperature <math>\geq 37.5^{\circ}\text{C}</math></b>	18 (23.7%)	76 (34%)	
- Temperature, median (IQR)	36.7 (36.5 - 37.9)	37 (36.5 - 37.7)	0.84
<b>Oxygen saturation</b>			
- $\leq 95$	38 (50%)	162 (72%)	< 0.001
- % Median (IQR)	95.5 (90 - 97)	94 (90 - 96)	0.009
- Received supplemental oxygen at triage	7 (9.2%)	8 (4%)	
<b>Respiratory Rate <math>\geq 25</math> breaths/min</b>	16 (21.1%)	58 (26%)	0.52
<b>Heart Rate <math>&gt; 100</math> beats/min</b>	11 (14.5%)	37 (17%)	0.55
<b>Laboratory measures (median, IQR)</b>			
- White blood cell count, $\times 10^9/\text{L}$	6.82 (4.67 - 10.47)	6.85 (4.73 - 9)	0.71
- Sodium, mmol/L	137 (135 - 139)	137 (135 - 139)	0.62
- Aspartate aminotransferase, U/L	49 (36.5 - 72.5)	39.5 (32 - 56)	0.06
- Alanine aminotransferase, U/L	28.5 (19.5 - 44)	28 (18.5 - 40)	0.57
- C-reactive protein, mg/L	83.9 (17.1 - 136.4)	76.1 (29.9 - 144.7)	0.38

**Supplementary Table S5. Variables independently associated with mortality in separate stepwise logistic regression models including a) CXR extent grade (1-3); b) CT extent grade (1-3); c) limited/extensive disease with combined CXR/CT extent (1-2). Odds ratios (OR) are stated with 95% confidence intervals in parentheses and statistical significance (p values). Model accuracy for is quantified from ROC curves and re-examined with the exclusion of duration of symptoms. Gender and the cormorbidity score were not retained in stepwise models.**

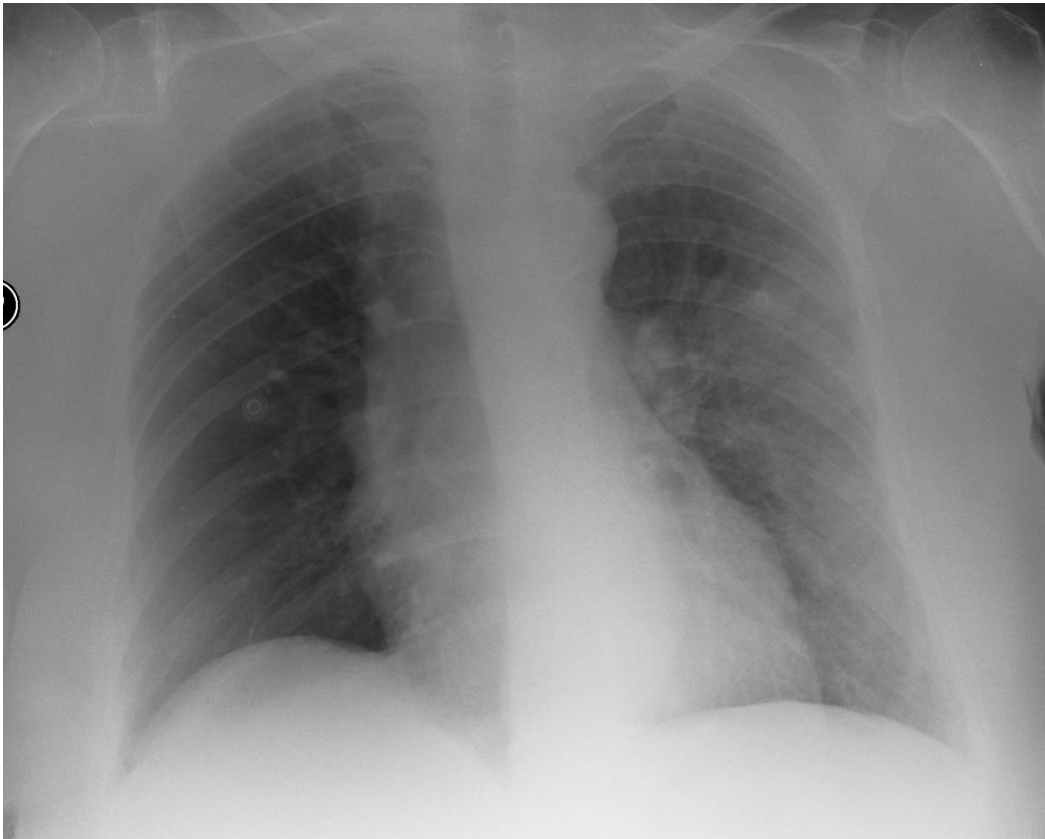
	Extent of disease	Age (years)	Duration of symptoms (days)	ROC values
<b>Index cohort</b>				
CXR extent model	OR = 2.76 (1.76, 4.10) p<0.0005	OR = 1.06 (1.03, 1.10) p<0.0005	OR = 0.90 (0.82, 0.99) p=0.03	0.81
Simplified CXR extent model	OR = 2.61 (1.69, 4.02) p<0.0005	OR=1.02 (1.05, 1.11) p<0.0005		0.79
CT extent model (ROC = 0.82)	OR = 2.85 (1.73, 4.70) p<0.0005	OR = 1.06 (1.03, 1.10) p<0.0005	OR = 0.90 (0.81, 0.98) p=0.02	0.82
Simplified CT extent model	OR = 2.63 (1.62, 4.28)	OR = 1.07 (1.04, 1.11)		0.80
Limited/extensive disease model	OR = 6.62 (3.05, 14.34) p<0.0005	OR = 1.06 (1.03, 1.10) p<0.0005	OR = 0.90 (0.82, 0.99) p=0.03	0.82
Simplified limited/extensive model	OR = 6.04 (2.85, 12.79) p<0.0005	OR = 1.08 (1.04, 1.11) p<0.0005		0.80

**Supplementary Table S6. The prognostic scoring system consists of a five point scale (0-4, with 0-2 for age and 0-2 for imaging) with associated mortalities.**

Points	Age/r-CXR mortality	Age/HRCT mortality
0	2/32 (6%)	0/21 (0%)
1	10/53 (19%)	7/49 (14%)
2	18/69 (26%)	15/56 (27%)
3	16/28 (57%)	24/53 (45%)
4	15/17 (88%)	15/20 (75%)

**Fig. S1.** R-CXR(A), and standard supine CXR (B) coronal HRCT (C) of a 79 years old female patient with unilateral COVID-19 pneumonia. R-CXR (A) was considered very similar to standard supine CXR (B), both showing a large ground-glass opacity in the left lung. This finding was confirmed on HRCT (C).





**Fig. S2.** R-CXR(A), standard supine CXR (B) and coronal HRCT (C) of a 50 years old male patient with unilateral COVID-19 pneumonia. R-CXR (A) was considered very similar to standard supine CXR (B). However, either r-CXR or supine standard CXR showed bilateral consolidations that were confirmed on HRCT (C).



