EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting

Nicola Sverzellati, Christopher J Ryerson, Gianluca Milanese, Elisabetta A Renzoni, Annalisa Volpi, Paolo Spagnolo, Francesco Bonella, Ivan Comelli, Paola Affanni, Licia Veronesi, Carmelinda Manna, Andrea Ciuni, Carlotta Sartorio, Giulia Tringali, Mario Silva, Emanuele Michieletti, Davide Colombi, U Athol Wells

Please cite this article as: Sverzellati N, Ryerson CJ, Milanese G, *et al*. Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting. *Eur Respir J* 2021; in press (https://doi.org/10.1183/13993003.04188-2020).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting

Sverzellati Nicola,[1] Ryerson Christopher J,[2] Milanese Gianluca,[1] Renzoni Elisabetta A,[3] Volpi Annalisa,[4] Spagnolo Paolo,[5] Bonella Francesco,[6] Comelli Ivan,[7] Affanni Paola,[8] Veronesi Licia,[8] Manna Carmelinda,[1] Ciuni Andrea, [1] Sartorio Carlotta, [1] Tringali Giulia, [1] Silva Mario, [1] Michieletti Emanuele,[9] Colombi Davide,[9] Wells U Athol[3].

[1]Scienze Radiologiche, Dipartimento di Medicina e Chirurgia, University-Hospital of Parma, Parma, Italy.

[2]Department of Medicine, University of British Columbia and Centre for Heart Lung Innovation, St. Paul's Hospital.

[3]Interstitial Lung Disease Unit, Royal Brompton Hospital, Imperial College, London, UK.

[4]1st Anesthesia and Intensive Care Unit, University Hospital of Parma, Parma, Italy

[5]Respiratory Disease Unit, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova, Padua, Italy.

[6]Center for interstitial and rare lung diseases, Pneumology Department, Ruhrandklinik University Hospital, University of Duiburg-Essen, Essen, Germany.

[7]Unità Operativa Pronto Soccorso e Medicina d' Urgenza, Azienda Ospedaliero-Universitaria di Parma, Parma, Italy.

[8] Department of Medicine and Surgery, University of Parma, Parma, Italy

[9]Department of Radiological Functions, Radiology Unit, "Guglielmo da Saliceto" Hospital, Via Taverna 49, 29121, Piacenza, Italy.

Corresponding author:

Prof. Nicola Sverzellati

Scienze Radiologiche, Dipartimento di Medicina e Chirurgia, Università di Parma

Address: Padiglione Barbieri, Scienze Radiologiche, Azienda Ospedaliero-Universitaria di Parma

Parma, 43126

Italy

Abstract

Introduction: for the management of patients referred to respiratory triage during the early stages of the SARS-CoV-2 pandemic, either chest radiograph (CXR) or computed tomography (CT) were used as first-line diagnostic tools. The aim of this study was to compare the impact on triage, diagnosis and prognosis of patients with suspected COVID-19 when clinical decisions are derived from reconstructed CXR or from CT.

Methods: we reconstructed CXR (r-CXR) from high-resolution CT (HRCT) scan. Five clinical observers independently reviewed clinical charts of 300 subjects with suspected COVID-19 pneumonia, integrated with either r-CXR or HRCT report in two consecutive blinded and randomized sessions: clinical decisions were recorded for each session. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and prognostic value were compared between r-CXR and HRCT. The best radiological integration was also examined to develop an optimized respiratory triage algorithm.

Results: interobserver agreement was fair (Kendall's W = 0.365; p<0.001) by r-CXR-based protocol and good (Kendall's W = 0.654; p<0.001) by CT-based protocol. NPV assisted by r-CXR (31.4%) was lower than that of HRCT (77.9%). In case of indeterminate or typical radiological appearence for COVID-19 pneumonia, extent of disease on r-CXR or HRCT were the only two imaging variables that were similarly linked to mortality by adjusted multivariable models

Conclusions: the present findings suggest that clinical triage is safely assisted by CXR. An integrated algorithm using first-line CXR and contingent use of HRCT can help optimize management and prognostication of COVID-19.

Funding: study was not funded.

Introduction

Despite worldwide efforts to halt its transmission, severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) has affected more than 30 million individuals and caused nearly 1 million deaths as of late September 2020 [1, 2]. After the initial outbreak, most countries have prepared their healthcare systems to face the pandemic. Although highly desirable, global and shared preparedness planning has faced political, institutional, social, environmental, and technological challenges [3, 4]. A recent International survey reported substantial heterogeneity in the diagnostic approach to coronavirus disease 2019 (COVID-19) pneumonia within and among countries and continents [4].

To date, molecular testing is used in both symptomatic or asymptomatic subjects with risk of contamination. However, the use of imaging, particularly the choice of imaging technique, is still a matter of debate [5-8]. Molecular and imaging testing are helpful in different aspects of the disease and their integration should be driven by dynamic protocols to be adapted as knowledge of the disease improves. The use of real-time reverse transcription-polymerase chain reaction (RT-PCR) was adapted to the massive needs by shortening of reaction time (and thus reporting time), yet it is still challenged by a substantial proportion of false negatives [9]. Conversely, imaging can show signs of pneumonia in patients with negative RT-PCR but clinically suspected COVID-19, thus offering a potential role in supporting rapid decision making [9, 10]. The use of imaging was thus recommended for patients who present at triage with moderate to severe features of COVID-19 pneumonia regardless of RT-PCR results [7].

Different imaging approaches are available and have been discussed in the recent COVID-19 literature, including chest X-ray (CXR) and computed tomography (CT). There is still no international consensus upon the integrated use of CXR or CT for clinical assessment and management of subjects with suspected COVID-19 pneumonia [4-6, 11-17]. Most concerns are focused on the accuracy of these tests, individual resources, and risk of infection for radiographers and other healthcare employees. However, the scientific debate lacks important evidence on the impact of CXR and CT on triage decisions and patient care.

In this study, we sought to use a post-processing imaging technique to retrospectively reconstruct CXR from CT scan, and compare the impact of these two imaging tools on the initial clinical triage, diagnosis and prognosis of patients with suspected COVID-19.

Methods

Study population

The study population comprised patients who had been evaluated with chest high-resolution computed tomography (HRCT) scan by the COVID-19 respiratory triage of the University Hospital of Parma, which is located in one of the most affected areas in Northern Italy. In brief, patients were screened for symptoms (e.g. fever and dyspnea) and oxygen saturation. Patients with moderate to severe pulmonary involvement (e.g. oxygen saturation $\leq$95%) underwent HRCT scan. Given the turnaround times for SARS-CoV-2 testing results (e.g. ranging from 2 to >48 hours), a presumptive diagnosis based on clinical-radiological findings was considered for swift decision making such as discharge, recommendation for home quarantine, or hospitalization in different hospital areas including either dedicated COVID-19 pavilions or non-COVID-19 wards. Details on clinical triage were previously reported [8].

The study derivation cohort was built up by including 300 patients who consecutively underwent HRCT at the Parma triage from 29th February to 7th March 2020, as follows: the first 200 patients consecutively admitted to the triage with moderate-severe respiratory clinical impairment were mixed up with the first 100 patients consecutively admitted with mild respiratory clinical impairment (e.g. oxygen saturation 96-98%, hyperthermia, tachypnea) (Fig. 1). The addition of this subgroup ensured that the full spectrum of COVID-19 severity was evaluated.

The study results were externally validated in a cohort of 104 patients (validation cohort) consecutively evaluated at a neighbouring hospital during the same time frame (Piacenza, Italy), which adopted a similar diagnostic protocol (see also supplementary material).

This retrospective study was approved by the referring local Review Board. Informed consent was obtained from the study patients.

Imaging technique

Details on CT scanners and HRCT technique are reported in the supplementary material. HRCT scans allow for various post-processing reconstruction algorithms, including Average Intensity Projection (AIP) [18]. AIP images represent the average of each component attenuation values encountered by the X-ray beam through an object. By manipulating the slab thickness of the coronal AIP mages, it was possible to obtain images analogous to frontal CXR, hereafter called reconstructed (r)-CXR (online videoclip, Fig. 2, 3, 4). Further illustrations, videos and technical

details of the conversion from HRCT to r-CXR imaging, as well as evaluation of r-CXR consistency with standard CXR are reported in the supplementary material.

HRCTs were prospectively scored by a senior chest radiologist (NS, 16-year experience in imaging of interstitial lung disease). He recorded individual HRCT abnormalities and graded the HRCTs into four diagnostic categories, as follows: normal, alternative diagnosis (to be specified), indeterminate, or typical for COVID-19 pneumonia [8]. The total extent of pulmonary disease was scored to the nearest 5%.

In keeping with the visual scoring of the HRCT, two radiologist observers (CM and AC with 11- and 4-year experience in chest imaging, respectively) recorded individual r-CXR abnormalities and graded the r-CXR as follows: normal, alternative diagnosis (to be specified), indeterminate, or typical for COVID-19 pneumonia. The total extent of pulmonary disease on r-CXR was evaluated through an overall visual impression, using a four-point-scale: 0) no parenchymal abnormality; 1) <20%; 2) 20-50%; 3) >50%. Interobserver discrepancies for both diagnostic categories and disease extent were resolved by consensus.

Clinical assessment in a triage-like setting

Clinical data for each patient were jointly reported in data sheets by a consultant anesthesiologist (AV) and a radiologist (GM) Table 1. These data were assembled into individual clinical charts that were given for simulation of clinical management to each clinical observer of this study. RT-PCR test results were not included in the clinical charts in order to simulate an environment where turnaround times were long.

Five clinical observers from three different countries participated in the study. These comprised three physicians (AV, IC, FB) who worked at a HRCT-based triage, and two physicians (PS, EAR) whose hospital protocol proposed a CXR-based triage. Details on the study observer characteristics are given in the supplementary material. Each clinical observer was asked to independently read twice the full set of clinical charts according to two different settings, as follows:

1) The first review was designed to simulate *r-CXR-based integrated clinico-radiological protocol*: each clinical observer reviewed the data sheet and r-CXR report for each patient. Each clinical observer was asked to provide a clinical decision according to one of the following options:
    a. discharge,

b. hospitalization in non-COVID-19 area,
c. home quarantine,
d. hospitalization in COVID-19 area,
e. further work-up by chest HRCT.

2) The second review was designed to simulate *HRCT-based integrated clinico-radiological protocol*: after two days from the first review completed, each clinical observer started the second of data sheet including HRCT report for each patient. Each clinical observer was asked to provide a clinical decision according to one of the following options:

a. discharge,
b. hospitalization in non-COVID-19 area,
c. home quarantine,
d. hospitalization in COVID-19 area.

The study observers were informed that this triage setting was supposed to simulate a pandemic environment with high influx of subjects with suspected COVID-19 pneumonia, and that the clinical decision could be expressed in the absence of any resource constraints.

Statistical analysis

The study analysis compared the frequency of each clinical decision category and its consistency between "r-CXR-based protocol" and "HRCT-based protocol", by means of both intra- and inter-observers analysis. Either Chi-squared or McNemar test were used to compare r-CXR and HRCT diagnostic categories and RT-PCR results. Radiological data were compared between r-CXR and HRCT by the weighted-kappa coefficient to evaluate interobserver agreement. Kendall W test was used to evaluate the overall interobserver clinical decision agreement by r-CXR-based or HRCT-based protocol. Further details are provided in the supplementary material.

Spearman's rank correlation coefficient disease was used to evaluate the correlation between r-CXR-extent and HRCT-extent of disease.

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of both r-CXR and HRCT diagnostic categories were calculated against RT-PCR, by grouping normal to alternative diagnosis and indeterminate to typical for COVID-19 pneumonia diagnostic categories, respectively.

Unadjusted and multivariable logistic regression analyses were used to identify the contribution of clinical and radiological variables to mortality prediction. Multivariable models included age, sex, duration of symptoms at triage, and a comorbidity score of 0-4, obtained by summing the presence (1 point for each) of individual comorbidities consistently associated with a poor outcome in previous reports (diabetes, hypertension, cardio-vascular disease, obesity). The classification performances of the models were evaluated by the area under the receiver operating characteristic curve (AUC).

Based on these models, we constructed a user-friendly five-point scale for both CT and CXR, integrating age and extent of disease on imaging (the only two variables emerging as strong independent determinants of mortality). These scales were externally tested in the validation cohort.

P values of less than 0.05 were considered to indicate statistically significant differences. Analyses were performed using STATA software (STATA Version 14; Computing Resource Centre, Santa Monica, CA).

Results

Study population

The derivation cohort included 300 patients (188 men and 112 women, 66.8 ±15.8, age range 23.1-97.6 years). Clinical and laboratory characteristics are summarized in Table 1 and in the supplementary material. The mean time interval between the onset of symptoms and the triage assessment (and also HRCT scanning) was 6.3 days (±4.4). A total of 248 patients underwent RT-PCR testing. A total of 162/248 (65.3%) subjects tested positive at initial RT-PCR. Repeat RT-PCR testing was performed if the first RT-PCR was negative, identifying an additional 28 positive cases, thus resulting in an overall 190/248 (76.6%) positive RT-PCR. Vital status was ascertained for 295 out of 300 (98.3%) subjects. All patients were followed until death or a minimum of 30 days (range of follow up 1-153 days).

The characteristics of both derivation and validation cohort are summarized in Table 1.

Inter- and intra-observer agreement for the clinical decision

The clinical decision according to either r-CXR-based protocol or HRCT-based protocol is summarized in Tables 2 and 3. The overall interobserver agreement by r-CXR-based protocol was fair (Kendall's W = 0.365; $p<0.001$), and it improved to good by HRCT-based protocol (Kendall's W = 0.654; $p<0.001$). This was true even after stratification for categories of oxygen saturation levels (Supplementary Table 1). Of note, opposite trends in the agreement across the range of prespecified oxygen saturation levels were observed: the lower the oxygen saturation the worse the agreement by r-CXR and the better the agreement by HRCT-based protocol (Supplementary Table 1).

Paired interobserver agreement (Supplementary Table 2) by r-CXR ranged from poor (0.17 - Obs_1/Obs_5) to moderate (0.50 - Obs_2/Obs_4), and it ranged from fair (0.23 - Obs_1/Obs_5) to good (0.75 - Obs_2/Obs_4) by HRCT-based protocol.

Intra-observer agreement on the clinical decision by either r-CXR or HRCT ranged from fair (0.37) to good (0.71) (Supplementary Table 3).

In the r-CXR round, further work-up by HRCT scan was requested in 8% to 46% of patients by the study observers. Once the HRCT was provided, the most frequent decision in this group was COVID hospitalization (41.7% to 73.2%) (Table 3). In particular, HRCT scan was requested by at least one observer for a total of 224 (74.7%) patients, while a management decision without request of HRCT (only clinico-r-CXR findings) was expressed by all observers in 76 (25.3%) patients (Supplementary Table 4). In the latter subgroup, the interobserver agreement was similar

between r-CXR (Kendall's W = 0.763; p<0.001) and HRCT (Kendall's W = 0.725; p<0.001). When r-CXR was considered sufficient for the decision making by all the observers, patients had higher oxygen saturation, and reported more frequently a history of social contact with COVID-19 infected individuals, as compared to patients for whom HRCT was requested by any observer (Supplementary Table 4).

r-CXR vs. HRCT

r-CXR and HRCT data are summarized in Table 1.

Interobserver analysis data is given as supplementary data. Eighty-five r-CXR were classified normal, with 49/85 (57.6%) being false negative as compared to HRCT, notably 30/85 (35.3%) showing HRCT abnormalities consistent with COVID-19 pneumonia (ground-glass opacity in 26/30, 86.7%) with a mean disease extent on HRCT 23% (range 5-65%). Conversely, indeterminate r-CXR findings or r-CXR suggestive for a non-COVID-19 disease were each reported in 4/40 (10%) patients with normal HRCT, thus representing false positive r-CXR.

Of 190 patients with positive RT-PCR results, 156 (82.1%) had positive HRCT scans, and 133 (70%) had positive r-CXRs. Of 58 subjects with negative RT-PCR result, 28 (48.3%) had positive HRCT scans, and 31 (53.4%) had positive r-CXRs. The sensitivity, specificity, PPV, and NPV of HRCT in indicating COVID-19 infection were 95.2% (95%CI: 91.2%-97.8%), 32.8% (95%CI: 21%-46.3%), 82.2% (95%CI: 79.4%-84.7%) and 67.9% (95%CI: 50.3%-81.5%), respectively. The sensitivity, specificity, PPV, and NPV of r-CXR in indicating COVID-19 infection were 81.6% (95%CI 75.3%-86.8%), 27.6% (95%CI 16.7%-40.9%). 78.1% (95%CI 75%-80.9%) and 31.4% (95%CI 21.5%-43.3%), respectively.

Unadjusted associations between mortality and core clinical and imaging variables in patients with r-CXR findings compatible with COVID-19 pneumonia are summarized in Table 4. Two models examined mortality against the extent of disease on either r-CXR or HRCT, using the same 4-point categorical grading system (Table 4). In the logistic regression model with r-CXR, mortality was associated with extent of disease (OR=2.38; 95%CI 1.61-3.50; p<0.001). In the logistic regression model with HRCT, mortality was associated with extent of disease (OR=2.62; 95%CI 1.67-4.10; p<0.001). Both r-CXR and HRCT associations were robust with inclusion in models of age (also strongly linked to mortality), sex, duration of symptoms, and comorbidity score (Supplementary Table 5).

A further third model was built to explore a two-point grading system for risk stratification: the extent of disease was classified by contingent categories as either limited (r-CXR extent<20%; or r-CXR 20-50%, with HRCT extent 0-50%) or extensive (r-CXR extent >50%; or r-CXR 20-50%, with HRCT extent >50%). The distinction between limited and extensive disease was strongly associated with mortality (OR=5.24; 95%CI 2.69-10.22; p<0.001) and was robust with adjustment for age, sex, and comorbidity score (Supplementary Table 5).

Prognostic scoring systems using r-CXR, and HRCT

For each of the three abovementioned models, AUC values fell minimally when multivariable analysis was confined to age and extent of disease on imaging (Supplementary Table 5). Therefore, simplified scoring systems for r-CXR and HRCT were constructed based on these two variables. Age was categorized as <60 (score 0), 60-74 (score 1), and >74 (score 2) years, dividing the cohort into approximate thirds (OR=2.79; 95%CI 1.80-4.31; p<0.001). Imaging extent scores were categorized as <20% (score 0); 20-50% (score 1); >50% (score 2) (Fig. 2, 3, 4). Age and imaging scores were summed to provide five point scales (0-4) for age/r-CXR and age/HRCT scoring systems. Mortality in relation to the age/r-CXR and age/HRCT scores are shown in Fig. 5 and Supplementary Table 6.

In a logistic regression model, the age/r-CXR score was associated with mortality (OR=2.85; 95%CI 2.00-4.06; p<0.001), AUC=0.77. This finding was reproduced in the validation cohort (OR=2.48; 95%CI 1.47-4.17; p<0.001), AUC=0.73. In a separate logistic regression model, the age/HRCT score was associated with mortality (OR=2.71; 95%CI 1.92-3.84; p<0.005), AUC=0.76. This finding was reproduced in the validation cohort (OR=4.23; 95%CI 2.11-8.48; p<0.001), AUC=0.80.

Discussion

In this simulated COVID-19 pandemic triage for a high influx of patients with mostly moderate-to-severe clinical features suspicious of COVID-19 pneumonia, the use of r-CXR for morphology and extent of pneumonia allowed fair interobserver agreement for clinical management, however r-CXR with moderate extent displayed limited yield that could be assisted by integration by HRCT. The interobserver agreement considerably increased when using HRCT, which lead to an increase in the number of recommended hospitalizations. This observation is of

particular relevance as the observers were from different countries, used various COVID-19 triage strategies in their daily practice, and had different subspecialty interests.

The major source of disagreement for the r-CXR-based protocol was the frequency of the requested HRCT scans, which ranged from 8% to 46% of cases. When HRCT work-up was not recommended by any of the study observers, the agreement was similar to that obtained with the HRCT-based protocol. Although we observed a tendency in considering the r-CXR sufficient in patients with milder pulmonary dysfunction, abnormal r-CXR and high exposure risk, the variety in requesting HRCT scans had still no obvious reasons (e.g. age, disease extent on r-CXR etc.). Such observation emphasizes the need to define the clinical indications to CT scanning when the triage workflow relies on CXR as the first imaging modality. In fact, poorly defined criteria may unreasonably increase the number of CT scans after CXR, with potential detrimental effects on the workflow and healthcare worker safety.

In order to evaluate the clinical implications of the two imaging modalities, the sensitivity, the specificity, and the prognostic value of r-CXR and HRCT were compared. The levels of sensitivity for r-CXR were in keeping with previous findings [9, 16, 17, 19]. Sensitivity, specificity, and PPV of r-CXR were marginally lower than HRCT, whereas major discrepancy was found for NPV due to the remarkable proportion of false negative r-CXRs.

Any optimization process of COVID-19 triage protocols should consider mortality data and resource utilization predictions [20]. First, the proportion of deaths among patients with normal r-CXR was higher (12%) as compared to normal HRCT (5%). Such a discrepancy was also evident by comparing r-CXR- and HRCT-based prognostic scoring systems (Figure 5). Moreover, the distinction between typical COVID-19 pneumonia and indeterminate findings was not worthwhile for both r-CXR and HRCT as they were associated with similar outcomes. This suggests that, in a pandemic situation, indeterminate CXR or CT findings likely represent signs of COVID-19 pneumonia. Hence, after grouping indeterminate and typical COVID-19 categories, we found that the diagnostic categories had fairly major prognostic significance for both r-CXR and HRCT. This finding is important as the diagnostic categories used in the present study substantially overlap both CO-RADS categories and those proposed by the Radiologic Society of North America [8, 21].

Our findings do not suggest that HRCT extent scoring adds greatly to prognostic evaluation, over and above CXR scoring, and this is concordant with the need to first level approach by CXR to optimize use of radiology resources and maximize patient and healthcare employee safety. However, an alternative strategy – HRCT complement when CXR extent findings are

intermediate – allowed categorization into limited and extensive disease with a major prognostic separation. The proposed contingent staging algorithm - where r-CXR grades 1 (<20%) and 3 (>50%) are accepted and HRCT is used to adjudicate on r-CXR 2 (21-50%) - showed the highest prognostic value. Separation into a higher risk/lower risk dichotomy may have major clinical value but we suggest that this observation requires further exploration if CT is to be integrated into routine prognostic evaluation in selected cases, when CXR findings are not definitive.

The higher number of deaths in subjects with normal r-CXR still represents the main limitation of this staging system. In patients dying despite a r-CXR grade of 1, the HRCT extent grade was either intermediate or extensive in two thirds of cases. These data suggest that a first level approach by CXR will require some improvement in terms of sensitivity. Otherwise, first level approach by HRCT would provide major confidence in clinical decision and prognostication, therefore it might considered in the context of local logistics that minimize risk of contamination (e.g. dedicated scanner in proximity of the triage rooms).

This study has some limitations. It is worth emphasizing that our observations were derived from a pre-peak endemic environment (e.g. when RT-PCR demand was beyond capacity), while the triage process is being adjusted as the endemic changes. Nevertheless, the study findings may still be helpful in epidemic scenarios, the increased availability of the lab test results within a few hours could further mitigate the CXR limitations (e.g. lower diagnostic accuracy as compared to CT), encouraging its use. Moreover, the reported heterogeneity in the request of HRCT scans among the study observers when first interpreting a CXR might be overrated as a simulated triage setting does not likely reflect the routine clinical context, where multidisciplinary discussion (e.g. between radiologists and triagers) could support other clinical decisions with better integration of available data. The CXR was derived from HRCT data given the impossibility of running a two arm-randomized-controlled trial between CXR and CT in the COVID-19 setting, yet we tested this approach against original CXR (supplementary materials).

In conclusion, this study showed that r-CXR findings were often regarded as not sufficiently informative by clinicians in a COVID-19 pandemic triage setting. This observation suggests that chest CT should be considered after CXR in a substantial percentage of patients with suspected COVID-19 pneumonia, thus potentially causing detrimental effects in the absence of pre-defined diagnostic work-up criteria. Nevertheless, the present study findings suggest that clinicians could rely on positive CXR showing low or high extent of pneumonia, whereas intermediate extent by CXR should be complemented by CT for optimal classification into high and low risk group.

Data collected for the study will not be made available to others.

Declaration of interests

Dr. Bonella reports personal fees and non-financial support from Boehringer Ingelheim, Roche, Galapagos, GSK, Savara Pharma, BMS outside the submitted work. Dr. Spagnolo reports grants, personal fees and non-financial support from Roche, grants, personal fees and non-financial support from PPM Services, personal fees from Red X Pharma, personal fees from Galapagos, personal fees from Chiesi, grants, personal fees and non-financial support from BoehringerIngelheim, outside the submitted work; and Wife employee of Novartis. Dr. Wells reports personal fees and other from ROCHE, personal fees and other from Boehringer Ingelheim, personal fees and other from Bayer, outside the submitted work. The other Authors declare no conflict of interest.

Role of the funding source

The study was not funded.

References

1. Poston JT, Patel BK, Davis AM. Management of Critically Ill Adults With COVID-19. *JAMA* 2020.
2. World Health Organization: Director-General's opening remarks at the media briefing on COVID-19 - 29 June 2020. 2020 [cited April 1, 2020]; Updated March 11, 2020]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200629-covid-19-sitrep-161.pdf?sfvrsn=74fde64e_2
3. The Lancet Infectious D. Challenges of coronavirus disease 2019. *Lancet Infect Dis* 2020: 20(3): 261.
4. Blazic I, Brkljacic B, Frija G. The use of imaging in COVID-19-results of a global survey by the International Society of Radiology. *Eur Radiol* 2020.
5. ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. 2020 [cited; Available from: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection
6. Xu Q, Xu K, Yang C, Zhang X, Meng Y, Quan Q. Askin tumor: four case reports and a review of the literature. *Cancer Imaging* 2011: 11: 184-188.
7. Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, Schluger NW, Volpi A, Yim JJ, Martin IBK, Anderson DJ, Kong C, Altes T, Bush A, Desai SR, Goldin J, Goo JM, Humbert M, Inoue Y, Kauczor HU, Luo F, Mazzone PJ, Prokop M, Remy-Jardin M, Richeldi L, Schaefer-Prokop CM, Tomiyama N, Wells AU, Leung AN. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Chest* 2020.
8. Sverzellati N, Milanese G, Milone F, Balbi M, Ledda RE, Silva M. Integrated Radiologic Algorithm for COVID-19 Pandemic. *J Thorac Imaging* 2020.
9. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* 2020: 200642.
10. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, Ji W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* 2020: 200432.
11. Wang YXJ, Liu WH, Yang M, Chen W. The role of CT for Covid-19 patient's management remains poorly defined. *Ann Transl Med* 2020: 8(4): 145.
12. Kooraki S, Hosseiny M, Myers L, Gholamrezanezhad A. Coronavirus (COVID-19) Outbreak: What the Department of Radiology Should Know. *J Am Coll Radiol* 2020.
13. Davarpanah AH, Mahdavi A, Sabri A, Langroudi TF, Kahkouee S, Haseli S, Kazemi MA, Mehrian P, Mahdavi A, Falahati F, Tuchayi AM, Bakhshayeshkaram M, Taheri MS. Novel Screening and Triage Strategy in Iran During Deadly Coronavirus Disease 2019 (COVID-19) Epidemic: Value of Humanitarian Teleconsultation Service. *J Am Coll Radiol* 2020.
14. Mossa-Basha M, Medverd J, Linnau K, Lynch JB, Wener MH, Kicska G, Staiger T, Sahani D. Policies and Guidelines for COVID-19 Preparedness: Experiences from the University of Washington. *Radiology* 2020: 201326.
15. Oudkerk M, Buller HR, Kuijpers D, van Es N, Oudkerk SF, McLoud TC, Gommers D, van Dissel J, Ten Cate H, van Beek EJ. Diagnosis, Prevention, and Treatment of Thromboembolic Complications in COVID-19: Report of the National Institute for Public Health of the Netherlands. *Radiology* 2020: 201629.
16. Schiaffino S, Tritella S, Cozzi A, Carriero S, Blandi L, Ferraris L, Sardanelli F. Diagnostic Performance of Chest X-Ray for COVID-19 Pneumonia During the SARS-CoV-2 Pandemic in Lombardy, Italy. *J Thorac Imaging* 2020.

17. Balbi M, Caroli A, Corsi A, Milanese G, Surace A, Di Marco F, Novelli L, Silva M, Lorini FL, Duca A, Cosentini R, Sverzellati N, Bonaffini PA, Sironi S. Chest X-ray for predicting mortality and the need for ventilatory support in COVID-19 patients presenting to the emergency department. *European Radiology* 2020.
18. Lee KH, Kim YH, Hahn S, Lee KW, Kim TJ, Kang SB, Shin JH. Computed tomography diagnosis of acute appendicitis: advantages of reviewing thin-section datasets using sliding slab average intensity projection technique. *Invest Radiol* 2006: 41(7): 579-585.
19. Wong HYF, Lam HYS, Fong AH, Leung ST, Chin TW, Lo CSY, Lui MM, Lee JCY, Chiu KW, Chung T, Lee EYP, Wan EYF, Hung FNI, Lam TPW, Kuo M, Ng MY. Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. *Radiology* 2019: 201160.
20. Maves RC, Downar J, Dichter JR, Hick JL, Devereaux A, Geiling JA, Kissoon N, Hupert N, Niven AS, King MA, Rubinson LL, Hanfling D, Hodge JG, Jr., Marshall MF, Fischkoff K, Evans LE, Tonelli MR, Wax RS, Seda G, Parrish JS, Truog RD, Sprung CL, Christian MD, Care ATFfMC. Triage of Scarce Critical Care Resources in COVID-19 An Implementation Guide for Regional Allocation: An Expert Panel Report of the Task Force for Mass Critical Care and the American College of Chest Physicians. *Chest* 2020.
21. Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, Henry TS, Kanne JP, Kligerman S, Ko JP, Litt H. Radiological Society of North America Expert Consensus Statement on Reporting Chest[ZERO WIDTH SPACE] CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. *J Thorac Imaging* 2020.

Table 1. Baseline characteristics, presentation vitals and laboratory results of Patients admitted to the respiratory triage for COVID-19 (derivation and validation cohort).

	Derivation cohort (n = 300)	Validation cohort (n= 104)
Age (mean, SD)	66.8 (± 15.8)	72.5 (61-81)
Gender	188 M (62.7%); 112 F (37.3%)	71 M (68.3%); 33 F (31.7%)
Contact with suspected or confirmed COVID-19 cases		
- *Yes*	72 (24%)	23 (22.2%)
- *No*	49 (16.3%)	77 (74%)
- *Not reported*	179 (59.7%)	4 (3.8 %)
History of Fever	231 (77%)	92 (88.5%)
- *Days with fever (median, IQR)*	6 (3 - 7)	5.5 (2 - 7.5)
- *Temperature median (IQR)*	37 (36.5 - 37.8)	37.3 (36.5 - 38.2)
- *Temperature ≥ 37.5°C*	90 (30%)	44 (46.3%)
History of Cough	171 (57%)	42 (40%)
- *Days with cough (median, IQR)*	5 (3 - 7)	3 (0 - 7)
History of Dyspnoea	109 (36.3%)	63 (60.6%)
- *Days with dyspnoea (median, IQR)*	3 (3 - 7)	1 (0 - 4)
Respiratory Rate ≥ 25 breaths/min	74 (24.7%)	16 (22.9%)
Oxygen saturation		
- *≤ 95%*	200 (66.7%)	73 (75.3%)
- *%, median (IQR)*	94 (90 - 96)	92 (88 - 95.2)
Hearth Rate > 100 beats/min	66 (22%)	15 (18.5%)
Other symptoms	92 (30.7%)	22 (21.2%)
Hypertension	114 (38%)	58 (55.8%)
Diabetes	42 (14%)	18 (17.3%)
Obesity	18 (6%)	9 (8.7%)
Other comorbidities		
- Cancer	33 (11%)	19 (18.3%)
- Chronic respiratory disease	44 (14.7%)	15 (14.4%)
- Kidney disease	12 (4%)	13 (12.5%)
- Liver disease	1 (0.3%)	2 (1.9%)
- No other comorbidities reported	210 (70%)	55 (52.9%)
Laboratory measures		
- *White blood cell count, × 10^6/L*	6.4 (4.6 - 8.7)	7.4 (5.1 - 9.5)
- *Sodium, mmol/L*	137 (135 - 139)	136 (133 - 139)
- *Aspartate aminotransferase, U/L*	42 (32 - 62)	44 (30 - 63)
- *Alanine aminotransferase, U/L*	28 (19 - 42)	28 (20 - 49)
- *C-reactive protein, mg/L*	83 (27 - 148)	82 (36 – 150)
r-CXR categories		
- *Normal*	85	18
- *Alternative diagnosis*	7	0
- *Indeterminate*	40 (median extent: ≤20%)	16 (median extent: ≤20%)
- *Typical for COVID-19 pneumonia*	168 (median extent: 21-50%)	70 (median extent: 20-50%)
HRCT categories		

- *Normal*	40	6
- *Alternative diagnosis*	35 (mean extent: 5.6%)	1
- *Indeterminate*	41 (mean extent: 28.4%)	13 (mean extent: 32.7%)
- *Typical for COVID-19 pneumonia*	184 (mean extent: 35%)	84 (mean extent: 36.5%)

Table 2. Management according to r-CXR-based protocol and HRCT-based protocol.

	MANAGEMENT	Obs_1 (n=300)	Obs_2 (n=300)	Obs_3 (n=300)	Obs_4 (n=300)	Obs_5 (n=300)
r-CXR	**DISCHARGE**	11 (4%)	24 (8%)	13 (4%)	4 (1%)	48 (16%)
	NON-COVID HOSPITALIZATION	10 (3%)	23 (8%)	27 (9%)	18 (6%)	49 (16%)
	HOME QUARANTINE	111 (37%)	27 (9%)	99 (33%)	47 (16%)	18 (6%)
	COVID HOSPITALIZATION	103 (34%)	153 (51%)	137 (46%)	164 (55%)	47 (16%)
	NEED OF CHEST CT	65 (22%)	73 (24%)	24 (8%)	67 (22%)	138 (46%)
HRCT	**DISCHARGE**	5 (2%)	17 (6%)	4 (1%)	8 (3%)	25 (8%)
	NON-COVID HOSPITALIZATION	10 (3%)	46 (15%)	64 (21%)	53 (18%)	76 (25%)
	HOME QUARANTINE	119 (40%)	30 (10%)	101 (34%)	41 (13%)	29 (10%)
	COVID HOSPITALIZATION	166 (55%)	207 (69%)	131 (44%)	198 (66%)	170 (57%)

Table 3. HRCT-based recommendation for subjects for whom HRCT was requested at the r-CXR-based round

MANAGEMENT	Obs_1 *(n = 65)*	Obs_2 *(n = 73)*	Obs_3 *(n = 24)*	Obs_4 *(n = 67)*	Obs_5 *(n = 138)*
DISCHARGE	0 (0%)	3 (4%)	0 (0%)	2 (3%)	2 (1%)
NON-COVID HOSPITALIZATION	2 (3%)	15 (21%)	8 (33%)	22 (33%)	27 (20%)
HOME QUARANTINE	19 (29%)	9 (12%)	6 (25.0%)	10 (15%)	8 (6%)
COVID HOSPITALIZATION	44 (68%)	46 (63.0%)	10 (42%)	33 (49%)	101 (73%)

Table 4. Unadjusted associations between mortality and core clinical and imaging variables in patients with chest radiographic findings compatible with COVID infection examined using logistic regression. Relationships are expressed as odds ratios for mortality, with 95% confidence intervals and p values.

	Odds ratio	95% confidence intervals	Statistical significance
Age (years)	1.07	1.04 - 1.10	p<0.001
Gender	0.82	0.44 - 1.53	p=0.53
Duration of dyspnoea (days)	1.01	0.92 - 1.11	p=0.78
Duration of cough (days)	0.91	0.83 - 0.99	p=0.03
Duration of fever (days)	0.85	0.78 - 0.93	p<0.001
Duration of symptoms (days)	0.86	0.78 - 0.93	p<0.001
Comorbidity score (0-4)	1.52	1.12 - 2.05	p<0.01
CXR severity (1-3)	2.38	1.61 - 3.50	p<0.001
CT severity (1-3)	2.62	1.67 - 4.10	p<0.001
Imaging (CXR/CT) severity (0/1)	5.24	2.69 - 10.22	p<0.001

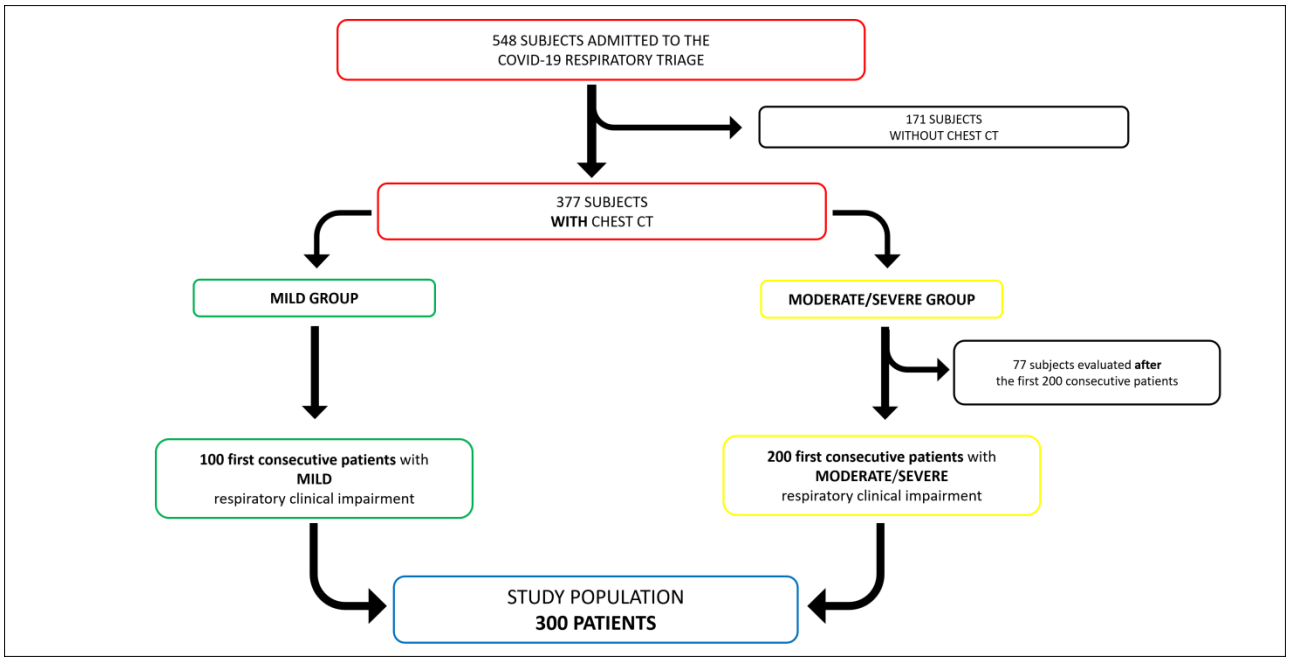
Fig. 1. Diagram illustrating the selection of the derivation study cohort.

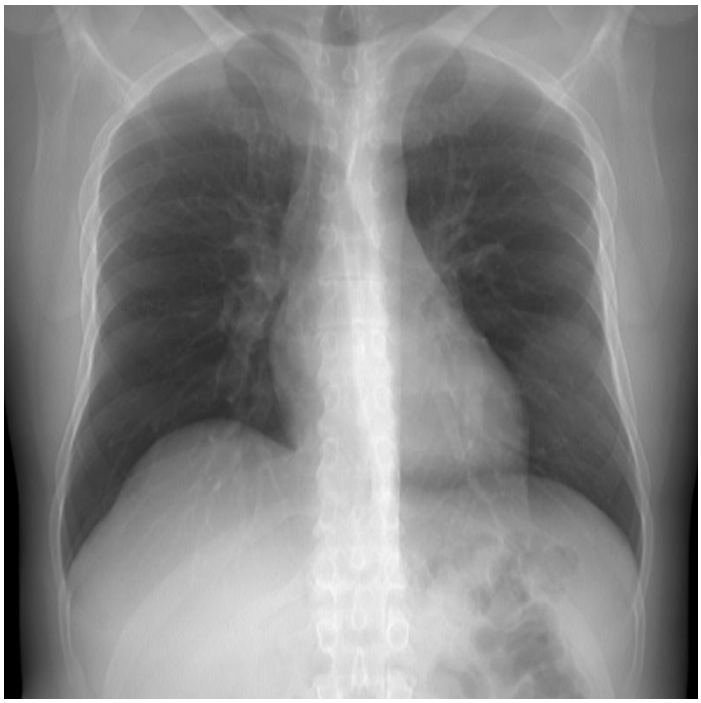
Fig. 2. Reconstructed chest X-ray (r-CXR) (**A**) and coronal multiplanar reconstruction (MPR) of high-resolution computed tomography (HRCT) (**B**) of a 41 years old male patient with unilateral COVID-19 pneumonia. Focal, ovoid area of ground glass opacity in the left lower lobe was identified at both r-CXR and HRCT by the study observers. The extent of disease at r-CXR was scored as category 0 (<20%) by both the study observers when using the prognostic scoring system.

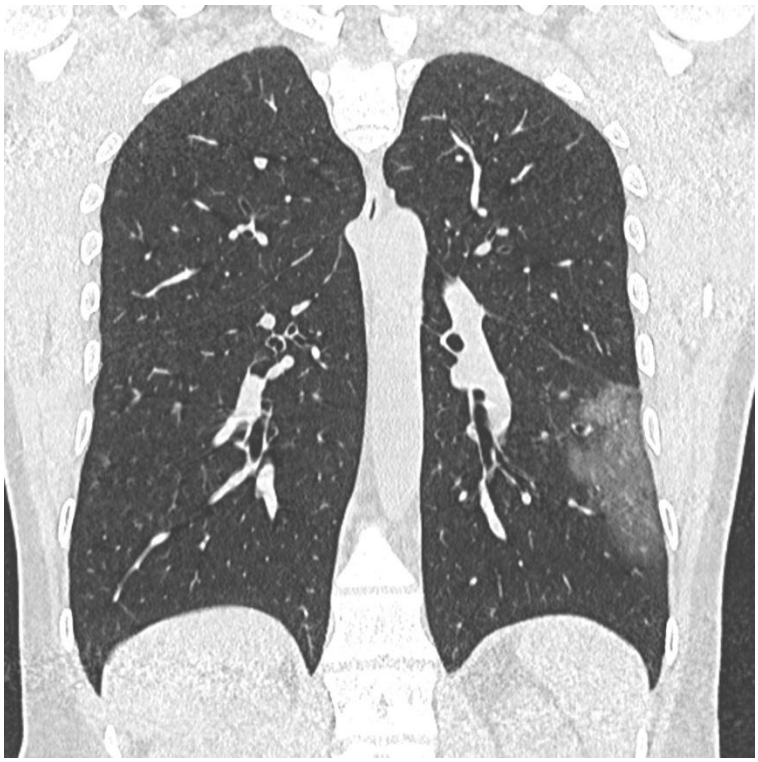
Fig. 3. Reconstructed chest X-ray (r-CXR) (**A**) and coronal multiplanar reconstruction (MPR) of high-resolution computed tomography (HRCT) (**B**) of a 71 years old male patient with bilateral COVID-19 pneumonia. Bilateral, peripheral ground glass opacities were reported at both r-CXR and HCRT. However, disease extent was scored as category 0 (<20%) on r-CXR and category 1 (20-50%) on HRCT.

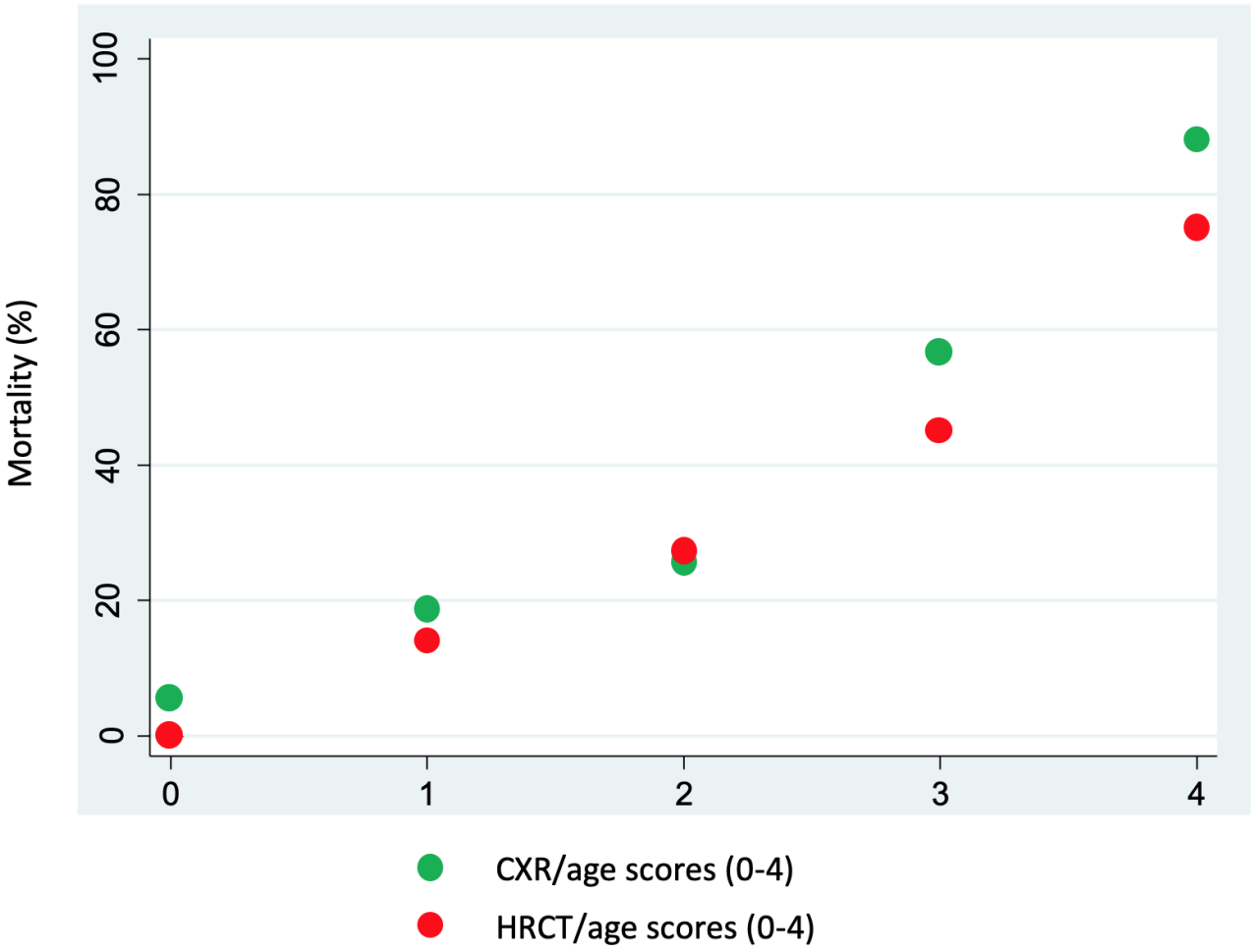
Fig.4. Reconstructed chest X-ray (r-CXR) (**A**) and coronal multiplanar reconstruction MPR of high-resolution computed tomography (HRCT) (**B**) of a 41 years old female patient with severe COVID-19 pneumonia. Bilateral, diffuse ground glass opacities and consolidations occupy most of the lung parenchyma, and the disease extent was scored as category 2 (>50%) both on r-CXR and HRCT.

Fig. 5. Mortality in relation to the age/reconstructed chest X-ray (r-CXR) and age/HRCT scores. Age and imaging scores into a five point scales (0-4), both for age/r-CXR and age/HRCT scoring systems.


548 SUBJECTS ADMITTED TO THE
COVID-19 RESPIRATORY TRIAGE
171 SUBJECTS
WITHOUT CHEST CT
377 SUBJECTS
WITH CHEST CT
MILD GROUP
MODERATE/SEVERE GROUP
77 subjects evaluated **after**
the first 200 consecutive patients
100 first consecutive patients with
MILD
respiratory clinical impairment
200 first consecutive patients with
MODERATE/SEVERE
respiratory clinical impairment
STUDY POPULATION
300 PATIENTS






Mortality (%)
0
20
40
60
80
100
0
1
2
3
4
CXR/age scores (0-4)
HRCT/age scores (0-4)

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 (Bl57 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 fur 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 ≤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 ≤ 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 ($R2 = 0.23$, $p<0.0005$).

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 (indterminate 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 1. 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.

	r-CXR	*p-values*	**HRCT**	*p-values*
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 2. Inter-observer agreement for the whole study population based on r-CXR-round and HRCT-round (weighted kappa coefficients).

		r-CXR				
	Reader	**R_1**	**R_2**	**R_3**	**R_4**	**R_5**
HRCT	**R_1**		0.23 (0.15 - 0.31)	0.23 (0.15 - 0.31)	0.28 (0.20-0.37)	0.17 (0.11 - 0.24)
	R_2	0.31 (0.23 - 0.40)		0.34 (0.25 - 0.42)	0.51 (0.42 - 0.60)	0.31 (0.23 - 0.39)
	R_3	0.44 (0.36 - 0.52)	0.51 (0.44 - 0.59)		0.27 (0.18 - 0.35)	0.19 (0.12 - 0.26)
	R_4	0.33 (0.24 - 0.41)	0.75 (0.68 - 0.81)	0.57 (0.49 - 0.65)		0.13 (0.05 - 0.21)
	R_5	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 3. 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*
R_1	0.59 (0.50-0.69)	235
R_2	0.71 (0.62-0.80)	227
R_3	0.37 (0.27-0.47)	276
R_4	0.53 (0.42-0.65)	233
R_5	0.54 (0.44-0.64)	162

Supplementary Table 4. 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
Age (mean, SD)	65.2 (± 18)	67.4 (± 15)	0.40
Gender	47 M (62%); 29 F (38%)	141 M (63%); 83 F (37%)	0.86
Contact with COVID-19 cases			
- *Yes*	41 (54%)	31 (14%)	**< 0.001**
- *No*	6 (8%)	43 (19%)	
- *Not reported*	29 (38%)	150 (67%)	
History of Fever	54 (71.1%)	177 (79%)	0.82
- *Days with fever (median, IQR)*	5 (3 - 7)	7 (3 - 8)	0.28
History of Cough	43 (56.6%)	128 (57%)	0.36
- *Days with cough (median, IQR)*	6 (3 - 8)	5 (3 - 7)	0.89
History of Dyspnoea	25 (33.9%)	84 (38%)	0.20
- *Days with dyspnoea (median, IQR)*	5 (2.5 - 7)	3 (3 - 7)	0.59
Other symptoms	22 (28.9%)	76 (34%)	
Hypertension	27 (35.5%)	87 (39%)	0.61
Diabetes	8 (10.5%)	34 (15)	0.31
Obesity	2 (0.3%)	17 (8)	0.12
Comorbidities			
- 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
Temperature ≥ 37.5°C	18 (23.7%)	76 (34%)	
- *Temperature, median (IQR)*	36.7 (36.5 - 37.9)	37 (36.5 - 37.7)	0.84
Oxygen saturation			
- *≤ 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%)	
Respiratory Rate ≥ 25 breaths/min	16 (21.1%)	58 (26%)	0.52
Hearth Rate > 100 beats/min	11 (14.5%)	37 (17%)	0.55
Laboratory measures *(median, IQR)*			
- *White blood cell count, × 10^9/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 5. 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
Index cohort				
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 6. 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%)

sFig. 1. 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).

sFig. 2. 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 consilidations that were confirmed on HRCT (C).