Chest Radiograph Scoring Alone or Combined with Other Risk Scores for Predicting Outcomes in COVID-19

Iain Au-Yong, BM BCh, FRCR • *Yutaro Higashi, MBBS, FRCR* • *Elisabetta Giannotti, MD* • *Andrew Fogarty, DM* • *Joanne R. Morling, PhD* • *Matthew Grainge, PhD* • *Andrea Race* • *Irene Juurlink, MSc* • *Mark Simmonds, BMBS, BMedSci* • *Steve Briggs, BSc* • *Simon Cruikshank, BSc* • *Susan Hammond-Pears* • *Joe West, PhD* • *Colin J. Crooks, PhD* • *Timothy Card, PhD*

From the Department of Radiology, Nottingham University Hospitals National Health Service (NHS) Trust, Nottingham, England (I.A.Y., Y.H., E.G.); Respiratory Medicine (A.F.), Gastroenterology (C.J.C., T.C., J.W.), Department of Critical Care (M.S.), Department of Digital Services (I.J.), Department of Corporate Services (A.R., S.H.P.), Specialist Support Directorate (S.C.), Nottingham University Hospitals NHS Trust, Nottingham, England (J.R.M., S.B.); Translational Medical Sciences, School of Medicine, University of Nottingham, Nottingham, England NG7 2UH (C.J.C.); Population and Lifespan Sciences, School of Medicine, University of Nottingham, Nottingham NG5 1PB, England (A.F., J.R.M., M.G., J.W., T.C.); NIHR Nottingham Biomedical Research Centre (BRC), Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, England (A.F., J.R.M., J.W., C.J.C., T.C.); and East Midlands Academic Health Science Network, University of Nottingham, Nottingham, England (J.W.). Received April 16, 2021; revision requested June 1; revision received August 8; accepted August 16. **Address correspondence to** T.C. (e-mail: *tim.card@nottingham.ac.uk*).

Background: Radiographic severity may help predict patient deterioration and outcomes from COVID-19 pneumonia.

Purpose: To assess the reliability and reproducibility of three chest radiograph reporting systems (radiographic assessment of lung edema [RALE], Brixia, and percentage opacification) in patients with proven SARS-CoV-2 infection and examine the ability of these scores to predict adverse outcomes both alone and in conjunction with two clinical scoring systems, National Early Warning Score 2 (NEWS2) and International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium (ISARIC-4C) mortality.

Materials and Methods: This retrospective cohort study used routinely collected clinical data of patients with polymerase chain reaction–positive SARS-CoV-2 infection admitted to a single center from February 2020 through July 2020. Initial chest radiographs were scored for RALE, Brixia, and percentage opacification by one of three radiologists. Intra- and interreader agreement were assessed with intraclass correlation coefficients. The rate of admission to the intensive care unit (ICU) or death up to 60 days after scored chest radiograph was estimated. NEWS2 and ISARIC-4C mortality at hospital admission were calculated. Daily risk for admission to ICU or death was modeled with Cox proportional hazards models that incorporated the chest radiograph scores adjusted for NEWS2 or ISARIC-4C mortality.

Results: Admission chest radiographs of 50 patients (mean age, 74 years ± 16 [standard deviation]; 28 men) were scored by all three radiologists, with good interreader reliability for all scores, as follows: intraclass correlation coefficients were 0.87 for RALE (95% CI: 0.80, 0.92), 0.86 for Brixia (95% CI: 0.76, 0.92), and 0.72 for percentage opacification (95% CI: 0.48, 0.85). Of 751 patients with a chest radiograph, those with greater than 75% opacification had a median time to ICU admission or death of just 1–2 days. Among 628 patients for whom data were available (median age, 76 years [interquartile range, 61–84 years]; 344 men), opacification of 51%–75% increased risk for ICU admission or death by twofold (hazard ratio, 2.2; 95% CI: 1.6, 2.8), and opacification greater than 75% increased ICU risk by fourfold (hazard ratio, 4.0; 95% CI: 3.4, 4.7) compared with opacification of 0%–25%, when adjusted for NEWS2 score.

Conclusion: Brixia, radiographic assessment of lung edema, and percentage opacification scores all reliably helped predict adverse outcomes in SARS-CoV-2 infection.

©RSNA, 2021

Online supplemental material is available for this article.

In December 2019, a strain of coronavirus (SARS-CoV-2) was isolated from bronchoalveolar lavage samples in a cluster of patients with a febrile illness. This illness spread around the world and was declared a pandemic by the World Health Organization on March 11, 2020.

The reference standard for diagnosis of SARS-CoV-2 infection is reverse transcription polymerase chain reaction (RT-PCR) testing because of the test's high specificity, despite limitations in its sensitivity (1). Chest radiography has limited sensitivity and specificity for identification

of patients with SARS-CoV-2 infection (2,3) but can help identify patients with the disease (3–5). Several studies have demonstrated that the severity of lung involvement at chest radiography in SARS-CoV-2 infection is closely correlated with many key outcomes for patients, including intensive care unit (ICU) admission and death (6–20). Several scoring systems for assessing the severity of lung involvement in SARS-CoV-2 infection have been described in the literature. The most widely used and studied systems include Brixia (7,12,21) and radiographic assessment of lung edema

Abbreviations

ICU = intensive care unit, IQR = interquartile range, ISARIC-4C = International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium, NEWS2 = National Early Warning Score 2, RALE = radiographic assessment of lung edema, RT-PCR = reverse transcription polymerase chain reaction

Summary

Brixia, radiographic assessment of lung edema, and percentage opacification produced reliable and reproducible severity scores for chest radiographs in patients with SARS-CoV-2 infection that led to improved accuracy for predicting adverse outcomes when incorporated into National Early Warning Score 2 and International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium clinical scoring systems.

Key Results

- Among 787 patients with confirmed COVID-19, three chest radiograph scores (Brixia, radiographic assessment of lung edema, and percentage opacification) all had good interreader reliability, with intraclass correlation coefficients of 0.87, 0.86, and 0.72, respectively.
- Radiograph scores helped to predict admission to the intensive care unit (ICU) or death after COVID-19 diagnosis; opacification of 51%–75% (vs 0%–25%) was associated with a 2.2-fold increase in these outcomes among patients eligible for ICU care after adjustment for clinical risk scoring.

(RALE/modified RALE) (6,13,14,16,22). No large studies have compared these systems with percentage opacification.

During the pandemic, severity of illness and risk for need for escalation of care could be predicted by using clinical scoring systems. In England, two examples are National Early Warning Score 2 (NEWS2) and the International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium (ISARIC-4C) mortality score (23,24). The utility of these scores has been established independently of chest radiology in previously published studies (23–27).

We aimed to assess the reliability, reproducibility, and ability to predict ICU admission or death (adverse outcomes) of three chest radiograph reporting systems (RALE, Brixia, and percentage opacification) in patients with SARS-CoV-2 infection. We also sought to examine the ability of these scores to predict adverse outcomes of SARS-CoV-2 infection, both alone and in conjunction with two clinical scoring systems: NEWS2 and ISARIC-4C mortality.

Materials and Methods

This work was approved by Nottingham University Hospitals Clinical Effectiveness Team (references 20–153C and 20–501C), the Nottingham University Hospitals Caldicott Guardian, Data Protection Impact Assessment (reference 436), and the National Health Service Health Research Authority (reference 282490). The Health Reference Authority confirmed that individual patient consent was not required.

Study Design and Patients

A retrospective, observational cohort study was conducted at Nottingham University Hospitals National Health Service

Trust, England. Nottingham University Hospitals serve a population of approximately 2.5 million and has 1500 inpatient hospital beds. By using electronic health records within Nottingham University Hospitals, all patients admitted with RT-PCR–proven SARS-CoV-2 infection and who had undergone chest radiography were selected consecutively between February 2020 and July 2020, as previously described (28). Patients who first underwent chest radiography after ICU admission were excluded because they had already reached a study end point. There were no additional exclusions, and a sample size calculation indicated that more than 500 patients were required for a multivariate prediction model with more than 10 parameters (28). We obtained follow-up data for discharge, subsequent admissions to Nottingham University Hospitals, and death both in and out of the hospital (by the National Health Service Patient Demographics Service).
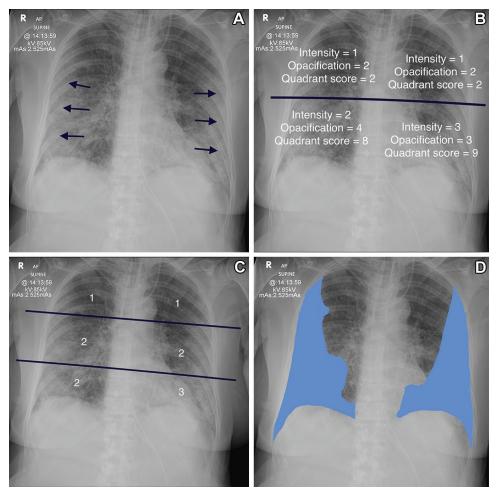
Chest Radiography

At Nottingham University Hospitals, single-view posteroanterior or anteroposterior chest radiography was routinely performed for patients admitted with proven or suspected SARS-CoV-2 infection. The first chest radiographic examination performed during the admission relating to RT-PCR was selected for analysis. During the early part of the pandemic, reporting was not uniform, and no specific scoring system was used. Thus, all chest radiographs were reassessed by using the RALE (22), Brixia (7,21), and percentage opacification scores (18). Table E1 (online) and Figure 1 describe these scoring systems. Chest radiographs were interpreted by using a Jusha monitor (Nanjing Jusha Commercial and Trading), with a picture communication and archiving system (GE Healthcare). Radiography was performed by using dual-detector system (DigitalDiagnost; Philips), digital system (Ysio Max; Siemens), and the FDR Nano Mobile (Fuji) system.

Implementation of Scoring in the Cohort

Assessment of interreader and intrareader reliability.—Chest radiographs of the first 50 patients chronologically were assessed for all three scoring systems independently by all three consultant radiologists contributing to the study (I.A.Y., Y.H., and E.G., with 16, 15, and 10 years of experience and subspecialty interests in chest, gastrointestinal, and breast radiology, respectively) to assess interreader reliability. The three radiologists measured the time to score images (totaled across multiple chest radiographs) by using a stopwatch. For validation of intrareader reliability, each radiologist performed a second read of each chest radiograph, independent of the first read, at least 1 week later.

Assessment of performance of scores in predicting severity.—Once reliability of scoring was established, the entire cohort of consecutive patients with RT-PCR tests that were positive for SARS-CoV-2 was divided so that each chest radiograph was scored by one of the radiologists. Each radiologist therefore applied all three scores to approximately one-third of the chest ra-

diographs. The radiologists were aware of the RT-PCR status but were blinded to all other clinical information.

Collection of other data.—Data collection beyond the scoring of chest radiographs has been fully described previously (28). Briefly, data were extracted from electronic records and included demographic characteristics, clinical decisions at eligibility for escalation (an assessment that had no fixed rules and was on the basis of a physician's assessment of the patient's frailty, comorbidity, and preferences), ward type (standard inpatient ward or ICU), and all other data needed to calculate ISARIC-4C mortality and NEWS2 (Table E1 [online]). Data were extracted from the date of relevant hospital admission (or of suspicion of SARS-CoV-2 if preceding the date of admission from an emergency department visit) until admission to the ICU, discharge from hospital, or in-hospital death. The time between both earliest suspected diagnosis and earliest confirmed diagnosis and the time of the chest radiograph were also calculated. These data were anonymized before delivery to the research team and were used to calculate NEWS2 and ISARIC-4C mortality.

Figure 1: Determination of the various scoring systems. **(A)** Anteroposterior radiograph in 78-year-old woman shows the classic changes of COVID-19 pneumonitis, which consist of opacification in a peripheral and basal distribution (arrows). **(B)** Radiograph shows calculation of the radiographic assessment of lung edema (RALE) score. The radiograph is divided into four quadrants. Each quadrant is assigned an intensity score and an opacification score. These are multiplied together for each quadrant, and all four scores are added together. The patient has a RALE score of 21. **(C)** Radiograph shows calculation of the Brixia score. The lungs are divided into six zones, and the degree of opacification is scored as follows: interstitial opacities, interstitial and alveolar opacities (interstitial predominate), and interstitial and alveolar opacities (alveolar predominate), scored as 1, 2, and 3, respectively. The patient has a Brixia score of 11 (1 + 2 + 2 + 1 + 2 + 3). The highest possible Brixia score is 18. **(D)** Radiograph shows percentage opacification, a simple visual estimate of the total percentage of lung parenchymal opacification.

Statistical Analysis

A *P* value less than .05 indicated statistical significance where relevant.

One author (T.C.) examined inter- and intraobserver reliability by using intraclass correlation coefficient for the RALE, Brixia, and percentage opacification scores. The intraclass correlation coefficient for interreader agreement was calculated by using a two-way random-effects model for absolute agreement and individual raters. A similar model with mixed effects was used for assessing intrareader agreement.

One author (C.J.C.) performed the cohort analysis. Follow-up started on the date of the reviewed chest radiograph and continued for 60 days, with complete follow-up for death or ICU admission.

Baseline demographic characteristics on the day of chest radiography were tabulated against the worst outcome observed during the follow-up period and were stratified by the initial decision on the eligibility of the patient for escalation because this assessment alters potential patient outcomes.

The median and interquartile range (IQR) for each chest radiograph score were tabulated and categorized (for percentage opacification, 0 = none, 1 ≤ 25%, 2 = 26%–50%, 3 = 51%–75%, 4 = 76%–100%; for RALE scores, 0–12, 13–24, 25–36, and 37–48; for Brixia scores, 0–5, 6–10, 11–15, and 16–18).

Kaplan-Meier survival curves were plotted and stratified by the categories of each score. The 95% CIs were calculated and estimated median survival times were presented where plausible for each category. Differences in overall survival by opacity score quantile were tested by using the log-rank test.

Finally, we assessed the additional predictive contribution of the opacity scores compared with previously published scores. Cox proportional hazard models were fitted
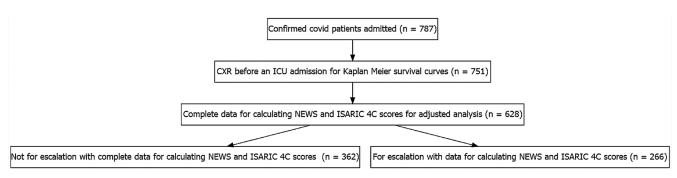
Figure 2: Flowchart of patients admitted with reviewed chest radiograph (CXR) from February to July 2020 at Nottingham University Hospitals. ICU = intensive care unit, ISARIC-4C = International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium, NEWS2 = National Early Warning Score 2.

Table 1: Baseline Demographic Characteristics by Worst Outcome during 60-Day Follow-up from Time of Earliest Chest Radiographic Examination

| Characteristic | Not for Escalation to ICU | | Eligible for Escalation to ICU | | |
	Survived	Died	Not Escalated	ICU Admission	Died
No. of patients	149	273	215	63	51
Age (y)*	82 (77–87)	82 (76–88)	55 (44–68)	57 (46–65)	69 (62–84)
No of men	71 (48)	168 (62)	115 (53)	37 (59)	28 (55)
Ethnicity					
Other or not stated ethnic group	22 (15)	33 (12)	57 (27)	23 (37)	12 (24)
Black or mixed ethnic group	<5	9 (3)	17 (8)	5 (8)	<5
Indian or Pakistani ethnic group	<5	6 (2)	22 (10)	4 (6)	<5
White ethnic group	119 (80)	225 (82)	119 (55)	31 (49)	36 (71)
30-day readmission	63 (42)	52 (19)	66 (31)	11 (17)	...
30-day mortality	...	197 (72)	...	...	42 (82)
Died out of hospital	...	18 (7)	...	...	6 (12)
30-day ICU admission	...	...	...	60 (95)	28 (55)
Length of stay (d)*	12 (7–18)	10 (4–18)	3 (1–6)	23 (12–37)	10 (4–22)
Opacification score*	10 (0–30)	20 (5–45)	20 (5–30)	40 (22–62)	40 (10–65)
Brixia score*	3 (0–6)	5 (1–8)	5 (1–8)	10 (7–12)	8 (3–12)
RALE score*	4 (0–12)	8 (1–17)	7 (1–15)	17 (10–26)	15 (4–25)
NEWS2					
ISARIC-4C recorded*	11 (9–13)	12 (10–14)	5 (3–8)	7 (4–8)	10 (8–12)
BMI < 20 kg/m^2	32 (21)	73 (27)	70 (33)	31 (49)	21 (41)
BMI > 30 kg/m^2	34 (23)	47 (17)	64 (30)	14 (22)	7 (14)
Smoking	9 (6)	24 (9)	10 (5)	1 (2)	5 (10)
Vaping	7 (5)	8 (3)	5 (2)	0 (0)	2 (4)
Alcohol risk	15 (10)	29 (11)	38 (18)	7 (11)	3 (6)
Charlson comorbidity index*	2 (1–3)	2 (1–4)	1 (0–2)	1 (0–1)	1 (0–2)

Note.—Unless otherwise indicated, data are numbers of patients and data in parentheses are percentages. BMI = body mass index, ICU = intensive care unit, IQR = interquartile range, ISARIC-4C = International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium, NEWS2 = National Early Warning Score 2, RALE = radiographic assessment of lung edema.
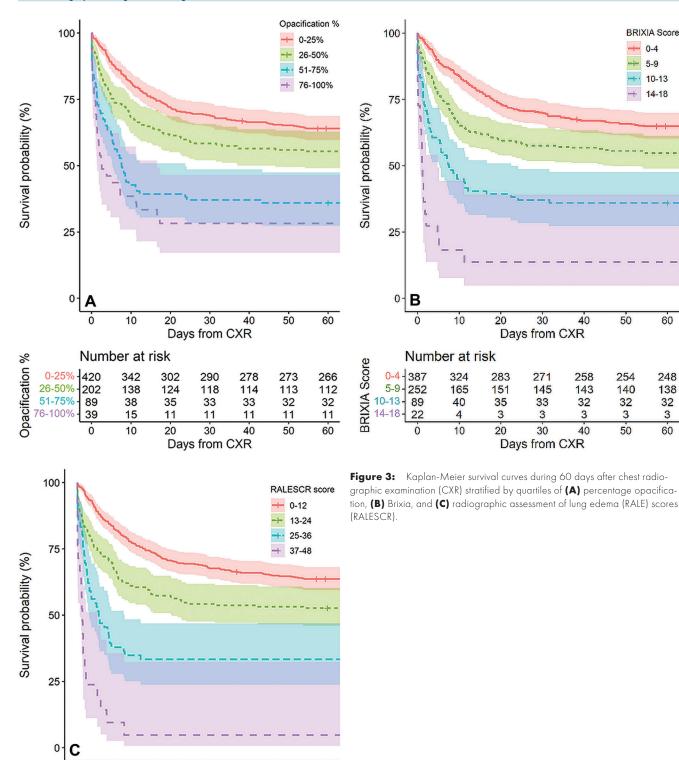
*Data are median; data in parentheses are interquartile range.

by using ISARIC-4C mortality or NEWS2 calculated on the day the chest radiograph was reviewed, and then each opacity score was added to the model. Akaike information criterion was used to compare models, and the Harrell C statistic was used to measure changes in discrimination. Models were stratified by the initial escalation decision. R version 4.0.3 (R Project for Statistical Computing) was used for statistical analysis.

Results

Interrater and Intrarater Reliability and Speed of Reading

The initial 50 selected chest radiographs were obtained in 50 different patients between February 22 and May 27, 2020. The patients ranged in age from 40 to 95 years (mean, 74 years ± 16 [standard deviation]), and 28 of 50 patients (56%) were men. Interrater reliability was good for all three scores, with intraclass correlation co-

Figure 3: Kaplan-Meier survival curves during 60 days after chest radiographic examination (CXR) stratified by quartiles of **(A)** percentage opacification, **(B)** Brixia, and **(C)** radiographic assessment of lung edema (RALE) scores (RALESCR).

Table 2: Cox Proportional Hazard Models Predicting Escalation to Intensive Care Unit or Death within 60 Days for All Patients

Parameter	ISARIC-4C				NEWS2			
	No Chest Radiography	Percentage Opacification	Brixia Score	RALE Score	No Chest Radiography	Percentage Opacification	Brixia Score	RALE Score
Percentage opacification								
26%–50%		1.3 (1.0, 1.6)				1.2 (1.0, 1.5)		
51%– 75%		2.5 (2.2, 2.8)				1.9 (1.6, 2.2)		
76%–100%		4.0 (3.6, 4.5)				2.8 (2.3, 3.2)		
Brixia score								
6–10			1.3 (1.0, 1.6)				1.1 (0.9, 1.4)	
11–15			2.6 (2.3, 3.0)				2.0 (1.6, 2.3)	
16–18			8.7 (8.2, 9.2)				5.9 (5.4, 6.4)	
RALE score								
13–24				1.4 (1.1, 1.6)				1.2 (0.9, 1.5)
25–36				3.3 (2.9, 3.6)				2.2 (1.8, 2.5)
37–48				9.9 (9.4, 10.5)				7.5 (7.0, 8.0)
ISARIC-4C	1.1 (1.1, 1.1)	1.1 (1.1, 1.1)	1.1 (1.1, 1.1)	1.1 (1.1, 1.1)				
NEWS2					1.2 (1.2, 1.3)	1.2 (1.2, 1.3)	1.2 (1.2, 1.3)	1.2 (1.2, 1.3)
C statistic	0.58	0.65	0.65	0.66	0.67	0.69	0.69	0.69
Observations	628	628	628	628	628	628	628	628
R^2	0.06	0.13	0.15	0.17	0.14	0.17	0.19	0.20
Akaike information criterion	3411	3368	3357	3341	3356	3337	3324	3313

Note.—Unless otherwise noted, values are hazard ratios; data in parentheses are 95% CIs. Data were in all patients (n = 628); there was admission to the intensive care unit or death in 276 patients. Data are reported by the opacity score of the initial reviewed chest radiograph, adjusted for ISARIC-4C or NEWS2 calculated on the day of the initial chest radiographic examination. Opacity scores have been divided into strata for analysis (percentage opacification: 0%–25%, 26%–50%, 51%–75%, and 76%–100%; Brixia: 0–5, 6–10, 11–15, and 16–18; RALE: 0–12, 13–24, 25–36, and 37–48); in each case, the lowest category is the reference group for analysis. ICU = intensive care unit, ISARIC-4C = International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium, NEWS2 = National Early Warning Score 2, RALE = radiographic assessment of lung edema.

efficients of 0.87 (95% CI: 0.80, 0.92) for RALE score, 0.86 (95% CI: 0.76, 0.92) for Brixia score, and 0.72 (95% CI: 0.48, 0.85) for percentage opacification (Table E2 [online]). Intrarater reliability was also good, with intraclass correlation coefficients of 0.86 (95% CI: 0.75, 0.92) for RALE score, 0.86 (95% CI: 0.75, 0.92) for Brixia score, and 0.88 (95% CI: 0.82, 0.91) for percentage opacification (Table E3 [online]). Across the studies timed by the three readers, the scoring of the studies took a mean of 58 seconds (standard error of the mean, 5.7 seconds) for RALE, 37 seconds for Brixia (standard error of the mean, 3.6 seconds), and 33 seconds for percentage opacification (standard error of the mean, 3.2 seconds).

Description of Cohort for Assessing Performance of Scores in Predicting Severity

A total of 787 consecutive patients were admitted to Nottingham University Hospitals with a positive RT-PCR result for SARS-CoV-2 and a chest radiograph available for review in the study period (Fig 2). For the unadjusted survival analysis, 751 patients (95.4%) had a chest radiograph available before any escalation to the ICU; the median age of this group was 76 years (IQR, 59–84 years), and 419 of 751 (56%) were men. This cohort was stratified by the decision for escalation; 329 patients were eligible for escalation to the ICU and 422 patients were not eligible for escalation. For the further adjusted survival analysis by ISARIC-

4C mortality or NEWS2, 628 patients had all the necessary data to calculate the scores (median age, 76 years; IQR, 61–84 years; 344 [57%] were men). The adjusted analysis was also stratified by the decision for escalation, with 266 eligible patients and 362 ineligible patients.

Patients were followed up for 60 days from the date of their initial chest radiographic examination. The median duration between the absolute time of chest radiography and earliest suspected diagnosis was 6 hours 40 minutes (IQR, 72 minutes to 69 hours). The median time between confirmed diagnosis and chest radiograph was 10 hours 43 minutes (IQR, 1 hour 38 minutes to 60 hours).

Patients' baseline demographic characteristics are shown by their worst recorded outcomes during this follow-up period in Table 1. Among the patients not eligible for escalation who eventually died, the median percentage opacification was 20% (IQR, 4%–45%); among those who survived, the median percentage opacification was lower, at 10% (IQR, 0%–30%). Overall median percentage opacification was higher among patients who were eligible for escalation (median percentage opacification, 20%; IQR, 5%–30%) among those who did not need escalation and 40% among those who were escalated and transferred to the ICU (IQR, 22%–62%) or died (IQR, 10%–65%). There was a similar pattern for the other opacity scores (Table 1).

Table 3: Cox Proportional Hazard Models Predicting Escalation to Intensive Care Unit or Death within 60 Days for Patients Eligible for Escalation

Variable	ISARIC-4C				NEWS2			
	No Chest Radiography	Percentage Opacification	Brixia Score	RALE Score	No Chest Radiography	Percentage Opacification	Brixia Score	RALE Score
Percentage opacification								
26%–50%		1.5 (1.0, 2.0)				1.4 (0.8, 1.9)		
51%–75%		4.1 (3.5, 4.6)				2.2 (1.6, 2.8)		
76%–100%		6.0 (5.4, 6.7)				4.0 (3.4, 4.7)		
Brixia score								
6–10			1.7 (1.2, 2.2)				1.2 (0.6, 1.7)	
11–15			4.8 (4.3, 5.4)				2.9 (2.3, 3.4)	
16–18			8.4 (7.7, 9.2)				4.6 (3.8, 5.4)	
RALE score								
13–24				1.7 (1.2, 2.2)				1.2 (0.7, 1.7)
25–36				5.4 (4.8, 5.6)				3.1 (2.5, 3.6)
37–48				6.8 (6.1, 7.5)				4.7 (4.0, 5.4)
ISARIC-4C	1.1 (1.1–1.2)	1.1 (1.1, 1.2)	1.1 (1.1, 1.2)	1.1 (1.1, 1.2)				
NEWS2					1.4 (1.3, 1.4)	1.3 (1.2, 1.4)	1.3 (1.2, 1.4)	1.3 (1.2, 1.4)
C statistic	0.63	0.72	0.72	0.71	0.74	0.76	0.77	0.77
Observations	266	266	266	266	266	266	266	266
R^2	0.07	0.19	0.20	0.21	0.20	0.25	0.26	0.27
Akaike information criterion	964	932	928	928	925	913	908	906

Note.—There were 266 patients eligible for escalation to intensive care unit (ICU); there were 90 patients with ICU admission or death. Unless otherwise noted, values are hazard ratios, with 95% CIs in parentheses. Data are reported by the opacity score of the initial reviewed chest radiograph for patients eligible for escalation, adjusted for ISARIC-4C or NEWS2 calculated on the day of the chest radiograph. Opacity scores have been divided into strata for analysis (percentage opacification: 0%–25%, 26%–50%, 51%–75%, and 76%–100%; Brixia: 0–5, 6–10, 11–15, and 16–18; RALE: 0–12, 13–24, 25–36, and 37–48); in each case, the lowest category is the reference group for analysis. ICU = intensive care unit, ISARIC-4C = International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium, NEWS2 = National Early Warning Score 2, RALE = radiographic assessment of lung edema.

Survival Curves

Figure 3 shows crude survival during the 60 days after undergoing chest radiography for the 751 patients with a chest radiograph during the admission before any ICU admission. Kaplan-Meier plots show each percentage opacification score stratified into categories (Fig 3). Each categorized score separated out the patients' overall survival. Median escalation-free survival was 7.6 days (95% CI: 5.4, 23.7) for percentage opacification of 51%–75% and 2.6 days (95% CI: 1.5, 16.6) for 76%–100% opacity ($P < .001$, log-rank test). The corresponding median survival times were 7.5 days (95% CI: 5.4, 22.4) for the categorized Brixia score of 11–15 and 1.15 days (95% CI: 0.7, 4.8) for a Brixia score of 16–20 ($P < .001$, log-rank test). Similarly, median survival was 5.18 days (95% CI: 2.9, 11.4) for a RALE score of 25–36 and 1.21 days (95% CI: 0.6, 4.8) for a RALE score of 36–48 ($P < .001$, log-rank test).

Adjusted Survival Analysis

A total of 628 patients had data regarding the required blood tests and observations on the day of the radiographic chest examination to enable the calculation of these measures. Compared with 0%–25% opacification, risk for ICU admission or death was increased by more than twofold (hazard ratio, 2.17; 95% CI: 1.57, 2.77) with 51%–75% opacification and by more than fourfold (hazard ratio, 4.04; 95% CI: 3.4, 4.69) with 76%–100% opacification.

Discrimination increased for percentage opacification, Brixia score, and RALE score. For example, the percentage opacification, Brixia score, and RALE score, adjusted for ISARIC-4C mortality, increased the C statistic from 0.58 to 0.65, 0.65, and 0.66, respectively, and improved the corresponding goodness of fit from an R^2 value of 0.06 to 0.13, 0.15, and 0.15, respectively (Table 2).

This was also observed when the results were stratified by eligibility for escalation in Tables 3 and 4. For example, in patients eligible for escalation, for each score adjusted for ISARIC-4C mortality, the C statistic improved from 0.63 to 0.72, 0.72, and 0.71 respectively, and the corresponding R^2 value improved from 0.07 to 0.19, 0.20, and 0.21, respectively. Models and opacity scores had greater discrimination and goodness of fit in patients eligible for escalation than in those who were ineligible; for example, in a comparison of patients ineligible for escalation adjusted for the ISARIC-4C mortality score, the C statistic improved from 0.56 to 0.60, 0.60, and 0.61, respectively, for each score, and the corresponding R^2 value improved from 0.03 to 0.07, 0.09, and 0.11, respectively. The highest discrimination was among patients eligible for escalation according to RALE scores and NEWS2 (C statistic, 0.77; R^2 = 0.27).

Table 4: Cox Proportional Hazard Models Predicting Death within 60 Days for Patients Not Eligible for Escalation

Parameter	ISARIC-4C				NEWS2			
	No Chest Radiography	Percentage Opacification	Brixia Score	RALE Score	No Chest Radiography	Percentage Opacification	Brixia Score	RALE Score
Percentage opacification								
26%–50%		1.2 (0.9, 1.6)				1.3 (0.9, 1.6)		
51%–75%		2.0 (1.6, 2.4)				1.8 (1.4, 2.2)		
76%–100%		2.6 (1.9, 3.2)				1.9 (1.2, 2.6)		
Brixia score								
6–10			1.3 (0.9, 1.6)				1.2 (0.9, 1.5)	
11–15			1.9 (1.4, 2.4)				1.7 (1.3, 2.2)	
16–18			9.5 (8.7, 10.3)				7.7 (6.8, 8.5)	
RALE score								
13–24				1.3 (0.9, 1.6)				1.3 (1.0, 1.7)
25–36				2.6 (2.2, 3.1)				2.1 (1.6, 2.6)
37–48				13.2 (12.4, 14.0)				11.5 (10.6, 12.3)
ISARIC-4C	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)	1.1 (1.0, 1.2)				
NEWS2					1.2 (1.1, 1.2)	1.2 (1.1, 1.2)	1.2 (1.1, 1.2)	1.2 (1.1, 1.2)
C statistic	0.56	0.6	0.6	0.61	0.64	0.65	0.65	0.66
Observations	362	362	362	362	362	362	362	362
R^2	0.03	0.07	0.09	0.11	0.10	0.13	0.15	0.16
Akaike information criterion	2086	2078	2070	2060	2057	2054	2046	2038

Note.—There were 362 patients who were not eligible for escalation to the intensive care unit; there were 186 deaths. Data are reported by the opacity score of the initial reviewed chest radiograph for patients ineligible for escalation, adjusted for ISARIC-4C or NEWS2 calculated on the day of the chest radiograph. Opacity scores have been divided into strata for analysis (percentage opacification: 0%–25%, 26%–50%, 51%–75%, and 76%–100%; Brixia: 0–5, 6–10, 11–15, and 16–18; RALE: 0–12, 13–24, 25–36, and 37–48); in each case, the lowest category is the reference group for analysis. ICU = intensive care unit, ISARIC-4C = International Severe Acute Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium, NEWS2 = National Early Warning Score 2, RALE = radiographic assessment of lung edema.

Discussion

The identification of patients with SARS-CoV-2 at high risk for deterioration is important for treatment. Such prediction may help triage, enable timely prevention of further deterioration, and aid in resource allocation. This cohort study in a single English city aimed to examine the effect of combining established clinical scores to predict adverse outcomes with a chest radiography severity score to predict key outcomes. We demonstrated that severity on chest radiographs in patients with SARS-CoV-2 can be rapidly and reliably reported with radiographic assessment of lung edema (RALE), Brixia, and percentage opacification scoring systems. All three scores were calculated in less than 1 minute; percentage opacification was the fastest to be calculated. Higher scores were strongly associated with admission to the intensive care unit (ICU) or death up to 60 days after SARS-CoV-2 diagnosis. Opacification on a radiograph of 76%–100% was associated with an overall adjusted risk for ICU admission or death that was three- to fourfold higher than that seen with opacification of 0%–25%. Patients in the highest categories of each score had a median time to ICU admission or death of just 1–2 days. Our data demonstrate that all methods of chest radiograph severity scoring can improve prediction of admission to the ICU or death when added to International Severe Acute

Respiratory and Emerging Infection Consortium: Coronavirus Clinical Characterization Consortium (ISARIC-4C) mortality score and National Early Warning Score 2 (NEWS2). In our cohort, the best approach, by discrimination, for predicting outcome on admission with SARS-CoV-2 infection was to combine a chest radiograph severity score with NEWS2 (C statistic, 0.77 for patients eligible for escalation with use of RALE score and NEWS2).

Our cohort had a full 60-day follow-up without loss of any patients to follow-up, stratified by eligibility for escalation of care, in contrast to other published reports (6,7,12–14,16,21,22). Notably, the chest radiograph scores were less discriminative in patients who were not eligible for escalation. This presumably relates to greater contribution of other factors, such as increasing age, frailty, and comorbidity, in this group. Scoring systems, such as NEWS2 (29,30) and ISARIC-4C mortality (31), can help predict the need for escalation in these patients. But neither they nor the closely related ISARIC-4C deterioration score (32) (assessing requirement of ventilatory support or critical care, or death) incorporates a chest radiograph scoring system.

There has been recent interest in the potential of scoring radiographic severity of SARS-CoV-2 pneumonitis changes to predict important patient outcomes. Several studies have reported such data (6–20). Our study focused on percentage

opacification, a simple and readily applicable scoring system, and compared it directly to Brixia and RALE scores, which have been studied more extensively. To our knowledge, only one other study has examined a percentage opacification score in relation to patient outcomes (18). One advantage of percentage opacification is that it is quick and straightforward to calculate, as reflected in the speed of its calculation compared with that of Brixia and RALE scores. The latter scores are also more complex to apply in clinical practice because of incorporation of zonal scoring and the need to assess the nature and intensity of the parenchymal opacification. Thus, RALE and Brixia may be more challenging for junior medical staff to apply in practice, and increased complexity may add to existing pressures on staff given the increasingly recognized risks of burnout (33).

Two other studies have also examined the effect of radiologic severity scoring systems in predicting patient outcome. Maroldi et al (12) examined the Brixia score in a cohort of similar size. That study incorporated radiographic scoring for radiographs obtained during the patients' SARS-CoV-2 admission to determine whether the Brixia score could predict outcomes. The results suggested that the Brixia score can be used to predict mortality. However, the study by Maroldi et al did not stratify by eligibility for escalation, had only in-hospital death as the outcome, and did not compare different scoring systems. Balbi et al (18) examined a smaller cohort of 340 patients from Italy and compared two scores: Brixia and percentage opacification on the admission chest radiograph. The authors concluded that Brixia helps predict mortality, whereas percentage opacification helps predict the need for ventilation. This study categorized the scores into binary variables, did not stratify by eligibility for escalation, and analyzed only in-hospital mortality.

Finally, there has been interest in artificial intelligence in scoring of chest radiograph (6,34) by comparing the performance of artificial intelligence systems to severity scores in predicting patient outcomes. The studied artificial intelligence algorithm incorporates scoring that appears similar to percentage opacification. The results of these artificial intelligence studies suggest a potential future role for artificial intelligence in calculating radiographic severity of SARS-CoV-2 pneumonitis, particularly if it could be combined with the calculation of an early-warning score, such as NEWS2.

Our study had limitations. It was a retrospective single-center study, all patients were RT-PCR positive (clinical diagnoses of SARS-CoV-2 infection were not included), and only admitted patients who underwent chest radiography were studied. These factors may limit the generalizability of our results, particularly to mild cases. In addition, the participating radiologists were not all chest radiologists but were all highly experienced in chest radiograph reporting, especially during the pandemic, thus reflecting real-world practice.

In conclusion, we show that three scoring systems for assessing chest radiograph severity can be calculated quickly for SARS-CoV-2 infection with good reliability and reproducibility. Overall, the higher the score, the worse the outcome for patients with SARS-CoV-2. Incorporation of all three scores into well-described risk prediction models substantially improved the ability to predict adverse outcomes of SARS-CoV-2 infection at hospital admission.

References

1. Dramé M, Tabue Teguo M, Proye E, et al. Should RT-PCR be considered a gold standard in the diagnosis of COVID-19? J Med Virol 2020;92(11):2312–2313.
2. Wong HYF, Lam HYS, Fong AH-TT, et al. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. Radiology 2020;296(2):E72–E78.
3. Cozzi A, Schiaffino S, Arpaia F, et al. Chest x-ray in the COVID-19 pandemic: radiologists' real-world reader performance. Eur J Radiol 2020;132:109272.
4. Ng MY, Wan EYF, Wong HYF, et al. Development and validation of risk prediction models for COVID-19 positivity in a hospital setting. Int J Infect Dis 2020;101:74–82.
5. Lechien JR, Chiesa-Estomba CM, Place S, et al. Clinical and epidemiological characteristics of 1420 European patients with mild-to-moderate coronavirus disease 2019. J Intern Med 2020;288(3):335–344.
6. Mushtaq J, Pennella R, Lavalle S, et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: analysis of 697 Italian patients. Eur Radiol 2021;31(3):1770–1779.
7. Borghesi A, Zigliani A, Golemi S, et al. Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy. Int J Infect Dis 2020;96:291–293.
8. Reeves RA, Pomeranz C, Gomella AA, et al. Performance of a severity score on admission chest radiograph in predicting clinical outcomes in hospitalized patients with coronavirus disease (COVID-19). AJR Am J Roentgenol 2021;217(3):623–632.
9. Xiao N, Cooper JG, Godbe JM, et al. Chest radiograph at admission predicts early intubation among inpatient COVID-19 patients. Eur Radiol 2021;31(5):2825–2832.
10. Kaleemi R, Hilal K, Arshad A, et al. The association of chest radiographic findings and severity scoring with clinical outcomes in patients with COVID-19 presenting to the emergency department of a tertiary care hospital in Pakistan. PLoS One 2021;16(1):e0244886.
11. Maguire D, Woods M, Richards C, et al. Prognostic factors in patients admitted to an urban teaching hospital with COVID-19 infection. J Transl Med 2020;18(1):354.
12. Maroldi R, Rondi P, Agazzi GM, Ravanelli M, Borghesi A, Farina D. Which role for chest x-ray score in predicting the outcome in COVID-19 pneumonia? Eur Radiol 2021;31(6):4016–4022.
13. Cocconcelli E, Biondini D, Giraudo C, et al. Clinical features and chest imaging as predictors of intensity of care in patients with COVID-19. J Clin Med 2020;9(9):2990.
14. Toussie D, Voutsinas N, Finkelstein M, et al. Clinical and chest radiography features determine patient outcomes in young and middle-aged adults with COVID-19. Radiology 2020;297(1):E197–E206.
15. Berrill M, Karaj J, Zamfir G, et al. Chest radiographs may assist in predicting the outcome in the early phase of Covid-19. UK district general hospital experience of Covid-19 first wave. Expert Rev Respir Med 2021;15(4):537–541.
16. Joseph NP, Reid NJ, Som A, et al. Racial and ethnic disparities in disease severity on admission chest radiographs among patients admitted with confirmed coronavirus disease 2019: a retrospective cohort study. Radiology 2020;297(3):E303–E312.

17. Monaco CG, Zaottini F, Schiaffino S, et al. Chest x-ray severity score in CO-VID-19 patients on emergency department admission: a two-centre study. Eur Radiol Exp 2020;4(1):68 [Published correction appears in Eur Radiol Exp 2021;5(1):17.].

18. Balbi M, Caroli A, Corsi A, et al. Chest X-ray for predicting mortality and the need for ventilatory support in COVID-19 patients presenting to the emergency department. Eur Radiol 2021;31(4):1999–2012.

19. De Sanctis V, Bedair EMA, Soliman AT, Nair AP, Al Masalamani MA, Yassin M. Proposed scoring system for evaluating clinico-radiological severity of CO-VID- 19 using plain chest x-ray (CXR) changes (CO X-RADS): preliminary results. Acta Biomed 2020;91(4):e2020172.

20. Al-Smadi AS, Bhatnagar A, Ali R, Lewis N, Johnson S. Correlation of chest radiography findings with the severity and progression of COVID-19 pneumonia. Clin Imaging 2021;71:17–23.

21. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiol Med (Torino) 2020;125(5):509–513.

22. Warren MA, Zhao Z, Koyama T, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax 2018;73(9):840–846.

23. Covino M, De Matteis G, Burzo ML, et al. Predicting in-hospital mortality in COVID-19 older patients with specifically developed scores. J Am Geriatr Soc 2021;69(1):37–43.

24. Holten AR, Nore KG, Tveiten CEVWK, Olasveengen TM, Tonby K. Predicting severe COVID-19 in the emergency department. Resusc Plus 2020;4:100042.

25. Myrstad M, Ihle-Hansen H, Tveita AA, et al. National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19 - a prospective cohort study. Scand J Trauma Resusc Emerg Med 2020;28(1):66.

26. Bradley P, Frost F, Tharmaratnam K, Wootton DG; NW Collaborative Organisation for Respiratory Research. Utility of established prognostic scores in COVID-19 hospital admissions: multicentre prospective evaluation of CURB-65, NEWS2 and qSOFA. BMJ Open Respir Res 2020;7(1):e000729.

27. Rigoni M, Torri E, Nollo G, Delle Donne L, Cozzio S. NEWS2 is a valuable tool for appropriate clinical management of COVID-19 patients. Eur J Intern Med 2021;85:118–120.

28. Crooks CJ, West J, Fogarty A, et al. Predicting the need for escalation of care or death from repeated daily clinical observations and laboratory results in patients with SARS-CoV-2 during 2020: A retrospective population-based cohort study from the United Kingdom. medRxiv. https://www.medrxiv.org/content/10.1101/2020.12.14.20248181v1. Posted December 16, 2020. Accessed December 16, 2020.

29. Carr E, Bendayan R, Bean D, et al. Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. BMC Med 2021;19(1):23.

30. Royal College of Physicians. National Early Warning Score (NEWS) 2: standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. London, England: Royal College of Physicians, 2017.

31. Knight SR, Ho A, Pius R, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 2020;370:m3339 [Published correction appears in BMJ 2020;371:m4334.].

32. Gupta RK, Harrison EM, Ho A, et al. Development and validation of the ISARIC 4C Deterioration model for adults hospitalised with COVID-19: a prospective cohort study. Lancet Respir Med 2021;9(4):349–359.

33. Morgantini LA, Naha U, Wang H, et al. Factors contributing to healthcare professional burnout during the COVID-19 pandemic: A rapid turnaround global survey. PLoS One 2020;15(9):e0238217.

34. Ebrahimian S, Homayounieh F, Rockenbach MABC, et al. Artificial intelligence matches subjective severity assessment of pneumonia for prediction of patient outcome and need for mechanical ventilation: a cohort study. Sci Rep 2021;11(1):858.

Erratum

Originally published in:

https://doi.org/10.1148/radiol.2021210986

Chest Radiograph Scoring Alone or Combined with Other Risk Scores for Predicting Outcomes in COVID-19

Iain Au-Yong, Yutaro Higashi, Elisabetta Giannotti, Andrew Fogarty, Joanne R. Morling, Matthew Grainge, Andrea Race, Irene Juurlink, Mark Simmonds, Steve Briggs, Simon Cruickshank, Susan Hammond-Pears, Joe West, Colin J. Crooks, Timothy Card

Erratum in:

https://doi.org/10.1148/radiol.219029

In the Appendix E1 (online), links to Figure 3B, 3C, and 3D were incorrect and have been changed to **1B**, **1C**, and **1D**.