Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Respiratory Medicine 176 (2021) 106271

Contents lists available at ScienceDirect

Respiratory Medicine

journal homepage: http://www.elsevier.com/locate/rmed

Predictive value of chest CT scoring in COVID-19 patients in Wuhan, China: A retrospective cohort study

Shu Li [a], Shaoyu Liu [a], Ben Wang [b], Qiuyu Li [c], Hua Zhang [d], Lin Zeng [d], Hongxia Ge [a], Qingbian Ma [a], Ning Shen [c,*]

[a] Department of Emergency Medicine, Peking University Third Hospital, Beijing, China
[b] Department of Orthopedics, Peking University Third Hospital, Beijing, China
[c] Department of Pulmonary and Critical Care Medicine, Peking University Third Hospital, Beijing, China
[d] Clinical Epidemiology Research Center, Peking University Third Hospital, Beijing, China

ARTICLE INFO

Keywords:
COVID-19
Computed tomography
Pneumonia
Prognosis
Severity of illness

ABSTRACT

Background: Computed tomography (CT) findings of COVID-19 patients were demonstrated by cases series and descriptive studies, but quantitative analysis performed by clinical doctors and studies on its predictive value were rarely seen. The aim of the study is to analyze CT score in COVID-19 patients and explore its predictive value.

Materials and methods: We conducted a retrospective cohort study among confirmed COVID -19 patients with available CT images between February 8, 2020 and March 7, 2020. The lung was divided into six zones by the level of tracheal carina and the level of inferior pulmonary vein bilaterally on CT. Ground-glass opacity (GGO), consolidation, crazy-paving pattern and overall lung involvement were rated by Likert scale of 0–4 or binary as 0 or 1. Global severity score for each targeted pattern was calculated as total score of six zones.

Results: There were 53 patients and 137 CT scans included in the study. There were 18(34%) of the patients classified as moderate cases while 35(66%) patients were severe/critical cases. Severe/critical patients had higher CT scores in several types of abnormalities than moderate patients from the second week to the fourth week post symptom onset. Overall lung involvement score in the second week demonstrated predictive value for severity with a sensitivity of 81.0% and specificity of 69.2%.

Conclusions: Our modified semi-quantitative CT scoring system for COVID-19 patients demonstrated feasibility. Overall lung involvement score on the second week had predictive value for clinical severity and could be indicator for further treatment.

1. Introduction

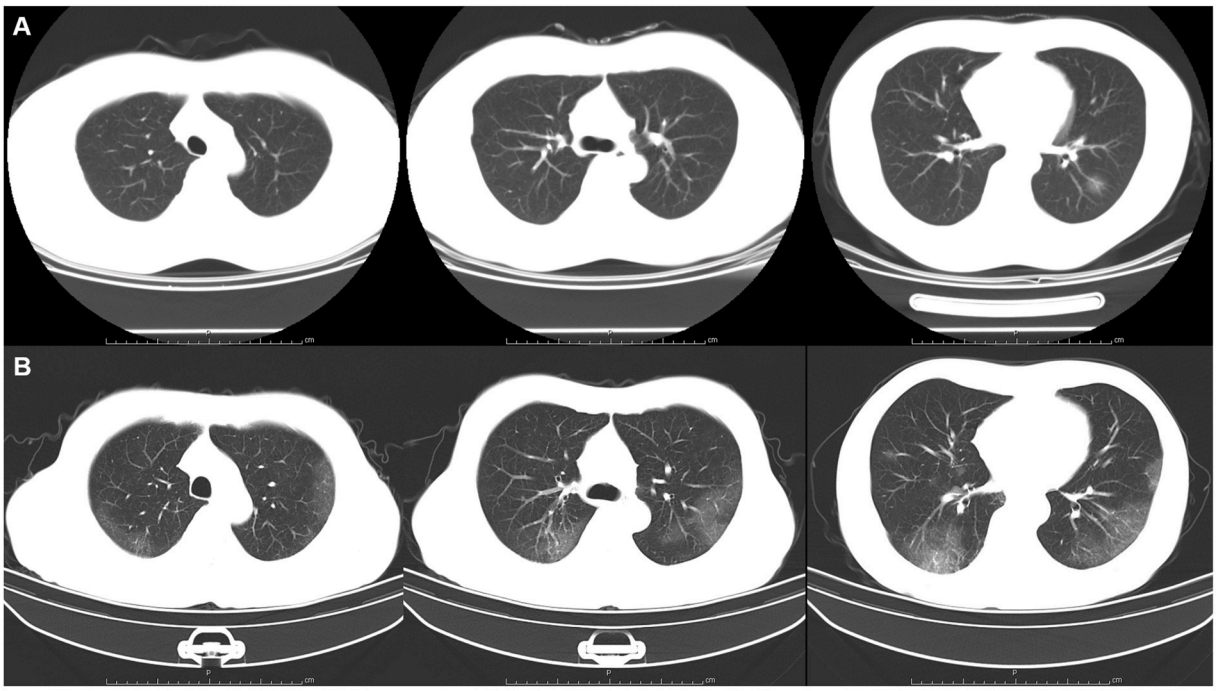
By late April 2020, approximately over 3,000,000 patients were diagnosed with COVID-19 globally since outbreak of the disease at the end of December 2019. It was reported that there could be up to over 20,000 cases newly diagnosed or over 900 death in one country in a single day and there have been more than 10 countries all over the world with mortality rate exceeding 10% [1]. One of the existing barriers for COVID-19 treatment is to detect patients who might present to have stable vital sign as mild/moderate cases on arrival but suffered from dramatic exacerbation or even fatal outcome. It is of great importance to identify insidious onset and implement intervention and resource allocation at an early stage to reduce overall mortality rate.

Chest computed tomography (CT) is supposed to be golden standard and important diagnostic tool for lung diseases. CT has been widely used for the diagnosis and severity evaluation for clinical purpose during the pandemic. There were quite a few case series and descriptive studies of CT manifestations and evolvement in COVID-19 patients. However, quantitative and subgroup analysis was not reported yet. Furthermore, almost all the studies on CT score for COVID-19 patients were senior radiologists rather than clinical doctors.

Here, we present the first application of CT score by clinical doctors to analysis the serial CT features in confirmed COVID-19 patients in different clinical severity groups and explore its predictive value.

* Corresponding author. Department of Pulmonary and Critical Care Medicine, Peking University Third Hospital, No. 49 North Garden Road, Haidian District, Beijing, 100191, China.
E-mail address: puh3shenning@bjmu.edu.cn (N. Shen).

https://doi.org/10.1016/j.rmed.2020.106271
Received 7 May 2020; Received in revised form 4 November 2020; Accepted 25 November 2020
Available online 28 November 2020
0954-6111/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).



Patients	CT scoring item	LUL	RUL	LML	RML	LLL	RLL	Total
Patient A	GGO score	0	0	0	0	1	0	1
	Consolidation score	0	0	0	0	0	0	0
	Crazy-paving pattern score	0	0	0	0	0	0	0
	Overall lung involvement score	0	0	0	0	1	0	1
Patient B	GGO score	1	1	1	1	1	1	6
	Consolidation score	0	0	0	0	0	0	0
	Crazy-paving pattern score	0	0	0	0	0	0	0
	Overall lung involvement score	1	1	1	1	1	1	6

For GGO, consolidation, overall lung involvement, use Likert scale (0=absent; 1=1 –25%; 2=26–50%; 3=51–75%; 4=76–100%). Total 0-24. For crazy-paving pattern, use 1 for present, 0 for absent. Total 0-6.

Fig. 1. CT scoring system demonstration. A: CT images of a 40-year-old male patient who presented with fever, cough and mild hemoptysis and was identified as moderate case by clinical criteria. The axial non-contrast CT images showed unilateral focal pure ground glass opacity (GGO) with no crazy-paving pattern or consolidation. B: CT images of a 73-year-old male who presented with fever, cough, fatigue, and dyspnea and was defined as severe case by clinical criteria. The axial non-contrast CT images showed bilateral pure ground glass opacity (GGO) with no crazy-paving pattern or consolidation. The lesions were distributed in the peripheral and posterior part of the lung (severe case). Each of the three images of the same patient showed bilateral upper zone, median zone, and lower zone.

2. Methods

2.1. Study design

The study protocol was reviewed by the Institutional Review Board of the hospital and received a waiver of informed consent (IRB00006761-M2020060). This was a single-center, retrospective, observational cohort study between February 8, 2020 and March 7, 2020 at a tertiary medical center. Admitted adult COVID-19 patients with pneumonia were consecutively recruited. Patients without obtainable CT images within five weeks post symptom onset were excluded.

2.2. Diagnosis and severity of disease

Clinical data was collected from electronic medical records. The diagnosis of COVID -19 were confirmed via laboratory testing with positive result of real-time reverse transcriptase polymerase chain reaction (rRT-PCR), or retrospectively by positive IgM and IgG [2]. The severity of disease was retrospectively classified as severe/critical or moderate base on the worst situation during hospitalization [2]. Criteria for severe cases were as following:(1) respiratory distress, RR $\geq$ 30

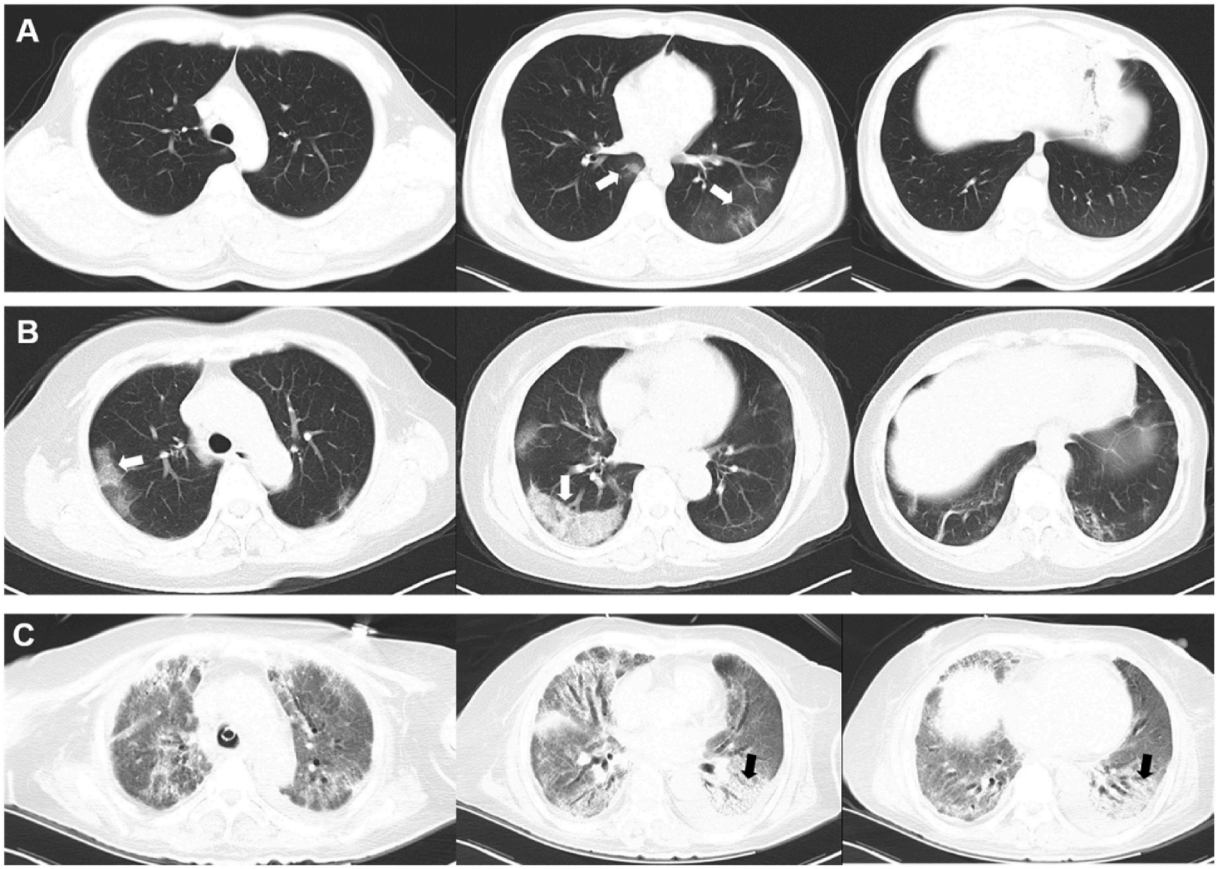


Fig. 2. CT images from patients of different disease severity. A: CT images of a 43-year-old male patient who presented with fever, shiver, cough and headache, and was identified as moderate case by clinical criteria. The axial non-contrast CT images showed multi-focal ground glass opacity (GGO) with curvilinear lines(arrow) but no consolidation. B: CT images of a 64-year-old male patient who presented with fever, shiver, cough, diarrhea and dyspnea, and was classified as sever case by clinical criteria. The axial non-contrast CT images showed bilateral peripheral distributed ground glass opacity with reticulation, presenting as crazy-paving pattern (arrows). C: CT images of a 73-year-old female patient who was transferred to our hospital with sever hypoxia who was later intubated and put on ECMO. The axial non-contrast CT images showed diffuse ground glass opacity (GGO) with bilateral posterior distributed crazy-paving pattern and consolidation (arrows).

beats/min; (2) resting blood oxygen saturation $\leq$ 93%; or (3) partial pressure of arterial blood oxygen (PaO_2)/oxygen concentration (FiO_2) $\leq$ 300 mmHg. Criteria for critical patients were: (1) respiratory failure needing mechanical ventilation; (2) shock; (3) other organ failure needing ICU monitoring treatment.

2.3. CT scan protocol

All of the CT scans were performed during a single breath-hold with 1–10 mm slice thickness and 100-120kv voltage without contrast on BrightSpeed/LightSpeed scanner (GE Medical Systems; Milwaukee, WI, the US), SOMATOM Force scanner (Siemens Healthineers; Erlangen, Germany), uCT530/780 scanner (United Imaging Healthcare Co., Shanghai, China), Aquilion scanner (Toshiba, Tokyo, Japan), and Ingenuity scanner (Philips, Amsterdam,Netherlands).

2.4. CT evaluation

The lung was divided into six zones (upper, middle, and lower on both sides) by the level of the tracheal carina and the level of the inferior pulmonary vein bilaterally on CT, using a modified scoring system [3,4]. The observers recognized ground-glass opacity (GGO), consolidation and crazy-paving pattern following Fleischner Society definitions [5,6]. Bronchiectasis, cavity, pleural effusion, etc., were not included in CT reading and analysis because of low incidence [7].The reviewers evaluated the extent of the targeted patterns and overall affected lung parenchyma for each zone, using Likert scale (0 = absent; 1 = 1–25%; 2 = 26–50%; 3 = 51–75%; 4 = 76–100%). Thus, GGO score, consolidation score, and overall lung involvement score were sum of 6 zones ranging from 0 to 24. For crazy-paving pattern, it was only coded as absent or present (0 or 1) for each zone and therefore ranging from 0 to 6 (See Fig. 1). Existence of certain abnormality, bilateral involvement or multi-zone involvement was identified as present for a single patient if ever observed in any of his/her CT. Multi-zone involvement was considered as $\geq$ 3 zones to avoid double counting from adjacent zones.

A test set of 50 CT images were reviewed by two radiologists of over 5 years' experience (YQZ and GG) and two ED attending physicians with approximately 10 years' experience (SL and SYL) independently. Then, after optimal inter-rater reliability was reached between radiologists and clinicians, all CT images were reviewed by the two ED attending physicians. A second round of image reading was performed to resolve discrepancy. Final score was determined by average score from the two reviewers if consensus was not reached.

2.5. Statistical analysis

The hypothesis of this study was that CT score can effectively distinguish severe/critical patients from moderate cases and the area under receiver operating characteristic curves (ROC curves) was greater than 0.5 with $\alpha = 0.05$ (one side), $\beta = 0.1$. The proportion between groups was set as 1:1. At least 16 patients were to be included for each group. Median with interquartile range (IQR) was used for continuous variables and counts and frequencies for categorical variables. Continuous data were compared with Mann-Whitney *U* Test, and categorical

Table 1
Demographics and clinical features of patients.

Variable	Total (n = 53) n (%)		Moderate (n = 18) n (%)		Severe/critical (n = 35) n (%)		*p value**
Demographics							
Age, years, median (IQR)	64.8	(49.0,73.0)	50.7	(42.0,69.5)	66.8	(55.8,73.2)	0.066
Male	27	(50.9)	7	(38.9)	20	(57.1)	0.208
Comorbidity							
Hypertension	24	(45.3)	7	(38.9)	17	(48.6)	0.502
Diabetes	8	(15.1)	3	(16.7)	5	(14.3)	0.820
Smoker	7	(13.2)	1	(5.6)	6	(17.2)	0.225
Clinical characteristics							
Wuhan resident	46	(86.8)	16	(88.9)	30	(85.7)	>0.999
Clustered onset	26	(49.1)	6	(3.3%)	20	(57.1)	0.101
Symptoms							
Fever	41	(77.4)	13	(72.2)	28	(80.0)	0.526
T_{max}, °C, median (IQR)	38.5	(38.0,39.0)[a]	38.4	(38.0,39.1)[b]	38.5	(38.0,39.0)[c]	0.922
Cough	48	(90.6)	14	(77.8)	34	(97.1)	**0.040**
Dyspnea	41	(77.4)	12	(66.7)	29	(82.9)	0.190
Diarrhea	28	(52.8)	11	(61,6)	17	(48.6)	0.386
Time before first CT, median (IQR)	9.0	(4.5,12.0)	9.0	(3.0,12.3)	9.0	(6.0,12.0)	0.770
Time before admission, days, median (IQR)	15.0	(10.5,19.5)	14.5	(6.8,18.3)	15.0	(11.0,22.0)	0.288
Initial WBC count, $\times 10^9$/L, median (IQR)	5.1	(3.9,6.5)[d]	5.3	(3.8,6.5)[e]	5.1	(3.8, 6.7)[f]	0.740
Initial lymphocyte count, $\times 10^9$/L, median (IQR)	1.0	(0.6,1.4)[g]	1.1	(0.8,1.5)[h]	0.7	(0.5,1.3)[i]	0.134
Lymphocyte count on admission, $\times 10^9$/L, median (IQR)	1.3	(0.7,1.6)	1.4	(1.1,1.7)	1.1	(0.5,1.5)	0.051
Initial PLT count, $\times 10^9$/L, median (IQR)	134.0	(118.0,199.0)[j]	132.0	(104.0,175.0)[k]	164.0	(130.0, 225.8)[l]	0.118
BUN on admission, mmol/L, median (IQR)	4.2	(3.1,6.4)	4.0	(3.0,5.0)	4.3	(3.2,7.9)	0.264
qSOFA score on admission, median (IQR)	0	(0,1)	0	(0,0)	0	(0,1)	**0.029**
CURB65 score on admision, median (IQR)	1	(0,2)	0	(0,1)	1	(1,2)	**0.002**

*All bold values were statistically significant.
Abbreviations: T_{max}: maximum body temperature before admission; WBC: white blood cell; PLT: platelet; BUN: blood urea nitrogen; qSOFA: Quick Sequential (Sepsis-Related) Organ Failure Assessment; CURB 65: Confusion, Urea, Respiratory Rate and Age 65.

data with Chi-Square Test or Fisher's exact test. The inter-rater reliability was measured by intraclass correlation coefficient (ICC) among physicians. Values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 were indicative of poor, moderate, good, and excellent reliability, respectively [8]. To evaluate the predictive power of CT score on severity scale of COVID-19, receiver operating characteristic curves were plotted, and corresponding areas under the curve (AUCs) were determined. The 95% confidence intervals for the corresponding AUCs and *p* values were calculated from Mann Whitney *U* test. Analyses were two-sided with a significance level of 0.05 and performed using IBM SPSS Statistics software, version 22.0 for Windows. ROC curve comparison was made using MedCalc, version 19.1 for Windows. PASS (version 11 for Windows) was used to estimate the sample size.

3. Results

There were 58 admitted patients confirmed with COVID-19 pneumonia between February 8, 2020 and March 7, 2020 while 5 of them with no CT scan qualified for further analysis. There were 53 patients with 168 available CT collected. However, there were 18 CT scans performed after 5 weeks post symptom onset and another 13 repeated CT scans within one week excluded. 53 patients with 137 CT scans entered final analysis.

3.1. Clinical characteristics

Of all the 53 patients, 46(86.8%) were Wuhan residents and 26 (49.1%) were clustered onset cases. The median age was 65(IQR 49, 73) years. Approximately half (50.9%) of them were male patients. As initial symptom, cough occurred in 48(90.6%) patient and 41 (77.4%) patients had fever. The median time to hospital admission was 15 days. There were 21(39.6%) patients underwent deterioration and escalation of care after admission. Ultimately, there were 18(34%) moderate (Fig. 2A) and 35(66%) severe/critical cases in our cohort (Fig. 2B and C). Other patient characteristics are listed in Table 1.

The median time from symptom onset to first CT scan was 9 days while the median time of interval for CT scan in our study was 7 days (IQR 5.5, 10). There were 51 (96.2%) patients had GGO, 38 (71.7%) had consolidation, and 42(79.2%) had crazy-paving pattern throughout

Table 2
Characteristics of CT scans.

Variable	Total n = 53		Moderate n = 18		Severe/critical n = 35		*P value*
Number of CT scan, n	137		47		90		
1st week, n	20		7		13		
2nd week, n	34		13		21		
3rd week, n	29		11		18		
4th week, n	30		10		20		
5th week, n	24		6		18		
Lung involvement							
Bilateral involvement, n (%)	49	(92.5)	15	(83.3)	34	(97.1)	0.108
Multiple lung zones involvement, n (%)	48	(90.6)	16	(88.9)	32	(91.4)	>0.999
GGO, n (%)	51	(96.2)	18	(100)	33	(94.3)	0.543
Consolidation, n (%)	38	(71.1)	11	(61.1)	27	(77.1)	0.220
Crazy-paving pattern, n (%)	42	(79.2)	13	(72.2)	29	(82.9)	0.373
Time consumed for CT reading, sec, median (IQR)	118.5	(85.0,164.3)	91.5	(70.0,120.5)	133.8	(92.3179.6)	

*GGO: ground-glass opacity.

Table 3
Inter-rater reliability.

	ICC[a]	Lower 95% CI	Upper 95% CI	*P value*
ICC between two radiologists				
GGO score	0.801	0.674	0.882	<0.001
Consolidation score	0.703	0.530	0.820	<0.001
Crazy-paving pattern score	0.593	0.379	0.747	<0.001
Overall lung involvement score	0.814	0.693	0.890	<0.001
ICC between two clinicians				
GGO score	0.784	0.648	0.781	<0.001
Consolidation score	0.794	0.663	0.877	<0.001
Crazy-paving pattern score	0.442	0.189	0.640	0.001
Overall lung involvement score	0.861	0.768	0.919	<0.001
ICC between radiologists and clinicians				
GGO score	0.907	0.842	0.946	<0.001
Consolidation score	0.887	0.809	0.934	<0.001
Crazy-paving pattern score	0.756	0.607	0.854	<0.001
Overall lung involvement score	0.933	0.885	0.962	<0.001

[a] ICC: intraclass correlation coefficient.

clinical course. Bilateral lung involvement was observed in 49 (92.5%) patients (See Table 2).

3.2. The feasibility of CT score for clinicians

The intraclass correlation coefficients (ICC) between two radiologists ranged from 0.593 to 0.814. The ICCs between two clinicians were 0.442–0.861. The ICCs between radiologists and clinicians were 0.756–0.933, which were classified as good or excellent (See Table 3). The median time consumed for each CT by clinicians was 118.5 s.

3.3. Evolution of CT score and differences between moderate and severe/critical patients

In order to analyze changes of lung involvement with time and differences between moderate and severe/critical patients, clinical course was divided into five stages by week (1–5 week) according to time from initial symptom onset. Temporal changes of CT scores roughly depicted bell-shaped curves. In severe/critical patients, all the metrics reached their peak value in the third week and declined later. In moderate patients, crazy-paving pattern score reached its peak value in the second week. The consolidation score presented at a relatively constant low level. Overall lung involvement score and GGO score stayed at a moderate level with minor variation (See Fig. 3 and Table 4). However, further quantitative analysis of data trending by ANOVA for repeated measurement design was not applicable because of considerable missing data.

Furthermore, overall lung involvement score, GGO score, consolidation score, and crazy-paving pattern score were compared by week between groups of different clinical severity. Severe/critical patients had higher overall lung involvement score than moderate patients from the second week to the fourth week, so it was with GGO score (See Fig. 3A and B and Table 4). Additionally, consolidation score and crazy-

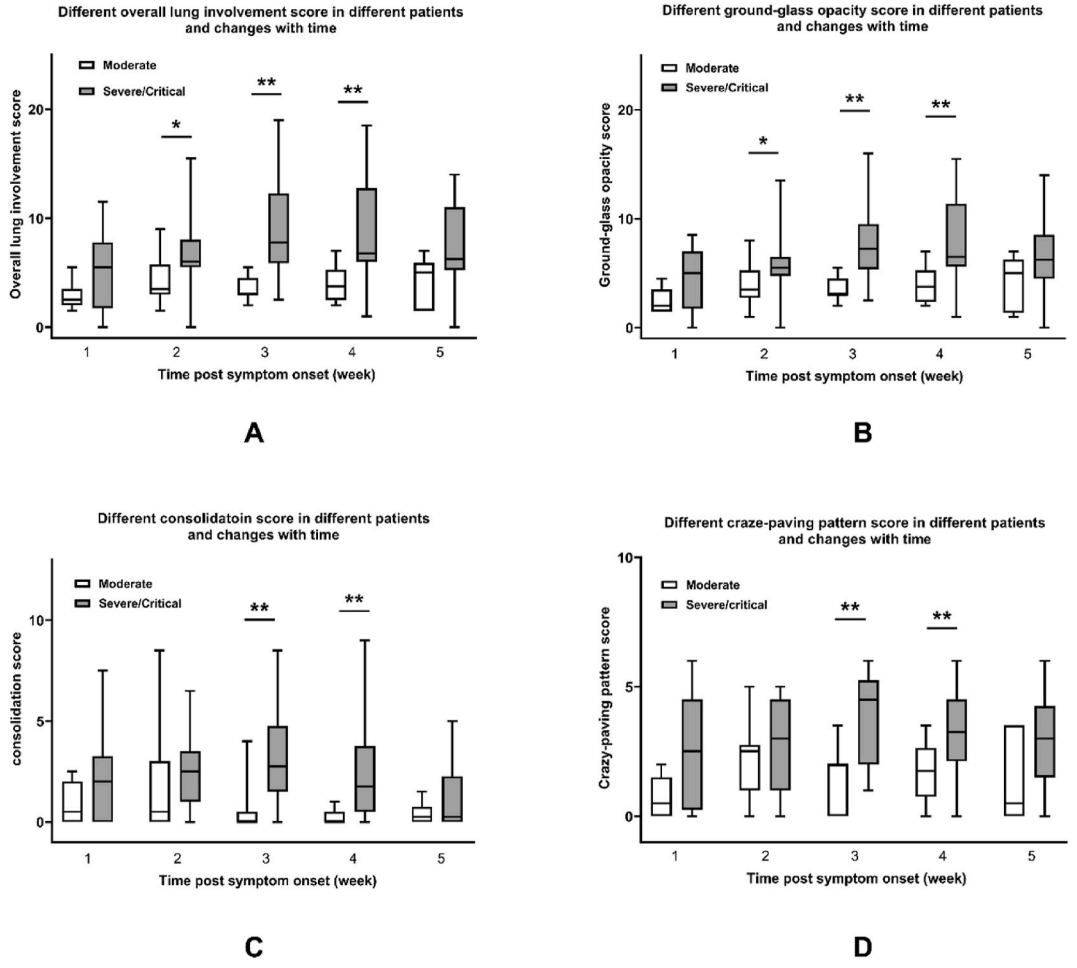


Fig. 3. CT scores in COVID-19 patients of different severity based on time course. A: Overall lung involvement score B: Ground-glass opacity score C: Consolidation score D: Crazy-paving pattern score.

Table 4
Longitudinal changes of CT score in COVID-19 patients and differences between groups.

Time post symptom onset	Clinical severity	Type of CT findings							
		Ground-glass opacity		Consolidation		Crazy-paving pattern		Overall lung involvement	
Week 1	Total n = 20	3.75	(1.63,5.88)	1.50	(0.00,2.50)	1.25	(0.13,2.88)	3.50	(2.00,5.88)
	Moderate n = 7	2.00	(1.50,3.50)	0.50	(0.00,2.00)	0.50	(0.00,1.50)	2.50	(2.00,3.50)
	Severe/critical n = 13	5.00	(1.75,7.00)	2.00	(0.00,3.25)	2.50	(0.25,4.50)	5.50	(1.75,7.75)
	*p**	0.111		0.223		0.101		0.189	
Week 2	Total n = 34	5.50	(3.38,6.00)	1.50	(0.50,3.50)	2.50	(1.25,4.00)	5.50	(3.50,7.62)
	Moderate/critical n = 13	3.50	(2.75,5.25)	0.50	(0.00,3.00)	2.50	(1.00,2.75)	3.50	(3.00,5.75)
	Severe n = 21	5.50	(4.75,6.50)	2.50	(1.00,3.50)	3.00	(1.00,4.50)	6.00	(5.00,8.00)
	p	**0.018**		0.218		0.292		**0.016**	
Week 3	Total n = 29	5.05	(3.25,8.50)	2.00	(0.00,3.75)	2.50	(1.50,5.00)	5.50	(3.25,8.75)
	Moderate n = 11	3.00	(3.00,4.50)	0.00	(0.00,0.50)	2.00	(0.00,2.00)	3.00	(3.00,4.00)
	Severe/critical n = 18	7.25	(5.38,9.50)	2.75	(1.50,4.75)	4.50	(2.00,5.25)	7.75	(5.88,12.25)
	p	**<0.001**		**0.003**		**0.002**		**<0.001**	
Week 4	Total n = 30	6.00	(3.00,7.63)	0.75	(0.00,2.13)	3.00	(1.50,4.00)	6.00	(3.00,9.13)
	Moderate n = 10	3.75	(2.38,5.25)	0.00	(0.00,5.00)	1.75	(0.75,2.63)	3.75	(2.5,5.25)
	Severe/critical n = 20	6.50	(5.625,11.38)	1.75	(0.50,3.75)	3.25	(2.13,4.50)	6.75	(6,12.75)
	p	**0.007**		**0.002**		**0.008**		**0.005**	
Week 5	Total n = 24	6.00	(4.50,7.00)	0.25	(0.00,1.50)	3.00	(1.00,4.00)	6.00	(4.50,7.00)
	Moderate n = 6	5.00	(1.38,6.25)	0.25	(0.00,0.75)	0.50	(0.00,3.50)	5.00	(1.5,5.88)
	Severe/critical n = 18	6.25	(4.50,8.50)	0.25	(0.00,2.25)	3.00	(1,50,4.25)	6.25	(5.25,11.00)
	p	0.141		0.568		0.064		0.076	

*p value $\leq$ 0.05 was considered statistically significant. Therefore, all bold values were statistically significant.

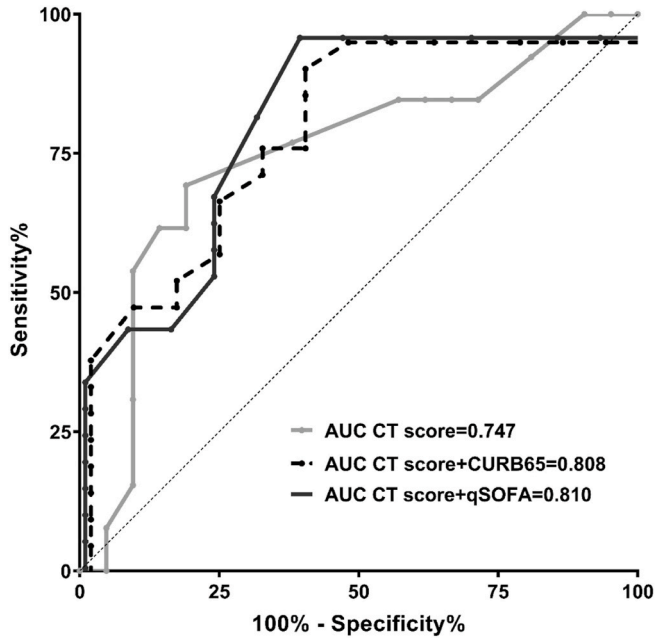


Fig. 4. Receiver operating characteristic curves and AUCs of the predictors for severity of disease.

paving pattern score in moderate patients were lower than severe/critical patients during the third and fourth week (See Fig. 3C and D and Table 4).

From all above, the second week was the earlies time point to distinguish these two groups of patients. To evaluate predictive value of CT score on clinical severity, 34 sets of CT scores on the second week were utilized to generate ROC curve (Fig. 4). The best AUC 0.747 (0.566,0.928), p = 0.017, was obtained for overall lung involvement score. The optimum cut-off value was higher than 5.25, with a sensitivity of 81.0% and specificity of 69.2%. Combined model was developed in order to improve predict capacity. qSOFA and CURB 65 score were selected as significant variables. Combined model which included either qSOFA or CURB 65 score, increased AUC to 0.810 (95% CI 0.654, 0.956) and 0.808(95% CI 0.658, 0.958), with specificity of 61.5% for both and sensitivity of 95.2% and 90.5%, respectively (See Table 5, Fig. 4). However, ROC comparison analysis failed to demonstrate significant differences of AUCs between original model and combined models.

4. Discussion

This is the first study by clinical doctors that compared the longitudinal changes of CT manifestations between moderate and severe/critical COVID-19 patients through a semi-quantitative visual scoring system. Severe/critical patients had higher overall lung involvement score and GGO score than moderate patients since the second week while consolidation score and crazy-paving pattern score reached their

Table 5
AUCs of ROC curves.

	AUC	Standard error	*p**	95%CI		cut-off value	sensitivity	specificity
				Lower	upper			
Predict factor								
Overall lung involvement score	0.747	0.092	**0.017**	0.566	0.928	5.25	81.0%	69.2%
Ground glass opacity score	0.744	0.090	**0.018**	0.567	0.920	4.25	81.0%	69.2%
Consolidation score	0.626	0.107	0.221	0.417	0.836	0.75	81.0%	53.8%
Crazy-paving pattern score	0.608	0.098	0.296	0.416	0.800	3.25	47.6%	84.6%
Combined model								
Overall lung involvement score + CURB65 score	0.808	0.077	**0.003**	0.658	0.958	–	90.5%	61.5%
Overall lung involvement score + qSOFA score	0.810	0.079	**0.003**	0.654	0.965	–	95.2%	61.5%

*All bold values were statistically significant.
Abbreviations: AUC: areas under the curve; ROC: receiver operating characteristic.

separating point later on the third week. Overall lung involvement score on the second week appeared to have predictive value for whole-course clinical severity with optimal cut-off of 5.25 points.

4.1. Patient characteristic

Our patients were all confirmed cases with pneumonia who were admitted in early February in a university affiliated tertiary hospital in Wuhan. Clustered onset was frequently seen. There were 22(41.5%) severe and 13(24.5%) critical cases added up to two thirds of our patient population. Half of them was male, similar to previous study [9]. The incident of hypertension and diabetes in our patient cohort was as high as 45% and 15%, respectively, which might indicate vulnerability of this group of patients to COVID-19. However, this may also relate to a relatively senior age. Association of cardiovascular comorbidities with clinical severity and prognosis in hospitalized patients remains to be further investigated and would have profound impact on patient management.

4.2. Application and feasibility of CT score system for physicians

High resolution CT severity scoring system was widely used in interstitial lung disease and pneumonia for medical decision-making and prognosis [3,10–12]. There were several innovative applications of the system in our study. First, this was the first pilot study of CT scoring by clinical doctors rather than radiologists. The two reviewers were both attending physician of emergency medicine by training with over ten years' clinical experience in a university affiliated tertiary hospital. As we saw in test set consisted of 50 patients, the inter-rater reliability between the two ED doctors was ranked as moderate or good while inter-rater reliability between radiologists and clinicians were good or excellent as measured by ICCs. Overall lung involvement score was proved to have the highest ICCs. Also, the values and evolving trend of CT score was similar to those reported by radiologists or deep-learning approach in previous studies [4,13–15]. Additionally, the median time for visual assessment was only about 2 min without complex protocol. Second, the six zones of the lung were much easier to recognize than lobes or segments, especially when high resolution CT was not available. Third, three major types of lesions, GGO, consolidation and crazy-paving pattern, were evaluated separately to show evolving patterns and difference presentation between patients of different severity with predictive value. Rare manifestations were excluded from analysis to avoid inaccuracy and time consumption. Thus, CT score demonstrated remarkable feasibility and efficiency used by experienced clinicians directly as a quick diagnostic tool.

4.3. CT score changes and predict value upon diagnostic capacity

CT scan at an early stage showed preferable diagnostic value with a sensitivity as high as 80–90% compared with rRT-PCR at around 70% [16,17]. The peak of lung opacification occurred about 10–13 days after symptom onset with bimodal phases from GGO predominant to crazy-paving pattern and consolidation predominant before final remission [4,13–15,18,19]. But those conclusions were made based on a majority of mild/moderate patients. This might introduce patient selection bias. Our study was the first one to differentiate clinical severity of patients by CT score and to analyze their dynamic changing over time at once, based on the considerable number of severe cases with subgroups analysis, which has rarely been explored. This idea originated from tough issues the authors encountered during patient care. There were 21(39.6%) patients in our study underwent escalation of care during hospital stay, higher than that reported at 20% [20]. COVID-19 patients compensated well with no or low oxygen demand in the early stage, sometimes suffered from sudden, unexpected deterioration, or even ended up with intubation or in-hospital death ultimately, due to mild physical activity or mood swing. The discrepancy between normal saturation and considerably affected lung on CT initially drew our attention to the predictive value of CT scan. It is noteworthy that the disadvantages of CT scan might be its financial burden and radiation exposure for the patients, CT screening for the detection of COVID-19 is not recommended by radiologist either [21]. However, the prognostic value of CT on the second week was extremely inspiring in that an overall lung involvement score exceeding 5 points at this time could be taken as alarming signal before oxygenation reserve crashed or clinical decompensation occurred. Physicians should provide sufficient oxygen support for the patient in advantage to prevent sudden deterioration. Furthermore, it might even be reasonable to include CT score as one of the criteria for clinical severity classification together with clinical indicators. Therefore, the authors suggested that importance of CT scan still outweighed its adverse impact for the patients during early stage of disease. The interval of follow-up CT should be considered based on patient status and contextual factors though.

4.4. Limitation

There were several limitations in our study. First of all, the study was a pilot study with only 53 patients enrolled. The relatively small sample size was inadequate to disclose further potential mechanism or to include other predictive factors for prognosis. Further study is needed with larger sample size and external validation. Second, the application of artificial intelligence (AI) assisted diagnostic technology is developing rapidly on reproductivity, sensitivity, and accuracy of quantitative evaluation [22,23]. However, numerous procedural requirements are mandatory, including end-inspiration scan, scanner calibration, unified section thickness and reconstruction protocol, manual segmentation adjustment, lung volume correction and so on. The systemic error is also concerning. Furthermore, software analysis mainly focused on small airway diseases such as COPD and is less applied to the evaluation of GGO or other respiratory diseases. In addition, the authors, unfortunately, couldn't manage to get access to software analysis due to limited resources during the initial stage of epidemic. Software analysis was not applicable in such urgent circumstances or in remote areas when visual assessment was proved to be a simple, rapid, and relatively reliable method.

5. Conclusion

Our modified semi-quantitative CT scoring system for COVID-19 patients demonstrated efficiency and feasibility for clinical use. Severe/critical patients had higher scores for GGO, consolidation, crazy-paving pattern, and overall lung involvement than moderate cases during 2–4 weeks of clinical course. Overall lung involvement score on the second week appeared to have predictive value for whole-course clinical severity.

CRediT authorship contribution statement

Shu Li: Writing - review & editing, Writing - original draft, Conceptualization, Project administration, Formal analysis, Methodology. **Shaoyu Liu:** Project administration. **Ben Wang:** Data curation. **Qiuyu Li:** Data curation. **Hua Zhang:** Formal analysis, Methodology. **Lin Zeng:** Formal analysis. **Hongxia Ge:** Writing - review & editing. **Qingbian Ma:** Resources. **Ning Shen:** Conceptualization, Supervision, Resources, Funding acquisition, Methodology, Writing - original draft.

Declaration of competing interest

The authors have disclosed that there is no financial, consultant, institutional, and other relationships that might lead to bias or a conflict of interest.

Acknowledgements

We wish to thank our radiologists Yuqing Zhao and Ge Guo for analyzing the images and constructive criticisms. We greatly appreciate the friendship with colleagues from Branch of Sino-French, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. Thanks for their great support during patient management. Also, we are grateful to offer our deepest and warmest thanks for all the colleagues from B11 west isolation unit for their determination to risk their own lives to fight together during the epidemic and their passionate work with high quality and efficiency as a team. The study was funded by National Key Research and Development Program of China (2018YFC1311900) and Fundamental Research Funds for the Central Universities (BMU2020HKYZX011).

References

[1] Johns Hopkins University and Medicine, Coronavirus resource center[EB/OL], Available from: https://coronavirus.jhu.edu/map.html, 2020 April 30.
[2] National Health Commission & National Administration of Traditional Chinese Medicine, Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 7) [EB/OL], Available from, https://www.chinadaily.com.cn/pdf/2020/1.Clinical.Protocols.for.the.Diagnosis.and.Treatment.of.COVID-19.V7.pdf, 2020 Mar 29.
[3] K. Fujimoto, H. Taniguchi, T. Johkoh, et al., Acute exacerbation of idiopathic pulmonary fibrosis: high-resolution CT scores predict mortality, Eur. Radiol. 22 (1) (2012) 83–92, https://doi.org/10.1007/s00330-011-2211-6.
[4] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z.A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li, S. Li, H. Shan, A. Jacobi, M. Chung, Chest CT findings of COVID-19: Relationship with duration, Radiology (2020 Feb 20), https://doi.org/10.1148/radiol.2020200463 ([Epub ahead of print]).
[5] D.M.B.A. Hansell, H. MacMahon, T.C. McLoud, N.L. Muller, Remy Jacques, Fleischner Society glossary of terms for thoracic imaging, Radiology 246 (3) (2008) 697–722, https://doi.org/10.1055/s-0035-1553216.
[6] T. Franquet, Imaging of pulmonary viral pneumonia, Radiology 260 (1) (2011) 18–39, https://doi.org/10.1148/radiol.11092149.
[7] H.X. Guan, Y. Xiong, N.Q. Shen, Y.Q. Fan, J.B. Shao, H.J. Li, X.M. Li, D.Y. Hu, W. Z. Zhu, Z.Y. Jin, Clinical and thin-section CT features of patients with 2019-nCoV-pneumonia, Radiologic Practice (2020 Feb 3), https://doi.org/10.13609/j.cnki.1000-0313.2020.02.001 [Epub ahead of print].
[8] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, Journal of Chiropractic Medicine 15 (2) (2016) 155–163.
[9] N. Chen, M. Zhou, X. Dong, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 395 (10223) (2020) 507–513, https://doi.org/10.1016/s0140-6736(20)30211-7.
[10] C.W. Chou, H.S. Chao, F.C. Lin, H.C. Tsai, W.H. Yuan, S.C. Chang, Clinical usefulness of hrct in assessing the severity of pneumocystis jirovecii pneumonia: a cross-sectional study, Medicine (Baltim.) 94 (16) (2015) e768, https://doi.org/10.1097/MD.0000000000000768.
[11] S. Wangkaew, J. Euathrongchit, P. Wattanawittawas, N. Kasitanon, Correlation of delta high-resolution computed tomography (HRCT) score with delta clinical variables in early systemic sclerosis (SSc) patients, Quant. Imag. Med. Surg. 6 (4) (2016) 381–390, https://doi.org/10.21037/qims.2016.08.08.
[12] M.A. Man, E. Dantes, B. Domokos Hancu, et al., Correlation between transthoracic lung ultrasound score and HRCT features in patients with interstitial lung diseases, J. Clin. Med. 8 (8) (2019), https://doi.org/10.3390/jcm8081199.
[13] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R.L. Hesketh, L. Yang, C. Zheng, Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia, Radiology (2020 Feb 13), https://doi.org/10.1148/radiol.2020200370 [Epub ahead of print].
[14] Y. Wang, C. Dong, Y. Hu, C. Li, Q. Ren, X. Zhang, H. Shi, M. Zhou, Temporal changes of CT findings in 90 patients with COVID-19 pneumonia: a longitudinal study, Radiology (2020 Mar 19), https://doi.org/10.1148/radiol.2020200843 ([Epub ahead of print]).
[15] H.R. Huang Lu, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest CT assessment of COVID-19: deep-learning approach, Radiology: Cardiothoracic Imaging (2020 Mar 30), https://doi.org/10.1148/ryct.2020200075 ([Epub ahead of print]).
[16] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest CT for COVID-19: comparison to RT-PCR, Radiology (2020 Feb 19), https://doi.org/10.1148/radiol.2020200432 ([Epub ahead of print]).
[17] P. Huang, T. Liu, L. Huang, H. Liu, M. Lei, W. Xu, X. Hu, J. Chen, B. Liu, Use of chest CT in combination with negative RT-PCR assay for the 2019 novel coronavirus but high clinical suspicion, Radiology (2020 Feb 12), https://doi.org/10.1148/radiol.2020200330 [Epub ahead of print].
[18] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, C. Zheng, Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study, Lancet Infect. Dis. (2020 Feb 24), https://doi.org/10.1016/S1473-3099(20)30086-4 ([Epub ahead of print]).
[19] S. Salehi, A. Abedi, S. Balakrishnan, A. Gholamrezanezhad, Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients, AJR Am. J. Roentgenol. (2020) 1–7, https://doi.org/10.2214/AJR.20.23034.
[20] Z.B. Lyu, C.S. Guan, S. Yan, T. Cui, A. Zhou, R.M. Xie, B.D. Chen, Value of CT findings in predicting transformation of clinical types of COVID-19, Chin. J. Radiol. (2020 Mar 13), https://doi.org/10.3760/cma.j.issn.1005-1201.2020.0020 ([Epub ahead of print]).
[21] S.K.F. Simpson, S. Abbara, S. Bhalla, J.H. Chung, M. Chung, T.S. Henry, J.P. Kanne, S. Kligerman, J.P. Ko, H. Litt, Radiological Society of North America Expert Consensus Statement on Reporting:Chest CT Findings Related to COVID-19, Endorsed by the Society of Thoracic and Radiology, the American College of Radiology, and RSNA. Radiology: Cardiothoracic Imaging, 2020 Mar 25, https://doi.org/10.1148/ryct.2020200152 ([Epub ahead of print]).
[22] H. Shi, X. Han, Y. Fan, B. Liang, F. Yang, P. Han, C. Zheng, Radiologic features of patients with 2019-nCoV infection, Journal of Clinical Radiology (2020 Feb 6), https://doi.org/10.13437/j.cnki.jcr.20200206.002 [Epub ahead of print].
[23] M.J. Willemink, W.A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. R. Folio, R.M. Summers, D.L. Rubin, M.P. Lungren, Preparing medical imaging data for machine learning, Radiology: Cardiothoracic Imaging (2020 Feb 18), https://doi.org/10.1148/radiol.2020192224 ([Epub ahead of print]).