CHEST

Observer agreement and clinical significance of chest CT reporting in patients suspected of COVID-19

Marie-Pierre Debray [1,2] · Helena Tarabay [1] · Lisa Males [1,3] · Nisrine Chalhoub [1] · Elyas Mahdjoub [1,3] ·
Thomas Pavlovsky [4] · Benoît Visseaux [5,6] · Donia Bouzid [4,6] · Raphael Borie [2,7] · Catherine Wackenheim [1] ·
Bruno Crestani [2,3,7] · Christophe Rioux [8] · Loukbi Saker [1] · Christophe Choquet [4] · Jimmy Mullaert [6,9] · Antoine Khalil [1,2,3]

Abstract

Objectives To assess interobserver agreement and clinical significance of chest CT reporting in patients suspected of COVID-19.
Methods From 16 to 24 March 2020, 241 consecutive patients addressed to hospital for COVID-19 suspicion had both chest CT and SARS-CoV-2 RT-PCR. Eight observers (2 thoracic and 2 general senior radiologists, 2 junior radiologists, and 2 emergency physicians) retrospectively categorized each CT into one out of 4 categories (evocative, compatible for COVID-19 pneumonia, not evocative, and normal). Observer agreement for categorization between all readers and pairs of readers with similar experience was evaluated with the Kappa coefficient. The results of a consensus categorization were correlated to RT-PCR.
Results Observer agreement across the 4 categories was good between all readers (κ value 0.61 95% CI 0.60–0.63) and moderate to good between pairs of readers (0.54–0.75). It was very good (κ 0.81 95% CI 0.79–0.83), fair (κ 0.32 95% CI 0.29–0.34), moderate (κ 0.56 95% CI 0.54–0.58), and moderate (0.58 95% CI 0.56–0.61) for the categories evocative, compatible, not evocative, and normal, respectively. RT-PCR was positive in 97%, 50%, 31%, and 11% of cases in the respective categories. Observer agreement was lower ($p < 0.001$) and RT-PCR positive cases less frequently categorized evocative in the presence of an underlying pulmonary disease ($p < 0.001$).
Conclusion Interobserver agreement for chest CT reporting using categorization of findings is good in patients suspected of COVID-19. Among patients considered for hospitalization in an epidemic context, CT categorized evocative is highly predictive of COVID-19, whereas the predictive value of CT decreases between the categories compatible and not evocative.

Key Points

• *In patients suspected of COVID-19, interobserver agreement for chest CT reporting into categories is good, and very good to categorize CT "evocative."*
• *Chest CT can participate in estimating the likelihood of COVID-19 in patients presenting to hospital during the outbreak, CT categorized "evocative" being highly predictive of the disease whereas almost a third of patients with CT "not evocative" had a positive RT-PCR in our study.*
• *Observer agreement is lower and CTs of positive RT-PCR cases less frequently "evocative" in presence of an underlying pulmonary disease.*

✉ Marie-Pierre Debray
marie-pierre.debray@aphp.fr

1 Department of Radiology, Hôpital Bichat Claude-Bernard, APHP, 46 rue Henri Huchard, 75018 Paris, France

2 Inserm UMR1152, Physiopathology and Epidemiology of Respiratory Diseases, Paris, France

3 Faculty of Medicine, Paris Diderot University, Bichat Campus, Paris, France

4 Department of Emergency, Hôpital Bichat Claude-Bernard, APHP, 46 rue Henri Huchard, F-75018 Paris, France

5 Laboratoire de Virologie, Hôpital Bichat Claude Bernard, APHP, 46 rue Henri Huchard, F-75018 Paris, France

6 Université de Paris, IAME, INSERM, Paris, France

7 Department of Pneumology A, Centre de Référence Constitutif des Maladies Pulmonaires Rares, Hôpital Bichat Claude-Bernard, Assistance Publique-Hôpitaux de Paris, 46 rue Henri Huchard, F-75018 Paris, France

8 Department of Infectious Diseases, Hôpital Bichat Claude-Bernard, APHP, 46 rue Henri Huchard, F-75018 Paris, France

9 Department of Epidemiology, Biostatistics and Clinical Research, AP-HP, Hôpital Bichat, F-75018 Paris, France

Abbreviations

COVID-19	Coronavirus disease 2019
CT	Computed tomography
GGO	Ground-glass opacities
RT-PCR	Real-time reverse transcription-polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2

Introduction

Since December 2019, a new respiratory disease related to a new coronavirus, SARS-CoV-2, developed in China and rapidly spread to other countries, reaching a pandemic stage in March 2020 [1, 2]. Even if the disease follows a benign course in many cases, some patients develop respiratory difficulties requiring hospitalization, leading to a large amount of patients with clinical suspicion of coronavirus disease 2019 (COVID-19) presenting to the emergency departments [3]. Accurate identification of COVID-19 patients is crucial to isolate them from not infected patients and to limit the diffusion of the outbreak. The reference standard is the positivity of the real-time reverse transcription-polymerase chain reaction (RT-PCR); nevertheless, the sensitivity of this test remains unclear, having been reported between 42 and 71% in some early series [4–6], because of suboptimal sampling technique, limitations in performance assay, or low viral load in the nasopharyngeal area. Chest CT shows abnormalities in a large majority of cases, with some signs described as typical or very evocative of the disease in the current outbreak context [7–11]. Sensitivity of chest CT has been reported as high as 97% as compared to RT-PCR [4] and CT abnormalities could precede RT-PCR positivity [12]. Because it is readily available, chest CT may assist first-line triage of patients presenting to hospital [3]. Several radiology societies [6, 13–15] have proposed structured reporting of CT into categories, defined according to the typical or less typical appearance of lung involvement, to facilitate communication with physicians. In routine practice, categorization is based on each reader individual impression supported by numerous papers having described imaging signs of COVID-19 pneumonia [7–11]. Such categorization may directly impact the clinical decision-making. However, the reproducibility of the categorization is unknown and the clinical significance of the different categories is unclear. Thus, the objectives of our study were to assess interobserver agreement to categorize CT findings as well as performances of chest CT across the different categories in patients suspected of COVID-19 presenting to hospital.

Material and methods

Setting

This is a monocentric retrospective study conducted in a University Hospital (Bichat Claude-Bernard Hospital, Paris, France) between March 16, 2020, and March 24, 2020. Institutional review board was approved and written informed consent waived. During this period of COVID-19 outbreak, patients presenting at our hospital for COVID-19 suspicion and for whom hospitalization was considered had both chest CT scan and SARS-CoV-2 RT-PCR. Diagnosis of COVID-19 relied on the positivity of the RT-PCR and CT could assist early triage in critically ill patients or with clinically overt pneumonia. Patients with a negative RT-PCR result could have a subsequent RT-PCR test and/or another chest CT during the next few days, depending on the physician's judgment.

Patients

Consecutive adult patients attending the emergency room or the infectious diseases department of our hospital with clinical suspicion of COVID-19 and having both chest CT and SARS-CoV-2 RT-PCR were included.

Clinical and laboratory data collection

Demographic, clinical and laboratory data at presentation, and follow-up data when available were extracted from electronic medical records. Clinical data included symptoms, any need for oxygen supply, time from symptom onset to CT, comorbidities, and pre-existing pulmonary diseases. RT-PCR was performed on nasopharyngeal swabs or aspiration, using RealStar® SARS-CoV-2 RT-PCR Kit (Altona Diagnostics) or Cobas® SARS-CoV-2 Test (Roche).

Chest CT protocols

Chest CT scans were acquired on a multidetector-row CT (Aquilion One Genesis or Prime, Canon Medical Systems Corporation) without contrast medium injection. They were performed in the supine position at full inspiration. The scanning parameters were as follows: 120 kVp, automatic exposure control for tube current (SD:15), exposure time 0.27–0.35 s per rotation depending on the CT unit, collimation 40 mm. Images were reconstructed with 1-mm slice thickness and 0.8-mm inter-slice gap, using a high-frequency reconstruction algorithm.

Image analysis

All CT scans were analyzed by 8 readers, including 2 senior emergency physicians (T.P., D.B., with 10 years of experience each), 2 radiology residents (H.T., N.C., 4 and 5 years of experience), 2 senior general radiologists (L.M., E.M., 6 and 9 years of experience), and 2 senior thoracic radiologists (M.P.D., A.K., 23 and 25 years of experience), blindly to RT-PCR results and final diagnosis. All readers classified each examination into one out of 4 categories, as follows: evocative, compatible, not evocative of COVID-19, and normal following the recommendations of the French Society of Radiology [14]. The global impression of each reader was supported by previous typical and less typical reported signs in the literature. A guide was provided, recalling these signs. The "evocative" category included multifocal ground-glass opacities (GGO), being nodular or not, or crazy-paving with or without consolidations, with a bilateral, peripheral, or mixed distribution and involvement of the posterior zones. The intermediate category "compatible" corresponded to cases showing abnormalities already reported in COVID-19 but that may be encountered in other diseases or very limited in extent. It included GGO and/or consolidations with very few lesions and unilateral distribution, exclusive central distribution or absence of posterior lung area involvement, halo sign as main abnormality, and association of typical opacities, with atypical signs or other lesions. The category "not evocative" corresponded to cases showing abnormalities very rarely reported in COVID-19 or typical of another diagnosis as isolated systematized consolidation, discrete centrilobular nodules with tree-in-bud appearance or lung cavitation in favor of other lung infection, centrally distributed GGO with septal lines and pleural effusion in favor of cardiogenic pulmonary edema, peripheral reticulations with or without honeycombing, traction bronchiectasis, and GGO in favor of interstitial lung disease. This category also included non-specific abnormalities as sub-segmental atelectasis or opacities considered to be sequelae. Because the distinction of such minor non-specific or typically sequellar abnormalities from normal parenchyma may have little clinical relevance in the present study, findings were analyzed according both to the 4 categories and to the 3 categories that resulted from merging of "not evocative" and "normal." Any disagreement between the 4 senior radiologists was analyzed in consensus of these 4 readers giving a final consensus categorization for all cases.

Finally, all chest CTs were described by one thoracic radiologist (M.P.D.) for presence and distribution of various elementary signs, as well as signs of any underlying pulmonary disease (significant pulmonary emphysema, interstitial lung disease, bronchiectasis, parenchymal sequelae, bronchial carcinoma).

Statistical analysis

Categorical variables were described by numbers and percentage for each category. The agreement between two and more than two readers was evaluated with the Cohen's kappa coefficient and the Fleiss' kappa, respectively, and their 95% confidence interval, which measures the excess proportion of agreement after taking chance into account. Interobserver agreement was considered poor for a kappa value < 0.20, fair for 0.21–0.40, moderate for 0.41–0.60, good for 0.61–0.80, and very good for 0.81–1.00. Comparisons between dependent kappas (e.g., for different couple of readers for the same images) were performed with bootstrapping ($N = 10000$ samples) and the p value corresponds to the proportion of bootstrap samples that yield a couple of kappa value in a different order than the observed one. Comparisons between independent kappa (e.g., for different levels of a categorical variable) were performed according to the method proposed in [16].

Comparison of the frequency of radiologic signs between categories was performed with the fisher exact test. All analysis were done using R v4.0.2.

Results

General description of the population

In total, 241 patients were included. Their demographic, signs at presentation, and comorbidities are in Table 1. COVID-19 was confirmed in 158 patients by RT-PCR positivity. COVID-19 was deemed likely despite two negative RT-PCR in 2 cases with clinical and CT follow-up strongly supportive of this diagnosis. Fifteen patients were considered non-COVID-19 because of at least 2 consecutive negative RT-PCR and absence of clinical and radiological signs favoring COVID-19 during follow-up. Sixty-six patients were considered non-COVID-19 with only one negative RT-PCR but including 38 with CT and/or clinical follow-up (Fig. 1).

Interobserver agreement for chest CT categorization

Kappa coefficient between all readers across the 4 CT categories was good (0.61, 95% CI 0.60–0.63). It was moderate to good for each pair of readers, significantly better between any pair of radiologists as compared to emergency physicians ($p < 0.001$) (Table 2). The kappa value between all readers was lower when abnormalities were unilateral as compared to bilateral lesions ($p = 0.002$) and in the presence of underlying pulmonary lesions ($p < 0.001$). It was lower for patients older as compared to those younger than 70 years ($p \leq 0.001$), and when time from symptom onset to CT was shorter than
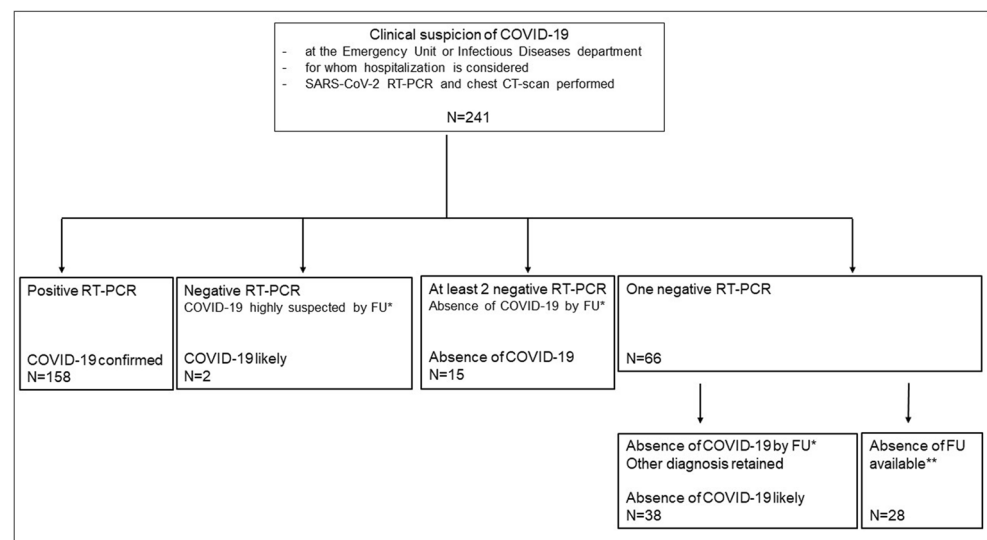
Table 1 Clinical characteristics of 241 included patients with clinical suspicion of COVID-19

Characteristics	No. (%) or median [Q1–Q3]
Age (years)	64 [52–79]
Male (%)	155 (64)
Comorbidities	
COPD	24 (10)
Respiratory insufficiency	37 (15)
Any other	170 (71)
Signs at presentation	
Respiratory (any)	202 (84)
Cough	152 (63)
Expectoration	27 (11)
Dyspnea	149 (62)
Chest pain	34 (14)
Need for oxygen supply	143 (59)
Digestive (abdominal pain, diarrhea, nausea)	65 (27)
Neurologic (confusion, headache)	53 (22)
Asthenia	92 (38)
Fever, chills, sweats	115 (48)
Onset-to-CT-delay (days)	4 [2–7]

Data are numbers with percentages in brackets or medians with lower and upper quartiles in square brackets

5 days ($p \leq 0.001$). Observer agreement was very good between all readers for the category "evocative" (0.81, 95% CI 0.79–0.83), fair for the category "compatible" (0.32, 95% CI 0.29–0.34), moderate for each category "not evocative" and "normal" (0.56, 95% CI 0.54–0.58; and 0.58, 95% CI 0.56–0.61, respectively), and good (0.74, 95% CI 0.71–0.76) for chest CTs classified either not evocative or normal (Table 3).

Correlation of RT-PCR and categories on chest CT

The RT-PCR positivity rate was highly significantly different among the 4 categories ($p < 0.0001$): 119 out of 123 (97%), 15 out of 30 (50%), 22 out of 70 (31%), and 2 out of 18 (11%) cases considered evocative, compatible, not evocative, and normal by the consensus reading had a positive RT-PCR, respectively (Table 3). The rate of RT-PCR positivity was 27% (24 out of 88) among CTs categorized either not evocative or normal.

With RT-PCR as reference, chest CT classified evocative had 75% sensitivity (95% CI 68–81%) and 95% specificity (95% CI 87–98%) whereas chest CT classified evocative or compatible had 85% sensitivity (95% CI 78–90%) and 77% specificity (95% CI 66–85%) for COVID-19. Sensitivity for evocative CT was significantly lower for patients with a delay since symptom onset lower than 5 days (68% 95% CI [60–78%] vs 83% 95% CI [73–91%], $p = 0.038$) and for patients with an underlying pulmonary disease (36% 95% CI [21–54%] vs 87% 95% CI [80–92%], $p < 0.001$), but the difference was not significant between younger and older patients (age > 70, $p = 0.23$).

Among 90 patients with a first negative RT-PCR, 22 were re-tested, including 5 out of 6 patients (83%) with CT considered evocative, 5 out of 17 patients (29%) with CT considered compatible, and 12 out of 67 patients (18%) with CT considered not evocative or normal. Of these subsequent RT-PCR, 2 out of 5 were positive for each category "evocative" and "compatible" and 3 out of 12 for the third category.

Clinical and chest CT findings in the different categories

CT features of the whole population, of RT-PCR positive cases, and of the different CT categories are in Table 4. The

Fig. 1 Flow chart of the study. SARS-CoV-2 RT PCR and chest CT were performed at presentation, except in 7 cases for which RT-PCR had been performed 1 to 3 days earlier. Among 158 COVID-19 confirmed cases, RT-PCR was positive at admission in 151 cases and in subsequent days during hospitalization in 7 cases. *Clinical and chest CT follow-up. **Absence of follow-up available because the patient had been transferred to another hospital or turned back home

Table 2 Chest CT categorization of the 8 readers and consensus reading, with observer agreement between pairs and all readers

	Evocative N (%)	Compatible N (%)	Not evocative N (%)	Normal N (%)	κ value[a] (95% CI) 4 categories[b]	κ value[a] (95% CI) 3 categories[c]
Thoracic senior radiologists						
Reader 1	117 (49)	32 (13)	74 (31)	18 (7.5)	0.68*	0.69
Reader 2	126 (52)	52 (22)	45 (19)	18 (7.5)	(0.60–0.75)	(0.62–0.76)
General senior radiologists						
Reader 3	119 (49)	18 (7.5)	79 (33)	25 (10)	0.68* (0.61–0.75)	0.75
Reader 4	112 (51)	38 (11)	62 (26)	29 (12)		(0.68–0.82)
Resident radiologists						
Reader 5	122 (51)	26 (11)	50 (21)	43 (18)	0.75* (0.68–0.82)	0.85**
Reader 6	114 (47)	29 (12)	68 (28)	30 (12)		(0.79–0.90)
Senior emergency physicians						
Reader 7	119 (49)	37 (15)	50 (21)	35 (15)	0.52 (0.43–0.60)	0.64
Reader 8	131 (54)	35 (15)	47 (20)	28 (12)		(0.56–0.72)
All readers					0.61 (0.60–0.63)	0.68 (0.67–0.70)
Consensus reading	123 (51)	30 (12)	70 (29)	18 (7.5)		

Data are numbers with percentages, or 95% confidence interval for the κ value, in brackets

[a] Cohen's kappa value for agreement between 2 readers, and Fleiss' kappa value for agreement between all readers

[b] Kappa value calculated for the 4 following categories: evocative, compatible, not evocative, and normal

*Observer agreement significantly better between any pair of radiologists as compared to agreement between emergency physicians ($p < 0.001$) and not significantly different between the 3 pairs of radiologists

[c] Kappa value calculated for the 3 following categories: evocative, compatible, and not evocative or normal

**Observer agreement significantly better between resident radiologists as compared to agreement between thoracic senior radiologists ($p < 0.001$)

most frequent pattern of CT considered evocative was mixed with predominant GGO. Typical bilateral and peripheral distribution with posterior involvement was almost constant. Some centrally distributed lesions were associated to peripheral lesions in 72% of cases (Fig. 2). The 30 chest CTs considered compatible more frequently showed pure GGO as compared to evocative cases ($p = 0.0012$). Among these 30 cases, 12 and 6 showed features of an underlying pulmonary disease and of an associated pulmonary edema, respectively (Figs. 3, 4). As compared with cases classified "not evocative" (Figs. 5, 6), cases classified compatible more frequently showed a typical distribution and atypical signs were absent among those with positive RT-PCR. Time from symptom onset to CT was longer, patients were older, and need for

oxygen supply was more frequent in patients whose CT was categorized evocative as compared to other patients.

Discussion

The current study describes observer agreement for chest CT reporting in a large series of consecutive patients suspected of COVID-19. We found that categorization of CT reports was reproducible and meaningful in patients considered for hospitalization. With SARS-CoV-2 RT-PCR as reference, chest CT reported "evocative of COVID-19 pneumonia" was highly predictive of the disease in this population during the outbreak and agreement for this category between observers of various

Table 3 Interobserver agreement for each category and correlation to SARS-CoV-2 RT-PCR results

	Evocative	Compatible	Not evocative	Normal	Not evocative or normal	p value
Kappa value, all readers (95% CI)	0.81 (0.79–0.83)	0.32 (0.29–0.34)	0.56 (0.54–0.58)	0.58 (0.56–0.61)	0.74 (0.71–0.76)	
Effective for comparison to RT-PCR (N) (consensus reading)	123	30	70	18	88	
Positivity of RT-PCR (%)	119 (97)	15 (50)	22 (31)	2 (11)	24 (27)	< 0.0001*

Data are numbers with percentages, or 95% confidence interval for the κ value, in brackets

*p value is < 0.0001 for 4 categories (evocative, compatible, not evocative, normal) and for 3 categories (evocative, compatible, not evocative or normal)

experiences and sub-specialties was very good. The positivity rate of RT-PCR was highly significantly different among the categories, suggesting that CT may help in disease likelihood stratification. It should be emphasized that almost a third of patients with chest CT classified "not evocative" had a positive RT-PCR, highlighting that no CT pattern can rule out COVID-19 in an epidemic context. As previously reported [4, 17], RT-PCR may also be positive in patients without lung abnormalities on CT.

We observed only fair observer's agreement for the category "compatible." This may be explained by the method used to classify patients. Thus, categorization of chest CT was mainly based on the global impression of each reader, according to the routine practice in our hospital, where more than thousand chest CTs have been performed to date in COVID-19 patients. The guide we provided to all observers before readings to help case classification partly relied on interpretation of findings which may vary according to each reader experience. Cases classified "compatible" encompassed only 12% of all, i.e., 30 cases at all. More than half of them showed features of an underlying pulmonary disease or of a mixed pattern with some opacities that could be attributable to pulmonary edema. Such mixed features complicate interpretation and categorization of findings and lower the reader's confidence. Indeed we observed significant lower observer agreement in cases showing an underlying pulmonary disease. These mixed and complex cases are part of the routine practice.

The recent Radiological Society of North America proposal for CT findings related to COVID-19 includes 4 categories: typical, indeterminate, atypical appearance, or CT negative for pneumonia [6], the "indeterminate" category appearing similar to the "compatible" category we used in the present study. Terms and categories proposed by other radiology societies vary, according to whether or not they individualize a normal category, and to the intermediate category being named either "compatible" or "indeterminate" [13–15]. We herein retained 4 categories, including a normal category, following the recommendations of the French Society of Radiology but also performed an analysis on 3 categories, resulting from merging of a normal appearance of the lung parenchyma with non-specific lung abnormalities or features suggesting an alternative diagnosis, in accordance with the guidelines of the European Society of Radiology. The surprisingly only moderate interobserver agreement we observed for the category "normal" may be explained by minor disagreements between minor non-specific abnormalities, as plate-like atelectasis or even dependent-induced opacities and strictly normal findings.

Table 4 CT features in all cases, SARS-CoV-2 RT-PCR positive cases and in the different CT categories

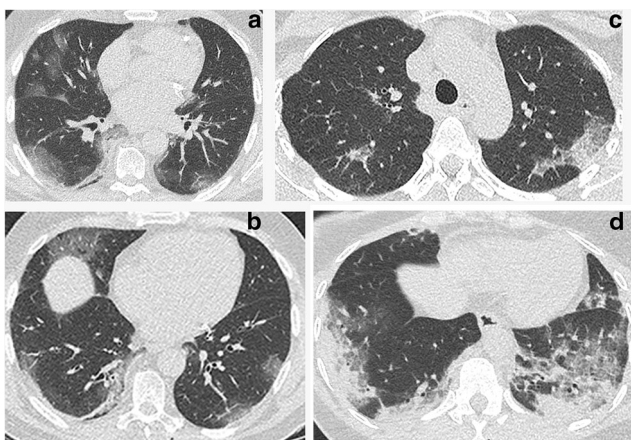|	All $N = 241$ N (%)	Positive RT-PCR $N = 158$ N (%)	Evocative $N = 123$ N (%)	Compatible $N = 30$ N (%)	Not evocative $N = 70$ N (%)	p value*
Pattern						
Isolated GGO	51 (21)	31 (20)	21 (17)	14 (47)	16 (23)	0.004
Isolated consolidations	24 (10)	8 (5)	1 (1)	4 (13)	19 (27)	< 0.001
Mixed, predominant GGO	91 (38)	82 (52)	81 (66)	7 (23)	3 (4)	< 0.001
Mixed, predominant consolidations	30 (12)	24 (15)	20 (16)	4 (13)	6 (9)	0.33
Distribution						
Bilateral	177 (73)	140 (89)	122 (99)	24 (80)	31 (44)	< 0.001
Posterior	181 (75)	139 (88)	122 (99)	25 (83)	34 (49)	< 0.001
Peripheral	186 (78)	143 (91)	122 (99)	24 (80)	40 (59)	< 0.001
Central	145 (61)	114 (72)	107 (87)	15 (50)	23 (34)	< 0.001
Atypical features						
Pleural effusion	35 (15)	13 (8)	10 (8)	8 (27)	17 (24)	0.002
Systematized consolidation	7 (3)	4 (3)	1 (1)	1 (3)	5 (7)	0.035
Centrilobular nodules	14 (6)	4 (3)	1 (1)	2 (7)	10 (14)	< 0.001
Mucoid impaction	18 (7)	6 (4)	2 (2)	4 (13)	12 (17)	< 0.001
Underlying pulmonary disease	61 (25)	36 (23)	13 (11)	14 (47)	33 (47)	< 0.001

Cases showing normal lung parenchyma are not included

*p value is for comparison between CT categories
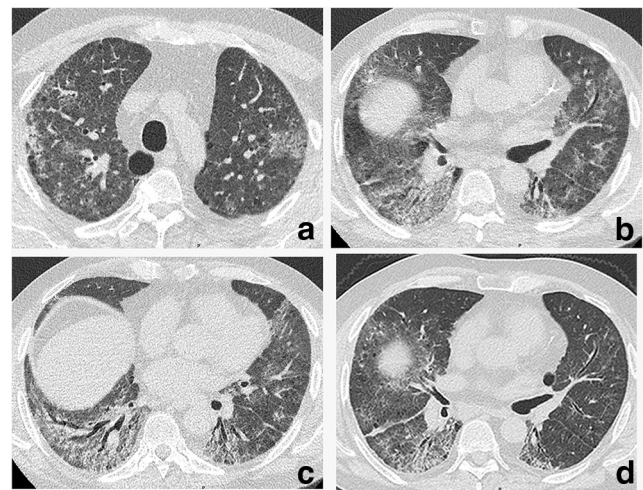
GGO: ground-glass opacities

Underlying pulmonary disease includes significant emphysema, interstitial lung disease, bronchiectasis, sequelae

By showing significant differences in the RT-PCR positivity rate among CT categories, our study supports that chest CT can participate in estimating the likelihood of COVID-19, in association with contact history, clinical presentation, and prevalence of the disease in the population [18]. The role of chest CT for patients suspected of COVID-19 is not completely established. Despite limitations in sensitivity and result delays, the RT-PCR remains the diagnostic reference and chest CT is not recommended for screening by most radiology societies [6, 15, 19, 20]. According to a recent consensus statement from the Fleischner Society [20], imaging may be indicated for diagnosis when RT-PCR is negative or unavailable in patients having risk factors for worsening or moderate-to-severe respiratory signs. In our study, most patients who had chest CT at the emergency room had indeed either moderate or severe clinical features or comorbidities. Chest CT helped addressing or transferring some patients showing a chest CT evocative of COVID-19 into the proper COVID-19 hospitalization area, especially those needing urgent decision, before the RT-PCR result was provided. However, none of the other categories could rule out the disease, even the "normal" category whose RT-PCR positivity rate was 11%. Chest CT could favor re-testing in cases with negative RT-PCR [12]. Patients with a first negative RT-PCR and a chest CT considered "compatible" tended to be more frequently re-tested and to have a subsequent PCR more frequently positive, as compared to patients with a first negative RT-PCR and a chest CT considered "not evocative."

To date, performances of chest CT have been analyzed according to a binary consideration, i.e., CT positive or negative for COVID-19 pneumonia, using RT-PCR as reference. Most studies have reported high sensitivity, up to 97% [4, 5, 21], and specificity between 25 and 56%, with pooled
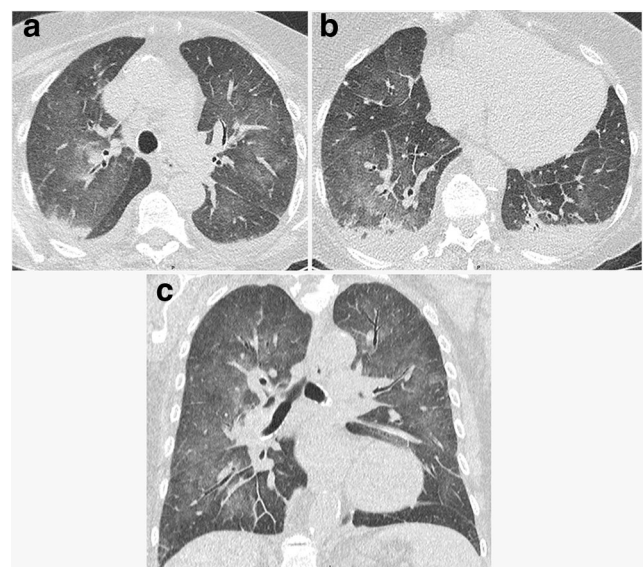

Fig. 3 Chest CT scan categorized "compatible with COVID-19 pneumonia," in association with fibrosing interstitial lung disease (ILD) showing a non-specific interstitial pneumonia pattern (panel **c**, axial plane through the lung bases). Pure ground-glass opacities, both centrally and peripherally distributed, have appeared in the left upper lobe (panels **a**, **b**), as compared to the previous CT performed 4 months earlier (panel **d**). Such new opacities could be attributable to COVID-19, another infection, or acutisation of ILD. Nasopharyngeal SARS-CoV-2 RT-PCR positive

sensitivity and specificity of 94% and 37%, respectively, according to a recent meta-analysis [22]. Whether positive CTs in these studies showed typical imaging features is unclear. Our results differ, chest CT classified evocative having 75%
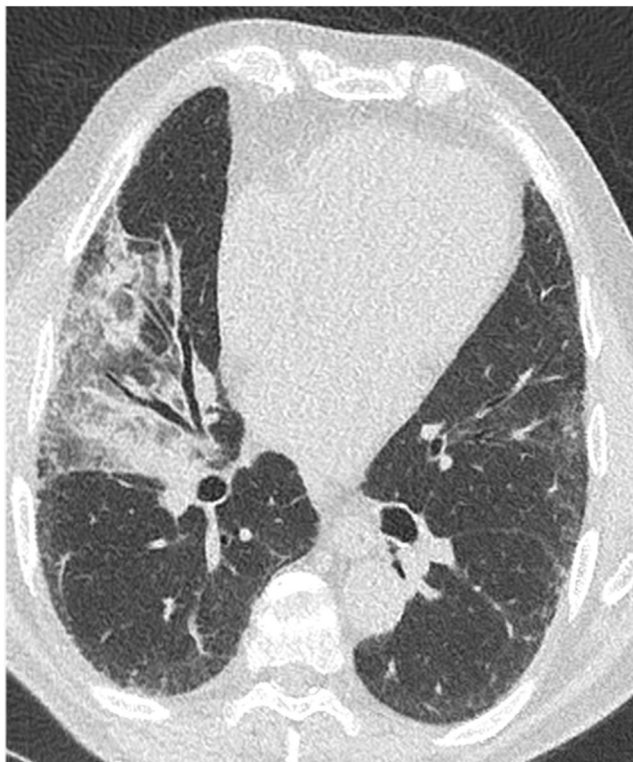

Fig. 2 Chest CT scan categorized "evocative of COVID-19 pneumonia" in two different patients with positive nasopharyngeal SARS-CoV-2 RT-PCR at admission and secondarily (panels **a** and **b** and panels **c** and **e**, respectively). Multifocal bilateral ground-glass opacities with subpleural and posterior predominance, associated with band-like (panel **b**) or more extensive consolidations (panel **d**)

Fig. 4 Chest CT scan categorized "compatible with with COVID-19 pneumonia," in association with pulmonary edema, manifesting as ground-glass opacities with a predominant central distribution, septal lines, and bilateral pleural effusion (panels **a**, **b**, axial plane; panel **c**, coronal plane) in a patient with history of chronic renal insufficiency on dialysis. Subpleural consolidation (panel **a**) in the posterior zone of the right upper lobe is consistent with associated COVID-19 pneumonia. Nasopharyngeal SARS-CoV-2 RT-PCR positive

Fig. 5 Chest CT scan categorized "not evocative of COVID-19 peumonia," showing small consolidations in the left lower lobe associated with bronchial thickening and endobronchial filling, in favor of bronchopneumonia. Nasopharyngeal SARS-CoV-2 RT-PCR negative

sensitivity and 95% specificity and chest CT classified evocative or compatible having 85% sensitivity and 77% specificity. These differences may be attributable to differences in CT features between an "evocative" CT and a "positive" CT as

Fig. 6 Chest CT scan categorized "not evocative of COVID-19 pneumonia" showing combination of areas of ground-glass opacity and consolidation systematized in the middle lobe (arrows), evoquing a lobar pneumonia. Despite positivity of the SARS-CoV-2-RT-PCR, bacterial co-infection was suspected and the patient received antibiotics with favorable evolution

well as differences in characteristics of the population having chest CT. The prevalence of the disease in the population, severity and type of clinical presentation, time from symptom to CT, age of patients, and any underlying pulmonary pathology may modify the performances of chest CT for diagnosing COVID-19 pneumonia [20, 23]. Indeed we observed that the presence of an underlying pulmonary disease lowered the sensitivity for an evocative CT. This concerned a quarter of the whole population in our study and almost a quarter of the patients with positive RT-PCR. CT reporting in several categories seems best suited to the routine practice than a binary conclusion, when CT abnormalities may mix different types of lesions or are very limited in extent. It allows identifying a category with typical features, whose high specificity can be useful in an epidemic context, allowing relying on chest CT for diagnosis in some cases.

Our study has some limitations. Firstly, because it is monocentric and because of various presentation and prevalence of the disease around the world, caution should be taken to extrapolate CT predictive values, which vary according to the disease prevalence, to other populations and periods [24, 25]. We may assume that the high positive predictive value we report for the category "evocative" would be lower if the prevalence of COVID-19 decreased and the one of other viral pneumonia or some interstitial lung diseases, as drugs or connective tissue diseases related, increased. On the contrary, in a very low prevalence context, we may expect the negative predictive value of chest CT classified "not evocative" or "normal" would be very high and CT would be useful for triage of negative cases. Secondly, the clinical significance of CT reporting should integrate the risk level of each patient that we have not precisely taken into account, even if the study period took place during the outbreak. Thirdly, the performances of chest CT have been evaluated in comparison to the results of the RT-PCR, whose sensitivity is imperfect. Two patients with repeated negative tests but typical clinical and CT features, driving a likely diagnosis of COVID-19 although remaining uncertain, were merged in the analysis with other unequivocal negative RT-PCR cases.

In conclusion, interobserver agreement to report chest CT findings into categories for clinical suspicion of COVID-19 is good, among readers of various experience levels and subspecialties. Chest CT can participate in estimating the likelihood of COVID-19 in patients presenting to hospital during the outbreak. CT categorized evocative of COVID-19 pneumonia were highly predictive of the disease, whereas the predictive value of CT decreased between the categories "compatible", "not evocative" and "normal," from 50 to 11%. Category reports need to be integrated to the clinical presentation and risk level for COVID-19.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Antoine Khalil.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors, Jimmy Mullaert, has significant statistical expertise.

Informed consent Written informed consent was not required for this observational non interventional study. It was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic
- performed at one institution

References

1. Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506
2. World Health Organization (2020) Director General's speeches. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19%2D%2D-11-march-2020. Accessed 27 Apr 2020
3. Sverzellati N, Milone F, Balbi M (2020) How imaging should properly be used in COVID-19 outbreak: an Italian experience. Diagn Interv Radiol 26:204–206
4. Ai T, Yang Z, Hou H et al (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. https://doi.org/10.1148/radiol.2020200642
5. Fang Y, Zhang H, Xie J et al (2020) Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 295:208–209
6. Simpson S, Kay FU, Abbara S et al (2020) Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. Radiology: Cardiothoracic Imaging. https://doi.org/10.1148/ryct.2020200152
7. Ye Z, Zhang Y, Wang Y, Huang Z, Song B (2020) Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. Eur Radiol 19:1–9
8. Zu ZY, Jiang MD, Xu PP et al (2020) Coronavirus disease 2019 (COVID-19): a perspective from China. Radiology. https://doi.org/10.1148/radiol.2020200490
9. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A (2020) Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. AJR Am J Roentgenol 14:1–7
10. Zhao W, Zhong Z, Xie X, Yu Q, Liu J (2020) Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. AJR Am J Roentgenol 3:1–6
11. Zhou Z, Guo D, Li C et al (2020) Coronavirus disease 2019: initial chest CT findings. Eur Radiol 24:1–9
12. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J (2020) Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. Radiology. https://doi.org/10.1148/radiol.2020200343
13. British Society of Thoracic Imaging (2020) Reporting templates COVID-19. https://www.bsti.org.uk/media/resources/files/BSTI_COVID_CT_Proforma_v2_13.04.2020.pdf. Accessed 27 Apr 2020
14. French Society of Radiology (2020) COVID-19 Compte rendu TDM thoracique. https://ebulletin.radiologie.fr/actualités-covid-19/compte-rendu-tdm-thoracique-iv. Accessed 27 Apr 2020
15. Revel MP, Parkar AP, Prosch H et al (2020) COVID-19 patients and the radiology department - advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI). Eur Radiol 20:1–7
16. Fleiss JL (1981) Statistical methods for rates and proportions, 2nd edn. John Wiley, New York
17. Bernheim A, Mei X, Huang M et al (2020) Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. Radiology. https://doi.org/10.1148/radiol.2020200463
18. Zhou TT, Wei FX (2020) Primary stratification and identification of suspected corona virus disease 2019 (COVID-19) from clinical perspective by a simple scoring proposal. Mil Med Res 7:16
19. Nair A, Rodrigues JCL, Hare S et al (2020) A British Society of Thoracic Imaging statement: considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic. Clin Radiol 75:329–334
20. Rubin GD, Ryerson CJ, Haramati LB et al (2020) The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. Chest. https://doi.org/10.1016/j.chest.2020.04.003
21. Caruso D, Zerunian M, Polici M et al (2020) Chest CT features of COVID-19 in Rome, Italy. Radiology. https://doi.org/10.1148/radiol.2020201237
22. Kim H, Hong H, Yoon SH (2020) Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: a meta-analysis. Radiology. https://doi.org/10.1148/radiol.2020201343
23. Lei P, Fan B, Wang P (2020) Differential diagnosis for coronavirus disease (COVID-19): beyond radiologic features. AJR Am J Roentgenol. https://doi.org/10.2214/AJR.20.23119
24. Wáng YXJ (2020) A call for caution in extrapolating chest CT sensitivity for COVID-19 derived from hospital data to patients among general population. Quant Imaging Med Surg 10:798–799
25. Raptis CA, Hammer MM, Short RG et al (2020) Chest CT and coronavirus disease (COVID-19): a critical review of the literature to date. AJR Am J Roentgenol 16:1–4