



Research article

Standardized reporting systems of chest computed tomography in a population with low coronavirus disease 2019 prevalence: A retrospective comparative study



Ryo Kurokawa^a, Shohei Inui^a, Wataru Gono^{a,*}, Yudai Nakai^a, Masanori Ishida^a, Yusuke Watanabe^a, Takatoshi Kubo^a, Yosuke Amano^b, Koh Okamoto^c, Hidenori Kage^b, Sohei Harada^d, Goh Tanaka^b, Takuya Kawahara^e, Takahide Nagase^b, Kyoji Moriya^c, Osamu Abe^a

^a Department of Radiology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

^b Department of Respiratory Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

^c Department of Infectious Diseases, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

^d Department of Infection Control and Prevention, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

^e Clinical Research Promotion Center, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

ARTICLE INFO

Keywords:

Diagnostic performance
Interobserver agreement
Chest computed tomography
COVID-19

ABSTRACT

Purpose: To compare the diagnostic performance and interobserver agreement of three reporting systems for computed tomography findings in coronavirus disease 2019 (COVID-19), namely the COVID-19 Reporting and Data System (CO-RADS), COVID-19 Imaging Reporting and Data System (COVID-RADS), and Radiological Society of North America (RSNA) expert consensus statement, in a low COVID-19 prevalence area.

Method: This institutional review board approval single-institutional retrospective study included 154 hospitalized patients between April 1 and May 21, 2020; 26 (16.9 %; 63.2 ± 14.1 years, 21 men) and 128 (65.7 ± 16.4 years, 87 men) patients were diagnosed with and without COVID-19 according to reverse transcription-polymerase chain reaction results, respectively. Written informed consent was waived due to the retrospective nature of the study. Six radiologists independently classified chest computed tomography images according to each reporting system. The area under receiver operating characteristic curves, sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and interobserver agreements were calculated and compared across the systems using paired t-test and kappa analysis.

Results: Mean area under receiver operating characteristic curves were as follows: CO-RADS, 0.89 (95 % confidence interval [CI], 0.87–0.90); COVID-RADS, 0.78 (0.75–0.80); and RSNA expert consensus statement, 0.88 (0.86–0.90). Average kappa values across observers were 0.52 (95 % CI: 0.45–0.60), 0.51 (0.41–0.61), and 0.57 (0.49–0.64) for CO-RADS, COVID-RADS, and RSNA expert consensus statement, respectively. Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were the highest at 0.71, 0.53, 0.72, 0.96, and 0.56 in the CO-RADS; 0.56, 0.31, 0.54, 0.95, and 0.35 in the COVID-RADS; 0.83, 0.49, 0.61, 0.96, and 0.55 in the RSNA expert consensus statement, respectively.

Conclusions: The CO-RADS exhibited the highest specificity, positive predictive value, which are especially important in a low-prevalence population, while maintaining high accuracy and negative predictive value, demonstrating the best performance in a low-prevalence population.

1. Introduction

The prevalence of coronavirus disease 2019 (COVID-19) differs significantly across countries and regions. In addition, the degree of

ongoing local transmission has been changing very dynamically [1]. In Tokyo, Japan, disease prevalence was 268 per million population in April 2020 [2, 3]. Pending the development and delivery of highly effective medications and vaccines, public health interventions constituted the

* Corresponding author.

E-mail address: gonoiw@gmail.com (W. Gono).

<https://doi.org/10.1016/j.heliyon.2021.e07743>

Received 1 November 2020; Received in revised form 8 April 2021; Accepted 5 August 2021

2405-8440/© 2021 The Author(s). 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/>).

primary transmission-reduction strategy [4, 5, 6]. The pandemic will very likely continue with a fluctuation of new cases. In high-prevalence areas of COVID-19, rapid diagnostic tools with high sensitivity and negative predictive value (NPV) are crucial to isolate the patients with COVID-19 and initiate specialized care without delay. Conversely, in low-prevalence areas, both high specificity and positive predictive value (PPV) are also required to avoid excessive isolation and precautions. Although the specificity (0.95–1 with 32.0–38.0 % of COVID-19 cases) [7, 8] and PPV (0.95–0.96 with 17.7–19.9 % of COVID-19 cases [9]) of reverse transcription-polymerase chain reaction (RT-PCR) are high, specific equipment is required to perform, involving a lengthy result-acquisition process, making it unsuitable for a rapid determination of the necessity of specialized management for COVID-19.

Chest computed tomography (CT) can determine the presence and extent of pneumonia rapidly and has been sometimes applied for screening or a follow-up study for COVID-19. While COVID-19 can present a broad spectrum of CT findings, several standardized reporting systems of CT findings in COVID-19 have been proposed to facilitate the standard CT diagnosis with an acceptable inter-observer agreement. Among these systems, the Radiological Society of North America (RSNA) expert consensus statement [10], COVID-19 Reporting and Data System (CO-RADS) [11], and the COVID-19 Imaging Reporting and Data system (COVID-RADS) [12], have been commonly used. These three systems were initially designated as simple and sensitive surveillance tools to aid the diagnosis of COVID-19 during the ongoing community transmission. However, the application of these standardized reporting systems to low-prevalence study populations is unknown. Moreover, no study has conducted a detailed diagnostic performance and interobserver agreement comparison among these reporting systems.

To fill the knowledge gap, we aimed to evaluate the diagnostic performance and interobserver agreement of the three sets of a reporting system for chest CT in a low-prevalence patient population.

2. Material and methods

2.1. Ethical approval

This single-center retrospective comparative study was conducted with the approval of our Institutional ethics review board and was performed in accordance with the 1964 Declaration of Helsinki and its later amendments. Written informed consent was waived due to the retrospective nature of the study. The privacy of all patients was protected.

2.2. Sample size calculation

The primary endpoint of this cohort study was assessed by both an area under a receiver operating characteristic curve (AUC) and inter-observer correlation (as indicated by the kappa-value) with expected values of AUC and kappa-value set at 0.91 and 0.47 respectively by referring to the original study of CO-RADS [11]. The sample size was determined for the lower limit of 95 % confidence intervals of 0.86 (for AUC) and 0.4 (for kappa) when each indicator of the primary endpoints reached the expected values. This estimate resulted in the need for a total of 137 people, including 23 positive cases at a low positive rate, with 16.7 % of COVID-19 cases.

2.3. Study population

In our hospital, patients with one or more symptoms suggestive of COVID-19 (fever (≥ 37.5 °C), cough, dyspnea, tachypnea, malaise, hypoxemia) were reported by their attending physicians to a dedicated COVID-19 team, consisting of pulmonologists and infectious disease specialists for the purpose of COVID-19 control. Patients who reported to the dedicated COVID-19 team between April 1 and May 21, 2020, were included in the study population after confirming the presence of COVID-19-suggestive symptoms on the medical chart. A total of 192 patients met the criteria, excluding 38 due to being hospitalized longer than 2 weeks upon symptom prevalence to avoid including cases of nosocomial infection ($n = 9$), being younger than 20 years old to unify subjects with adult patients ($n = 9$), having no RT-PCR or CT examination performed ($n = 11$ and $n = 9$, respectively). Finally, 154 patients, including 26 with COVID-19 (16.9 %), were planned for further evaluation (Figure 1). Patients with COVID-19 in the hospital during the period were included in the study population.

Patients were classified either as COVID-19 or as non-COVID-19 according to RT-PCR results. RT-PCR for the detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from nasopharyngeal or sputum samples was performed using LightMix® Modular SARS and Wuhan CoV E-gene kit (TIB Molbiol, Berlin, Germany) and LightMix® Modular Wuhan CoV RdRP-gene kit (TIB Molbiol) with LightCycler Multiplex RNA Virus Master (Roche, Basel, Switzerland) according to the manufacturer's instructions [13]. Some patients underwent testing by RT-PCR multiple times, as a high index of clinical suspicion remained despite the negative result. Patients were diagnosed with COVID-19 if they tested positive on RT-PCR at least once.

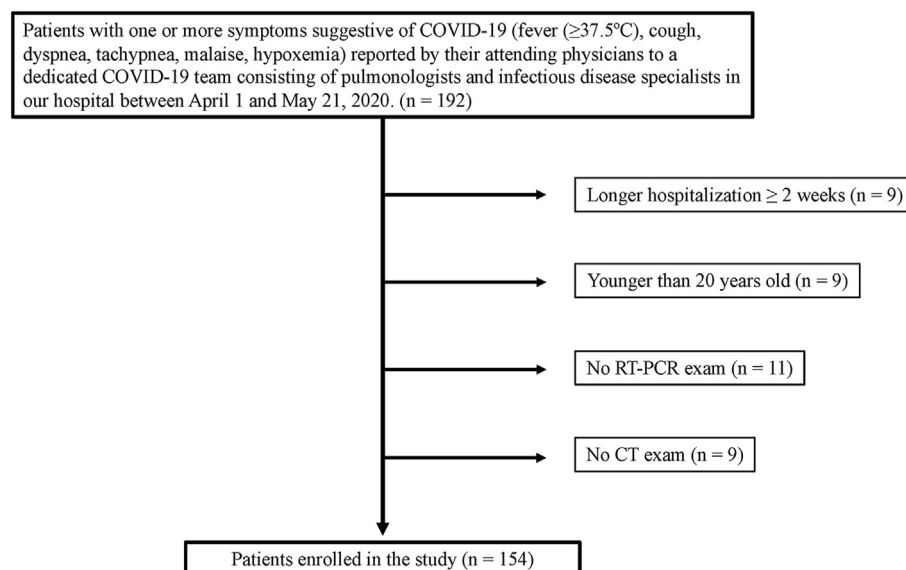


Figure 1. Flowchart of patient selection. COVID-19, coronavirus disease 2019; CT, computed tomography; RT-PCR, reverse transcription-polymerase chain reaction.

Table 1. Characteristics of the patient cohort.

Parameter	COVID-19 cases (N = 26)	Non-COVID-19 cases (N = 128)	P-value
Age [years, mean \pm SD (range)]	63.2 \pm 14.1 (24–89)	65.7 \pm 16.4 (16–95)	0.47
Gender (%)			0.19
Men	21 (80.8 %)	87 (68.0 %)	
Women	5 (19.2 %)	41 (32.0 %)	
Smoking (%)			0.52
Yes	10 (38.5 %)	59 (46.1 %)	
No	16 (61.5 %)	69 (63.9 %)	
RT-PCR test [times, number (%)]			<0.001
One	26 (100 %)*	47 (36.7 %)	
Two	0 (0 %)	71 (55.5 %)	
Three	0 (0 %)	10 (7.8 %)	
Symptom onset to CT [days, median (range)]	7 (0–14)	2 (0–34)	0.017
Symptom onset to positive RT-PCR result [days, median (range)]	7 (0–15)	–	
Comorbidities (%)			>0.99
Absent	5 (19.2 %)	24 (18.8 %)	
Present (any)**	21 (80.8 %)	104 (81.2 %)	
Respiratory disease	6 (23.1 %)	39 (30.5 %)	
Diabetes	7 (26.9 %)	30 (23.4 %)	
Hypertension	9 (34.6 %)	54 (42.2 %)	
Chronic kidney disease (grade 5)	1 (3.8 %)	18 (14.1 %)	
Cardiac disease	8 (30.8 %)	45 (35.2 %)	
Malignancy	3 (11.5 %)	41 (32.0 %)	
Immune deficiency	0 (0 %)	0 (0 %)	

COVID-19, coronavirus disease; CT, computed tomography; SD, standard deviation; RT-PCR, reverse transcriptase-polymerase chain reaction.

* Number of examinations until the first positive result was obtained.

** Multiple answers included.

2.4. Clinical data

The clinical and imaging findings extracted from the patient medical records were reviewed. We focused on demographic data, presence or absence of smoking history, underlying comorbidities, symptoms and signs, and duration from onset of symptoms to CT.

2.5. Chest CT acquisition

Chest CT was performed using a 4–320-row multi-detector CT unit (Aquilion PRIME, Aquilion ONE, Aquilion Precision, Alexion, Aquilion, Activion 16, Asteion; Canon, Tochigi, Japan; Revolution CT, Revolution Frontier, Revolution EVO, Discovery CT750 HD, LightSpeed VCT,

Table 2. RT-PCR positive and negative rates for each category.

	Sum of single observations	RT-PCR positive	RT-PCR negative	kappa (95 % CI)*
CO-RADS				
CO-RADS 1	170	10 (5.9 %)	160 (94.1 %)	0.58 (0.54–0.62)
CO-RADS 2	256	8 (3.1 %)	248 (96.9 %)	0.39 (0.35–0.43)
CO-RADS 3	107	2 (1.9 %)	105 (98.1 %)	0.25 (0.21–0.29)
CO-RADS 4	236	25 (10.6 %)	211 (89.4 %)	0.35 (0.31–0.39)
CO-RADS 5	155	111 (71.6 %)	44 (28.4 %)	0.76 (0.72–0.80)
COVID-RADS				
COVID-RADS 0	44	0 (0 %)	44 (100 %)	0.45 (0.41–0.49)
COVID-RADS 1	209	13 (6.2 %)	196 (93.8 %)	0.28 (0.24–0.32)
COVID-RADS 2A	35	1 (2.9 %)	34 (97.1 %)	0.19 (0.15–0.23)
COVID-RADS 2B	472	54 (11.4 %)	418 (88.6 %)	0.35 (0.31–0.39)
COVID-RADS 3	164	88 (53.7 %)	76 (46.3 %)	0.61 (0.57–0.65)
RSNA expert consensus statement				
Cov19Neg	150	8 (5.3 %)	142 (94.7 %)	0.56 (0.52–0.60)
Cov19Aty	246	9 (3.7 %)	237 (96.3 %)	0.38 (0.34–0.42)
Cov19Ind	315	10 (3.2 %)	305 (96.8 %)	0.38 (0.34–0.43)
Cov19Typ	213	129 (60.6 %)	84 (39.4 %)	0.69 (0.65–0.73)

RT-PCR, reverse transcriptase-polymerase chain reaction; CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data System; CI, confidence interval, RSNA, Radiological Society of North America; Cov19Neg/Aty/Ind/Typ, negative for pneumonia/atypical findings/indeterminate findings/typical findings in the RSNA expert consensus statement.

* Fleiss' kappa.

Table 3. Comparison of AUC and Cohen's kappa-values between the three criteria.

Criteria	AUC [mean, (95 % CI)]	Cohen's kappa [mean, (95 % CI)]
CO-RADS	0.89 (0.87–0.90)	0.52 (0.45–0.60)
COVID-RADS	0.78 (0.75–0.80)	0.51 (0.41–0.61)
RSNA consensus statement	0.88 (0.86–0.90)	0.57 (0.49–0.64)
P-value for one-way repeated measures analysis of variance	<0.001*	0.076
P-value for paired t-test		
CO-RADS vs COVID-RADS	0.0019**	
CO-RADS vs RSNA consensus statement	0.49	
COVID-RADS vs RSNA consensus statement	0.0050**	

Each AUC and Cohen's kappa value are derived from the average of each observer's value.

AUC, area under receiver operating characteristic curves; CI, confidence interval; CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data system; RSNA, Radiological Society of North America.

* Statistically significant.

** Statistically significant after family-wise error correction.

Optima CT660; GE, Tokyo, Japan; SOMATOM Emotion 6, SOMATOM go.Up; Siemens, Tokyo, Japan; IQon-Spectral; Philips, Tokyo, Japan) applying the following parameters: tube voltage, 120–140 kVp; effective current 69–749 mA; spiral pitch factor, 0.798–1.5, helical pitch, 7.2–127, the field of view, 27.5–45 cm; matrix size, 512 × 512. Acquisition parameters were modified to minimize patient radiation exposure while maintaining sufficient resolution for chest CT evaluation. A 1.0–5.0-mm gapless section was reconstructed before being reviewed on the picture archiving and communication system monitor. In total, 142 patients (COVID-19, n = 24; non-COVID-19, n = 118) were evaluated on CT with a slice thickness of 3 mm or less; 12 patients (COVID-19, n = 2; non-COVID-19, n = 10) were evaluated on CT with a slice thickness of 5 mm. Contrast-enhanced CT was performed in 28 patients, while the other 126 patients underwent plain CT alone.

2.6. CT image interpretation

Six radiologists (W.G., M.I., Y.W., T.K., R.K., S.I.) served as observers from a hospital in Japan. Five observers were board-certified diagnostic radiologists.

Randomization was stratified by the patients' disease status (patients with vs. without COVID-19). All patient information was de-identified, and observers were blinded to all clinical information. Each observer independently categorized the same CT according to the three reporting systems, namely, CO-RADS, COVID-RADS, and RSNA expert consensus statement in a row, and recorded all data using a spreadsheet prepared in advance (Microsoft, Redmond, WA, USA).

2.7. Statistical analysis

For quantitative analysis, findings on COVID-RADS and the RSNA expert consensus statement were converted into numerical grades. The conversion of both the COVID-RADS and RSNA expert consensus statement was as follows: for COVID-RADS, COVID-RADS 0 into grade 0, 1 into grade 1, 2A into grade 2, 2B into grade 3, and 3 into grade 4; for the RSNA expert consensus statement, negative for pneumonia (Cov19Neg) into grade 0, atypical appearance (Cov19Aty) into grade 1, indeterminate appearance (Cov19Ind) into grade 2, and typical appearance (Cov19Typ) into grade 3.

The AUC was calculated for each system by each of the observers. Using the RT-PCR results as the gold standard of COVID-19 diagnosis, the AUC was used to assess each of the three reporting systems' performance; moreover, we calculated both the mean AUC across observers and 95 % confidence interval (CI). Finally, we determined the average percentage of patients assigned to each category, including 95 % CI, for each criterion.

CT findings were considered radiologically positive for each system if classified as CO-RADS 5, COVID-RADS 3, or Cov19Typ; they were considered negative if classified as CO-RADS 1 or 2, COVID-RADS 0 or 1, or Cov19Neg or Cov19Aty. We considered only CO-RADS 5 as positive because a higher false-positive rate of CO-RADS 4 was expected in the

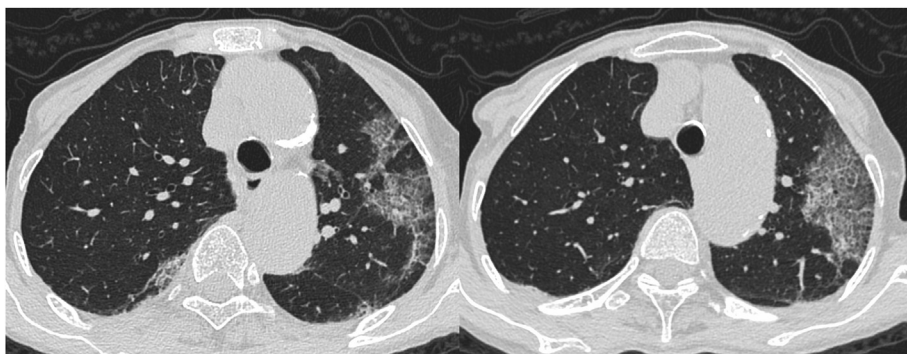


Figure 2. Typical coronavirus disease 2019 (COVID-19) case classified as positive with all three sets of reporting systems. Axial chest computed tomography images of a 77-year-old woman with COVID-19 obtained 5 days after the onset of symptoms show bilateral, peripheral multifocal ground-glass opacity. Crazy-paving appearance was found in the left lung. Our majority classification was CO-RADS 5, COVID-RADS 3, and typical for pneumonia in the RSNA expert consensus statement (true positive). CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data system; RSNA, Radiological Society of North America.

Table 4. Comparison of sensitivity, specificity, PPV, NPV, and accuracy between the three sets of criteria.^a

Criteria	Sensitivity	Specificity	PPV	NPV	Accuracy
CO-RADS	0.71 (111/156)	0.53 (408/768)	0.72 (111/155)	0.96 (408/426)	0.56 (519/924)
COVID-RADS	0.56 (88/156)	0.31 (240/768)	0.54 (88/164)	0.95 (240/253)	0.35 (328/924)
RSNA expert consensus statement	0.83 (129/156)	0.49 (379/768)	0.61 (129/213)	0.96 (379/396)	0.55 (508/924)

PPV, positive predictive value; NPV, negative predictive value; CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data System; RSNA = Radiological Society of North America.

^a Each case was diagnosed as positive when radiologists selected CO-RADS 5, COVID-RADS 3, or Cov19Typ for CO-RADS, COVID-RADS, or RSNA expert consensus statement, respectively, whereas negative when radiologists selected CO-RADS 1 or 2, COVID-RADS 0 or 1, or Cov19Neg or Aty.

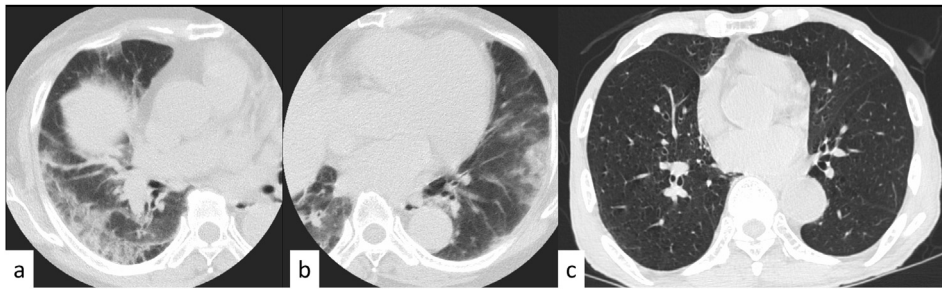


Figure 3. Representative false-positive and false-negative coronavirus disease 2019 (COVID-19) cases. (a, b) Axial chest computed tomography (CT) images of a 74-year-old woman without COVID-19 obtained 7 days after the onset of symptoms show bilateral, peripheral multifocal ground-glass opacity. Our majority classification was CO-RADS 5, COVID-RADS 3, and typical for pneumonia in the RSNA expert consensus statement (false positive on all the three reporting systems). (c) Axial chest CT image of an 89-year-old man with COVID-19 obtained on the same day as that of onset of symptoms shows no abnormal findings. Our majority classification was CO-RADS 1, COVID-RADS 0, and negative for pneumonia in the RSNA expert consensus statement (false negative on all the three reporting systems). CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data system; RSNA, Radiological Society of North America.

low-prevalence study population due to the definition, potentially overlapping with other diseases.

The interobserver agreement was quantified using two types of kappa value (Fleiss' kappa and Cohen's kappa) calculated across observers. Cohen's kappa values were obtained by comparing the grades of each observer to the median grade of the other remaining observers to compare with the original article of CO-RADS (i.e., to calculate the kappa value for observer 1, grades of observer 1 was compared to the median grades of the observer 2 to 6.) [11]. The overall agreement was quantified using Fleiss' kappa. The degree of interobserver agreement was judged using kappa: 0–0.20: poor; 0.21–0.40: fair; 0.41–0.60: moderate; 0.61–0.80: good; and 0.81–1.00: excellent agreement.

Statistical analysis was conducted using the JMP statistical software program (JMP Pro, version 15.0.0; SAS, Cary, NC, USA) and R software (R version 4.0.0, The R Foundation for Statistical Computing, Vienna, Austria). Quantitative variables were expressed as mean \pm standard deviation (range) or median and interquartile range based on the normality of data. Categorical variables were presented as the percentage of the total. The comparisons of quantitative variables were evaluated using a non-paired t-test or Mann-Whitney U-test and categorical data using the Pearson χ^2 test. We compared AUC and kappa-values using one-way repeated-measures analysis of variance, according to the normal

distribution assessed by the Shapiro-Wilk test; post hoc family-wise error (FWE)-corrected comparisons between each pair of the systems was conducted using a paired t-test while using Bonferroni's method in multiple comparisons. All *P* values corresponded to two-sided tests, while the statistical significance level was set at FWE-corrected *P* < 0.05.

3. Results

3.1. Clinical findings

Demographics and clinical characteristics of the study population are summarized in Table 1. The study population comprised 154 patients, with 26 patients with COVID-19 (16.9%; 21 men; mean age, 63.2years \pm 14.1; range, 24–89) and 128 patients without COVID-19 (87 men; mean age, 65.7years \pm 16.4, range; 16–95).

3.2. Diagnostic performances of the reporting systems

Six observers scored the CT images of 154 patients, resulting in a total of 924 observations for each reporting system. The probability of COVID-19 diagnosis of each category for each reporting system is summarized in Table 2, whose numbers are derived from the sum of all single

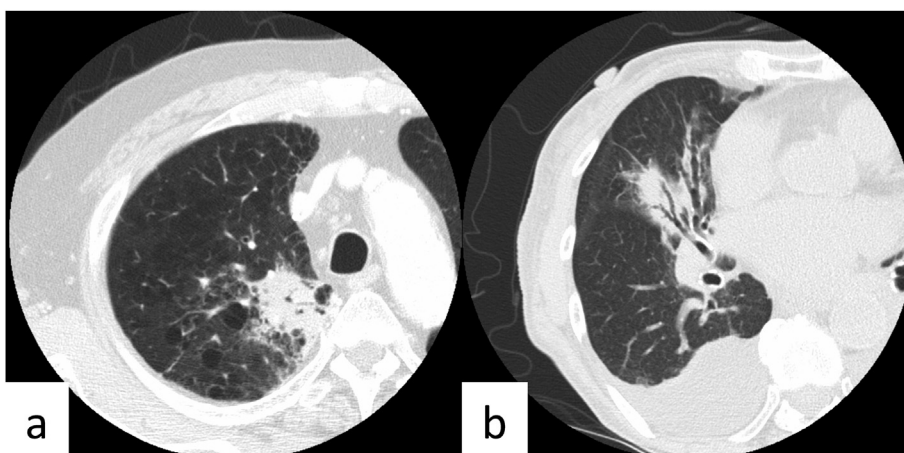


Figure 4. Cases illustrative of upgrading in COVID-RADS despite atypical patterns seen in coronavirus disease (COVID-19). (a) Axial chest computed tomography (CT) image of a 72-year-old man without COVID-19 obtained one day after the onset of symptoms shows consolidation along with bronchial ducts in the right upper lobe against a background of emphysema. Our majority classification was as follows: CO-RADS 2: typical for lobar pneumonia, COVID-RADS 2A + 1 (upgraded to 2B): consolidation without GGO (2A) + emphysema (1), RSNA expert consensus statement: atypical appearance. (b) Axial chest CT image of an 83-year-old woman without COVID-19 obtained on the same day as that of the onset of symptoms shows consolidation along with bronchial ducts in the right upper lobe with pleural effusion. Our majority classification was as follows: CO-RADS 2: typical for lobar pneumonia, COVID-RADS 2A + 1 (upgraded to 2B): consolidation without GGO (2A) + pleural effusion (1), RSNA expert consensus statement: atypical appearance, CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data system; RSNA, Radiological Society of North America.

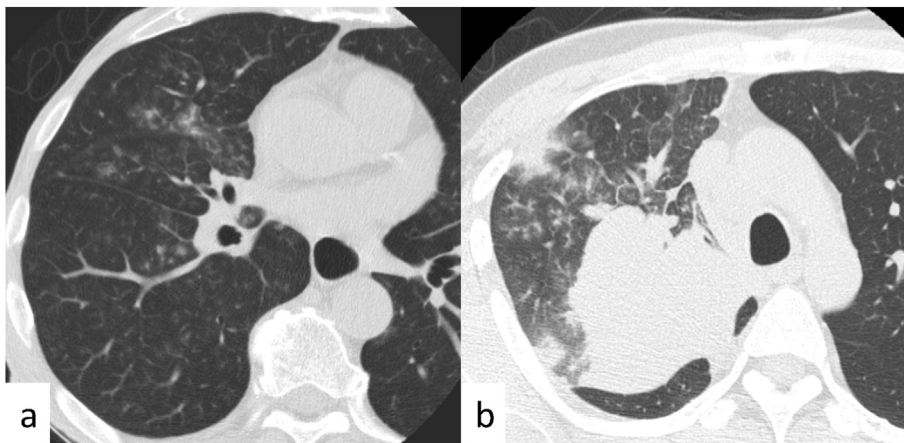


Figure 5. Cases illustrative of correct classification in CO-RADS other than the other two reporting systems. (a) Axial chest computed tomography (CT) image of a 65-year-old man without coronavirus disease 2019 (COVID-19) obtained 7 days after the onset of symptoms shows centrilobular nodules with surrounding ground-glass opacity (GGO) in the right lung. Our majority classification was as follows: CO-RADS 2: typical for bronchopneumonia, COVID-RADS 3 + 1 (2B): multifocal GGO (3) + centrilobular distribution (1), RSNA expert consensus statement: indeterminate appearance. (b) Axial chest CT image of a 47-year-old man without COVID-19 obtained on the same day as that of the onset of symptoms shows centrilobular nodules with surrounding GGO and a large mass in the right upper lobe of the lung. Our majority classification was as follows: CO-RADS 2: typical for obstructive bronchopneumonia caused by a tumor, COVID-RADS 2A + 1 (2B): multifocal GGO (3) + centrilobular distribution (1), RSNA expert consensus statement: indeterminate appearance. CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data system; RSNA, Radiological Society of North America.

categorizations by the six radiologists. The diagnostic performance and interobserver agreements for each reporting system set are summarized in Table 3. Average AUCs with 95 % CI for each criterion were: CO-RADS, 0.89 (95 % CI: 0.87–0.90); COVID-RADS, 0.78 (0.75–0.80); and RSNA scoring system, 0.88 (0.86–0.90). The AUC values were significantly lower in COVID-RADS than in CO-RADS or the RSNA scoring system ($P = 0.002$ or 0.0005 , respectively), with no significant difference between the CO-RADS and the RSNA scoring systems ($P = 0.49$).

As summarized in Table 2, radiological false negatives were noted in 10/170 (5.9 %) and 8/256 (3.1 %) in CO-RADS 1 and CO-RADS 2, respectively; false positives were observed in 44/155 (28.4 %) in CO-RADS 5. False negatives were found in 0/44 (0 %) and 13/209 (6.2 %) in COVID-RADS 0 and 1, respectively; false positives were noted in 76/164 (46.3 %) in COVID-RADS 3. False negatives were observed in 8/150 (5.3 %) and 9/246 (3.7 %) in Cov19Neg and Cov19Aty, respectively, whereas false positives were noted in 84/213 (39.4 %) in Cov19Typ.

As summarized in Table 4, sensitivity, specificity, PPV, NPV, and accuracy were the highest, at 0.71, 0.53, 0.72, 0.96, and 0.56, respectively, in CO-RADS vs. 0.56, 0.31, 0.54, 0.95, and 0.35 in COVID-RADS vs. 0.83, 0.49, 0.61, 0.96, and 0.55 in the RSNA scoring system.

Representative true-positive, false-positive, and false-negative cases are shown in Figures 2 and 3. Figure 4 shows CT images from non-COVID-19 patients with COVID-RADS 2B, CO-RADS 2, and Cov19Aty. Figure 5 shows

CT images from non-COVID-19 patients with CO-RADS downgraded to category 2, while COVID-RADS and RSNA expert consensus statements were classified as COVID-RADS 2B and Cov19Ind, respectively.

3.3. Inter-observer variabilities of the criteria

The inter-observer variabilities of COVID-19 diagnosis with 95 % CI of each scoring system were summarized in Table 3. The overall agreement was moderate; the average Cohen's kappa in all observers was 0.52 (95 % CI: 0.45–0.60), 0.51 (0.41–0.61), and 0.57 (0.49–0.64) for CO-RADS, COVID-RADS, and RSNA scoring system, respectively. No significant difference was found in Cohen's kappa between the three sets of reporting systems ($P = 0.08$). A representative case with non-uniform classifications in each of the three reporting system is shown in Figure 6.

The details of all 924 observations by all six observers are summarized in Supplementary Tables S1–S4. Associations between each system are summarized in Supplementary Tables S5–S7.

4. Discussion

We compared three standardized reporting systems of CT findings in COVID-19 (i.e., CO-RADS, COVID-RADS, and the RSNA expert consensus statement) in a low-prevalence study population. The AUC of COVID-RADS

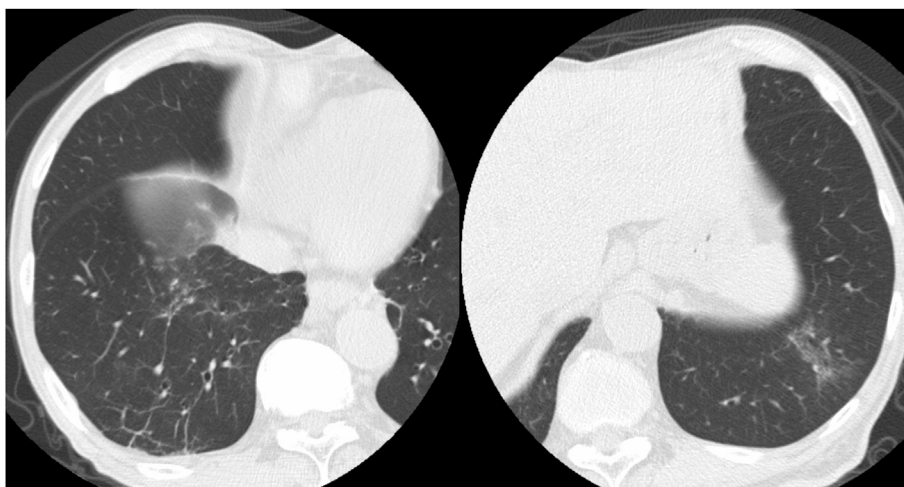


Figure 6. Cases illustrative of ununified classification in all reporting systems. Axial chest computed tomography image of a 77-year-old man with coronavirus disease 2019 (COVID-19) obtained 3 days after the onset of symptoms shows linear and reticulated opacity in the right lower lobe and ground-glass opacity along with bronchial duct in the left lower lobe. Our classification was divided as follows: CO-RADS: divided into 2, 3, and 4 (classified by two radiologists for each), COVID-RADS: divided into 1 (classified by one radiologist), 2A (two), 2B (two), or 3 (one), RSNA expert consensus statement: divided into Neg (classified by one radiologist who considered these findings not indicating pneumonia), Aty (two), and Ind (three). CO-RADS, COVID-19 Reporting and Data System; COVID-RADS, COVID-19 Imaging Reporting and Data system; RSNA, Radiological Society of North America.

peak was significantly lower than the CO-RADS's and the RSNA expert's consensus statement. Interobserver agreement was moderate in all systems, showing no significant differences between them. Specificity, PPV, and accuracy were the highest in CO-RADS while maintaining a high NPV.

The three standardized reporting systems were originally designed for simple and sensitive detection of patients with COVID-19 during the spread of the pandemic, potentially interpreted as a non-satisfactory diagnostic performance in a low-prevalence study population [10, 11, 12]. The overall diagnostic performance of each reporting system could have been lowered under the low prevalence of 268 cases per million populations in April in Tokyo. Actually, the overall diagnostic performance of each reporting system was observed to be high with AUC values of 0.89 and 0.88 with CO-RADS and the RSNA expert consensus statement, respectively. While both the AUCs of CO-RADS and the RSNA expert consensus statement were similar, they were significantly higher than the COVID-RADS's, which may be due to the wide range of applicability of the COVID-RADS 2B category, which potentially includes typical aspiration pneumonia or pulmonary edema by definition. Indeed, more than half of all single observations (472/924, 51.1 %) were classified into COVID-RADS 2B, and 88.6 % were without COVID-19 (418/472). As a result, cases classified as COVID-RADS 2B were given a less specific classification than in the other two systems. While cases classified as CO-RADS 4 ($n = 236$) or Cov19Ind ($n = 315$) were classified mostly as COVID-RADS 2B (193/236 (81.8 %) or 272/315 (86.3 %), respectively), cases classified as COVID-RADS 2B were classified as CO-RADS 4 and Cov19Ind in only 40.9 % (193/472) and 57.6 % (272/472), respectively (Supplementary Tables S5-S7). Both the high frequency and the wide applicability range of the COVID-RADS 2B category compared to that of CO-RADS 4 and Cov19Ind may have resulted in lower AUC, specificity, and PPV.

The CT image category in each reporting system is intended for determination either by the presence or absence of pre-defined imaging findings, regardless of the RT-PCR results; however, no high Fleiss' kappa values (0.39–0.49) were found in any assessed system in this study, which was likely directly related to the low prevalence of COVID-19 pneumonia, where the majority of patients had different and more diverse background features.

Kim et al. reported that the PPV of chest CT ranged between 0.02 and 0.27 in study populations with a low positive RT-PCR rate (<20 %) and an overall specificity of 0.25–0.33 in study populations with a prevalence of 25.1–33.3 % in their meta-analysis [9]. Because they included studies between January 1, 2020, and April 3, 2020, no studies included in their meta-analysis diagnosed COVID-19 on chest CT according to any of CO-RADS (published online on April 27, 2020), COVID-RADS (published online on April 28, 2020), or the RSNA expert consensus statement (published online on March 25, 2020). Comparison of the results between their meta-analysis and this study indicates both an enhanced specificity and PPV with CO-RADS (specificity, 0.52), COVID-RADS (specificity, 0.48), and with all three systems (PPV: 0.58–0.84), respectively. Among the three reporting systems in this study, the diagnostic performance of CO-RADS was the numerically highest regarding AUC, specificity, PPV, NPV, and accuracy. Especially, when classifying the CT image as CO-RADS 5, the RT-PCR positive rate was 111/155 (71.6 %), with the highest category kappa of 0.76 throughout the three systems. These results suggest that CO-RADS may deliver the best performance in a low-prevalence study population.

The limitations of this study are as follows: first, it is a single institutional retrospective study; second, some included cases incorporated no high-resolution CT images, potentially affecting the classifications; and third, the possibility that the reporting systems' high AUCs were attributable to the high non-COVID-19 rates of the study population cannot be excluded; however, these reporting systems showed both a higher specificity and PPV compared to the values reported in the meta-analysis for chest CT diagnosis with no reporting systems used in low-prevalence populations; as such, using these systems can be considered preferable.

5. Conclusion

Among the three standardized reporting systems, the AUCs were significantly higher with a CO-RADS and RSNA expert consensus statement than with COVID-RADS in a low-prevalence study population. The interobserver agreement was moderate with all of the systems. The specificity, PPV, and accuracy were the highest in CO-RADS while maintaining a high NPV. These results suggest that CO-RADS may deliver the best performance in a low-prevalence study population.

Declarations

Author contribution statement

Ryo Kurokawa Shohei Inui and Wataru Gono: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Masanori Ishida, Yusuke Watanabe and Takatoshi Kubo: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Yudai Nakai: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Goh Tanaka: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Takuya Kawahara: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Yosuke Amano, Koh Okamoto, Hidenori Kage, Sohei Harada and Osamu Abe: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Takahide Nagase and Kyoji Moriya: Contributed reagents, materials, analysis tools or data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare the following conflict of interests: Sohei Harada reports personal fees from BD, Meiji, Shionogi, Sumitomo Dainippon Pharma, Astellas, Beckman Coulter Diagnostics, FUJIFILM Toyama Chemical, grants and personal fees from MSD, outside the submitted work.

Osamu Abe: Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: institution received grants from Bayer Yakuhin, Canon Medical Systems, Eisai, FUJIFILM Toyama Chemical, GE.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2021.e07743>.

References

- [1] WHO Situation Report, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. (Accessed 9 August 2020).
- [2] COVID-19 the Information Website (Tokyo), 2020. <https://stopcovid19.metro.tykyo.lg.jp/en>. (Accessed 9 August 2020).

- [3] Statistics of Tokyo. <https://www.toukei.metro.tokyo.lg.jp/jsuikai/js-index.htm>. (Accessed 9 August 2020).
- [4] S. Flaxman, S. Mishra, A. Gandy, H.J.T. Unwin, T.A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J.W. Eaton, M. Monod, Imperial College COVID-19 Response Team, A.C. Ghani, C.A. Donnelly, S. Riley, M.A.C. Vollmer, N.M. Ferguson, L.C. Okell, S. Bhatt, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, *Nature* 584 (2020) 257–261.
- [5] A. Pan, L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, S. Wei, T. Wu, Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China, *J. Am. Med. Assoc.* 323 (2020) 1915–1923.
- [6] K. Prem, Y. Liu, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, M. Jit, P. Klepac, the effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, *Lancet Public Health* 5 (2020) e261–e270.
- [7] J.L. He, L. Luo, Z.D. Luo, J.X. Lyu, M.Y. Ng, X.P. Shen, Z. Wen, Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China, *Respir. Med.* 168 (2020) 105980.
- [8] J. Altamirano, P. Govindarajan, A.L. Blomkalns, L.E. Kushner, B.A. Stevens, B.A. Pinsky, Y. Maldonado, Assessment of sensitivity and specificity of patient-collected lower nasal specimens for sudden acute respiratory syndrome coronavirus 2 testing, *JAMA Netw Open* 3 (2020), e2012005.
- [9] H. Kim, H. Hong, S.H. Yoon, Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: a meta-analysis, *Radiology* 296 (2020) E145–E155.
- [10] S. Simpson, F.U. Kay, 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 radiology, the American college of radiology, and RSNA, *Radiol. Cardiothorac. Imag.* 2 (2020), e200152.
- [11] M. Prokop, W. van Everdingen, T. van Rees Vellinga, H. Quarles van Ufford, L. Stöger, L. Beenen, B. Geurts, H. Gietema, J. Krdzalic, C. Schaefer-Prokop, B. van Ginneken, M. Brink, CO-RADS, A categorical CT assessment scheme for patients suspected of having COVID-19—definition and evaluation, *Radiology* 296 (2020) E97–E104.
- [12] S. Salehi, A. Abedi, S. Balakrishnan, A. Gholamrezanezhad, Coronavirus disease 2019 (COVID-19) imaging reporting and data system (COVID-RADS) and common lexicon: a proposal based on the imaging data of 37 studies, *Eur. Radiol.* 30 (2020) 4930–4942.
- [13] C.C.Y. Yip, S. Sridhar, A.K.W. Cheng, K.H. Leung, G.K.Y. Choi, J.H.K. Chen, R.W.S. Poon, K.H. Chan, A.K.L. Wu, H.S.Y. Chan, S.K.Y. Chau, T.W.H. Chung, K.K.W. To, O.T.Y. Tsang, I.F.N. Hung, V.C.C. Cheng, K.Y. Yuen, J.F.W. Chan, Evaluation of the commercially available LightMix(R) Modular E-gene kit using clinical and proficiency testing specimens for SARS-CoV-2 detection, *J. Clin. Virol.* 129 (2020) 104476.