CO-RADS – A categorical CT assessment scheme for patients with suspected COVID-19: definition and evaluation

Original research

Mathias Prokop, MD, PhD* (1); Wouter van Everdingen, MD, PhD* (1), Tjalco van Rees

Vellinga, MD (2); Jet Quarles van Ufford, MD, PhD (2); Lauran Stöger, MD, PhD (3); Ludo

Beenen, MD (4); Bram Geurts, MD (1), Hester Gietema (5,6), Jasenko Krdzalic, MD, PhD (7),

Cornelia Schaefer-Prokop, MD, PhD (8), Bram van Ginneken MD, PhD (1), Monique Brink,

MD, PhD (1); for The "COVID-19 Standardized Reporting" Working Group of the Dutch

Radiological Society

1. Department of Radiology, Nuclear Medicine and Anatomy Radboudumc,

   Nijmegen, The Netherlands

2. Department of Radiology and Nuclear Medicine, Haaglanden Medical Center,

   The Hague, The Netherlands

3. Department of Radiology, Leiden University Medical Centre, Leiden, The

   Netherlands

4. Department of Radiology and Nuclear Medicine, Amsterdam UMC, location

   AMC, Amsterdam, The Netherlands

5. Department of Radiology and Nuclear Medicine, Maastricht University Medical

   Center+, The Netherlands

6. GROW School of Oncology and Developmental Biology, Maastricht, The

   Netherlands

7. Department of Radiology, Zuyderland MC, Heerlen, The Netherlands

8. Department of Radiology, Meander Medical Center, Amersfoort, The

   Netherlands

* M.P. and W.v.E. contributed equally to this work.

Word count: 3953 (without references and figure legends), abstract: 295

Running title: CO-RADS – A categorical CT assessment scheme for patients with suspected COVID-19: definition and evaluation

Address for correspondence:

Monique Brink, MD PhD

Department of Radiology Nuclear Medicine and Anatomy, Radboud University Medical Center

P.O. Box 9101

6500 HB Nijmegen, The Netherlands

Email: monique.brink@radboudumc.nl

Tel.: +31-(0)24 36 140 11

Summary statement

CO-RADS, for COVID-19 Reporting and Data System, is a categorical assessment scheme for chest CT in patients suspected of COVID-19, representing the level of suspicion for pulmonary involvement. The substantial agreement among observers and its discriminatory value make it well-suited for use in clinical practice.

Key results

- CO-RADS, for COVID-19 Reporting and Data System, provides a standardized assessment scheme that simplifies reporting with a five-point scale of suspicion for pulmonary involvement of COVID-19 on chest CT

- CO-RADS has a moderate to substantial agreement among observers with an overall Fleiss' kappa of 0.47 (95% CI 0.45-0.49).

- The discriminatory power of CO-RADS for diagnosing COVID-19 was high, with a mean area under the ROC curve of 0.91 (95% CI 0.85-0.97) for positive RT-PCR results.

Abbreviations

RT-PCR: reverse transcriptase-polymerase chain reaction.

ROC: receiver operating characteristics.

AUC: area under the ROC curve.

CI: confidence interval.

CT: computed tomography.

IQR: interquartile range.

Abstract

Purpose

To introduce the COVID-19 Reporting and Data System (CO-RADS) for standardized assessment of pulmonary involvement of COVID-19 on non-enhanced chest CT and report its initial interobserver agreement and performance.

Methods

The Dutch Radiological Society (NVvR) developed CO-RADS based on other efforts for standardization, such as Lung-RADS or BI-RADS. CO-RADS assesses the suspicion for pulmonary involvement of COVID-19 on a scale from 1 (very low) to 5 (very high). The system is meant to be used in patients presenting with moderate to severe symptoms of COVID-19. The system was evaluated using 105 chest CTs of patients admitted to the hospital with clinical suspicion of COVID-19 in whom RT-PCR was performed (62 +/- 16 years, 61 men, 53 with positive RT-PCR). Eight observers assessed the scans using CO-RADS. Fleiss' kappa was calculated, and scores of individual observers were compared to the median of the remaining seven observers. The resulting area under the receiver operating characteristics curve (AUC) was compared to results from RT-PCR and clinical diagnosis of COVID-19.

Results

There was absolute agreement among observers in 573 (68.2%) of 840 observations. Fleiss' kappa was 0.47 (95% confidence interval (CI) 0.45-0.47), with the highest kappa for CO-RADS categories 1 (0.58, 95% CI 0.54-0.62) and 5 (0.68, 95% CI 0.65-0.72). The average AUC was 0.91 (95% CI 0.85-0.97) for predicting RT-PCR outcome and 0.95 (95% CI 0.91-0.99) for

clinical diagnosis. The false negative rate for CO-RADS 1 was 9/161 (5.6%, 95% CI 1.0-10%), and the false positive rate for CO-RADS 5 was 1/286 (0.3%, 95% CI 0-1.0%).

Conclusions

CO-RADS is a categorical assessment scheme for pulmonary involvement of COVID-19 on non-enhanced chest CT providing very good performance for predicting COVID-19 in patients with moderate to severe symptoms and has a substantial interobserver agreement, especially for categories 1 and 5.

Introduction

Definitive diagnosis of COVID-19 is usually made using a reverse transcriptase-polymerase chain reaction (RT-PCR) assay, which performs accurately in a laboratory setting. However, reported sensitivities in clinical practice range between 42% and 83% depending on symptom duration, viral load, and test sample quality (1–5). Cases are increasingly reported in which the assay yielded a positive result only after multiple negative ones in patients with typical clinical and imaging signs of COVID-19 (6,7). Also, RT-PCR takes hours, or even days, before the results are available, putting strain on the holding units where patients are kept before being sent to a normal or COVID-19 ward. Increasingly, settings occur in which RT-PCR tests are scarce and will not be used on every patient.

In this light, the role of chest computed tomography (CT) in COVID-19 is constantly evolving with modest scientific evidence but substantial differences in opinion on when and how the technique should be used for clinical workup or treatment decisions. Whereas the *7th Chinese Novel Coronavirus Pneumonia Diagnosis and Treatment Plan* incorporates CT imaging into the criteria that clinically define COVID-19 (8), the American College of Radiology, among others, discouraged the use of CT in the initial workup of patients (9,10), only advocating its use for problem-solving. The Fleischner Society, in their recent statement, however, sees a role for imaging in various scenarios, with imaging, and in particular, CT scanning, as a major tool if symptoms worsen or in an environment that is resource-constrained for RT-PCR (11).

COVID-19 presents with CT findings that partially overlap with other diseases, mainly viral infections, but also shows characteristic features seen less frequently in other settings (12). Various attempts have been made to standardize reporting of CT for suspected COVID-19. The recent *RSNA expert consensus statement on reporting* (13), for example, proposes standardized nomenclature and an imaging classification for COVID-19 pneumonia that

involves four categories (i.e. typical appearance, indeterminate appearance, atypical appearance, and negative for pneumonia).

In early March 2020, the Dutch Radiological Society (Nederlandse Vereniging voor Radiologie, NVvR) initiated a COVID-19 network to facilitate development and nationwide dissemination of COVID-19-related information and tools. Within this network, a "COVID-19 standardized reporting" working group was formed. The authors developed a standardized assessment scheme for pulmonary involvement of COVID-19 to be able to compare data across institutions and populations and, thus, provide a basis for gathering scientific evidence and imporoved communication with referring physicians.

Because the system is based on other efforts for standardization, such as Lung-RADS, PI-RADS or BI-RADS, the authors chose the acronym CO-RADS, for COVID-19 Reporting and Data System (14). The system was iteratively refined through feedback from members and input from clinical partners. This type of system has been shown to work well in clinical practice and to allow for selection of optimal cutoff points for various clinical decisions depending on the tasks at hand. The current version represents the consensus formed on April 7th, 2020.

In this article, the definitions of CO-RADS are presented, along with the results of an initial observer study to assess the interobserver variability and diagnostic accuracy of the proposed system in the hands of observers with variable experience in reading chest CT for suspected COVID-19.

CO-RADS, the COVID-19 Reporting and Data System

CO-RADS provides a level of suspicion for pulmonary involvement of COVID-19 based on the features seen on a non-enhanced chest CT. The level of suspicion increases from very low (CO-RADS 1) to very high (CO-RADS 5). Two additional categories respectively encode a

technically insufficient examination (CO-RADS 0) and RT-PCR-proven SARS-CoV-2 infection at the time of examination (CO-RADS 6).

It should be noted that CO-RADS is a CT-based system that assesses the suspicion of pulmonary involvement in COVID-19. The actual interpretation of whether a patient suffers from COVID-19 needs to include other data, such as laboratory tests, clinical findings, and type and duration of symptoms. At present, the reference standard for diagnosing COVID-19 remains a positive RT-PCR. In clinical practice, however, this may require repeated testing including deep bronchial and fecal samples and may be hampered by scarcity of tests in high-prevalence areas.

An overview of CO-RADS is given in Table 1, and a pictorial overview is presented in Supplement 1.

CO-RADS 0

CO-RADS 0 is chosen if none of the five categories can be assigned because of scans that are incomplete or of insufficient quality, for example because of severe artifacts due to coughing or breathing.

CO-RADS 1

CO-RADS 1 implies a *very low level of suspicion* for pulmonary involvement by COVID-19 based on either a normal CT or CT findings of unequivocal non-infectious etiology. This was modelled on Lung-RADS, where cases with no nodules or with nodules with definitely benign features are reported together. As opposed to BI-RADS, where category 1 refers to normal only, we consider this approach more suitable for potential COVID-19 patients; concomitant findings are frequent in the lung and there is considerable interobserver variability regarding which findings are considered normal or not. Using our definition, mild or severe

emphysema, perifissural nodules, lung tumors, or fibrosis are classified as CO-RADS 1. The category is identical to the "negative for pneumonia" category of the RSNA consensus statement (13).

CO-RADS 2

CO-RADS 2 implies a *low level of suspicion* for pulmonary involvement by COVID-19 based on CT findings in the lungs that are typical of infectious etiology that are considered not compatible with COVID-19. Examples are bronchitis, infectious bronchiolitis, bronchopneumonia, lobar pneumonia, and pulmonary abscess. Features include tree-in-bud sign, a centrilobular nodular pattern, lobar or segmental consolidation, and lung cavitation. These features are similar to the ones in the "atypical appearance" category of the RSNA consensus statement (13). Cases with smooth interlobular septal thickening with pleural effusion, which is also part of this RSNA category, are assigned CO-RADS 1 if considered typical for interstitial pulmonary edema, or CO-RADS 3 if ground-glass opacities that may mimic pulmonary involvement by COVID-19 are also present. This choice was made because CO-RADS describes the pulmonary and not the cardiac involvement of COVID-19.

CO-RADS 3

CO-RADS 3 implies *equivocal findings* for pulmonary involvement of COVID-19 based on CT features that can also be found in other viral pneumonias or non-infectious etiologies. Findings include perihilar ground-glass, homogenous extensive ground glass with or without sparing of some secondary pulmonary lobules, or ground glass together with smooth interlobular septal thickening with or without pleural effusion in absence of other typical CT findings. CO-RADS 3 also includes small ground glass opacities that are not centrilobular (otherwise CO-RADS 2) or not located close to the visceral pleura (otherwise CO-RADS 4). In addition, it contains patterns of consolidation compatible with organizing pneumonia

without other typical findings of COVID-19. This category partially overlaps with the "indeterminate appearance" category of the RSNA consensus statement but includes those cases with lower likelihood for COVID-19 (13).

CO-RADS 4

CO-RADS 4 implies a *high level of suspicion* for pulmonary involvement by COVID-19 based on CT findings that are typical for COVID-19 but showing some overlap with other (viral) pneumonias. Findings are similar to CO-RADS 5 but are *not* located in contact with the visceral pleura or are located strictly unilaterally, are in a predominant peribronchovascular distribution, or are superimposed on severe diffuse pre-existing pulmonary abnormalities. CO-RADS 4 comprises the features of the "indeterminate appearance" category of the RSNA consensus statement that are associated with a higher likelihood of COVID-19 (13).

CO-RADS 5

CO-RADS 5 implies a *very high level of suspicion* for pulmonary involvement by COVID-19 based on typical CT findings (Table 2). *Mandatory features* are ground-glass opacities, with or without consolidations, in lung regions close to visceral pleural surfaces, including the fissures, and a multifocal bilateral distribution. Other classifications only describe a peripheral location, but we found that the vicinity to the minor or major fissure is also typical. Subpleural sparing is allowed to be present. We found that the previously described lower lobe predominance is frequently not present in otherwise typical RT-PCR-positive cases and therefore lower lobe predominance was excluded as a required feature.

CO-RADS 5 requires the presence of at least one *confirmatory pattern*, which aligns with the temporal evolution of the disease (15). The pattern that has been described early in the course of COVID-19 is dominated by multiple *ground-glass* areas, which often show (half) rounded and unsharp demarcation, but can be accompanied by sharply delineated ground-

glass areas that outline the shape of multiple adjacent secondary pulmonary lobules. The *crazy paving* pattern, which has been described to appear later in the course of the disease, shows visible intralobular lines. As the disease progresses, increasing *consolidations* occur within the ground-glass areas. Finally, opacities occur that resemble *organizing pneumonia*, such as reverse halo signs or ground glass with extensive subpleural consolidations and air bronchograms. Subpleural curvilinear bands or bands of ground glass with or without consolidation in a tethered, arching pattern with small connections to the pleura are also considered typical. Thickened vessels within lung abnormalities are typical and frequently found in all other confirmatory patterns. CO-RADS 5 is largely identical to the "typical appearance" of the RSNA consensus statement (13).

CO-RADS 6

CO-RADS 6, similar to BI-RADS 6, was introduced to indicate proven COVID-19 as signified by a positive RT-PCR test for virus-specific nucleic acid.

Methods

Study group

An observer study was performed on a set of 105 randomly selected chest CT scans obtained in a group of consecutive patients presenting to the emergency ward between March 14th 2020 and March 25th 2020 with suspected SARS-CoV-2 infection, in whom RT-PCR was performed. Patient inclusion, CT protocol, and radiation parameters are described in Supplement 2. Medical ethics committee approval was obtained prior to the study. Informed consent was waived, and data collection and storage were carried out in accordance with local guidelines.

Patient characteristics (age, gender, comorbidities), clinical follow up, including a multidisciplinary clinical diagnosis, if applicable, and RT-PCR results were extracted from electronic patient records. These data allowed stratification of all patients into one of the following three groups: patients with at least one positive RT-PCR result for SARS-CoV-2 within five days after CT (PCR+), patients with one or multiple negative RT-PCR results but a clinical diagnosis of COVID-19 according to clinical records (PCR-/Clinical+), and patients with one or multiple negative RT-PCR results and a clinical course not consistent with COVID-19, or consistent with an alternative diagnosis (PCR-/Clinical-).

CT scoring procedure

CT images were extracted from the picture archive and communication system, anonymized, and imported in a browser-based dedicated viewing system for CT scans (CIRRUS Core available at https://grand-challenge.org/reader-studies/). The software facilitated reading and scoring of anonymized CT images in the three orthogonal views, providing reading tools such as average or maximum intensity projections, window width-window level adaptation, panning and zooming.

Eight observers from seven hospitals in the Netherlands participated in the study (B.G., J.K., L.F.B., M.P., H.A.G., J.L.S., C.M.S-P., T.R.V.). Four observers had <5 years of experience in reading chest CTs, while the others had 5 to 27 years. All observers were familiar with the CO-RADS score from clinical experience interpreting at least 30 CT scans. Observers scored CO-RADS using a scoring software with drop-down lists, blinded for RT-PCR results and patient information except for age and gender. In addition, they were blinded for the prevalence of COVID-19 in the selected cohort, medical history, and clinical follow-up.

Statistical analysis

Statistical analysis was performed using SPSS statistics version 25 (IBM, Armonk, New York, USA). Data is presented as mean ± standard deviation or median and interquartile range (IQR), based on normality of data. A 5x5 confusion matrix was made separately per observer, in which the CO-RADS score of the observer was compared to the median CO-RADS score of the remaining seven observers. A similar matrix was calculated using the sum of all individual 5x5 tables.

For each observer, a receiver operating characteristics (ROC) curve was calculated, and the area under the ROC curve (AUC) was used to assess the performance of CO-RADS relative to two reference standards for the diagnosis of COVID-19: a positive RT-PCR test (PCR+) and a reference that combined the results of the RT-PCR test with a clinical COVID-19 diagnosis (PCR+ and PCR-/clinical+). Mean AUC across observers and 95% confidence intervals (95% CI) were calculated. In addition, the average percentage of cases assigned to each CO-RADS category, including 95% CI, were determined for PCR+, PCR-/clinical+, and PCR-/clinical- cases.

To quantify interobserver agreement, the Fleiss' kappa was determined across observers. Kappa values were obtained by comparing the CO-RADS scores of each observer to the

median of the remaining seven observers. Interobserver agreement was considered slight

for a kappa value of 0.01–0.20, fair for 0.21–0.40, moderate for 0.41– 0.60, substantial for

0.61–0.80, and almost perfect for 0.81–1.00 (16).

Results

Table 3 depicts baseline characteristics of the 105 included patients (62 +/- 16 years, 61 men, 53 with positive RT-PCR). In 21 patients at least one repeated RT-PCR was performed because of high clinical suspicion for COVID-19 but with a negative initial RT-PCR. An additional five patients had a clinical diagnosis of COVID-19 despite between one (n=2) and five RT-PCR negative tests (PCR-/Clinical+).

Interobserver variability of CO-RADS

There was absolute agreement in assigned CO-RADS category in 573 of all 840 (68.2%) observations. A discrepancy by a single CO-RADS category was seen in 235/840 (28.0%) observations, of which pairs of CO-RADS 4 and 5 and CO-RADS 1 and 2 occurred in 128/840 (15.2%) observations. A difference of two CO-RADS categories was found in 31 (3.7%) and of three categories in 1 (0.1%) of the 840 observations. The resulting 5x5 table is given in Supplement 3. The Fleiss' kappa of all observers on CO-RADS was 0.47 (95% CI 0.45-0.49). Kappa values and 95% confidence intervals for the individual categories were as follows: CO-RADS 1: 0.58 (0.54-0.62), CO-RADS 2: 0.36 (0.32-0.40), CO-RADS 3: 0.31 (0.28-0.35), CO-RADS 4: 0.20 (0.17-0.24), CO-RADS 5: 0.68 (0.65-0.72). The kappa value for each observer is provided in Table 4. Agreements of individual observers with the median of the remaining observers were either substantial (n=4) or moderate (n=4).

Diagnostic performance of CO-RADS

CO-RADS was able to distinguish between patients with PCR+ from those with PCR- with an average AUC of 0.91 (95% CI, 0.85-0.97). Average AUC increased to 0.95 (95% CI, 0.91-0.99) if a clinical diagnosis of COVID-19 was also accepted (Table 4). The proportion of cases with a positive PCR or clinical diagnosis of COVID-19 increased from CO-RADS 1 to 5, as shown in

Figure 1. All five CT scans of PCR-/Clinical+ patients were assigned CO-RADS categories 3 to 5.

False negative CO-RADS 1 relative to a combined clinical and RT-PCR reference standard were found in 9/161 (5.6%, 95% CI 1.0-10 %) of ratings. This occurred in four patients: one patient CT with bronchial wall thickening and very subtle ground-glass opacities in both lower lobes (1 observer), one CT with multiple ground-glass opacities only in the right lower lobe and a pleural lesion in the right upper lobe (2 observers), and two CTs with concurrent pre-existing disease (i.e. hypersensitivity pneumonitis and silicosis) (6 observers). CO-RADS 2 assessments were false negative for COVID-19 in 22/159 ratings (average 13.8%, 95% CI 9-18%), in 13 cases, including all four false negative cases with CO-RADS 1 scores. Correspondingly, 6 out of 97 (6.2%, 95% CI 1.5-10%) CO-RADS 4 observations, and 1 out of 286 (0.3%, 95% CI 0-1.0%) CO-RADS 5 observations were false positive, as assessed by 6 observers for 4 cases. RT-PCR and clinical diagnosis were almost equally distributed among observations of CO-RADS 3; 57 out of 137 (41.6%, 95% CI 31-54%) observations of CO-RADS 3 were COVID-19 positive. Figure 2 shows examples of RT-PCR positive and RT-PCR negative CO-RADS 3 observations, and examples of CO-RADS 4 and CO-RADS 5 observations are shown in Figure 3 and 4.

Discussion

CO-RADS was developed as a categorical system to assess the suspicion of lung involvement by COVID-19 on chest CT, and to provide standardized communication. With soaring case numbers and increasing logistic constraints, CO-RADS was readily embraced by clinicians for ease of communication and workflow optimization. CO-RADS assesses the suspicion for pulmonary involvement on CT. Accordingly, it has to be interpreted together with the duration and type of symptoms, as well as clinical and laboratory findings, when it comes to building a clinical diagnosis of COVID-19 with or without lung involvement before RT-PCR tests are available.

Our evaluation in a random sample of patients with symptoms suggestive of COVID-19 showed moderate to substantial interobserver agreement, despite the fact that all observers were from different hospitals and had different levels of exposure to CT in COVID-19 patients. Observer agreement was highest in categories CO-RADS 1 and 5, with kappa values of 0.58 and 0.68 respectively. In 68.2% of observations there was absolute agreement of scores, and in 15.2% scores varied between CO-RADS 1 and 2 or CO-RADS 4 and 5, indicating that observers agreed in more than 80% of the cases on the suspicion for pulmonary involvement of COVID-19 being either low to very low, or high to very high. The CO-RADS level of interobserver agreement (Fleiss' kappa 0.47) lies between values for PI-RADS (0.24) and Lung-RADS (0.67) (17,18).

In our setting with a high pre-test probability of disease in the acute phase of the pandemic, the performance of CO-RADS was very good, with an average AUC of 0.91, when compared to RT-PCR, and an AUC of 0.95, when compared to a combined RT-PCR and clinical reference standard. However, our results also indicate that the diagnosis of COVID-19 on CT remains difficult in a subset of patients, which underlines the importance of a reporting tool that

includes diagnostic confidence. Bernheim et al. described that CT can be negative at the early stages of COVID-19 (19), which might be the case for some of the 13/58 COVID-19 patients whose CTs were rated CO-RADS 1 or 2 by at least one observer. Therefore, a CO-RADS score of 1 and 2 should be interpreted with caution within the first days of disease presence. CO-RADS 3 encompasses a category in which CT alone offers little for the diagnosis of COVID-19. Presumably, knowledge of the prevalence of disease within the patient population, prior imaging studies, or a higher level of experience may decrease the number of equivocal calls. While CT findings are not specific for COVID-19 (12), they appear highly suggestive, which is underlined by only 1 false positive rating out of 286 CO-RADS 5 ratings.

Several caveats concern the performance of CO-RADS, mainly because the system was developed in the acute stage of the COVID-19 pandemic with rapidly rising case numbers and a parallel restriction in resources. Whether its accuracy remains high in other settings may depend on the prevalence of the disease, the duration of the pandemic, and the prevalence of other diseases with overlapping CT morphology. CO-RADS was developed in a high-prevalence setting, which implies that the positive predictive value is much higher than in a low prevalence situation. Also, no patients with residual abnormalities, such as subpleural banding after previous COVID infection, existed yet at the beginning of the pandemic. At the time that SARS-CoV-2 spread throughout Europe, the "influenza season" was coming to an end, reducing the number of overlapping patterns due to other viruses. Finally, whether this system suffices for patients with mild or no symptoms, has not been validated.

Our observer study has limitations. First, the study group is relatively small. Second, it is representative of a population presenting to the emergency ward in the acute phase of the SARS-CoV-2 outbreak and requiring hospital admission for clinical reasons. This increases the

disease prevalence substantially over a population with fewer symptoms. Third, observers had limited experience compared to areas with a larger outbreak of SARS-CoV-2. Finally, the diagnosis of COVID-19 was based on clinical decision despite negative RT-PCR results, but this occurred in a small subset of patients (n=5). In this multidisciplinary decision, the report of the CT scan was known, introducing an affirmation bias. Nevertheless, we included those patients in the study because it reflects current clinical practice.

Conclusion

The COVID-19 Reporting and Data System CO-RADS, developed by the Dutch Radiological Society, provides a framework that builds on other reporting schemes for COVID-19 but expands the concept in a way similar to systems like Lung-RADS. Its categories 1 to 5 provide increasing suspicion for pulmonary involvement of COVID-19 on non-contrast chest CT, thus allowing for task-specific cutoff points for clinical decision making. It provides very good performance for predicting COVID-19 in patients with moderate to severe symptoms and has substantial interobserver agreement, especially for categories 1 and 5. Therefore, the system fulfills the need for a structured and fast reporting system that decreases ambiguity in the communication with referring physicians and facilitates collection of CT performance data for further research of this worldwide healthcare problem.

Acknowledgments

References

1. Li Y, Yao L, Li J, et al. Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. J Med Virol. 2020;http://doi.wiley.com/10.1002/jmv.25786. Accessed April 15, 2020.

2. Ai T, Yang Z, Hou H, et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology. 2020;200642.

3. Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. 2020;https://jamanetwork.com/journals/jama/fullarticle/2762997. Accessed April 15, 2020.

4. Zou L, Ruan F, Huang M, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med. 2020;382(12):1177–1179.

5. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance. 2020;25(3)https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.3.2000045. Accessed April 16, 2020.

6. Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology. 2020;200432.

7. Long C, Xu H, Shen Q, et al. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? Eur J Radiol. 2020;126:108961.

8. Novel Coronavirus Pneumonia Diagnosis and Treatment Plan (Provisional 7th Edition), 04-03-2020; Chinese: http://www.gov.cn/zhengce/zhengceku/2020-03/04/5486705/files/ae61004f930d47598711a0d4cbf874a9.pdf English translation: https://www.chinalawtranslate.com/en/coronavirus-treatment-plan-7/. .

9. Hope MD, Raptis CA, Shah A, Hammer MM, Henry TS. A role for CT in COVID-19? What data really tell us so far. The Lancet. 2020;395(10231):1189–1190.

10. American College of Radiology. ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. March 11, 2020. https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection. Updated March 22, 2020. .

11. Rubin GD, Ryerson CJ, Haramati LB, et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology. Radiological Society of North America; 2020;201365.

12. Bai HX, Hsieh B, Xiong Z, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. Radiology. 2020;200823.

13. Simpson S, Kay FU, Abbara S, et al. 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 Imaging. 2020;2(2):e200152.

14. An JY, Unsdorfer KML, Weinreb JC. BI-RADS, C-RADS, CAD-RADS, LI-RADS, Lung-RADS, NI-RADS, O-RADS, PI-RADS, TI-RADS: Reporting and Data Systems. RadioGraphics. 2019;39(5):1435–1436.

15. Wang Y, Dong C, Hu Y, et al. Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. Radiology. 2020;200843.

16. Fleiss, J. L., & Cohen, J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educational and Psychological Measurement, 1973; 33(3), 613–619. https://doi.org/10.1177/001316447303300309. .

17. Smith CP, Harmon SA, Barrett T, et al. Intra- and interreader reproducibility of PI-RADSv2: A multireader study. J Magn Reson Imaging. 2019;49(6):1694–1703.

18. van Riel SJ, Jacobs C, Scholten ETh, et al. Observer variability for Lung-RADS categorisation of lung cancer screening CTs: impact on patient management. Eur Radiol. 2019;29(2):924–931.

19. Bernheim A, Mei X, Huang M, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology. 2020;200463.

Tables

Table 1. Overview of CO-RADS categories and the corresponding level of suspicion for

pulmonary involvement in COVID-19

	Level of suspicion for pulmonary involvement of COVID-19	Summary
CO-RADS 0	not interpretable	scan technically insufficient for assigning a score
CO-RADS 1	very low	normal or non-infectious
CO-RADS 2	low	typical for other infection but not COVID-19
CO-RADS 3	equivocal / unsure	features compatible with COVID-19, but also other diseases
CO-RADS 4	high	suspicious for COVID-19
CO-RADS 5	very high	typical for COVID-19
CO-RADS 6	proven	RT-PCR positive for SARS-CoV-2

Table 2. Features – typical for COVID-19

Typical features for pulmonary involvement of COVID-19

Obligatory features:
 	
● ground-glass opacities, with or without consolidations, in lung regions close to visceral pleural surfaces, including the fissures (subpleural sparing is allowed) *AND*
 	
● multifocal bilateral distribution
 	
Confirmatory patterns:
 	
● ground-glass regions
 	
    - unsharp demarcation, (half) rounded shape
 	
    - sharp demarcation, outlining the shape of multiple adjacent secondary pulmonary lobules
 	
● crazy paving
 	
● patterns compatible with organizing pneumonia
 	
● thickened vessels within parenchymal abnormalities found in all confirmatory patterns

Table 3. Baseline characteristics of the patient study group

Parameter	Value (n=105)
Age (mean years ± SD)	62 ±16
Sex (n, % men)	61 (58%)
Comorbidities (n, %)	
Diabetes	15 (14%)
Lung disease	41 (39%)
Cancer	22 (21%)
Immune deficiency	17 (16%)
Cardiovascular	46 (44%)
Duration of symptoms in days (IQR)	6.0 (2.0 – 10.0)
Number of RT-PCR assays	
1	84 (80%)
2	14 (13%)
3	4 (4%)
4	2 (2%)
5	1 (1%)
Positive RT-PCR results (n, %)	53 (51%)

Characteristics of the cases selected for the observer study. IQR: interquartile range. SD:

Standard deviation

Table 4. CO-RADS interobserver comparison and performance

Observer	kappa*	AUC vs RT-PCR*	AUC vs Diagnosis*
Observer 1	0.58 (0.47-0.69)	0.93 (0.88-0.98)	0.96 (0.93-0.99)
Observer 2	0.63 (0.52-0.74)	0.92 (0.86-0.97)	0.95 (0.91-0.99)
Observer 3	0.68 (0.57-0.78)	0.90 (0.84-0.96)	0.94 (0.90-0.99)
Observer 4	0.45 (0.34-0.57)	0.88 (0.81-0.95)	0.92 (0.87-0.98)
Observer 5	0.48 (0.37-0.59)	0.92 (0.87-0.98)	0.96 (0.93-1.00)
Observer 6	0.65 (0.55-0.76)	0.93 (0.88-0.98)	0.96 (0.93-1.00)
Observer 7	0.61 (0.51-0.72)	0.92 (0.87-0.98)	0.96 (0.93-1.00)
Observer 8	0.60 (0.49-0.71)	0.87 (0.80-0.94)	0.92 (0.86-0.97)

Overall	0.47 (0.45-0.49)**	0.91 (0.85-0.97)	0.95 (0.91-0.99)

Kappa characteristics of CO-RADS of each observer compared to the median of the other observers. Area under the curve of the receiver operating characteristics curve for each observer is given, separated for the reference standards defined by RT-PCR alone and RT-PCR together with clinical diagnosis.

*Data between parentheses are 95% confidence intervals

** Fleiss kappa

Figures & legends
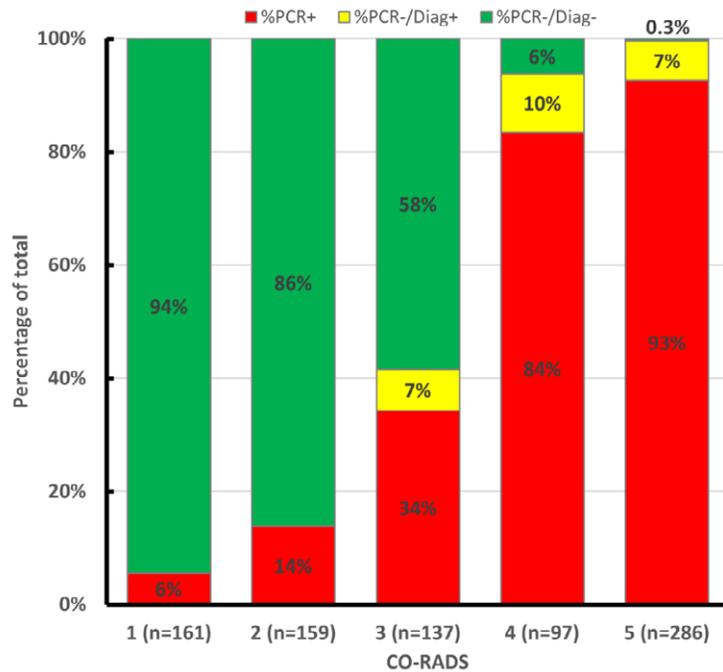
Figure 1. Cumulative CO-RADS score vs. RT-PCR results and clinical diagnosis.

The red columns display cases with a positive RT-PCR result. Yellow columns represent cases with a negative RT-PCR result, but a clinical diagnosis of COVID-19. Green columns display the percentage of cases with a negative RT-PCR result for SARS-Cov-2 and no clinical COVID-19 diagnosis.
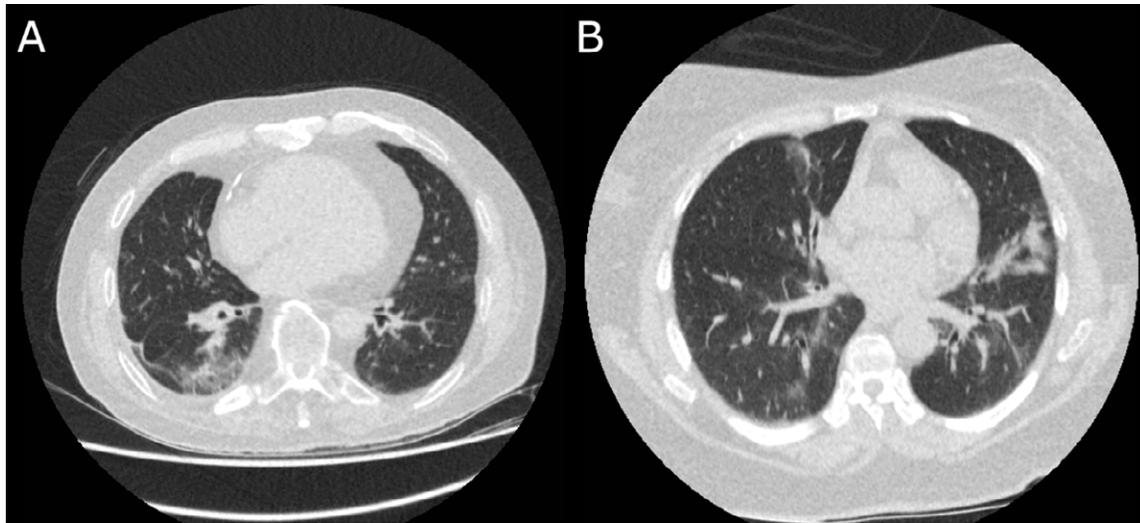
Figure 2. Examples of CO-RADS 3

Axial slices of the basal lungs of two cases with a majority of CO-RADS 3 observations. (A) 72 year-old male with a history of cardiovascular disease and COPD, who presented with fever and a productive cough since one day. He had a negative RT-PCR test for SARS-CoV-2 and a clinical diagnosis of community-acquired pneumonia He was treated with antibiotics and discharged after eight days. (B) 63 year-old female with diabetes, chronic kidney failure and hypertension, who presented with fever and cough since 3 days. RT-PCR was positive for SARS-CoV-2 and she was treated with oxygen therapy and discharged after two days with an advice for quarantine until full resolution of symptoms. Symptoms improved after a few days.
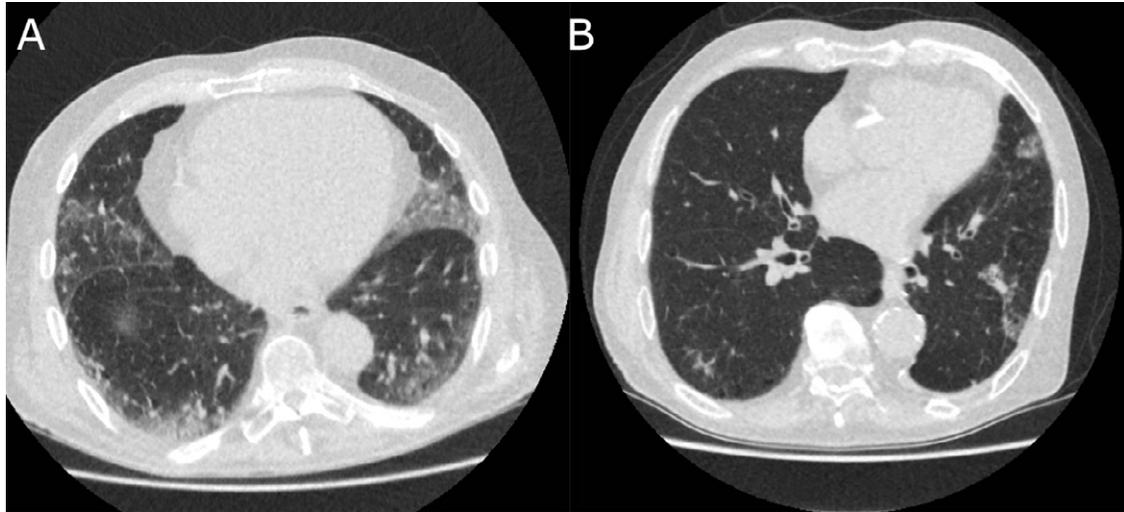
Figure 3. Examples of CO-RADS 4

Axial slices of the basal lungs of two cases with a majority of CO-RADS 4 observations. (A) 79 year-old male with a history of pulmonary embolism, who presented with a cough since 7 days and fever upon presentation. RT-PCR was positive for SARS-CoV-2, for which the patient was admitted and treated. Patient deceased despite treatment after fourteen days. (B) 78 year-old male with a history of COPD, lung cancer and hypertension, who presented with a productive cough and dyspnea since five days, with fever upon presentation. Despite initial negative RT-PCR test for SARS-CoV-2, a clinical diagnosis of COVID-19 was stated based on typical symptoms, CT characteristics and absence of an alternative diagnosis. The patient was discharged and advised to stay in quarantine until full resolution of symptoms.
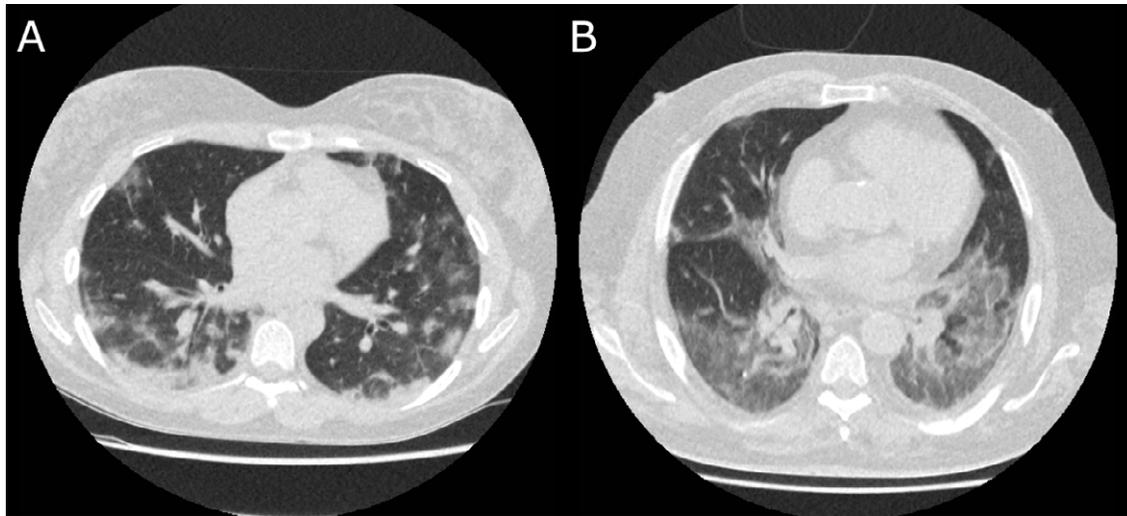
Figure 4. Examples of CO-RADS 5

Axial slices of the basal lungs of two cases with CO-RADS 5 observations by all eight observers.

(A) 30 year-old female, RT-PCR positive for SARS-CoV-2, who presented with fever and cough since 12 days. Patient was admitted to the COVID-19 ward. She was discharged after seven days, with resolution of symptoms. (B) 51 year-old male, presenting after eight days of fever, dyspnea and cough. A clinical diagnosis of COVID-19 was stated due to clinical symptoms and laboratory findings, despite a negative result at repeated RT-PCR. Patient was admitted for two days due to hypoxia with alleviation of symptoms after five days.

Supplement 1: Figures illustrating CO-RADS 1-5

Figure 1. CO-RADS 1

(a) Normal low-dose CT.

(b) Interstitial pulmonary edema.
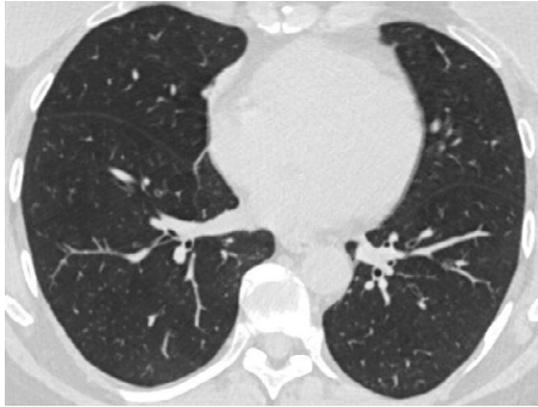
a   b

Figure 2. CO-RADS 2

(a) Dilated bronchi with thickened bronchial walls in chronic bronchitis combined with

tree-in-bud pattern indicative of infectious bronchiolitis.

(b) Consolidation in lobar pneumonia.
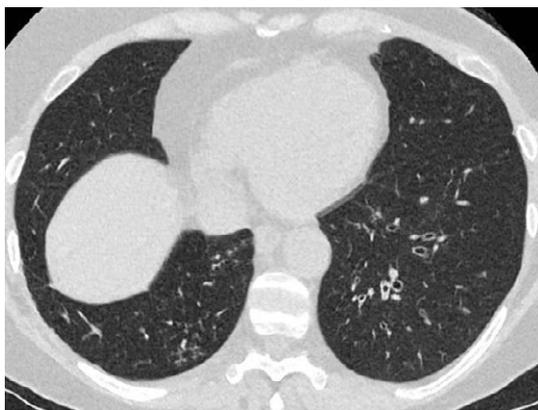
a   b

Figure 3. CO-RADS 3

a) Single focus of ground-glass opacity in a COVID-19 RT-PCR positive patient

b) Ground-glass with superimposed consolidations in a centrilobular pattern in a COVID-19
   RT-PCR negative patient
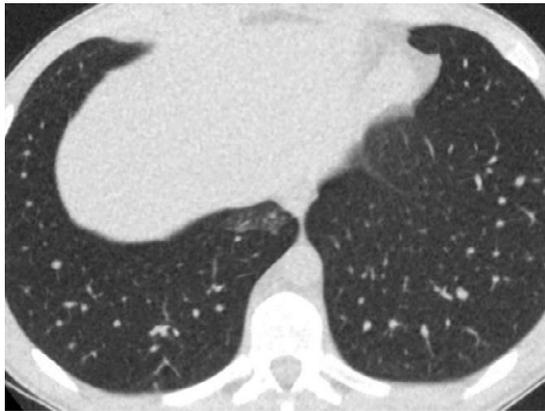
a   b

Figure 4. CO-RADS 4

(a) Unilateral multifocal ground-glass opacities with pleural contact in a COVID-19 RT-
    PCR positive patient.

(b) Ground glass superimposed on severe centrilobular emphysema, visceral pleural
   contact, unsharply demarcated in a COVID-19 RT-PCR positive patient.
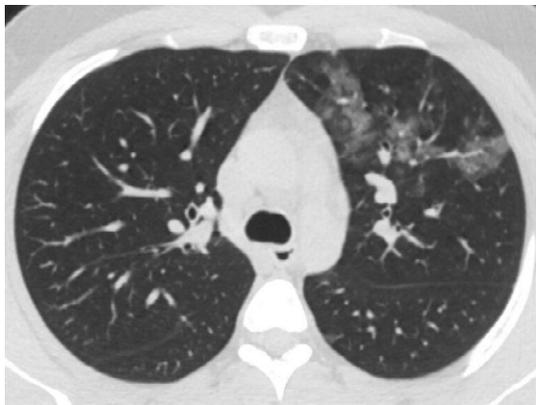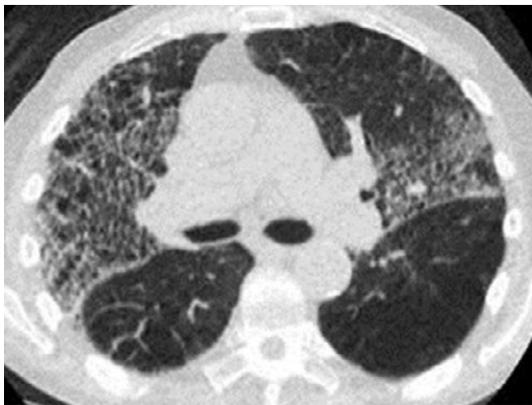
a   b

**Figure 5.** CO-RADS 5 in COVID-19 RT-PCR positive patients

(a) Multifocal bilateral subpleural ground-glass opacities, some with minimal subpleural sparing.

(b) Multifocal bilateral subpleural ground-glass opacities with curvilinear bands.

(c) Multifocal ground-glass opacities with inter- and intralobular septal thickening ("crazy paving") and consolidation.

(d) Multifocal bilateral ground-glass opacities with partial arcuated contact to the visceral pleura, consolidation and subpleural sparing.
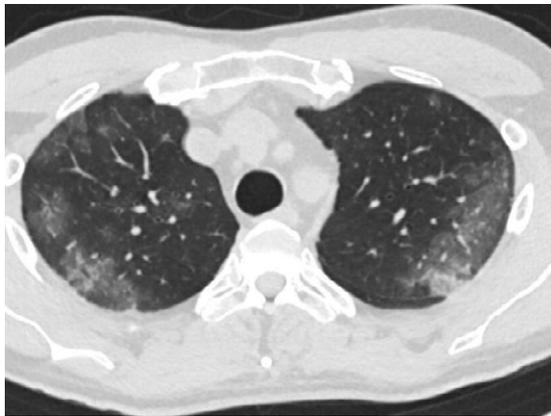

a

b
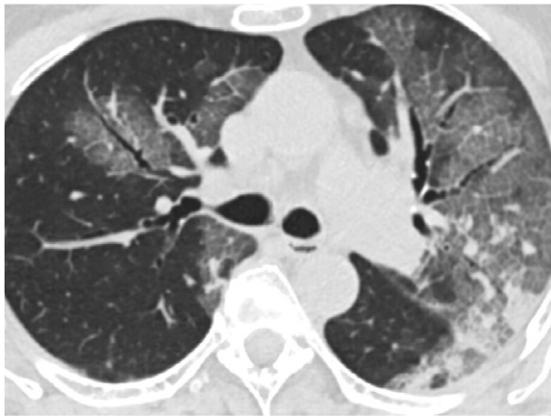
c
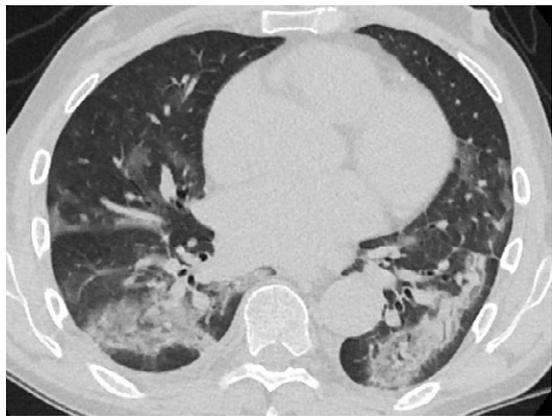
d

Supplement 2: Study group and CT protocol

We retrospectively collected imaging data of patients who presented with clinical suspicion of moderate to severe COVID-19 at a single academic center in the Netherlands (Radboud University Medical Centre, Nijmegen) between March 14th 2020 and March 25th 2020.

The study group consisted of a randomly selected subgroup of 105 of all consecutive patients undergoing a chest CT scan at the emergency ward with suspected SARS-CoV-2 infection, based on lower respiratory tract infection symptoms including cough and clinically relevant dyspnea requiring hospital admission with or without fever >38 °C. Patients were excluded if there was no RT-PCR test result for detection of viral SARS-CoV-2 RNA (target: RrRP-gen and E-gen) within five days after CT. A flow chart of patient inclusion is shown in the figure below.
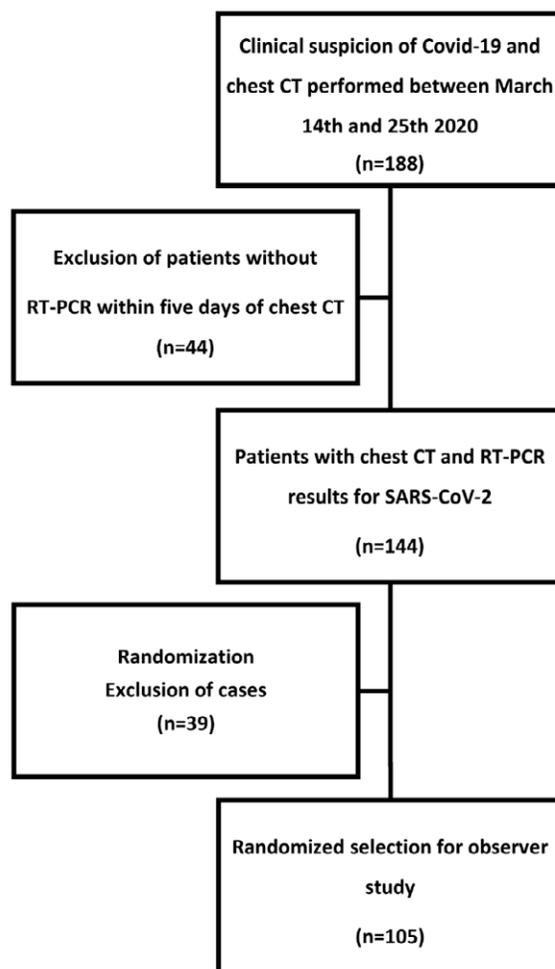

Figure: Flow-chart presenting the selection procedure of patients used for the observer study.

CT scans were obtained on a 320 multidetector CT (Canon Aquilion Vision, Canon Medical Systems, Otawara, Japan) without contrast injection using a low dose protocol with a fast

area-detector data acquisition protocol (collimation of 80 x 0.5 mm, rotation time of 0.275 s, tube energy of 100, 120 or 135 kV, and hybrid iterative reconstruction (AIDR 3D) with a pre-set noise tolerance of 35). A deep-inspiration breath-hold technique was applied whenever feasible.

For the included CT scans, radiation dose parameters were extracted from DICOM files. Tube energy was 100 kV in 31/105 patients (30%), 120 kV in 73/105 (70%) and 135 kV in 1 (1%) patient. Median $CTDI_{vol}$ was 1.0 mGy (IQR 0.7 – 1.6), and median DLP was 39.4 mGy*cm (IQR 25.4 – 56.5).

Supplement 3: Results

Table: 5x5 table of CO-RADS score of all observers

		Median of observers				
		CO-RADS 1	**CO-RADS 2**	**CO-RADS 3**	**CO-RADS 4**	**CO-RADS 5**
Sum of all single observers	**CO-RADS 1 (n=161, 19%)**	114/840 (14%)	40/840 (4.8%)	7/840 (0.8%)	0/840 (0%)	0/840 (0%)
	CO-RADS 2 (n=159, 19%)	23/840 (2.7%)	97/840 (12%)	37/840 (4.4%)	2/840 (0.2%)	0/840 (0%)
	CO-RADS 3 (n=137, 16%)	7/840 (0.8%)	41/840 (4.9%)	65/840 (7.7%)	15/840 (1.8%)	9/840 (1.1%)
	CO-RADS 4 (n=97, 12%)	0/840 (0%)	1/840 (0.1%)	14/840 (1.7%)	38/840 (4.5%)	44/840 (5.2%)
	CO-RADS 5 (n=286, 34%)	0/840 (0%)	1/840 (0.1%)	5/840 (0.6%)	21/840 (2.5%)	259/840 (31%)

For each observer the CO-RADS score was compared to the median of the CO-RADS scores of the other seven observers. The sum of all CO-RADS scores of single observers compared to the other observers is shown.