



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

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



Validation of the British Society of Thoracic Imaging guidelines for COVID-19 chest radiograph reporting



S.S. Hare, A.N. Tavare, V. Dattani, B. Musaddaq, I. Beal, J. Cleverley, C. Cash, E. Lemoniati, J. Barnett*

Department of Radiology, Royal Free NHS Foundation Trust, London, UK

ARTICLE INFORMATION

Article history:

Received 26 May 2020

Accepted 8 June 2020

AIM: To validate the British Society of Thoracic Imaging issued guidelines for the categorisation of chest radiographs for coronavirus disease 2019 (COVID-19) reporting regarding reproducibility amongst radiologists and diagnostic performance.

MATERIALS AND METHODS: Chest radiographs from 50 patients with COVID-19, and 50 control patients with symptoms consistent with COVID-19 from prior to the emergence of the novel coronavirus were assessed by seven consultant radiologists with regards to the British Society of Thoracic Imaging guidelines.

RESULTS: The findings show excellent specificity (100%) and moderate sensitivity (44%) for guideline-defined Classic/Probable COVID-19, and substantial interobserver agreement (Fleiss' $k=0.61$). Fair agreement was observed for the "Indeterminate for COVID-19" ($k=0.23$), and "Non-COVID-19" ($k=0.37$) categories; furthermore, the sensitivity (0.26 and 0.14 respectively) and specificity (0.76, 0.80) of these categories for COVID-19 were not significantly different (McNemar's test $p=0.18$ and $p=0.67$).

CONCLUSION: An amalgamation of the categories of "Indeterminate for COVID-19" and "Non-COVID-19" into a single "not classic of COVID-19" classification would improve interobserver agreement, encompass patients with a similar probability of COVID-19, and remove the possibility of labelling patients with COVID-19 as "Non-COVID-19", which is the presenting radiographic appearance in a significant minority (14%) of patients.

© 2020 Published by Elsevier Ltd on behalf of The Royal College of Radiologists.

Introduction

In December 2019, Wuhan City (Hubei Province, China) reported a cluster of patients displaying a febrile respiratory tract illness of unknown origin. Bronchoalveolar lavage of the patients isolated the pathogen as a novel strain of

coronavirus (SARS-coronavirus-2 [SARS-CoV-2]). The pulmonary infection caused by SARS-CoV-2 was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO).

In early 2020, the unprecedented surge in UK COVID-19 cases saw the chest radiograph (CXR) emerge as the

* Guarantor and correspondent: J. Barnett, Department of Radiology, Royal Free NHS Foundation Trust, London, UK. Tel. +020 7794 0500. E-mail address: josephbarnett@nhs.net (J. Barnett).

frontline diagnostic imaging test, in conjunction with clinical history and key blood biomarkers: C-reactive protein (CRP) and lymphopenia. The British Society of Thoracic Imaging (BSTI) developed a simple, internationally recognised CXR reporting template¹ to help facilitate consistency of reporting with embedded CXR reporting codes and allow retrospective radiology information system keyword searches for audit purposes. Frontline doctors have found this standardised reporting method a useful adjunct to clinical assessment, particularly when CXRs are “hot-reported”. The BSTI reporting template has been incorporated into an NHS England (NHSE) endorsed radiology decision tool for suspected COVID-19.^{1,2} Moreover, CXR has also emerged as a pivotal triage tool in proposed infection-control management strategies of inpatients as a result of reverse transcription-polymerase chain reaction (RT-PCR) result delays of up to 24–48 h and suspected initial false-negative COVID-19 nose/throat swab RT-PCR results in patients with a high clinical suspicion of COVID-19 infection.^{1,3}

Despite the prominence of CXR in the management of patients suspected to have SARS-CoV-2 pulmonary infection in Great Britain, to date very few studies have examined the CXR findings of COVID-19.^{4,5} Saliently, no study to date has examined the diagnostic utility of CXR against non-COVID-19 controls. Furthermore, the four diagnostic categories of chest radiograph introduced in the BSTI guidelines have not been validated regarding their utility or reproducibility.

The aim of the present study was to validate the BSTI COVID-19 CXR classification criteria with regards firstly to their reproducibility amongst consultant radiologists involved in the front-line care of patients with COVID-19, and secondly, to their diagnostic utility against symptomatic control patients without COVID-19.

Materials and methods

Patient selection

Consecutive adult patients with nose/throat swab RT-PCR-confirmed SARS-CoV-2 infection were identified from the microbiology database at Barnet General Hospital. Fifty consecutive patients were selected following exclusion of patients <18 years ($n=0$); patients with multiple organisms identified on PCR ($n=0$); and patients without admission CXR available on the picture archiving and communication system (PACS; $n=4$). As a retrospective evaluation of routinely collected clinical data, ethical approval was not required.

Given the limited application of RT-PCR testing at this time in England, and the reported non-trivial false-negative rate of RT-PCR testing for SARS-CoV-2,⁶ it is difficult to identify patients who are definitively negative for SARS-CoV-2; therefore, a control cohort of patients was selected from November 2019, prior to the emergence of SARS-CoV-2. Fifty consecutive adult patients with symptoms consistent with COVID-19 (new cough and fever) and available admission chest radiograph were selected from the PACS records.

Imaging evaluation

Images were anonymised regarding both patient identifiable data and date of image acquisition and stored in a random order on the Trust's PACS.

Seven consultant radiologists (median length of time on the specialist register 10 years, range 1–22 years) were recruited to participate in the study. They received training consisting of a review of the educational material available on the BSTI website⁷ regarding COVID-19 classification of CXRs. Participants were informed of the presenting complaint (new cough and fever, query COVID-19), and asked to categorise each CXR with regards to the BSTI guidelines⁷ (Fig 1). Patients were categorised as Classic/probable COVID-19, Indeterminate for COVID-19, Normal, and Non-COVID-19 according to Box 1.

In order to reach a final categorisation for each radiograph, the scores of two fellowship-trained thoracic radiologists (SSH, JB) were used, with consensus reached by post-hoc agreement.

Data collection and analysis

Data were extracted electronically from clinical records and analysed in R (<https://www.R-project.org>). Tests of comparative statistics were selected using Levene's test for homoscedasticity of variables and quintile–quintile plots for normality. Student's *t*-test was utilised for homoscedastic normally distributed data. Kruskal–Wallis test was used for non-normal homoscedastic data and Welch's *t*-test was used in the case of heteroscedasticity. Agreement was measured using unweighted Cohen's kappa for the final categorisation, with unweighted Fleiss' kappa used for assessment all participants. Only test statistics independent of disease prevalence (sensitivity and specificity) are reported. Confidence intervals for test statistics were calculated using the methods of Simel.⁸ McNemar's test was used to compare test statistics.

Results

Patient demographics are displayed in Table 1. Patients with SARS-CoV-2 infection were significantly older than those without, but there was no significant difference in gender. Although neither lymphocyte count nor lymphopenia (defined as lymphocyte count $<1 \times 10^9 \text{ l}^{-1}$) were significantly different in COVID-19 patients than controls, patients with COVID-19 had significantly greater CRP levels at presentation.

Amongst all radiologists, overall agreement of CXR categorisation was moderate (fleiss $K = 0.50$). Agreement for individual diagnostic categories was substantial for 'Classic/Probable COVID-19', ($k=0.61$) and 'Normal' ($k = 0.68$).⁹ Fair agreement was observed for the 'Indeterminate for COVID-19' ($k=0.23$), and 'Non-COVID-19' ($k=0.37$) categories. Post-hoc combination of the 'Indeterminate for COVID-19' and 'Non-COVID-19' codes into a single category was associated with improved inter-observer agreement ($k=0.58$).

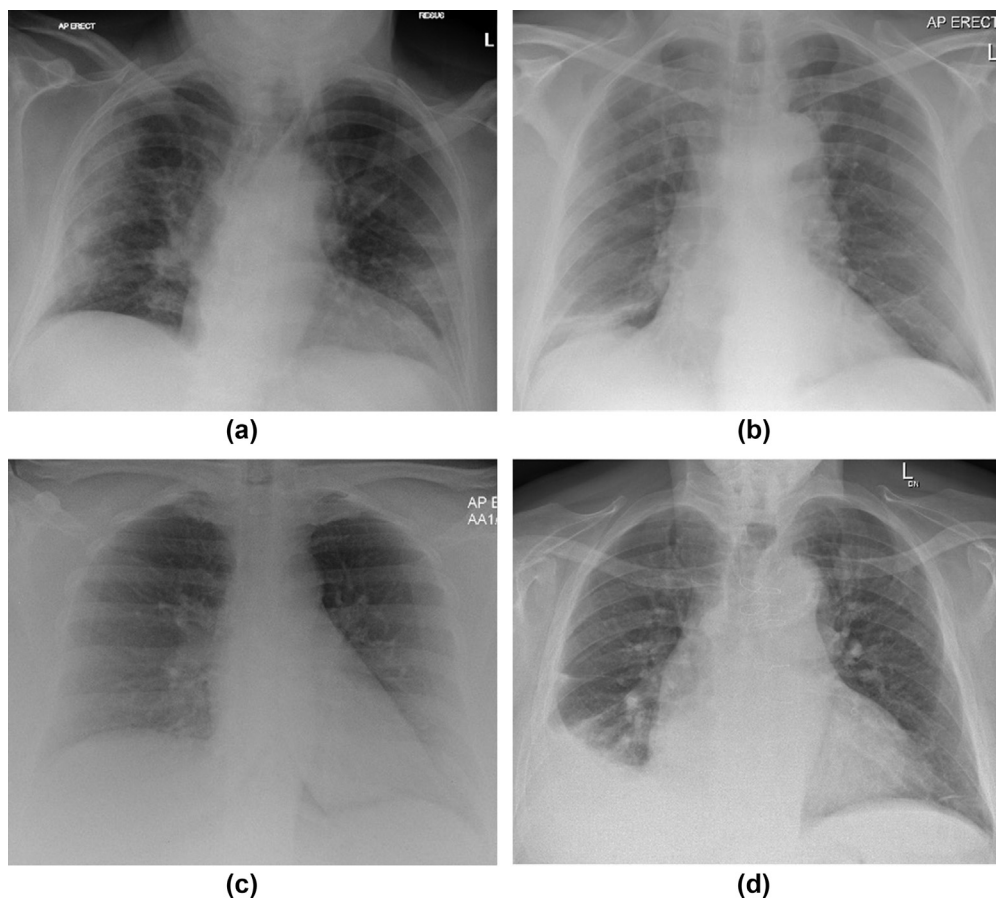


Figure 1 Examples of the COVID BSTI categories for plain films, in each case all radiologists agreed on the categorisation. (a) Anteroposterior (AP) erect radiograph demonstrating “Classic COVID-19”. (b) AP erect chest radiograph “Indeterminate for COVID-19”. (c) AP erect radiograph classified as “COVID normal”. (d) AP erect radiograph classified as “Non-COVID”.

For the purposes of final classification of CXRs, scores from two fellowship-trained thoracic radiologists (SSH, JB) were used, with disagreements arbitrated by consensus. Agreement amongst these radiologists was almost perfect for “Classic/Probable COVID-19” ($k=0.83$), substantial for “Normal” ($k=0.70$), moderate for “Non-COVID-19” ($k=0.50$) and slight for “Indeterminate” ($k=0.25$). The final classifications of patients are given in Table 2. The “Classic/Probable COVID-19” category was associated with 100% specificity for COVID-19, and detected 44% of patients with RT-PCR-confirmed SARS-CoV-2 infection. Normal CXRs

were significantly more frequent in controls ($p<0.001$ after adjustment for multiple testing), but still occurred in 16% of patients with RT-PCR-confirmed SARS-CoV-2 infection. The frequency of “Indeterminate for COVID-19” and “Non-COVID-19” chest radiographs was not significantly different between COVID-19 patients and controls, indeed the sensitivity and specificity of these categories for COVID-19 were not significantly different (McNemar’s test $p=0.18$

Box 1

The British Society of Thoracic Imaging chest radiography reporting criteria.

Normal	COVID-19 not excluded, please correlate with PCR
Classic/probable COVID-19	Lower lobe and peripheral predominant multiple opacities that are bilateral (>> unilateral)
Indeterminate for COVID-19	Does not fit Classic or Non-COVID-19 descriptors” or “poor quality film
Non-COVID-19	Pneumothorax/lobar pneumonia/pleural effusion(s)/pulmonary oedema/other

Table 1

Demographic and radiological data.

	COVID	Non-COVID	p-Value
Demographic data			
<i>n</i>	50	50	
Male, <i>n</i> (%)	33 (66.0%)	24 (48.0%)	0.07
Age, years (\pm SD)	68.6 (\pm 17.3)	55.4 (\pm 21.3)	0.001
Lymphocyte count, $\times 10^9 \text{ l}^{-1}$ (\pm sd)	1.06 (\pm 0.68)	1.22 (\pm 0.68)	0.23
Lymphopenia, <i>n</i> (%)	27 (54%)	23 (46%)	0.54
CRP, (IQR)	77 (114)	38 (100)	0.01

Proportions compared using fisher test; age and lymphocyte count expressed as mean and standard deviation and compared using Student’s *t*-test; CRP expressed as median and interquartile range and compared with Kruskal–Wallis test. Lymphopenia defined as lymphocyte count $<1.0 \times 10^9 \text{ l}^{-1}$.

CRP, C-reactive protein; SD, standard deviation; IQR, interquartile range.

Table 2

Proportions of patients compared using fisher test with Benjamini and Hochberg method of adjustment for multiple testing.

	COVID	Non-COVID	p-Value	Sensitivity	Specificity
Final CXR categorisation					
COVID classic	22	0	<0.001	0.44 (0.30, 0.59)	1.00 (0.93, 1.00)
COVID indeterminate	13	12	1	0.26 (0.15, 0.40)	0.76 (0.62, 0.87)
COVID normal	8	28	<0.001	0.16 (0.07, 0.29)	0.44 (0.30, 0.59)
Non-COVID	7	10	1	0.14 (0.06, 0.27)	0.80 (0.66, 0.90)

Sensitivity and specificity expressed with 95% confidence intervals.
CXR, chest radiography.

and $p=0.67$, respectively). Finally, 7/50 patients (14%) with RT-PCR-confirmed SARS-CoV-2 infection had an admission CXR classified as “Non-COVID-19”, examples of such patients are demonstrated in Fig 2.

Discussion

The results of the present study demonstrate that the BSTI “Classic/Probable COVID-19” categorisation is very specific and moderately sensitive for patients with RT-PCR-

confirmed SARS-CoV-2 pulmonary infection on admission CXR, as opposed to symptom-matched controls. Furthermore, this classification is substantially agreed upon by consultant radiologists.

A significant minority of patients in this study with SARS-CoV-2 infection presented with normal CXRs, findings that reinforce the BSTI “Normal” categorisation, which states that COVID-19 cannot be excluded and that RT-PCR may be required. These results, however, highlight that some refinement of the BSTI COVID-19 classification criteria

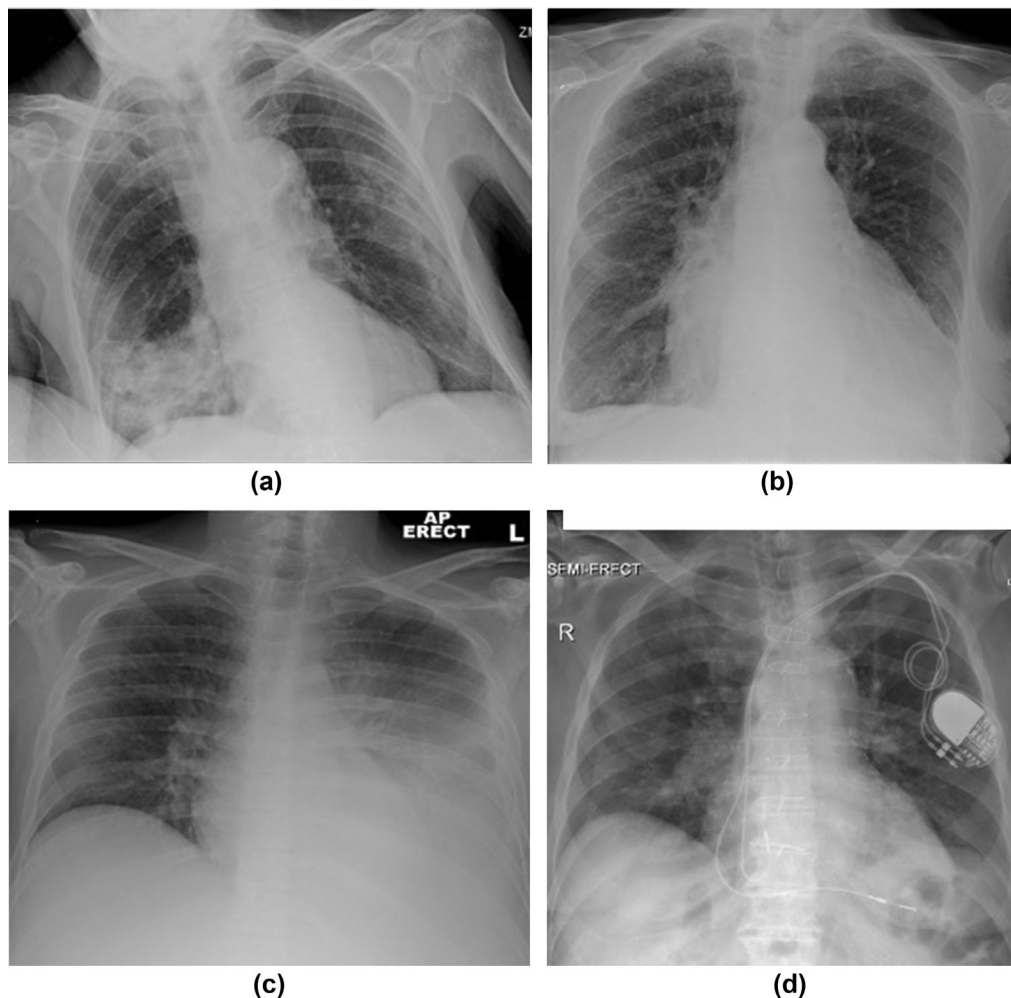


Figure 2 Examples of patients with SARS-CoV-2 infection, but admission chest radiographs classified as “Non-COVID”. (a) AP erect radiograph demonstrating lobar pneumonia. (b) AP erect radiograph showing congestive cardiac failure. (c) AP erect chest radiograph with unilateral pleural effusion. (d) AP semi-erect chest radiograph with left lower lung airways disease/thickening.

may be needed, specifically the categories of “Indeterminate for COVID-19” and “Non-COVID-19”. Only fair inter-observer agreement was observed for these categories, which in the case of “Indeterminate for COVID-19” fell to slight agreement when only the categorisation of fellowship-trained thoracic radiologists were used. In addition, these categories have similar diagnostic performance regarding SARS-CoV-2 infection.

The potential need for an iteration of these two categories is highlighted by an inherent overlap of CXR appearances between “Indeterminate for COVID-19” and “Non-COVID-19” categories. Examples of this overlap exist in patients with limited or unilateral consolidation, which could be SARS-CoV-2 or bacterial in aetiology; and in patients with multiple radiological abnormalities, for example, fluid overload and alveolar opacity.

In equivocal cases, it is expected that radiologists may also reasonably be informed by the pre-test probability of SARS-CoV-2 infection in assigning cases to the “Indeterminate” or “Non-COVID” categories. Thus, the categorisation of the same radiograph may differ depending on whether the patient presented in the first peak of the COVID-19 pandemic in London, when up to 80% of emergency department admissions were COVID related, as opposed to during a relative trough.

Inclusion of non-diagnostic examinations in the “Indeterminate for COVID-19” category also adds variation to this group of patients, compounding interobserver variation with regards to diagnosis and that regarding acceptable film quality. A non-diagnostic examination should be reported as such, and no statement about COVID-19 classification is necessary or possible in this situation. There is also variability in the recommendation implied by this diagnostic category; patients with a non-diagnostic film are more likely to benefit from a repeated attempt at imaging, whereas those with a diagnostic-quality examination revealing a non-specific abnormality may not.

As the prevalence of COVID-19 increases and as health-seeking behaviours of the population respond to the presence of a pandemic, atypical radiographic presentations of SARS-CoV-2 infection will become more frequent relative to non-coronavirus disease. Indeed, 14% of patients in this study with SARS-CoV-2 infection had a CXR categorised as “Non-COVID”. The amalgamation of the “Indeterminate for COVID-19” and “Non-COVID-19” categories into a single “not classic of COVID-19” category would have several advantages. Firstly, this category would increase consultant radiologist agreement. Secondly, the category would encompass a group of patients with similar probability of SARS-CoV-2 infection. Thirdly, it would remove the possibility of potentially mislabelling patients with SARS-CoV-2 infection as “Non-COVID-19”.

The present study has a number of limitations. Firstly, performance of a given test varies with disease prevalence. In the present study, patients and controls were matched at a 1:1 ratio. Throughout the height of the pandemic, the authors’ anecdotal experience is of patients with COVID-19 outnumbering those without. In order to minimise the effect of varying prevalence of SARS-CoV-2 on the results,

only sensitivity and specificity have been presented, which are statistics independent of disease prevalence. Using PCR as the reference standard diagnosis in a study examining radiological diagnosis is necessary to avoid incorporation bias, but introduces the biases of PCR testing strategy, namely towards those patients with more severe disease requiring admission. This may have the effect of overstating the sensitivity of the “Classic/Probable COVID” category, but should not affect the specificity. It is currently uncertain whether patients with false-negative SARS-CoV-2 RT-PCR have a distinct radiological phenotype to those who are RT-PCR positive; this could also introduce bias into the data.

In conclusion, the results of the present demonstrate variable performance of the BSTI COVID-19 classification criteria. The guideline defined “Classic/Probable” appearance of COVID-19 has excellent specificity and moderate sensitivity for SARS-CoV-2 pulmonary infection, and furthermore, is associated with substantial interobserver agreement. The categories of “Indeterminate for COVID-19” and “Non-COVID-19”, however, suffer from greater interobserver variability, and furthermore, have similar sensitivity and specificity for COVID-19. The authors suggest an amalgamation of these categories into a “not classic of COVID-19” category, which would increase interobserver agreement; encompass patients with similar probability of COVID-19; and remove the potential for mislabelling patients with SARS-CoV-2 infection as “Non-COVID”.

Conflicts of interest

Dr Hare is on the committee of the British Society of Thoracic Imaging.

Acknowledgements

The authors thank Mr Jack Gaskell for his help in image anonymisation. S.S.H. is on the committee of the British Society of Thoracic Imaging. The authors declare that they have no further relevant conflicts of interest.

References

1. Hare S, Rodrigues J, Nair A, et al. The continuing evolution of COVID-19 imaging pathways in the UK: a British Society of Thoracic Imaging expert reference group update. *Clin Radiol* 2020;75:399–404. <https://doi.org/10.1016/j.crad.2020.04.002>.
2. NHS England. Specialty guides for patient management during the coronavirus pandemic Clinical guide for the management of Radiology patients during the coronavirus pandemic. Available at: <https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/specialty-guide-radiology-and-coronavirus-v1-20-march-2020.pdf>. [Accessed 14 May 2020].
3. Taware AN, Braddy A, Brill S, et al. Managing high clinical suspicion COVID-19 inpatients with negative RT-PCR: a pragmatic and limited role for thoracic CT. *Thorax* April 2020, <https://doi.org/10.1136/thoraxjnl-2020-214916>.
4. Yoon SH, Lee KH, Kim JY, et al. Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): analysis of nine patients treated in Korea. *Korean J Radiol* 2020;21(4):498–504. <https://doi.org/10.3348/kjr.2020.0132>.

5. Wong HYF, Lam HYS, Fong AHT, et al. Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology* 2019;2011160. <https://doi.org/10.1148/radiol.2020201160>.
6. Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* 2020, <https://doi.org/10.1148/radiol.2020200432>. 200432.
7. BSTI. Thoracic imaging in COVID-19 infection. Guidance for the Reporting Radiologist British Society of thoracic imaging. Available at: https://www.bsti.org.uk/media/resources/files/BSTI_COVID-19_Radiology_Guidance_version_2_16.03.20.pdf. [Accessed 13 April 2020].
8. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;**44**(8):763–70.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74. <https://doi.org/10.2307/2529310>.