Chest X-Ray Has Poor Diagnostic Accuracy and Prognostic Significance in COVID-19: A Propensity Matched Database Study

Authors: Aditya Borakati (0000-0003-0457-4944), Adrian Perera (0000-0002-1753-2586), James Johnson (0000-0001-7796-659X), Tara Sood (0000-0003-1620-916X)

Emergency Department, Royal Free Hospital, Pond Street, London, UK NW3 2QG

Aditya Borakati, Academic Foundation Doctor, Division of Surgery and Interventional Science, Royal Free Campus, University College London

Adrian Perera, Emergency Medicine and Intensive Care Medicine Specialist Registrar, Emergency Department, Royal Free Hospital

James Johnson, Foundation Doctor, Emergency Department, Royal Free Hospital

Tara Sood, Consultant in Emergency Medicine, Emergency Department, Royal Free Hospital

Correspondence:

Dr Aditya Borakati

Academic Foundation Doctor

Emergency Department, Royal Free Hospital, London, UK NW3 2QG

a.borakati@doctors.org.uk

Author contribution (CRediT) statement:

Aditya Borakati: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project Administration

Adrian Perera: Conceptualization, Methodology, Investigation, Writing- Review & Editing, Supervision, Project Administration

James Johnson: Investigation

Tara Sood: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project Administration

Abstract

Objectives: To identify the diagnostic accuracy of common imaging modalities, chest X-ray (CXR) and computed tomography (CT) for diagnosis of COVID-19 in the general emergency population in the UK and to find the association between imaging features and outcomes in these patients.

Design: Retrospective analysis of electronic patient records

Setting: Tertiary academic health science centre and designated centre for high consequence infectious diseases in London, UK.

Participants: 1,198 patients who attended the emergency department with paired RT-PCR swabs for SARS-CoV 2 and CXR between 16th March and 16th April 2020

Main outcome measures: Sensitivity and specificity of CXR and CT for diagnosis of COVID-19 using the British Society of Thoracic Imaging reporting templates. Reference standard was any reverse transcriptase polymerase chain reaction (RT-PCR) positive naso-oropharyngeal swab within 30 days of attendance. Odds ratios of CXR in association with vital signs, laboratory values and 30-day outcomes were calculated.

Results: Sensitivity and specificity of CXR for COVID-19 diagnosis were 0.56 (95% CI 0.51-0.60) and 0.60 (95% CI 0.54-0.65), respectively. For CT scans these were 0.85 (95% CI 0.79-0.90) and 0.50 (95% CI 0.41-0.60), respectively. This gave a statistically significant mean increase in sensitivity with CT compared with CXR, of 29% (95% CI 19%-38%, $p<0.0001$). Specificity was not significantly different between the two modalities.

Chest X-ray findings were not statistically significantly or clinical meaningfully associated with vital signs, laboratory parameters or 30-day outcomes.

Conclusions: Computed tomography has substantially improved diagnostic performance over CXR in COVID-19. CT should be strongly considered in the initial assessment for suspected COVID-19. This gives potential for increased sensitivity and considerably faster turnaround time, where capacity allows and balanced against excess radiation exposure risk.

Key words: X-Rays, Computed Tomography, COVID-19, severe acute respiratory syndrome coronavirus 2, Emergency Medicine, Diagnostic Imaging

Statistical review: The statistical methods in this manuscript and associated code have been reviewed by Dr Federico Ricciardi of the Department of Statistical Science at University College London and confirmed as robust and accurate.

Ethical approval: This study was registered with the local institutional review board as a service evaluation using anonymised data only. No formal ethics committee review was required.

Data availability: Anonymised data is available on reasonable request from the corresponding author. Analysis scripts are attached as a supplementary file.

Declarations: The authors have no relevant conflicts of interest to declare.

Funding: No funding was received for this study.

Introduction

SARS-CoV 2 and its resulting disease, COVID-19, have propagated exponentially worldwide, with over 10 million cases in 188 countries at the time of writing [1,2].

The gold standard for diagnosis of the virus is the detection of viral RNA through reverse transcriptase polymerase chain reaction (RT-PCR) of respiratory tract samples. However, this method has several limitations including: (1) low sensitivity at 59-71% [3,4], (2) relatively slow turnaround times ranging from a few hours to several days [5], (3) high expense and (4) limited capacity for testing in many countries.

Computed tomography (CT) has been shown to be more sensitive than RT-PCR for diagnosis of COVID-19 [3,4], while being significantly faster and cheaper. This comes with a large radiation dose and capacity is still lacking in many countries.

Plain film chest X-ray (CXR) is ubiquitous worldwide, with a 30-70x lower dose of radiation[6] and is commonly performed as an initial investigation in COVID-19.

Studies have so far only evaluated imaging in those with confirmed infection, it is therefore, not possible to calculate the specificity of these modalities. In the context of the global pandemic, infection may be widespread in the community, often with subclinical infection [7,8]. A reliable and rapid method to detect infection in the general population, who may present to medical personnel with other complaints, is needed.

Despite its extensive use, the specificity and sensitivity of CXR in the general emergency population for diagnosis of COVID-19 is unknown, nor how imaging features correlate with severity.

This study evaluated the performance of CXR in diagnosing COVID-19 in the emergency department (ED) of a tertiary care hospital.

Methods

This study was conducted at the Royal Free Hospital, London, UK, an academic health science centre and nationally designated centre for High Consequence Infectious Diseases [9].

All individuals attending the emergency department who had paired plain film chest radiographs and RT-PCR nasopharyngeal swabs for COVID-19 at initial attendance between 16th March 2020 and 16th April 2020 were included.

All chest radiographs were reported by a Consultant Radiologist and rated on an ordinal scale for probability of COVID-19: Alternative pathology identified, not COVID-19; Clear chest, unlikely COVID; Indeterminate findings for COVID-19; Classical findings of COVID-19, based on the British Society of Thoracic Imaging's (BSTI) reporting templates (table 1) [10].

RT-PCR of swabs were performed in laboratories either at our centre or at a public health laboratory (PHE Collindale, UK), according to published national standard operating procedures [11]. Subsequent RT-PCR swabs taken within 30 days of initial ED attendance were also included.

CT scans performed within 30 days of attendance were retrieved. These were also reported according to the BSTI template. CT pulmonary angiogram was performed in the ED if the D-dimer was >5000 to exclude pulmonary emboli as per the locally agreed protocol. Subsequent CT chest imaging (whether pulmonary angiogram, contrast or non-contrast) was performed on the basis of clinical suspicion.

Prospectively recorded data was extracted from the Cerner Millennium electronic patient record system (Cerner Corp., Kansas City, MO).

Primary Outcome

The primary outcome is sensitivity and specificity of initial CXR, where it is reported as having classic COVID-19 features in the ED. This is compared with RT-PCR swab as the reference standard for diagnosis of COVID-19.

In the event of multiple RT-PCR swabs during one attendance, a single positive swab was taken as an overall positive test during one admission.

Secondary Outcomes

In those patients who also had CT scans of the thorax, the diagnostic accuracy was compared with CXR, with RT-PCR again as the reference standard. Sensitivity and specificity of CXR when X-rays reported as indeterminate or atypical for COVID-19 were classed as positive was also calculated.

Chest x-ray findings were correlated with vital signs at attendance and blood results, including: neutrophil counts, D-dimer and C-reactive protein, which have been associated with poor prognosis in COVID-19 [12]. Hazard ratios for clinical outcomes including direct admission to the intensive treatment unit (ITU) from ED and 30-day mortality rates were also calculated for CXR reporting categories.

Statistical Analysis

In the event of missing data, multiple imputation was conducted using a Predictive Mean Matching algorithm, via the MICE R package, as described previously [13]. Briefly, this uses a linear regression model (or logistic regression model for categoric data), to find a random value based on already observed data, to replace missing fields [14]. Variables without missing data fields were not modified. The number of imputed datasets was similar in number to the percentage of missing data as suggested by White and colleagues [15]. Balance diagnostics with density plots are available in supplementary file 1, adequate balance was assessed via visual inspection of imputed distributions with respect to the original dataset.

The propensity for a CXR being reported as positive or negative for COVID-19 was calculated for several plausible covariates that may influence image characteristics such as Age, Gender, Ethnicity, pre-existing morbidities and the respiratory rate of the patient using a generalised linear model [16]. X-ray positive and negative groups were then matched in each imputed dataset using the nearest neighbour algorithm, with a calliper of 0.2 of the propensity score standard deviation, without replacement and in random sequential order to obtain a 1:1 match as described elsewhere [17].

The balance of the match data was assessed quantitatively with mean differences of covariates in each of the X-ray groups pre- and post-matching, with a difference of less than 0.1% considered a good match (supplementary figure 2). Visual inspection of matches was also conducted to ensure balance (supplementary figures 2, 3 and 4).

After matching, outcome data were adjusted for covariates including age, gender, ethnicity and presence of co-morbidities as well as C-reactive protein, D-dimer, troponin and vital signs. This was achieved by generalised linear regression for continuous outcome data, binomial logistic regression for binary categoric outcomes, or ordinal logistic regression in the case of CXR where it is the outcome variable.

These regression models were run on each imputed dataset and outcomes were pooled together across each imputed data set according to Rubin's rules [18] to give an overall estimate.

Diagnostic Accuracy Statistics

Chest X-rays reported as classical for COVID-19 as per the BSTI guidelines were considered a positive test in the primary analysis. In a secondary analysis X-rays reported as 'Indeterminate' or 'Atypical' for COVID-19 were also considered positive. All other reports were classified as a negative test. These were compared to nasopharyngeal aspirate RT-PCR results, which were taken as the gold standard for diagnosis of COVID-19. Where more than one swab was taken during the study period (up to 30 days after initial attendance), a single positive result was taken as a positive result for calculation of diagnostic accuracy statistics.

Sensitivity, specificity, predictive values and diagnostic accuracy were calculated using the propensity matched data after imputation and pooled across imputed datasets with 95% confidence intervals. Apparent and true prevalence based on this dataset are also given for interpretation of the predictive values.

Chest CTs were also reported according to the BSTI guidelines as with X-ray. Diagnostic statistics were calculated on raw, unmatched and non-imputed data (due to a low volume of

data for imputation and matching) in the same manner as X-ray. Mean differences and 95% confidence intervals between CT and X-ray for each of the diagnostic statistics are given, with a p-value calculated from the confidence intervals.

Agreement between the modalities was assessed on the unmatched dataset, in the sample where CT, CXR and RT-PCR were all available using Cohen's (for two group agreement) and Fleiss' Kappa (when all 3 are compared).

Data Presentation

Descriptive statistics are given as means and standard deviations for normally distributed data and as medians and interquartile ranges for non-normally distributed data, before and after matching and multiple imputation (for the latter these statistics are pooled across imputations).

Association of explanatory variables with SARS-CoV 2 and Chest X-ray findings are given as odds ratios in uni- and multi-variate configurations.

Data was considered statistically significant if $p < 0.05$. Given the large number of analyses in this paper, data is separately highlighted if $p<0.001$ as a secondary threshold to address the potential for false positives with multiple testing.

Analyses were conducted using R 4.0.0 (R Foundation for Statistical Computing, Vienna, Austria) and code for the analyses is given in supplementary file 2.

Sample size calculation

In this study, the lower confidence interval for sensitivity of CXR as reported by Wong et al.[19] (56%) was used as an estimate of likely sensitivity for COVID-19 . A power of 80% at an alpha of 0.05 was used to calculate the sample size for sensitivities and specificities of 56%. This gave an estimated sample size of 165 in each of the COVID-19 negative and positive groups by RT-PCR (total 330).

Ethical approval

This study was registered with the local institutional review board as a service evaluation using anonymised data only. No formal ethics committee review was required.

Reporting Guidelines

This study is reported according to the STARD guidelines [20] for diagnostic accuracy studies.

Results

1,198 eligible patients were identified in the study period. Their characteristics, stratified by positivity for SARS-CoV 2 infection by RT-PCR is summarized in table 2. This showed that those with confirmed SARS-CoV 2 infection were more likely to be male, older (mean 66.2 vs 62.7), have lower saturations, higher respiratory rates, whilst being more likely to be admitted and die within 30 days. There was a signification association with X-ray images and SARS-CoV 2 at baseline, with 59.6% having classic imaging features of COVID-19 in those with positive swabs versus 39.1% in those with negative swabs. There was 8.6% missing data overall in the dataset when variables with >50% missing data were removed and 15 imputations were performed on these remaining variables only.

After multiple imputation for missing data and pooled propensity score matching for plausible covariates that may affect CXR reporting, there were 430 patients in each of the X-ray positive and X-ray negative groups, for a total of 860 patients. Adequate balance was achieved for relevant covariates with a mean difference of <0.1 between groups (supplementary table 2).

Computed tomography (CT) was performed in 302 patients with paired RT-PCR during the same time period, with a median serial interval of 4.5 days (inter quartile range 0-17) after the initial attendance in ED and of these 30.1% were within one day of attendance.

Diagnostic Accuracy

The pooled sensitivity and specificity of CXR was 0.56 (95% CI 0.51-0.60) and 0.60 (95% CI 0.54-0.65), respectively (table 4). This gave an overall diagnostic accuracy of 0.57 (95% CI 0.54-0.61) for CXR.

In comparison, sensitivity and specificity for CT was 0.85 (95% CI 0.79-0.90) and 0.50 (95% CI 0.41-0.60), respectively. This gave a statistically significant mean increase in sensitivity with CT compared with CXR by 29% (95% CI 19%-38%, p<0.0001). Specificity was not significantly different between the two modalities. Diagnostic accuracy and negative predictive values were also significantly increased with CT at 0.15 and 0.22, respectively, while the negative likelihood ratio was significantly decreased at -0.44. This shows that the post-test odds of being negative for SARS-CoV 2 by RT-PCR with a negative CT is significantly lower.

Taking X-rays reported as indeterminate as positive increased the sensitivity of CXR to 0.80 (95% CI 0.77-0.84), however reduced specificity to 0.40 (95% CI 0.35-0.46). When CT scans reported as indeterminate are also considered positive the sensitivity of CT increased to 0.93 (95% CI 0.89-0.96), whilst mean specificity reduced to 0.37 (95% CI 0.28-0.47), although this was not statistically different from when indeterminate CTs are considered negative. Sensitivity of CT remained significantly higher than CXR (when indeterminates are considered positive for both) by 0.13 (95% CI 0.05-0.19, p<0.001), specificity was not significantly different between the two.

When comparing only the unimputed, unmatched subset of data where CT, RT-PCR and CXR were all performed (n=287), the agreement between CT and CXR was poor (Cohen's kappa 0.406, p<0.0001). Agreement between all three was also poor (Fleiss' kappa 0.361, p=0).

Association of CXR with Markers of Severity and Outcomes

Association of covariates with RT-PCR results is shown in table 4 and figure 1. Those who tested positive for SARS-CoV 2 by RT-PCR were significantly more likely to have a classical X-ray (OR 1.79 95% CI 1.25-2.56, $p<0.002$) as would be expected by the diagnostic accuracy statistics (table 4). When the CXR report is considered as an ordered scale, worsening grades of report were associated more strongly with RT-PCR positivity, with a 1.94 x increase in odds for each grade.

Positive chest X-rays for COVID-19 were significantly associated with lower oxygen saturations (OR 0.94 95% CI 0.92-0.97, $p<0.001$) and temperatures (2.30 95% CI 1.46-3.63, $p<0.001$) in the ED following propensity score matching and multivariate regression (table 5 and figure 2).

They also had higher rates of admission to a general ward from the ED (OR 2.30 95% CI 1.46-3.63, $p<0.001$) but no significant association with 30 day outcomes. There was a statistically significant increase in C-reactive protein with a positive X-ray, however, this is unlikely to be clinically meaningful due to the minimal association (OR 1.00 95% CI 1.00-1.01).

Discussion

This study is the first to report the diagnostic accuracy of CXR and CT in the general emergency population during the COVID-19 pandemic.

We show that CXR has poor sensitivity and specificity for diagnosis of COVID-19, whilst CT has 29% higher sensitivity. Many international radiological guidelines advise against CT scanning for the initial assessment of COVID-19 [21–23] or where there are equivocal CXRs, whilst in other countries CT scanning is performed as a routine first line investigation. Our results suggest that CT should be considered in the initial assessment of COVID-19 and that CXR findings poorly correlate with CT findings in this setting. We also show that indeterminate and non-classical features of COVID-19 significantly increase the sensitivity of these imaging modalities, without a significant decrease in specificity. Further, we demonstrate the limited prognostic value of CXR in COVID-19.

These findings mirror what has previously been reported in the literature on individuals with confirmed COVID-19. Wong et al. [19] showed a sensitivity of 59% for initial X-ray in confirmed COVID-19 infection, similarly initial case series in China also reported a sensitivity of 59.1%[12].

A recent in press article from Italy reported a much higher sensitivity of 89% for CXR in a smaller general emergency population (n=535) without confirmed COVID-19 at attendance [24]. However, this used telephone follow up for clinical symptoms of COVID-19 as a reference standard in individuals with an initial negative RT-PCR swab and appeared to classify any abnormal X-ray as positive, which may inflate this figure. When indeterminate CXRs are counted as positive in this study, the sensitivity would be in line with this Italian data. In the US, a study of patients attending an urgent care centre with confirmed COVID-19, showed a much lower sensitivity at 41.7% for CXR where any abnormality was found on the images [25]. In this study 97/636 reports were re-classified from ‘possible pneumonia’ to ‘normal’ on second reading from a radiologist, highlighting the importance of inter-rater agreement and possibly explaining this low estimate.

Computed tomography has been reported in previous studies as being up to 98% sensitive for the diagnosis of COVID-19 in confirmed patients, when RT-PCR is used as the reference standard in confirmed patients [3,4]. These studies used any potential features of COVID-19 (e.g. ground glass opacification, crazy paving) as a positive scan, regardless of spatial distribution or features more characteristic of alternate pathology, unlike the BSTI guidelines used in this study. When we classified indeterminate CTs as positive like these latter studies, our estimates match their sensitivity values.

Consequently, a much lower specificity of 25% was found with initial RT-PCR in previous literature; however, it is reported that 10 out of 15 (67%) of these negatives subsequently tested positive. This would give an adjusted specificity of 75%, considering subsequent swabs as a reference standard, which combined with the wider CIs in these smaller studies, would bring estimates in line with the specificity in this paper. More recent meta-analyses have placed the pooled sensitivity of CT in populations with confirmed COVID-19 only, at 89.76% (95% CI 84.42%-93.84%) [26], in line with the estimates identified here.

There is limited coverage in the literature on association of X-ray findings with clinical and laboratory parameters and outcomes in the COVID-19 pandemic. This study demonstrates that classic appearances of COVID-19 were associated with initial lower saturations and lower

temperature. Volume opacification of the lung fields were not quantified as a surrogate of severity; however, the use of the BSTI grading templates does this somewhat. When the X-ray report is considered as a graded scale from low likelihood of COVID-19 and severity to high likelihood and severity of disease there was no significant difference in association with vital signs or laboratory parameters compared with when the X-ray report is merely considered as a binary positive and negative outcome for COVID-19.

Borghesi and colleagues have devised a grading system for severity in patients with confirmed SARS-CoV 2 infection [27]. They further found a significant difference in the severity of CXR by this scoring system in those who were discharged versus those who died [28]. However, no numerical difference was given in their study, only that there was a significant difference, so it is not possible to quantify the strength or clinical significance of this association.

Here, there were no relevant associations between CXR and laboratory values. This analysis also found no association with positive X-rays and 30 day outcomes after multivariate analyses, unlike Borghese et al. This is also in contrast to Guan et al. who found higher rates of ITU admission and death in those with positive imaging findings. However, these studies analysed only those with confirmed SARS-CoV 2 infection. The divergence observed in this study may be due to classifying those with 'Alternate pathology/ Indeterminate' or 'CVXC3/ CVXC2' as per the BSTI templates, negative for COVID-19 in these analyses. Other studies classified X-rays with any abnormality as a positive for COVID-19. These alternate distributions may still be reflective of underlying COVID-19 and we show significantly higher sensitivity for both CT and CXR when these are classed as positive. It may be that correlating indeterminate X-rays (in addition to classical images) with vitals, laboratory markers and 30 day outcomes would yield significant associations. However this may be unlikely, Xu and Zhang et al. found that those with classical bilateral and diffuse involvement in upper and lower lobes had more severe disease than those without [29,30].

There were a total of 70 confirmed pulmonary emboli in our dataset out of 114 CT pulmonary angiograms (61.0%, 5.84% of all patients attending) performed in the emergency department. The incidence of venous thromboembolism is reported as ranging from 20-30% in admitted confirmed SARS-CoV 2 positive patients [31]. Although we have not focused on this cohort of patients in this paper for the sake of brevity and simplicity, this high incidence represents a further advantage for CT over CXR.

CT, even with the absence of contrast has been shown to have strong accuracy in the diagnosis of pulmonary emboli and many imaging features correlate with the presence of pulmonary emboli. Sensitivities of non-contrast CT for diagnosis of PE have been reported at 96.9% and specificity at 71.9% [32,33].

We therefore see the advantages of CT scanning in COVID-19 as threefold over other diagnostic techniques: 1) The rapid turnaround; 2) Increased sensitivity and 3) The possibility to identify pulmonary emboli in COVID-19, which are a significant burden in this group.

This must be balanced against the excess radiation exposure with CT. Radiation from CT and its association with carcinogenesis is difficult to quantify and no definitive epidemiological studies have confirmed excess risk of cancer[34]. Modern CT scanners and software reconstruction techniques continue to minimise radiation exposure and many ways of shielding parts of the body from radiation also exist. Nevertheless, the excess risk of lifetime cancer is estimated at 1 per 5,000 CT examinations[35].

Strengths and Limitations

This study is the largest conducted on imaging in the COVID-19 pandemic and one of the only studies conducted in the general population during the pandemic rather than only in confirmed patients. This enables greater applicability to the clinical setting where the diagnosis is uncertain, in addition to being able to calculate specificity, which is not possible in most studies. This study was planned to be powered to detect a sensitivity and specificity of 56% for CXR and greatly exceeded the sample size necessary for this.

Comprehensive statistical analyses were conducted to account for confounders in both factors influencing reporting of CXR and in factors affecting outcomes. The data was collected from prospectively maintained electronic records; however, the retrieval took place retrospectively with its inherent disadvantages. We were not able to collect data on several relevant covariates such as specific comorbidities or markers of severity such as lymphocytes. Furthermore, there was a significant amount of missing data that required multiple imputation to replace, although the fit of this imputed data was good, actual, observed data would be ideal.

Inter-rater reliability of imaging reports was not analysed in this paper and there was the potential for individual radiologists to have greater or lesser accuracy in the diagnosis of COVID-19. The literature has so far suggested a strong degree of agreement between radiologists in reporting of COVID-19 images [28].

The single centre nature of this study further limits generalisability and the potential for inter-hospital disagreement in imaging, in addition to inter-rater disagreement.

Finally, the median time for patients to receive a CT scan was 4.5 days following initial attendance to ED. Thus, the scans may not have been directly comparable to the initial CXR, both because of the progression of disease and because the SARS-CoV 2 status may have been confirmed at this point, biasing the reporting of these scans.

Future Research

Although this study used RT-PCR of nasopharyngeal swabs as a reference standard, newer methods exist for diagnosis of the disease. Serological assays for antibodies against SARS-CoV 2 are increasingly available and may represent a better gold standard in diagnosis for future research [36]. RT-PCR is limited by swabbing technique for nasopharyngeal samples and the fact that the virus is more avid in the lower respiratory tract [37]. However, many patients may not seroconvert prior to death limiting this test to survivors only.

Point of care lung ultrasound is a new technique for diagnosis of COVID-19 which may mitigate many of the issues noted with the modalities discussed so far. It has no radiation, is fast, cheap and may be able to detect lower respiratory tract disease unlike nasopharyngeal swab. However, there is limited evidence beyond small case series on its diagnostic accuracy [38–40]. Further, like other ultrasound techniques accuracy will likely be operator dependent [41] and experience will need to be built up for robust results in evaluating suspected COVID-19.

Finally, much research has been conducted in the use of artificial intelligence techniques to correctly diagnose COVID-19 based on imaging [42–44]. These techniques would obviate capacity limitations in reporting imaging as well as eliminate inter-reporter variability. However, as with any supervised machine learning technique, large, generalisable datasets, with correctly

pre-classified positive and negative cases (which in turn will depend on a truly accurate reference standard) are needed [45].

Conclusion

Chest X-ray has poor sensitivity and specificity in diagnosing COVID-19 in the general population during the pandemic. CT scanning has demonstrated excellent sensitivity and should strongly be considered during the pandemic in the initial assessment of COVID-19. This needs to be balanced against the risk of excess radiation with CT, where capacity allows.

Summary box

What is already known on this topic

-Small observational studies, predominantly in China, have reported on imaging features in COVID-19 after a confirmed RT-PCR swab test

-These studies have shown limited sensitivity for chest X-ray, but excellent sensitivity for CT scans, it is not possible to calculate the specificity of these modalities as they only included patients with confirmed COVID-19, therefore it is not possible to assess their utility in the general population who may or may not have COVID-19

-Literature on this general population attending emergency departments and the accuracy of these imaging techniques is limited

-International guidelines including from the British Society of Thoracic Imaging and American College of Radiology do not recommend the use of CT in initial evaluation of suspected COVID-19, largely due to capacity concerns

What this study adds

-This study shows that Chest x-ray has poor sensitivity and specificity in patients with suspected COVID-19 attending the emergency department, whilst CT has excellent sensitivity and is 29% more sensitive than CXR in our study cohort; there was also poor agreement between CT and CXR findings in COVID-19

-Patients with indeterminate imaging without classical distribution of COVID-19 should still be considered at high risk of having the disease

-Our data suggest that CT should be employed more widely as an initial investigation, where capacity allows and balanced against the risk of excess radiation exposure

Acknowledgements

We would like to thank Scott Wilson from the Royal Free Hospital's clinical practice group analytics department for retrieving the data from the hospital's data warehouse.

We would like to thank Dr Federico Ricciardi of the Department of Statistical Science and PRIMENT Clinical Trials Unit at University College London for reviewing the statistical methods in this study.

Declarations of Interest

The authors declare no conflicts of interest.

References

1 COVID-19 Map. Johns Hopkins Coronavirus Resour. Cent. https://coronavirus.jhu.edu/map.html (accessed 30 Jun 2020).

2 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;**0**. doi:10/ggnsjk

3 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. doi:10/ggmw6p

4 Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology 2020;:200432. doi:10/ggnnkj

5 Konrad R, Eberle U, Dangel A, et al. Rapid establishment of laboratory diagnostics for the novel coronavirus SARS-CoV-2 in Bavaria, Germany, February 2020. Eurosurveillance 2020;**25**. doi:10/ggp6bw

6 Lin EC. Radiation Risk From Medical Imaging. Mayo Clin Proc 2010;**85**:1142–6. doi:10/c445mk

7 Mizumoto K, Kagaya K, Zarebski A, et al. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Eurosurveillance 2020;**25**:2000180. doi:10/ggn4bd

8 Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med 2020;**382**:2081–90. doi:10/ggszfg

9 High consequence infectious diseases (HCID). GOV.UK. https://www.gov.uk/guidance/high-consequence-infectious-diseases-hcid (accessed 24 May 2020).

10 Desai S. COVID-19 BSTI Reporting templates | The British Society of Thoracic Imaging. Br. Soc. Thorac. Imaging. 2020.https://www.bsti.org.uk/covid-19-resources/covid-19-bsti-reporting-templates/ (accessed 29 Apr 2020).

11 NHS England. Guidance and Standard Operating Procedure: COVID-19 virus testing in NHS Laboratories. 2020.https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/guidance-and-sop-covid-19-virus-testing-in-nhs-laboratories-v1.pdf (accessed 24 May 2020).

12 Guan W, Ni Z, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020;**0**:null. doi:10/ggm6dh

13 Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data. J Stat Softw 2011;**45**. doi:10/gdqc9c

14 Ginkel JR van, Linting M, Rippe RCA, et al. Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data. J Pers Assess 2020;**102**:297–308. doi:10/gftj5w

15 White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;**30**:377–99. doi:10.1002/sim.4067

16 He H, McDermott MP. A robust method using propensity score stratification for correcting verification bias for binary tests. Biostat Oxf Engl 2012;**13**:32–47. doi:10/c4jzn6

17 Ho DE, Imai K, King G, et al. MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. J Stat Softw 2011;**42**. doi:10/gdwtnq

18 Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009;**9**:57. doi:10.1186/1471-2288-9-57

19 Wong HYF, Lam HYS, Fong AH-T, et al. Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. Radiology 2020;:201160. doi:10/ggqbp4

20 Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;**6**:e012799. doi:10.1136/bmjopen-2016-012799

21 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 2020;:201365. doi:10/ggrmg4

22 ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection (accessed 5 Jun 2020).

23 British Society of Thoracic Imaging. COVID-19: BSTI STATEMENT AND GUIDANCE. 2020;:1.https://www.bsti.org.uk/media/resources/files/COVID11.3.20_2.pdf (accessed 5 Jun 2020).

24 Schiaffino S, Tritella S, Cozzi A, et al. Diagnostic Performance of Chest X-Ray for COVID-19 Pneumonia During the SARS-CoV-2 Pandemic in Lombardy, Italy. J Thorac Imaging 2020;**Publish Ahead of Print**. doi:10/ggx268

25 Weinstock MB, Echenique A, Russell JW, et al. Chest X-Ray Findings in 636 Ambulatory Patients with COVID-19 Presenting to an Urgent Care Center: A Normal Chest X-Ray Is no Guarantee. ;:10.

26 Bao C, Liu X, Zhang H, et al. Coronavirus Disease 2019 (COVID-19) CT Findings: A Systematic Review and Meta-analysis. J Am Coll Radiol 2020;**17**:701–9. doi:10/ggr28p

27 Borghesi A, Zigliani A, Masciullo R, et al. Radiographic severity index in COVID-19 pneumonia: relationship to age and sex in 783 Italian patients. Radiol Med (Torino) 2020;**125**:461–4. doi:10/ggtvwp

28 Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiol Med (Torino) 2020;**125**:509–13. doi:10/ggtvwn

29 Xu Y-H, Dong J-H, An W-M, et al. Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. J Infect 2020;**80**:394–400. doi:10/ggqwf3

30 Zhang J-J, Dong X, Cao Y-Y, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy Published Online First: 19 February 2020. doi:10/ggpx6g

31 Lodigiani C, Iapichino G, Carenzo L, et al. Venous and arterial thromboembolic complications in COVID-19 patients admitted to an academic hospital in Milan, Italy. Thromb Res 2020;**191**:9–14. doi:10/ggvcft

32 Chien C-H, Shih F-C, Chen C-Y, et al. Unenhanced multidetector computed tomography findings in acute central pulmonary embolism. BMC Med Imaging 2019;**19**:65. doi:10/ggzg85

33 Mohamed N, Othman MoustafaHM, Hassan L, et al. The accuracy of non-contrast chest computed tomographic Scan in the detection of pulmonary thromboembolism. J Curr Med Res Pract 2019;**4**:61. doi:10/ggzg83

34 McCollough CH, Bushberg JT, Fletcher JG, et al. Answers to Common Questions About the Use and Safety of CT Scans. Mayo Clin Proc 2015;**90**:1380–92. doi:10/f3jgqx

35 Moser JB, Sheard SL, Edyvean S, et al. Radiation dose-reduction strategies in thoracic CT. Clin Radiol 2017;**72**:407–20. doi:10/f95q7p

36 Long Q-X, Liu B-Z, Deng H-J, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nat Med 2020;:1–4. doi:10.1038/s41591-020-0897-1

37 Wang W, Xu Y, Gao R, et al. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA 2020;**323**:1843–4. doi:10/ggpp6h

38 Smith MJ, Hayward SA, Innes SM, et al. Point-of-care lung ultrasound in patients with COVID-19 – a narrative review. Anaesthesia;**n/a**. doi:10/ggr2p7

39 Haaksma ME, Heldeweg MLA, Matta JEL, et al. Lung ultrasound findings in patients with novel SARS-CoV2. medRxiv 2020;:2020.05.18.20105775. doi:10.1101/2020.05.18.20105775

40 Benchoufi M, Bokobza J, Chauvin AA, et al. Lung injury in patients with or suspected COVID-19 : a comparison between lung ultrasound and chest CT-scanner severity assessments, an observational study. medRxiv 2020;:2020.04.24.20069633. doi:10.1101/2020.04.24.20069633

41 Fine D, Perring S, Herbetko J, et al. Three-dimensional (3D) ultrasound imaging of the gallbladder and dilated biliary tree: reconstruction from real-time B-scans. Br J Radiol 1991;**64**:1056–7. doi:10/fqr9mh

42 Shi F, Wang J, Shi J, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation and Diagnosis for COVID-19. IEEE Rev Biomed Eng 2020;:1–1. doi:10/ggs2km

43 Li L, Qin L, Xu Z, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology 2020;:200905. doi:10/ggpdgp

44 Wang L, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. ArXiv200309871 Cs Eess Published Online First: 11 May 2020.http://arxiv.org/abs/2003.09871 (accessed 13 Jun 2020).

45 Kotsiantis SB. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artif Intell Rev 2012;**37**:331–44. doi:10/fmbng4

Tables

Ordinal scale for study	BSTI grade	Features on X-ray
0	CVCX3- Non-COVID-19	Alternative pathology such as pneumothorax with no features of COVID-19 identified
1	CVCX0- Normal	No pathology seen
2	CVCX2- Indeterminate for COVD-19 or atypical features	Poor quality film or central/ basal consolidation
3	CVCX1- Classic findings of COVID-19	Peripheral ground glass opacities

Table 1- Ordinal scale used in this study based on the British Society of Thoracic Imaging (BSTI) Reporting Template [10]

	SARS-CoV 2 RT-PCR		p-value	Missing (%)
	Negative	Positive		
n (%)	435 (36.3)	763 (63.7)		
Number of Swabs (%)	810 (48.3)	868 (51.7)		
Age (mean (SD))	62.74 (17.72)	66.18 (17.58)	**0.001***	0
Ethnicity			0.097	19
Other- Asian (%)	29 (8.0)	72 (11.8)		
South- Asian (%)	27 (7.5)	38 (6.2)		
Black (%)	41 (11.4)	91 (14.9)		
Mixed (%)	6 (1.7)	6 (1.0)		
Other (%)	56 (15.5)	105 (17.2)		
White (%)	202 (56.0)	297 (48.8)		
Sex – Male (%)	233 (53.6)	480 (62.9)	**0.002***	0
Oxygen Saturation (median (IQR))	95 (6)	93 (8)	**<0.001****	6.3
Respiratory Rate (median (IQR))	22 (8)	26 (12)	**<0.001****	6.3
Glasgow Coma Scale (median (IQR))	15 (0)	15 (0)	**0.043***	6.6
Systolic BP (median (IQR))	134 (32)	130 (30)	**0.009***	15.8
Heart Rate (median (IQR))	96 (27)	94 (27)	0.092	6.4
Temperature (median (IQR))	37.1 (1.4)	37.7 (1.4)	**<0.001****	6.7
Chest X-ray report			**<0.001****	0
Alternative pathology (%)	4 (0.9)	3 (0.4)		
No abnormalities (%)	178 (40.9)	136 (17.8)		
Indeterminate (%)	83 (19.1)	169 (22.1)		
Classic COVID-19 (%)	170 (39.1)	455 (59.6)		
Presence of comorbidities (%)	297 (79.0)	482 (80.3)	0.669	18.5
Dyspnoea (%)	274 (69.4)	497 (75.5)	0.034	12.1
Neutrophils (median (IQR))	6.42 (4.56)	5.25 (3.92)	**<0.001****	2.3
D-Dimer (median (IQR))	1250 (2440)	1105 (1803)	0.204	23.2
Albumin (median (IQR))	39 (7)	37 (6)	**<0.001****	10
C-Reactive Protein (median (IQR))	91.0 (115)	146.5 (264.8)	**<0.001****	3
Creatine Kinase (median (IQR))	51 (104)	145 (260)	**<0.001****	23.3
Troponin (median (IQR))	19 (46)	20 (44)	0.278	19.1
Admitted (%)	331 (76.0)	635 (83.2)	**0.003***	0.1
Admitted to ITU (%)	5 (1.3)	32 (4.8)	**0.005***	12.4
Thirty Day Follow Up Status			**<0.001****	24
Discharged (%)	219 (78.2)	367 (58.3)		
On Ambulatory Follow Up (%)	14 (5.0)	49 (7.8)		
Admitted (%)	18 (6.4)	60 (9.5)		
Died (%)	29 (10.4)	154 (24.4)		
CT report			**<0.001****	0
No pathology identified (%)	23 (22.1)	6 (3.3)		
Classic COVID-19 findings (%)	52 (50.0)	157 (85.8)		
Indeterminate for COVID-19 (%)	14 (13.5)	14 (7.7)		
Alternative pathology identified (%)	15 (14.4)	6 (3.3)		
Day of Symptoms (mean (SD))	9.84 (9.63)	8.56 (15.80)	0.368	69.2

Table 2- Baseline characteristics of dataset stratified by overall SARS-CoV 2 RT-PCR status, including subsequent swabs during the study period- NB there were 480 additional swabs on 399 unique patients with a median of 2 and mean of 3.5 per patient; *significant at p< 0.05; **significant at p< 0.001

	Chest X-ray	CT Chest	Mean Difference	p-value
Total (n)	860	302		
True Positives (n)	305	162	-	-
False Positives (n)	125	55	-	-
True Negatives (n)	187	56	-	-
False Negatives (n)	243	29	-	-
Apparent prevalence (95% CI)	0.50 (0.47-0.53)	0.72 (0.66-0.77)	**0.22 (0.04-0.21)**	**<0.0001****
True prevalence (95% CI)	0.64 (0.60-0.67)	0.63 (0.58-0.69)	-0.00 (-0.09-0.03)	0.111
Sensitivity (95% CI)	0.56 (0.51-0.60)	0.85 (0.79-0.90)	**0.29 (0.19-0.38)**	**<0.0001****
Specificity (95% CI)	0.60 (0.54-0.65)	0.50 (0.41-0.60)	-0.10 (-0.25-0.04)	0.119
Positive Predictive Value (95% CI)	0.71 (0.66-0.75)	0.75 (0.68-0.80)	0.04 (-0.06-0.14)	0.492
Negative Predictive Value (95% CI)	0.43 (0.39-0.48)	0.66 (0.55-0.76)	**0.22 (0.06-0.37)**	**0.005***
Positive Likelihood Ratio (95% CI)	1.39 (1.19-1.62)	1.71 (1.41- 2.08)	0.32 (-0.22-0.89)	0.258
Negative Likelihood Ratio (95% CI)	0.74 (0.64-0.84)	0.30 (0.21-0.44)	**-0.44 (-0.64 - -0.21)**	**0.022***
Diagnostic Accuracy (95% CI)	0.57 (0.54-0.61)	0.72 (0.66-0.77)	**0.15 (0.06-0.23)**	**<0.0001****

Table 3- Diagnostic Accuracy Metrics for CXR and CT Chest with RT-PCR for SARS-CoV 2, as the reference standard; *significant difference at the <0.05 level; **significant difference at the <0.0001 level

		SARS-CoV 2 RT-PCR		OR (univariable)	OR (multivariable)
		Negative	Positive		
n		312	548		
Chest X-ray report	Alternative pathology (%)	3 (0.8)	3 (0.5)	-	-
	No abnormalities (%)	123 (39.6)	104 (19.1)	0.76 (0.08-6.82, p=0.801)	0.48 (0.03-8.82, p=0.620)
	Indeterminate/ atypical findings (%)	61 (19.5)	136 (4.8)	1.99 (0.22-17.81, p=0.535)	0.92 (0.05-16.88, p=0.952)
	Classic COVID (%)	125 (40.1)	305 (55.6)	2.17 (0.24-19.19, p=0.484)	1.14 (0.06-20.98, p=0.927)
Age	Mean (SD)	61.8 (17.9)	67.0 (17.7)	**1.02 (1.01-1.02, p<0.001)****	**1.02 (1.00-1.03, p=0.028)***
Sex	Female (%)	138 (44.3)	212 (38.7)	-	-
	Male (%)	174 (55.7)	336 (61.3)	1.26 (0.93-1.70, p=0.137)	1.19 (0.83-1.71, p=0.340)
Ethnicity	Other Asian (%)	31 (9.9)	66 (12.0)	-	
	White (%)	164 (52.7)	270 (49.2)	0.76 (0.44-1.31, p=0.326)	0.73 (0.38-1.40, p=0.339)
	Black (%)	39 (12.4)	84 (15.3)	1.01 (0.52-1.98, p=0.974)	0.92 (0.43-1.97, p=0.827)
	Mixed (%)	6 (1.8)	4 (0.8)	0.36 (0.08-1.62, p=0.184)	0.74 (0.11-4.94, p=0.754)
	South Asian (%)	22 (7.0)	36 (6.6)	0.77 (0.34-1.76, p=0.531)	0.68 (0.28-1.65, p=0.390)
	Other (%)	51 (16.2)	89 (16.2)	0.82 (0.43-1.55, p=0.535)	0.88 (0.45-1.74, p=0.716)
Comorbidity	No (%)	65 (20.8)	95 (17.4)	-	-
	Yes (%)	247 (79.2)	453 (82.6)	1.25 (0.82-1.89, p=0.296)	1.00 (0.53-1.88, p=0.993)
Dyspnoea on attendance	No (%)	90 (28.8)	139 (25.4)	-	-
	Yes (%)	222 (71.2)	409 (74.6)	1.19 (0.82-1.73, p=0.356)	0.84 (0.53-1.32, p=0.447)
Oxygen Saturation	Median (IQR)	96 (6)	93 (8)	**0.94 (0.91-0.97, p<0.001****	0.97 (0.93-1.00, p=0.072)
Respiratory rate	Median (IQR)	23 (8)	25 (8)	**1.04 (1.01-1.07, p=0.002)***	1.01 (0.98-1.05, p=0.462)
Glasgow Coma Scale	Median (IQR)	15 (0)	15 (0)	1.02 (0.89-1.17, p=0.819)	1.21 (0.98-1.48, p=0.073)
Temperature	Mean (SD)	37.2 (1.4)	37.7 (1.1)	**1.48 (1.26-1.73, p<0.001)****	**1.44 (1.20-1.74, p<0.001)****
Heart Rate	Mean (SD)	96.7 (20.5)	94.9 (21.5)	1.00 (0.99-1.00, p=0.305)	1.00 (0.99-1.01, p=0.702)
Systolic Blood Pressure	Mean (SD)	136.2 (25.8)	132.6 (24.5)	0.99 (0.99-1.00, p=0.086)	0.99 (0.98-1.00, p=0.097)
Neutrophils	Median (IQR)	6.26 (4.52)	5.05 (3.93)	**0.92 (0.89-0.96, p<0.001)****	**0.87 (0.82-0.91, p<0.001)****
D-Dimer	Median (IQR)	1220 (2343)	1061 (1814)	1.00 (1.00-1.00, p=0.403)	1.00 (1.00-1.00, p=0.419)
C-Reactive Protein	Median (IQR)	45 (100)	77 (107)	**1.00 (1.00-1.01, p<0.001)****	**1.00 (1.00-1.01, p=0.021)***
Troponin	Median (IQR)	20 (55)	21 (46)	1.00 (1.00-1.00, p=0.890)	1.00 (1.00-1.00, p=0.667)
Albumin	Median (IQR)	39 (7)	37 (6)	0.97 (0.94-1.00, p=0.071)	1.02 (0.98-1.06, p=0.432)
Creatine Kinase	Median (IQR)	94 (131)	145 (263)	1.00 (1.00-1.00, p=0.119)	1.00 (1.00-1.00, p=0.152)
Admitted from ED	Admitted (%)	235 (75.2)	453 (82.7)	-	-
	Discharged (%)	77 (24.8)	95 (17.3)	**1.56 (1.06 -2.33, p=0.022)****	1.35 (0.79-2.30, p=0.272)
Admitted To ITU from ED	No (%)	307 (98.5)	532 (97.1)	-	-
	Yes (%)	5 (1.5)	16 (2.9)	1.92 (0.60-6.18, p=0.274)	1.06 (0.25-4.40, p=0.940)
Thirty Day Follow up Status	Discharged (%)	259 (83.0)	368 (67.1)	-	-
	Admitted (%)	22 (6.9)	47 (8.5)	1.53 (0.82-2.87, p=0.181)	1.64 (0.77-3.51, p=0.198)
	Dead (%)	31 (10.1)	133 (24.4)	**3.00 (1.86-4.84, p<0.001)****	**2.81 (1.22-6.50, p=0.017)***

Table 4- Association of covariates with RT-PCR status for SARS-CoV 2, following propensity score matching and binomial logistic regression; SD- Standard deviation; IQR- Interquartile Range; *p<0.05; **p<0.001

		X-ray report		OR (univariable)	OR with XR as binary outcome (multivariable)	OR with XR as ordinal variable (multivariable)
		Other X-ray Findings	Classical COVID-19			
n		430	430			
RT-PCR for SARS-CoV 2	Negative (%)	187 (43.4)	125 (29.1)	-	-	-
	Positive (%)	243 (56.6)	305 (70.9)	**1.85 (1.36-2.56, p<0.001)****	**1.79 (1.25-2.56, p<0.002)***	**1.94 (1.37-2.76, p<0.001)****
Age	Mean (SD)	65.0 (18.9)	65.3 (16.9)	1.00 (0.99-1.01, p=0.849)	0.99 (0.98-1.00, p=0.164)	1.00 (0.99-1.01, p=0.542)
Sex	Female (%)	176 (40.9)	175 (40.6)	-	-	
	Male (%)	254 (59.1)	255 (59.3)	1.01 (0.75-1.37, p=0.940)	0.87 (0.63-1.20, p=0.400)	1.02 (0.49-2.09, p=0.967)
Ethnicity	Other Asian (%)	49 (11.4)	48 (11.2)	-	-	
	South Asian (%)	29 (6.7)	29 (6.7)	1.04 (0.52-2.04, p=0.912)	1.02 (0.47-2.17, p=0.965)	1.02 (0.49-2.09, p=0.967)
	Black (%)	61 (14.2)	61 (14.2)	1.02 (0.55-1.85, p=0.957)	0.88 (0.46-1.69, p=0.719)	0.92 (0.52-1.65, p=0.789)
	Mixed (%)	5 (1.2)	5 (1.2)	0.92 (0.21-4.00, p=0.911)	0.86 (0.18-4.17, p=0.853)	0.85 (0.17-4.30, p=0.838)
	Other (%)	70 (16.3)	70 (16.3)	1.02 (0.58-1.79, p=0.943)	0.98 (0.52-1.82, p=0.942)	0.93 (0.53-1.64, p=0.810)
	White (%)	216 (50.2)	217 (50.5)	1.03 (0.63-1.67, p=0.913)	0.97 (0.57-1.67, p=0.926)	0.90 (0.55-1.47, p=0.666)
Comorbidity	No (%)	82 (19.1)	78 (18.1)	-	-	
	Yes (%)	348 (80.9)	352 (81.9)	0.95 (0.66-1.36, p=0.777)	0.93 (0.59-1.49, p=0.782)	0.88 (0.57-1.37, p=0.592)
Dyspnoea	No (%)	191 (29.3)	103 (24.0)	-	-	
	Yes (%)	304 (70.7)	327 (76.0)	1.31 (0.92-1.88, p=0.123)	1.20 (0.80-1.82, p=0.380)	1.22 (0.83-1.80, p=0.301)
Oxygen Saturation	Median (IQR)	95 (7)	93 (7)	**0.94 (0.91-0.96, p<0.001)****	**0.94 (0.92-0.97, p<0.001)****	**0.94 (0.91-0.97, p<0.001)****
Respiratory rate	Median (IQR)	24 (10)	24 (10)	1.01 (0.99-1.02, p=0.570)	0.97 (0.94-1.00, p=0.063)	0.98 (0.96-1.01, p=0.157)
Glasgow Coma Scale	Median (IQR)	15 (0)	15 (0)	1.04 (0.92-1.19, p=0.524)	1.05 (0.90-1.23, p=0.503)	1.05 (0.92-1.21, p=0.464)
Temperature	Mean (SD)	37.6 (1.1)	37.5 (1.3)	0.93 (0.83-1.06, p=0.297)	**0.79 (0.67-0.93, p=0.006)***	**0.85 (0.73-0.99, p=0.031)***
Heart Rate	Mean (SD)	95.7 (21.4)	95.5 (21.0)	1.00 (0.99-1.01, p=0.888)	1.00 (0.99-1.01, p=0.864)	1.00 (0.99-1.01, p=0.872)
Systolic Blood Pressure	Mean (SD)	133.8 (25.0)	134.0 (25.6)	1.00 (0.99-1.01, p=0.907)	1.00 (0.99-1.01, p=0.335)	1.00 (1.00-1.01, p=0.478)
Neutrophils	Median (IQR)	5.44 (4.54)	5.67 (4.03)	1.00 (0.97-1.04, p=0.892)	0.96 (0.92-1.01, p=0.143)	0.96 (0.92-1.01, p=0.115)
D-Dimer	Median (IQR)	1119 (2221)	1119 (1850)	1.00 (1.00-1.00, p=0.513)	1.00 (1.00-1.00, p=0.568)	1.00 (1.00-1.00, p=0.385)
C-Reactive Protein	Median (IQR)	46 (93)	88 (110)	**1.00 (0.99-1.00, p<0.001)****	**1.00 (1.00-1.01, p<0.001)****	**1.00 (1.00-1.01, p<0.001)****
Troponin	Median (IQR)	23 (54)	20 (46)	1.00 (1.00-1.00, p=0.231)	1.00 (1.00-1.00, p=0.277)	1.00 (1.00-1.00, p=0.059)
Albumin	Median (IQR)	39 (7)	37 (6)	**0.93 (0.90-0.96, p<0.001)****	**0.93 (0.90-0.97, p=0.001)***	**0.94 (0.91-0.97, p=0.001)***
Creatine Kinase	Median (IQR)	110 (183)	134 (239)	1.00 (1.00-1.00, p=0.535)	1.00 (1.00-1.00, p=0.242)	1.00 (1.00-1.00, p=0.186)
Admitted from ED	Admitted (%)	315 (73.3)	373 (86.7)	**2.37 (1.63-3.46, p<0.001)****	**2.30 (1.46-3.63, p<0.001)****	**2.22 (1.47-3.33, p<0.001)****
	Discharged (%)	115 (26.7)	57 (13.3)	-	-	-
Admitted to ITU from ED	No (%)	423 (98.4)	416 (96.7)	-	-	
	Yes (%)	7 (1.6)	14 (3.3)	2.17 (0.69-6.67, p=0.181)	1.27 (0.32-5.00, p=0.732)	1.34 (0.36-5.00, p=0.653)
30 Day Follow Up Status	Discharged (%)	316 (73.5)	311 (72.3)	-	-	
	Admitted (%)	34 (7.9)	34 (7.9)	1.31 (0.81-2.13, p=0.282)	1.32 (0.69-2.53, p=0.392)	1.43 (0.78-2.63, p=0.653)
	Dead (%)	80 (18.6)	85 (19.8)	1.03 (0.73-1.45, p=0.886)	1.38 (0.80-2.37, p=0.247)	1.41 (0.87-2.27, p=0.157)

Table 5- Association of covariates with CXR report following propensity score matching and either binomial or ordinal logistic regression; SD- Standard deviation; IQR- Interquartile Range; *p<0.05; **p<0.001

Figures

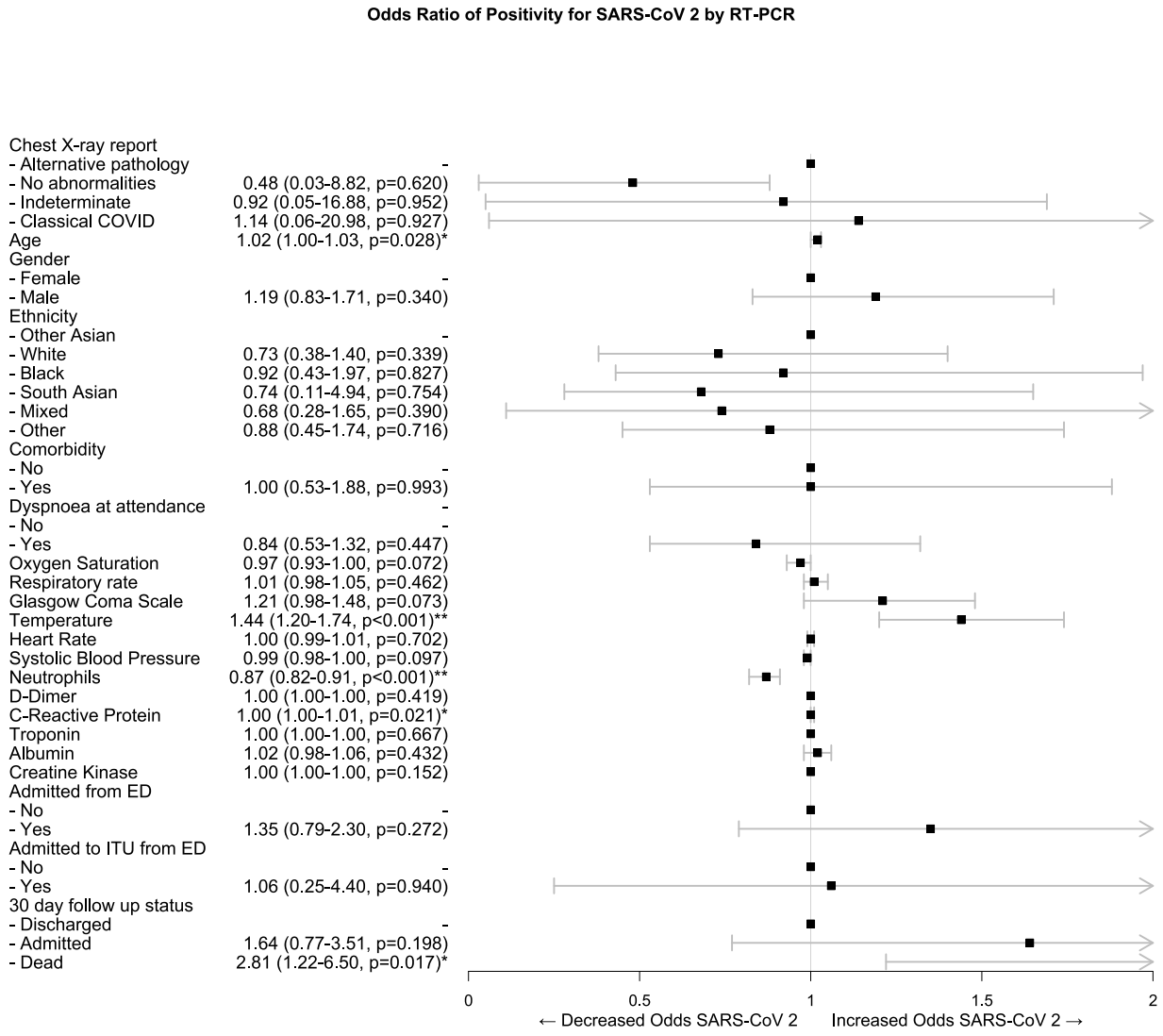


Figure 1- Forest plot of odds ratios of variables associated with RT-PCR positivity for SARS-CoV 2, following multiple imputation, propensity score matching and binomial logistic regression; *significant difference at the <0.05 level; **significant difference at the <0.001 level

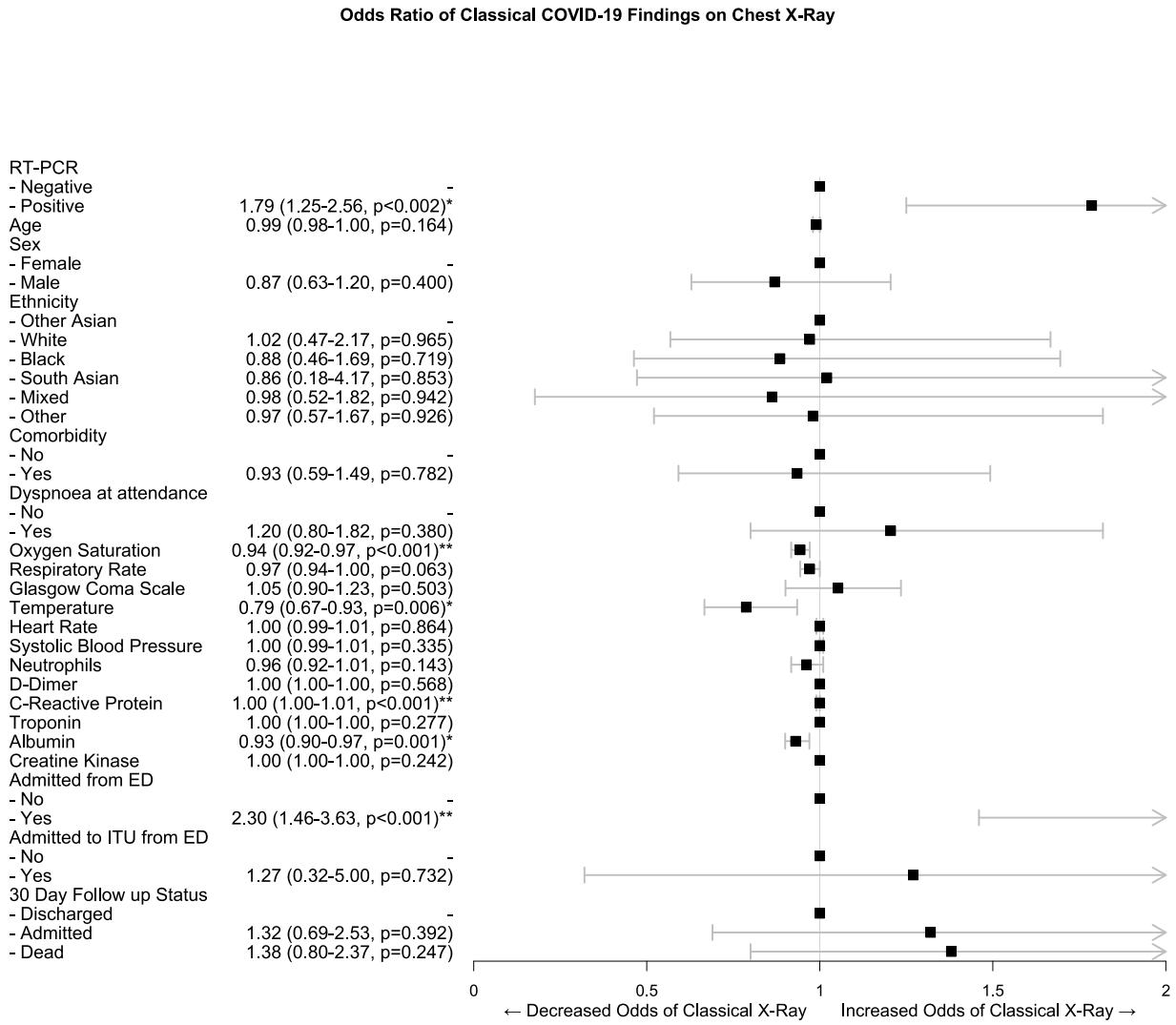
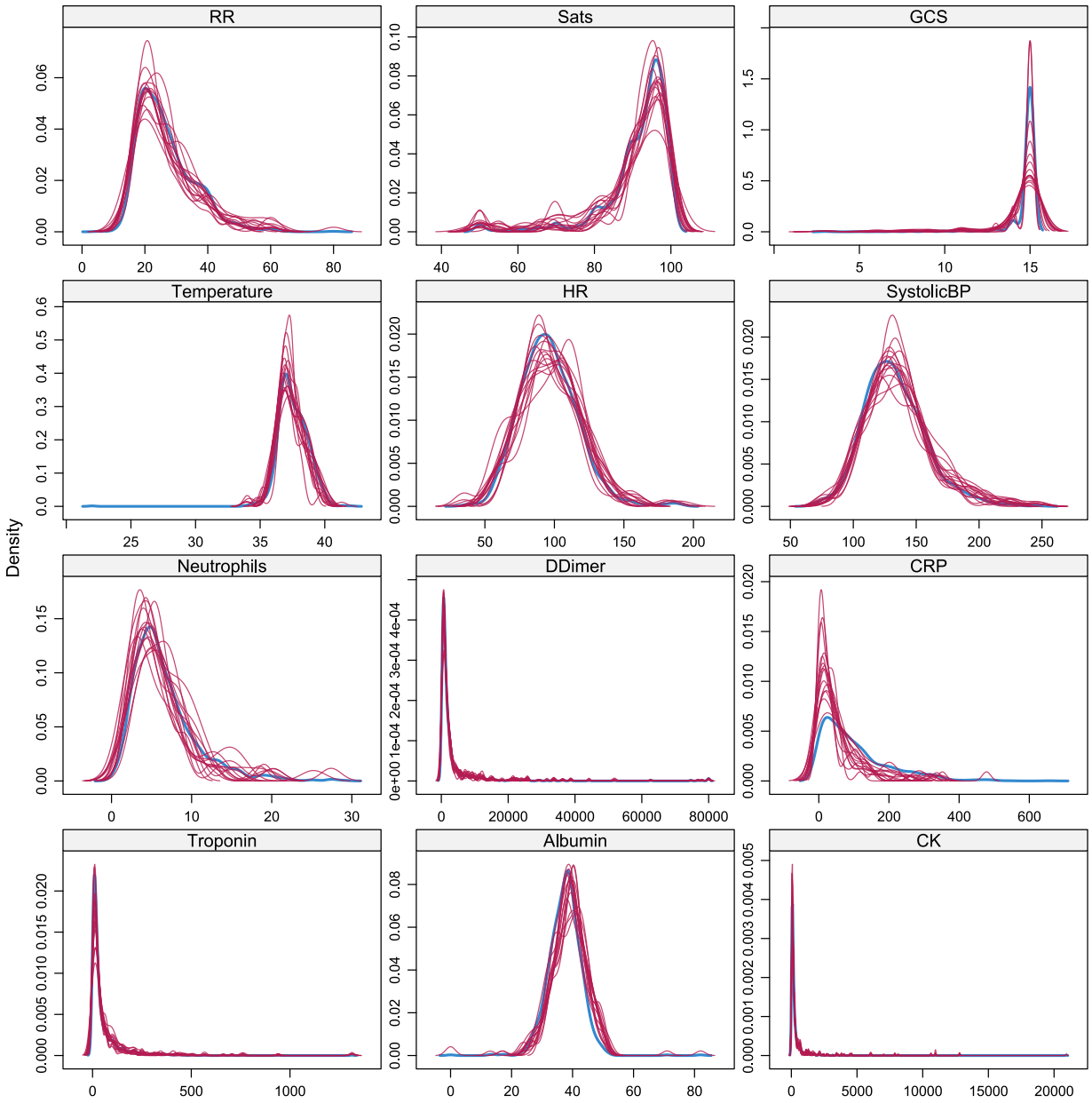


Figure 2- Forest plot of odds ratios of variables associated with classical Chest X-ray features COVID-19 following propensity score matching and binomial logistic regression; *significant difference at the <0.05 level; **significant difference at the <0.001 level

Supplementary file 1



Supplementary figure 1- Density plots of imputed datasets; Blue represents original dataset; other colours are individual imputed datasets (n=15)

Covariate:	Means Treated	Means Control	Standard Deviation Control	Mean Difference
Overall Propensity Score	0.422997940	0.53935303	0.1449627	-0.1163550897
Female	36.3782051	45.026178	0.4979547	-8.64797288
Male	63.6217949	54.973822	0.4979547	8.64797288
Age	63.796474359	66.19022688	18.5893357	-23.937525171
Comorbidity- Yes	76.1217949	84.467714	0.3625287	-8.34591892
Ethnicity- South Asian	6.5705128	6.631763	0.2490539	-0.06124983
Ethnicity- Black	16.1858974	11.518325	0.3195219	4.66757283
Ethnicity- Mixed	0.9615385	1.396161	0.1174340	-0.43462210
Ethnicity- Other	18.9102564	13.263525	0.3394765	5.64673110
Ethnicity- White	46.6346154	57.766143	0.4943635	-11.13152772
Respiratory Rate	29.214743590	24.01745201	7.2639816	5.1972915828

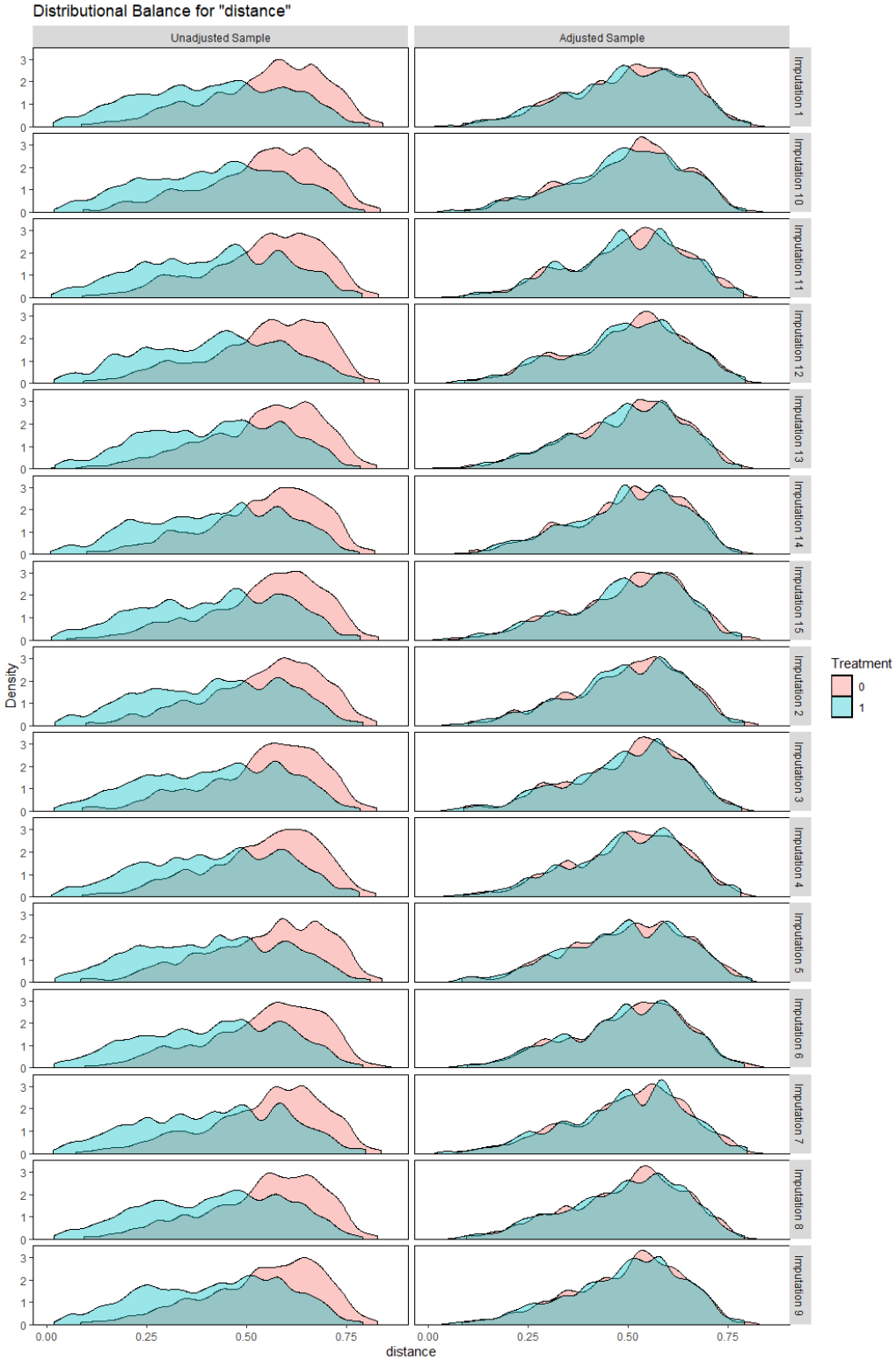
Supplementary table 1- Means of data before multiple imputation and propensity score matching

	Type	Minimum Difference Adjusted	Mean Difference Adjusted	Maximum Difference Adjusted
Distance	Distance	0.016988	0.027107	0.040963
Sex = Male	Binary	-0.03917	-0.0028	0.015982
Age	Contin.	-0.04586	-0.01371	0.027589
Comorbidity = Yes	Binary	-0.02331	-0.00778	0.004598
Ethnicity = Other Asian	Binary	-0.01392	0.002362	0.016471
Ethnicity = South Asian	Binary	-0.01399	-0.00136	0.011905
Ethnicity = Black	Binary	-0.01852	0.000443	0.015982
Ethnicity = Mixed	Binary	-0.00464	0.001403	0.007042
Ethnicity = Other	Binary	-0.01152	4.30E-06	0.00939
Ethnicity = White	Binary	-0.02353	-0.00285	0.018433
Respiratory Rate	Contin.	-0.06157	-0.03478	-0.00442

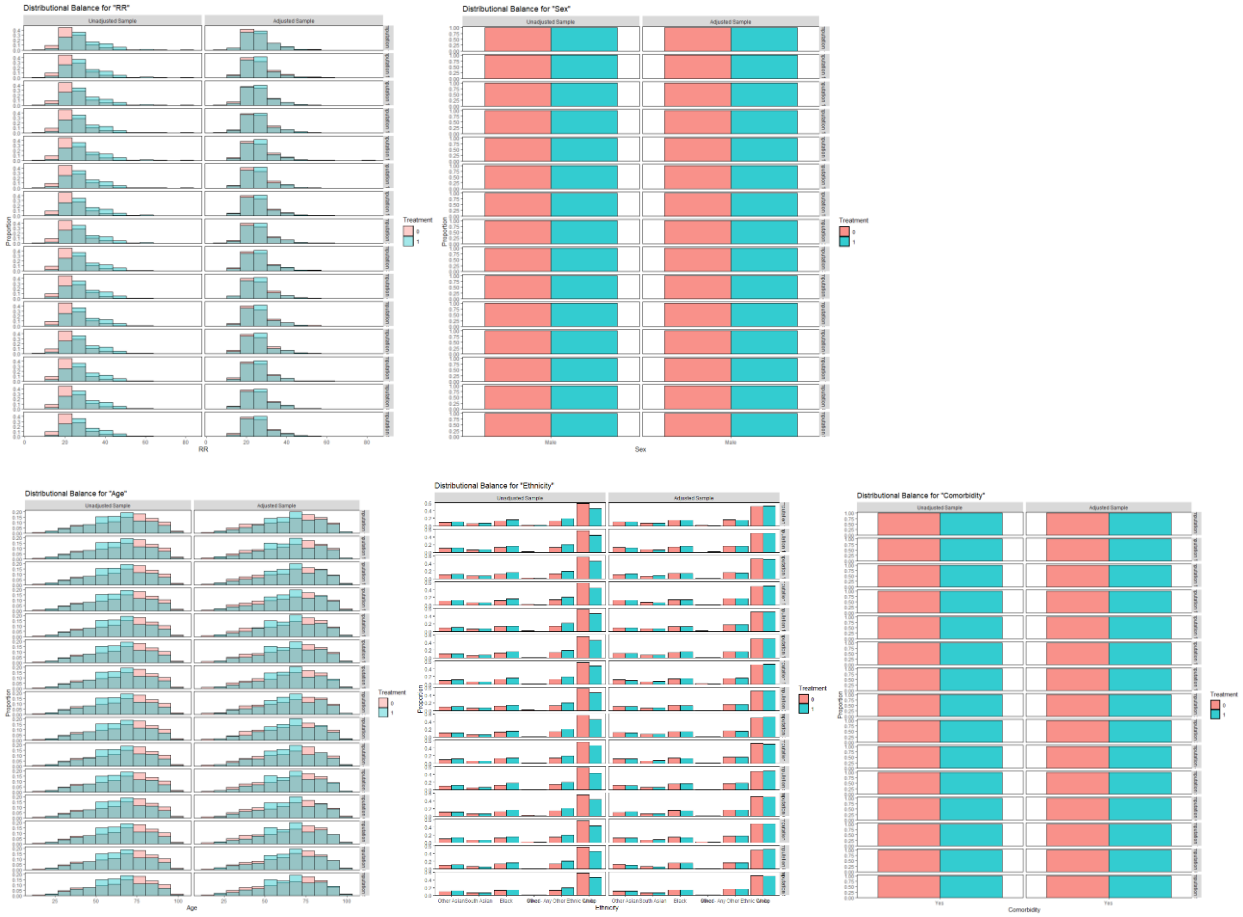
Supplementary table 2- Balance summary across imputations

	XR- Negative	XR- Positive	Total
All	573	625	1,198
Matched	430	430	860
Unmatched	143	195	338
Discarded	0	0	0

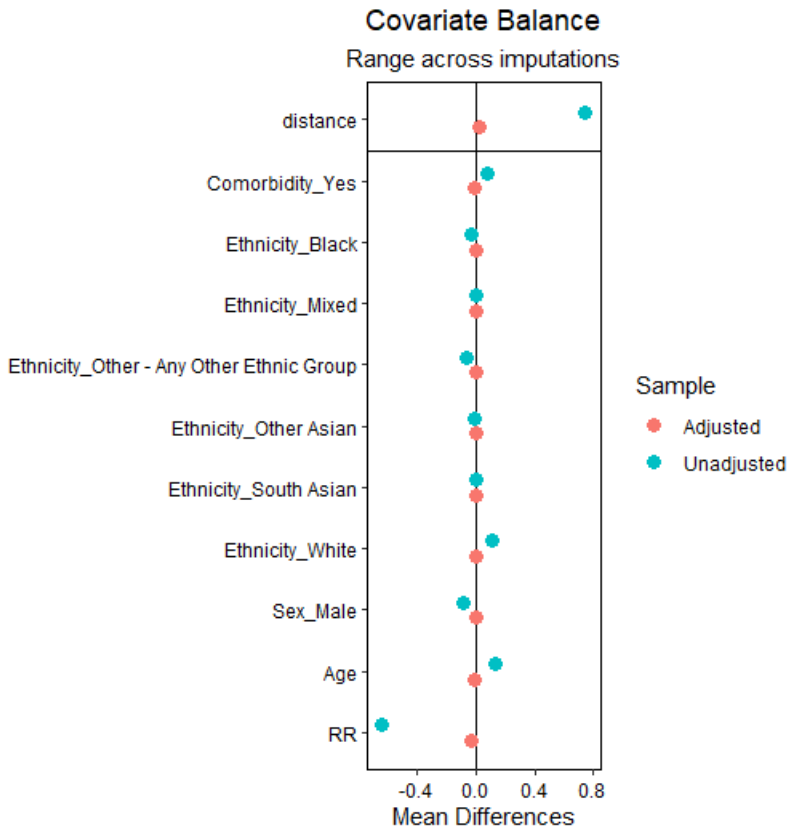
Supplementary table 3- Average Sample sizes pre- and post- matching across imputed data sets



Supplementary figure 2- Density plot of propensity scores pre- and post- matching in each imputed dataset; treatment units represent a positive X-ray for COVID-19, whereas a control unit represents a negative X-ray



Supplementary figure 3- Histogram of distributions for each matching covariate pre- and post- matching in each imputed dataset; treatment units represent a positive X-ray for COVID-19, whereas a control unit represents a negative X-ray



Supplementary figure 4- Love plot of pooled balances across imputed datasets in matching covariates after matching

Supplementary file 2- Software environment and packages

R version 4.0.0 (2020-04-24)

Platform: x86_64-w64-mingw32/x64 (64-bit)

Running under: Windows 10 x64 (build 19041)

Matrix products: default

locale:

LC_COLLATE=English_United Kingdom.1252 LC_CTYPE=English_United Kingdom.1252

LC_MONETARY=English_United Kingdom.1252 LC_NUMERIC=C

LC_TIME=English_United Kingdom.1252

attached base packages:

stats graphics grDevices utils datasets methods base

other attached packages:

corrplot 0.84

Taiyun Wei and Viliam Simko (2017). R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot

MKmisc 1.6

Kohl M (2019). MKmisc: Miscellaneous functions from M. Kohl_. R package version 1.6, http://www.stamats.de

epiR 1.0-14

Mark Stevenson with contributions from Telmo Nunes, Cord Heuer, Jonathon Marshall, Javier Sanchez, Ron Thornton, Jeno Reiczigel, Jim Robison-Cox, Paola Sebastiani, Peter Solymos, Kazuki Yoshida, Geoff Jones, Sarah Pirikahu, Simon Firestone, Ryan Kyle, Johann Popp, Mathew Jay and Charles Reynard. (2020). epiR: Tools for the Analysis of Epidemiological Data. R package version 1.0-14. https://CRAN.R-project.org/package=epiR

Matching 4.9-7

Jasjeet S. Sekhon (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. Journal of Statistical Software, 42(7), 1-52. URL http://www.jstatsoft.org/v42/i07/.

MASS 7.3-51.5

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Ordinal 2019.12-10

Christensen, R. H. B. (2019). ordinal - Regression Models for Ordinal Data. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

Hmisc 4.4-0

Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2020). Hmisc: Harrell Miscellaneous. R package version 4.4-0. https://CRAN.R-project.org/package=Hmisc

Formula 1.2-3

Achim Zeileis, Yves Croissant (2010). Extended Model Formulas in R: Multiple Parts and Multiple Responses. Journal of Statistical Software 34(1), 1-13. doi:10.18637/jss.v034.i01

lattice 0.20-41

Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5

mice 3.8.0

Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL https://www.jstatsoft.org/v45/i03/.

readxl 1.3.1

Hadley Wickham and Jennifer Bryan (2019). readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

finalfit 1.0.1

Ewen Harrison, Tom Drake and Riinu Ots (2020). finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling. R package version 1.0.1. https://CRAN.R-project.org/package=finalfit

MatchIt 3.0.2

Daniel E. Ho, Kosuke Imai, Gary King, Elizabeth A. Stuart (2011). MatchIt:

Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software, Vol. 42, No. 8, pp. 1-28. URL http://www.jstatsoft.org/v42/i08/

tableone 0.11.1

Kazuki Yoshida (2020). tableone: Create 'Table 1' to Describe Baseline Characteristics. R package version 0.11.1. https://CRAN.R-project.org/package=tableone

forcats 0.5.0

Hadley Wickham (2020). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.0. https://CRAN.R-project.org/package=forcats

stringr 1.4.0

Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr

dplyr 0.8.5

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 0.8.5. https://CRAN.R-project.org/package=dplyr

purrr 0.3.4

Lionel Henry and Hadley Wickham (2020). purrr: Functional Programming Tools. R package version 0.3.4. https://CRAN.R-project.org/package=purrr

readr 1.3.1

Hadley Wickham, Jim Hester and Romain Francois (2018). readr: Read Rectangular Text Data. R package version 1.3.1. https://CRAN.R-project.org/package=readr

tidyr 1.0.2

Hadley Wickham and Lionel Henry (2020). tidyr: Tidy Messy Data. R package version 1.0.2. https://CRAN.R-project.org/package=tidyr

tibble 3.0.0

Hadley Wickham and Lionel Henry (2020). tidyr: Tidy Messy Data. R package version 1.0.2. https://CRAN.R-project.org/package=tidyr

ggplot2 3.3.0

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

tidyverse 1.3.0

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

forestplot 1.9

Max Gordon and Thomas Lumley (2019). forestplot: Advanced Forest Plot Using 'grid' Graphics. R package version 1.9. https://CRAN.R-project.org/package=forestplot

MatchThem 0.9.3

Farhad Pishgar and Noah Greifer (2020). MatchThem: Matching and Weighting Multiply Imputed Datasets. R package version 0.9.3. https://CRAN.R-project.org/package=MatchThem

miceadds 3.9-14

Robitzsch, A., & Grund, S. (2020). miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'. R package version 3.9-14. https://CRAN.R-project.org/package=miceadds

cobalt 4.2.2

Noah Greifer (2020). cobalt: Covariate Balance Tables and Plots. R package version 4.2.2. https://CRAN.R-project.org/package=cobalt