Epidemiological and Clinical Predictors of COVID-19

Yinxiaohe Sun[5*], Vanessa Koh[1,2*], Kalisvar Marimuthu[1,2,4], Oon Tek Ng[1,2,3], Barnaby Young[1,2,7], Shawn Vasoo[1,2], Monica Chan[1,2], Vernon JM Lee[6,5], Partha P De[2], Timothy Barkham[2,4], Raymond TP Lin[1], Alex R Cook[5,1], Yee Sin Leo[1,2,3,4,5] on behalf of the National Centre for Infectious Diseases COVID-19 Outbreak Research Team

[1] Department of Infectious Diseases, National Centre for Infectious Diseases, 16 Jalan Tan Tock Seng, 308442, Singapore

[2] Department of Infectious Diseases, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, 308433, Singapore

[3] Lee Kong Chian School of Medicine, Nanyang Technological University, 11 Mandalay Road, 308232, Singapore

[4] Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, 10 Medical Drive, 117597, Singapore

[5] Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, 12 Science Drive 2, #10-01, 117549, Singapore

[6] Communicable Disease Division, Ministry of Health, 12 College Road, 169852, Singapore

[7] Lee Kong Chian School of Medicine, Nanyang Technological University, 11 Mandalay Road, 308232, Singapore.

[*] These authors contributed equally to the manuscript

Corresponding author: oon_tek_ng@ncid.sg (Oon Tek Ng)

Key points: A risk score incorporating easily ascertainable demographic, clinical evaluation and clinical testing covariates to identify patients at high risk for COVID-19 can help prioritize subjects for testing and public health measures to prevent onward transmission, especially in resource-limited settings.

ABSTRACT

Background

Rapid identification of COVID-19 cases, which is crucial to outbreak containment efforts, is challenging due to the lack of pathognomonic symptoms and in settings with limited capacity for specialized nucleic acid-based reverse transcription polymerase chain reaction (PCR) testing.

Methods

This retrospective case-control study involves subjects (7 to 98 years) presenting at the designated national outbreak screening centre and tertiary care hospital in Singapore for SARS-CoV-2 testing from January 26 to February 16, 2020. COVID-19 status was confirmed by PCR testing of sputum, nasopharyngeal swabs or throat swabs. Demographic, clinical, laboratory and exposure-risk variables ascertainable at presentation were analyzed to develop an algorithm for estimating the risk of COVID-19. Model development used Akaike's information criterion in a stepwise fashion to build logistic regression models, which were then translated into prediction scores. Performance was measured using receiver operating characteristics curves, adjusting for over-confidence using leave-out-one cross validation.

Results

The study population included 788 subjects, of whom 54 (6.9%) were SARS-CoV-2 positive and 734 (93.1%) were SARS-CoV-2 negative. The median age was 34 years and 407 (51.7%) were female. Using leave-out-one cross validation, all the models incorporating clinical tests (Models 1, 2 and 3) performed well with areas under the receiver operating characteristics curve (AUC) of 0.91, 0.88 and 0.88 respectively. In comparison, Model 4 had an AUC of 0.65.

Conclusions

Rapidly ascertainable clinical and laboratory data could identify individuals at high risk of COVID-19 and enable prioritization of PCR-testing and containment efforts. Basic laboratory test results were crucial to prediction models.

Keywords: COVID-19, SARS-CoV-2, Risk factors, Prediction model

INTRODUCTION

On December 31, 2019, a cluster of atypical pneumonia cases was reported in Wuhan City, China[1]. The etiologic agent was subsequently identified as a novel coronavirus[2], severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[3]. The disease, named coronavirus disease 2019 (COVID-19)[4], can progress to acute respiratory distress in severe cases[5]. The basic reproduction number of SARS-CoV-2 has been estimated to be 2.2[6], and human-to-human transmission has since occurred to other parts of China and beyond, affecting 87,137 cases in 59 countries worldwide as of March 1, 2020[6–10].

The clinical spectrum of COVID-19 is broad and the majority of infected individuals experience only a mild or subclinical illness, especially in the early phase of illness[11,12]. Approximately 16 to 26% of hospitalized patients diagnosed with COVID-19 develop severe acute respiratory distress requiring oxygen supplementation and/or intensive care. Disease severity and mortality is associated with older age and underlying comorbidities such as diabetes, hypertension and cardiovascular disease.

In the absence of a vaccine or effective prophylaxis, the containment of SARS-CoV-2 is contingent on interrupting transmission through rapid identification and isolation of all infected individuals. Symptomatic contacts must be isolated early, while close contacts of cases who may be incubating infection need to be quarantined and monitored[13]. Currently case identification relies on specialized nucleic acid-based reverse transcription polymerase chain reaction (PCR) testing, which is not readily available in resource-limited settings[14,15]. Even in well-resourced settings the broad range of clinical presentation presents a challenge in deciding who to test and could strain laboratory testing resources if criteria for testing are overly expansive.

To allow for assessment of the probability of milder cases having COVID-19, we conducted risk factor analysis on a case-control cohort of 54 COVID-19 cases and 734 controls to determine the epidemiological and clinical risk factors that correlate with COVID-19, and determine the accuracy of risk scoring systems based on readily available clinical information.

METHODS

Study design and setting

This retrospective case-control study was conducted in Singapore at the National Centre for Infectious Diseases (NCID), a 330-bed infectious diseases treatment facility with the onsite National Public Health Laboratory, which develops certified testing protocols for emerging infectious diseases for the country[16]. This work was completed as part of outbreak operational evaluation and did not require institutional research board review. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guideline[17].

Study population

Patients presenting to NCID for SARS-CoV-2 testing between January 26 and February 16, 2020 were analyzed. Patients were either self-referred, referred from primary care facilities, or were at-risk cases identified by national contact tracing efforts (Supplementary Table 1). Cases were defined as individuals who had a positive SARS-CoV-2 PCR test and controls were defined as individuals for whom all SARS-CoV-2 PCR results were negative. (Figure 1 and Supplementary Table 2).

Data collection

We collected the following data recorded at initial presentation for testing from the electronic medical records: demographic characteristics, medical comorbidities, exposure risk factors (including contact with a known COVID-19 case, contact with travellers from China, recent travel history, and visit to hospital in China within 14 days prior to symptom onset), symptom days prior to presentation, vital signs at first clinical encounter (respiratory rate,

blood pressure, temperature, and pulse rate); respiratory symptoms, gastrointestinal symptoms, physical examination finding of pneumonia, radiologic evidence of pneumonia and blood investigation results (complete blood count, creatinine, sodium and potassium).

Investigation for SARS-CoV-2

We collected respiratory specimens in the following order of preference: Sputum or endotracheal aspirate, nasopharyngeal swab, and throat swab. For subjects with more than one specimen, the first and last specimens were collected at least 24 hours apart. High-risk patients were tested at least twice while low-risk patients were tested at least once according to a predefined algorithm[18]. SARS-CoV-2 tests were performed using one of the methods described in Supplementary Methods.

Statistical analyses

Study variables from the four abovementioned categories were analysed for differences between SARS-CoV-2 positive and negative subjects using Mann-Whitney-Wilcoxon Test or Yates' corrected chi-squared test. All tests were 2-tailed and a $P < 0.05$ was considered to be statistically significant.

Development of risk-scoring models

A preliminary filtering of variables was conducted by removing those without sufficient variability (fewer than five positive readings or scores) or with too many missing values (more than 80% missing). We also assessed variables for collinearity using variance inflation factor and correlation. We defined a lack of multicollinearity between predictors as a variance inflation factor of less than 2.5 or a correlation coefficient of less than 0.6. When

two variables were found to be colinear, we selected variables for inclusion based on magnitude of effect and clinical relevance.

Predictors of SARS-CoV-2-positive status were classified into four categories: exposure risk factors, demographic variables, clinical findings and clinical test results (Table 2). Two datasets were created: one comprising of 788 subjects with complete reporting for demographic variables, exposure risk factors, clinical findings and radiological tests (excluding other clinical tests such as blood tests), the other comprising of a subset of 292 subjects with complete reporting for demographic variables, exposure risk factors, clinical findings and all clinical tests (Figure 1).

Four prediction models were developed based on these two overlapping datasets: Model 1 included covariates from all four categories; Model 2 included demographic variables, clinical findings and clinical test results; Model 3 included demographic variables, clinical findings and clinical test results (excluding radiology); Model 4 included only demographic variables and clinical findings. Model 4 was built using all 788 subjects (54 cases and 734 controls). Of these 788 subjects, complete blood count was not performed for 481, testing for creatinine, sodium and potassium was not performed for 13, and 2 subjects had incomplete creatinine, sodium and potassium test results. The dataset for Models 1, 2 and 3, which included laboratory blood tests, comprised of a subset of 292 subjects (49 cases and 243 controls) (Figure 1).

The variables for our final models were selected through stepwise use of Akaike's Information Criterion (AIC) to build multivariate logistic regression models, which were then translated into prediction scores.

Evaluation of risk-scoring models

The predictive performance of our final models in determining whether a patient is positive for SARS-CoV-2 was assessed using receiver operating characteristic (ROC) curves and the corresponding area under the ROC (AUC) values with confidence intervals for the specificity at a given sensitivity derived using bootstrapping. We performed leave-out-one cross validation to obtain corrected estimates of sensitivity, specificity and AUC of the risk-scoring models. Specifically, each individual was withheld in turn, the model refit to the remaining individuals, and then used to estimate the withheld patient's risk of COVID-19. This provides a good estimate of the out of sample performance of each model. An AUC of 1.00 corresponds to perfect discrimination, whereas an AUC of 0.50 corresponds to no discriminating ability.

RESULTS

A total of 991 patients were referred to NCID for SARS-CoV-2 testing between January 26 to February 16, 2020. We excluded 193 patients whose SARS-CoV-2 results were not yet available, 3 patients whose electronic medical records were not yet available, and 7 patients with unavailable vital sign records. Of the 788 patients included in the analysis, 54 were COVID-19 cases, and 734 were controls (Figure 1). The median age was 34 years (range: 7 to 98 years; inter-quartile range [IQR]: 27 to 45 years). The majority were female (407, 51.7%). The majority were Singapore citizens (414, 52.5%) or Chinese nationals (145, 18.4%). Of the 54 cases, the median age was 42 years (range: 16 to 79 years; IQR: 34 to 54 years), 29 (53.7%) were male and 48 (88.9%) were ethnic Chinese. Singapore citizens and Chinese nationals comprised of 34 (63%) and 13 (24.1%) cases, respectively. In the control group, the median age was 34 years (range: 7 to 98 years; IQR: 27 to 43 years), 351 (47.9%) were male and 553 (75.3%) were ethnic Chinese. Singapore citizens and Chinese nationals comprised of 379 (51.7%) and 132 (18.0%) cases, respectively (Table 1).

Positive cases were more likely to be older compared with controls ($p<0.001$). Positive cases were not more likely to have any of the comorbidities documented than controls. In terms of exposure risk factors, positive cases were more likely to have contact with a known COVID-19 case (32 out of 54 cases [59.3%]; 126 out of 734 controls [17.2%]) or have recently travelled to Wuhan, China (15 out of 54 cases [27.8%]; 42 out of 734 controls [5.7%]). Positive cases were more likely to have an elevated body temperature ($p=0.003$) at clinical presentation. Of clinical test results, positive cases were more likely to have radiological findings suggestive of pneumonia (23 out of 54 cases [42.6%]; 81 out of 734 controls [11.1%]) as well as lower blood counts of white blood cells, platelets, neutrophils, lymphocytes, eosinophils and basophils (all $p<0.001$) (Table 1).

Significant predictors of SARS-CoV-2 positive test

The final covariate risk estimates of each of the four multivariable models are detailed in Table 2. In Model 1, exposure risk factors most predictive for COVID-19 were travel to Wuhan Province in China since December 1, 2019, around the time of the first outbreak in Wuhan[6] (AOR, Model 1: 23.05, 95% CI: 3.29–268.08) and contact with a confirmed COVID-19 case in Singapore (AOR, Model 1: 6.04, 95% CI: 1.54–27.61).

The other three models exclude exposure risk factors. Clinically, elevated body temperature (AOR, Model 1: 4.81, 95% CI: 1.97–13.12; AOR, Model 2: 2.55, 95% CI: 1.32–5.21; AOR, Model 3: 2.43, 95% CI: 1.25–5.02; AOR, Model 4: 2.27, 95% CI: 1.5–3.44) was the strongest predictor across all four models, except Model 2 where gastrointestinal symptoms fared slightly better (AOR, Model 2: 2.69, 95% CI: 1.08–6.89). Gastrointestinal symptoms was also selected in Model 1 and Model 3 (AOR, Model 1: 3.73, 95% CI: 1.23–12.45; AOR, Model 3: 2.31, 95% CI: 0.92–5.93). Elevated respiratory rate (AOR, Model 1: 1.21, 95% CI: 0.93–1.5; AOR, Model 2: 1.29, 95% CI: 1.07–1.59; Model 3: 1.3, 95% CI: 1.07–1.6) and absence of symptoms such as sore throat (AOR, Model 1: 0.35, 95% CI: 0.1–1.06; AOR, Model 3: 0.53, 95% CI: 0.22–1.25; Model 4: 0.63, 95% CI: 0.34–1.14) and sputum production (AOR, Model 1: 0.23, 95% CI: 0.06–0.78; AOR, Model 2: 0.29, 95% CI: 0.1–0.72; Model 3: 0.3, 95% CI: 0.11–0.79) were strong predictors in the models in which they were selected.

In terms of clinical test results, radiologic evidence of pneumonia (AOR, Model 1: 6.18, 95% CI: 1.68–25.75) was the overall strongest predictor in Model 1 and also contributed significantly to Model 2 (AOR, Model 1: 2.86, 95% CI: 1.09–7.69). Radiology results were excluded in Models 3 and 4. Interestingly, blood parameters were found to contribute significantly to the predictive value of all the models in which they were selected

(Models 1, 2 and 3). The white blood count subsets most closely correlated with risk were lower neutrophil (AOR, Model 1: 0.32 per $1x10^9$/L, 95% CI: 0.19–0.49; AOR, Model 2: 0.39 per $1x10^9$/L, 95% CI: 0.26–0.54; Model 3: 0.38 per $1x10^9$/L, 95% CI: 0.25–0.53) and eosinophil (AOR, Model 1: 0.85 per $1x10^9$/L, 95% CI: 0.78–0.91); AOR, Model 2: 0.89 per $1x10^9$/L, 95% CI: 0.83–0.94; Model 3: 0.9 per $1x10^9$/L, 95% CI: 0.84–0.96) counts.

Model performance of the prediction models

The optimism-bias-corrected performance of Models 1, 2, 3 and 4 differentiated between patients who did and did not have COVID-19 with AUCs of 0.91 (95% CI: 0.86–0.96), 0.88 (95% CI: 0.83–0.93), 0.88 (95% CI: 0.83–0.93), 0.65 (95% CI: 0.57–0.73) respectively (Figure 2). All models incorporating clinical test results had comparable AUCs (0.88 and above). Additionally, comparing Model 2 with Model 3, the exclusion of chest radiology did not result in an appreciable decrease in AUC.

DISCUSSION

Although the epidemiological and clinical characteristics of patients with COVID-19 are well described[19,20], it is challenging for healthcare workers in the primary care or emergency room setting to determine individuals that are more likely to have COVID-19 for isolation and testing. Model 1 incorporating all easily ascertainable data at presentation for SARS-CoV-2 testing performed exceptionally well with an AUC of 0.91. Additionally, the performance of Model 2 suggests that even in the absence of exposure risk factors, clinical findings and tests can identify subjects at high risk of COVID-19. Furthermore, exclusion of radiologic evidence of pneumonia (Model 3) did not significantly impact model performance. However, when basic blood test results such as complete blood count were excluded (Model 4), predictive accuracy was reduced substantially.

The contact risk factors and clinical findings associated with a positive SARS-CoV-2 test are consistent with the known epidemiology and clinical features of COVID-19. Clinical findings strongly associated with a positive SARS-CoV-2 in our sample were higher temperature, higher respiratory rate, gastrointestinal symptoms and decreased sputum production. Our results corroborate with a recent analysis[11] incorporating 1,099 cases throughout China that found fever (87.9%) and non-productive cough (67.7%) to be the dominant symptoms. Diarrhoea (3.7%), although also reported, was less common. In another study involving 138 SARS-CoV-2-positive inpatients from a hospital in Wuhan, a large proportion of patients presented with fever (98.6%) and dry cough (59.4%). Diarrhoea (10.1%) was also reported[12].

Our findings suggest a strong association of reduced white blood cell count with diagnosis of COVID-19. In the above study of 1,099 cases, leukopenia was observed in 33.7% of patients on admission and was more prominent in severe cases[11].

The rapid global dissemination of COVID-19 which has significant morbidity with no proven treatment or vaccine presents a major concern for resource-limited settings with minimal or no access to PCR testing. For well-resourced settings, COVID-19 presents a challenge for healthcare resources to cope with the large numbers of at-risk subjects in need of precautionary (often inpatient) isolation and rapid testing. A risk scoring system would help prioritize high-risk individuals in primary care and emergency room settings for clinical care, isolation precautions and contact tracing efforts.

Most risk scoring systems for infectious pathogens include exposure risk variables, which are sensitive to the local epidemiologic context and phase of the global outbreak. Our current pilot analysis suggests that it is feasible to derive risk-scoring systems for COVID-19 diagnosis, which are reliant mainly on clinical findings and simple test results and hence robust to changes in transmission risk factors.

The current proposed model is based on limited dataset and additional validation in larger datasets and across different contexts would increase confidence in its performance and implementation. Trade-off between sensitivity and specificity will also need to be considered – a higher sensitivity will result in larger number of individuals needing to be isolated and tested, while a higher specificity will exclude some COVID-19 cases.

CONCLUSION

Prediction models which include rapidly ascertainable clinical findings and clinical tests, especially basic blood tests, have sufficient predictive value to identify individuals with a higher probability for COVID-19 and should be considered to stratify at-risk populations for laboratory testing (where available), isolation and contact tracing measures.

ACKNOWLEDGEMENTS

We thank the Centre for Health Protection, Department of Health, Hong Kong for the SARS-CoV N protocol, which was modified to specifically target SARS-CoV-2.

Members of the National Centre for Infectious Diseases COVID-19 Outbreak Research Team

Poh Lian Lim, Brenda Ang, Cheng Chuan Lee, David Chien Boon Lye, Li Min Ling, Lawrence Soon-U Lee, Sapna Sadarangani, Chen Seong Wong, Tau Hong Lee, Ray Junhao Lin, Po Ying Chia, Mucheli Sharavan Sadasiv, Deborah Hee Ling Ng, Chiaw Yee Choy, Tsin Wen Yeo, Glorijoy Shi En Tan, Yu Kit Chan, Jun Yang Tay, Pei Hua Lee, Sean Wei Xiang Ong, Stephanie Sutjipto, Ian Liang En Wee, Dimatatac Frederico, Chi Jong Go, Florante Santo Isais

DISCLAIMER

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of MOH/NMRC.

FUNDING

This work was supported by the following grants from Singapore Ministry of Health's National Medical Research Council: Collaborative Solutions Targeting Antimicrobial Resistance Threats in Health Systems (CoSTAR-HS) [NMRC CGAug16C005], NMRC Clinician Scientist Award [MOH-000276] and NMRC Clinician Scientist Individual Research Grant [MOH-CIRG18nov-0006].

POTENTIAL CONFLICTS:

Dr. Young reports personal fees from Sanofi Pasteur and Roche, outside the submitted work. All other authors have no potential conflicts.

REFERENCES

1. Wuhan Municipal Health Commission. Report of clustering pneumonia of unknown etiology in Wuhan City. 2019. Available at: http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989. Accessed 19 February 2020.

2. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med **2020**; Available at: https://doi.org/10.1056/NEJMoa2001017. Accessed 10 February 2020.

3. Gorbalenya AE, Baker SC, Baric RS, et al. *Severe acute respiratory syndrome-related coronavirus* : The species and its viruses – a statement of the Coronavirus Study Group. Microbiology, 2020. Available at: http://biorxiv.org/lookup/doi/10.1101/2020.02.07.937862. Accessed 19 February 2020.

4. World Health Organization. Novel Coronavirus(2019-nCoV) Situation Report – 22. 2020. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2. Accessed 19 February 2020.

5. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA **2020**; Available at: https://jamanetwork.com/journals/jama/fullarticle/2762130. Accessed 3 March 2020.

6. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med **2020**; :NEJMoa2001316. Available at: http://www.nejm.org/doi/10.1056/NEJMoa2001316. Accessed 30 January 2020.

7. Chan JF-W, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The Lancet **2020**; Available at: https://linkinghub.elsevier.com/retrieve/pii/S0140673620301549. Accessed 30 January 2020.

8. Phan LT, Nguyen TV, Luong QC, et al. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. N Engl J Med **2020**; :NEJMc2001272. Available at: http://www.nejm.org/doi/10.1056/NEJMc2001272. Accessed 30 January 2020.

9. Liu Y-C, Liao C-H, Chang C-F, Chou C-C, Lin Y-R. A Locally Transmitted Case of SARS-CoV-2 Infection in Taiwan. N Engl J Med **2020**; :NEJMc2001573. Available at: http://www.nejm.org/doi/10.1056/NEJMc2001573. Accessed 20 February 2020.

10. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 41. 2020. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2. Accessed 2 March 2020.

11. Guan W, Ni Z, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med **2020**; :NEJMoa2002032. Available at: http://www.nejm.org/doi/10.1056/NEJMoa2002032. Accessed 4 March 2020.

12. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. JAMA **2020**; Available at: https://jamanetwork.com/journals/jama/fullarticle/2761044. Accessed 13 February 2020.

13. Wong JEL, Leo YS, Tan CC. COVID-19 in Singapore—Current Experience: Critical Global Issues That Require Attention and Action. JAMA **2020**; Available at: https://jamanetwork.com/journals/jama/fullarticle/2761890. Accessed 26 February 2020.

14. Wang S, Lifson MA, Inci F, Liang L-G, Sheng Y-F, Demirci U. Advances in addressing technical challenges of point-of-care diagnostics in resource-limited settings. Expert Review of Molecular Diagnostics **2016**; 16:449–459. Available at: http://www.tandfonline.com/doi/full/10.1586/14737159.2016.1142877. Accessed 28 February 2020.

15. Pang J, Wang MX, Ang IYH, et al. Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systematic Review. JCM **2020**; 9:623. Available at: https://www.mdpi.com/2077-0383/9/3/623. Accessed 28 February 2020.

16. Ng OT, Lee V, Marimuthu K, et al. A case of imported Monkeypox in Singapore. The Lancet Infectious Diseases **2019**; 19:1166. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1473309919305377. Accessed 27 February 2020.

17. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med **2015**; 162:W1. Available at: http://annals.org/article.aspx?doi=10.7326/M14-0698. Accessed 26 February 2020.

18. Tay J-Y, Lim PL, Marimuthu K, et al. De-isolating COVID-19 Suspect Cases: A Continuing Challenge. Clinical Infectious Diseases **2020**; Available at: https://doi.org/10.1093/cid/ciaa179. Accessed 3 February 2020.

19. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet **2020**; :7.

20. Chang D, Lin M, Wei L, et al. Epidemiologic and Clinical Characteristics of Novel Coronavirus Infections Involving 13 Patients Outside Wuhan, China. JAMA **2020**; Available at: https://jamanetwork.com/journals/jama/fullarticle/2761043. Accessed 20 February 2020.

TABLES

Table 1. Baseline characteristics of SARS-CoV-2 positive and SARS-CoV-2 negative subjects.

Characteristic	**All, no. (%) (n=788)**	**Cases, no. (%) (n=54)**	**Controls, no. (%) (n=734)**	**P-value[a]**
Demographics				
Age, median (years)	34	42	34	<0.001
Gender				
Male	380 (48.7)	29 (53.7)	351 (47.9)	0.488
Female	407 (51.7)	25 (46.3)	382 (52.1)	
Ethnicity				
Chinese	601 (76.3)	48 (88.9)	553 (75.3)	0.045
Malay	59 (7.5)	1 (1.9)	58 (7.9)	
Indian	69 (8.8)	5 (9.3)	64 (8.7)	
Others	59 (7.5)	0	59 (8.0)	
Nationality				
Singaporean	414 (52.5)	34 (63.0)	380 (51.8)	0.027
Chinese	145 (18.4)	13 (24.1)	132 (18.0)	
Malaysian	79 (10.0)	0	79 (10.8)	
Others	150 (19.1)	7 (13)	143 (19.5)	
Comorbidities				

Any	75 (9.5)	5 (9.3)	70 (9.5)	1.000
Obstructive pulmonary disease	10 (1.3)	0	10 (1.4)	0.815
Congestive heart failure	1 (0.1)	0	1 (0.1)	1.000
Connective tissue disease	4 (0.5)	0	4 (0.5)	1.000
Cerebrovascular disease	7 (0.9)	0	7 (1.0)	1.000
Dementia	4 (0.5)	0	4 (0.5)	1.000
Myocardial infarction	9 (1.1)	0	9 (1.2)	0.877
Leukaemia	1 (0.1)	0	1 (0.1)	1.000
Solid tumour	14 (1.8)	0	14 (1.9)	0.624
Chronic kidney disease	8 (1.0)	0	8 (1.1)	0.946
Diabetes mellitus	54 (6.9)	5 (9.3)	49 (6.7)	0.655
Chronic liver disease	3 (0.4)	0	3 (0.4)	1.000
Exposure risk factors				
Healthcare worker	79 (10.0)	0	79 (10.8)	0.021
Contact with:				
a known COVID-19 case	158 (20.1)	32 (59.3)	126 (17.2)	<0.001
a traveller from China (from December 1, 2019)	174 (22.1)	11 (20.4)	163 (22.2)	0.885
a group of travellers from China (from December 1, 2019)	84 (10.7)	7 (13)	77 (10.5)	0.734
History of travel (from December 1, 2019) to:				

Wuhan, China	57 (7.2)	15 (27.8)	42 (5.7)	<0.001
China (including Wuhan)	236 (30.0)	17 (31.5)	219 (29.8)	0.920
Other countries (other than China)	216 (27.4)	18 (33.3)	198 (27)	0.394
Visited any hospital in China recently (14 days since onset of symptoms)	6 (0.8)	0	6 (0.8)	1.000
Clinical signs and symptoms				
Number of subjects with >5 days of symptoms (n=758)[b]	252 (33.2)	20 (38.5)	232 (32.9)	0.38
Body temperature, median (ºC)	37.1	37.5	37.1	0.003
Heart rate, median (beats per minute)	89	87	89	0.379
Respiration rate, median (breaths per minute)	18	18	18	0.159
Systolic blood pressure, median (mmHg)	131	131	131	0.502
Diastolic blood pressure, median (mmHg)	78	78	78	0.596
Cough	564 (71.5)	36 (66.7)	528 (71.9)	0.502
Sputum production	212 (26.9)	13 (24.1)	199 (27.1)	0.744
Shortness of breath	100 (12.7)	7 (13)	93 (12.7)	1.000
Rhinnorhea or nasal congestion	238 (30.2)	12 (22.2)	226 (30.8)	0.242
Sore throat	350 (44.4)	18 (33.3)	332 (45.2)	0.120
Auscultation finding of pneumonia (e.g.	42 (5.3)	6 (11.1)	36 (4.9)	0.100

crackles)				
Respiratory symptoms (other than those listed above)	45 (5.7)	2 (3.7)	43 (5.9)	0.723
Gastrointestinal symptoms	258 (32.8)	20 (37)	238 (32.4)	0.585
Clinical Tests				
CXR/CT suggestive of pneumonia (n=788)	104 (13.2)	23 (42.6)	81 (11.1)	<0.001
Complete blood count (n=307)[c]				
White blood cells, median (x10^9/L)	7.1	4.7	7.8	<0.001
Haemoglobin, median (g/dL)	13.5	13.9	13.4	0.102
Platelets, median (x10^9/L)	242	205	249	<0.001
Neutrophils, median (x10^9/L)	4.4	2.5	4.9	<0.001
Lymphocytes, median (x10^9/L)	1.6	1.2	1.7	<0.001
Eosinophils, median (x10^9/L)	0.09	0.02	0.10	<0.001
Basophils, median (x10^9/L)	0.03	0.02	0.04	<0.001
Renal panel (n=294)[d]				
Creatine, median (µmol/L)	63	64	62	0.977
Sodium, median (mmol/L)	141	141	141	0.600
Potassium, median (mmol/L)	3.6	3.5	3.6	0.156

Abbreviations: CXR, chest X-ray; CT, chest computed tomography scan.

[a]The Yates' corrected χ^2 test and Mann-Whitney-Wilcoxon test were used to calculate P values for categorical and continuous variables, respectively

[b]There were a total of 758 subjects that were symptomatic on presentation (52 cases and 706 controls). 30 subjects were asymptomatic on presentation (2 cases and 28 controls)

[c]Complete blood count was performed for 307 subjects (out of 788), of which 52 were cases (out of 54) and 255 were controls (out of 734)

[d]Renal panel results were obtained for 294 subjects (out of 788), of which were 51 were cases (out of 54) and 243 were controls (out of 734)

Table 2. Final covariates in the four multivariate models for COVID-19 infection.

Variable	Model 1		Model 2		Model 3		Model 4	
	AOR (95% CI)	P-value	AOR (95% CI)	P-value	AOR (95% CI)	P-value	AOR (95% CI)	P-value
Age							1.03 (1.02 - 1.05)	<0.001
Male sex	5.98 (1.23 - 36.05)	0.038	3.67 (1.03 - 14.12)	0.051	3.51 (0.97 - 13.89)	0.063		
Contact with a COVID-19 case	6.04 (1.54 - 27.61)	0.013						
Travel to Wuhan since December 1, 2019	23.05 (3.29 - 268.08)	0.004						
Travel to China (including Wuhan) since December 1, 2019	0.02 (0 - 0.19)	0.002						
Temperature	4.81 (1.97 - 13.12)	0.001	2.55 (1.32 - 5.21)	0.007	2.43 (1.25 - 5.02)	0.011	2.27 (1.5 - 3.44)	<0.001
Heart rate	0.95 (0.91 - 1)	0.044	0.95 (0.92 - 0.99)	0.01	0.96 (0.92 - 0.99)	0.029	0.97 (0.95 - 0.99)	0.01
Respiration rate	1.21 (0.93 - 1.5)	0.079	1.29 (1.07 - 1.59)	0.005	1.3 (1.07 - 1.6)	0.004		
Systolic blood pressure							0.97 (0.95 - 0.99)	0.016
Diastolic blood pressure	1.04 (0.99 - 1.1)	0.103	1.04 (1 - 1.09)	0.061	1.05 (1 - 1.1)	0.044	1.03 (1 - 1.06)	0.102

Sore throat	0.35 (0.1 - 1.06)	0.073			0.53 (0.22 - 1.25)	0.149	0.63 (0.34 - 1.14)	0.132
Sputum production	0.23 (0.06 - 0.78)	0.024	0.29 (0.1 - 0.72)	0.011	0.3 (0.11 - 0.79)	0.019		
Shortness of breath					2.76 (0.67 - 10.7)	0.145		
Gastrointestinal symptoms	3.73 (1.23 - 12.45)	0.024	2.69 (1.08 - 6.89)	0.035	2.31 (0.92 - 5.93)	0.076		
CXR/CT suggestive of pneumonia	6.18 (1.68 - 25.75)	0.008	2.86 (1.09 - 7.69)	0.033				
Lymphocytes (per $1x10^9$/L)					0.56 (0.25 - 1.12)	0.117		
Neutrophils (per $1x10^9$/L)	0.32 (0.19 - 0.49)	<0.001	0.39 (0.26 - 0.54)	<0.001	0.38 (0.25 - 0.53)	<0.001		
Eosinophils (per $1x10^9$/L)	0.85 (0.78 - 0.91)	<0.001	0.89 (0.83 - 0.94)	<0.001	0.9 (0.84 - 0.96)	0.002		
Creatinine (per µmol/L)	0.96 (0.9 - 1)	0.111	0.96 (0.91 - 1)	0.062	0.96 (0.92 - 1)	0.079		
Sodium (per mmol/L)	1.17 (0.96 - 1.43)	0.133						

Abbreviations: AOR, adjusted odds ratio; CXR, chest X-ray; CT, chest computed tomography scan

FIGURE LEGENDS

Figure 1. Study subject disposition. NCID: National Centre for Infectious Diseases, Singapore.

Figure 2. Performance of Models 1, 2, 3 and 4 measured using receiver operating characteristics curves, adjusting for over-confidence using leave-out-one cross validation.

Figure 1

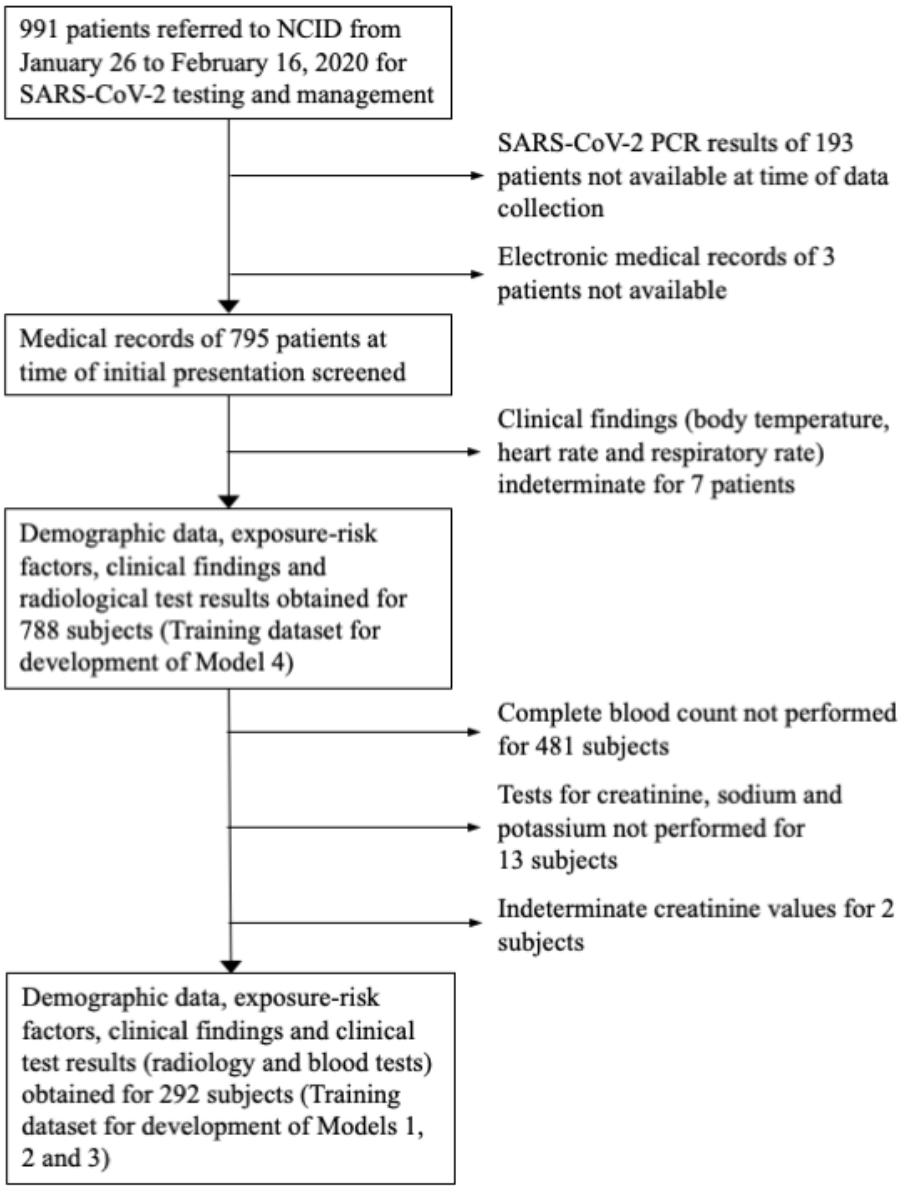

Figure 2