



A predictive model for disease progression in non-severely ill patients with coronavirus disease 2019

To the Editor:

Over the past 3 months, coronavirus disease 2019 (COVID-19) has emerged across China and developed into a worldwide outbreak [1]. The disease has caused varying degrees of illness. The proportion of patients with COVID-19 with non-severe illness was 84.3% on admission, and severe cases accounted for 15.7% [2]. Most of the non-severe pneumonia patients would gradually alleviate and be cured with treatment, while others would rapidly progress to severe illness, which has a poor prognosis [3, 4]. As recently reported, the cumulative risk of the composite end-point was 3.6% in all COVID-19 patients, and the cumulative risk was 20.6% for severe illness [2].

However, it is still unknown whether early identification and intervention for non-severe patients with COVID-19 could prevent progression into severe disease. According to the experience of treating other diseases, there might be a large promoting effect of treatment. In this paper, we aim to build a predictive model for identifying high-risk non-severe pneumonia patients at an early stage.

86 patients with COVID-19 and non-severe pneumonia on admission were recruited as the training cohort at Renmin Hospital of Wuhan University from 2 to 20 January, 2020, and another 62 patients were prospectively enrolled as the validation cohort from 28 January to 9 February, 2020. COVID-19 was confirmed by real-time PCR. Disease severities of COVID-19 were defined as severe and non-severe pneumonia based on the criteria of American Thoracic Society guidelines for community-acquired pneumonia [2, 5]. The exclusion criteria included: 1) degrees of severity were not available on admission or during follow-up; 2) diagnosed with severe illness at the time of admission; 3) confirmed with COVID-19 and treated at other hospitals; 4) medication was administered within 15 days before admission; 5) received oxygen support during follow-up. Patients were divided into “progressed” or “non-progressed” groups, based on whether they progressed to severe illness or not during the 14-day follow-up period. Comorbidity included diabetes, hypertension, cardiovascular and cerebrovascular diseases, COPD, malignant tumour, chronic liver disease, chronic kidney disease, tuberculosis and immunodeficiency diseases, *etc.*

Clinical characteristics and laboratory findings were extracted from electronic medical records. Radiological features were extracted from chest computed tomography (CT) imaging using a double-blind method [6]. To evaluate the lesion size accurately, a diagnosis system for COVID-19 based on artificial intelligence (AI) was employed to measure volume ratio of pneumonia automatically by analysing CT values [7, 8].

Logistic regression was used as the classifier to build the predictive model. The discriminative performance of the predictive model was quantified by the value of the area under the receiver operating characteristic curve (AUC) in the cross-validation of the training and validation datasets. Risk index calculated with the weight of each variable in the model was used to identify high-risk groups. All analyses were performed using R-3.6.0.

The median age of the 148 patients was 46.5 years (interquartile range (IQR) 35.8–58.0 years), and 81 (54.7%) were female. A total of 60 (40.5%) non-severe patients progressed to severe illness, and the median time of progression was 5.0 days (IQR 2.8–9.0 days). For training cohort, 60 (40.5%) non-severe patients progressed to severe illness, and 26 (41.9%) cases were in validation cohort. The median time of

 @ERSpublications
A predictive model for COVID-19 <https://bit.ly/2WHiEkJ>

Cite this article as: Ji M, Yuan L, Shen W, *et al.* A predictive model for disease progression in non-severely ill patients with coronavirus disease 2019. *Eur Respir J* 2020; 56: 2001234 [<https://doi.org/10.1183/13993003.01234-2020>].

TABLE 1 Description of clinical characteristics and multivariate analysis in training cohort

Variables	Clinical characteristics		Multivariate analysis in training cohort	
	Training cohort (n=86)	Validation cohort (n=62)	OR (95% CI)	Score
Group				
Non-progressed group	52 (60.5%)	36 (58.1%)		
Progressed group	34 (39.5%)	26 (41.9%)		
Time of progression days	5.5 (1.0–9.0)	5.0 (3.0–9.8)		
Age years	50.5 (37.0–60.5)	44.5 (35.0–53.0)		
Age range years				
<40	27 (31.4%)	21 (33.9%)		
40–49	15 (17.4%)	14 (22.6%)		
50–59	22 (25.6%)	15 (24.2%)		
60–69	13 (15.1%)	9 (14.5%)		
70–79	7 (8.1%)	2 (3.2%)		
≥80	2 (2.3%)	1 (1.6%)		
Female	45 (52.3%)	36 (58.1%)		
Comorbidity	42 (48.8%)	15 (24.2%)	3.436 (1.084–10.896)	12× [0/1; no=0, yes=1]
Dyspnoea on admission	11 (12.8%)	6 (9.7%)	4.869 (0.760–31.212)	16× [0/1; no=0, yes=1]
Temperature on admission °C	36.8 (36.5–37.2)	36.8 (36.5–37.1)		
Respiratory rate on admission	19.0 (18.0–20.0)	20.0 (19.0–20.0)		
Lactate dehydrogenase U·L⁻¹	214.0 (187.8–275.8)	201.5 (160.3–247.0)	1.008 (1.001–1.014)	0.07× per unit (U·L ⁻¹)
Procalcitonin ng·mL⁻¹	0.04 (0.03–0.07)	0.03 (0.02–0.05)		
Lymphocyte count ×10⁹ L⁻¹	1.2 (0.9–1.6)	1.3 (1.0–1.7)	0.134 (0.038–0.471)	–20× per unit (10 ⁹ L ⁻¹)
White blood cells ×10⁹ L⁻¹	4.8 (3.7–6.1)	4.7 (4.0–6.1)		
Neutrophil count ×10⁹ L⁻¹	3.1 (2.2–4.1)	3.0 (2.0–3.9)		
Platelet count ×10⁹ L⁻¹	159.3 (132.5–204.0)	164.5 (120.3–210.4)		
Haemoglobin concentration g·L⁻¹	138.5 (127.0–156.6)	143.3 (130.0–152.8)		
Arterial oxygen saturation %	97.0 (95.3–98.8)	96.0 (95.0–98.0)		
Radiological abnormality				
GGOSS	36 (41.9%)			
Pure ground-glass opacity	32 (37.2%)			
Consolidation	12 (14.0%)			
Other	6 (7.0%)			
Number of affected segments	7.0 (2.3–12.0)			
Lesion size				
<1 cm	4 (4.7%)			
1–3 cm	32 (37.2%)			
3 cm to 50% lobe	45 (52.3%)			
>50% lobe	5 (5.8%)			
AI-based volume ratio of pneumonia				
–700 to 500 HU	0.18 (0.11–0.27)			
–600 to 500 HU	0.11 (0.07–0.17)			
Treatment				
Corticosteroid agents	55 (64.0%)	19 (30.6%)		
Anti-infection agents	85 (98.8%)	52 (83.9%)		
Interferon agents	34 (39.5%)	7 (11.3%)		
Antiviral agents	74 (86%)	61 (98.4%)		
Gamma globulin agents	54 (62.8%)	21 (33.9%)		

Data are presented as medians (interquartile ranges) and n (%). Percentages may not total 100 because of rounding. Variables in the validation cohort were not completely collected, as some of them did not appear in the model of the training cohort. GG OSS: ground-glass opacities overlapped with striped shadows; AI: artificial intelligence.

progression in these two cohorts were 5.5 days (IQR 1.0–9.0 days) and 5.0 days (IQR 3.0–9.8 days). Description of variables was provided in the table 1.

To build the predictive model, we tested all the clinical, laboratory and radiological variables, except for characteristics about treatment. Four variables were finally included in the model, including comorbidity ($\beta=1.234$, $p=0.036$), dyspnoea on admission ($\beta=1.583$, $p=0.095$), lactate dehydrogenase ($\beta=0.007$, $p=0.027$) and lymphocyte count ($\beta=-2.012$, $p=0.002$). The Hosmer Lemeshow test of the training dataset was done (Chi-squared=10.451, $p=0.235$). The AUC value in the cross-validation of training dataset was 0.819 (95% CI 0.731–0.907). It was 0.759 (95% CI 0.635–0.884) in the validation dataset. According to the regression

coefficients, the four variables were given different weights. Comorbidity was 12 points per unit, dyspnoea was 16, lactate dehydrogenase was 0.07, and lymphocyte count was -20 . Then, total scores for each person were calculated, and different scores showed different risks. AUC value based on the risk scores in training dataset was 0.856 (95% CI 0.776–0.935). Patients were divided into high-risk and low-risk groups (total score >-6.0 and ≤ -6.0) based on the best cut-off value determined by the Youden index; the sensitivity was 0.941, specificity was 0.635. More details can be found in table 1.

In our prediction model, comorbidity was associated with disease progression, which meant that patients with comorbidities were more likely to progress to severe disease than those without. Previous studies have shown a higher proportion of patients with comorbidities in those with more severe disease [9]. We further confirmed that non-severe patients with comorbidities were more likely to progress. It should be explained that the p value for dyspnoea on admission was not less than 0.05 in the multivariate regression, which might be due to the relationship between dyspnoea and the outcome in this study not being strictly linear after adjusting for other variables. Although we did try other models with better performance earlier, we finally chose the logistic model because of its interpretability and simplicity of application. Patients who progressed have been found to be more likely to accompany this with a decrease in lymphocyte count and an increase in lactate dehydrogenase [2, 10]. Our research further confirmed that these two indicators were also related to disease progression. A decrease in lymphocyte count usually indicated the decline of immune function, and multiple organ dysfunction might lead to an increase in lactate dehydrogenase [11], which are consistent with the phenomena we have observed clinically.

Previous reports have pointed out that advanced age was one of the risk factors for poor prognosis in patients with COVID-19 [2, 3]. However, age was not included in the model. It suggests that treatment for young non-severe illness patients should not be neglected in at an early stage. We speculate that the contribution of age to disease progression was reflected in comorbidities and dyspnoea. In addition, some studies reported the correlations between radiological indicators and COVID-19 disease [12]. Although radiological features in CT images on admission were described in detail, they were not included into the model. We speculate that multiple images during treatment instead of a single image could indicate further progression of the disease. Although variables extracted with AI from CT imaging were not included in the model, this was showed promise and will be the focus of our subsequent research.

There were some limitations to this study. First, patients with COVID-19 included in this study were from a single hospital, which is a potential constraint for the generalisation of our model. Second, critically ill patients were transferred to other designated hospitals according to the regulations of the local government. We were unable to track these patients' deaths in the short term, and the association between the model and overall survival could not be evaluated, which unfortunately was a major limitation of this study.

Conclusively, the progression of non-severe patients with COVID-19 could be predicted by our model based on clinical characteristics on admission. The model was further verified with a prospective validation cohort with good performance. With the help of our model, clinicians could easily identify high-risk non-severe patients on admission with few routine clinical indicators, thereby contributing to the treatment and prevention of COVID-19.

Mengyao Ji^{1,10}, Lei Yuan^{2,10}, Wei Shen^{3,4,10}, Junwei Lv^{3,4}, Yong Li^{5,10}, Jia Chen^{3,4}, Chaonan Zhu^{3,4}, Bo Liu^{3,4}, Zhenzhen Liang⁶, Qiang Lin^{3,4}, Wenjie Xie⁷, Ming Li², Zhifan Chen⁸, Xuefang Lu⁹, Yijuan Ding¹, Ping An¹, Sheng Zhu^{3,4}, Mengting Gao², Hao Ni^{3,4}, Lanhua Hu⁹, Guanglei Shi^{3,4}, Lei Shi^{3,4,11} and Weiguo Dong^{1,11}

¹Dept of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China. ²Dept of Information Center, Renmin Hospital of Wuhan University, Wuhan, China. ³Shanghai Key Laboratory of Artificial Intelligence for Medical Image and Knowledge Graph, Shanghai, China. ⁴Hangzhou YITU Healthcare Technology Co., Ltd, Hangzhou, China. ⁵Dept of Radiology, Suining Central Hospital, Suining, China. ⁶Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China. ⁷Dept of Critical Care Medicine, Renmin Hospital of Wuhan University, Wuhan, China. ⁸Dept of Radiology, Affiliated Hospital of Chengdu University, Chengdu, China. ⁹Dept of Radiology, Renmin Hospital of Wuhan University, Wuhan, China. ¹⁰Mengyao Ji, Lei Yuan, Wei Shen and Yong Li contributed equally to this paper. ¹¹Weiguo Dong and Lei Shi contributed equally to this article as lead authors and supervised the work.

Correspondence: Weiguo Dong, Renmin Hospital of Wuhan University, 238 Jiefang Road, Wuhan City, Wuhan 430060, China. E-mail: weiguodong_66@163.com

Received: 16 March 2020 | Accepted after revision: 10 May 2020

Support statement: This work was supported by the National Natural Science Foundation of China (grant: 81901817), Natural Science Foundation of Hubei Province (grant: 2018CFB136), and Innovation Seed Funding of Wuhan University (grant: TFZZ2018020). Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: None declared.

References

- 1 Chinese Center for Disease Control and Prevention. Distribution of Covid-19. <http://2019ncov.chinacdc.cn/2019-nCoV/>
- 2 Guan WJ, Ni ZY, Hu Y, *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; 382: 1708–1720.
- 3 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; 323: 1061–1069.
- 4 Yang X, Yu Y, Xu J, *et al.* Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020; 8: 475–481.
- 5 Metlay JP, Waterer GW, Long AC, *et al.* Diagnosis and treatment of adults with community-acquired pneumonia. an official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med* 2019; 200: e45–e67.
- 6 Song F, Shi N, Shan F, *et al.* Emerging 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology* 2020; 295: 210–217.
- 7 Cohen JG, Goo JM, Yoo RE, *et al.* Software performance in segmenting ground-glass and solid components of subsolid nodules in pulmonary adenocarcinomas. *Eur Radiol* 2016; 26: 4465–4474.
- 8 Kitami A, Kamio Y, Hayashi S, *et al.* One-dimensional mean computed tomography value evaluation of ground-glass opacity on high-resolution images. *Gen Thorac Cardiovasc Surg* 2012; 60: 425–430.
- 9 Liang W, Guan W, Chen R, *et al.* Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *Lancet Oncol* 2020; 21: 335–337.
- 10 Liu Y, Yang Y, Zhang C, *et al.* Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *SciChina Life Sci* 2020; 63: 364–374.
- 11 Khan AA, Allemailem KS, Alhumaydhi FA, *et al.* The biochemical and clinical perspectives of lactate dehydrogenase: an enzyme of active metabolism. *Endocr Metab Immune Disord Drug Targets* 2019; in press [<https://doi.org/10.2174/1871530320666191230141110>].
- 12 Arabi YM, Arifi AA, Balkhy HH, *et al.* Clinical course and outcomes of critically ill patients with Middle East respiratory syndrome coronavirus infection. *Ann Intern Med* 2014; 160: 389–397.

Copyright ©ERS 2020.

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.