



A predictive model for respiratory distress in patients with COVID-19: a retrospective study

Xin Zhang^{1,2,3,4#}, Wei Wang^{5#}, Cheng Wan^{1,3#}, Gong Cheng⁶, Yuechuchu Yin^{1,2,3}, Kaidi Cao^{1,2,3}, Xiaoliang Zhang^{1,2,3}, Zhongmin Wang^{1,2,3}, Shumei Miao^{1,2,3}, Yun Yu^{1,3}, Jie Hu^{1,3}, Ruochen Huang^{1,2,3}, Yun Ge⁴, Ying Chen⁴, Yun Liu^{1^}

¹Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China; ²Department of Information, the First Affiliated Hospital, Nanjing Medical University, Nanjing, China; ³Institute of Medical Informatics and Management, Nanjing Medical University, Nanjing, China; ⁴School of Electronic Science and Engineering, Nanjing University, Nanjing, China; ⁵Network Information Center, Wuhan No. 1 Hospital, Wuhan, China; ⁶State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

Contributions: (I) Conception and design: Y Liu, X Zhang, G Cheng; (II) Administrative support: Y Liu, W Wang, Z Wang; (III) Provision of study materials or patients: W Wang; (IV) Collection and assembly of data: X Zhang, Y Yin, K Cao, X Zhang; (V) Data analysis and interpretation: X Zhang, C Wan, S Miao, Y Yu, J Hu, R Huang, Y Ge, Y Chen; (VI) Manuscript writing: All authors.; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Yun Liu. Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, No. 300 Guang Zhou Road, Nanjing, China. Email: liuyun@njmu.edu.cn.

Background: Coronavirus disease 2019 (COVID-19), associated with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a global public health crisis. We retrospectively evaluated 863 hospitalized patients with COVID-19 infection, designated IWCH-COVID-19.

Methods: We built a successful predictive model after investigating the risk factors to predict respiratory distress within 30 days of admission. These variables were analyzed using Kaplan-Meier and Cox proportional hazards (PHs) analyses. Hazard ratios (HRs) and performance of the final model were determined.

Results: Neutrophil count $>6.3 \times 10^9/L$, D-dimer level ≥ 1.00 mg/L, and temperature ≥ 37.3 °C at admission showed significant positive association with the outcome of respiratory distress in the final model. Complement C3 (C3) of 0.9–1.8 g/L, platelet count $>350 \times 10^9/L$, and platelet count of $125\text{--}350 \times 10^9/L$ showed a significant negative association with outcomes of respiratory distress in the final model. The final model had a C statistic of 0.891 (0.867–0.915), an Akaike's information criterion (AIC) of 567.65, and a bootstrap confidence interval (CI) of 0.866 (0.842–0.89). This five-factor model could help in early allocation of medical resources.

Conclusions: The predictive model based on the five factors obtained at admission can be applied for calculating the risk of respiratory distress and classifying patients at an early stage. Accordingly, high-risk patients can receive timely and effective treatment, and health resources can be allocated effectively.

Keywords: Coronavirus disease 2019 (COVID-19); coronavirus; predictive model; respiratory distress; severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

Submitted Jun 27, 2020. Accepted for publication Sep 25, 2020.

doi: 10.21037/atm-20-4977

View this article at: <http://dx.doi.org/10.21037/atm-20-4977>

[^] ORCID: Xin Zhang: 0000-0003-0047-3134; Yun Liu: 0000-0002-4311-3772.

Introduction

The World Health Organization (WHO) designated the pneumonia caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the 2019 coronavirus disease (COVID-19). COVID-19 has become a global public health crisis. As of May 8, 2020, more than 3.5 million cases of COVID-19 and 250,000 deaths have been reported to the WHO. As had said, *“We cannot end the pandemic until we address the inequalities that are fueling it”* (1).

Although most patients only exhibit mild symptoms, a considerable proportion of patients have serious disease progression, with the aggravation of hypoxia until respiratory distress or in some cases respiratory failure. Approximately 5% of patients need intensive care, including mechanical ventilation (MV) (2-4). The progression of respiratory distress is important to the survival and prognosis of patients. As patients with respiratory distress usually need equipment such as ventilators, if respiratory distress could be predicted in advance, medical personnel will be better equipped to deal with the emergency condition of some patients. Meanwhile, this will provide a basis for the allocation of medical resources and help improve the prognosis of patients. Because of the variability of COVID-19 virus and the shortage of healthcare resources in the affected areas, it is very important to effectively allocate resources to high-risk patients with worsening disease (5). Respiratory distress should be prepared in advance to avoid emergency intubation or cardiopulmonary resuscitation, so as to avoid endangering the lives and safety of medical staff.

Although several studies have analyzed the prognostic factors for mortality outcomes in COVID-19 (2,6,7), the predictors of respiratory distress have not been reported fully and in detail. Thus, in this study, we retrospectively evaluated patients with COVID-19 to determine the factors that can predict respiratory distress and the need for MV in the early stages of the disease. This will facilitate reasonable early triage of patients and adequate preparation for the provision of the required medical resources.

We present the following article in accordance with the TRIPOD reporting checklist (available at <http://dx.doi.org/10.21037/atm-20-4977>).

Methods

Study design and data collection

This was a retrospective study of 1,150 patients, 12–97 years

of age, who were hospitalized in the infection ward of Wuhan No. 1 Hospital, China. The infection ward was established on February 12, 2020, and cleaned up on March 19, 2020. After 286 patients were excluded because they were not diagnosed at the time of discharge from the hospital, the remaining 864 patients who were included were diagnosed as having COVID-19 pneumonia (ICD-10 code U07.100x001) or COVID-19 infection (ICD-10 code U07.100x002), or had a clinical diagnosis of COVID-19 (ICD-10 code U07.100x003) according to Diagnosis and Treatment of COVID-19 (trial version 7) (8). COVID-19 infection was confirmed in the laboratory by local health authorities. In China, pediatricians take care of children <14 years old, and those above 14 years old are treated in an adult specialty. Therefore, we excluded one patient who was under the age of 14.

Derivation cohort: data for this cohort were extracted from the electronic medical records (EMRs) designed for patients with “respiratory distress” at the infection ward of Wuhan No. 1 Hospital, China (briefly designated as IWCH-COVID-19 in this paper). The clinical characteristics, laboratory findings, treatment, procedure, and outcome data of 863 patients who were discharged from the hospital (including patients who died or were transferred to another hospital) before March 19, 2020, were extracted from EMRs.

The anonymized data were subjected to privacy-free and cleaning treatment. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (No. 2020-SR-163), and informed consent from study participants was waived.

Outcomes

Patients were divided into two groups depending on whether they developed respiratory distress within 30 days: patients without respiratory distress (NRD group) and patients with respiratory distress before discharge (RD group). Respiratory distress was defined as the need for MV or being diagnosed with acute respiratory distress syndrome (ARDS, ICD-10 code J80.x00) or respiratory failure (ICD-10 codes J96.000, J96.900). Patient survival time in the RD group was defined as the number of days from the time of admission to the time of starting MV, while the patient survival time in the NRD group was counted as the number of days from the day of admission to the day of discharge

(or day of death if the patient in the NRD group had died). Due to the recommendation of immediate high-flow nasal oxygen and prone positioning for patients with moderately severe hypoxemia (9) or respiratory distress (10), the patients undergoing these procedures were placed in the NRD group.

Predictors

In our analysis, we included variables that were available in the hospitals' EMR system and known or thought to affect respiratory distress risk according to results reported in related research. The laboratory test value closest to the entry date to the cohort for each patient was used, and missing values were imputed where necessary, as described below. Two demographic characteristics (age and gender); five clinical measurements [systolic blood pressure (SBP), diastolic blood pressure (DBP), breath, pulse, and temperature at admission]; seven comorbidities (hypertension, diabetes, coronary heart disease, chronic obstructive lung disease, cerebrovascular disease, carcinoma and chronic kidney disease) and fourteen laboratory test results [white blood cell count, neutrophil count, lymphocyte count, hemoglobin, platelet count, albumin, alanine aminotransferase (ALT), aspartate aminotransferase, serum creatinine, lactate dehydrogenase, creatine kinase, D-dimer, interleukin-6 (IL-6), and complement C3 (C3)] that are routinely measured were initially identified as potential predictors on the basis of existing research (2,6,7) (Table S1). The first laboratory test results obtained in 3 days from the day of patient admission were used to present the patient admission status, and the results were divided into classes using the reference values below:

- ❖ White blood cell count: $(3.50-9.50) \times 10^9/L$;
- ❖ Neutrophil count: $(1.80-6.30) \times 10^9/L$;
- ❖ Lymphocyte count: $(1.10-3.20) \times 10^9/L$;
- ❖ Hemoglobin: 130–175 g/L;
- ❖ Platelet count: $(125-350) \times 10^9/L$;
- ❖ Albumin: 40.0–55.0 g/L;
- ❖ ALT: 9–60 IU/L;
- ❖ Aspartate aminotransferase: 15–40 IU/L;
- ❖ Serum creatinine concentration: 53–106 $\mu\text{mol/L}$;
- ❖ Lactate dehydrogenase: 114–250 IU/L;
- ❖ Creatine kinase: 0–171 IU/L;
- ❖ D-dimer: 0–1.00 mg/L;
- ❖ IL-6: 0–5.90 pg/mL;
- ❖ C3: 0.90–1.80 g/L.

Missing data

The combinations of missing values were analyzed by cluster analysis of patterns of missingness. The IWCH-COVID-19 cohort had missing information on SBP (1.64%), DBP (1.76%), pulse (0.23%), white blood cell count (7.03%), neutrophil count (7.03%), lymphocyte count (7.03%), hemoglobin (7.03%), platelet count (7.03%), albumin (9.50%), ALT (9.50%), aspartate aminotransferase (9.50%), serum creatinine concentration (10.20%), lactate dehydrogenase (30.95%), creatine kinase (30.95%), D-dimer (48.77%), C3 (89.21%), and IL-6 (96.95%). The completeness of predictors can be found in Table S2.

Missing predictor data were imputed using multiple imputations, assuming data were missing at random (R package mice). Because of the high rate of missing data (>90%), we excluded IL-6 test results. The model contained all the prediction factors except IL-6, and was used to generate the estimation dataset of 20 missing variables. It was evaluated by Marshall's adaption of Rubin's rules (11).

Statistical analysis

Grouped Kaplan-Meier analysis, which is a nonparametric approach, was applied to compare the different characteristics of the IWCH-COVID-19 cohort. The time unit was set as days.

The hypothesis of proportional hazard (PH) was tested by the statistical significance test and graphic diagnosis based on the scaled Schoenfeld residuals. A univariate Cox PH model was used to evaluate the risk of respiratory distress in the IWCH-COVID-19 cohort with associations expressed as hazard ratios (HRs) (Figure S1, step 1). When appropriate, the 95% confidence interval (CI) was calculated, and $P < 0.01$ was considered statistically significant. Each model calculated Akaike's information criterion (AIC). The top 12 statistical significance predictors with lower AIC values were included in a multivariable Cox PH model (Figure S1, step 2). A backward stepwise procedure was used to evaluate the final statistical significance of the predictors.

Performance of risk prediction

The predictive performance of the selected risk scores was assessed by internal validation. Discrimination of the final model was indicated by the Harrell C statistic. Discrimination describes the ability of the model to

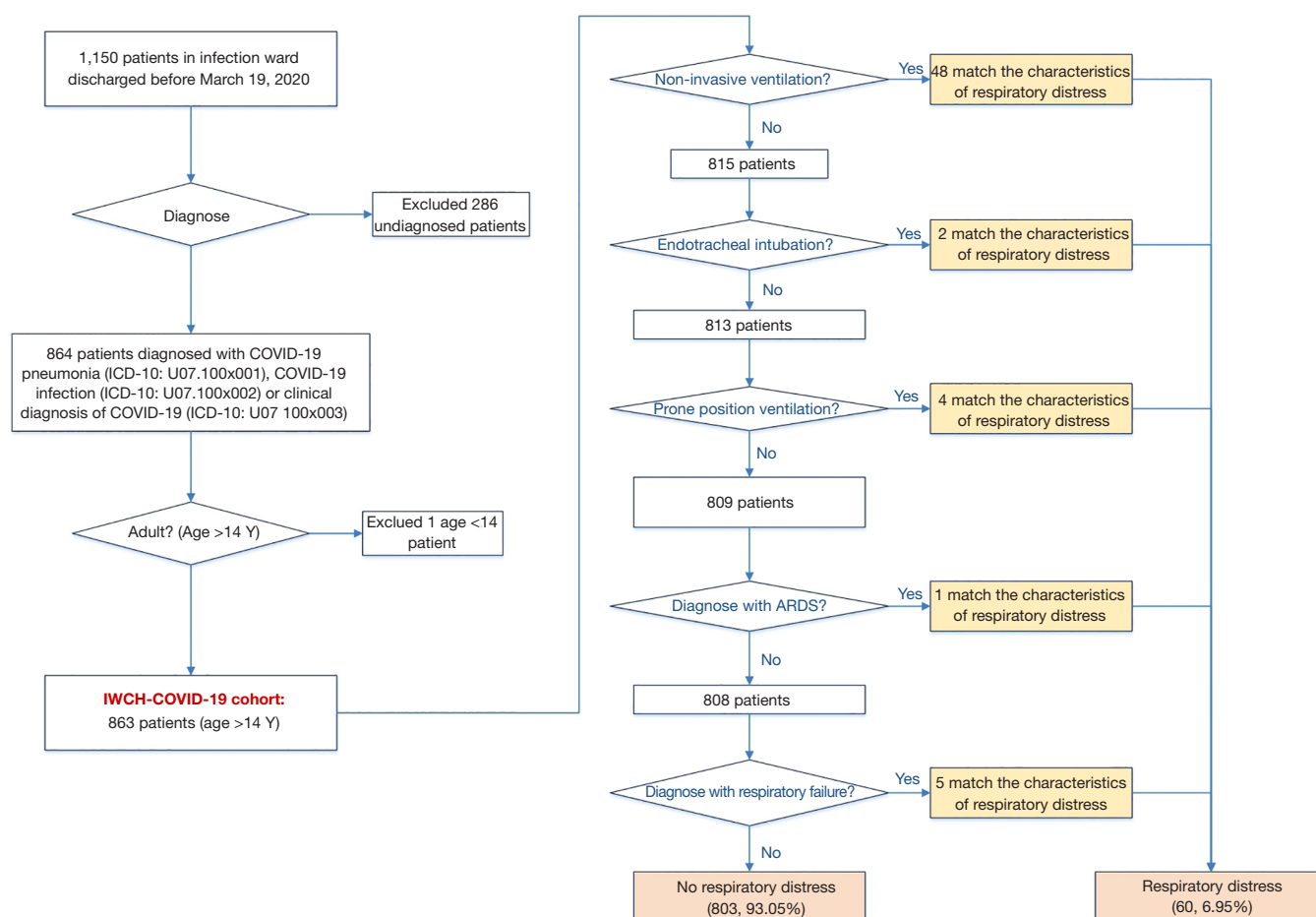


Figure 1 Data processing workflow diagram. COVID-19, coronavirus disease 2019; ARDS, acute respiratory distress syndrome.

distinguish between patients with and without respiratory distress. A higher C statistic indicated better performance, and a value of 0.5 denoted a prediction model that is not clinically useful. AIC was used to evaluate the steps used in model development. Bootstrap values were applied in the internal validation.

All statistical analyses were carried out in R software version 3.5.3. The rms, survival, survminer, and pec packages in R were used in the analysis.

Results

Participant characteristics

A total of 1,150 patients were hospitalized at the infection ward of Wuhan No. 1 Hospital. After excluding 287 (24.96%) patients that were not confirmed by COVID-19 detection as of March 19, 2020 and children (<14 years

of age, 1 patient), the remaining 863 (70.04%) patients were placed in the IWCH-COVID-19 cohort in the final analysis (Figure 1). The median age of the 863 patients was 62.0 years [interquartile range (IQR), 51.0–70.0], with a range of 16–97 years, and most patients were female, with 388 men (44.96%). There were 323 (37.43%) older people (age >65 years). Of the 863 patients, 319 suffered from comorbidities such as hypertension (228 patients, 26.42%), diabetes (89 patients, 10.31%), coronary heart disease (56 patients, 6.49%), chronic obstructive lung disease (6 patients, 0.70%), cerebrovascular disease (16 patients, 1.85%), carcinoma (11 patients, 1.29%), and chronic kidney disease (4 patients, 0.46%). IWCH-COVID-19 cohort characteristics are presented in Table 1.

Survival analysis

In the IWCH-COVID-19 cohort, 60 patients (6.95%)

Table 1 Demographics, baseline and clinical characteristics of patients with COVID-19

Variables	N. IWCH-COVID-19 cohort	P. IWCH-COVID-19 cohort	N. group NRD	P. group NRD	N. group RD	P. group RD
Demographics and clinical characteristics						
Age	59.3587	14.4664	58.75	14.4688	68.5472	10.9765
Gender						
Female	465	54.51%	446	55.75%	19	35.85%
Male	388	45.49%	354	44.25%	34	64.15%
Temperature	36.6632	0.5438	36.6471	0.5332	36.9057	0.6434
<37.3 °C	769	90.15%	729	91.12%	40	75.47%
≥37.3 °C	84	9.85%	71	8.88%	13	24.53%
SBP	131.2392	16.5038	131.1375	16.4898	132.7736	16.7982
DBP	79.8195	11.3002	79.855	11.346	79.283	10.6724
Breaths per min	20.6448	5.5607	20.5362	5.6141	22.283	4.4088
Pulse beats per min	84.7737	13.9137	84.4	13.6533	90.4151	16.5232
Comorbidity						
Hypertension	225	26.38%	209	26.12%	16	30.19%
Diabetes	89	10.31%	79	9.88%	10	18.87%
Coronary heart disease	53	6.21%	49	6.12%	4	7.55%
Chronic obstructive lung disease	6	0.7%	4	0.5%	2	3.77%
Cerebrovascular disease	14	1.64%	13	1.62%	1	1.89%
Carcinoma	11	1.29%	11	1.38%	0	0%
Chronic kidney disease	4	0.47%	4	0.5%	0	0%
Laboratory findings						
White blood cell count, ×10 ⁹ /L						
<3.5	41	4.81%	39	4.88%	2	3.77%
3.5–9.5	744	87.22%	714	89.25%	30	56.6%
>9.5	68	7.97%	47	5.88%	21	39.62%
Neutrophil count, ×10 ⁹ /L						
<1.8	59	6.92%	57	7.12%	2	3.77%
1.8–6.3	709	83.12%	683	85.38%	26	49.06%
>6.3	85	9.96%	60	7.5%	25	47.17%
Lymphocyte count, ×10 ⁹ /L						
<1.1	208	24.38%	176	22%	32	60.38%
1.1–3.2	633	74.21%	612	76.5%	21	39.62%
>3.2	12	1.41%	12	1.5%	0	0%

Table 1 (continued)

Table 1 (continued)

Variables	N. IWCH-COVID-19 cohort	P. IWCH-COVID-19 cohort	N. group NRD	P. group NRD	N. group RD	P. group RD
Albumin, g/L						
<40	726	85.11%	673	84.12%	53	100%
40–55	127	14.89%	127	15.88%	0	0%
>55	0					
Platelet count, $\times 10^9/L$						
<125	42	4.92%	30	3.75%	12	22.64%
125–350	711	83.35%	674	84.25%	37	69.81%
>350	100	11.72%	96	12%	4	7.55%
D-dimer, mg/L						
≤ 1	586	68.7%	576	72%	10	18.87%
>1	267	31.3%	224	28%	43	81.13%
High-sensitive cardiac troponin I, $\mu g/L$						
0–0.026	787	92.26%	751	93.88%	36	67.92%
>0.026	66	7.74%	49	6.12%	17	32.08%
Lactate dehydrogenase, IU/L						
<114	6	0.7%	6	0.75%	0	0%
114–250	672	78.78%	655	81.88%	17	32.08%
>250	175	20.52%	139	17.37%	36	67.92%
Creatine kinase, IU/L						
0–171	803	94.14%	763	95.38%	40	75.47%
>171	50	5.86%	37	4.62%	13	24.53%
IL-6, pg/mL						
0–5.9	296	34.7%	278	34.75%	18	33.96%
>5.9	557	65.3%	522	65.25%	35	66.04%
Hemoglobin, g/L						
<130	469	54.98%	440	55%	29	54.72%
130–175	383	44.9%	360	45%	23	43.4%
>175	1	0.12%	0	0%	1	1.89%
Serum creatinine concentration, $\mu mol/L$						
<53	188	22.04%	178	22.25%	10	18.87%
53–106	624	73.15%	587	73.38%	37	69.81%
>106	41	4.81%	35	4.38%	6	11.32%

Table 1 (continued)

Table 1 (continued)

Variables	N. IWCH-COVID-19 cohort	P. IWCH-COVID-19 cohort	N. group NRD	P. group NRD	N. group RD	P. group RD
C3, g/L						
<0.9	155	18.17%	137	17.13%	18	33.96%
0.9–1.8	698	81.83%	663	82.88%	35	66.04%
ALT, IU/L						
<9	28	3.28%	28	3.5%	0	0%
9–60	733	85.93%	689	86.12%	44	83.02%
>60	92	10.79%	83	10.38%	9	16.98%
Aspartate aminotransferase, IU/L						
<15	57	6.68%	56	7%	1	1.89%
15–40	675	79.13%	646	80.75%	29	54.72%
>40	121	14.19%	98	12.25%	23	43.4%

RD group, patients with respiratory distress before discharge (including patients who died or were transferred to another hospital); NDR group, patients with no respiratory distress before discharge (including patients who died or were transferred to another hospital). COVID-19, coronavirus disease 2019; SBP, systolic blood pressure; DBP, diastolic blood pressure; IL-6, interleukin-6; C3, complement C3; ALT, alanine aminotransferase.

developed respiratory distress within 30 days of admission. Among these, 33 of the patients who needed ventilatory support were men (62.26%) compared with 20 women (37.74%). The survival curves of eight predictors (gender, comorbidity diabetes, comorbidity carcinoma, comorbidity chronic obstructive lung disease, neutrophil count, D-dimer, platelet count, and C3) are shown in *Figure 2*, and other predictors are shown in *Figure S2*.

Risk prediction model development and performance

In univariate Cox models, gender, age, pulse, temperature, white blood cell count, neutrophil count, lymphocyte count, hemoglobin, platelet count, creatine kinase, D-dimer, C3, and high-sensitive cardiac troponin I were all statistically significant predictors ($P < 0.01$). The top 12 predictors according to the AIC value were included in a multivariate Cox model using a backward stepwise selection procedure. Five predictors—temperature, platelet count, neutrophil count, D-dimer, and C3—were found to be statistically significant ($P < 0.01$) in predicting respiratory distress in COVID-19. All these procedures and results can be found in *Table S3*.

The final five predictors were included in a multivariate

Cox model (called the final model). Neutrophil count $> 6.3 \times 10^9/L$, D-dimer ≥ 1.00 mg/L, and temperature at admission ≥ 37.3 °C had a significant positive association with outcomes of respiratory distress in the final model [HR, 8.286 (1.833–37.446), $P < 0.01$; HR, 7.835 (3.737–16.423), $P < 0.001$; and HR, 3.299 (1.673–6.503), $P < 0.001$, respectively; *Table 2*]. C3 level of 0.9–1.8 g/L, platelet count $> 350 \times 10^9/L$, and platelet count of 125 – $350 \times 10^9/L$ had a significant negative association with outcomes of respiratory distress in the final model [HR, 0.268 (0.143–0.504), $P < 0.001$; HR, 0.096 (0.028–0.326), $P < 0.001$; and HR, 0.209 (0.106–0.411), $P < 0.001$, respectively; *Table 2*]. *Figure 3* shows the forest plot of the HR.

Table 2 shows the performance of the final model with a C statistic of 0.891 (0.867–0.915) and an AIC of 567.65. The bootstrap CI of the final model was 0.866 (0.842–0.89).

Application of the model

As an example to illustrate the use of this risk model, consider a male patient aged 58 years for whom the following data were recorded at admission: temperature, 37.8 °C; platelet count, $137 \times 10^9/L$; D-dimer, 0.89 mg/L; C3, 1.1 g/L; neutrophil count, $6.5 \times 10^9/L$ (these values were obtained from

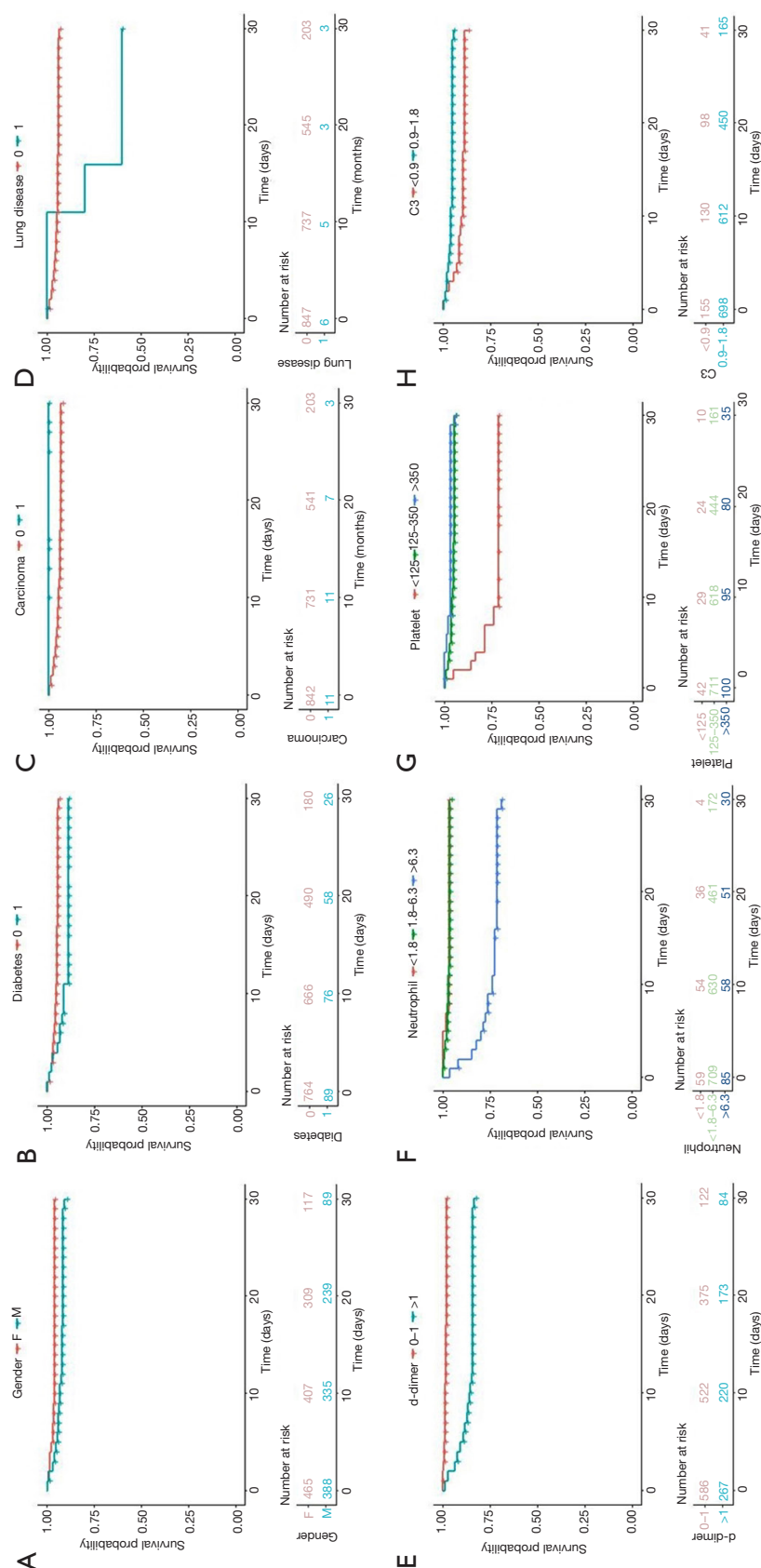


Figure 2 Kaplan-Meier curves for respiratory distress in 30 days following admission based on demographic and clinical data. (A) The median follow-up times for women (n=465) and men (n=388) were 24 (IQR, 15–30) and 23 (IQR, 15–29) days, respectively; (B) the median follow-up times for patients with (n=89) and without diabetes (n=764) were 25 (IQR, 16–30) and 24 (IQR, 15–29) days, respectively; (C) the median follow-up times for patients with (n=11) and without (n=842) carcinoma were 25 (IQR, 15.5–29) and 24 (IQR, 15–29) days, respectively; (D) the median follow-up times for patients with chronic obstructive lung disease were 23 (IQR, 12.25–30) and 24 (IQR, 15–29) days, respectively; (E) the median follow-up times for patients with D-dimer levels of 0–1.0 mg/L (n=586) and ≥ 1.00 mg/L (n=267) were 24 (IQR, 15–29) and 25 (IQR, 13–30) days, respectively; (F) the median follow-up times for patients with neutrophil counts $<1.8 \times 10^9/L$ (n=59), $1.8\text{--}6.3 \times 10^9/L$ (n=709), and $>6.3 \times 10^9/L$ (n=165) were 23 (IQR, 15–26), 24 (IQR, 15–29), and 24 (IQR, 7–30) days, respectively; (G) the median follow-up times for patients with platelet counts $<125 \times 10^9/L$ (n=42), $125\text{--}350 \times 10^9/L$ (n=711), and $>350 \times 10^9/L$ (n=100) were 23.5 (IQR, 7.50–29), 24 (IQR, 14–29), and 26 (IQR, 21–30) days, respectively; (H) the median follow-up times for patients with C3 <0.9 g/L (n=155) and between 0.9 and 1.8 g/L (n=698) were 25 (IQR, 15–29) and 24 (IQR, 14.50–30) days, respectively. IQR, interquartile range.

Table 2 Beta coefficients and SE for predictors in the final model and the performance of the final model

Predictors	Final model [$S_0(30) = 0.9783886$]			
	Parameter estimates (SE)	HR (95% CI)	P value	P for PH
Temperature at admission time ≥ 37.3 vs. < 37.3 °C	1.1518 (0.3288)	3.164 (1.661–6.027)	< 0.001	0.0005
Platelet count, $\times 10^9/L$				
>350 vs. <125	–2.9404 (0.6037)	0.053 (0.016–0.173)	< 0.001	< 0.0001
125–350 vs. <125	–1.7335 (0.3432)	0.177 (0.09–0.346)	< 0.001	< 0.0001
D-dimer, mg/L				
>1.00 vs. ≤ 1.00	2.1551 (0.3697)	8.629 (4.18–17.811)	< 0.001	< 0.0001
C3, g/L				
0.9–1.8 vs. < 0.9	–1.4017 (0.3023)	0.246 (0.136–0.445)	< 0.001	< 0.0001
Neutrophil count, $\times 10^9/L$				
>6.3 vs. < 6.3	2.7162 (0.7474)	15.122 (3.495–65.44)	< 0.001	0.0003
1.8–6.3 vs. < 6.3	0.7149 (0.7425)	2.044 (0.477–8.76)		0.3356

SE, standard error; $S_0(30)$, 30-year baseline survival; HR, hazard ratio; CI, confidence interval; PH, proportional hazard; C3, complement C3.

the first test after the patient was admitted to the hospital). For this patient, the risk is calculated as follows:

$$\text{Thirty-day respiratory distress risk} = (1 - 0.97839^{\exp[1.1518 \times (1 - 0.0985) - 1.7335 \times (1 - 0.8335) + 2.1551 \times (0 - 0.687) - 1.4017 \times (1 - 0.8183) + 2.7162 \times (1 - 0.0996)]}) \times 100\% = 8.98\%$$

Furthermore, the 15- and 30-day survival estimates can be calculated based on the nomogram of the final model shown in *Figure 4*.

Discussion

Main findings

This paper proposes a predictive risk model of respiratory distress in COVID-19 patients. Using this model for analysis, the results revealed that in the IWCH-COVID-19 cohort, a neutrophil count $> 6.30 \times 10^9/L$, D-dimer > 1.00 mg/L, temperature at admission ≥ 37.3 °C, platelet count $< 125 \times 10^9/L$, and C3 < 1.80 g/L were risk factors for predicting the likelihood of respiratory distress in COVID-19 inpatients. These five predictors yielded a statistically significant result and demonstrated useful discrimination and excellent calibration of the model.

Comparison with previous studies

Some Kaplan-Meier survival plots for different prognostic

factors have been mentioned in previous studies (12). The trends of platelet, neutrophil, and lymphocyte counts were similar to those in the plots by Chen (12). In our Kaplan-Meier survival curves, male gender, D-dimer ≥ 1.00 mg/L, and C3 < 0.90 g/L also indicated a higher risk of respiratory distress. The trends of IL-6 and comorbidity with coronary heart disease were not so obvious in our dataset.

Some previous research analyzed the risk factors for fatal outcome or critical illness in COVID-19 patients admitted to hospital. Zhou *et al.* (2) and Huang *et al.* (13) found higher D-dimer levels on admission could predict a poor prognosis. Fan *et al.* (14) indicated that a higher neutrophil count and a lower platelet count were associated with poor outcomes. Lippi *et al.* (15) and Ruan *et al.* (16) also reported that low platelet count was associated with increased risk of severe disease and mortality in patients with COVID-19. Our predictive model was consistent with the results reported in these studies. Furthermore, in a study by Ko *et al.* (17), it was concluded that Middle East respiratory syndrome patients with a lower platelet count were at higher risk of developing respiratory failure, which is consistent with the findings of our research.

Some studies have mentioned older age, comorbidities such as diabetes (18), hypertension (14,19,20), lymphocytopenia (14,21), leukocytosis, IL-6 (22), and high hypersensitive troponin I (16) to be risk predictors

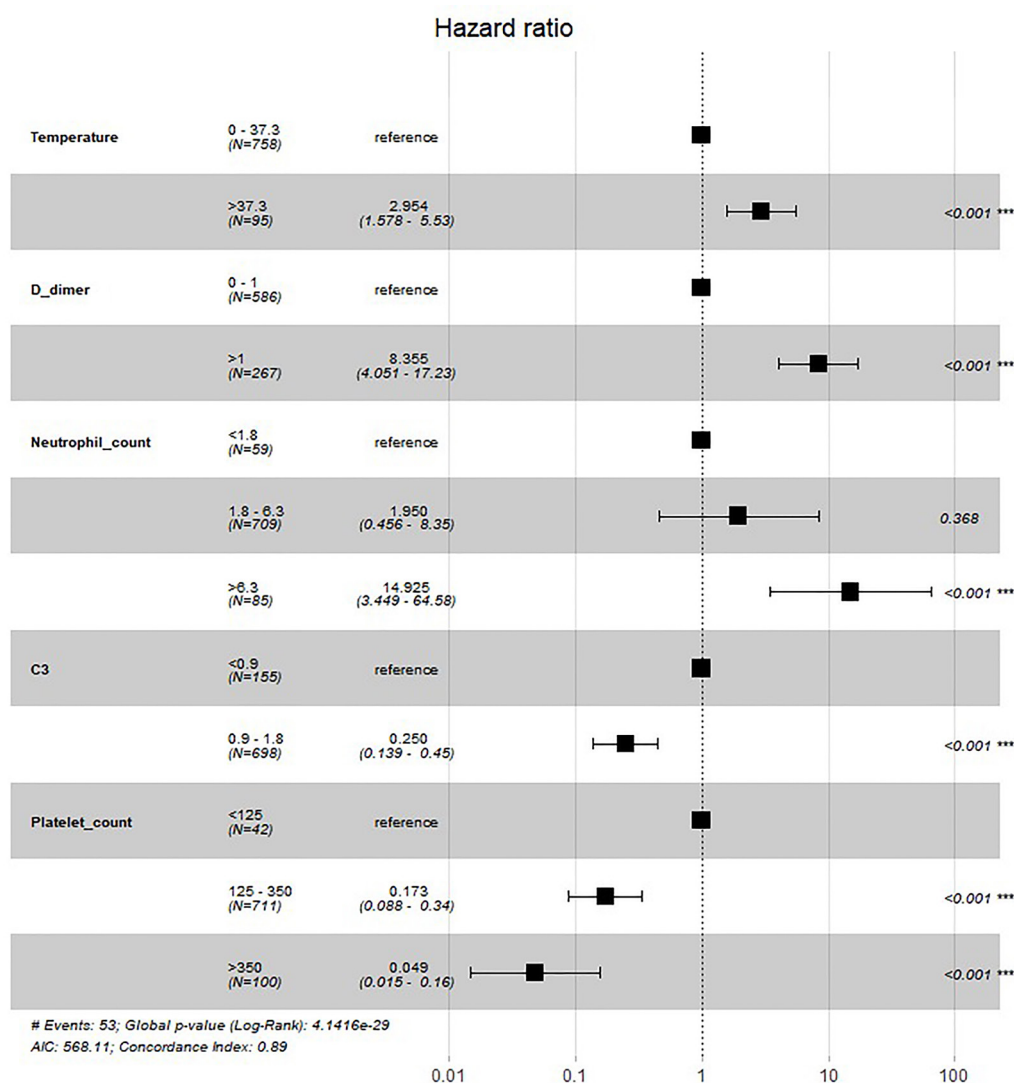


Figure 3 Forest plot for HR of the predictors. Neutrophil count $>6.3 \times 10^9/L$, HR: 8.286 (1.833–37.446), $P < 0.01$; D-dimer ≥ 1.00 mg/L, HR: 7.835 (3.737–16.423), $P < 0.001$; temperature at admission ≥ 37.3 °C, HR: 3.299 (1.673–6.503), $P < 0.001$; C3 between 0.9 and 1.8 g/L, HR: 0.268 (0.143–0.504), $P < 0.001$; platelet count $>350 \times 10^9/L$, HR: 0.096 (0.028–0.326), $P < 0.001$; platelet count of $(125-350) \times 10^9/L$, HR: 0.209 (0.106–0.411), $P < 0.001$. HR, hazard ratio; C3, complement C3; AIC, Akaike's information criterion.

for fatality. Other research made no definite conclusions on the association between diabetes and the morbidity or mortality of COVID-19 patients (23). We also used these predictors as potential predictors. Gender, age, number of breaths, pulse, comorbid diabetes, and high-sensitive cardiac troponin I >0.026 were statistically significant in our univariate Cox models. However, these factors were not included in our final multivariate Cox model.

Strengths and limitations

One strength of this study is the comprehensive risk prediction model for respiratory distress in COVID-19 patients proposed in this study. We not only investigated the presence of respiratory distress but also performed survival analysis based on the time of the respiratory distress occurring for the first time because we were able to trace the exact time of these events.

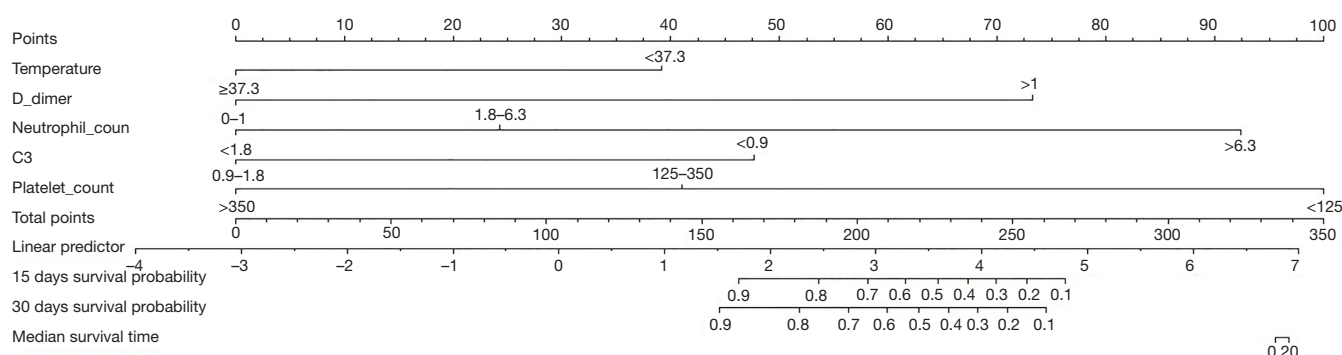


Figure 4 Nomogram of the predictive model to calculate 15- and 30-day survival. For example, a patient with temperature 37.8 °C, platelet count $137 \times 10^9/L$, D-dimer 0.89 mg/L, C3 1.1 g/L, and neutrophil count $6.5 \times 10^9/L$ has a total score of $0+41+0+48+92=181$. This is indicative of a 15- and 30-day survival of 82% and 79% respectively. C3, complement C3.

One weakness of this study is that we used data from the infection department of only one grade A hospital as we obtained authorized access to this data alone. It is hard to study a multicenter cohort that could reflect information on predictors and outcomes more completely. Meanwhile, we considered an interactive model, but could not achieve this because of the small number of positive samples. Furthermore, smoking and body mass index (BMI) have been mentioned in the literature several times as potential predictors of respiratory distress in COVID-19 patients, but due to the lack of records, smoking and BMI were not included as potential predictors in our research. Another limitation is that patients aged >80 years may not have the same reference values for their laboratory test results, as there is no standard method for analyzing them. Additionally, COVID-19 virus infection mainly affects older adults, so there are not enough samples to correct for the lack of homogeneity across age groups.

Conclusions

The predictive model generated in this study based on the factors obtained at admission can be used to calculate the risk of respiratory distress within 30 days of admission; specifically, the nomogram can be used to calculate the risk and classify patients at an early stage. The model can be very helpful for the early allocation of medical resources, attention to severe disease progression, and improved prognosis. The results may help guide the clinical management of patients with severe COVID-19, especially when limited resources need to be strategically allocated.

Acknowledgments

Funding: This work was supported by grants from the National key Research & Development Plan of the Ministry of Science and Technology of the People's Republic of China (Grant No. 2018YFC1314900, 2018YFC1314901), 2019 Provincial Special Guide Fund Project for the Development of Modern Service Industry [2019 (783)], the 2018 Projects of Jiangsu Province Department of Industry and Information Technology (Grant No. 2018419) and the 2016 Projects of Nanjing Science Bureau (Grant No. 201608003). YL is the guarantor of this paper.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <http://dx.doi.org/10.21037/atm-20-4977>

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/atm-20-4977>

Peer Review File: Available at <http://dx.doi.org/10.21037/atm-20-4977>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-20-4977>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related

to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (No. 2020-SR-163), and informed consent from study participants was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. WHO. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. 2020. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (Accessed 5 April 2020).
2. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054-62.
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;323:1061-9.
4. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;382:1708-20.
5. Emanuel EJ, Persad G, Upshur R, et al. Fair allocation of scarce medical resources in the time of Covid-19. *N Engl J Med* 2020;382:2049-55.
6. 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;8:475-81.
7. Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;180:934-43.
8. Department of Medical Administration. Diagnosis and Treatment Protocol for COVID-19 (Trial Version 7). Available online: <http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml> (Accessed May 8 2020).
9. Matthay MA, Aldrich JM, Gotts JE. Treatment for severe acute respiratory distress syndrome from COVID-19. *Lancet Respir Med* 2020;8:433-4.
10. WHO. Clinical management of COVID-19. Available online: [https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-\(ncov\)-infection-is-suspected](https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected) (Accessed April 8 2020).
11. 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.
12. Chen R, Sang L, Jiang M, et al. Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J Allergy Clin Immunol* 2020;146:89-100.
13. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497-506.
14. Fan BE, Chong VCL, Chan SSW, et al. Hematologic parameters in patients with COVID-19 infection. *Am J Hematol* 2020;95:E131-4.
15. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin Chim Acta* 2020;506:145-8.
16. Ruan Q, Yang K, Wang W, et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* 2020;46:846-8.
17. Ko JH, Park GE, Lee JY, et al. Predictive factors for pneumonia development and progression to respiratory failure in MERS-CoV infected patients. *J Infect* 2016;73:468-75.
18. Huang I, Lim MA, Pranata R. Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia - A systematic review, meta-analysis, and meta-regression. *Diabetes Metab Syndr* 2020;14:395-403.
19. Xie J, Covassin N, Fan Z, et al. Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clin Proc* 2020;95:1138-47.
20. Chen R, Liang W, Jiang M, et al. Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in China. *Chest*

- 2020;158:97-105.
21. Shi SJ, Li H, Liu M, et al. Mortality prediction to hospitalized patients with influenza pneumonia: PO₂ / FiO₂ combined lymphocyte count is the answer. Clin Respir J 2017;11:352-60.
22. Liu F, Li L, Xu M, et al. Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. J Clin Virol 2020;127:104370.
23. Hussain A, Bhowmik B, do Vale Moreira NC. COVID-19 and diabetes: knowledge in progress. Diabetes Res Clin Pract 2020;162:108142.

Cite this article as: Zhang X, Wang W, Wan C, Cheng G, Yin Y, Cao K, Zhang X, Wang Z, Miao S, Yu Y, Hu J, Huang R, Ge Y, Chen Y, Liu Y. A predictive model for respiratory distress in patients with COVID-19: a retrospective study. Ann Transl Med 2020;8(23):1585. doi: 10.21037/atm-20-4977

Table S1 Potential predictors to derive the prediction score

No.	Predictors	Terms included before backward elimination	Included in the final model 1	Included in the final model 2
1	Gender	Binary (male; female)	No	Yes
2	Age	Value (unit: year)	No	No
3	SBP	Value (unit: mmHg)	No	No
4	DBP	Value (unit: mmHg)	No	No
5	Breaths	Value (unit: per min)	No	No
6	Pulse	Value (unit: beats per min)	No	Yes
7	Temperature	Binary (=0 if temperature <37.3 °C, =1 if temperature ≥37.3 °C)	Yes	Yes
8	Comorbidity hypertension	Binary (=0 if patient had no hypertension before admission, =1 if patient had hypertension before admission)	No	No
9	Comorbidity diabetes	Binary (=0 if patient had no diabetes before admission, =1 if patient had diabetes before admission)	No	No
10	Comorbidity coronary heart disease	Binary (=0 if patient had no coronary heart disease before admission, =1 if patient had coronary heart disease before admission)	No	No
11	Comorbidity chronic obstructive lung disease	Binary (=0 if patient had no chronic obstructive lung disease before admission, =1 if patient had chronic obstructive lung disease before admission)	No	No
12	Comorbidity cerebrovascular disease	Binary (=0 if patient had no cerebrovascular disease before admission, =1 if patient had cerebrovascular disease before admission)	No	No
13	Comorbidity carcinoma	Binary (=0 if patient had no carcinoma before admission, =1 if patient had carcinoma before admission)	No	No
14	Comorbidity chronic kidney disease	Binary (=0 if patient had no chronic kidney disease before admission, =1 if patient had chronic kidney disease before admission)	No	No
15	White blood cell count	Binary (<3.50×10 ⁹ /L, 3.50–9.50×10 ⁹ /L, ≥9.50×10 ⁹ /L)	No	No
16	Neutrophil count	Binary (<1.80×10 ⁹ /L, 1.80–6.30×10 ⁹ /L, ≥6.30×10 ⁹ /L)	Yes	No
17	Lymphocyte count	Binary (<1.10×10 ⁹ /L, 1.10–3.20×10 ⁹ /L, ≥3.20×10 ⁹ /L)	No	No
18	Hemoglobin	Binary (<130 g/L, 130–175 g/L, ≥175 g/L)	No	No
19	Platelet count	Binary (<125×10 ⁹ /L, 125–350×10 ⁹ /L, ≥350×10 ⁹ /L)	Yes	Yes
20	Albumin	Binary (<40.0 g/L, 40.0–55.0 g/L, ≥55.0 g/L)	No	No
21	Aspartate aminotransferase	Binary (<15 IU/L, 15–40 IU/L, ≥40 IU/L)	No	No
22	Serum creatinine concentration	Binary (<53 μmol/L, 53–106 μmol/L, ≥106 μmol/L)	No	No
23	Creatine kinase	Binary (<171 IU/L, ≥171 IU/L)	No	No
24	High-sensitive cardiac troponin I	Binary (<0.026 μg/L, ≥0.026 μg/L)	No	No
25	D-dimer	Binary (<1.00 mg/L, ≥1.00 mg/L)	Yes	Yes
26	IL-6	Binary (<5.90 pg/mL, ≥5.90 pg/mL)	No	No
27	C3	Binary (<0.90 g/L, 0.90–1.80 g/L, ≥1.80 g/L)	Yes	Yes

SBP, systolic blood pressure; DBP, diastolic blood pressure; IL-6, interleukin-6; C3, complement C3.

Table S2 Rate of missing data

Predictors	Missing rate (%)
Gender	0.00
Age	0.00
SBP	1.64
DBP	1.76
Breaths	0.00
Pulse	0.23
Temperature	0.00
Comorbidity hypertension	0.00
Comorbidity diabetes	0.00
Comorbidity coronary heart disease	0.00
Comorbidity chronic obstructive lung disease	0.00
Comorbidity cerebrovascular disease	0.00
Comorbidity carcinoma	0.00
Comorbidity chronic kidney disease	0.00
White blood cell count	7.03
Neutrophil count	7.03
Lymphocyte count	7.03
Hemoglobin	7.03
Platelet count	7.03
Albumin	9.50
Aspartate aminotransferase	9.50
Serum creatinine concentration	10.20
Creatine kinase	30.95
D-dimer	48.77
High-sensitive cardiac troponin I	55.22
C3	89.21
IL-6	96.95

SBP, systolic blood pressure; DBP, diastolic blood pressure; C3, complement C3; IL-6, interleukin-6.

Table S3 Results of univariate Cox models and the multivariate Cox model with backward stepwise procedure

Predictors	HR (95% CI) in univariate Cox models (step 1)	P value of univariate Cox models	AIC of univariate Cox models	Included in the multivariate Cox model	HR (95% CI) in multivariate Cox model with backward stepwise procedure (step 2)	P value of multivariate Cox model	Included in the final model
Gender	2.197 (1.253–3.852)	<0.01	698.136	Yes	2.079 (1.145–3.774)	<0.05	No
Age	1.055 (1.031–1.079)	<0.001	682.16	Yes			No
SBP	1.006 (0.99–1.022)			No			No
DBP	0.997 (0.973–1.021)			No			No
Breaths	1.027 (1.002–1.052)	<0.05		No			No
Pulse	1.029 (1.012–1.047)	<0.001	696.372	Yes	1.019 (0.999–1.04)	<0.1	No
Temperature	3.122 (1.67–5.838)	<0.001	695.807	Yes	3.299 (1.673–6.503)	<0.001	Yes
Comorbidity hypertension	1.176 (0.654–2.114)			No			No
Comorbidity diabetes	1.993 (1.002–3.967)	<0.05		No			No
Comorbidity coronary heart disease	1.232 (0.445–3.415)			No			No
Comorbidity chronic obstructive lung disease	5.949 (1.446–24.465)	<0.05		No			No
Comorbidity cerebrovascular disease	1.053 (0.145–7.62)			No			No
Comorbidity carcinoma	0 (0–Inf)			No			No
Comorbidity chronic kidney disease	0 (0–Inf)			No			No
White blood cell count							
>9.5	7.342 (1.721–31.335)	<0.01	662.797	Yes			No
3.5–9.5	0.83 (0.198–3.475)						No
Neutrophil count							
>6.3	9.993 (2.364–42.244)	<0.01	654.499	Yes	8.286 (1.833–37.446)	<0.01	Yes
1.8–6.3	1.077 (0.255–4.537)				1.49 (0.348–6.379)		
Lymphocyte count							
>3.2	0 (0–Inf)				0 (0–Inf)		No
1.1–3.2	0.207 (0.119–0.359)	<0.001	674.913	Yes	0.493 (0.278–0.874)	<0.05	No
Hemoglobin							
>175	61.593 (8.034–472.199)	<0.001	701.846	No			No
130–175	0.986 (0.57–1.704)			No			No
Platelet count							
>350	0.117 (0.038–0.363)	<0.001	685.45	Yes	0.096 (0.028–0.326)	<0.001	Yes
125–350	0.16 (0.084–0.308)	<0.001	685.45	Yes	0.209 (0.106–0.411)	<0.001	Yes
Albumin							
40–55	0 (0–Inf)			No			No
Aspartate aminotransferase							
>40	11.191 (1.511–82.872)	<0.05		No			No
15–40	2.466 (0.336–18.105)			No			No
Serum creatinine concentration							
>106	2.778 (1.009–7.643)	<0.05		No			No
53–106	1.116 (0.555–2.244)			No			No
High-sensitive cardiac troponin I							
>0.026	6.311 (3.543–11.241)	<0.001	676.809	Yes	1.609 (0.843–3.072)		No
Creatine kinase							
>171	5.761 (3.081–10.775)	<0.001	684.566	Yes			No
D-dimer							
>1	9.915 (4.981–19.736)	<0.001	647.472	Yes	7.835 (3.737–16.423)	<0.001	Yes
IL-6							
>5.9	1.063 (0.602–1.877)			No			No
C3							
0.9–1.8	0.423 (0.239–0.747)	<0.01	698.217	Yes	0.268 (0.143–0.504)	<0.001	Yes

HR, hazard ratio; CI, confidence interval; AIC, Akaike's information criterion; SBP, systolic blood pressure; DBP, diastolic blood pressure; IL-6, interleukin-6; C3, complement C3.

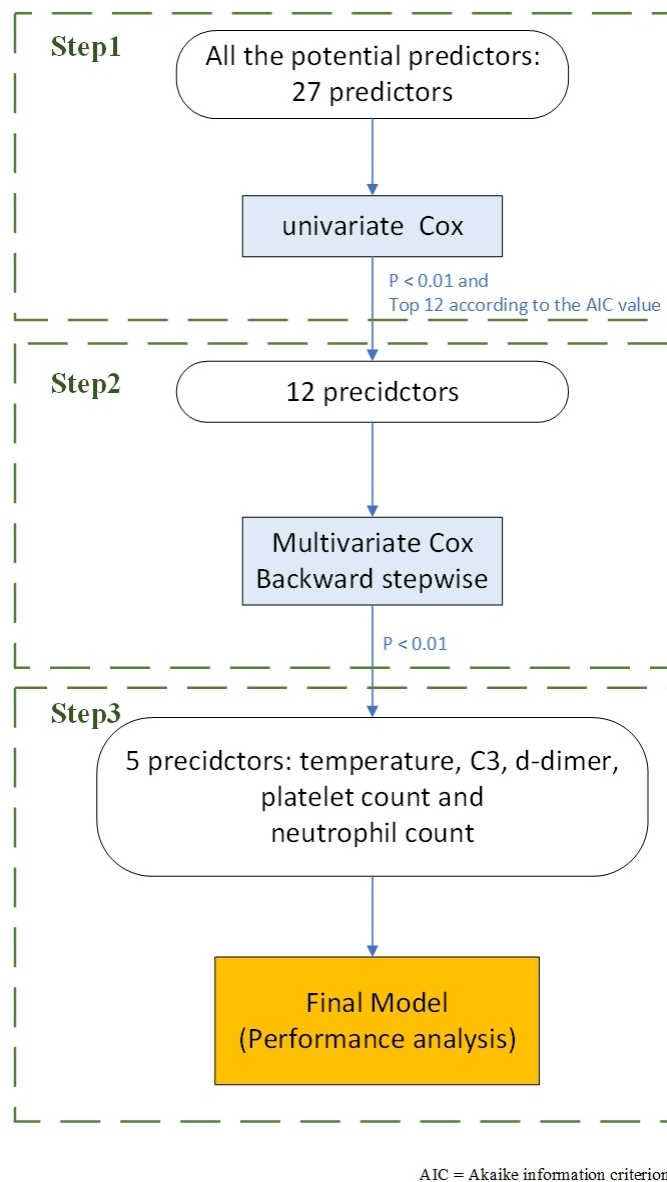


Figure S1 Workflow diagram for predictors screening. C3, complement C3.

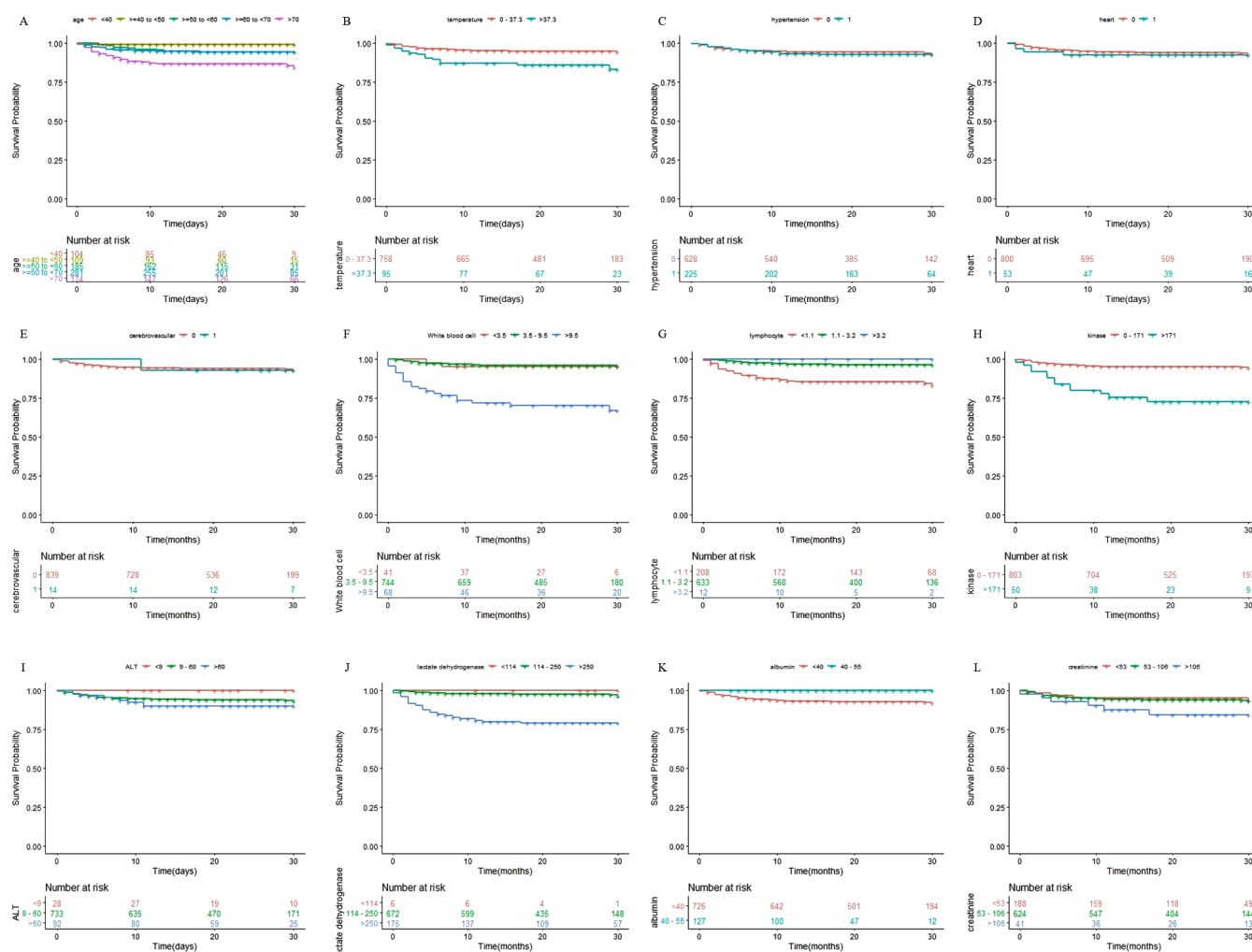


Figure S2 Kaplan-Meier curves for respiratory distress in 30 days following admission based on demographic and clinical data. (A) The median follow-up times for patients <40 years of age (n=104), 40–50 years of age (n=109), 50–60 years of age (n=185), 60–70 years of age (n=281) and >70 years of age (n=174) were 16.5 (IQR, 11–24.25), 20 (IQR, 13–27), 23 (IQR, 15–27), 25 (IQR, 19–30) and 27 (IQR, 19–30) days, respectively; (B) the median follow-up times for patients with temperature >37.3 (n=95) and with temperature ≤37.3 (n=758) were 24 (IQR, 15–29) and 24 (IQR, 18–29) days, respectively; (C) the median follow-up times for patients with (n=225) and without (n=628) hypertension were 25 (IQR, 19–30) and 23 (IQR, 14–29) days, respectively; (D) the median follow-up times for patients with (n=53) and without (n=800) coronary heart disease were 25 (IQR, 19–30) and 24 (IQR, 14.75–29) days, respectively; (E) the median follow-up times for patients with (n=14) and without (n=839) cerebrovascular disease were 29.50 (IQR, 26–30) and 24 (IQR, 15–29) days, respectively; (F) the median follow-up times for patients with white blood cell count <3.5×10⁹/L (n=41), 3.5–9.5×10⁹/L (n=744), and >9.5×10⁹/L (n=68) were 24 (IQR, 19–26), 24 (IQR, 15–29), and 22 (IQR, 7–30) days, respectively; (G) the median follow-up times for patients with lymphocyte count <1.1×10⁹/L (n=208), 1.1–3.2×10⁹/L (n=633), and >3.2×10⁹/L (n=12) were 26 (IQR, 16.75–30), 23 (IQR, 15–29), and 17.50 (IQR, 12.25–26.50) days, respectively; (H) the median follow-up times for patients with creatine kinase <171 IU/L (n=803) and ≥171 IU/L (n=50) were 24 (IQR, 15–29) and 18.5 (IQR, 10.25–27.50) days, respectively; (I) the median follow-up times for patients with ALT <9 IU/L (n=28), 9–60 IU/L (n=733), and >60 IU/L (n=92) were 22.5 (IQR, 13–30), 24 (IQR, 15–29), and 25 (IQR, 14.75–30) days, respectively; (J) the median follow-up times for patients with lactate dehydrogenase <114 IU/L (n=6), 114–250 IU/L (n=672), and >250 IU/L (n=175) were 25 (IQR, 18.25–26.75), 24 (IQR, 15–29), and 25 (IQR, 12–30) days, respectively; (K) the median follow-up times for patients with albumin <40.0 g/L (n=726), and 40.0–55.0 g/L (n=127) were 25 (IQR, 17–30), and 15 (IQR, 11–23) days, respectively; (L) the median follow-up times for patients with serum creatinine concentration <53 μmol/L (n=188), 53–106 μmol/L (n=624), and >106 μmol/L (n=41) were 24 (IQR, 13–30), 24 (IQR, 15–29), and 23 (IQR, 14–30) days, respectively. IQR, interquartile range.