

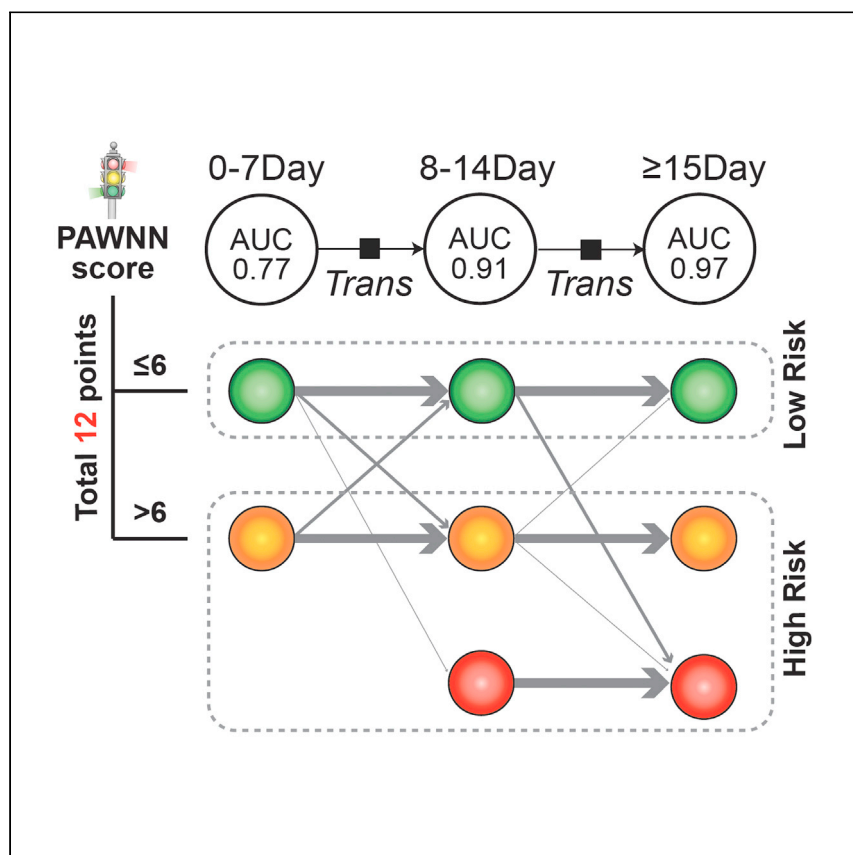


Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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

Clinical and Translational Article

Development and validation of a risk score using complete blood count to predict in-hospital mortality in COVID-19 patients



Hui Liu, Jing Chen, Qin Yang, ...,
Yibin Wang, Jingjing Cai,
Hongliang Li

caijingjing83@hotmail.com (J.C.)
lih1@whu.edu.cn (H.L.)

HIGHLIGHTS

The blood count-based PAWNN score can accurately predict mortality risk of COVID-19

PAWNN can properly stratify COVID-19 patients with limited medical resources

Liu et al. developed and validated a complete blood count-based score (PAWNN) that can predict mortality during the entire course of hospitalization in a cohort of 13,138 COVID-19 patients. The model can serve as a valuable tool to properly monitor and stratify COVID-19 patients with very limited medical resources.



Liu et al., Med 2, 435–447
April 9, 2021 © 2020 Elsevier Inc.
<https://doi.org/10.1016/j.medj.2020.12.013>



Clinical and Translational Article

Development and validation of a risk score using complete blood count to predict in-hospital mortality in COVID-19 patients

Hui Liu,^{1,5,18} Jing Chen,^{3,4,18} Qin Yang,^{7,18} Fang Lei,^{1,2,3} Changjiang Zhang,^{2,3,6} Juan-Juan Qin,^{2,3} Ze Chen,^{1,2,3} Lihua Zhu,^{2,3} Xiaohui Song,³ Liangjie Bai,³ Xuwei Huang,^{2,3} Weifang Liu,³ Feng Zhou,^{3,7} Ming-Ming Chen,^{2,3} Yan-Ci Zhao,^{2,3} Xiao-Jing Zhang,^{1,2,3} Zhi-Gang She,^{2,3} Qingbo Xu,⁸ Xinliang Ma,⁹ Peng Zhang,^{1,3,7} Yan-Xiao Ji,^{3,7} Xin Zhang,³ Juan Yang,³ Jing Xie,² Ping Ye,¹⁰ Elena Azzolini,^{11,12} Alessio Aghemo,^{11,12} Michele Ciccarelli,^{11,12} Gianluigi Condorelli,^{11,12} Giulio G. Stefanini,^{11,12} Jiahong Xia,¹³ Bing-Hong Zhang,¹⁴ Yufeng Yuan,¹⁵ Xiang Wei,¹³ Yibin Wang,¹⁷ Jingjing Cai,^{1,16,*} and Hongliang Li^{1,2,3,7,19,*}

SUMMARY

Background: To develop a sensitive risk score predicting the risk of mortality in patients with coronavirus disease 2019 (COVID-19) using complete blood count (CBC).

Methods: We performed a retrospective cohort study from a total of 13,138 inpatients with COVID-19 in Hubei, China, and Milan, Italy. Among them, 9,810 patients with ≥ 2 CBC records from Hubei were assigned to the training cohort. CBC parameters were analyzed as potential predictors for all-cause mortality and were selected by the generalized linear mixed model (GLMM).

Findings: Five risk factors were derived to construct a composite score (PAWNN score) using the Cox regression model, including platelet counts, age, white blood cell counts, neutrophil counts, and neutrophil:lymphocyte ratio. The PAWNN score showed good accuracy for predicting mortality in 10-fold cross-validation (AUROCs 0.92–0.93) and subsets with different quartile intervals of follow-up and preexisting diseases. The performance of the score was further validated in 2,949 patients with only 1 CBC record from the Hubei cohort (AUROC 0.97) and 227 patients from the Italian cohort (AUROC 0.80). The latent Markov model (LMM) demonstrated that the PAWNN score has good prediction power for transition probabilities between different latent conditions.

Conclusions: The PAWNN score is a simple and accurate risk assessment tool that can predict the mortality for COVID-19 patients during their entire hospitalization. This tool can assist clinicians in prioritizing medical treatment of COVID-19 patients.

Funding: This work was supported by National Key R&D Program of China (2016YFF0101504, 2016YFF0101505, 2020YFC2004702, 2020YFC0845500), the Key R&D Program of Guangdong Province (2020B1111330003), and the medical flight plan of Wuhan University (TFJH2018006).

INTRODUCTION

The outbreak of coronavirus disease 2019 (COVID-19) continues to escalate, with particular intensity in developing countries, where the surging demand for

Context and significance

Researchers from Wuhan, China developed a complete blood count-based risk score (PAWNN score) that can predict mortality during the entire course of hospitalization in a large cohort with 13,138 COVID-19 patients. This risk score was validated in two independent cohorts from China and Italy and in a longitudinal model. The model can serve as a valuable tool for physicians with very limited medical resources to properly monitor and stratify COVID-19 patients and to reduce mortality.



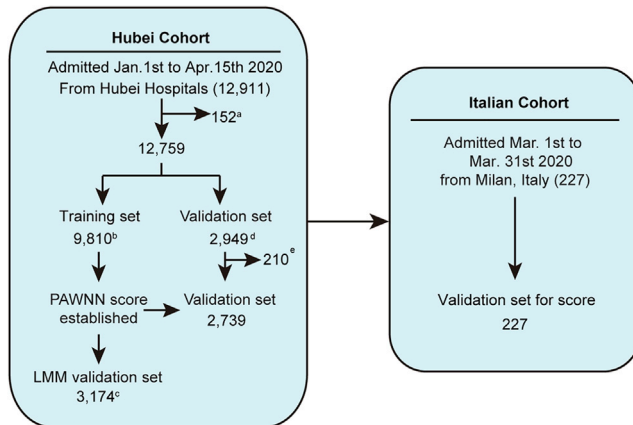


Figure 1. Flowchart for patient selection and distribution of the training and the validation cohorts

^aExcluded due to leukemia, ^bwith at least 2 complete blood count (CBC) records, ^cwith CBC records at arbitrary time points, ^dwith only 1 CBC test, and ^eremained in hospitals at the end of follow-up date. LMM, latent Markov model.

pandemic prevention and control has put a major strain on underresourced national health systems.^{1,2} Considering the heterogeneity in pathogenic manifestation among COVID-19 patients, there is an urgent need to develop an accurate and robust risk assessment tool to evaluate the disease prognosis that is also easy and economical to implement. Such a tool would help frontline clinicians to optimize medical interventions with very limited medical resources.¹ Several prognostic models for COVID-19 have been reported in the past few months. However, most of them are considered to have a high risk of bias due to deficiencies in the methodologies used.³ Furthermore, some predictors in these models rely on tests that are time-consuming and costly, thus reducing their applicability in regions with limited medical resources.³

Extensive and dynamic changes in peripheral blood cells, such as lymphopenia and neutrophilia, have been observed in patients with COVID-19, and are considered to be closely associated with the severity of the disease.^{4–6} Furthermore, complete blood count (CBC) is one of the most commonly available tests in the clinic, with minimal time and cost involved. Here, we have collected and analyzed longitudinal data about CBC from a large sample of COVID-19 cases and found that a composite score based on a few selected CBC parameters could dynamically and with high accuracy predict the risk of imminent death during hospitalization. We also revealed the longitudinal trajectories of CBC parameters and the composite score for the disease severity of COVID-19 across the duration of hospitalization. The findings of our study would be particularly helpful for optimizing clinical decision making and potentially reducing the mortality rate in countries suffering from a significant shortage of medical resources.

RESULTS

Clinical characteristics of patients from Hubei province

As shown in Figure 1, a final total of 12,759 patients (6,593 non-severe survivors, 4,246 severe survivors, 984 deaths and 936 censored) were included in this analysis from the Hubei cohort. The median follow-up day was 17 (interquartile range [IQR], 11–26). The baseline clinical characteristics, preexisting chronic diseases, and laboratory examinations at admission are described in Table 1. The median age of the

¹Department of Cardiology, Zhongnan Hospital of Wuhan University, Wuhan, China

²Department of Cardiology, Renmin Hospital of Wuhan University, Wuhan, China

³Institute of Model Animal, Wuhan University, Wuhan, China

⁴School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan, China

⁵Department of Gastroenterology, Wuhan Third Hospital & Tongren Hospital of Wuhan University, Wuhan, China

⁶The Central Hospital of Enshi Tujia and Miao Autonomous Prefecture, Enshi, China

⁷Medical Science Research Center, Zhongnan Hospital of Wuhan University, Wuhan, China

⁸Centre for Clinic Pharmacology, the William Harvey Research Institute, Queen Mary University of London, London, UK

⁹Department of Emergency Medicine, Thomas Jefferson University, Philadelphia, PA 19004, USA

¹⁰Department of Cardiology, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

¹¹Humanitas Clinical and Research Hospital IRCCS, Rozzano-Milan, Italy

¹²Department of Biomedical Sciences, Humanitas University, Pieve Emanuele-Milan, Italy

¹³Division of Cardiothoracic and Vascular Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

¹⁴Department of Neonatology, Renmin Hospital of Wuhan University, Wuhan, China

¹⁵Department of Hepatobiliary and Pancreatic Surgery, Zhongnan Hospital of Wuhan University, Wuhan

¹⁶Department of Cardiology, The Third Xiangya Hospital, Central South University, Changsha, China

¹⁷Department of Anesthesiology, Cardiovascular Research Laboratories, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

¹⁸These authors contributed equally

¹⁹Lead contact

*Correspondence: caijing83@hotmail.com (J.C.), lihl@whu.edu.cn (H.L.)

<https://doi.org/10.1016/j.medj.2020.12.013>

Table 1. Baseline characteristics of the Hubei cohort

| Variables | All (n = 12,759) | Non-severe survivor (n = 6,593) | Severe survivor (n = 4,246) | Death (n = 984) |
|---|---------------------|---------------------------------|-----------------------------|---------------------|
| Clinical characteristics of admission | | | | |
| Median age (IQR), y | 59 (46–68) | 56.0 (42.0–66.0) | 60.0 (48.0–68.0) | 70.0 (63.0–78.0) |
| Male sex, no./total no. (%) | 6,157 (48.3) | 3,016/6,593 (45.8) | 2,031/4,246 (47.8) | 648/984 (65.9) |
| Median heart rate (IQR), bpm | 84 (78–96) | 81.0 (77.0–90.0) | 94.0 (80.0–107.0) | 89.0 (78.0–101.0) |
| Median respiratory rate (IQR) | 20 (19–21) | 20.0 (19.0–20.0) | 20.0 (20.0–22.0) | 21.0 (20.0–25.0) |
| Median systolic blood pressure (IQR), mmHg | 128 (120–140) | 128.0 (120.0–140.0) | 127.0 (117.0–140.0) | 131.0 (120.0–146.0) |
| Median diastolic blood pressure (IQR), mmHg | 79 (71–87) | 80.0 (72.0–88.0) | 78.0 (70.0–85.0) | 77.0 (69.0–86.0) |
| Fever, no./total no. (%) | 9,351 (73.3) | 4,617/6,593 (70.0) | 3,278/4,246 (77.2) | 782/984 (79.5) |
| Median SpO ₂ (IQR), % | 97 (95–98) | 98.0 (96.0–99.0) | 97.0 (95.0–98.0) | 90.0 (81.0–96.0) |
| Days from symptom to hospitalization (IQR), days | 11 (7–20) | 12.0 (6.0–22.0) | 11.0 (7.0–19.0) | 10.0 (6.0–14.0) |
| Median follow-up time (IQR), days | 17 (11–26) | 16.0 (10.0–24.0) | 22.0 (14.0–32.0) | 9.0 (5.0–17.0) |
| Comorbidities on admission | | | | |
| Diabetes mellitus (any type), n (%) | 2,200 (17.2) | 852 (12.9) | 872 (20.5) | 290 (29.5) |
| Chronic obstructive pulmonary disease, n (%) | 150 (1.18) | 51 (0.8) | 50 (1.2) | 36 (3.7) |
| Hypertension, n (%) | 4,648 (36.4) | 2,020 (30.6) | 1,680 (39.6) | 559 (56.8) |
| Coronary arterial disease, n (%) | 1,168 (9.2) | 454 (6.9) | 397 (9.4) | 202 (20.5) |
| Heart failure, n (%) | 99 (0.8) | 11 (0.2) | 31 (0.7) | 46 (4.7) |
| Cerebrovascular disease, n (%) | 413 (3.2) | 148 (2.2) | 125 (2.9) | 81 (8.2) |
| Renal insufficiency, n (%) | 500 (3.9) | 168 (2.6) | 132 (3.1) | 150 (15.2) |
| Neoplastic disease, n (%) | 385 (3.0) | 180 (2.7) | 110 (2.6) | 61 (6.2) |
| Liver disease, n (%) | 276 (2.2) | 125 (1.9) | 103 (2.4) | 32 (3.3) |
| Laboratory examination on admission | | | | |
| WBC count >9.5 × 10 ⁹ /L, no./total no. (%) | 1,296/12,759 (10.2) | 320/6,593 (4.9) | 431/4,246 (10.2) | 411/984 (41.8) |
| Neutrophil count >6.3 × 10 ⁹ /L, no./total no. (%) | 1,980/12,759 (15.5) | 501/6,593 (7.6) | 695/4,246 (16.4) | 574/984 (58.3) |
| Lymphocyte count <1.1 × 10 ⁹ /L, no./total no. (%) | 5,150/12,759 (40.4) | 1,832/6,493 (27.8) | 1,958/4,246 (46.1) | 841/984 (85.5) |
| Platelet count <125 × 10 ⁹ /L, no./total no. (%) | 1,236/12,759 (9.7) | 439/6,493 (6.7) | 374/4,246 (8.8) | 304/984 (30.9) |
| CRP > ULN, no./total no. (%) ^a | 3,565/7,247 (49.2) | 1,550/4,157 (37.3) | 1,055/1,871 (56.4) | 529/542 (97.6) |
| Procalcitonin > ULN, no./total no. (%) ^a | 4,303/10,212 (42.1) | 1,444/4,927 (29.3) | 1,859/3,629 (51.2) | 696/854 (81.5) |
| Alanine transaminase >40 U/L, no./total no. (%) | 2,717/12,213 (22.3) | 1,260/6,245 (20.2) | 1,005/4,159 (24.2) | 261/949 (27.5) |
| BUN > ULN, no./total no. (%) ^a | 1,178/12,361 (9.5) | 302/6,308 (4.8) | 296/4,188 (7.1) | 449/959 (46.8) |
| CK-MB > ULN, no./total no. (%) ^a | 462/8,339 (5.5) | 138/3,870 (3.6) | 109/2,960 (3.7) | 177/748 (23.7) |
| Total cholesterol >5.17 mmol/L, no./total no. (%) | 1,302/10,301 (12.6) | 762/4,936 (15.4) | 409/3,745 (10.9) | 49/820 (6.0) |
| D-dimer > ULN, no./total no. (%) ^a | 5,497/11,384 (48.3) | 2,038/5,618 (36.3) | 2,175/3,985 (54.6) | 819/930 (88.1) |
| LDL-C >3.37 mmol/L, no./total no. (%) | 1,302/9,220 (14.1) | 731/4,604 (15.9) | 421/3,103 (13.6) | 57/720 (7.9) |

bpm, beats per minute; BUN, blood urea nitrogen; CK-MB, creatinine kinase-myocardial band; CRP, C-reactive protein; IQR, interquartile range; LDL-C, low-density lipoprotein cholesterol; SpO₂, peripheral oxygen saturation; ULN, upper limit of normal; WBC, white blood cell.
^aULN was defined according to criteria in each hospital.

participants was 59 (IQR, 46–68) years, and 48.3% were males. The median period from symptom onset to hospitalization was 11 (IQR, 7–20) days. The median peripheral oxygen saturation (SpO₂) was 97% (IQR, 95–98), and 73.3% of the patients had fever. Hypertension (36.4%), diabetes (17.2%), and coronary arterial disease (9.2%)

were the most common coexisting chronic diseases in this cohort. On admission, 40.4% of the patients showed decreased lymphocyte counts. Increased neutrophil counts and decreased platelet counts occurred in 15.5% and 9.7% of the patients, respectively. Inflammation markers, including C-reactive protein (CRP) and procalcitonin, were elevated in 49.2% and 42.1% of the patients, respectively. The elevated alanine aminotransferase (ALT), blood urea nitrogen (BUN), and creatinine kinase-myocardial band (CK-MB) indicating liver, kidney, and cardiac impairment were reported in 22.3%, 9.5%, and 5.5% of the participants, respectively. The differences in baseline characteristics among patients from non-severe survivor, severe survivor, and death groups are detailed in [Table 1](#).

Dynamic trajectories of CBC parameters

[Figure 2](#) illustrates the linear fitting curve for dynamic trajectories of 13 CBC parameters from admission to day 30 of hospitalization, grouped by disease severity. The levels of white blood cell (WBC) counts, neutrophil counts, neutrophil percentage, and neutrophil:lymphocyte ratio (NLR) increased along with the severity of the disease. These severity-correlated elevations were discernible early on at admission and persisted throughout the 30-day period of hospitalization. For all four parameters, an upward trend was observed during the first 2 weeks, followed by a decline in the later period. In contrast, the levels of lymphocyte counts, lymphocyte percentage, monocyte counts, eosinophil counts, basophil counts, and platelet counts were decreased along with the severity of the disease. The differences among the non-severe survivor, the severe survivor, and the death groups were also evident early on at admission and persisted during hospitalization. Finally, for red blood cell (RBC) counts, hematocrit, and hemoglobin concentrations, the differences were not significant on admission. However, during hospitalization, the values of these three parameters decreased steadily, and the slope and the magnitude of the decline correlated with the disease severity.

Predictor selection and score development

All of the 9,810 patients (8,311 discharged, 773 deaths, and 726 censored) in the training cohort were included for variable selection and risk score development. Variable selection process was detailed in [Figure 3](#) and [Tables S1](#) and [S2](#). A total of 38 variables—25 categorized variables and 13 continuous parameters after being scaled and centered—were used for subsequent variable selection. The generalized linear mixed models (GLMMs) with each variable as a fixed effect were ranked by the Akaike information criterion (AIC) and are reported in [Table S1](#). Comparisons of hierarchical models with different random effects (models 1, 2, and 3) are shown in [Table S2](#), and model 1, with patient-specific random slope and random intercept, was chosen due to its lowest AIC value. To further select the fixed-effect predictors with the most impact, we used multivariate GLMMs with a forward stepwise approach based on the AIC rankings in [Table S1](#). As shown in [Table S2](#), model 4 showed the middle step of the forward selection process, and model 5 was the final model selected with fixed-effects NLR (3 categories), platelet counts decrease, WBC counts increase, and neutrophil counts increase when maintaining the hierarchical structure of patient-specific random intercept and random slope. Model 6 was based on model 5, with an additional controlled age category. Age was categorized based on COVID-19 fatality ratios⁷—that is, category 1 (also as the reference category): ages 0–49 when the fatality ratio was <1%, category 2: ages 50–59, category 3: ages 60–69, and category 4: ages ≥ 70. All of the parameters remained significant, including all age categories in model 6.

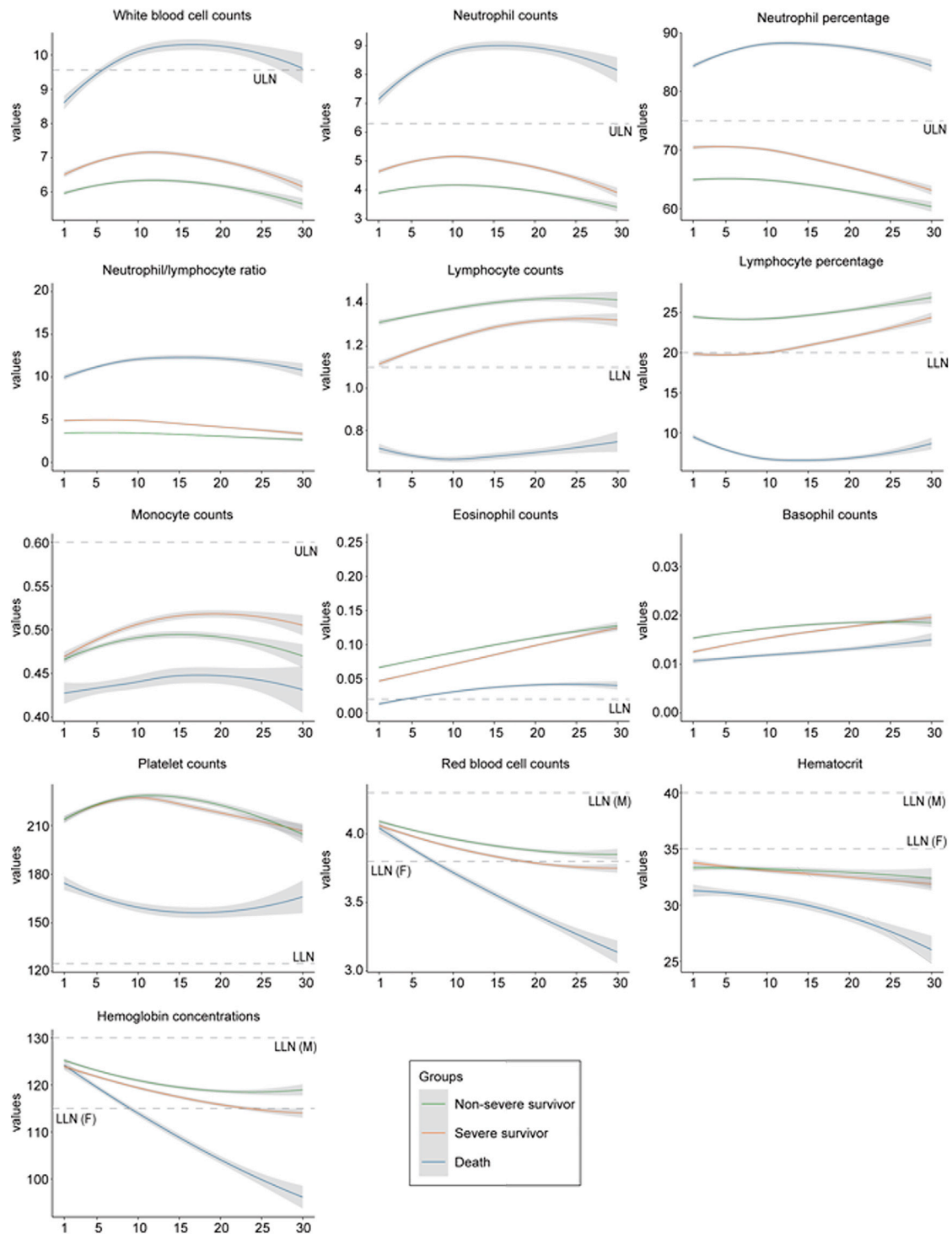


Figure 2. Dynamic trajectories of 13 CBC parameters in patients with COVID-19

Smooth trajectories of the values of CBC parameters by the severity of the disease with 95% confidence intervals were plotted based on locally weighted regression and smoothing scatterplots. The horizontal dotted lines represent the empirical upper limit of normal (ULN) or lower limit of normal (LLN) of these CBC parameters. M, males; F, females.

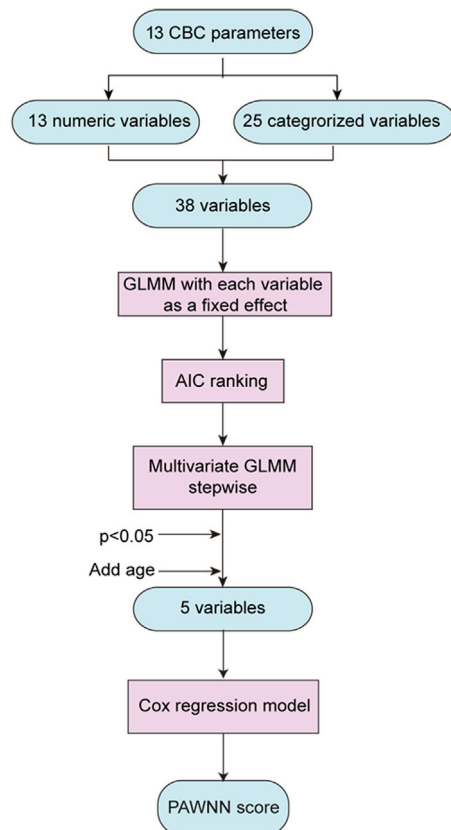


Figure 3. Flowchart for variable selection

A total of 38 CBC factors—13 numeric and 25 categorized variables—were included in the selection process. Generalized linear mixed models (GLMMs) with each variable as a fixed effect were built. The multivariate GLMM with stepwise forward selection following the Akaike information criterion (AIC) ranking established from the univariate fixed-effects models was applied. The significance levels for entry and stay were set to 0.05, and the multivariate GLMM was further controlled for age. A multivariate model with 5 variables was selected as the optimal model and was used to develop the risk assessment score.

We assigned each of the factors derived from model 6 a numeric score by rounding up the number of its specific coefficient in the Cox proportional hazards regression model. Thus, a risk-assessment scoring model based on PAWNN parameters (platelet, age, WBC counts, neutrophil counts, and NLR) was established, with the possible scores ranging from 0 to 12 points (Table 2). A PAWNN score of 6 was the cutoff value for discriminating the risk of death in the Cox model and in the Kaplan-Meier curve.

Performance of PAWNN score in the training and validation cohorts

We performed internal validation for the accuracy and specificity of the PAWNN score by 10-fold cross-validation among patients with outcomes. The area under the receiver operating characteristic (AUROC) curves based on randomly divided 10 subsets from the training cohort ranged from 0.92 (95% confidence interval [CI] 0.91–0.93) to 0.93 (95% CI 0.92–0.94) (Figure S1A). The trend plot of the PAWNN score showed a clear distinction of levels among the non-severe survivor, severe survivor, and death groups during the course of hospitalization (Figure S1B). By dividing the follow-up period into quartiles, we found that the PAWNN score remained highly

Table 2. Factors generating from GLMM and point distribution according to the coefficient in the Cox model

| Covariates | Estimates | SE | z | p | Points |
|----------------------------|-----------|------|-------|--------|--------|
| NLR > 4.06 | 3.50 | 0.11 | 14.46 | <0.001 | 6 |
| NLR 2.22–4.06 | 1.03 | 0.12 | 4.35 | <0.001 | 2 |
| Platelet counts decrease | 0.75 | 0.03 | 10.42 | <0.001 | 2 |
| Neutrophil counts increase | 0.46 | 0.04 | 7.59 | <0.001 | 1 |
| WBC counts increase | 0.15 | 0.04 | 2.86 | 0.004 | 1 |
| Age, y | | | | | |
| 50–59 | 0.51 | 0.06 | 2.76 | 0.006 | 1 |
| 60–69 | 0.47 | 0.06 | 2.85 | 0.004 | 1 |
| ≥70 | 0.75 | 0.06 | 4.68 | <0.001 | 2 |

Platelet count decrease indicates platelet count $<100 \times 10^9/L$, neutrophil count increase indicates neutrophil count $>6.3 \times 10^9/L$, and WBC count increase indicates WBC count $>9.5 \times 10^9/L$. GLMM, generalized linear mixed model; NLR, neutrophil:lymphocyte ratio; SE, standard error; WBC, white blood cell.

Cox model: $h(t;X) = h_0(t)\exp(3.5[\text{NLR} > 4.06] + 1.03[\text{NLR} 2.22\text{--}4.06] + 0.75[\text{platelet counts decrease}] + 0.46[\text{neutrophil counts increase}] + 0.15[\text{WBC counts increase}] + 0.51[\text{age } 50\text{--}59] + 0.47[\text{age } 60\text{--}69] + 0.75[\text{Age} \geq 70])$.

accurate for predicting mortality at different time intervals. The lowest AUROC, 0.89 (95% CI 0.88–0.90), was observed for 0–1 day after hospital admission, with a cutoff value of 6 points, and the highest AUROC, 0.94 (95% CI 0.93–0.94/0.95), was reached after 8 days after admission, with a cutoff value of 6 points (Figure S1C; Table S3).

In the validation dataset of 2,949 patients (2,528 discharged, 211 died, and 210 censored) from Hubei Province with only a single CBC test during hospitalization, the PAWNN score had an AUROC of 0.97 (95% CI 0.96–0.98), a sensitivity of 93.84% (95% CI 90.51–98.10), and a specificity of 90.90% (95% CI 85.13–92.84) (Table 3). The performance of the PAWNN score was further tested in a cohort of COVID-19 patients from Milan, Italy, where CBC data were collected at admission. Baseline characteristics of this cohort were described in Table S4. The predictability performance in the Italian cohort remained high, with an AUROC of 0.80 (95% CI 0.74–0.86), a sensitivity of 68.83% (95% CI, 58.44–94.81), and a specificity of 80.67% (95% CI, 49.33–87.33) (Table 3).

Latent Markov Model (LMM)

The LMM consists of a structural model for the latent disease status and a measurement model for the observed indicators, which are the parameters selected in GLMMs. By fitting the proposed model with the LMest package in R,⁸ we found that the model with three statuses resulted in the lowest AIC (Table S5). In the following, we report the results obtained with this number of statuses.

The parameters' effects on the logit of initial probabilities are presented in Table S6. Table S7 reports the model initial probability configurations, which indicate that the two outcome groups performed similarly at the early phase of the test. These probabilities allowed us to characterize the latent status. The first latent status (LS 1) (low-risk group) and the second status (LS 2) (medium-risk group) showed the highest probability to respond to survived outcomes, and the third status (LS 3) (high-risk group) response was mostly to dead outcomes. These probability distributions represent response strategies as a function of the latent component or LS of the disease. In particular, in our context, LS 1 may be easily understood as a safety status in which LS 2 seems to characterize a type of endangered status. Finally, LS 3 indicates a nearly dying disease status.

Table 3. Performance of PAWNN score in Wuhan and Italian validation datasets

| | Hubei participants with 1 CBC test | Italian participants |
|----------------------------|------------------------------------|----------------------|
| No. patients | 2,739 | 227 |
| AUROC (95% CI) | 0.97 (0.96–0.98) | 0.80 (0.74–0.86) |
| Total accuracy, % (95% CI) | 91.13 (86.05–92.85) | 76.65 (65.20–81.94) |
| Sensitivity, % (95% CI) | 93.84 (90.51–98.10) | 68.83 (58.44–94.81) |
| Specificity, % (95% CI) | 90.90 (85.13–92.84) | 80.67 (49.33–87.33) |
| PPV, % (95% CI) | 46.30 (35.27–52.01) | 64.77 (49.32–73.97) |
| NPV, % (95% CI) | 99.44 (99.13–99.81) | 83.67 (79.00–93.94) |

AUROC, area under the subject operating characteristic curve; CBC, complete blood count; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

The structure of the three-status model with the prevalence of each status is shown in [Figure 4](#). The means and standard deviations (SDs) of the PAWNN score were calculated by status to evaluate the ability of the PAWNN score to identify LS at different time points. These represent, at the given time point, the probability of transiting from a current status LS to a different status LS or remaining in the same status LS.

The transition trajectories showed a clear pattern that all patients with dead outcomes either went through the medium-risk to the high-risk process or directly from the low-risk group, whereas fewer patients with medium risk transferred back to the low-risk status. The transition probabilities from the low risk at time 1 to the medium-risk and high-risk status at time 2 were 18% and 1%, respectively. The transition probabilities from the medium risk at time 1 to the low-risk group at time 2 were 27%. There were 17% patients at the low risk and 1% at the medium risk at time 2 transit to the high-risk group at time 3.

In general, the PAWNN score was significantly higher for the high-risk group (LS 3) compared to the survived at all time points. The PAWNN score was a good predictor of death at all 3 time points, with the AUROC 0.77 at time 1, 0.91 at time 2, and 0.97 at time 3 ([Tables S8](#) and [S9](#)). The PAWNN score also demonstrated good discrimination power for time 2 LS, with an AUROC of 0.86 and time 3 LS, with an AUROC of 0.81. For prediction transition probability, an AUROC value of 0.86 was attributed from the time 1 PAWNN score on time 2 LS, as well as an AUROC of 0.81 from the time 2 PAWNN score on time 3 LS ([Table S9](#)).

DISCUSSION

In this study, we have developed and validated a composite score (PAWNN score) using age and 4 CBC parameters (platelet, WBC counts, neutrophil counts, and NLR) to predict mortality in patients with COVID-19 during hospitalization. The performance of this risk score was satisfactory in accuracy based on AUROCs in both the training and the validation cohorts from China and an external validation cohort from Italy. Therefore, the PAWNN score may serve as an accurate and reliable tool to quickly quantify the risk of imminent death across different clinical cohorts. Since the CBC is among the most commonly available and low-cost tests, the PAWNN score can be readily determined and implemented as a very simple and economic tool to prioritize patients for physicians with diverse backgrounds and specialties quickly. In the subgroup analysis, we showed that the AUROC was lower in the patients admitted before February 12 compared to those admitted after February 12 ([Table S10](#)). It is suggested that the limited medical resources and lack of therapeutic experience may disrupt proper treatment for patients with COVID-19 before February 12 (the first peak day of daily new cases in China). While the performance

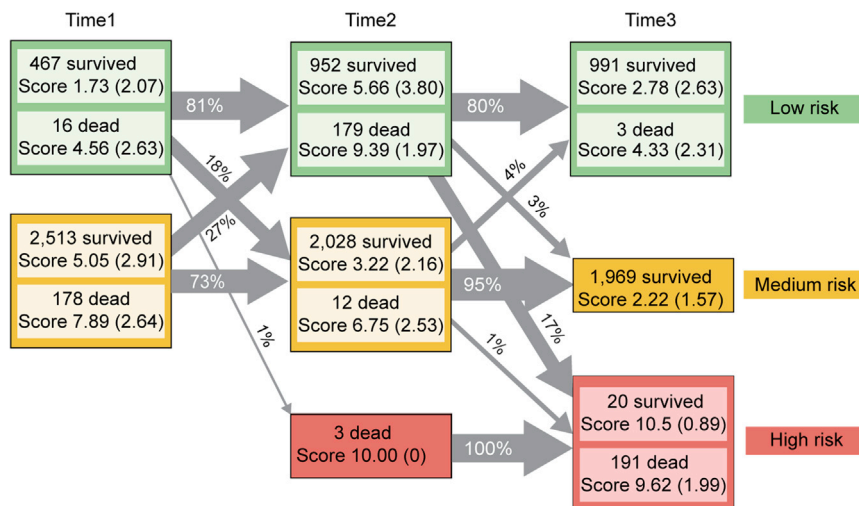


Figure 4. Status prevalence and transition probabilities between the statuses at subsequent time points

The number of survivors and non-survivors with their means and standard deviations of the PAWNN score of each status at each time point are recorded in the boxes. The transition probabilities represent the probability that a member of a given status at a specified time point will transition to another given status at the next time point. Transition probabilities are represented by arrows. Time 1, time 2, and time 3 were selected based on days 0–7 after admission (time 1), value of CBC records during days 8–14 (time 2) of hospitalization, and CBC during hospitalization ≥ 15 days (time 3).

of the PAWNN score was stable in patients with different age, gender, and preexisting comorbidities (Tables S10 and S11), a PAWNN score of 6 demonstrated good performance in discriminating the risk of death in all of the groups (Figure S2). If a patient's predicted risk for death is low (e.g., PAWNN score ≤ 6), the physician may choose to monitor at peripheral or district hospitals, while a high-risk estimate may support more aggressive intervention or early transfer to tertiary centers and admission to the intensive care unit. Our PAWNN score is particularly helpful for physicians to allocate limited resources to the most needy in areas with less advanced healthcare systems during a surge of COVID-19.

In Table S3, we categorize patients with CBC tests into quartiles according to the number of hospital days from admission. The death rate was higher in the first two quartiles but lower in the last two quartiles. A possible explanation for this phenomenon may be associated with the stringent rules for patient discharge in China. After COVID-19-associated symptoms are significantly relieved, hospital discharge is authorized only after 2 consecutive negative viral PCR tests. Thus, patients who survive have a long hospital stay. In the first quartile, there was a higher proportion of sick patients and higher mortality. The PAWNN score can dynamically predict mortality during the entire course of hospitalization. At different intervals during the hospital stay, the PAWNN score shows high predictive accuracy for death based on ROC curves. In the clinical setting, evaluation at admission decides medical resource allocation. The PAWNN score demonstrates decent but slightly lower prediction accuracy in the earliest phase of hospitalization, which indicates a larger variability in the earlier phases of the disease; the closer to the outcome, the more accurate the PAWNN score in its prediction. Furthermore, using a data-driven approach (i.e., LMM), we defined longitudinal statuses of COVID-19 patients based on selected CBC parameters. The PAWNN score was a distinguishing feature of the defined

statuses at each time point. LMM also allowed us to examine the transition probabilities between statuses over time simultaneously. One of the most interesting findings of this analysis is the existence of transitioning from the low-risk status at time 1 to the medium-risk status at time 2 and then to the high-risk status at time 3. These transitioning trajectories are correlated with a marked increase in mortality risk, and the PAWNN score showed good predictive accuracy for class transitioning at the subsequent time point. Thus, early prediction of transition probabilities based on the current PAWNN score can help to modify interventions in advance. This further supports the significance of the PAWNN score for dynamic monitoring of prognosis in COVID-19 patients during the entire course of hospitalization.

Our study systematically analyzed the dynamic trajectories of 13 CBC parameters in patients with different severities of COVID-19. In general, the temporal patterns of changes in all CBC parameters are distinctly different between non-survivors and survivors. Moreover, the dynamic changes in all parameters are closely associated with the course of disease progression. These data can help us assess the status of the disease in COVID-19 patients during hospitalization.

The identified predictors in PAWNN score from our study have been implicated in several previous studies as potential risk factors for mortality or severe illness related to COVID-19. Liu et al.⁹ have proposed that NLR on admission could serve as an independent predictor of in-hospital mortality for COVID-19 patients. Both lymphocytopenia and neutrophilia were often observed in earlier reports of COVID-19 symptoms, which showed more prominent manifestations among non-survivors versus survivors.^{5,10} An overt lymphocytopenia suggests that lymphocyte deficiency or incapacity is a critical cellular pathology of COVID-19, and less robust immune responses following severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) infection may contribute to disease progression. Moreover, previous studies demonstrate that ~7–14 days from symptom onset, there is a surge in the clinical manifestations of COVID-19-related complications along with a pronounced systemic increase of inflammatory mediators and cytokines, the so-called cytokine storm.⁴ Neutrophils are the main source of chemokines and cytokines.¹¹ In our study, neutrophil counts in the high-risk group were above the normal range throughout hospitalization, which reflects a sustained status of inflammatory overactivation. This may induce the cytokine storm and subsequently contribute to the development of acute respiratory distress syndrome (ARDS) and death. In our study, the levels of platelet counts in non-survivors were substantially lower than those in survivors, which agreed with the findings of Zhou et al.⁶ and a recent meta-analysis.¹² The mechanisms of thrombocytopenia in coronavirus infections may be multifactorial, but a surge in platelet consumption in response to endothelial damage caused by coronavirus infection and mechanical ventilation is plausible.¹³

Currently, the severity and risk of death in patients with COVID-19 are often graded with the use of only a single parameter, such as lymphopenia. Tan et al.¹⁴ analyzed the time courses of lymphocyte counts in a small sample of COVID-19 patients and proposed a model that found that patients with <20% lymphocytes at days 10–12 from illness onset and <5% at days 17–19 have the worst prognosis. Nevertheless, this model was based on a small sample size and lacked the minimal level of test and validation of its robustness. Furthermore, a single parameter may not adequately reflect disease progression at the function level. Thus, a sensitive and quantitative composite score may be more valuable for risk stratification. Recently, Liang et al.¹⁵ developed a clinical risk score (COVID-GRAM) to predict the occurrence of critical illness based on 1,590 hospitalized patients with COVID-19.

However, some of the parameters in this score, such as chest radiography, lactate dehydrogenase, and direct bilirubin, require additional time and cost, which undermines its application as a simple, economic, yet broadly applicable tool in regions with scarce medical resources. Furthermore, this score was based only on variables at hospital admission; therefore, it may not be dynamically monitored during the course of hospitalization.

In this study, we developed and validated a readily applicable risk assessment tool, the PAWNN score, to dynamically estimate the risk of imminent death for patients with COVID-19 during the course of hospitalization by using CBC parameters. As CBC is the most commonly available test, the PAWNN score may assist frontline clinicians from diverse backgrounds and specialties to optimize the use of limited medical resources in areas experiencing a surge in COVID-19.

Limitations of study

Although our study developed and validated a simple and dynamically applicable composite risk score based on a large retrospective in-hospital cohort of COVID-19, several limitations should be noted in interpreting the results. First, our study was based on retrospectively collected CBC parameters and focused on prediction. Hence, no causal conclusions could be drawn from our algorithm. Second, this simple CBC model could be a convenient tool for stratifying patients at risk in the clinic; however, it may have a limited ability to explain all of the variances across patient populations. Third, repeated tests for CBC were carried out at different time intervals for each patient; thus, bias may occur due to an increased number of tests in patients with severe illness. Fourth, due to the limited longitudinal data for patients in the external validation sets, we were not able to validate the PAWNN score indicating the probability of transition between classes during hospitalization. Fifth, the patient cohort was recruited from hospitalized patients from only one province in China, and the sample size in the Italy cohort was relatively small. Whether this model could be generalizable to outpatients and patients with different genetic and geographic backgrounds would require further external validation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Participants and Study Procedure
- METHOD DETAILS
 - Data Collection and Definitions
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Statistical Analysis
 - Predictor Selection and Model Development
 - Score Development and Validation
 - Latent Markov Model

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.medj.2020.12.013>.

ACKNOWLEDGEMENTS

This work was supported by grants from National Key R&D Program of China, China (2016YFF0101504, 2016YFF0101505, 2020YFC2004702, 2020YFC0845500), the Key R&D Program of Guangdong Province, China (2020B1111330003), and the medical flight plan of Wuhan University, China (TFJH2018006).

AUTHOR CONTRIBUTIONS

H. Liu, J. Chen, and Q.Y. designed the study, collected and analyzed the data, and wrote the manuscript. C.Z., J.-J.Q., Z.C., L.Z., X.S., L.B., X.H., W.L., M.-M.C., Y.-C.Z., Q.X., X.M., P.Z., Y.-X.J., X.Z., J.Y., J. Xia, P.Y., J. Xie, B.-H.Z., Y.Y., and X.W. collected, reviewed, interpreted, and checked the clinical, laboratory, and radiological data. J. Chen, F.L., and F.Z. performed the statistical analysis. X.-J.Z., Z.-G.S., and Y.W. wrote the manuscript and provided valuable suggestions for study design and data analysis. M.C., E.A., A.A., G.G.S., and G.C. collected, reviewed, and analyzed the clinical data in Italy and provided valuable suggestions for study design. J. Cai and H. Li contributed equally, designed the project, edited the manuscript, and supervised the study. All of the authors approved the final version of this article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 6, 2020

Revised: November 10, 2020

Accepted: December 21, 2020

Published: January 8, 2021

REFERENCES

- Ayebare, R.R., Flick, R., Okware, S., Bodo, B., and Lamorde, M. (2020). Adoption of COVID-19 triage strategies for low-income settings. *Lancet Respir. Med.* 8, e22.
- Lloyd-Sherlock, P., Ebrahim, S., Geffen, L., and McKee, M. (2020). Bearing the brunt of COVID-19: older people in low and middle income countries. *BMJ* 368, m1052.
- Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M.J., Damen, J.A.A., Debray, T.P.A., De Vos, M., et al. (2020). Prediction models for diagnosis and prognosis of COVID-19 infection: systematic review and critical appraisal. *BMJ* 369, m1328.
- Terpos, E., Ntanasis-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T.N., Politou, M., Psaltopoulou, T., Gerotziafas, G., and Dimopoulos, M.A. (2020). Hematological findings and complications of COVID-19. *Am. J. Hematol.* 95, 834–847.
- Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., et al. (2020). Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern. Med.* 180, 934–943.
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395, 1054–1062.
- Ruan, S. (2020). Likelihood of survival of coronavirus disease 2019. *Lancet Infect. Dis.* 20, 630–631.
- Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017). LMest: An R package for latent Markov models for longitudinal categorical data. *J. Stat. Softw.* 81, <https://doi.org/10.18637/jss.v081.i04>.
- Liu, J., Liu, Y., Xiang, P., Pu, L., Xiong, H., Li, C., Zhang, M., Tan, J., Xu, Y., Song, R., et al. (2020). Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J. Transl. Med.* 18, 206.
- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in wuhan, china. *JAMA* 323, 1061–1069.
- Channappanavar, R., and Perlman, S. (2017). Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin. Immunopathol.* 39, 529–539.
- Lippi, G., Plebani, M., and Henry, B.M. (2020). Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin. Chim. Acta* 506, 145–148.
- Yang, M., Ng, M.H.L., and Li, C.K. (2005). Thrombocytopenia in patients with severe acute respiratory syndrome (review). *Hematology* 10, 101–105.
- Tan, L., Wang, Q., Zhang, D., Ding, J., Huang, Q., Tang, Y.Q., Wang, Q., and Miao, H. (2020). Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduct. Target. Ther.* 5, 33.
- Liang, W., Liang, H., Ou, L., Chen, B., Chen, A., Li, C., Li, Y., Guan, W., Sang, L., Lu, J., et al. (2020). Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern. Med.* 180, 1081–1089.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York: Springer-Verlag).
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, 77.
- Bates, D., Machler, M., Bolker, B.M., and Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.

19. Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression*, Third Edition (Thousand Oaks, CA: Sage).
20. Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *J. Stat. Softw.* *8*, 1–27.
21. Fox, J., and Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *J. Stat. Softw.* *32*, 1–24.
22. Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017). *Lmest: An r package for latent markov models for longitudinal categorical data*. *J. Stat. Softw.* *81*, 1–28.
23. Saefken, B., Ruegamer, D., Kneib, T., and Greven, S. (2018). Conditional Model Selection in Mixed-Effects Models with cAIC4. *arXiv*, 1803.05664.
24. Fop, M., Smart, K.M., and Murphy, T.B. (2017). Variable selection for latent class analysis with application to low back pain diagnosis. *Ann. Appl. Stat.* *11*, 2080–2110.
25. Dean, N., and Raftery, A.E. (2010). Latent class analysis variable selection. *Ann. Inst. Stat. Math.* *62*, 11–35.
26. Therneau, T.M., and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model* (New York: Springer).
27. World Health Organization (2020). Laboratory testing for 2019 novel coronavirus (2019-ncov) in suspected human cases interim guidance. <http://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance-publications>.
28. National Health Commission of China (2020). *New Coronavirus Pneumonia Prevention and Control Program, Fifth Edition* (National Health Commission of China).
29. Zhang, P., Zhu, L., Cai, J., Lei, F., Qin, J.J., Xie, J., Liu, Y.M., Zhao, Y.C., Huang, X., Lin, L., et al. (2020). Association of inpatient use of angiotensin converting enzyme inhibitors and angiotensin II receptor blockers with mortality among patients with hypertension hospitalized with COVID-19. *Circ. Res.* *126*, 1671–1681.
30. Liu, Y.M., Xie, J., Chen, M.M., Zhang, X., Cheng, X., Li, H., Zhou, F., Qin, J.J., Lei, F., Chen, Z., et al. (2020). Kidney function indicators predict adverse outcomes of COVID-19. *Med (N Y)*. <https://doi.org/10.1016/j.medj.2020.09.001>.
31. Cheng, X., Liu, Y.M., Li, H., Zhang, X., Lei, F., Qin, J.J., Chen, Z., Deng, K.Q., Lin, L., Chen, M.M., et al. (2020). Metformin is associated with higher incidence of acidosis, but not mortality, in individuals with COVID-19 and pre-existing type 2 diabetes. *Cell Metab.* *32*, 537–547.e3.
32. Zhang, X.J., Qin, J.J., Cheng, X., Shen, L., Zhao, Y.C., Yuan, Y., Lei, F., Chen, M.M., Yang, H., Bai, L., et al. (2020). In-hospital use of statins is associated with a reduced risk of mortality among individuals with COVID-19. *Cell Metab.* *32*, 176–187.e4.
33. Zhu, L., She, Z.G., Cheng, X., Qin, J.J., Zhang, X.J., Cai, J., Lei, F., Wang, H., Xie, J., Wang, W., et al. (2020). Association of blood glucose control and outcomes in patients with COVID-19 and pre-existing type 2 diabetes. *Cell Metab.* *31*, 1068–1077.e3.
34. Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining r^2 from generalized linear mixed-effects models. *Methods Ecol. Evol.* *4*, 133–142.
35. Johnson, P.C.D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods Ecol. Evol.* *5*, 944–946.
36. Ma, Y., Shi, N., Fan, Y., Wang, J., Zhao, C., Li, G., Zhang, X., Bai, W., Xiao, F., Wang, J., et al. (2020). Predictive Value of the Neutrophil-to-Lymphocyte Ratio (NLR) for Diagnosis and Worse Clinical Course of the COVID-19: Findings from Ten Provinces in China. <https://ssrn.com/abstract=3569838>.
37. Hsieh, F.I., Lien, L.M., Chen, S.T., Bai, C.H., Sun, M.C., Tseng, H.P., Chen, Y.W., Chen, C.H., Jeng, J.S., Tsai, S.Y., et al.; Taiwan Stroke Registry Investigators (2010). Get With the Guidelines-Stroke performance indicators: surveillance of stroke care in the Taiwan Stroke Registry: Get With the Guidelines-Stroke in Taiwan. *Circulation* *122*, 1116–1123.
38. Jha, A.K., Orav, E.J., Li, Z., and Epstein, A.M. (2007). The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Aff. (Millwood)* *26*, 1104–1110.
39. Heagerty, P.J., Lumley, T., and Pepe, M.S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* *56*, 337–344.
40. Heagerty, P.J., and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* *61*, 92–105.
41. Hand, D.J., and Till, R.J. (2001). A simple generalisation of the area under the roc curve for multiple class classification problems. *Mach. Learn.* *45*, 171–186.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------------|---|---|
| Software and algorithms | | |
| R-3.6.3 | R Foundation for Statistical Computing | https://www.r-project.org/ |
| Adobe illustrator CC 2019 | Adobe company | https://www.adobe.com/cn |
| Ggplot-3.3.2 | Wickham ¹⁶ | https://cran.r-project.org/web/packages/ggplot2/index.html |
| pRoc-1.16.2 | Robin et al. ¹⁷ | https://cran.r-project.org/web/packages/pROC/index.html |
| lme4- 1.1-26 | Bates et al. ¹⁸ | https://cran.r-project.org/web/packages/lme4/index.html |
| Caret- 6.0-86 | Max Kuhn | https://cran.r-project.org/web/packages/caret/index.html |
| Data.table- 1.13.4 | Matt Dowle | https://cran.r-project.org/web/packages/data.table/ |
| Effects- 4.2-0 | Fox and Weisberg, ¹⁹ Fox, ²⁰ Fox and Hong ²¹ | https://cran.r-project.org/web/packages/effects/ |
| Lmest-3.0.1 | Bartolucci et al. ²² | https://cran.r-project.org/web/packages/LMest/index.html |
| cAIC4-0.9 | Saefken et al. ²³ | https://cran.r-project.org/web/packages/cAIC4/index.html |
| LCAvarese-1.1 | Fop et al., ²⁴ Dean and Raftery ²⁵ | https://cran.r-project.org/web/packages/LCAvarese/index.html |
| Survival- 3.2-7 | Therneau and Grambsch ²⁶ | https://cran.r-project.org/web/packages/survival/index.html |
| Survminer-0.4.8 | Alboukadel Kassambara | https://cran.r-project.org/web/packages/survminer/index.html |

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Hongliang Li (lihl@whu.edu.cn).

Materials availability

The study did not generate any new reagents or materials.

Data and code availability

Data and codes involved in this research are available from the corresponding author upon reasonable request. The research team will provide an email address for communication once the information sharing is approved. The proposal should include detailed aims, statistical plan, and other information/materials to guarantee the rationality of requirement and the security of the data. The related patient data will be shared after review and approval of the submitted proposal and any related requested materials. Of note, data with patient names, national identification number, and other identifiers cannot be shared.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants and Study Procedure

We performed a retrospective cohort study from a total of 13,138 in-hospital patients with confirmed COVID-19 and at least one CBC test in Hubei Province, China, and Milan, Italy. Among them, 12,911 patients with COVID-19 who were

admitted to 16 COVID-19 designated hospitals in Hubei province, China from January 1st, 2020 to April 15th, 2020, and had CBC recorded at least once during hospitalization were initially evaluated for inclusion. The Italian cohort initially included 227 patients with confirmed COVID-19 admitted between March 1st, 2020, and March 31st, 2020, in Humanitas Research Hospital in Milan, Italy. All patients were consecutively enrolled in the study in each designated hospital. COVID-19 was diagnosed by clinical manifestations, computerized chest tomography (CT), or reverse transcription-polymerase chain reaction (RT-PCR) according to the New Coronavirus Pneumonia Prevention and Control Program (5th edition) published by the National Health Commission of China and WHO interim guidance.^{27,28} The classification of non-severe and severe cases was according to the New Coronavirus Pneumonia Prevention and Control Program (5th edition). During the hospitalization, patients with fever or suspected respiratory infection, plus one of the following clinical manifestations including respiratory rate > 30 breaths/min, severe respiratory distress, or SpO₂ < 93% on room air were classified as severe cases²⁹.

Of the initial 12,911 patients in Hubei province, 152 with leukemia were excluded. The end follow-up date was April 26th, 2020. 11,823 of the 12,759 patients who were either discharged or died were included in our analysis. 936 remained in hospital at the end of follow up were treated as censored during model building process. Among them, 9,810 who had at least two CBC tests during hospitalization were designated as the training cohort to establish a dynamically applicable risk score based on the longitudinal data. 3,174 patients who had at least 3 times CBC tests on 3 different phases during hospitalization were included in the development of a LMM.

2,949 patients who had only one CBC record in Hubei Province were used for externally validate the performance of the PAWNN score in discriminating the risk of death. The C-statistic of PAWNN score was performed in 2,739 patients with clear outcomes at the end of the follow-up date (Figure 1). To test the generality of PAWNN score in Western population, the performance of the PAWNN score in discriminating the risk of death and C-statistic were also external validated in 227 patients from the Italy. There were 77 patients died and 150 patients discharged by the end of the follow-up date (Figure 1).

The study design was approved by the central ethics boards and was accepted or approved by each collaborating hospital. Patient informed consent was waived by the ethics committees from each hospital. A part of the baseline data in this manuscript have been used in our previous articles.^{30–33}

METHOD DETAILS

Data Collection and Definitions

Demographic information, clinical characteristics, medical history, laboratory tests, and outcome data were obtained with data collection forms from patients' electronic medical records. CBC parameters including RBC counts, hemoglobin concentrations, hematocrit, WBC counts, neutrophil counts, neutrophil percentage, lymphocyte counts, lymphocyte percentage, basophil counts, eosinophil counts, monocyte counts, and platelet counts during hospitalization were extracted. NLR was calculated and used as a candidate predictor due to its significant changes reported in COVID-19 patients.⁹ There were no missing data on the 13 CBC parameters in all included patients.

The outcome was defined as all-cause mortality during hospitalization. All included patients were stratified into non-severe survivor, severe survivor, and death groups. Personal identification information (e.g., name and ID) of each participant was anonymized before data extraction by giving a new study ID through a coding system. A team of experienced physicians carefully interpreted and double-checked all data to ensure accuracy. Mortality and collected medical information were evaluated by two independent groups of physicians.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analysis

All statistical analyses were conducted using R-3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). Continuous variables were presented as the median and IQR or mean and SD. Categorical variables were presented as frequency and percentage (%). The linear fitting curve for dynamic trajectories of CBC parameters from admission to Day 30 of hospitalization by the severity of COVID-19 was performed using locally weighted regression and smoothing scatterplots (LOESS). A two-sided P value less than 0.05 was used to define statistical significance.

Predictor Selection and Model Development

We followed the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) statement for reporting multivariable prediction model development and validation (Methods S1). The variable selection process was depicted in Figure 3. All patients with COVID-19 in the training cohort were included for variable selection and risk score development. Before fixed effects selection, patient-specific and site-specific random effects were compared to no random effect model by AIC and the marginal and conditional R^2 recommended by Nakagawa and Schielzeth³⁴ and Johnson.³⁵ Association with CBC records and the outcome was evaluated using hierarchical GLMMs with selected random effects. To better understand the contribution of each CBC variable, all CBC variables except NLR were transformed into three forms: numeric that scaled and centered, and two categorical variables represent > upper limit of normal (ULN) or < lower limit of normal (LLN) according to their reference ranges set at each participating hospital. NLR was transformed into numeric that scaled and centered, and a categorical variable with three categories (< 2.22, 2.22-4.06, > 4.06) based on a recent study.³⁶ Thus, a total of 38 variables were used for subsequent variable selection. After that, GLMMs with each variable as a fixed effect was built. Then we applied these variables, as well as their two-way interactions with time, to the multivariate GLMM with stepwise forward selection following the AIC ranking established from the univariate fixed effect models.^{37,38} The minimum significance levels for entry and stay at the selection process for the parameters were set at 0.05. To ensure quality of the developed model, the best candidates for the final regression model were identified manually by dropping the covariates with a p value > 0.05 one at a time until all of the regression coefficients were significantly different from 0. Lastly, age was added to the final chosen model as a confounder.

Predictive ability was summarized with the marginal and conditional R^2 , which represents the proportion of variation explained by the fixed effects only and the proportion of variation explained by both fixed and random effects. The conditional R^2 can be viewed as representing the total amount of variability in the patient's outcome that can be explained by the model when accounting for both fixed and random effects. Meanwhile, the marginal R^2 helps characterize the ability of fixed effects to predict the outcome.

Score Development and Validation

To further validate the predicting ability of the fixed effects and to facilitate clinical usage, we assigned each of the final selected predictors a numeric score that was proportional to its specific coefficient in the Cox proportional hazards regression model. A composite score was therefore developed, and time-varying scores were calculated for each patient from their CBC records during hospitalization. The temporal patterns of changes in patient's score among the non-severe survivor, severe survivor, and death groups were demonstrated by trend plot. The covariate specific time-dependent ROC analysis was conducted to assess the overall performances of the model for predicting the risk of death in the training cohort.^{39,40} The PAWNN was derived from the Cox-model and further internally and externally validated. First, we conducted an internal validation of the score by 10-fold cross-validation in the training cohort to estimate the accuracy of the score. Moreover, follow up days were divided by quartile to evaluate the prediction accuracy by follow up time from AUROC as well as sensitivity and specificity of the max score in each quartile. Then, the performance of the score was further validated in 2,949 patients from Hubei Province with CBC tested only once. The score's performance was also validated externally in an Italian cohort of 227 patients with only baseline CBC records.

Latent Markov Model

To unmask a "latent" (i.e., unobserved) construct of patient statuses and further investigate how the score characterizes patient's condition and predicts the transitioning between statuses over time, a longitudinal analysis with fixed effects selected by multivariate GLMM was conducted using LMM on patients with 3 and more CBC test records on 3 and more different dates during hospitalization ($n = 3,174$).⁸ This analysis allowed us to investigate whether the current composite score can early predict transition probabilities and risk of death, thus help to modify interventions in advance. Three time points were selected for patient assessments in LMM including CBC from day 0 to 7 after admission (Time 1), value of CBC records during day 8-14 (Time 2) of hospitalization, and CBC during hospitalization over 15 days (Time 3). Three assessment time points enabled us to obtain the same CBC profile for each patient's status membership over time. To select the best model with certain number of latent status (LS), we optimally combined goodness of fit and parsimony as measured using the AIC, when a lower AIC means the model is better. We used LMM to simultaneously estimate status prevalence at each time point and transition between the statuses over time. The mean and SD of scores were also calculated for each status at each time point to characterize status characteristics. Furthermore, the discrimination accuracy of the score on status membership and its prediction accuracy on status transitioning were evaluated by ROC and multiple class ROC.⁴¹