Development and external validation of a COVID-19 mortality risk prediction algorithm: a multicentre retrospective cohort study

Jin Mei,1 Weihua Hu,2 Qijian Chen,3 Chang Li,4 Zaishu Chen,5 Yanjie Fan,6 Shuwei Tian,6 Zhuheng Zhang,6 Bin Li,6 Qifa Ye,7 Jiang Yue,8 Qiao-Li Wang9

ABSTRACT

Objective This study aimed to develop and externally validate a COVID-19 mortality risk prediction algorithm.

Design Retrospective cohort study.

Setting Five designated tertiary hospitals for COVID-19 in Hubei province, China.

Participants We routinely collected medical data of 1364 confirmed adult patients with COVID-19 between 8 January and 19 March 2020. Among them, 1088 patients from two designated hospitals in Wuhan were used to develop the prognostic model, and 276 patients from three hospitals outside Wuhan were used for external validation. All patients were followed up for a maximal of 60 days after the diagnosis of COVID-19.

Methods The model discrimination was assessed by the area under the receiver operating characteristic curve (AUC) and Somers’ D test, and calibration was examined by the calibration plot. Decision curve analysis was conducted.

Main outcome measures The primary outcome was all-cause mortality within 60 days after the diagnosis of COVID-19.

Results The full model included seven predictors of age, respiratory failure, white cell count, lymphocytes, platelets, D-dimer and lactate dehydrogenase. The simple model contained five indicators of age, respiratory failure, coronary heart disease, renal failure and heart failure. After cross-validation, the AUC statistics based on derivation cohort were 0.96 (95% CI, 0.96 to 0.97) for the full model and 0.92 (95% CI, 0.89 to 0.95) for the simple model. The AUC statistics based on the external validation cohort were 0.97 (95% CI, 0.96 to 0.98) for the full model and 0.88 (95% CI, 0.80 to 0.96) for the simple model. Good calibration accuracy of these two models was found in the derivation and validation cohort.

Conclusion The prediction models showed good model performance in identifying patients with COVID-19 with a high risk of death in 60 days. It may be useful for acute risk classification.

Web calculator We provided a freely accessible web calculator (https://www.whuylia.com/).

INTRODUCTION

The pandemic of COVID-19 has spread rapidly across the world since December 2019. The number of confirmed cases is continuing to rise and the related deaths pile up, making it a great challenge for the healthcare resources to meet the increased demand for hospital beds and medical equipment (eg, ventilators). A prediction model for prognostic is needed for the risk stratification of confirmed cases. Early identification and intervention of patients with COVID-19 can reduce mortality and morbidity as well as mitigating the burden on the healthcare system. There are two assessment tools to evaluate community-acquired pneumonia including CURB-65 and Pneumonia Severity Index for adults. However, these tools were not specific for COVID-19 and they do not include known risk factors for COVID-19-related prognosis. Previous studies have well documented the association between laboratory indicators or comorbidities and COVID-19 severity. The reported risk factors associated with the poor prognosis include older age, cardiovascular metabolic diseases, respiratory disease and...
increased blood lactate dehydrogenase level. However, there are currently few models to predict mortality risk among patients with COVID-19, and these models had a high risk of selection and detection bias as well as model overfitting. Model calibration and external validation of risk prediction models are also lacking in the models. Therefore, this study aimed to develop a valid and easy-to-use risk prediction algorithm that estimates the risk of short-term mortality and to externally validate the model in an independent population.

METHODS
Study design, participants and data collection
We performed a multicentre retrospective cohort study of 1364 confirmed cases from designated tertiary hospitals in Hubei province, China. All confirmed 1088 cases of COVID-19 in the derivation cohort were from two designated tertiary hospitals in Wuhan (the Fifth Hospital of Wuhan and Hubei No. 3 People’s Hospital of Jianghan University) during the period of 8 January 2020–19 March 2020. All diagnosed patients with COVID-19 (n=276) from three designated tertiary hospitals outside Wuhan (Jiayu People’s Hospital, Jingzhou First People’s Hospital and People’s Hospital of Nanzhang County) were included as an independent validation cohort.

All the patients were diagnosed by the confirmatory testing for COVID-19, a real-time reverse transcription-PCR assay with nasal and pharyngeal swab specimens according to the WHO interim guidance. The virus detection was repeated twice for each patient. The patients were followed up for maximal 60 days after the diagnosis. We extracted the medical records and collected the information using a standardised case report form. Data collection included demographic factors (eg, age at diagnosis and sex), medical history (e.g., COVID-19 diagnosis date, death or discharge status and comorbidities at diagnosis), symptoms and vital signs at admission, and laboratory indicators (e.g., C-reactive protein and D-dimer) for each participant. Inclusion criteria were confirmed COVID-19 cases aged ≥20 years old during the study period.

Outcomes
The primary outcome was all-cause mortality, using the date of death recorded on the medical records. Patients were followed up for a maximum of 60 days until the occurrence of death or discharge.

Candidate predictor variables
We examined candidate predictor variables based on risk factors highlighted in related literature and routine blood laboratory tests. All demographic and epidemiological variables (e.g., age, sex and smoking status), symptoms at diagnosis (e.g., fever, headache and cough), comorbidities (e.g., hypertension, diabetes and respiration failure), vital signs (e.g., temperature, pulse rate and respiration rate) and laboratory indicators (e.g., white cell count, neutrophils and lymphocytes) were collected at the time point for the patients’ first hospital admission.

Derivation of the models
For both the derivation cohort and external validation cohort, multiple imputations based on the multivariate normal distribution were conducted for variables with more than 5% missing rate. Ten imputations were conducted for missing variables. We identified potential auxiliary variables that had absolute correlations greater than 0.4 with variables with missing data. The convergence of multiple imputation models performed well, and it was assessed by trace plots and autocorrelation plots. To explore the risk pattern of short-term mortality among patients with COVID-19, univariate logistic regression was conducted to estimate the odds ratio (OR) with a 95% CI for each of the 51 variables.

We initially included all predictor variables in a multivariate logistic regression model. Then the stepwise selection approach was applied for prediction selection, with a predefined nominal significance level of 0.05 for both model entry and retention. To avoid substantial improvement of the goodness of model fit in the likelihood ratio test by the omitted predictors, the excluded predictors from the first step were later re-entered into the model and re-evaluated one by one. Age was included without any evaluation as older age has been reported to be strongly associated with death in patients with COVID-19. Interaction terms of predictors were tested and added to the model. The risk equation for predicting the log odds of short-term mortality after COVID-19 infection was computed using the estimated β estimates multiplied by the corresponding selected predictors, along with the average intercept. The predicted log odds of mortality (marked as µ) from the derivative model were further used for computing the predicted absolute risk of short-term mortality with the algorithm of predicted risk=1/(1+e⁻µ).

For a quick classification of patients with COVID-19 with a high risk of short-term death, we also developed a simple model excluding laboratory tests but including comorbidities which had been previously reported to be associated with the poor prognosis among patients with COVID-19. The simple model was developed without any predictor evaluation because all included predictors have been previously reported to be risk factors for mortality among patients with COVID-19. The model performance of both the full and simple models was assessed.

Test of model performance and external validation
The prediction accuracy of the model was assessed by the area under the receiver operating characteristic curve (AUC) and the Somers’ D statistic. The AUC assessed the model’s ability to distinguish patients with from patients without the outcome of interest. Somers’ D statistic measured the strength and direction of correlations between observed outcomes and predicted probabilities. To avoid model...
overfitting, a leave-one-out cross-validation strategy was conducted to retest the model performance. Unbiased AUC and Somers’ D statistics were thus estimated with the predicted probability for each patient by a model ignoring this patient. Model calibration was assessed by comparing the predicted risk of 60 days of death with the observed risks by 10th of the predicted risk. To evaluate the predicted risk distribution in various centiles, we computed the sensitivity, specificity, positive and negative predictive number of deaths in the model.

We applied the developed risk prediction algorithms on the independent validation cohort. We accessed both discrimination and calibration accuracies in these patients.

**Decision curve analysis**
To evaluate the clinical utility, a decision curve analysis was conducted in the external validation cohort. We assessed the net benefit of the prediction model by estimating the differences between the proportion of true positive and false positive value and they were later multiplied by the risk odds. The decision curve evaluated the net benefit of correctly identifying patients who would have an event with the relative harms of a false-positive prediction across a wide range of threshold probabilities. The strategy of risk prediction model application was then compared with strategies of ‘treat all patients’ and ‘treat no patient’.

This study followed the Transparent Reporting of a multivariable Prediction model for Individual Prognosis or Diagnosis guidelines and checklist. The statistical software package SAS 9.4 for Windows was used for statistical analysis.

**Patient and public involvement**
No patients were involved in the study design or in setting the research questions or in the outcome measures directly. No patients were asked to advise on interpretation or the writing up of results.

**RESULTS**

### Clinical characteristics of patients in the derivation cohort
In the derivation cohort, 50% of the inpatients were women, and the median age was 58 years (interquartile range [IQR], 46–66) (online supplemental table S1). About 82% of the patients had comorbidity, and the most common comorbidities were hypertension (31%), diabetes (17%), respiratory failure (7%), coronary heart disease (7%) and liver disease (7%). The most common symptoms at admission were fever, cough, fatigue and breathlessness (70%, 57%, 27% and 27%, respectively). A total of 103 deaths (9.5%) were reported from the patients in the derivation cohort (online supplemental table S2).

### Predictor variables
The ORs and 95% CIs of 51 predictor variables in the univariable logistic regression models were shown in online supplemental table S3. The full model included predictors of age (per year increase, continuous), respiratory failure (yes vs no), white cell count (per $10^9/L$ increase, continuous), lymphocytes (per $10^9/L$ increase, continuous), platelets (per $10^9/L$ increase, continuous), D-dimer (per 1 µg/mL increase, continuous), lactate dehydrogenase (per 1 U/L increase, continuous), and two interactions of white cell count with platelets and D-dimer with lactate dehydrogenase. In this model, an increased risk of mortality was markedly associated with respiratory failure (OR 53; 95% CI, 22 to 128) (table 1).

**Table 1** Association between included predictor variables and 60-day mortality in the full model, expressed as ORs with 95% CIs and beta coefficients in the model

<table>
<thead>
<tr>
<th>Factors</th>
<th>OR (95% CI)</th>
<th>P value</th>
<th>β coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (per year increase)</td>
<td>1.07 (1.03 to 1.10)</td>
<td>0.0001</td>
<td>0.069800</td>
</tr>
<tr>
<td>Respiratory failure</td>
<td>4.120200</td>
<td>&lt;0.0001</td>
<td>-2.045900</td>
</tr>
<tr>
<td>No</td>
<td>Refernce</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>52.60 (21.54 to 128.47)</td>
<td>0.8000</td>
<td>0.034600</td>
</tr>
<tr>
<td>White cell (per $10^9/L$ increase)</td>
<td>1.03 (0.83 to 1.27)</td>
<td>0.0001</td>
<td>-2.045900</td>
</tr>
<tr>
<td>Lymphocytes (per $10^9/L$ increase)</td>
<td>0.13 (0.05 to 0.37)</td>
<td>0.0001</td>
<td>0.016000</td>
</tr>
<tr>
<td>Platelets (per $10^9/L$ increase)</td>
<td>0.98 (0.97 to 0.99)</td>
<td>0.0021</td>
<td>0.048500</td>
</tr>
<tr>
<td>D-dimer (per 1 µg/mL increase)</td>
<td>0.96 (0.88 to 1.05)</td>
<td>0.4000</td>
<td>0.0000910</td>
</tr>
<tr>
<td>Lactate dehydrogenase (per 1 U/L increase)</td>
<td>1.00 (1.00 to 1.00)</td>
<td>0.6900</td>
<td></td>
</tr>
<tr>
<td>Interaction terms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White cell count×platelets</td>
<td>1.00 (1.00 to 1.00)</td>
<td>0.0306</td>
<td>0.001150</td>
</tr>
<tr>
<td>Lactate hydrogenase×D-dimer</td>
<td>1.00 (1.00 to 1.00)</td>
<td>0.0295</td>
<td>0.000322</td>
</tr>
<tr>
<td>Constant intercept</td>
<td>-4.795000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
interaction between age and renal failure. Both risk prediction algorithms can be found in the supplemental document.

Model performance

The receiver operating characteristic curves for the prediction models were shown in figure 1 and table 2. The AUC statistics without cross-validation based on the derivation cohort were 0.97 (95% CI, 0.96 to 0.97) for the full model and 0.92 (95% CI, 0.90 to 0.95) for the simple model. After cross-validation, the AUC statistics were lightly declined to 0.96 (95% CI, 0.96 to 0.97) in the full model and 0.92 (95% CI, 0.89 to 0.95) in the simple model, indicating the comparable performance of the simple model with the full model.

A good performance was observed between the observed and predicted proportion of events in both the full and simple models from the derivation cohort and external validation cohort, indicating that the algorithms were well calibrated (figure 2). The sensitivity, specificity, positive predictive value and negative predictive value of the risk prediction model across various risk thresholds based on the derivation cohort were shown in online supplemental table S4. For example, with a risk threshold of 30%, the model has a sensitivity for identifying deaths of 74.2%, a specificity of 97.2%, a positive predictive value of 71.7% and a negative predictive value of 97.5%.

External validation

In the validation cohort, 55% of the patients were women (online supplemental table S1). The median age was 49 years (IQR, 36–62) and the proportion of death was 5.1% (14/276) (online supplemental table S2). About 39% of the patients had comorbidity, and the most common comorbidities were hypertension (18%), liver disease (8%) and diabetes (5%). The AUC statistics based on the external validation cohort were 0.97 (95% CI, 0.96 to 0.98) for the full model and 0.88 (95% CI, 0.80 to 0.96) for the simple model.

Decision curve analysis

The decision curve analysis showed the net benefit in 60 days after COVID-19 diagnosis in adults using the short-term mortality algorithm of the full model. The figure indicated an overall higher net benefit of the risk
prediction algorithm compared with the approaches based on considering either no patients or all patients for intervention, across all risk thresholds (figure 3).

**Model presentation**
We constructed an interactive web tool that integrated the risk prediction algorithms based on either the full model or the simple model (https://www.whuyijia.com/). For COVID-19-confirmed cases, the risk calculator provides the probabilities of mortality in 60 days with a numerical scale.

**DISCUSSION**
This study developed a full model to predict individual risk of short-term mortality after COVID-19 diagnosis, with predictors of age, respiratory failure, white cell count, lymphocytes, platelets, D-dimer and lactate dehydrogenase. The model showed good performance regarding discrimination and calibration accuracy in both the derivation and validation cohorts. We also developed a simple and easy-to-use model with only five readily available variables including age, comorbidities of respiratory failure, coronary heart disease, renal failure and heart failure, which can be used as a clinical risk stratification tool. A higher net benefit of the prediction model was observed compared with treat-all or treat-none approaches at various risk thresholds, indicating the potential clinical usefulness. The risk prediction algorithms were integrated as an online risk calculator.

Strengths of this study included the cohort design with complete follow-up and limited bias from patients’ selection or disease detection. We involved all patients with COVID-19 in the defined hospitals during the study period and followed them up for the coming 60 days at hospitals. An independent population from other cities was applied for external validation of the prediction model. The model performance was assessed using a cross-validation strategy, and the calibration plots of both the derivation and external validation cohorts indicated a good performance of the full model. The calibration plot was preferred over the Hosmer-Lemeshow tests because the latter are powerless in detecting the overfitting of predictor effects and sensitive to sample size.32–34 There are also limitations. Misclassification is unavoidable for self-reported variables, for example, smoking history. Yet they were not selected in the full model and all the final included predictors were clinically relevant data that were directly retrieved from clinical medical records, which ensured its accuracy. Both the mortality and comorbidity rates were higher in the derivation cohort compared with the validation cohort because the derivation cohort was based on the data from Wuhan, which was the centre of the COVID-19 outbreak and had more severe patients. The differences in mortality and comorbidity rate might affect the calibration of the validation cohort. Missing predictor variables from both the derivation and validation cohorts were reported. Although these variables with a missing rate of more than 5% were imputed for 10 times by the multiple imputations approach, potential information bias cannot be ruled out. Co-linearity was identified, which complicated the full model. Therefore, we introduced the interaction terms. In addition, concerns of representativeness of the validation cohort might be possible given the cohort sample size. Since the prediction models were developed based on the Chinese population, model validation in other populations might be necessary before its direct application.

A recent systematic review of prediction models for diagnosis and prognosis of COVID-19 pointed out that the existing prognostic models had high or unclear risks of both bias and overfitting.6 Moreover, the mortality prediction models did not present any applicable equations or web calculators, which made it impossible to use or verify. Among eight mortality prediction models, only one model assessed the calibration.78 35–40 It has been found that the predicted mortality risk of that model was too high for low-risk patients and too low for high-risk patients when applied to new patients.38 The disagreement between the observed and predicted proportion of events may be due to the small size of cases and the selection bias in a case-control setting. A mortality model based on 117000 confirmed cases was developed using artificial intelligence, and the accuracy of the prediction rate was 93%. However, the predictors included in the final model were unclear, and no application equation was provided.36 Although comorbidities have been reported to be associated with the worse outcome, only a few comorbidities have been screened during the model development in these studies. Specifically, the simple model developed in this study was based on the well-known risk factors of prognosis of COVID-19 (ie, individual comorbidities conditions such as coronary heart disease and renal failure). This model also presented a good model performance and is easy to use with only five readily available predictors.

Our full model included seven key determinants of death after COVID-19 infection, such as age, respiratory failure, white cell count, lymphocytes, platelets, D-dimer...
and lactate dehydrogenase. These variables have been documented to be associated with the mortality risk of COVID-19. The data from different regions suggest that the risk of severity and mortality of COVID-19 increases with age. It has been reported that the average death rate for adults over 80 was about 9.3%, but the death rate for adults under 60 was less than 0.2%. Despite the higher OR found among men in the univariate analysis of this study, sex was not included in the full model after multivariate analysis as it did not reach the predefined nominal significance level of entry and retention criteria. COVID-19 mainly affected the respiratory system, and the mortality risk significantly increased in patients who had severe respiratory failure. The mortality rate in critical cases of COVID-19 with respiratory failure has been reported between 26% and 61.5%, and Lymphopenia, D-dimer and lactate dehydrogenase have also been shown as independent risk factors associated with the severity and mortality of COVID-19. Lymphopenia was associated with a 2.99-fold higher risk of COVID-19 severity and an increase per 1 U/L of lactate dehydrogenase was independently related to 1.012 higher risk for disease severity. The OR of COVID-19 mortality was 2.14 higher when D-dimer reached 0.5 μg/ml or higher. These laboratory predictors themselves might also serve as indicators of other severe diseases, such as heart failure or renal failure.

This prediction model might be useful for clinicians to identify confirmed patients with COVID-19 who are at a very high risk of mortality. We have provided a novel algorithm to predict the 60-day mortality risk of confirmed patients with COVID-19, which may help clinicians do the objective decision-making based on medical and epidemiological evidence. The web-based risk calculator is freely accessible and could serve as a resource to support patient education and inform discussions around outcome expectations and management, including rehabilitation needs. It can also serve as a data-driven tool to enable patients and their relatives to effectively participate in making clinical decisions together with clinicians. Our simple model including age, respiratory failure, coronary heart disease, renal failure and heart failure is potentially useful for the quick risk classification of patients at admission. The laboratory markers included in the algorithm indicated that these indicators might be involved in the pathophysiological mechanism of COVID-19 infection.

CONCLUSION

In this study, we developed the prediction algorithms of 60-day mortality risk among patients with COVID-19. The easy-to-use model presented good discrimination and calibration ability and was well externally validated in an independent population. The online risk calculator could provide immediate risk prediction for clinical use.

Author affiliations

1Central Laboratory, Ningbo First Hospital, Zhejiang University, Ningbo, China
2Department of Respiratory and Critical Care, Jingzhou First People’s Hospital, Jingzhou, China
3Emergency Department, Fifth Hospital in Wuhan, Wuhan, Hubei, China
4Department of Cardiology, Hubei No.3 People’s Hospital of Jianghan University, Wuhan, Hubei, China
5Department of Cardiology, Internal Medicine, Jiaju People’s Hospital, Jiaju, China
6Department of Pharmacology, School of Basic Medical Sciences, Wuhan University, Wuhan, China
7Institute of Hepatobiliary Diseases of Wuhan University, Zhongnan Hospital of Wuhan University, Wuhan, China
8Hubei Province Key Laboratory of Allergy and Immunology, Wuhan, China
9Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Acknowledgements We thank Maoping Ye, Mengya Fan, Niuniu Gao, Fangfang Wang, Yulu Feng, Rui Zhang, Shiyian Chen, Jianfeng Ye, Junlu Tong, Zhanjian Lu, Feina Cai, Mingxuan Li and Yiding Qi for their efforts in collecting the information.

Contributors QW and JM contributed equally and share the first authorship. QW, JM, WH, QC, CL, ZC, OY and JY—study conception and design. QW, JM, WH, QC, CL, ZC, YF, ST, ZZ, BL and JY—data collection, analysis and review. QW and JY—statistical analysis and interpretation. JM and JY—administrative and technical support. QW, JM and JY—drafting of the manuscript. QW, JM, WH, QC, CL, ZC, YF, ST, ZZ, BL, OY and JY—critical revision of the manuscript.

Funding This study was supported by the National Natural Science Foundation of China (No. 81970653) and Medical Science Advancement Program (Clinical Medicine) of Wuhan University (Grant No. TFLC 2018003).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the institutional board of Ningbo First Hospital of Zhejiang University (2020-R120) with a waiver of written informed consent. The study was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, built upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID id

Jiang Yue http://orcid.org/0000-0001-8906-1253

REFERENCES


