

University of Groningen

Contemporary issues in static and dynamic prediction

Chen, Yuntao

DOI:
[10.33612/diss.232289260](https://doi.org/10.33612/diss.232289260)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Chen, Y. (2022). *Contemporary issues in static and dynamic prediction: some applications and evaluation in the clinical context*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.232289260>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 3

Derivation and validation of a prognostic model for predicting in-hospital mortality in patients admitted with COVID-19 in Wuhan, China: the PLANS (Platelet Lymphocyte Age Neutrophil Sex) model

Jiong Li*, Yuntao Chen*, Shujing Chen*, Sihua Wang, Dingyu Zhang, Junfeng Wang, Douwe Postmus, Hesong Zeng, Guoyou Qin, Yin Shen, Jinjun Jiang, Yongfu Yu

BMC infectious diseases. 2020 Dec 17;20(1):959.

Abstract

Background

Previous published prognostic models for COVID-19 patients have been suggested to be prone to bias due to unrepresentativeness of patient population, lack of external validation, inappropriate statistical analyses, or poor reporting. A high-quality and easy-to-use prognostic model to predict in-hospital mortality for COVID-19 patients could support physicians to make better clinical decisions.

Methods

Fine-Gray models were used to derive a prognostic model to predict in-hospital mortality (treating discharged alive from hospital as the competing event) in COVID-19 patients using two retrospective cohorts (n=1008) in Wuhan, China from January 1 to February 10, 2020. The proposed model was internally evaluated by bootstrap approach and externally evaluated in an external cohort (n=1031).

Results

The derivation cohort was a case-mix of mild-to-severe hospitalized COVID-19 patients (43.6% females, median age 55). The final model (PLANS), including five predictor variables of platelet count, lymphocyte count, age, neutrophil count, and sex, had an excellent predictive performance (optimism-adjusted C-index: 0.85, 95% CI: 0.83 to 0.87; averaged calibration slope: 0.95, 95% CI: 0.82 to 1.08). Internal

validation showed little overfitting. External validation using an independent cohort (47.8% female, median age 63) demonstrated excellent predictive performance (C-index: 0.87, 95% CI: 0.85 to 0.89; calibration slope: 1.02, 95% CI: 0.92 to 1.12). The averaged predicted cumulative incidence curves were close to the observed cumulative incidence curves in patients with different risk profiles.

Conclusions

The PLANS model based on five routinely collected predictors would assist clinicians in better triaging patients and allocating healthcare resources to reduce COVID-19 fatality.

Keywords: COVID-19, in-hospital mortality, prognostic model, PLANS.

Background

The novel coronavirus disease 2019 (COVID-19) has become a pandemic worldwide since its first outbreak in Wuhan, China since December 2019.(1) As of July 3, 2020, more than 10 million cases are confirmed in over 200 countries, including 517 337 deaths.(2) Due to the high contagiousness and rapid progression of the disease, healthcare demand, in particular for critical care capacities, has often been overwhelming even in high-income areas.(3) Good support tools are needed for clinicians and other healthcare workers to respond promptly to urgent situations. It is crucial to accurately select severe patients for targeted treatment. For example, while it is essential to increase the intensive care unit (ICU) capacities and staff, ICU triage may be critical to prioritising severe patients for intensive care.(4) Therefore, early stratification of patients will facilitate targeted supportive care and appropriate allocation of medical resources.

Prognostic model that combines several clinical or non-clinical variables to estimate the future health outcomes of an individual could be a useful tool.(5) To respond quickly to the health crisis of COVID-19, a prognostic model based on robust evidence could be used as a simple and inexpensive tool to assist physicians in triaging the patients in the first place, which in turn may mitigate the burden of overwhelmed healthcare system and better allocate limited healthcare resources to reduce COVID-19 fatality.(6) Currently, several clinical prognostic models have been developed for COVID-19 patients.(7, 8) However, the quality of these models has been criticized and was prone to bias due to unrepresentativeness of patient

population, lack of external validation, inappropriate statistical analyses, or poor reporting.(7) Two of these prognostic models have been constructed with promising predictive performance for predicting mortality.(9, 10) However, they may not be highly reliable due to relatively small derivation cohorts (189 to 296 patients) and external validation cohorts (19 to 165 patients). Several studies used a time-to-event analysis to allow for administrative censoring.(11-13) However, censoring for other reasons, such as being discharged alive because of quick recovery, was seldom considered to be analyzed in a competing risk framework. For example, the CALL score(11) predicted the disease progression in hospitalized COVID-19 patients by using the standard Cox model. In this study, the risk of progression would be over-estimated because the patients discharged alive are no longer at risk of disease progression while the standard Cox model assumes they are still at risk.

In this study, we aimed to develop and validate a prognostic model to predict in-hospital mortality in COVID-19 patients using routinely measured demographic and clinical characteristics.

Methods

Study cohorts

Derivation cohort

The derivation cohort included 1008 COVID-19 patients admitted at Jinyintan Hospital (n=763) and Union Hospital (n=245) in Wuhan, China from January 1 to

February 10, 2020. Patients were followed up to March 20, 2020. Patients who were still hospitalized until March 20, 2020 were not included in the analyses. The Jinyintan hospital had mostly severe patients while Union Hospital had mostly mild patients, thus providing a case-mix of mild-severe COVID-19 patients.

Validation cohort

The validation cohort included 1031 COVID-19 patients aged ≥ 18 years at Tongji Hospital in Wuhan, China from January 14 to March 8, 2020. Since this cohort was designed to assess the potential risk factors related to acute cardiac injury in COVID-19 patients, patients with stage of chronic kidney disease ≥ 4 , chronic heart failure in the decompensatory stage, acute myocardial infarction during hospitalization, or having missing information on hypersensitive cardiac troponin I were excluded. Patients were followed up to March 30, 2020.

Data collection

A trained team of physicians retrospectively reviewed clinical electronic medical records and laboratory findings for all the patients. All patients met the diagnostic criteria according to the WHO interim guidance.(14) In the derivation cohort, we collected data on age, sex, the dates of admission and discharge or death, complete blood count at admission (neutrophil, lymphocyte, platelet count, haemoglobin), current smoking status (no, yes), chronic disease history (hypertension, digestive disease, kidney disease, coronary heart disease(CHD), chronic pulmonary disease,

cerebrovascular disease, diabetes, thyroid disease, malignancy, and other diseases). In the validation cohort, we collected data on age, sex, the dates of admission and discharge or death, complete blood count at admission (neutrophil, lymphocyte, platelet count), chronic disease history (hypertension, diabetes, CHD). All data were reviewed and collected by two physicians and a third researcher adjudicated any difference in interpretation between the two physicians.

Outcome and candidate predictors

The end point of interest was the time from hospital admission until in-hospital death (event of interest) or discharged alive (competing event) or 30-day after hospital admission (censored), whichever came first. Discharged alive was treated as a competing event because the event of discharged alive precludes the event of in-hospital death. Since conventional survival methods, such as Kaplan-Meier method and Cox model, assume two competing events (in-hospital death and discharged alive) are independent, they are not valid any more and more advanced methods accounting for competing risks should be used. Candidate predictor variables included a set of demographic variables (age, sex, current smoking status), laboratory findings (neutrophil count, lymphocyte count, platelet count), and comorbidities (hypertension, CHD, diabetes, cerebrovascular disease, and malignancy), which were selected according to clinical knowledge, literature,(7, 15) and data availability. While current smoking status was not considered due to high proportion of missing

data in the derivation cohort (46.3% missing), information on all other candidate predictor variables and outcome was complete for data analysis.

Model derivation

Fine-Gray models were used to develop the prognostic model, treating discharged alive from hospital as a competing event.⁽¹⁶⁾ The prognostic model derivation consisted of a prognostic index (PI) that captured the effect of the predictor variables on cumulative incidence function (CIF) for death, and a baseline CIF that determined the cumulative mortality of an “average” patient, i.e., a patient with the average value of PI. First, uni-variable Fine-Gray models with fractional polynomials (maximum permissible degree 1) were performed to investigate the potential non-linear relationship between continuous variables and CIF for death. Second, a multivariable Fine-Gray model with all the predictors was built. Backward elimination was applied to do the variable selection with significant level setting to 0.05, resulting in a final model in this step. PI was then calculated based on the combination of β coefficients and values of the corresponding predictors. The baseline CIF $CIF_0(t)$ corresponds to the cumulative mortality of an “average” patient with the average value of PI. The CIF for death of other patients can be computed via the formula: $CIF_i(t) = 1 - (1 - CIF_0(t))^{\exp(PI_i - \overline{PI})}$, where PI_i is the PI of patient i and \overline{PI} is the average value of PI in the derivation cohort. Details about the implementation and estimates of the Fine-Gray model, see the Appendix Text 1.

Model performance and internal validation

Model performance was assessed in terms of discrimination and calibration. Discrimination was assessed using the concordance statistic (C-index).(17) Calibration was assessed jointly by calibration slope and calibration plot. Calibration slope is a measure to estimate the regression coefficient on the PI in the validation dataset.(18) In the calibration plot, the averaged predicted mortality curves estimated by the proposed prognostic model were compared with the averaged observed mortality curves across several risk groups. The risk group was based on patients' PI (thresholds: 16th, 50th and 84th percentiles).(19)

We performed internal validation to estimate the optimism (the level of model overfitting) and adjusted measures of C-index and calibration slope by bootstrapping 1000 samples of the original data (a detailed description of implementation of bootstrap is provided in Appendix Text 2). Average calibration slope in the internal validation was obtained to be a uniform shrinkage factor. We multiplied the shrinkage factor by the raw PI (PI in the model derivation step) to obtain optimism-adjusted PI. Lastly, we developed the final model by re-estimating the baseline CIF for death based on the optimism-adjusted PI.

External validation

The final model was applied to each patient in the external validation cohort. PI was then calculated based on the combination of β coefficients and the corresponding predictor values of every patient. The discriminative accuracy of the proposed model was evaluated using C-index and visually checked by the distribution of PIs. The calibration accuracy of the proposed model was assessed using calibration slope and visually checked by calibration plot.

Statements about reporting and evaluation of our prognostic model

The reporting of this prognostic model study followed Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement (Appendix).(20) The risk of bias of the prognostic model was independently assessed by an expert (JW, who did not take part in the model development and validation) using PROBAST (prediction model risk of bias assessment tool).(21)

Results

Patient population

In the derivation cohort, the median age of 1008 patients was 55 (interquartile range [IQR] 44-65, youngest at 14 years of age and oldest at 98 years) and 43.6% patients

were females. During a median length of stay (LOS) of 12 days (IQR 8-16), 211 patients died in total, and 4 of which died beyond 30 days. 757 patients discharged alive from the hospital within 30 days. There were 438 (43.5%) patients with one or more comorbidities. Hypertension (N=232, 23.0%), diabetes (N=110, 10.9%), chronic digestive disease (N=78, 7.7%), and chronic pulmonary disease (N=40, 4.0%) were among the most frequent comorbidities (Table 1).

In the validation cohort, the 1031 patients included were older (63, IQR 52-70), had more females (47.8%), and were more prevalent with hypertension (N=383, 37.1%), CHD (N=83, 8.1%) and diabetes (N=189, 18.3%), compared to the derivation cohort (Table 1). Patients had a longer LOS (19, IQR 11-27). The in-hospital mortality of patients in the validation cohort was slightly lower compared with the derivation cohort (Appendix Figure 1).

Table 1. Basic characteristics

	Derivation cohort (n=1008)	Validation cohort (n=1031)
Age, years	55 (44-65)	63 (52-70)
Sex, female	439 (43.6%)	493 (47.8%)
Current smoke status*	57 (10.5%)	-
Neutrophil count, $\times 10^9/L$	4.40 (2.79-6.96)	3.90 (2.78-5.68)
Lymphocyte count*, $\times 10^9/L$	0.95 (0.61-1.34)	1.07 (0.70-1.49)
Platelet count*, $\times 10^9/L$	194 (145-256)	219 (164-288)
Haemoglobin*, g/L	126 (115-138)	-
Chronic pulmonary disease	40 (4.0%)	-

	Derivation cohort (n=1008)	Validation cohort (n=1031)
Hypertension	232 (23.0%)	383 (37.1%)
Coronary heart disease	32 (3.2%)	83 (8.1%)
Diabetes	110 (10.9%)	189 (18.3%)
Thyroid disease	31 (3.1%)	-
Chronic digestive disease	78 (7.7%)	-
Cerebrovascular disease	22 (2.2%)	-
Chronic kidney disease	25 (2.5%)	-
Malignancy	31 (3.1%)	-

* Current smoke status was missing in 467 (46.3%) patients in the derivation cohort, lymphocyte count was missing in 1 patient in the validation cohort, platelet count was missing in 2 patients in the validation cohort, and haemoglobin was missing in 376 (37.3%) patients in the derivation cohort.

In the validation cohort, the 1031 patients included were older (63, IQR 52-70), had more females (47.8%), and were more prevalent with hypertension (N=383, 37.1%), CHD (N=83, 8.1%) and diabetes (N=189, 18.3%), compared to the derivation cohort (Table 1). Patients had a longer LOS (19, IQR 11-27). The in-hospital mortality of patients in the validation cohort was slightly lower compared with the derivation cohort (Appendix Figure 1).

Coding of predictors

Categorical predictors (sex, hypertension, CHD, diabetes, cerebrovascular disease and malignancy) were coded as dummy variables. Among continuous predictors, we did not observe obvious violation of linearity assumption for age, neutrophil and

platelet count. We observed a non-linear relation between outcome and lymphocyte count. Therefore, we included the transformed lymphocyte count (square root of the lymphocyte count) in the model according to the results of fractional polynomial analyses.

Model derivation and internal validation

The PLANS model included five predictors: platelet count, lymphocyte count, age, neutrophil count, and sex. Cumulative incidence function for the in-hospital mortality was associated with older age, being male, higher neutrophil, lower lymphocyte and lower platelet count (Table 2). This model showed excellent apparent discriminative ability (C-index: 0.85, 95% CI: 0.83 to 0.88). After adjusting for overfitting, the model maintained excellent discriminative accuracy (optimism-adjusted C-index: 0.85, 95% CI: 0.83 to 0.87). The average calibration slope (uniform shrinkage factor) was 0.95 (95% CI: 0.82 to 1.08), suggesting little model overfit. The final PI was calculated as 0.95 (uniform shrinkage factor) times the raw PI and the formula for final PI was structured as

$$\text{PI} = -0.002 * \text{Platelet} - 2.399 * \text{Lymphocyte} + 0.044 * \text{Age} + 0.127 * \text{Neutrophil} + 0.468 * \text{Sex} \text{ (Formula 1)}$$

- Platelet: $\times 10^9/\text{L}$
- Lymphocyte: $\times 10^9/\text{L}$, transformed to $\text{lymphocyte}^{0.5}$
- Age: in years

- Neutrophil: $\times 10^9/L$
- Sex: female=0; male=1

The distribution of final PI suggested good discriminative ability of our model (upper panel of Figure 1). The relationship between PI and 7-day, 14-day and 30-day mortality are presented in Figure 2. While we observed a slight underestimate of the mortality in the highest risk group, the agreement between predicted mortality curves and the observed mortality curves in the other risk groups suggested good calibration of our model (left panel of Figure 3). The final formula for the PLANS model and a patient example of how it can be applied in the real clinical practice is depicted in Box 1. Furthermore, an online calculator can be accessed for this calculation: <https://plans.shinyapps.io/dynnomapp/>.

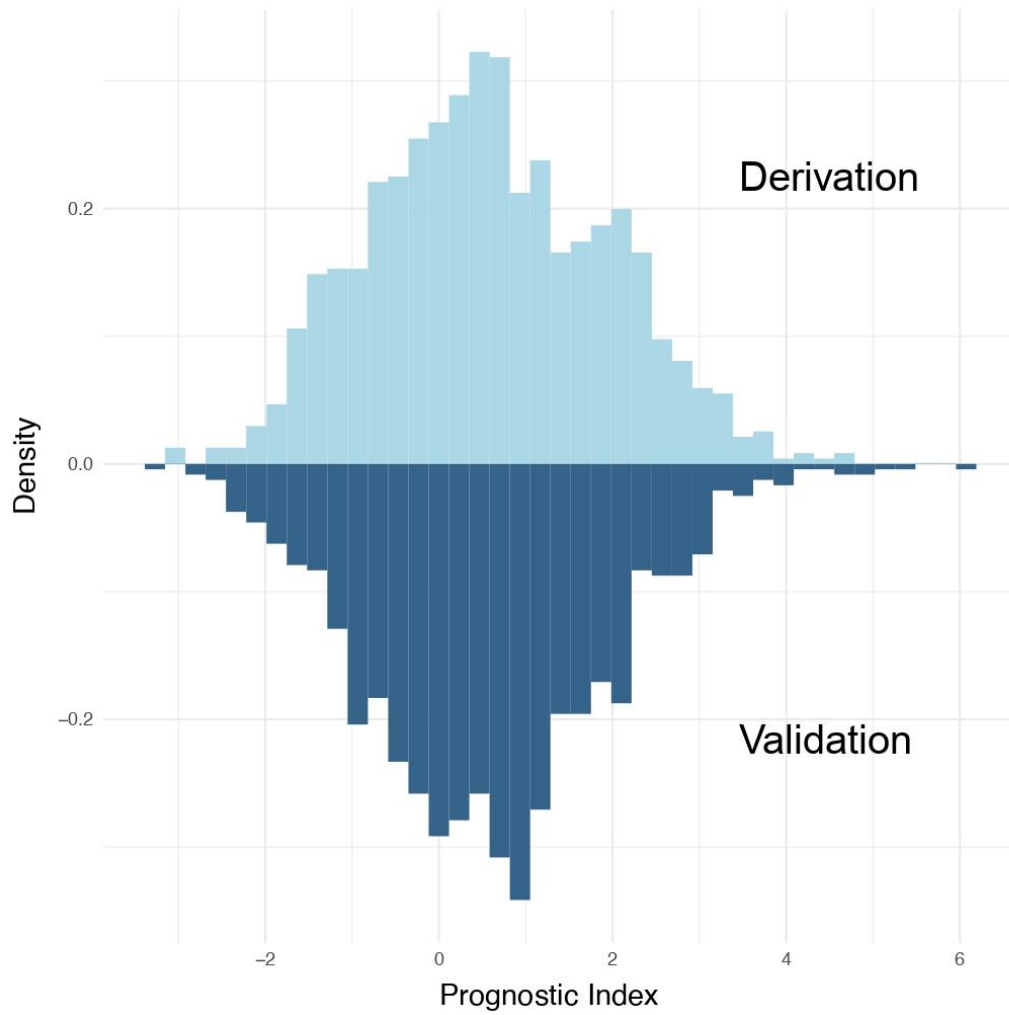


Figure 1. Distribution of the prognostic index of the prognostic model in the derivation and validation cohort; Upper part: derivation cohort; Lower part: validation cohort

Table 2. Results from multi-variable Fine-Gray model

Variables	Coding	Coefficient	95% CI	P
Age	=x*	0.046	0.036 - 0.057	<0.001
Sex	Dummy (0=Female, 1=Male)	0.490	0.179 - 0.802	0.002
Neutrophil count	=x	0.133	0.109 - 0.156	<0.001
Lymphocyte count	Lymphocyte count ^ 0.5	-2.514	-3.192- -1.835	<0.001
Platelet count	=x	-0.002	-0.004 - - 0.028 0.001	0.028

* x stands for original value.

Box 1: Final prognostic model (PLANS)

$$PI_i = -0.002 * \text{Platelet} - 2.399 * \text{Lymphocyte} + 0.044 * \text{Age} + 0.127 * \text{Neutrophil} + 0.468 * \text{Sex},$$
$$CIF_i(t) = 1 - (1 - CIF_0(t))^{\exp(PI_i - \overline{PI})}.$$

- Platelet: $\times 10^9/L$
- Lymphocyte: $\times 10^9/L$, transformed to lymphocyte $^{\wedge} 0.5$
- Age: in years
- Neutrophil: $\times 10^9/L$
- Sex: female=0; male=1
- $CIF_0(t)$: the mortality of an "average" patient with \overline{PI} equaling to 0.5662, which is given in Appendix Table 1
- $\overline{PI} = 0.5662$
- $CIF_i(t)$: the probability of death at time t

Example

- Suppose a 50 year old male patient was diagnosed with COVID-19 and was admitted in the hospital. He immediately went blood test and the results showed as follows:
 - Neutrophil count ($10^9/L$): 1.84,
 - Lymphocyte count ($10^9/L$): 0.69,
 - Platelet count ($10^9/L$): 184.

The 7-day and 30-day mortality can be estimated as follows:

- $PI_i = -0.002 * 184 - 2.399 * (0.69^{0.5}) + 0.044 * 50 + 0.127 * 1.84 + 0.468 * 1 = 0.5409$
- 7-day mortality:
 $CIF_i(7) = 1 - (1 - CIF_0(7))^{\exp(0.5409 - 0.5662)} = 1 - (1 - 0.0350)^{\exp(-0.0253)} = 3.41\%$
- 30-day mortality:
 $CIF_i(30) = 1 - (1 - CIF_0(30))^{\exp(0.5409 - 0.5662)} = 1 - (1 - 0.1197)^{\exp(-0.0253)} = 11.69\%$

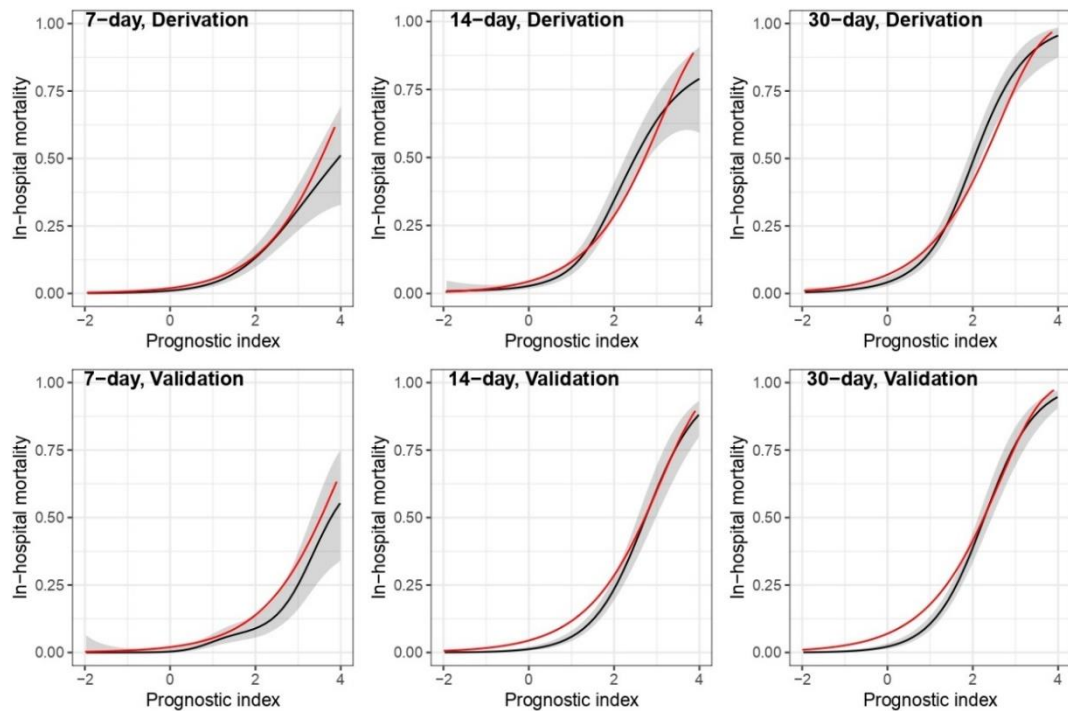


Figure 2. Prediction of 7-day, 14-day and 30-day mortality versus the prognostic index based on the PLANS model (red line); Prediction of 7-day, 14-day and 30-day mortality versus a smooth function of the prognostic index using generalized additive model (black line) together with the 95% confidence band; Upper part: derivation cohort; Lower part: validation cohort.

External validation

We applied the PLANS model to the independent cohort of 1031 patients from Tongji Hospital. The distribution of the PIs in the validation cohort was very similar to that in the derivation cohort, suggesting that the excellent discriminative accuracy of our model maintained in the validation cohort (Figure 1). The resulting C-index showed excellent discriminative accuracy of our model (C-index: 0.87, 95% CI: 0.85 to 0.89). Regarding the calibration accuracy, our model slightly overestimated

mortality in each risk group (right panel of Figure 3). Details about the thresholds and corresponding proportion and death toll included in each risk group are provided (Appendix Table 2). Jointly considering a close-to-one calibration slope (1.02, 95% CI: 0.92 to 1.12) and good agreement between predicted and observed cumulative incidence curves, our model still suggested good calibration accuracy in the validation cohort.

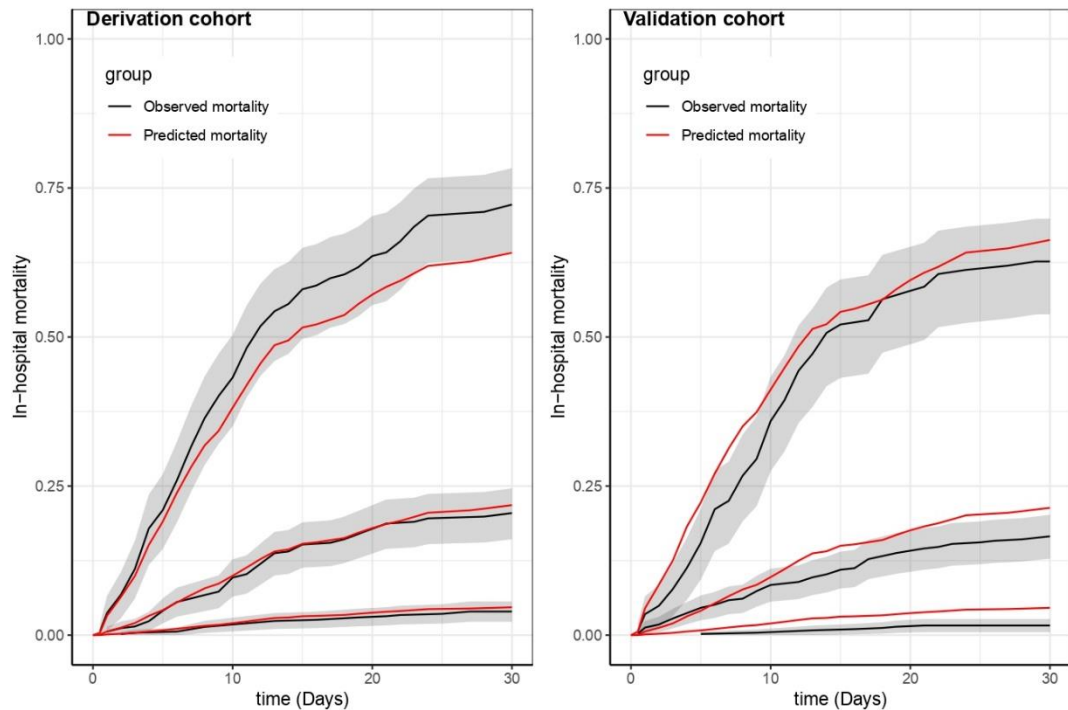


Figure 3. Predicted vs. Observed cumulative incidence curves per risk group in the derivation and validation cohort (Risk groups were defined based on PI. PI range from low risk to high risk group: ≤ -0.81 , -0.81 to 0.50 , 0.50 to 2.03 , and > 2.03 . Two lowest group were combined due to the limited death.)

Model update

The proposed model may not be directly applied to other areas where the distribution of predictive factors may be different from that in Wuhan. For instance, New York of USA and Lombardy of Italy could have a different distribution of predictor variables compared with Wuhan. Therefore, we used entropy balancing to update proposed model to generalize to their settings.(22) Details about the two updated models, see Appendix Text 3.

Methodology quality assessment

According to the PROBAST, the proposed model was rated as low risk of bias in all four domains: 17 of the total signaling questions were "Yes" and 3 were "Probably Yes". Rationales of answers were shown in Appendix Table 6.

Discussion

We developed a prognostic model (PLANS), using clinical readily available measures of platelet count, lymphocyte count, age, neutrophil count, and sex, to predict in-hospital mortality for COVID-19 patients using two retrospective cohorts in Wuhan, China. This model was first internally validated using bootstrap and then externally validated in an independent cohort in Wuhan. The PLANS model showed excellent discriminative and calibration accuracy.

All the five predictors are routinely collected and some of them have been already well established as the risk factors for in-hospital mortality in previous studies.(23) Recent studies from Italy, the USA, and China(24-26) have also reported that advanced age was a strong predictor of in-hospital mortality as suggested in our study. Compared to previous studies,(27, 28) our study had a more balanced gender composition. Our finding that male gender was associated with increased in-hospital mortality provided further evidence to support the hypothesis of male's vulnerability to COVID-19.(29, 30) Our study further confirmed that poor prognosis was associated with higher neutrophil and lower lymphocyte count.(31) On top of that, lymphopenia was found to have a non-linear relation with in-hospital mortality. A meta-analysis of nine studies had reported that thrombocytopenia was significantly associated with the severity of COVID-19 disease, but heterogeneity between studies was high.(32) Given a relatively large sample size and longer follow-up, our study indicated thrombocytopenia was associated with a higher risk of in-hospital mortality. Other studies have shown that several comorbidities (hypertension, diabetes, and coronary heart disease) were associated with poor prognosis.(24, 33) While none of the comorbidities were included in our model, we found that diabetes status would be incorporated when we excluded age from our model. It is plausible as the prevalence of most comorbidities, in particular diabetes, increases with age.(25)

Since the outbreak of COVID-19 in Wuhan, a number of prognostic models have been established.(7) A comprehensive systematic review conducted by Wynants and

colleagues found that most of these models were of high risk of bias due to several methodological limitations from participant domain to analysis domain.(7) Compared to the previous models, the PLANS model has several strengths. Our derivation cohort had a relatively large sample size with complete information on candidate predictors. While duration of follow up was unclear in most of the previous studies, the patients in our study were followed over a relatively long period, allowing us to perform a time-to-event analysis to predict in-hospital mortality. Furthermore, a competing risk analysis treating discharged alive as a competing event was done in this study to avoid overestimation of mortality. The similar distribution of age and sex in our study to recent large international reports(34, 35) indicates good representativeness of the patient population. External validation of the PLANS model to a large sample of patients showed excellent discrimination and calibration accuracy, indicating the generalizability of the PLANS model in the same city. Furthermore, we explored the possibility of generalizing the PLANS model to New York and Lombardy by using the published summary statistics. Though the adapted models are not recommended being applied before external validation, it might still be a good initiative to develop them and make use of them in the areas where the pandemic is still prevailing. The PLANS model was developed following high methodological standard and rated as low risks of bias in all four domains using PROBAST. Therefore, the PLANS model might be more reliable than most of the published prognostic models in making clinical decisions.

Several limitations should be noted. First, like most of the previous datasets and two main initiatives which created protocols for the investigators, namely, the ‘International Severe Acute Respiratory and emerging Infectious Consortium (ISARIC)’ and the ‘Lean European Open Survey on SARS-CoV-2 Infected Patients (LOESS)’, we only include closed (discharged or dead) COVID-19 cases. However, the resulting bias of unrepresentative sample could be largely offset by the long period of follow-up time. Second, we did have missing data on current smoking status for some patients. Inclusion of smoking status into the current model might improve the model performance. However, a reliable mechanism under the association between smoking and negative progression of COVID-19 is still missing.(36) Third, some potential risk factors confirmed by previous studies, such as D-dimer(31), C-reactive protein(37), lactate dehydrogenase(27, 38), and interleukin-6(27), were not available in our study. Respiration symptoms were not available either, and inclusion of which might improve the predictive accuracy. However, considering the practicality and validity in clinical application, a simple and interpretable model is usually preferred.(39) In addition, our model showed promising performances with five routinely available predictors, balancing the trade-off between model performance and model practicality.

Implication for practice

The availability of a prognostic model that can accurately predict in-hospital mortality in COVID-19 patients upon admission to hospital has important

implications for practice and policy. The PLANS model may assist physicians to early stratify the patients according to the estimated mortality at 7-day (14-day or 30-day) after admission, thus giving patients targeted supporting care and better allocating the limited medical facilities (e.g. ventilators), especially when critical care capacities are overwhelmed. Several studies showed that physicians have been experiencing guilt when they make clinical decisions that contravene the morals of those making them, e.g. one ventilator, two patients.(40, 41) The PLANS model might be useful to be incorporated into a protocol to assist physicians in making those difficult decisions. Our findings from the model update suggest that our model might be generalized to different countries as well. The model could be validated in the first place and then be used directly if it performs well or after being updated according to local settings.(42)

Conclusion

In summary, the PLANS model can be a guidance model for Chinese hospitals in case of the resurgence of COVID-19. It can also be a useful tool for predicting mortality or triage patients in the countries where COVID-19 is still a pandemic after being validated in their settings. Future studies are warranted about the impact of the PLANS model on clinical practice and decision.

References

1. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 2020;382(18):1708-20.
2. World Health Organization. Coronavirus disease (COVID-19) Pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed May 13, 2020.
3. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. 2020:eabb5793.
4. Phua J, Weng L, Ling L, Egi M, Lim CM, Divatia JV, et al. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *Lancet Respir Med*. 2020;8(5):506-17.
5. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375.
6. Coslovsky M, Takala J, Exadaktylos AK, Martinolli L, Merz TM. A clinical prediction model to identify patients at high risk of death in the emergency department. *Intensive Care Med*. 2015;41(6):1029-36.
7. Wynants L, Van Calster B, Bonten MMJ, Collins GS, Debray TPA, De Vos M, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.
8. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and laboratory predictors of in-hospital mortality in patients with COVID-19: a cohort study in Wuhan, China. *Clin Infect Dis*. 2020.
9. Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. 2020.
10. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool to Early Predict Severe 2019-Novel Coronavirus Pneumonia (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China. preprint. *Public and Global Health*; 2020 2020/03/20/.

11. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score. *Clinical Infectious Diseases*. 2020:ciaa414.
12. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, González-Díaz A, Márquez-Salinas A, et al. Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. *J Clin Endocrinol Metab*. 2020;105(8).
13. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J*. 2020;56(2).
14. World Health Organization. Clinical management of severe acute respiratory infection when COVID-19 is suspected. [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 March 6, 2020.; .
15. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *N Engl J Med*. 2020;0(0):null.
16. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks. *Stat Med*. 2016;35(22):4056-72.
17. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-87.
18. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401-15.
19. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13:33.
20. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102(3):148-58.
21. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med*. 2019;170(1):W1-W33.

22. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*. 2017;20(1):25-46.
23. Xu L, yaqian m, Chen G. Risk factors for severe corona virus disease 2019 (COVID-19) patients : a systematic review and meta analysis. preprint. *Infectious Diseases (except HIV/AIDS)*; 2020 2020/04/01/.
24. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*. 2020.
25. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020.
26. Ruan S. Likelihood of survival of coronavirus disease 2019. *Lancet Infect Dis*. 2020:S1473309920302577.
27. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, 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.
28. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, 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(10229):1054-62.
29. Wenham C, Smith J, Morgan R, Gender, Group C-W. COVID-19: the gendered impacts of the outbreak. *Lancet*. 2020;395(10227):846-8.
30. Sama IE, Ravera A, Santema BT, van Goor H, Ter Maaten JM, Cleland JGF, et al. Circulating plasma concentrations of angiotensin-converting enzyme 2 in men and women with heart failure and effects of renin-angiotensin-aldosterone inhibitors. *Eur Heart J*. 2020;41(19):1810-7.
31. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, et al. Hematological findings and complications of COVID-19. *Am J Hematol*. 2020.
32. 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.

33. Guan WJ, Liang WH, Zhao Y, Liang HR, Chen ZS, Li YM, et al. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J*. 2020;55(5).
34. Novel Coronavirus Pneumonia Emergency Response Epidemiology T. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2020;41(2):145-51.
35. The Italian Department of Infectious Diseases and the IT Service. Integrated surveillance of COVID-19 in Italy. https://www.epicentro.iss.it/en/coronavirus/bollettino/Infografica_11maggio%20ENG.pdf Accessed May 12, 2020.
36. Cai H. Sex difference and smoking predisposition in patients with COVID-19. *Lancet Respir Med*. 2020;8(4):e20.
37. Deng Y, Liu W, Liu K, Fang YY, Shang J, Zhou L, et al. Clinical characteristics of fatal and recovered cases of coronavirus disease 2019 in Wuhan, China: a retrospective study. *Chin Med J (Engl)*. 2020;133(11):1261-7.
38. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. 2020;395(10229):1054-62.
39. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338:b604.
40. Walton M, Murray E, Christian MD. Mental health care for medical staff and affiliated healthcare workers during the COVID-19 pandemic. *Eur Heart J Acute Cardiovasc Care*. 2020:2048872620922795.
41. The L. COVID-19: protecting health-care workers. *The Lancet*. 2020;395(10228):922.
42. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.

Appendix

TRIPOD Checklist

Text 1. Implementation and estimates of Fine-Gray model

Text 2. Internal validation by bootstrap

Text 3. Two updated models

Figure 1. Cumulative mortality curves for derivation and validation cohort.

Figure 2. The Schoenfeld residual plots for each predictor, test of proportional hazards

Table 1. “Baseline” mortality (Wuhan, China)

Table 2. Thresholds and corresponding proportion and death toll included in each risk group

Table 3. Basic characteristics used in entropy balancing in Derivation cohort, New York cohort and Lombardy cohort

Table 4. “Baseline” mortality (New York, USA)

Table 5. “Baseline” mortality (Lombardy, Italy)

Table 6. Methodology quality assessment based on PROBAST risk of bias assessment tool

Appendix Text 1. Implementation and estimates of Fine-Gray model

Since there were no patients lost to follow-up in our study, conventional Cox models gives the same estimate as Fine-Gray models by treating discharged patients as being right censored at maximum follow-up time (defined as 30 days after hospital admission) irrespective of the actual length of hospital stay.(1) A valid nonparametric estimates of the cumulative incidence function could also be obtained by Kaplan-Meier method using this data coding approach.(1)

Appendix Text 2. Internal validation by bootstrap

We performed internal validation to estimate the optimism (the level of model overfitting) and adjusted measures of C-index and calibration slope by bootstrapping 1000 samples of the original data. The model derivation process was repeated in each bootstrap dataset and then created 1000 bootstrap models (1000 apparent C-indice). We applied these 1000 bootstrap models to the original dataset (1000 validated C-indice). Optimism in C-index was determined through the difference between the average apparent C-index and the average validated C-index. Optimism-adjusted C-index was obtained by subtracting the optimism from the original C-index. Calibration slope was calculated by estimating the regression coefficient of the PI from the 1000 bootstrap models in the original dataset. Average calibration slope was obtained to be a uniform shrinkage factor.

Appendix Text 3. Two updated models

The proposed model may not be directly applied to other areas where the distribution of predictive factors may be different from that in Wuhan. For instance, New York

of USA and Lombardy of Italy could have a different distribution of predictor variables compared with Wuhan.(2,3) Therefore, we used entropy balancing to update proposed model to generalize to their settings.(4) First, entropy balancing approach was implemented to estimate a weight that made our derivation data comparable with the New York cohort in terms of the distribution of age, sex, hypertension, CHD, diabetes and malignancy. Second, a weighted Cox regression was used estimate the baseline mortality of the “average” patient in New York by offsetting the PI. Last, the updated prognostic model for New York can be obtained via $CIF_i(t) = 1 - (1 - CIF_0(t))^{\exp(PI_i - \overline{PI})}$, where $CIF_0(t)$ is a New York specific baseline survival, PI_i is the PI of patient i and \overline{PI} is the PI of the average patient in New York, which was assumed the same as that in Wuhan. Same procedures were implemented to obtain the updated model in Lombardy.

Compared with those in Wuhan, patients in New York and Lombardy were older and prevalent with comorbidities (Appendix Table 3). While Wuhan and New York had similar gender composition, Lombardy had a much higher of proportion of males (Appendix Table 3). With the same PI equaling to 0.566, patients in New York had similar mortality, while those in Lombardy had lower mortality compared with those in our derivation cohort in Wuhan. The final formula for calculating the survival probability for patient i is $CIF_i(t) = 1 - (1 - CIF_0(t))^{\exp(PI_i - \overline{PI})}$, where $CIF_0(t)$ is the mortality of the “average” patient in New York or Lombardy and is given in Appendix Table 4 and Table 5; respectively; PI_i is the prognostic index of patient i

and can be calculated by formula 1; \overline{PI} is the mean value of PIs in the derivation cohort and is 0.5662.

References

1. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks: Competing Risks and Risk Scores. *Statist Med.* 2016 Sep 30;35(22):4056–72.
2. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* [Internet]. 2020 Apr 22 [cited 2020 May 5]; Available from: <https://jamanetwork.com/journals/jama/fullarticle/2765184>
3. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA.* 2020 Apr 28;323(16):1574.
4. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Polit anal.* 2012;20(1):25–46.

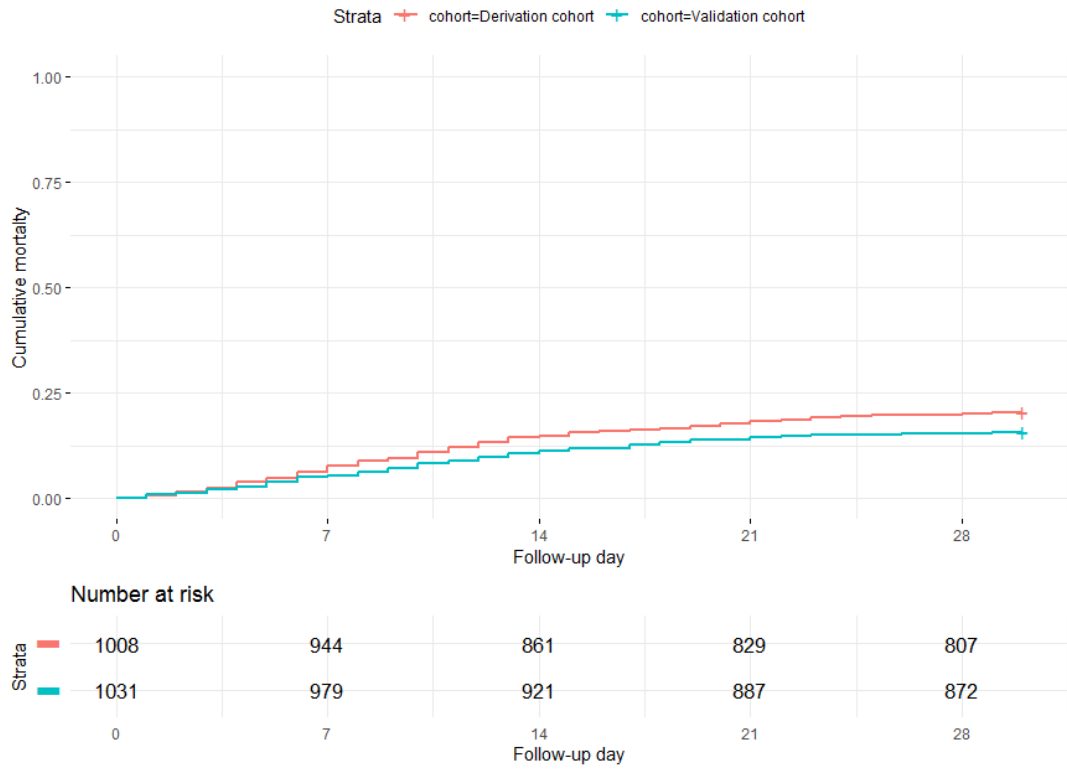


Figure 1. Cumulative mortality for derivation and validation cohort

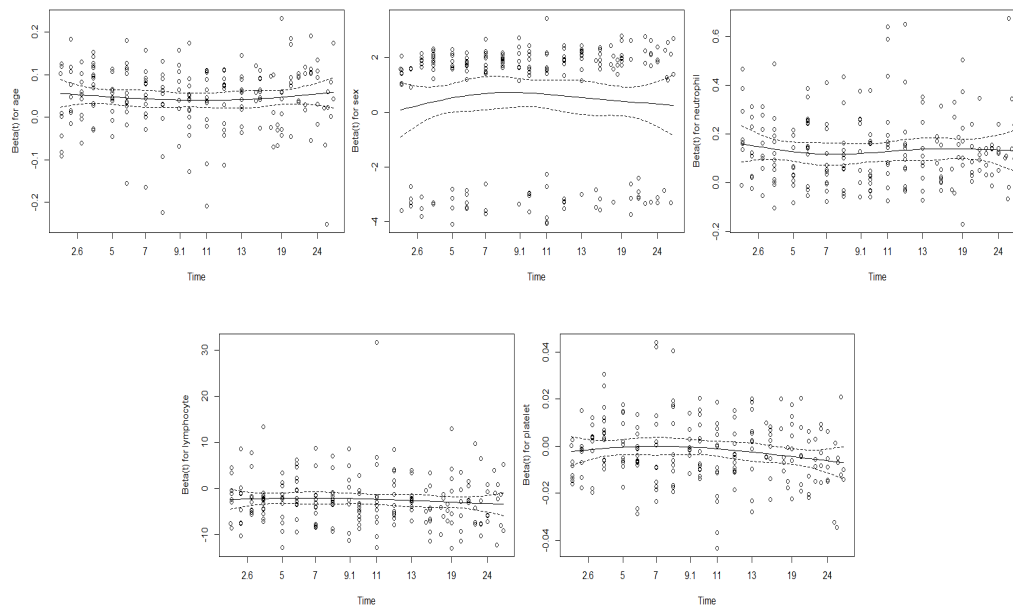


Figure 2. The Schoenfeld residual plots for each predictor, test of proportional hazards

Table 1. “Baseline”* mortality (Wuhan, China)

Day	Mortality	Day	Mortality
1	0.32%	16	8.38%
2	0.65%	17	8.58%
3	1.04%	18	8.79%
4	1.67%	19	9.29%
5	2.18%	20	9.74%
6	2.85%	21	10.12%
7	3.50%	22	10.43%
8	4.12%	23	10.82%
9	4.55%	24	11.22%
10	5.27%	25	11.31%
11	6.02%	26	11.39%
12	6.81%	27	11.47%
13	7.51%	28	11.63%
14	7.71%	29	11.80%
15	8.24%	30	11.97%

* The baseline refers to an “average” patient with PI 0.5662.

Table 2. Thresholds and corresponding proportion and death toll included in each risk group

Risk group	Range in PI	Derivation		Validation	
		Proportion	Death	Proportion	Death
1. Low risk	≤ -0.81	16.0%	3	15.6%	0
2. Low-Intermediate risk	-0.81 to 0.50	34.0%	17	32.5%	8
3. Moderate risk	0.50 to 2.03	34.0%	70	38.6%	65
4. High risk	> 2.03	16.0%	117	13.8%	89

Table 3. Basic characteristics used in entropy balancing in Derivation cohort, New York cohort and Lombardy cohort

	Derivation cohort (n=1008)	New York cohort (n=5700)	Lombardy cohort (n=1591)
Age, years	55 (44-65)	63 (52-70)	63 (56-70)
Sex, female	439 (43.6%)	2263 (39.7%)	287 (18%)
Hypertension	232 (23.0%)	3026 (56.6%)	509 (49%)
Coronary heart disease	32 (3.2%)	595 (11.1%)	223 (21%)
Diabetes	110 (10.9%)	1808 (33.8%)	180 (17%)
Malignancy	31 (3.1%)	320 (6%)	81 (8%)

Table 4. “Baseline”* mortality (New York, USA)

Day	Mortality	Day	Survival
1	0.33%	16	8.11%
2	0.95%	17	8.19%
3	1.33%	18	8.34%
4	2.04%	19	9.02%
5	2.42%	20	9.48%
6	3.31%	21	9.94%
7	4.01%	22	10.45%
8	4.43%	23	10.68%
9	4.82%	24	11.20%
10	5.40%	25	11.24%
11	6.04%	26	11.26%
12	6.52%	27	11.29%
13	7.19%	28	11.36%
14	7.36%	29	11.72%
15	8.01%	30	11.94%

* The baseline refers to an “average” patient with PI 0.5662.

Table 5. “Baseline”* mortality (Lombardy, Italy)

Day	Mortality	Day	Mortality
1	0.36%	16	7.48%
2	0.80%	17	7.59%
3	1.18%	18	7.72%
4	1.72%	19	8.34%
5	2.16%	20	8.60%
6	2.77%	21	8.72%
7	3.25%	22	9.51%
8	3.73%	23	9.82%
9	4.02%	24	10.11%
10	4.59%	25	10.13%
11	5.45%	26	10.16%
12	5.94%	27	10.21%
13	6.39%	28	10.32%
14	6.82%	29	10.44%
15	7.36%	30	10.72%

* The baseline refers to an “average” patient with PI 0.5662.

Table 6 Methodology quality assessment based on PROBAST risk of bias assessment tool

Question	Answer	Rationale
Domain 1: Participants		
1.1 Were appropriate data sources used, e.g., cohort, RCT or nested case-control study data?	Yes	Cohort design was used.
1.2 Were all inclusions and exclusions of participants appropriate?	Probably yes	The patients had no event (either death or discharge) at the end of follow-up date were excluded from the model development cohort. This is not model developers' choice, but due to data availability. Given the time between end of accrual and end of follow-up was 40 days, the actual proportion of excluded patients was very small, and the potential risk of bias can be mitigated.
Overall risk of bias of Domain 1	Low risk of bias	
Domain 2: Predictors		
2.1 Were predictors defined and assessed in a similar way for all participants?	Probably yes	Most of the predictors were patient characteristics and laboratory tests, which had standard definition. The scope of chronic disease history may be slightly different between the two hospitals. In general, all predictors can be considered as assessed in a similar way.
2.2 Were predictor assessments made without knowledge of outcome data?	Yes	All predictors were collected at patients admission.
2.3 Are all predictors available at the time the model is intended to be used?	Yes	All predictors included in the final model were patient characteristics and laboratory tests, which were easy to access.

Overall risk of bias of Domain 2	Low risk of bias	
Domain 3: Outcome		
3-1 Was the outcome determined appropriately?	Yes	In-hospital death was an event of interest for admitted patients with COVID-19. Discharge was properly considered as a competing risk event of in-hospital death, this is an advantage of this study compared with other studies (in which discharge was censored).
3-2 Was a pre-specified or standard outcome definition used?	Yes	In-hospital death was a hard end-point.
3-3 Were predictors excluded from the outcome definition?	Yes	In-hospital death was objective.
3-4 Was the outcome defined and determined in a similar way for all participants?	Yes	In-hospital death was objective.
3-5 Was the outcome determined without knowledge of predictor information?	Yes	In-hospital death was objective.
3-6 Was the time interval between predictor assessment and outcome determination appropriate?	Yes	All patients were staying in the hospitals and continuously followed up till discharge, death or the end of follow-up (which was 40 days after end of accrual). The time interval between predictor assessment and outcome determination is sufficient.
Overall risk of bias of Domain 3	Low risk of bias	
Domain 4: Analysis		
4-1 Were there a reasonable number of participants with the outcome?	Yes	The total number of candidate variables was 11, and the number of events was 211, so the number of events per variable (EPV) = $211/11 \approx 19$, which was reasonable.

4.2 Were continuous and categorical predictors handled appropriately?	Yes	Functional form of the relation between continuous variables and outcome was investigated, and suitable variable transformations were performed in case the linear assumption did not hold. No categorization was done for continuous variables.
4.3 Were all enrolled participants included in the analysis?	Yes	All patients met the inclusion criteria were included in the analysis.
4.4 Were participants with missing data handled appropriately?	Probably yes	Predictors with high percentage of missing value were excluded from the final model.
4.5 Was selection of predictors based on univariable analysis avoided?	Yes	Variable selection was based on multivariable analysis, and univariable was performed for investigating the non-linear relation.
4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of controls) accounted for appropriately?	Yes	Discharge was properly considered as a competing risk event of in-hospital death in the analysis.
4.7 Were relevant model performance measures evaluated appropriately?	Yes	Discrimination was assessed by C-statistics, and calibration curve was used to assess calibration.
4.8 Were model overfitting and optimism in model performance accounted for?	Yes	Optimism in model performance was adjusted with bootstrapping internal validation, a shrinkage factor was multiplied to the linear predictor to prevent overfitting.
4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	Yes	Coefficients of the multivariable Cox model was directly used in the calculation of the prognostic index (PI).
Overall risk of bias of Domain 4	Low risk of bias	

Part II: Contemporary issues in dynamic prediction