A COVID-19 Mortality Prediction Model for Korean Patients Using Nationwide Korean Disease Control and Prevention Agency Database

Yongho Jee (jyongho@eumc.ac.kr)
Advanced Biomedical Research Institute Ewha Womans University Seoul Hospital

Yi-Jun Kim
Institute of Convergence Medicine, Ewha Womans University Mokdong Hospital

Jongmin Oh
Department of Occupational and Environmental Medicine, Ewha Womans University College of Medicine, Seoul, Republic of Korea

Young-Ju Kim
Department of Obstetrics and Gynecology, Ewha Mokdong Hospital, Ewha Womans University, Seoul, Republic of Korea

Eun-Hee Ha
Department of Occupational and Environmental Medicine, Ewha Womans University College of Medicine, Seoul, Republic of Korea

Inho Jo
Department of Molecular Medicine, College of Medicine, Ewha Womans University, Seoul, Republic of Korea.

Research Article

Keywords: patients, confirmed, risk, covid, clinical

Posted Date: November 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-279161/v1

Version of Record: A version of this preprint was published at Scientific Reports on February 28th, 2022. See the published version at https://doi.org/10.1038/s41598-022-07051-4.

Abstract

The experience of the early nationwide COVID-19 pandemic in South Korea had led to an early shortage of medical resources. For efficient resource allocation, accurate prediction for the prognosis or mortality of confirmed patients is essential. Therefore, the aim of this study was to develop an accurate model for predicting COVID-19 mortality using epidemiolocal and clinical variables and for identifying high risk group of confirmed patients. Clinical and epidemiolocal variables of 4,049 patients with confirmed COVID-19 between January 20, 2020 and April 30, 2020 collected by Korean Disease Control and Prevention Agency were used. Among 4,049 total confirmed patients, 223 patients were dead while 3,826 patients were released from isolation. Patients who had the following risk factors showed significantly higher risk scores: age over 60 years, male, difficulty breathing, diabetes, cancer, dementia, change of consciousness, and hospitalized in intensive care unit. High accuracy was shown for both the development set (n = 2,467) and the validation set (n = 1,582), with AUC of 0.96 and 0.97, respectively. The prediction model developed in this study based on clinical features and epidemiological factors could be used for screening high risk group of patients and for evidence-based allocation of medical resources.

Introduction

Coronavirus Disease-2019 (COVID-19) has become a global pandemic that is threatening far more than a health crisis. It also affects societies and economics[1–3]. As the number of confirmed patients has explosively increased, there is a need for risk stratification both for preventing (i.e., home quarantine, social distancing) and for treating confirmed patients (i.e., hospitalization vs. community isolation) raised. While COVID-19 pandemic has caused almost 700,000 deaths worldwide, asymptomatic subpopulations also have a substantial number[4]. Identification of high risk confirmed patients is required to allow better allocation of existing available medical resources. According to Centers for Disease Control and Prevention (CDC), confirmed patients who are over 65 years old, who live in nursing homes, who have at least one of the following conditions: chronic lung disease, serious heart conditions, severe obesity, diabetes, liver disease, and immunocompromised are at a high risk of death by COVID-19[5]. Although the CDC guideline has been used as a reference for overall patients, more precise prediction using patient's multivariable data is required to evaluate individualized risk and to establish evidence for risk stratification[6]. In this context, an accurate model for predicting COVID-19 mortality and identifying risk factors could help stratify managements strategies for patients who have a high risk of death. Previous studies published in the early period of COVID-19 pandemic were not able to analyze individual level data. Barda et al. have established a prediction model by combining the development of a baseline respiratory infection risk predictor and a post-processing method using Israel data[7]. However, since they did not have individual data, they were not able to test its prediction performance. Since COVID 19 outbreak started from Hubei province of the People's Republic of China, the majority of early prediction studies were based on Chinese data[8–12]. According to Wynants' review on early reported prediction model

regarding COVID-19, these proposed models are poor with high risk of bias due to the lack of external validation of model[13].

Since South Korea is geopolitically close to China, it is one of countries most affected by COVID-19 during the early stage of the pandemic. In reality, Korea experienced explosive outbreak in the first two months since the first confirmed patient was detected on January 20,[14]. A mortality prediction model using machine method based on sociodemographic and medical information of national health insurance data has been proposed[15]. However, it was focused on socio-economic variables as predictors rather than clinical and epidemiological factors. Clinical experience and epidemiological characteristics have been reported as major factors associated with heterogeneity of prognosis after COVID-19 confirmation[16–17]. Therefore, the aim of this study was to establish a COVID-19 mortality prediction model using clinical and epidemiological variables nationally collected by Central Disease Control Headquarters.

Results

Baseline characteristics. Since the first patient confirmed with COVID-19 on January 20, 2020, 4,049 patients were managed by the government database and released from quarantine or dead until April 30, 2020. Among 4,049 released patients, the case mortality was 5.51% (223 deaths and 3,826 recoveries)

We compared the distribution of patients according to epidemiological and clinical characteristics. We also conducted a logistic regression analysis for mortality outcome by un-adjusting (univariable) or adjusting (multivariable) covariates. Results are shown in Table 1. In univariable analysis, age over 40, male sex, runny nose, and headache significantly increased the risk of mortality, while having abnormal change of consciousness (ACC), diabetes, hypertension, cancer history, dementia, and hospitalization in intensive care unit was protective. In multivariable analysis after adjusting for covariates, age over 40 and having s runny nose remained significant risk factors of mortality. Protective variables still remained protective after adjusting for covariates.

Factors associated with mortality from COVID-19. Table 2 summarizes differences in clinical characteristics for continuous variables and the risk of COVID-19 mortality by 1 unit increase of each clinical variable. Heart rate intensity (OR: 1.03, 95% CI: 1.02–1.04) and temperature (OR: 1.94, 95% CI: 1.55–2.43) were associated with an increased risk of COVID-19 mortality. Higher levels of hemoglobin, hematocrit, and lymphocytes were associated with a significantly lower risk of mortality. Based on exploratory analysis results shown in Tables 1 and 2, a prediction model for the development set was established as shown in Table 3. Odds ratio (regression coefficient) of mortality risk was determined to produce a risk score.

Table 1
Clinical and epidemiological characteristics of COVID-19 confirmed patients according to outcome status: categorical variable

		Quarantine release N (%)	Death N (%)	unadjusted OR (95% CI)	adjusted OR (95% CI)
Age	0–39	1144 (99.83%)	2 (0.17%)	1.0	1.0
(sex adjusted)	40–49	499 (99.60)	2 (0.40)	2.60 (0.37–18.55)	2.67 (0.22–33.09)
	50–59	839 (98.36)	14(1.64)	10.40 (2.35–45.90)	5.51 (0.67–45.38)
	60–69	732 (96.19)	29 (3.81)	23.49 (5.58–98.78)	8.48 (1.07–67.08)
	70–79	418 (86.36)	66 (13.64)	94.60(23.05–88.30)	17.61 (2.25-138.14)
	≥ 80	194 (63.82)	110 (36.18)	385.55 (94.09–999)	75.01(9.54–589)
Sex	Men	1454 (92.73)	114 (7.27)	1.0	1.0
(age adjusted)	Women	2372 (95.61)	109 (4.39)	0.45 (0.34–0.61)	0.54 (0.36–0.82)
Fever	Yes	849 (90.8)	86 (9.20)	1.0	1.0
	No	2977 (95.6)	137 (4.40)	0.37 (0.27–0.52)	0.73 (0.42–1.27)
Runny nose	Yes	386 (98.47)	6 (1.53)	1.0	1.0
	No	3440 (94.07)	217 (5.93)	2.87 (1.22–6.74)	2.95 (1.09–7.99)
SOB	Yes	453 (80.75)	108 (19.25)	1.0	1.0
	No	3373 (96.70)	115 (3.30)	0.22 (0.16–0.30)	0.38 (0.25–0.58)
headache	Yes	699 (98.45)	11 (1.55)	1.0	1.0
	No	3127 (93.65)	212 (6.35)	2.41 (1.28–4.57)	2.20 (1.02–4.74)
ACC	Yes	10 (31.25)	22 (68.75)	1.0	1.0
	No	3816 (95.00)	201 (5.00)	0.04 (0.01–0.10)	0.07 (0.02–0.21)

		Quarantine release N (%)	Death N (%)	unadjusted OR (95% CI)	adjusted OR (95% CI)
Diabetes	Yes	512 (84.91)	91 (15.09)	1.0	1.0
	No	3314 (96.17)	132 (3.83)	0.49 (0.36–0.66)	0.47 (0.31–0.72)
Hypertension	Yes	889 (86.90)	134 (13.10)	1.0	1.0
	No	2937 (97.06)	89 (2.94)	0.65 (0.47–0.89)	0.99 (0.66–1.50)
Heart failure	Yes	37 (69.81)	16 (30.19)	1.0	1.0
	No	3789 (94.82)	207 (5.18)	0.52 (0.27–1.03)	0.89 (0.36–2.21)
CKD	Yes	31 (65.96)	16 (34.04)	1.0	1.0
	No	3795 (94.83)	207 (5.17)	0.24 (0.11–0.52)	0.85 (0.44–1.67)
Cancer history	**Yes**	**113 (83.70)**	**22 (16.30)**	**1.0**	**1.0**
	No	3713 (94.86)	201 (5.14)	0.38 (0.22–0.66)	0.28 (0.13–0.57)
Dementia	Yes	138 (65.09)	74 (34.91)	1.0	1.0
	No	3688 (96.12)	149 (3.88)	0.40 (0.27–0.59)	0.21 (0.13–0.35)
sickbed	Yes	95 (55.23)	77 (44.77)	1.0	1.0
	No	3731 (96.23)	146 (3.77)	0.08 (0.05–0.12)	0.13 (0.08–0.21)

Table 2
Clinical and epidemiological characteristics of COVID-19 confirmed patients according to outcome status: continuous variable

	quarantine release Mean (Standard deviation)	Death Mean (Standard deviation)	OR (95% CI)
Systolic Blood Pressure	2.77 ± 1.33	2.99 ± 1.47	0.90 (0.80–1.01)
Diastolic blood pressure	2.01 ± 0.97	1.89 ± 1.01	0.91 (0.78–1.06)
Heart rate intensity	85.66 ± 15.00	89.40 ± 19.93	1.03 (1.02–1.04)
Temperature	36.92 ± 0.57	37.10 ± 0.76	1.94 (1.55–2.43)
hemoglobin (G/DL)	13.37 ± 1.69	11.76 ± 2.21	0.76 (0.69–0.82)
hematocrit (%)	39.51 ± 4.71	34.95 ± 6.68	0.91 (0.89–0.94)
Lymphocyte (%)	29.96 ± 11.18	15.34 ± 11.06	0.90 (0.88–0.92)

Table 3
Mortality Prediction Equation: Logistic model (Equation using development set:)

	Development set (N = 2,467) beta (se)
Intercept	20.3083 (2.6748)
Age 60–69	0.9596 (0.4714)
Age 70–79	1.4935 (0.4542)
Age ≥ 80	3.3010 (0.4538)
Men	-0.7845 (0.2828)
Fever	-0.8813 (0.2765)
shortness of breath	-0.9160 (0.2848)
Abnormal change of consciousness	-2.9806 (0.8839)
Dementia	-0.6318 (0.2689)
Cancer	-1.1150 (0.4675)
Dementia	-1.5940 (0.3426)
sickbed	-2.6019
Hematocrit	-0.0767 (0.0238)
Lymphocyte	-0.0694 (0.0131)

COVID 19 Mortality = 0.9596*(Age 60–69) + 1.4935*(Age 70–79) + 3.3010*(Age ≥ 80)-0.7845*(Sex)-0.8813*(Fever)-0.9160*(SOB)-2.9806*(ACC)-0.6318*(Diabetes)-1.1150*(Malignancy)-1.5940*(Dementia)-2.0619*(Sickbed type) -0.0767*(hematocrit)-0.0694*(Lymphocyte)

Performance of prediction model. We applied our risk score to our total set, the development set, and the validation set. Figures 1 to 3 show comparison results between the predicted mortality and the actual mortality by risk score stratified by decile.

Figure 1 shows results for the total set of participants.
Figure 2 describes results for the development set. Figure 3 shows results for the validation set. Performance of each dataset was evaluated using ROC curve. Results are shown in Fig. 4. Our prediction model showed fine performance for both the development set and the validation set, having area under the curve of 0.9656 and 0.9684, respectively.

Discussion

Our study developed and validated a COVID-19 mortality prediction model based on clinical and epidemiological data of COVID-19 4,049 confirmed patients recruited by Korea Centers for Disease Control and Prevention. The high AUC value of 0.9684 indicated a good reliability and performance of our model. The course of clinical symptoms of coronavirus ranges from asymptomatic infection to acute respiratory distress (ARDS) and death. As the period of COVID-19 global pandemic lasts longer, shortage of medical resources comes earlier. Therefore, differentiated patient management based on evidence is required. Risk stratification also suggests evidence to allocate resources efficiently when medical resources are limited[4]. Several previous Korean studies have reviewed characteristics of mortality cases of COVID-19. The Korean Society of Infectious Diseases and Korea Centers for Disease Control and Prevention has analyzed 54 COVID-19 mortality cases since the first mortality occurring from February 19 to March 10, 2020. The median age of mortality cases was 75.5 years. Of all mortality cases, 61.1% were men. The majority of such patients also had various underlying diseases such as hypertension, heart disease, diabetes, dementia, and stroke[18]. Another study reported in Korea was focused on 20 mortality cases in Gyeongbuk Province and Daegu city where the second outbreak wave occurred in February based on medical chart review[19]. Average age of mortality cases was 72 years. Of these mortality cases, 55.1% were women and 74.5% had an underlying disease. The median length from hospitalization to death was 8 days. Comorbidities such as diabetes, chronic lung disease, and chronic neurologic disease were significant risk factors associated with COVID-19 mortality. Clinical manifestations observed before death were abnormal heart rate intensity, systolic blood pressure, respiratory rate, oxygen saturated by pulse oximetry on room air, and altered mental status[19]. Although these two studies reported clinical characteristics of the deceased in detail at the level of descriptive epidemiology which contributed for overall understanding of COVID-19 patients, their numbers of cases were relatively small, was not enough for associational inference.

One study has developed an evidence-based COVID-19 prognostic model for military personnel in Korea[20]. Although there was a problem of generalization since it was developed for soldiers, age, body temperature, physical activity, history of cardiovascular disease, hypertension, visit to a region with an outbreak, feverishness, dyspnea, lethargy, and symptoms of chills were reported as significant predictors (overall C statistic: 0.963; 95% CI: 0.936–0.99) [15].

A COVID-19 mortality prediction model has been developed using machine learning after recruiting 10,237 COVID-19 confirmed patients and 228 mortality cases between January 20, 2020 and April 16, 2020[15]. This prediction model used various variables including socioeconomic status linked with National Health Insurance Service[15]. However, specific clinical and epidemiological variables were lacking since that study was focused on the linkage with NHIS data. For mortality prediction, LASSO and linear SVM were used in that study, with AUC values of 0.963 and 0.962, respectively[15]. The most significant factors in the mortality prediction model using LASSO were old age, preexisting DM, and cancer. The most significant factors in Random Forest were old age, infection route (cluster infection or infection

from personal contact), and underlying hypertension [15]. However, that model could not be immediately applied to the field or clinics due to the lack of specific clinical variables.

Previous foreign studies have reported that different clinical experience can lead to substantial heterogeneity in the prognostic trajectory of COIVD-19 confirmed patients spanning from patients who are asymptomatic to those with mild, moderate, and severe disease forms with low survival rates[21–23]. A COVID-19 mortality prediction model has been developed previously by analyzing data of 3,841 confirmed patients in New York, USA recruited from March 9 to April 6, 2020 using machine learning[20]. Sex, age, race, oxygen saturation, COPD, hypertension, and diabetes were found to be significant variables in that model with AUC of 0.91 to 0.94. However, blood test results were not included in that model. In that study, the minimum oxygen saturation was emphasized as a central factor in mortality prediction[20].

A study from Israel during the early period of COVID-19 pandemic estimated the risk of COVID-19 mortality when individual data were unavailable[7]. That study adopted a hybrid methodology under the hypothesis that the risk of severe respiratory infection or sepsis had a common etiology with the risk of COVID19. Major predictors were age, lymphocyte, and alnumin, with AUC value of 0.820[7]. Predictive factors found in the Israeli study were similar to those of our study. In terms of predictive power, the predictive power of the present study was much higher (at about 0.97).

A systematic review has been conducted based on 13 papers for the diagnosis and prognosis of COVID-19 infection. The majority of models used in that study failed to show sufficient performance as a predictive model due to a high risk of bias that required collaborative efforts with documented individual participant data[13].

A prediction model has been developed after analyzing 53,001 ICU patients requiring mechanical ventilation as well as those diagnosed with pneumonia from the US Medical Information Mart for Intensive Care (MIMIC). When that model was applied to 114 COVID-19 confirmed patients[24], AUC for 12, 24, 48, and 72 hours were reported to be 0.82, 0.81, 0.77, and 0.75, respectively[24]. Our study probably used the largest data set up to date to predict COVID-10 mortality involving specific clinical features of COVID-19 patients in Korea. The main advantage of our study was that we collected our clinical and epidemiological variables at the time confirmation was made. Results were obtained after a certain period of health system encounter or right after the diagnosis of COVID-19. Although we merely conducted logistic regression analysis, both development and validation sets showed high area under the curve (0.9656 and 0.9684, respectively).

Moreover, our model has the advantage of being able to easily interpret factors associated with high mortality rate of individuals according to the detailed algorithm shown in the model. In that context, our model has a high practical value for risk stratification in the clinical field.

The main limitation of our study was the issue of validation. Although our dataset was relatively large involving specific clinical features, we merely conducted an internal validation due to the lack of dataset

that had similar size and variables in Korea. Thus, the possibility of overestimation exists which requires cautious interpretation of our results. An external validation study using data of COVID-19 patients that occurred afterwards is required.

Subjects And Methods

Study population

Our study was based on the dataset established by Korean Disease Control and Prevention Agency Central Disease Countermeasure Headquarters. Individual level data for COVID-19 4,049 patients whose quarantine release was confirmed among patients infected between January 20, 2020 and April 30, 2020 were collected. Complete nationwide inpatient and outpatient data of patients who visited any medical institution with a confirmed diagnosis of COVID-19 during the study period were obtained. Definition of COVID-19 confirmation was determined by positive PCR-based clinical laboratory testing for SARS-Cov-2. Personal information de-identification measures were applied in accordance with governmental guidelines for non-identification measures and proceeded in accordance with adequacy evaluation.

Risk Factor Measurement

Collected data used in our study included 41 variables categorized to seven subtypes as follows: 1) basic data (age, sex, death/quarantine released, length of stay between infection to death/ quarantine released, pregnancy), 2) body index (height, weight), 3) initial examination finding (systolic/diastolic blood pressure, heart rate, body temperature), 4) clinical findings at hospitalization (history of fever, cough, sputum production, sore throat, runny nose/rhinorrhea, muscle aches/myalgia, fatigue/malaise, shortness of breath/dyspnea, headache, altered consciousness/confusion, vomiting/nausea, diarrhea), 5) comorbidity and past history (diabetes, hypertension, heart failure, chronic heart condition, asthma, chronic obstructive pulmonary disease, chronic kidney failure, cancer, chronic hepatic disease, rheumatism, autoimmune disease, dementia), 6) sickbed type and clinical severity, and 7) complete blood cell count. Each variable was either self-reported or recorded by professional health care providers. Mortality was defined when a patient with COVID-19 died during their encounter with the health system during the study period (January 1, 2020 ~ April 30, 2020).

Data usage and study design of our study were approved by the Institutional Review Board of Ewha Womans University Seoul Hospital with informed consent obtained from each subject (SEUMC 2020-09-009).

Statistical analysis

Risk scores for our COVID-19 mortality prediction model were developed by logistic regression analysis. We stratified our data into two groups: a 60% random sampling (development set data) for model

development and the remaining 40% (test data set) for internal validation. All analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

Declarations

Acknowledgments: This study was supported by the Korea Disease Control and Prevention Agency, KDCA by establishing dataset named “Clinical epidemiology information of COVID- 19 confirmed patients”. This study was supported by the 2020 Korea Medical Institute Research and Development Project fund.

Author contributions

Y. Jee planned the study, performed all statistical analyses, and wrote the paper. Y.J. Kim supported data analysis and contributed to revising the paper. J. Oh, Y. Kim, E. H Ha and I. Jo, helped to plan the study and to revise the manuscript.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to Y. Jee.

References

1. Mahase, E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ.* **368**, m641 (2020).
2. Ranney, M. L., Griffeth, V. & Jha, A. K. Critical supply shortages—the need for ventilators and personal protective equipment during the Covid-19 pandemic. *N. Engl. J. Med.* **382**, (2020). e41
3. Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).
4. Centers for Disease Control and Prevention. People Who Are at Higher Risk for Severe Illness, https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html (2020).
5. Chen, J. H. & Asch, S. M. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N. Engl. J. Med.* **376**, 2507–2509 (2017).
6. Wynants, L. *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj, 369(2020).
7. Caramelo, F., Ferreira, N. & Oliveiros, B. Estimation of risk factors for COVID-19 mortality-preliminary results.MedRxiv. Preprint Posted,25(2020).
8. Lu, J. *et al.* ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China medRxiv preprint doi: https://doi.org/10.1101/2020.02.20.20025510; (2020).

9. Shi, Y. *et al.* Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. *Critical care.* **24** (1.), 1–4 (2020).
10. Xie, J. *et al.* Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19.medRxiv(2020).
11. Yan, L. *et al.* Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. MedRxiv(2020).
12. Kim, J. Y. *et al.* The first case of 2019 novel coronavirus pneumonia imported into Korea from Wuhan, China: implication for infection prevention and control measures. *J Korean Med Sci.* **35** (5), e61 (2020).
13. An, C. *et al.* S. W. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. *Scientific reports.* **10** (1), 1–11 (2020).
14. Guan, W. J. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* **382**, 1708–1720 (2020).
15. Richardson, S. *et al.* Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA; 323: 2052–59(2020).
16. Park, S. *et al.* Analysis on 54 mortality cases of coronavirus disease 2019 in the Republic of Korea from January 19 to March 10, 2020. *J. Korean Med. Sci.* **35**, e132 (2020).
17. Lee, J. Y. *et al.* Risk factors for mortality and respiratory support in elderly patients hospitalized with COVID-19 in Korea.Journal of Korean Medical Science, 35(23) (2020).
18. Yadaw, A. S. *et al.* Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet Digital Health.* **2** (10), e516–e525 (2020).
19. Guan, W. J. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* **382**, 1708–1720 (2020).
20. Heo, J. *et al.* COVID-19 Outcome Prediction and Monitoring Solution for Military Hospitals in South Korea: Development and Evaluation of an Application. *Journal of medical Internet research.* **22** (11), e22131 (2020).
21. Barda, N. *et al.* Developing a COVID-19 mortality risk prediction model when individual-level data are not available. *Nature communications.* **11** (no. 1), 1–9 (2020).
22. Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E.,… van Smeden, M. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. bmj, 369 (2020).
23. Ryan, L. *et al.* Mortality prediction model for the triage of COVID-19, pneumonia, and mechanically ventilated ICU patients: a retrospective study. *Annals of Medicine and Surgery.* **59**, 207–216 (2020).

Figures

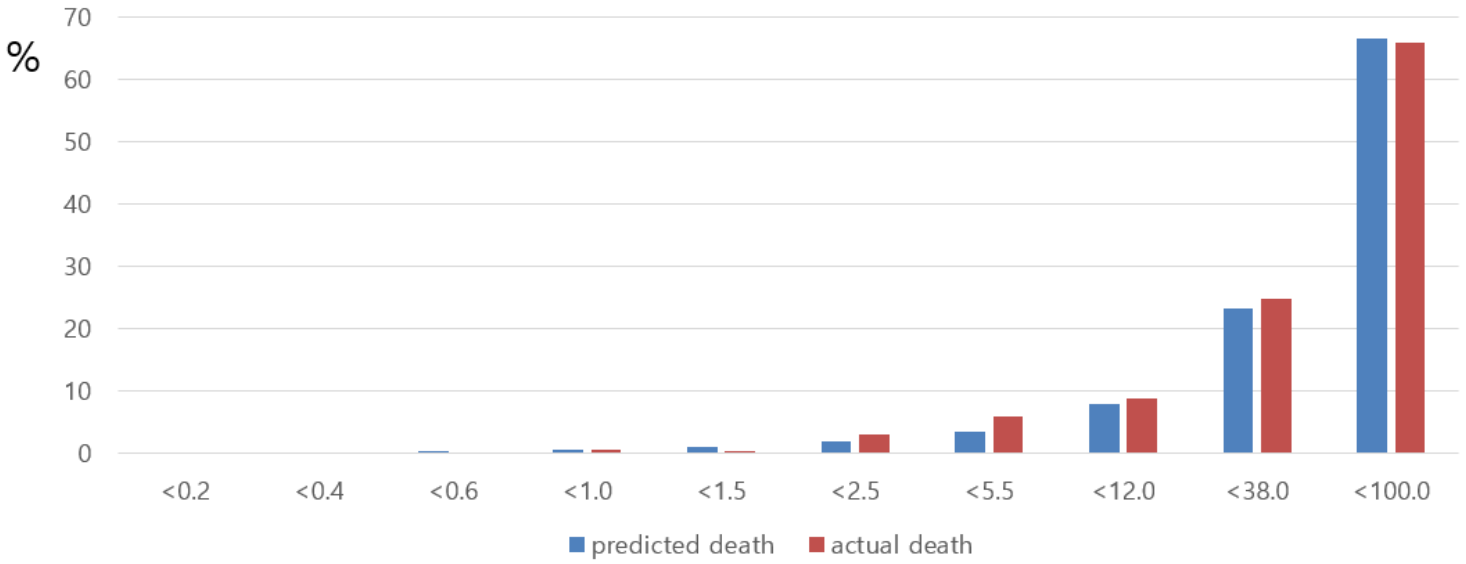

Figure 1

shows results for the total set of participants.

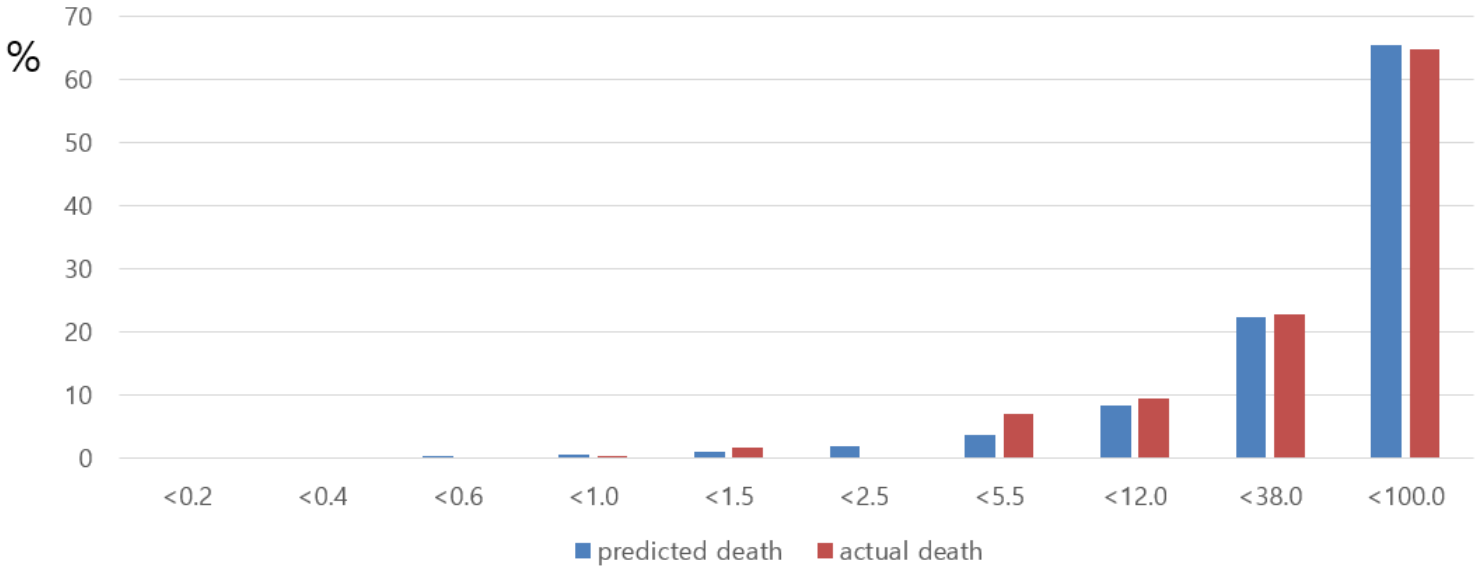

Figure 2

describes results for the development set.

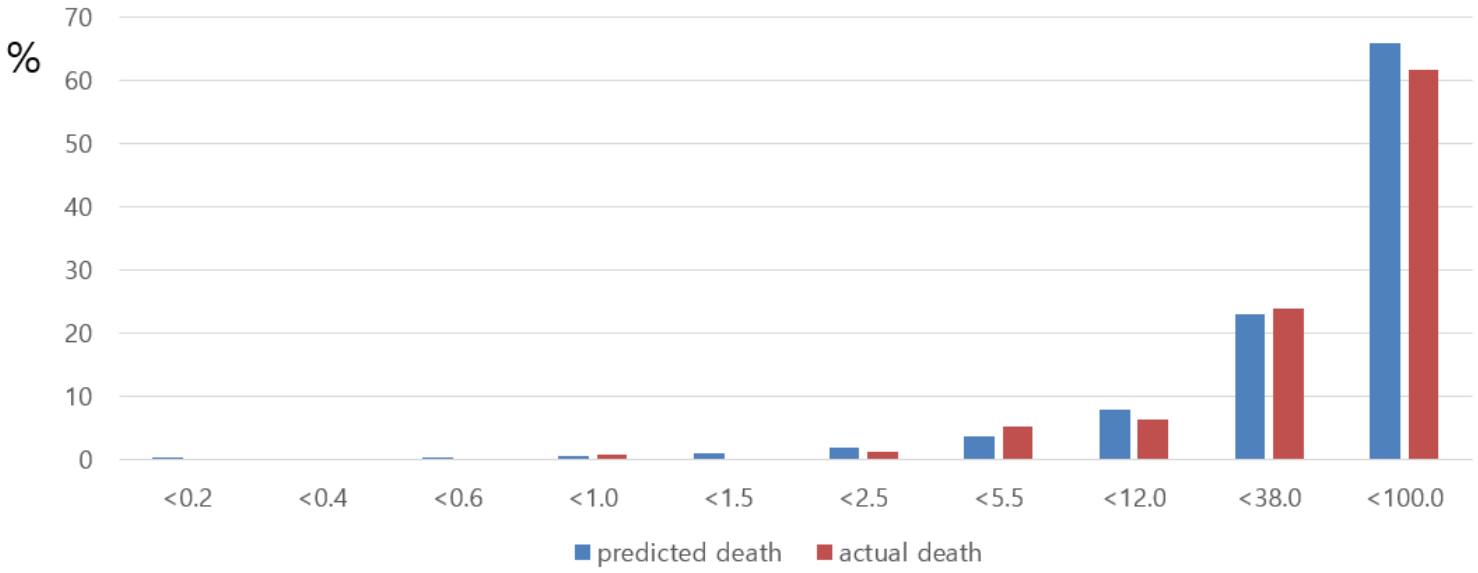

Figure 3

shows results for the validation set.

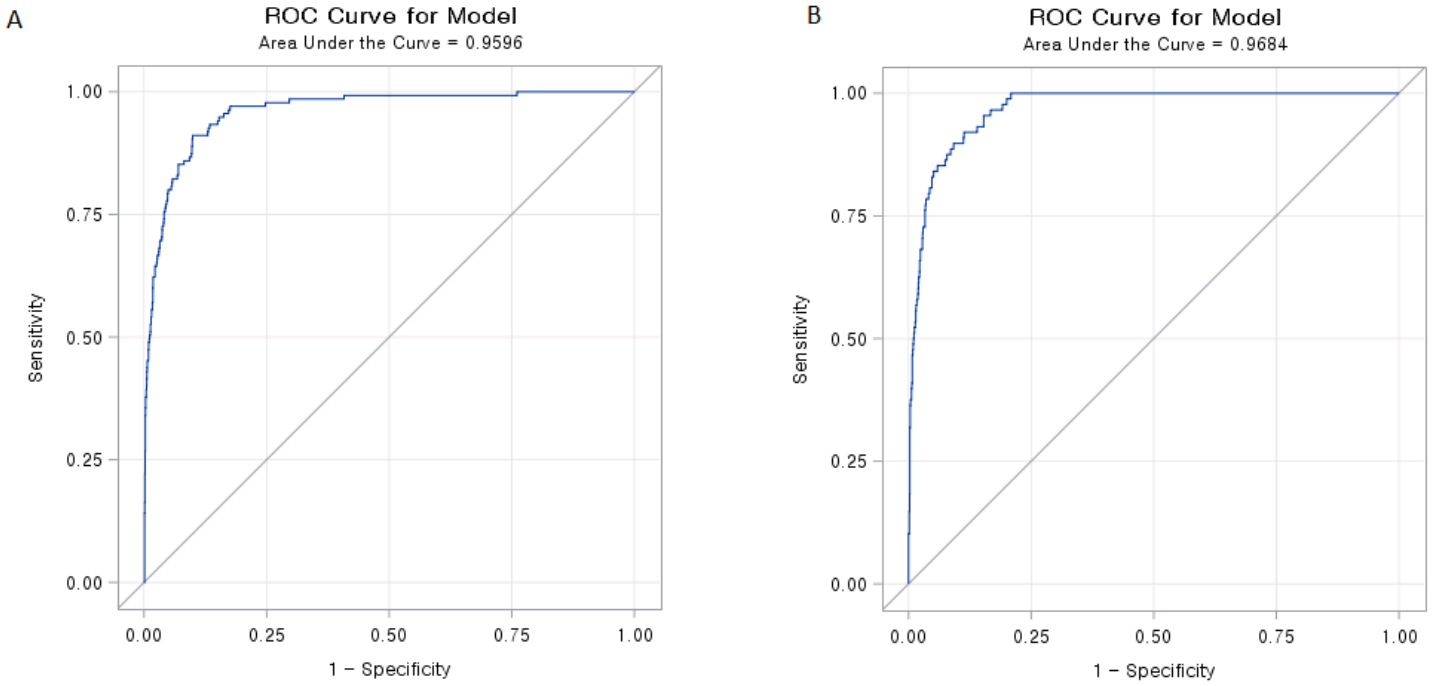

Figure 4

Performance of each dataset was evaluated using ROC curve.