Internal Medicine
The Japanese Society of Internal Medicine
doi: 10.2169/internalmedicine.0086-22
Intern Med 62: 201-213, 2023
http://internmed.jp

【 ORIGINAL ARTICLE 】

Mortality Prediction of COVID-19 in Hospitalized Patients Using the 2020 Diagnosis Procedure Combination Administrative Database of Japan

Shuko Nojiri[1-3], Yoshiki Irie[2,4], Rie Kanamori[3], Toshio Naito[5] and Yuji Nishizaki[1,3,5-7]

Abstract:

Objectives Numerous people have died from coronavirus disease 2019 (COVID-19) infection. Identifying crucial predictive biomarkers of disease mortality is critical to support decision-making and logistic planning in healthcare systems. This study investigated the association between mortality and medical factors and prescription records in 2020 in Japan, where COVID-19 prevalence and mortality remain relatively low.

Methods This retrospective cohort study analyzed anonymous administrative data from the Diagnosis Procedure Combination (DPC) database in Japan.

Results A total of 22,795 patients were treated in DPC hospitals in 2020 in Japan, and of these, 5,980 patients over 50 years old were hospitalized, with 299 (5.0%) dying. There were 2,399 severe patients among 11,440 total hospitalized patients (all ages). The results of a logistic model analysis revealed that an older age, male sex, Parkinson's disease, cerebrovascular diseases, and chronic kidney diseases were risk factors for mortality. A machine learning analysis identified an older age, male sex (mortality), pneumonia, drugs for acid-related disorders, analgesics, anesthesia, upper respiratory tract disease, drugs for functional gastrointestinal disorders, drugs for obstructive airway diseases, topical products for joint and muscular pain, diabetes, lipid-modifying agents, calcium channel blockers, drugs for diabetes, and agents acting on the renin-angiotensin system as risk factors for a severe status.

Conclusions This COVID-19 mortality risk tool is a well-calibrated and accurate model for predicting mortality risk among hospitalized patients with COVID-19 in Japan, which is characterized by a relatively low COVID-19 prevalence, aging society, and high population density. This COVID-19 mortality prediction model can assist in resource utilization and patient and caregiver education and be useful as a risk stratification instrument for future research trials.

Key words: COVID-19, DPC database, Japan, machine learning

(Intern Med 62: 201-213, 2023)
(DOI: 10.2169/internalmedicine.0086-22)

Introduction

The novel coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a pandemic. Although the first outbreak was attributed to zoonotic transmission in Wuhan, China, human-to-human transmission through respiratory droplets and aerosolization has resulted in rapid disease spread worldwide. Currently, the delta variant is fueling a surge in new COVID-19 cases globally, and accumulated COVID-19 infection cases in Japan reached more than

[1]Medical Technology Innovation Center, Juntendo University, Japan, [2]Clinical Research and Trial Center, Juntendo University, Japan, [3]Clinical Translational Science, Juntendo University Graduate School of Medicine, Japan, [4]Department of Information and Computer Technology, Graduate School of Engineering, Tokyo University of Science, Japan, [5]Department of General Medicine, Juntendo University School of Medicine, Japan, [6]Department of Cardiovascular Biology and Medicine, Juntendo University Graduate School of Medicine, Japan and [7]Division of Medical Education, Juntendo University School of Medicine, Japan
Received: April 10, 2022; Accepted: September 21, 2022; Advance Publication by J-STAGE: November 2, 2022
Correspondence to Dr. Shuko Nojiri, s-nojiri@juntendo.ac.jp

1,340,000, including 15,700 deaths, as of August 25, 2021 (1).

Clinical presentations of COVID-19 are heterogeneous, ranging from mild flu-like symptoms, such as a fever, cough, and fatigue, to severe respiratory symptoms and hypoxia, resulting in acute respiratory distress syndrome (2, 3). Patients with severe disease require intensive care and/or mechanical ventilation to prevent multisystem organ failure and death. The rate of needing intensive-care unit (ICU) transfer among hospitalized patients with COVID-19 is 33%, which is significantly higher than that among other hospitalized patients (11%) (4, 5).

Most countries, including Japan, are experiencing daily shortages of medical resources, such as medical staff, hospital beds, and personal protective equipment. Germany has only 8.2 such beds per 100,000 population, while Italy has 3.6 such beds. Although Japan has the highest number of hospital beds per population worldwide (13.7 per 100,000 population in 2018), the number of beds is still inadequate to accommodate all COVID-19 cases. Thus, Tokyo, the capital city of Japan, has started to quarantine people with mild symptoms in hotels. Japan, which also boasts the world's most rapidly aging population, has many nursing homes, but staffing shortages have become a major social problem, and some accommodation-type and day-care-type facilities have reported cluster infections.

Risk stratification is important to improve the usage of available resources and thereby improve patient outcomes. Several studies have attempted to clarify the relationship between risk factors and clinical prognoses in order to risk-stratify patients and prepare to allocate available-yet-limited healthcare resources. The Centers for Disease Control and Prevention (CDC) defined the following criteria as being associated with a high risk for a severe status: age ≥65 years, living in a nursing home, and having at least one of the conditions chronic lung disease, serious heart conditions, severe obesity, diabetes, chronic kidney disease, and liver disease or an immunocompromised state (6). Globally, observational studies (7, 8) have proposed that patients who are older or have various comorbidities, such as diabetes, cardiovascular disease, and hypertension, have a higher risk of in-hospital mortality from COVID-19 than others.

Regarding biomarkers, ferritin, lactate dehydrogenase (LDH), D-dimer, and C-reactive protein (CRP) reportedly predict COVID-19 severity (7, 8) Many studies have attempted to create prediction models that combine several variables to estimate the prognosis, including the use of scoring systems and machine learning. Unfortunately, many of these models are suboptimal because of the high risk of bias, restricted sample sizes, and limited number of outcomes of interest.

The clinical and radiologic characteristics of severe COVID-19 have been summarized (7). Recent reports examined the reasons for the low prevalence of COVID-19 in Japan. Iwasaki et al. proposed five hypotheses involving Japanese culture, previous exposure to a milder version of SARS-CoV-2 that conferred herd immunity, susceptibility reduction resulting from angiotensin-converting enzyme 2 (ACE2) receptor expression, distinct human leukocyte antigen that confers immune resistance to COVID-19, and Bacillus Calmette-Guérin vaccination (9). Given the low COVID-19 prevalence, the prognostic value of different variables remains unclear in Japan.

The present study therefore developed and validated a prognostic model based on the clinical and laboratory variables of patients with COVID-19 obtained from administrative data in Japan.

Materials and Methods

This retrospective cohort study used administrative databases as sources of information. We analyzed the following data collected from hospitalization records at Diagnosis Procedure Combination (DPC) hospitals in Japan: admission dates, healthcare beneficiaries, sex, birth date, residence, death date, and laboratory data. Chronic comorbidities were identified using the International Classification of Diseases 10th Revision (ICD-10) codes, with data from the outpatient care database and the drug prescription databases in addition to the abovementioned database. In the final analysis, immunosuppression/transplant and human immunodeficiency virus (HIV) patients were excluded because of their extremely small number. To clarify the mortality risk factors, the mortality rate was known to be high in the elderly, so we restricted the study patients to those over 50 years old for mortality risk assessments. For the severity risk analysis, the study patients were not restricted by age, as shown in Fig. 1.

Drugs were classified using the codes defined by the Anatomical Therapeutic Chemical Classification System, as summarized in Supplementary material 1b. Polypharmacy was defined as the administration of more than five medications within six months before COVID-19 infection hospitalization. Severity was defined based on a claim history of oxygen inhalation, high-flow therapy, and requirement for artificial respiration during hospitalization (Supplementary material 1c). We considered the severity to be Moderate II if only oxygen inhalation "140005610" in Supplementary material 1c was used, while if other procedures in Supplementary material 1c were performed, then the severity was considered to be Severe. Severe cases were defined as those with a Moderate II+Severe state, while non-Severe cases were those with a Mild+Moderate I state, according to the claim data.

The institutional review board of Juntendo University approved this study. The study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Statistical analyses

Baseline data are expressed as the median, and categorical variables are expressed as the frequency (%). Differences

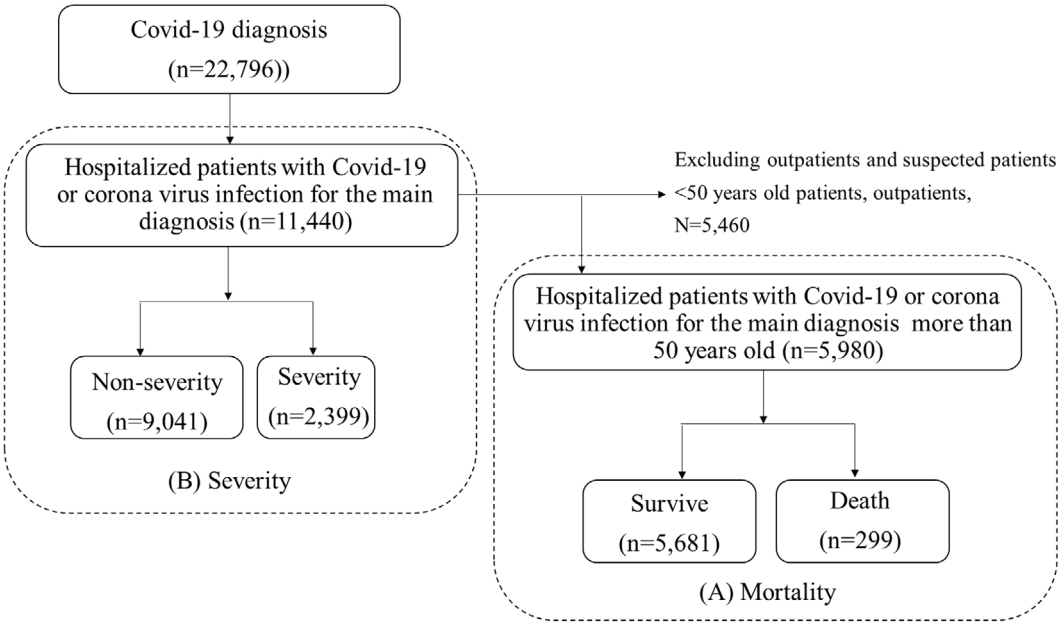

Figure 1. Study diagram indicating patient flow. (A) Hospitalized patients with COVID-19 infection (>50 years old). (B) Hospitalized patients with COVID-19 infection (all ages). We considered the severity as Moderate II if only oxygen inhalation "140005610" in Supplementary material 1c was used, and if other procedures in Supplementary material 1c were performed, then the severity was Severe. Severe cases were defined as those with a Moderate II+Severe state, while non-Severe cases were those with a Mild+Moderate I state, according to the claim data.

between the Severe group (moderate II+severe) and Non-severe (mild+moderate I) group were assessed using the Mann-Whitney *U* test for continuous data and Fisher's exact test for categorical data.

Patients' COVID-19 characteristics from administrative records were collected using several statistical methods, such as a logistic model and machine learning. First, from the development cohort, all patients hospitalized for COVID-19 were included in variable selection. We entered 67 variables into the selection process. The potential collinearity of variables measured from the same patient and the overfitting of variables were minimized using Least Absolute Shrinkage and Selection Operator (LASSO) regression, which reduce complexity by forcing less influential variables to have zero influence on the model. Recently, the concept of actively using sparsity to achieve simpler models has received much attention in fields such as statistical learning, data mining, and signal processing (10). We used L1-penalized LASSO regression for multivariable analyses, augmented with five-fold cross-validation for internal validation. A logistic regression model penalized the absolute size of the coefficients of a regression model according to the λ value. Only the strongest predictors remained in the model with larger penalties, and the estimates of weaker factors decreased to zero. The most predictive covariates were selected by the minimum (λ min). LASSO regression was performed using the R package "glmnet" statistical software program (R Foundation, R 4.0.1). The logistic model for identifying the risk factors of mortality and a severe disease state included clinically important variables, such as age, sex, cancer, diabetes, lipidemia, hypertension, dementia, schizophrenia, depression, anxiety, Parkinson's disease, cerebrovascular diseases, myocardial infraction, cardiac arrhythmia, peripheral artery disease, pneumonia, chronic obstructive pulmonary disease (COPD), asthma, liver disease, and chronic kidney disease (CKD) (Supplementary material 1).

Second, we applied the machine learning method. A gradient-boosting machine model built with decision tree base-learners was used to generate predictions; this model is utilized by many successful algorithms in the field of machine learning (11). Missing values were inherently handled by the gradient-boosting predictor trained with the XGBoost Python package. The training-validation set consisted of records from tested individuals within the study period and was further divided into training and validation sets at a 4:1 ratio. All p values were 2-sided, and p values <0.05 were considered significant.

All statistical analyses, modeling, and plotting were performed in R (version 3.5.3). The model was scored on the test set using the area under the receiver operating characteristic (ROC) curve (auROC). In addition, plots of the positive predictive values against sensitivity (precision-recall curve) were drawn across different thresholds. Metrics were calculated for all the thresholds from all ROC curves. Confidence intervals (CIs) for the various performance measures were derived through resampling using the bootstrap percentile method with 1,000 repetitions.

Results

Between January 1 and December 31, 2020, 22,796 patients were hospitalized due to COVID-19 infection. When the study population was limited to patients >50 years old and outpatients and those with suspected COVID-19 were excluded (Fig. 1), the number of hospitalized patients 50-65 years old was 2,550 (43%). A total of 2,241 (37%) patients were between 65 and 80 years old, and 1,189 (20%) patients were over 80 years old. Men accounted for 58% of the study population. In addition, 299 patients died during the first hospital admission for COVID-19 infection, and the mortality and severity rates were 5.0% and 34.7%, respectively (Table 1a). For a severe status, (Table 1b) shows that diabetes, chronic lung diseases, COPD, CKD, drugs for acid related disorders (A02), drugs for functional gastrointestinal disorders (A03), diuretics (C03), anti-inflammatory and antirheumatic products (M01), topical products for joint and muscular pain (M02) and age showed a statistically significant difference between the non-severe (mild+moderate I) and severe (moderate II+severe) patients.

Fig. 2 shows that most patients were in their 50s. The age distribution among patients who died of COVID-19 showed that death was highest in older patients (>75 years old) (Fig. 2A). In the Severe group, the distribution of age differed from that for mortality, starting among those in their 20s and peaking among those in their 70s (Fig. 2B). Differences in prognoses were found if patients' medical history included any of following: cancer, anemia, diabetes, hypertension, Parkinson's disease, cardiac arrhythmia, peripheral artery diseases, heart failure, chronic lung diseases, ulcer, or COPD (Table 1a). Furthermore, pharmaceutical treatment with the following classes of drugs before COVID-19 infection was associated with differences in prognoses: drugs for functional gastrointestinal disorders, drugs for constipation, drugs for diabetes, vitamin and mineral supplements, antithrombotic agents, diuretics, vasoprotectives, beta-blockers, calcium-channel blockers, lipid modifiers, anesthetics, antiemetic drugs, anti-Parkinson's disease drugs, and drugs for COPD ($p<0.05$) (Table 1a). The risk of Severe disease was affected by age, gender, and presence of diabetes, upper respiratory tract diseases, chronic lung diseases, COPD, CKD, drugs for acid-related disorders and functional gastrointestinal disorders, diuretics, anti-inflammatory and antirheumatic products, and topical products for joint and muscular pain (Table 1b).

As shown in Table 2, the mortality risk assessment using the LASSO model identified age, gender, and use of anti-Parkinson's disease drug as risk factors, and the auROC generated from the LASSO model analysis predicted mortality in hospitalized patients with COVID-19 [0.80; 95% confidence interval (CI): 0.778-0.822] (Fig. 3A). A Severe disease risk assessment using the LASSO model found that age, gender, pneumonia, COPD, CKD, drugs for diabetes, diuretics, and drugs for obstructive airway diseases were significant risk factors, and the corresponding auROC predicted mortality in hospitalized patients with COVID-19 with the LASSO model (0.80; 95% CI: 0.77-0.82) and with the XGBoost model (0.79; 95% CI: 0.77-0.81) (Fig. 3A).

In the prospective test set used for the machine learning method, older age and male sex were the strongest factors for mortality risk, and older age, male sex, pneumonia, drugs for acid-related disorders, analgesics, anesthetics, upper respiratory tract diseases, drugs for functional gastrointestinal disorders, drugs for obstructive airway diseases, topical products for joint and muscular pain, diabetes, lipid-modifying agents, calcium channel blockers, drugs for diabetes, and agents acting on the renin-angiotensin system were identified as risk factors for a severe disease state, with the corresponding auROC predicting mortality in hospitalized patients with COVID-19.

The results of the logistic model analysis for mortality, in which clinically important variables were forcibly added to the final model, identified Parkinson's disease (OR 3.57; 95% CI 1.08-10.2), CKD (OR 2.2; 95% CI 0.99-4.58), male sex (OR 2.4; 95% CI 1.85-3.14), and age as risk factors for mortality (Supplementary material 2a). Regarding the risk factors for a severe disease state, the results according to a logistic model analysis were older age, male sex, cancer, Parkinson disease, cerebrovascular diseases, peripheral artery diseases, pneumonia, and CKD (Supplementary material 2b).

Discussion

Our COVID-19 risk assessment model for the Japanese population accurately estimated the mortality risk of patients hospitalized for COVID-19. Presented by means of SHapley Additive exPlanations (SHAP), well-defined variables, such as an older age, male sex (mortality), pneumonia, drugs for acid-related disorders, analgesics, anesthetics, upper respiratory tract diseases, drugs for functional gastrointestinal disorders, drugs for obstructive airway diseases, topical products for joint and muscular pain, diabetes, lipid-modifying agents, calcium channel blockers, drugs for diabetes, and agents acting on the renin-angiotensin system were identified as risk factors for a severe disease state.

In the LASSO model, the COVID-19 mortality risk factors were older age, male sex, anti-Parkinson's disease drug use, and mineral supplement use, while severity was associated with older age, male sex, pneumonia, COPD, CKD, drugs for diabetes, mineral supplements, diuretics, and drugs for obstructive airway diseases. The present study developed a model that could identify patients at high risk for mortality and severe disease before the occurrence of irreversible clinical consequences during hospitalization. Using an independent COVID-19-positive population as a validation dataset, we showed that our prognostic model was consistent in predicting mortality risk.

In Japan, the first confirmed case of SARS-CoV-2 infection was recorded on January 16, 2020, when a Chinese national who had visited Wuhan tested positive (12). Subse-

Table 1. (a) Baseline Data (Mortality).

Characteristic	Overall, n=5,980	Death, n=299	Survival, n=5,681	p value
Influenza	360 (6.0%)	19 (6.4%)	341 (6.0%)	0.8
Cancer	**393 (6.6%)**	**29 (9.7%)**	**364 (6.4%)**	**0.025**
Anemia	**152 (2.5%)**	**16 (5.4%)**	**136 (2.4%)**	**0.002**
Diabetes	**451 (7.5%)**	**37 (12%)**	**414 (7.3%)**	**0.001**
Lipidemia	422 (7.1%)	28 (9.4%)	394 (6.9%)	0.11
Hypertension	**563 (9.4%)**	**43 (14%)**	**520 (9.2%)**	**0.003**
Dementia	80 (1.3%)	8 (2.7%)	72 (1.3%)	0.061
Schizophrenia	44 (0.7%)	2 (0.7%)	42 (0.7%)	>0.9
Depression & anxiety	138 (2.3%)	10 (3.3%)	128 (2.3%)	0.2
Parkinson's disease	**30 (0.5%)**	**5 (1.7%)**	**25 (0.4%)**	**0.015**
Sleeping disorder	268 (4.5%)	14 (4.7%)	254 (4.5%)	0.9
Cerebrovascular diseases	215 (3.6%)	14 (4.7%)	201 (3.5%)	0.3
Cardiovascular diseases	252 (4.2%)	17 (5.7%)	235 (4.1%)	0.2
Myocardial infraction	245 (4.1%)	16 (5.4%)	229 (4.0%)	0.3
Cardiac arrhythmia	**195 (3.3%)**	**16 (5.4%)**	**179 (3.2%)**	**0.037**
Peripheral artery diseases	**197 (3.3%)**	**16 (5.4%)**	**181 (3.2%)**	**0.041**
Heart failure	**354 (5.9%)**	**31 (10%)**	**323 (5.7%)**	**<0.001**
Upper respiratory tract diseases	379 (6.3%)	13 (4.3%)	366 (6.4%)	0.15
Pneumonia	307 (5.1%)	18 (6.0%)	289 (5.1%)	0.5
Acute lower tract infection	173 (2.9%)	12 (4.0%)	161 (2.8%)	0.2
Chronic lung diseases	**123 (2.1%)**	**12 (4.0%)**	**111 (2.0%)**	**0.014**
COPD	**42 (0.7%)**	**7 (2.3%)**	**35 (0.6%)**	**0.004**
Asthma	140 (2.3%)	8 (2.7%)	132 (2.3%)	0.7
Ulcer	**696 (12%)**	**46 (15%)**	**650 (11%)**	**0.038**
Liver diseases	158 (2.6%)	13 (4.3%)	145 (2.6%)	0.059
Rheumatoid diseases	43 (0.7%)	2 (0.7%)	41 (0.7%)	>0.9
Gout	167 (2.8%)	10 (3.3%)	157 (2.8%)	0.6
Kidney diseases	87 (1.5%)	8 (2.7%)	79 (1.4%)	0.079
Chronic kidney disease	**92 (1.5%)**	**11 (3.7%)**	**81 (1.4%)**	**0.006**
Allergy	206 (3.4%)	7 (2.3%)	199 (3.5%)	0.3
Polyvascular diseases	150 (2.5%)	9 (3.0%)	141 (2.5%)	0.6
Polypharmacy	524 (8.8%)	34 (11%)	490 (8.6%)	0.1
A02 DRUGS FOR ACID RELATED DISORDERS	713 (12%)	44 (15%)	669 (12%)	0.13
A03 DRUGS FOR FUNCTIONAL GASTROINTESTINAL DISORDERS	**404 (6.8%)**	**30 (10%)**	**374 (6.6%)**	**0.021**
A04 ANTIEMETICS AND ANTINAUSEANTS	**37 (0.6%)**	**6 (2.0%)**	**31 (0.5%)**	**0.009**
A05 BILE AND LIVER THERAPY	60 (1.0%)	6 (2.0%)	54 (1.0%)	0.12
A06 DRUGS FOR CONSTIPATION	**417 (7.0%)**	**32 (11%)**	**385 (6.8%)**	**0.009**
A07 ANTIDIARRHEALS, INTESTINAL ANTIINFLAMMATORY/ ANTIINFECTIVE AGENTS	**210 (3.5%)**	**17 (5.7%)**	**193 (3.4%)**	**0.036**
A09 DIGESTIVES, INCL. ENZYMES	17 (0.3%)	1 (0.3%)	16 (0.3%)	0.6
A10 DRUGS USED IN DIABETES	**214 (3.6%)**	**19 (6.4%)**	**195 (3.4%)**	**0.008**
A11 VITAMINS	**268 (4.5%)**	**25 (8.4%)**	**243 (4.3%)**	**<0.001**
A12 MINERAL SUPPLEMENTS	**71 (1.2%)**	**12 (4.0%)**	**59 (1.0%)**	**<0.001**
B01 ANTITHROMBOTIC AGENTS	**409 (6.8%)**	**31 (10%)**	**378 (6.7%)**	**0.013**
B02 ANTIHEMORRHAGICS	211 (3.5%)	16 (5.4%)	195 (3.4%)	0.08
B03 ANTIANEMIC PREPARATIONS	108 (1.8%)	9 (3.0%)	99 (1.7%)	0.11
C01 CARDIAC THERAPY	276 (4.6%)	24 (8.0%)	252 (4.4%)	0.004
C02 ANTIHYPERTENSIVES	42 (0.7%)	1 (0.3%)	41 (0.7%)	0.7
C03 DIURETICS	**137 (2.3%)**	**12 (4.0%)**	**125 (2.2%)**	**0.041**
C04 PERIPHERAL VASODILATORS	18 (0.3%)	0 (0%)	18 (0.3%)	>0.9
C05 VASOPROTECTIVES	**152 (2.5%)**	**14 (4.7%)**	**138 (2.4%)**	**0.016**
C07 BETA BLOCKING AGENTS	**163 (2.7%)**	**15 (5.0%)**	**148 (2.6%)**	**0.013**
C08 CALCIUM CHANNEL BLOCKERS	**298 (5.0%)**	**22 (7.4%)**	**276 (4.9%)**	**0.053**
C09 AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	270 (4.5%)	19 (6.4%)	251 (4.4%)	0.12
C10 LIPID MODIFYING AGENTS	**286 (4.8%)**	**23 (7.7%)**	**263 (4.6%)**	**0.016**
H02 CORTICOSTEROIDS FOR SYSTEMIC USE	229 (3.8%)	17 (5.7%)	212 (3.7%)	0.086
M01 ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS	479 (8.0%)	29 (9.7%)	450 (7.9%)	0.3
M02 TOPICAL PRODUCTS FOR JOINT AND MUSCULAR PAIN	305 (5.1%)	17 (5.7%)	288 (5.1%)	0.6
M03 MUSCLE RELAXANTS	193 (3.2%)	11 (3.7%)	182 (3.2%)	0.7
N01 ANESTHETICS	637 (11%)	42 (14%)	595 (10%)	0.051
N02 ANALGESICS	564 (9.4%)	36 (12%)	528 (9.3%)	0.11
N03 ANTIEPILEPTICS	**109 (1.8%)**	**12 (4.0%)**	**97 (1.7%)**	**0.004**
N04 ANTI-PARKINSON DRUGS	**25 (0.4%)**	**7 (2.3%)**	**18 (0.3%)**	**<0.001**
N05 PSYCHOLEPTICS	340 (5.7%)	24 (8.0%)	316 (5.6%)	0.073
N06 PSYCHOANALEPTICS	49 (0.8%)	4 (1.3%)	45 (0.8%)	0.3
N07 OTHER NERVOUS SYSTEM DRUGS	288 (4.8%)	20 (6.7%)	268 (4.7%)	0.12
R01 NASAL PREPARATIONS	66 (1.1%)	4 (1.3%)	62 (1.1%)	0.6
R03 DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	**225 (3.8%)**	**23 (7.7%)**	**202 (3.6%)**	**<0.001**
R05 COUGH AND COLD PREPARATIONS	389 (6.5%)	26 (8.7%)	363 (6.4%)	0.12
R06 ANTIHISTAMINES FOR SYSTEMIC USE	219 (3.7%)	14 (4.7%)	205 (3.6%)	0.3
Male	**3,439 (58%)**	**202 (68%)**	**3,237 (57%)**	**<0.001**
Age less than 65 yeas old	**2,550 (43%)**	**17 (5.7%)**	**2,533 (45%)**	**<0.001**
Age between 65 and 80 years old	**2,241 (37%)**	**111 (37%)**	**2,130 (37%)**	
Age more than 80 years old	**1,189 (20%)**	**171 (57%)**	**1,018 (18%)**	

Table 1. (b) Baseline Data (Severity Risk).

Characteristic	Overall, n=11,440	Non-severe, n=9,041	Severe, n=2,399	p value
Influenza	360 (6.0%)	225 (5.8%)	135 (6.5%)	0.300
Cancer	393 (6.6%)	267 (6.8%)	126 (6.1%)	0.200
Anemia	152 (2.5%)	102 (2.6%)	50 (2.4%)	0.600
Diabetes	**451 (7.5%)**	**274 (7.0%)**	**177 (8.5%)**	**0.037**
Lipidemia	422 (7.1%)	272 (7.0%)	150 (7.2%)	0.700
Hypertension	563 (9.4%)	354 (9.1%)	209 (10%)	0.200
Dementia	80 (1.3%)	48 (1.2%)	32 (1.5%)	0.300
Schizophrenia	44 (0.7%)	25 (0.6%)	19 (0.9%)	0.200
Depression & anxiety	138 (2.3%)	93 (2.4%)	45 (2.2%)	0.600
Parkinson's disease	30 (0.5%)	18 (0.5%)	12 (0.6%)	0.500
Sleeping disorder	268 (4.5%)	178 (4.6%)	90 (4.3%)	0.700
Cerebrovascular diseases	215 (3.6%)	144 (3.7%)	71 (3.4%)	0.600
Cardiovascular diseases	252 (4.2%)	164 (4.2%)	88 (4.2%)	>0.9
Myocardial infraction	245 (4.1%)	159 (4.1%)	86 (4.1%)	>0.9
Cardiac arrhythmia	195 (3.3%)	119 (3.0%)	76 (3.7%)	0.200
Peripheral artery diseases	197 (3.3%)	127 (3.3%)	70 (3.4%)	0.800
Heart failure	354 (5.9%)	218 (5.6%)	136 (6.5%)	0.140
Upper respiratory tract diseases	**379 (6.3%)**	**275 (7.0%)**	**104 (5.0%)**	**0.002**
Pneumonia	307 (5.1%)	185 (4.7%)	122 (5.9%)	0.059
Acute lower tract infection	173 (2.9%)	119 (3.0%)	54 (2.6%)	0.300
Chronic lung diseases	**123 (2.1%)**	**69 (1.8%)**	**54 (2.6%)**	**0.031**
COPD	**42 (0.7%)**	**20 (0.5%)**	**22 (1.1%)**	**0.016**
Asthma	140 (2.3%)	86 (2.2%)	54 (2.6%)	0.300
Ulcer	696 (12%)	459 (12%)	237 (11%)	0.700
Liver diseases	158 (2.6%)	103 (2.6%)	55 (2.6%)	>0.9
Rheumatoid diseases	43 (0.7%)	31 (0.8%)	12 (0.6%)	0.300
Gout	167 (2.8%)	106 (2.7%)	61 (2.9%)	0.600
Kidney diseases	87 (1.5%)	52 (1.3%)	35 (1.7%)	0.300
Chronic kidney disease	**92 (1.5%)**	**50 (1.3%)**	**42 (2.0%)**	**0.027**
Allergy	206 (3.4%)	143 (3.7%)	63 (3.0%)	0.200
Polyvascular diseases	150 (2.5%)	102 (2.6%)	48 (2.3%)	0.500
Polypharmacy	524 (8.8%)	354 (9.1%)	170 (8.2%)	0.200
A02 DRUGS FOR ACID RELATED DISORDERS	**713 (12%)**	**490 (13%)**	**223 (11%)**	**0.038**
A03 DRUGS FOR FUNCTIONAL GASTROINTESTINAL DISORDERS	**404 (6.8%)**	**283 (7.3%)**	**121 (5.8%)**	**0.036**
A04 ANTIEMETICS AND ANTINAUSEANTS	37 (0.6%)	26 (0.7%)	11 (0.5%)	0.500
A05 BILE AND LIVER THERAPY	60 (1.0%)	39 (1.0%)	21 (1.0%)	>0.9
A06 DRUGS FOR CONSTIPATION	417 (7.0%)	270 (6.9%)	147 (7.1%)	0.800
A07 ANTIDIARRHEALS, INTESTINAL ANTIINFLAMMATORY/ ANTIINFECTIVE AGENTS	210 (3.5%)	145 (3.7%)	65 (3.1%)	0.200
A09 DIGESTIVES, INCL. ENZYMES	17 (0.3%)	14 (0.4%)	3 (0.1%)	0.140
A10 DRUGS USED IN DIABETES	214 (3.6%)	127 (3.3%)	87 (4.2%)	0.065
A11 VITAMINS	268 (4.5%)	174 (4.5%)	94 (4.5%)	>0.9
A12 MINERAL SUPPLEMENTS	71 (1.2%)	39 (1.0%)	32 (1.5%)	0.066
B01 ANTITHROMBOTIC AGENTS	409 (6.8%)	262 (6.7%)	147 (7.1%)	0.600
B02 ANTIHEMORRHAGICS	211 (3.5%)	141 (3.6%)	70 (3.4%)	0.600
B03 ANTIANEMIC PREPARATIONS	108 (1.8%)	71 (1.8%)	37 (1.8%)	>0.9
C01 CARDIAC THERAPY	276 (4.6%)	178 (4.6%)	98 (4.7%)	0.800
C02 ANTIHYPERTENSIVES	42 (0.7%)	26 (0.7%)	16 (0.8%)	0.600
C03 DIURETICS	**137 (2.3%)**	**77 (2.0%)**	**60 (2.9%)**	**0.024**
C04 PERIPHERAL VASODILATORS	18 (0.3%)	15 (0.4%)	3 (0.1%)	0.110
C05 VASOPROTECTIVES	152 (2.5%)	101 (2.6%)	51 (2.5%)	0.800
C07 BETA BLOCKING AGENTS	163 (2.7%)	99 (2.5%)	64 (3.1%)	0.200
C08 CALCIUM CHANNEL BLOCKERS	298 (5.0%)	180 (4.6%)	118 (5.7%)	0.071
C09 AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	270 (4.5%)	162 (4.2%)	108 (5.2%)	0.064
C10 LIPID MODIFYING AGENTS	286 (4.8%)	174 (4.5%)	112 (5.4%)	0.110
H02 CORTICOSTEROIDS FOR SYSTEMIC USE	229 (3.8%)	152 (3.9%)	77 (3.7%)	0.700
M01 ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS	**479 (8.0%)**	**336 (8.6%)**	**143 (6.9%)**	**0.019**
M02 TOPICAL PRODUCTS FOR JOINT AND MUSCULAR PAIN	**305 (5.1%)**	**218 (5.6%)**	**87 (4.2%)**	**0.019**
M03 MUSCLE RELAXANTS	193 (3.2%)	135 (3.5%)	58 (2.8%)	0.200
N01 ANESTHETICS	637 (11%)	434 (11%)	203 (9.8%)	0.110
N02 ANALGESICS	564 (9.4%)	389 (10.0%)	175 (8.4%)	0.051
N03 ANTIEPILEPTICS	109 (1.8%)	68 (1.7%)	41 (2.0%)	0.500
N04 ANTI-PARKINSON DRUGS	25 (0.4%)	13 (0.3%)	12 (0.6%)	0.200
N05 PSYCHOLEPTICS	340 (5.7%)	225 (5.8%)	115 (5.5%)	0.700
N06 PSYCHOANALEPTICS	49 (0.8%)	32 (0.8%)	17 (0.8%)	>0.9
N07 OTHER NERVOUS SYSTEM DRUGS	288 (4.8%)	188 (4.8%)	100 (4.8%)	>0.9
R01 NASAL PREPARATIONS	66 (1.1%)	49 (1.3%)	17 (0.8%)	0.120
R03 DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	225 (3.8%)	135 (3.5%)	90 (4.3%)	0.092
R05 COUGH AND COLD PREPARATIONS	389 (6.5%)	264 (6.8%)	125 (6.0%)	0.300
R06 ANTIHISTAMINES FOR SYSTEMIC USE	219 (3.7%)	143 (3.7%)	76 (3.7%)	>0.9
Male	**3,439 (58%)**	**2,137 (55%)**	**1,302 (63%)**	**<0.001**
Age less than 65 years old	**2,550 (43%)**	**1,945 (50%)**	**605 (29%)**	**<0.001**
Age between 65 and 80 years old	**2,241 (37%)**	**1,355 (35%)**	**886 (43%)**	
Age more than 80 years old	**1,189 (20%)**	**602 (15%)**	**587 (28%)**	

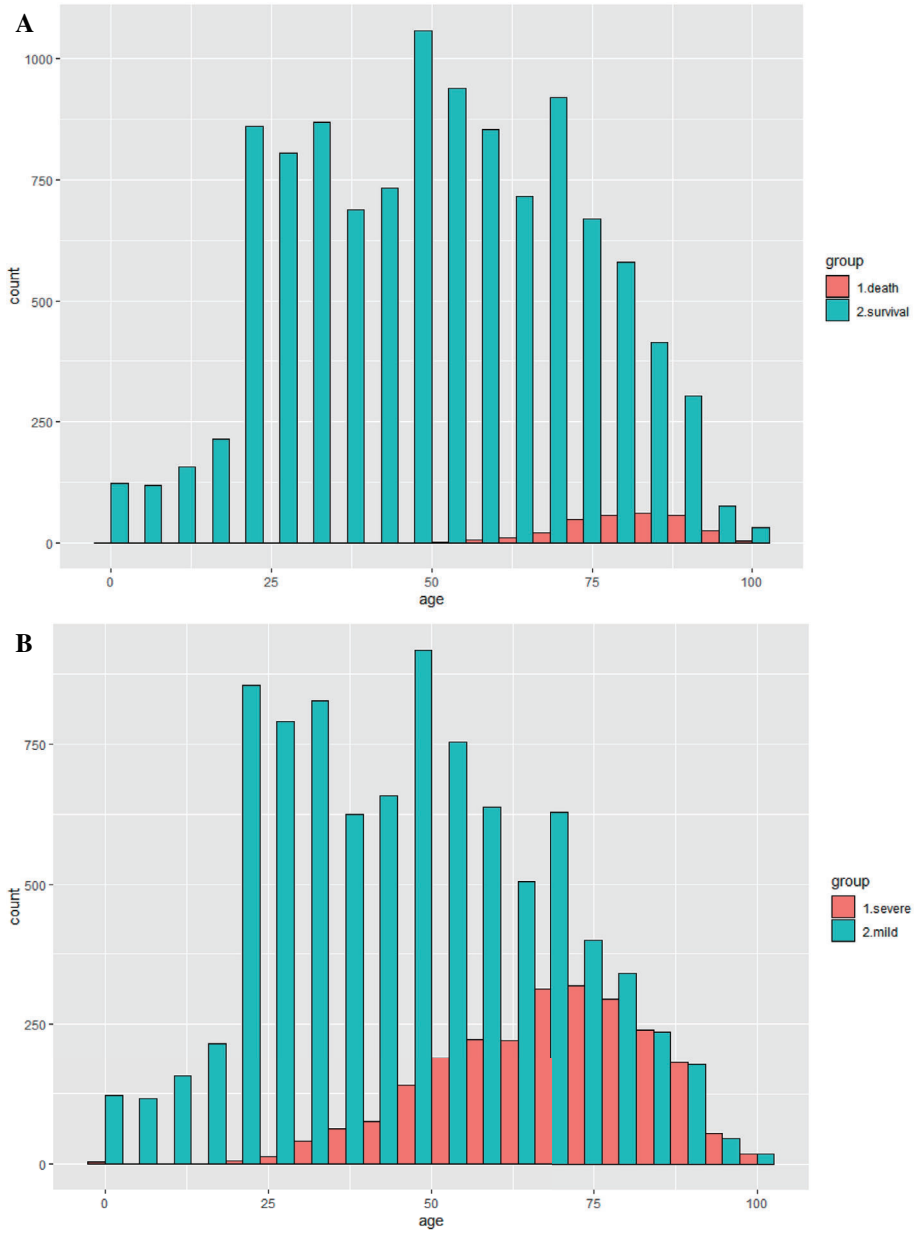

Figure 2. **Age distribution of hospitalized patients with COVID-19. (A) Mortality. Y-axis indicates the number of patients. (B) Severity. Y-axis indicates the number of patients. We considered the severity as Moderate II if only oxygen inhalation "140005610" in Supplementary material 1c was used, and if other procedures in Supplementary material 1c were performed, then the severity was Severe. Severe cases were defined as those with a Moderate II+Severe state, while non-Severe cases were those with a Mild+Moderate I state, according to the claim data.**

quently, many of the 3,711 passengers and crew members of a cruise liner named "Diamond Princess" were found to be infected on February 3, 2020; in a retrospective, single-center study involving 104 patients with laboratory-confirmed COVID-19, Tabata et al. reported that the LDH level was a potential predictor of symptom onset in COVID-19 patients and that an older age, consolidation on chest computed tomography, and lymphopenia might be risk factors for disease progression of COVID-19 (13).

Despite the high prevalence and mortality rates of COVID-19 in other regions, including the USA, South America, and Europe, Japan has continued to have low numbers of confirmed COVID-19 cases (Supplementary material 3). Japan has not applied for expansive testing across the nation, nor has strict contact tracing been implemented, as in Singapore and Hong Kong, but the growth rate of COVID-19 infection has been slow and the mortality rate low. In the present study, the mortality rate was 2.17%, which is considerably lower than that in other studies. For example, the US American Heart Association COVID-19 Cardiovascular Disease Registry study reported that out of 20,736 patients hospitalized for COVID-19 treatment between March and November of 2020, 3,271 (15.8%) died in the hospital (14).

Table 2. **(a) LASSO Model Analysis for the Mortality Risk.**

Variable	Estimate	Penalty
(Intercept)	-3.20668467	0.01
Influenza	0	0.01
Cancer	0	0.01
Anemia	0	0.01
Diabetes	0	0.01
Lipidemia	0	0.01
Hypertension	0	0.01
Dementia	0	0.01
Schizophrenia	0	0.01
Depression & anxiety	0	0.01
Parkinson's disease	0	0.01
Sleeping disorder	0	0.01
Cerebrovascular diseases	0	0.01
Myocardial infraction	0	0.01
Cardiac arrhythmia	0	0.01
Peripheral artery diseases	0	0.01
Upper respiratory tract diseases	0	0.01
Pneumonia	0	0.01
Acute lower tract infection	0	0.01
COPD	0	0.01
Asthma	0	0.01
Ulcer	0	0.01
Liver diseases	0	0.01
Rheumatoid diseases	0	0.01
Gout	0	0.01
Kidney diseases	0	0.01
Chronic kidney disease	0	0.01
Allergy	0	0.01
Polyvascular diseases	0	0.01
Polypharmacy	0	0.01
A02 DRUGS FOR ACID RELATED DISORDERS	0	0.01
A03 DRUGS FOR FUNCTIONAL GASTROINTESTINAL DISORDERS	0	0.01
A04 ANTIEMETICS AND ANTINAUSEANTS	0	0.01
A05 BILE AND LIVER THERAPY	0	0.01
A06 DRUGS FOR CONSTIPATION	0	0.01
A07 ANTIDIARRHEALS, INTESTINAL ANTIINFLAMMATORY/ANTIINFECTIVE AGENTS	0	0.01
A09 DIGESTIVES, INCL. ENZYMES	0	0.01
A10 DRUGS USED IN DIABETES	0	0.01
A11 VITAMINS	0	0.01
A12 MINERAL SUPPLEMENTS	**0.01559565**	**0.01**
B01 ANTITHROMBOTIC AGENTS	0	0.01
B02 ANTIHEMORRHAGICS	0	0.01
B03 ANTIANEMIC PREPARATIONS	0	0.01
C01 CARDIAC THERAPY	0	0.01
C02 ANTIHYPERTENSIVES	0	0.01
C03 DIURETICS	0	0.01
C04 PERIPHERAL VASODILATORS	0	0.01
C05 VASOPROTECTIVES	0	0.01
C07 BETA BLOCKING AGENTS	0	0.01
C08 CALCIUM CHANNEL BLOCKERS	0	0.01
C09 AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	0	0.01
C10 LIPID MODIFYING AGENTS	0	0.01
H02 CORTICOSTEROIDS FOR SYSTEMIC USE	0	0.01
M01 ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS	0	0.01
M02 TOPICAL PRODUCTS FOR JOINT AND MUSCULAR PAIN	0	0.01
M03 MUSCLE RELAXANTS	0	0.01
N01 ANESTHETICS	0	0.01
N02 ANALGESICS	0	0.01
N03 ANTIEPILEPTICS	0	0.01
N04 ANTI-PARKINSON DRUGS	**0.8593094**	**0.01**
N05 PSYCHOLEPTICS	0	0.01
N06 PSYCHOANALEPTICS	0	0.01
N07 OTHER NERVOUS SYSTEM DRUGS	0	0.01
R01 NASAL PREPARATIONS	0	0.01
R03 DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	0	0.01
R05 COUGH AND COLD PREPARATIONS	0	0.01
R06 ANTIHISTAMINES FOR SYSTEMIC USE	0	0.01
Male	**0.30830282**	**0.01**
Age less than 65 yeas old	**-0.9893284**	**0.01**
Age between 65 and 80 years old	0	0.01
Age more than 80 years old	**1.10548867**	**0.01**

Table 2. **(b) LASSO Model Analysis (Severity Risk).**

Variable	Estimate	Penalty
(Intercept)	-0.64249928	0.01
Influenza	0	0.01
Cancer	0	0.01
Anemia	0	0.01
Diabetes	0	0.01
Lipidemia	0	0.01
Hypertension	0	0.01
Dementia	0	0.01
Schizophrenia	0	0.01
Depression & anxiety	0	0.01
Parkinson's disease	0	0.01
Sleeping disorder	0	0.01
Cerebrovascular diseases	0	0.01
Myocardial infraction	0	0.01
Cardiac arrhythmia	0	0.01
Peripheral artery diseases	0	0.01
Upper respiratory tract diseases	-0.062951955	0.01
Pneumonia	**0.0447631**	**0.01**
Acute lower tract infection	0	0.01
COPD	**0.01874581**	**0.01**
Asthma	0	0.01
Ulcer	0	0.01
Liver diseases	0	0.01
Rheumatoid diseases	0	0.01
Gout	0	0.01
Kidney diseases	0	0.01
Chronic kidney disease	**0.08575941**	**0.01**
Allergy	0	0.01
Polyvascular diseases	0	0.01
Polypharmacy	0	0.01
A02 DRUGS FOR ACID RELATED DISORDERS	-0.053957032	0.01
A03 DRUGS FOR FUNCTIONAL GASTROINTESTINAL DISORDERS	0	0.01
A04 ANTIEMETICS AND ANTINAUSEANTS	0	0.01
A05 BILE AND LIVER THERAPY	0	0.01
A06 DRUGS FOR CONSTIPATION	0	0.01
A07 ANTIDIARRHEALS, INTESTINAL ANTIINFLAMMATORY/ANTIINFECTIVE AGENTS	0	0.01
A09 DIGESTIVES, INCL. ENZYMES	-0.153266672	0.01
A10 DRUGS USED IN DIABETES	**0.00109647**	**0.01**
A11 VITAMINS	0	0.01
A12 MINERAL SUPPLEMENTS	0	0.01
B01 ANTITHROMBOTIC AGENTS	0	0.01
B02 ANTIHEMORRHAGICS	0	0.01
B03 ANTIANEMIC PREPARATIONS	0	0.01
C01 CARDIAC THERAPY	0	0.01
C02 ANTIHYPERTENSIVES	0	0.01
C03 DIURETICS	**0.00481299**	**0.01**
C04 PERIPHERAL VASODILATORS	-0.065621736	0.01
C05 VASOPROTECTIVES	0	0.01
C07 BETA BLOCKING AGENTS	0	0.01
C08 CALCIUM CHANNEL BLOCKERS	0	0.01
C09 AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM	0	0.01
C10 LIPID MODIFYING AGENTS	0	0.01
H02 CORTICOSTEROIDS FOR SYSTEMIC USE	0	0.01
M01 ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS	0	0.01
M02 TOPICAL PRODUCTS FOR JOINT AND MUSCULAR PAIN	-0.081828354	0.01
M03 MUSCLE RELAXANTS	0	0.01
N01 ANESTHETICS	-0.102688582	0.01
N02 ANALGESICS	0	0.01
N03 ANTIEPILEPTICS	0	0.01
N04 ANTI-PARKINSON DRUGS	0	0.01
N05 PSYCHOLEPTICS	0	0.01
N06 PSYCHOANALEPTICS	0	0.01
N07 OTHER NERVOUS SYSTEM DRUGS	0	0.01
R01 NASAL PREPARATIONS	0	0.01
R03 DRUGS FOR OBSTRUCTIVE AIRWAY DISEASES	**0.00258742**	**0.01**
R05 COUGH AND COLD PREPARATIONS	0	0.01
R06 ANTIHISTAMINES FOR SYSTEMIC USE	0	0.01
Male	**0.39344302**	**0.01**
Age less than 65 yeas old	**-0.7094203**	**0.01**
Age between 65 and 80 years old	0	0.01
Age more than 80 years old	**0.4003379**	**0.01**

We considered the severity as Moderate II if only oxygen inhalation "140005610" in Supplementary material 1c was used, and if other procedures in Supplementary material 1c were performed, then the severity was Severe. Severe cases were defined as those with a Moderate II+Severe state, while non-Severe cases were those with a Mild+Moderate I state, according to the claim data.

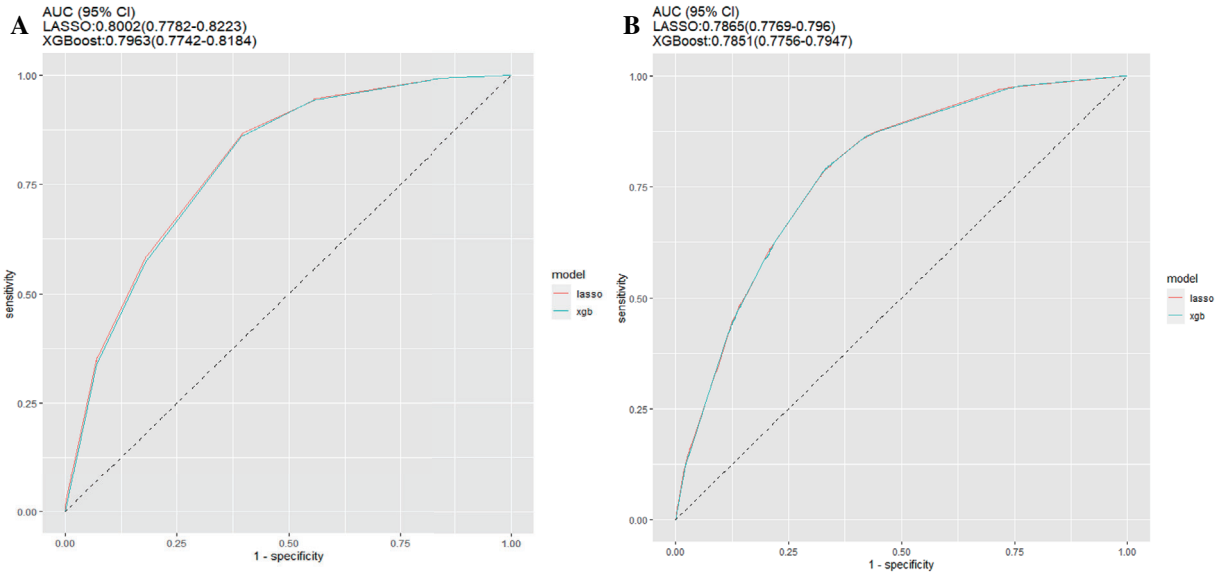

Figure 3. Model performance. ROC curves of the mortality predictive model on the prospective test set. (A) Mortality (>50 years old). (B) Severity (all ages). We considered the severity as Moderate II if only oxygen inhalation "140005610" in Supplementary material 1c was used, and if other procedures in Supplementary material 1c were performed, then the severity was Severe. Severe cases were defined as those with a Moderate II+Severe state, while non-Severe cases were those with a Mild+Moderate I state, according to the claim data.

COVID-GRAM, a prediction scoring tool, was proposed to predict critical illness development among hospitalized patients with COVID-19 in China, with 1,590 patients involved in the development of this model. The 10-item prediction rule (chest radiography abnormality, age, hemoptysis, dyspnea, unconsciousness, number of comorbidities, cancer history, neutrophil-to-lymphocyte ratio, LDH, and direct bilirubin) showed a good predictive value (auROC, 0.88), but it had several limitations that might affect its applicability and generalizability. For example, the study had a modest sample size for constructing the risk score and collecting data for score development, and validation was conducted in China alone (15).

In the present study, critical illness was defined as ICU entry or having respiratory aid records, and according to the machine learning results and logistic model analysis, the mortality risk factors for COVID-19 were older age, male sex, Parkinson's disease, and CKD, which was consistent with the findings of previous reports discussed above. An older age was the strongest independent risk factor according to the Shapley value. Imam et al. reported that an older age and the presence of comorbidities were independent mortality predictors in 1,305 patients with COVID-19 in Michigan, USA (16), which was consistent with our study results. Age-dependent immune cell defects leading to a more robust inflammatory response were associated with increased mortality in older patients (17). Our data identified the administration of anti-Parkinson disease drugs in addition to an older age and male sex as risk factors for mortality according to the LASSO model, which has unique characteristics. In the meta-analysis reported by Putri et al., Parkinson's disease was associated with poor in-hospital outcomes (odds ratio, 2.64; 95% CI, 1.75-3.99, I^2=81%) based on 12 studies with 103,874 patients with COVID-19 (18). Our data and the above meta-analysis have consistent results, suggesting that close monitoring should be provided to patients with Parkinson's disease in order to minimize the mortality risk, especially in patients with advanced age.

In the US, Hajifathalian et al. developed a risk model consisting of the patient age, hypoxia severity, mean arterial pressure, and presence of kidney dysfunction at hospital presentation (19). Other models employed a baseline model approach, with hospital admissions for respiratory disease used as a proxy for COVID-19 pneumonia. These models reported auROC values of 0.73-0.81, compared with 0.820 in our study. The models were all developed by DeCaprio et al. in a cohort of approximately 1.8 million Medicare members (20). Gao et al. used several machine learning methods, including logistic regression, support vector machine, a gradient-boosted decision tree, and a neural network; based on the early warning system to predict accurate mortality risk, clinical data from electronic health records were utilized for patient stratification by mortality risk on admission among 2,520 consecutive patients with COVID-19 in China. Eight features were positively associated with mortality (high risk: consciousness, sex, sputum, blood urea nitrogen, respiratory rate, D-dimer value, number of comorbidities, and age), and six features were negatively associated with mortality (low risk: platelet count, fever, albumin, SpO_2, lymphocyte, and CKD) (21). Although data on patient's symptoms were unavailable in our model, an older age, male sex, respiratory diseases, and kidney diseases were par-

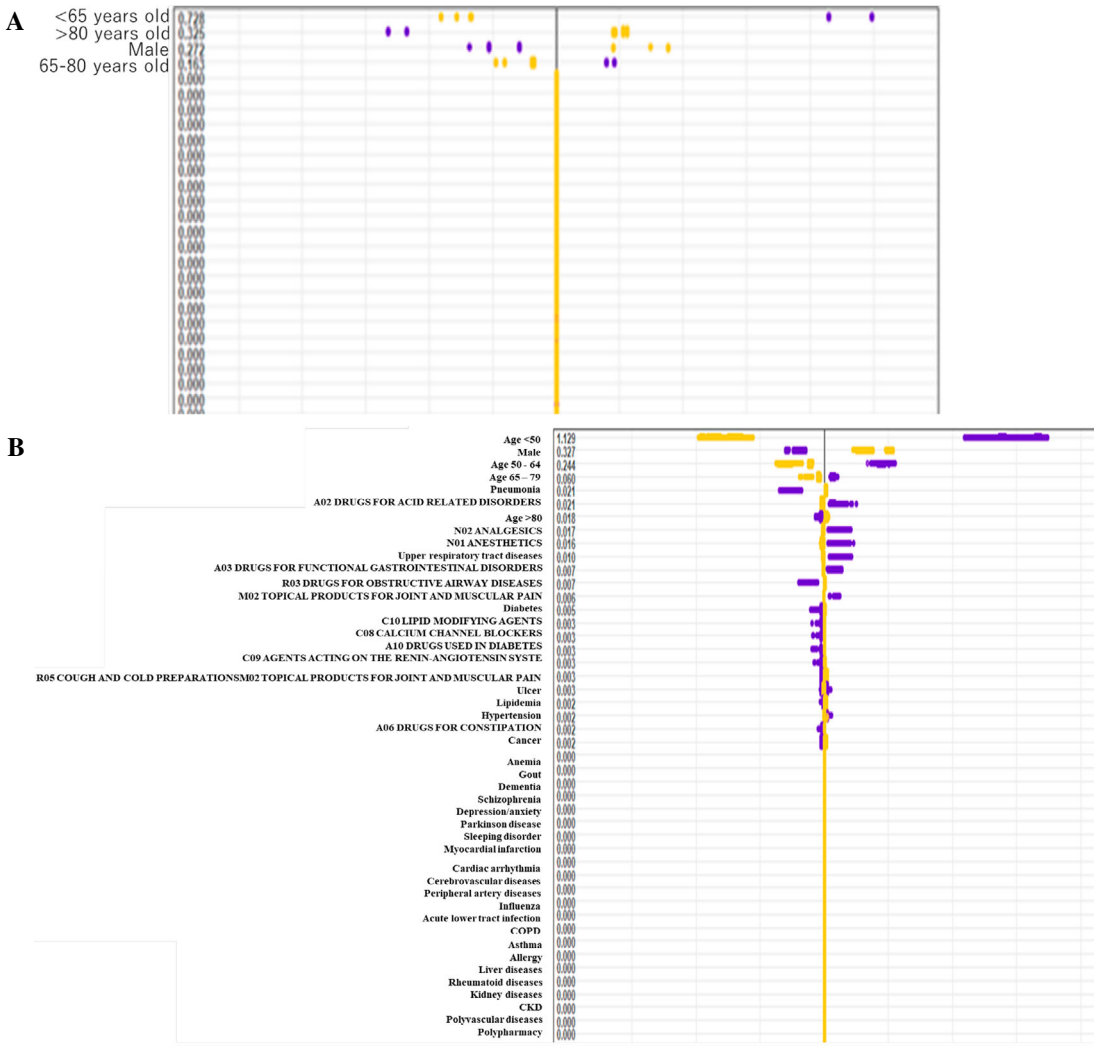

Figure 4. Important features. SHapley Additive exPlanations (SHAP) beeswarm plot. (A) Mortality (>50 years old). (B) Severity (all ages). We considered the severity as Moderate II if only oxygen inhalation "140005610" in Supplementary material 1c was used, and if other procedures in Supplementary material 1c were performed, then the severity was Severe. Severe cases were defined as those with a Moderate II+Severe state, while non-Severe cases were those with a Mild+Moderate I state, according to the claim data.

tially identified as risk factors for a severe disease state on admission. In addition, Yan et al. used machine learning tools to identify three biomarkers (LDH, lymphocyte, and high-sensitivity CRP) that predict the mortality of individual patients more than 10 days in advance and with more than 90% accuracy (22).

Light Gradient Boosting Machine (LightGBM) is a high-performance machine learning algorithm that has great interpretability potential because of its recursive tree-based decision system. We used LightGBM as our statistical method in the present study, owing to its high accuracy, fast training speed, large-scale data handling capability, and GPU learning-supported features (23). However, the internal model mechanisms of black box modeling strategies were difficult to interpret. To identify the principal features driving model prediction, we calculated SHAP values, which are suited for complex models, such as artificial neural networks and gradient-boosting machines (24).

Our study further demonstrated that the disease history and comorbidity status were independent risk factors of COVID-19 mortality during hospitalization in the 2020 infection period (pre-vaccination period). Our results are consistent with previous reports that suggested that pre-existing comorbidities were related to an increased risk of developing a severe state and an increased mortality rate of COVID-19 (5). The relevant comorbidities include hypertension, diabetes mellitus, cardiovascular disease, cerebrovascular disease, CKD, and COPD (25-27). Boehmer et al. reported that after adjusting for both the patient and hospital characteristics, patients with COVID-19 during March 2020-January 2021 had, on average, 15.7 times the risk for myocarditis compared with those without COVID-19 (95% CI=14.1-17.2) (28). Recent studies have elucidated the risk of acute cardiac injury occurrence in patients with COVID-19 (29).

In addition, cardiac injuries caused by pharmacological treatments may be related, and antiviral drugs can cause cardiac insufficiency, arrhythmias, and other cardiovascular disorders (30).

According to a systematic review by Iloanusi et al., polypharmacy and selected drug classes are associated with an increased risk of adverse clinical outcomes among patients with COVID-19 (31). Our analysis showed that polypharmacy was not associated with mortality risk. In our study, ulcer medication was an independent mortality risk factor, probably because patients with chronic gastrointestinal disease may be at high risk for severe COVID-19. The gastrointestinal tract may be susceptible to SARS-CoV-2 infection because of widely expressed ACE2 receptors in the intestine (ACE2 is a receptor for SARS-CoV-2 virus), and digestive symptoms associated with SARS-CoV-2 infection may be caused by direct viral attack resulting from the immune response (32).

Of note, ACE inhibitors, angiotensin receptor blockers (ARBs), and mineralocorticoid receptor antagonists were not risk factors of inpatient mortality risk, which was consistent with the findings of Baral et al.'s review of 52 studies involving 101,949 total patients that concluded that ACE inhibitor and ARB administration was not associated with a high risk of multivariable-adjusted mortality or severe adverse events among patients with COVID-19 who had either hypertension or multiple comorbidities (33). Our study population was limited to DPC patients, so further research is needed to conclude the effect of ACE inhibitors or ARBs on patients with COVID-19 in the Japanese population.

Several limitations associated with the present study warrant mention. A major limitation of these models is their reliance on administrative data, which do not contain sufficient clinical details to draw firm conclusions, such as patient symptoms and radiologic data. In particular, symptoms such as a lack of smell and taste have been identified as being highly predictive of COVID-19 infection by previous studies. Second, medical data analytics deal with collected data, which inherently include many variables with missing values, such as clinical test data and biomarkers, at each time point. Surmounting these challenges will be important for the wider application of big data in medical studies (34).

Nevertheless, our big-data analytics approach using machine learning is promising, as our administrative data have broader generalizability, considerably more patient records, and less attrition than clinical trial data. Our prediction model is most useful in a population health context where the only data available are administrative data in Japan. However, critical ill patients can hardly be identified using administrative records alone, and biomarkers generally change over time; such changes might reflect the patient symptoms and prognosis.

Conclusion

The machine learning method resulted in older age and male sex being identified as the strongest factors associated with mortality risk, and an older age, male sex, pneumonia, drugs for acid-related disorders, analgesics, anesthetics, upper respiratory tract diseases, drugs for functional gastrointestinal disorders, drugs for obstructive airway diseases, topical products for joint and muscular pain, diabetes, lipid-modifying agents, calcium channel blockers, drugs for diabetes, and agents acting on the renin-angiotensin system were identified as risk factors for a severe disease state in patients infected with COVID-19 in the first and the second epidemic waves in Japan.

Our model based on prehospital comorbid conditions and the prescription history might be a useful preliminary screening tool for assessing mortality risk in inpatients with COVID-19 in Japan, where the COVID-19 prevalence is relatively low.

The authors state that they have no Conflict of Interest (COI).

Acknowledgement

The author would like to thank the journal reviewers for their valuable comments concerning our study.

References

1. Wikipedia. COVID-19 pandemic data [Internet]. Available from: https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data
2. Covid CDC-19. 2021 [Internet]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.HTML
3. Weiss P, Murdoch DR. Clinical course and mortality risk of severe COVID-19. Lancet **395**: 1014-1015, 2020.
4. Myers LC, Parodi SM, Escobar GJ, Liu VX. Characteristics of hospitalized adults with COVID-19 in an integrated health care system in California. JAMA **323**: 2195-2198, 2020.
5. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA **323**: 2052-2059, 2020.
6. CDC. Assessing risk factors for severe COVID-19 illness [Internet]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/assessing-risk-factors.html
7. Yu C, Lei Q, Li W, et al. Clinical characteristics, associated factors, and predicting COVID-19 mortality risk: a retrospective study in Wuhan, China. Am J Prev Med **59**: 168-175, 2020.
8. Wu G, Yang P, Xie Y, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. Eur Respir J **56**: 2001104, 2020.
9. Iwasaki A, Grubaugh ND. Why does Japan have so few cases of COVID-19? EMBO Mol Med **12**: e12481, 2020.
10. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B **58**: 267-288, 1996.
11. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Springer, New York, NY, 2009.
12. Minister of Health. LaW. [First case related to COVID-19 infection] [Internet]. Available from: https://www.mhlw.go.jp/stf/newpage_08906.html
13. Tabata S, Imai K, Kawano S, et al. Clinical characteristics of COVID-19 in 104 people with SARS-CoV-2 infection on the Diamond Princess cruise ship: a retrospective analysis. Lancet Infect Dis **20**: 1043-1050, 2020.

14. Roth GA, Emmons-Bell S, Alger HM, et al. Trends in patient characteristics and COVID-19 in-hospital mortality in the united states during the COVID-19 pandemic. JAMA Netw Open **4**: e218828, 2021.
15. Liang W, Liang H, Ou L, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med **180**: 1081-1089, 2020.
16. Imam Z, Odish F, Gill I, et al. Older age and comorbidity are independent mortality predictors in a large cohort of 1305 COVID-19 patients in Michigan, United States. J Intern Med **288**: 469-476, 2020.
17. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet **395**: 1054-1062, 2020.
18. Putri C, Hariyanto TI, Hananto JE, Christian K, Situmeang RFV, Kurniawan A. Parkinson's disease may worsen outcomes from coronavirus disease 2019 (COVID-19) pneumonia in hospitalized patients: a systematic review, meta-analysis, and meta-regression. Parkinsonism Relat Disord **87**: 155-161, 2021.
19. Hajifathalian K, Sharaiha RZ, Kumar S, et al. Development and external validation of a prediction risk model for short-term mortality among hospitalized US COVID-19 patients: a proposal for the COVID-AID risk tool. PLOS ONE **15**: e0239536, 2020.
20. DeCaprio D, Gartner J, McCall CJ, et al. Building a COVID-19 vulnerability index. J Med Artif Intell **3**: 15, 2020.
21. Gao Y, Cai GY, Fang W, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun **11**: 5033, 2020.
22. Yan L, Zhang HT, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell **2**: 283-288, 2020.
23. Daoud EA. Comparison between XGBoost LightgBM and CatBoost using a home credit dataset. Int J Comput Inf Eng **13**: 6-10, 2019.
24. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng **2**: 749-760, 2018.
25. Huang I, Lim MA, Pranata R. Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia - a systematic review, meta-analysis, and meta-regression. Diabetes Metab Syndr **14**: 395-403, 2020.
26. Pranata R, Huang I, Lim MA, Wahjoepramono EJ, July J. Impact of cerebrovascular and cardiovascular diseases on mortality and severity of COVID-19 - systematic review, meta-analysis, and meta-regression. J Stroke Cerebrovasc Dis **29**: 104949, 2020.
27. Henry BM, Lippi G. Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection. Int Urol Nephrol **52**: 1193-1194, 2020.
28. Boehmer TK. Association between COVID-19 and myocarditis using hospital-based administrative data - United States, March 2020-January 2021. MMWR. Morbidity and mortality weekly report **70**: 1228-1232, 2021.
29. Aghagoli G, Gallo Marin B, Soliman LB, Sellke FW. Cardiac involvement in COVID-19 patients: risk factors, predictors, and complications: a review. J Card Surg **35**: 1302-1305, 2020.
30. Sakabe M, Yoshioka R, Fujiki A. Sick sinus syndrome induced by interferon and ribavirin therapy in a patient with chronic hepatitis C. J Cardiol Cases **8**: 173-175, 2013.
31. Iloanusi S, Mgbere O, Essien EJ. Polypharmacy among COVID-19 patients: a systematic review. J Am Pharm Assoc (2003) **61**: e14-e25, 2021.
32. Beyerstedt S, Casaro EB, Rangel ÉB. COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. Eur J Clin Microbiol Infect Dis **40**: 905-919, 2021.
33. Baral R, Tsampasian V, Debski M, et al. Association between renin-angiotensin-aldosterone system inhibitors and clinical outcomes in patients with COVID-19: a systematic review and meta-analysis. JAMA Netw Open **4**: e213594, 2021.
34. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract **36**: 3-11, 2017.