



OPEN

## Finding of the factors affecting the severity of COVID-19 based on mathematical models

Jiahao Qu<sup>1</sup>, Brian Sumali<sup>1</sup>, Ho Lee<sup>2</sup>, Hideki Terai<sup>2</sup>, Makoto Ishii<sup>2</sup>, Koichi Fukunaga<sup>2</sup>, Yasue Mitsukura<sup>1✉</sup> & Toshihiko Nishimura<sup>3</sup>

Since 2019, a large number of people worldwide have been infected with severe acute respiratory syndrome coronavirus 2. Among those infected, a limited number develop severe coronavirus disease 2019 (COVID-19), which generally has an acute onset. The treatment of patients with severe COVID-19 is challenging. To optimize disease prognosis and effectively utilize medical resources, proactive measures must be adopted for patients at risk of developing severe COVID-19. We analyzed the data of COVID-19 patients from seven medical institutions in Tokyo and used mathematical modeling of patient blood test results to quantify and compare the predictive ability of multiple prognostic indicators for the development of severe COVID-19. A machine learning logistic regression model was used to analyze the blood test results of 300 patients. Due to the limited data set, the size of the training group was constantly adjusted to ensure that the results of machine learning were effective (e.g., recognition rate of disease severity > 80%). Lymphocyte count, hemoglobin, and ferritin levels were the best prognostic indicators of severe COVID-19. The mathematical model developed in this study enables prediction and classification of COVID-19 severity.

In December 2019, a large number of viral pneumonia cases were noted in Wuhan, China. This pneumonia was found to be caused by a new type of coronavirus—severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)<sup>1,2</sup>, and the disease was named coronavirus disease 2019 (COVID-19)<sup>3</sup>. Subsequently, the disease was declared a pandemic and as of April 7, 2021, there were more than 132 million cases of COVID-19, with the death toll reaching 2.8 million worldwide<sup>4</sup>.

The symptoms of COVID-19 vary across patients<sup>5</sup>, and typical symptoms include fever, cough, and fatigue, while some individuals may also lose their sense of taste and smell. The severity of symptoms varies across individuals, and COVID-19 affects both the upper and lower respiratory tracts. The majority of individuals who are infected with SARS-CoV-2 develop mild-to-moderate cold-like symptoms or may even be asymptomatic. However, the probability of developing pneumonia is very high for elderly individuals.

Treatment of patients with severe COVID-19 is challenging. To date, to the best of our knowledge, no specific drug or treatment has been approved for COVID-19. Hydroxychloroquine was initially used for treatment during the early phase of the pandemic, but clinical trials subsequently showed that it was not effective<sup>6</sup>. Trials have also shown that dexamethasone is effective at reducing mortality in patients with severe COVID-19, but it had no effect on patients who did not require oxygen supplementation<sup>7</sup>. In a study by John et al.<sup>8</sup>, the recovery time of hospitalized patients with lower respiratory tract infection who received remdesivir was shorter than that of patients in the placebo control group. Among the 532 patients who received remdesivir treatment, 131 (24.6%) had serious adverse reactions. Remdesivir is currently the only drug that has been approved by the US Food and Drug Administration for treatment of COVID-19<sup>9</sup>. However, in a trial organized by the World Health Organization<sup>10</sup>, remdesivir as well as hydroxychloroquine, lopinavir, and interferon regimens did not affect the prognosis of hospitalized patients. Therefore, at present, there are no drugs or treatment methods that have proven to be completely effective for the treatment of COVID-19.

All patients with severe COVID-19 develop acute respiratory distress syndrome and frequently require mechanical ventilation to ensure sufficient oxygen supply, and some patients with severe COVID-19 may require extracorporeal membrane oxygenation (ECMO)<sup>11,12</sup>. Although ECMO can be life-saving in some cases, many complications are associated with the use of ECMO, including hemorrhage<sup>13</sup>, bacterial infections<sup>14</sup> and stroke<sup>15</sup>. Thus, there is a need to reduce the number of severe COVID-19 cases that require such interventions.

<sup>1</sup>School of Integrated Design Engineering, Keio University, Yokohama, Kanagawa, Japan. <sup>2</sup>Department of Neuropsychiatry, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo, Japan. <sup>3</sup>Department of Anesthesia, Stanford University School of Medicine, Stanford, CA, USA. ✉email: mitsukura@keio.jp

Symptom or sign	Disease severity	
	Non-severe (%) (N = 208)	Severe (%) (N = 104)
Altered level of consciousness	1	11
Fever ( $\geq 37.5$ °C)	65	89
Cough	44	52
Sputum	16	24
Sore throat	19	13
Nasal discharge	14	1
Dysgeusia	22	13
Dysosmia	18	9
Shortness of breath	14	50
Diarrhea	10	16
Nausea and vomiting	2	9
Malaise	27	52
Bacterial infection	2	27
Fungal infection	0	5
Heart failure	0	13
Thromboembolism (including pulmonary embolism and cerebral infarction)	0	10
Liver dysfunction	4	39
Renal dysfunction	0	17
Macrophage activation syndrome (including hemophagocytic syndrome)	0	2

**Table 1.** Clinical symptoms and signs according to disease severity (N = 312).

To limit the development of severe COVID-19, effective preventative measures should be identified and implemented in care settings. For this, it is necessary to understand the factors that contribute to the development of severe COVID-19. A previous study attempted to develop logistic regression models for this purpose<sup>16</sup>; however, the results were not universally applicable due to the limited size of the data set used in the analysis.

Therefore, the present study aimed to expand upon previous work to develop more effective models for predicting COVID-19 disease severity by analyzing the blood test results of 300 COVID-19 patients according to disease severity using mathematical models and evaluating the prognostic factors in the model. Models designated seven factors that are known to affect the severity of COVID-19 as model variables and simultaneously compared the influence of each factor to determine its impact on patient outcome. This allowed for identification of the three most influential factors contributing to COVID-19 severity, which were then used for further analysis.

## Results

In the model generated when 30% of the available patient data (86 patients) were used for training the machine learning model, the ratio of the training group to the test group was 7:3. Mathematical formulae of the lasso, ridge, and logistic regression models are shown in Eqs. (1), (2), and (3), respectively. The results of the analysis using this training model are shown in Table 1.

### Lasso regression model expression.

$$y = 0.15 * Cre - 0.02 * lymphocyte\ fraction + 0.08 * LDH + 0.06 * CRP + 0.10 \quad (1)$$

### Ridge regression model expression.

$$y = 0.15 * Cre - 0.03 * Hb - 0.007 * lymphocyte\ fraction + 0.07 * CRP + 0.08 * LDH - 0.01 * (-0.10) * ferritin + 0.008 * D - dimer + 0.1 \quad (2)$$

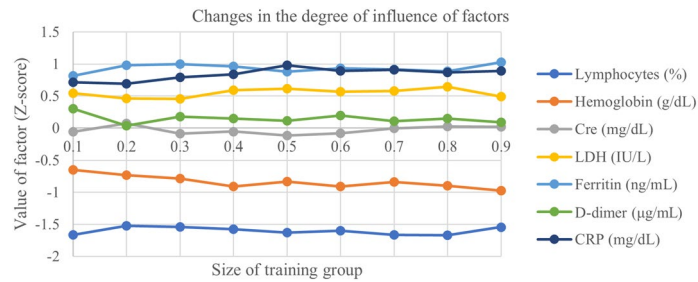
### Logistic regression model expression.

$$y = 5.05 * Cre - 0.28 * Hb - 0.90 * lymphocyte\ fraction + 0.99 * CRP + 2.21 * LDH - 1.43 * ferritin - 0.34 * D - dimer - 4.62 \quad (3)$$

The regression equation using 100% of the learning data of all patients is shown in Eq. (4).

$$y = 0.29 * Cre - 0.03 * Hb - 0.07 * lymphocyte\ fraction + 0.08 * CRP + 0.80 * LDH - 0.68 * ferritin - 0.30 * D - dimer - 0.06 \quad (4)$$

where Cre, CRP, Hb, and LDH indicate creatinine, C-reactive protein, hemoglobin, lactate dehydrogenase, respectively.



**Figure 1.** Association between the degree of influence of each factor according to the size of the training group. Cre, Creatine; CRP, C-reactive protein; LDH, lactate dehydrogenase.

## Discussion

The parameters in Eqs. (1), (2), and (3) are coefficients of each feature in each mathematical machine learning model. Using Eqs. (1), (2), and (3), we compared the coefficients of each impact factor, and the results of the comparisons are shown in Table 1.

Although the degree of influence of the prognostic factors varied in the different models, creatinine (Cre), lactate dehydrogenase (LDH), and C-reactive protein (CRP) were influential factors in all the three models. This result is consistent with results of previous research.<sup>16</sup> We attempted to combine the results of previous studies to determine the reasons why these three prognostic indicators were so influential.

**Abnormal increase in Cre levels.** Cre is a marker of kidney function. The results of the mathematical model showed that deterioration of renal function is related to COVID-19 severity. According to a study by Ishii et al.<sup>17</sup>, hyperuricemia and renal dysfunction are caused due to increased oxidative stress. Inflammation and oxidative stress are important indicators of COVID-19. Chronic kidney disease, which is associated with hyperuricemia, is likely to be a factor influencing COVID-19 severity, although the mechanism is unknown.

**Abnormal increase in LDH levels.** In one study, many COVID-19 patients had abnormally elevated liver function marker levels<sup>18</sup>. LDH is a marker of liver function, and a study by Shailendra et al.<sup>19</sup> found that the risk of death in patients with severe COVID-19 with liver dysfunction was threefold higher than that in patients without liver dysfunction and that the risk of death was increased 4.6-fold in individuals with cirrhosis. The angiotensin converting enzyme 2 receptor, by which SARS-CoV-2 gains entry to cells, is abundant in the liver and bile ducts where SARS-COV-2 infection can cause local damage. In the models to predict disease severity, we did not use other biochemical markers of liver function, such as aspartate aminotransferase, alanine aminotransferase, and gamma-glutamyl transpeptidase. A study by Kishaba et al.<sup>20</sup> found that LDH was a marker of the severity of idiopathic pulmonary fibrosis. Further, Yan et al.<sup>21</sup>, found that an isolated increase in LDH levels without an increase in the other liver enzyme levels is indicative of inflammation (including COVID-19 pneumonia) and that this level is elevated as a result of tissue breakdown.

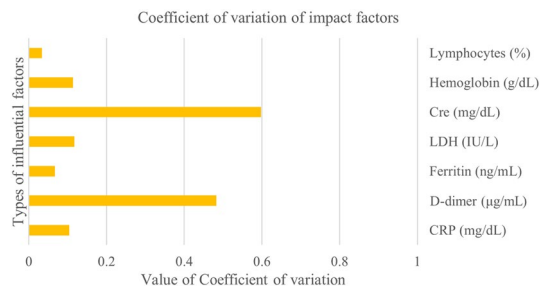
**Abnormal increase in CRP.** Cytokine storms (a form of hyperimmune response) cause death in patients with severe COVID-19<sup>22</sup>. Elevated CRP level is a marker of an acute inflammatory response, and elevated lymphocyte levels are an indicator of infection with viruses and other pathogens and immune function<sup>23</sup>. The immune system over-responds in severe cases of COVID-19, causing systemic acute inflammation that is associated with more severe diseases.

As shown in Table 1, the predictive ability of the logistic regression model considerably exceeded that of the lasso and ridge models. Therefore, the logistic regression model should be used as the basis for comparing the degree of influence of each factor on the model. Data from only 300 patients was included in our study, and this ensured that the model recognition rate was high (recognition rate > 80%) by changing the size of the training group of the machine learning by the logistic regression model. Increasing the number of simulations allowed for comparison of different training groups in terms of the degree of influence of the following factors (the coefficient of each factor in the mathematical model). Our method for this comparison was as follows:

1. The size of the training group was set at 10% intervals to 10%, 20%, 30%, ..... 90%, yielding a total of nine groups.
2. Twenty simulations were performed on the training data of each group.
3. After normalizing the coefficients (Z-score) of each factor obtained from each simulation, the value of the average of the 20 sets of results was calculated.

Figure 1 shows the results of these calculations. At the same time, we also calculated the coefficient of variation of the degree of influence of each factor, considering the absolute value, and the results are shown in Fig. 2.

As shown in Fig. 1, even if the amount of data in the training group changed significantly, there was very little change in the degree of influence of the factors. Based on the blood test results of 300 patients in our study, lymphocytes were the most influential factor, followed by ferritin, Hb, and CRP.



**Figure 2.** The coefficient of variation of each predictor in different training groups.

As shown in Fig. 2, the degree of change in lymphocytes was the smallest, followed by ferritin, whereas the degree of change in hemoglobin had a similar influence as that of LDH.

**Abnormal decrease in lymphocyte levels.** Although the amount of patient data in this study was small, lymphocyte counts consistently had the greatest predictive value for disease severity. Lymphopenia is common in patients with COVID-19<sup>24</sup>. A study by Fajgenbaum and June<sup>23</sup> showed that lymphopenia was not often present in diseases complicated by cytokine storms, but it is a marker of COVID-19 severity. However, it is not clear whether the lymphopenia observed in COVID-19 patients is due to tissue infiltration or lymphocyte destruction.

**Abnormal increase in ferritin levels.** In another study<sup>25</sup>, an abnormal increase in ferritin levels was found in patients who required mechanical ventilation. This is consistent with the definition of disease severity in our study (using ventilator as one of the definitions), wherein an abnormal increase in ferritin level was identified as another marker of severe COVID-19.

**Abnormal decrease in hemoglobin levels.** The oxygen saturation of hemoglobin determines the degree of hypoxia<sup>24</sup>. Most patients with severe disease needed supplemental oxygen to maintain an oxygen saturation above 90%. Therefore, a decrease in hemoglobin level is closely related to hypoxia. When there is a large reduction in Hb levels, patients will enter the hypoxemia state, which will affect the prognosis.

A machine learning logistic regression model was used to analyze the influence of blood biomarkers on COVID-19 severity using patient blood data. Cre, LDH, and CRP had the greatest influence on the model when the recognition rate of the model was maximum, whereas lymphocyte counts, ferritin, and hemoglobin consistently had a high influence on the model outcome when the number of simulations increased, while maintaining a high recognition rate.

In conclusion, the findings of these statistical models indicates that lymphocyte counts as well as ferritin and hemoglobin levels are important indicators of COVID-19 severity.

## Materials and methods

**Ethics and funding.** This study was approved by the Ethics Committee of Keio University and performed in accordance with the Declaration of Helsinki. The study protocol was approved by the Ethics Review Board of Keio University (approval ID 20,200,055). The study was funded by Keio Research Global Institute (grant number: UMIN00041186).

**Patients.** Patients with COVID-19 who were admitted to either Keio University Hospital or one of six community hospitals located in the greater Tokyo area (Saitama City Hospital, Tokyo Saiseikai Central Hospital, Federation of National Public Service Personnel Mutual Aid Associations, Tachikawa Hospital, National Hospital Organization Tokyo Medical Center, Nihon Koukan Hospital, and Saiseikai Utsunomiya Hospital) were included in the present study. In total, 312 patients meeting these criteria were identified, of whom, 12 patients were excluded due to incomplete data. Thus, patient data of 300 patients with COVID-19 were used in the present analysis.

Informed consent was obtained from all study subjects. Consent was obtained directly from adults aged  $\geq 20$  years. For minors aged  $< 20$  years, consent was obtained both from the patient and from the patient's parent or legal guardian.

Patients were divided into three major groups according to age: group 1 (0–30 years old), group 2 (31–60 years old), and group 3 (60 years old and above). Group 1 comprised 29 male patients (1 with severe COVID-19) and 44 female patients (1 with severe COVID-19), group 2 comprised 86 male patients (31 with severe COVID-19) and 45 female patients (7 with severe COVID-19), and group 3 comprised 63 male patients (40 with severe COVID-19) and 45 female patients (24 with severe COVID-19).

Patients who satisfied any of the following three criteria were defined as having severe COVID-19: 1) admitted to an intensive care unit, 2) required mechanical ventilation, or 3) died of COVID-19-related disease.

Regression type	Training group recognition rate	Test group recognition rate	1st	2nd	3rd	4th
Lasso	0.63	0.31	Cre	LDH	CRP	Hb
Ridge	0.64	0.3	Cre	Ferritin	CRP	LDH
Logistic	0.98	0.88	Cre	LDH	Ferritin	CRP

**Table 2.** Accuracy of the three evaluation models and comparison of the influence of each factor in each regression model from greatest to least. Cre, creatine; CRP, C-reactive protein; Hb, hemoglobin; LDH, lactate dehydrogenase.

**Symptoms and signs.** Direct symptoms of COVID-19 as well as types of complications associated with COVID-19 were summarized according to previous findings<sup>26–30</sup>. Accordingly, the following 19 items were assessed: 1. altered level of consciousness; 2. fever ( $\geq 37.5^{\circ}\text{C}$ ); 3. cough; 4. sputum production; 5. sore throat; 6. nasal discharge; 7. dysgeusia; 8. dyssomnia; 9. shortness of breath; 10. diarrhea; 11. nausea and vomiting; 12. malaise; 13. bacterial infection; 14. fungal infection; 15. heart failure; 16. thromboembolism (including pulmonary embolism and cerebral infarction); 17. liver dysfunction; 18. renal dysfunction; and 19. macrophage activation syndrome (including hemophagocytic syndrome). Patient records of the 300 study participants were reviewed for presence of the preceding 19 items, and association between these factors and COVID-19 severity is shown in Table 2.

From the results presented in Table 2, it is apparent that to reduce the incidence of each symptom in the non-severe group and the severe group separately, the difference may not exceed 50%. This is likely because the patient data were collected immediately after patient hospitalization; therefore, symptoms were likely to worsen as the disease developed further. In addition, patients with severe COVID-19 had a higher prevalence of most of the 19 factors analyzed in this study compared with patients with non-severe COVID-19. However, the number of patients with severe COVID-19 in this study was very small to verify whether these factors were generally predictive; therefore, early symptoms were not good predictors of disease severity in this study.

**Blood test results.** We analyzed patient blood test results and selected predictors from the 35 biomarkers assessed in these tests. These selected factors were used to predict disease severity. First, we summarized the clinical signs of 30% of the patients (86 patients) who had similar clinical manifestations as the entire patient population. These clinical signs appear to be manifestations of concurrent changes in blood biochemical indexes, as reflected in blood test results. Additionally, we reviewed previous studies of patients with severe COVID-19 and summarized their results. Previous studies have shown that diabetes mellitus and hypertension, among other factors, are associated with greater disease severity<sup>31</sup>. In a study by Liu et al.<sup>22</sup>, the immune function of patients was closely related to disease severity. Further, Mohammad et al.<sup>32</sup> reported that there was an interaction between liver function and the development of severe COVID-19.

Although previous studies have noted that age is an important indicator<sup>33–35</sup>, symptoms and age alone cannot help in appropriate classification of severe and non-severe COVID-19 patients during the early stages of COVID-19. Therefore, we decided to use patient blood test results to analyze the causes of severe COVID-19. In this analysis of 312 patients, the age of patients with severe disease patients was higher than that of patients with mild or moderate disease; however, we did not include age as a predictor of disease severity. As individuals get older, they are more likely to develop other underlying diseases, and their immune function declines. Thus, age has a strong influence on the development of severe disease, but its influence is indirect. Further, a study by Faust et al.<sup>36</sup> showed that older individuals as well as young individuals aged 25–44 years have increased mortality rates. This supports our speculation that the influence of age on the development of severe COVID-19 is limited as well as our decision to solely analyze patient blood test results to identify predictors of disease severity.

A total of 35 blood biochemical indicators were assessed in study participants. Based on expert medical opinion by physician and combined with the results of previous COVID-19 studies, we designated seven blood indicators as variables in the mathematical model, namely (1) lymphocyte count (%)<sup>37</sup>; (2) hemoglobin (g/dL)<sup>38</sup>; (3) Cre (mg/dL)<sup>39</sup>; (4) LDH (IU/L)<sup>40</sup>; (5) ferritin (ng/mL)<sup>41</sup>; (6) D-dimer ( $\mu\text{g/mL}$ )<sup>42</sup>; and (7) CRP (mg/dL)<sup>43</sup>.

**Statistical methods.** First, we assessed the results of three different mathematical models, namely a lasso regression model, ridge regression model, and logistic regression model. After comparing the results of these models, we decided to use the logistic regression model for the classification of COVID-19 severity using 30% of the available patient data for machine learning purposes. Additionally, we used lasso<sup>44</sup> and ridge regression<sup>45</sup> models to compare with the machine learning results obtained using the logistic regression model.

Finally, the resulting logistic regression model of classification was used to compare the influence of each factor on COVID-19 severity. The sample size of this study is small; therefore, when comparing the influence degree of each factor, it is possible that the results obtained will vary greatly. To reduce the impact of sample size on the results, we decided to change the size of the logistic regression model training group and increase the number of simulations to ensure accuracy of the model. To ensure that the classification accuracy of disease severity was  $> 80\%$ , the data size of the training group for each iteration of this process was changed (e.g. training group: 10%, 20%, ..., 90% of the total data volume), and the results of each simulation were then averaged.

## Data availability

Data and code are available from the corresponding author.

Received: 31 May 2021; Accepted: 7 December 2021

Published online: 20 December 2021

## References

- Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N Engl. J. Med.* **382**(8), 727–733 (2020).
- Xu, Z. *et al.* Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* **8**(4), 420–422 (2020).
- Naming the coronavirus disease (COVID-19) and the virus that causes it, World Health Organization. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it) (April, 7, 2020).
- COVID-19 MAP, Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html> (April, 7, 2020).
- Gandhi, R. T. *et al.* Mild or moderate Covid-19. *N Engl. J. Med.* **383**(18), 1757–1766 (2020).
- RECOVERY Collaborative Group. Effect of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl. J. Med.* **383**(21), 2030–2040 (2020).
- RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19 - Preliminary Report. *N Engl J Med* (2020)
- Beigel, J. H. *et al.* Remdesivir for the treatment of Covid-19—Final report. *N Engl. J. Med.* **383**(19), 1813–1826 (2020).
- Therapeutic management of patients with COVID-19, coronavirus Disease 2019 (COVID-19) treatment guidelines, National Institutes of Health. <https://www.covid19treatmentguidelines.nih.gov/therapeutic-management/> (April, 7, 2020).
- WHO Solidarity Trial Consortium. Repurposed antiviral drugs for Covid-19 - Interim WHO solidarity trial results. *N Engl. J. Med.* **384**(6), 497–511 (2021).
- de Souza DB, *et al.*, Possible solutions for oxygenation support in critically ill patients with COVID-19. *Res. Biomed. Eng.* (2021)
- MacLaren, G. *et al.* What's new in ECMO for COVID-19?. *Intens. Care Med.* **47**(1), 107–109 (2020).
- Fitzgerald, A. L. *et al.* The prolonged use of VV ECMO support in Covid-19: A case report. *Crit. Care Med. (Targu Mures)* **6**(4), 224–230 (2020).
- Nawa, K. *et al.* ECMO-its basics, clinical practice and prospects. *Artif. Organs* **12**, 982–990 (1983).
- Biancari, F. *et al.* Six-month survival after extracorporeal membrane oxygenation for severe COVID-19. *J. Cardiothorac. Vasc. Anesth.* **S1053–0770**(21), 00062–00068 (2021).
- Ayanian, S. *et al.* The association between biomarkers and clinical outcomes in novel coronavirus pneumonia in a US cohort. *Biomark. Med.* **14**(12), 1091–1097 (2020).
- Ishii, M. *et al.* Clinical characteristics of 345 patients with coronavirus disease 2019 in Japan: A multicenter retrospective study. *J. Infect.* **81**(5), e3–e5 (2020).
- Goyal, P. *et al.* Clinical characteristics of COVID-19 in New York City. *N Engl. J. Med.* **382**(24), 2372–2374 (2020).
- Singh, S. & Khan, A. Clinical characteristics and outcomes of coronavirus Disease 2019 among patients with preexisting liver disease in the United States: A multicenter research network study. *Gastroenterology* **159**(2), 768–771.e3 (2020).
- Kishaba, T. *et al.* Staging of acute exacerbation in patients with idiopathic pulmonary fibrosis. *Lung* **192**(1), 141–149 (2014).
- Yan, L. *et al.* An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* **2**(5), 283–288 (2020).
- Liu, F. *et al.* Prognostic value of interleukin-6, C-reactive protein, and procalcitonin in patients with COVID-19. *J. Clin. Virol.* **127**, 104370 (2020).
- Fajgenbaum, D. C. & June, C. H. Cytokine storm. *N Engl. J. Med.* **383**(23), 2255–2273 (2020).
- Berlin, D. A. *et al.* Severe Covid-19. *N Engl. J. Med.* **383**(25), 2451–2460 (2020).
- Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**(10229), 1054–1062 (2020).
- Mizrahi, B. *et al.* Longitudinal symptom dynamics of COVID-19 infection. *Nat. Commun.* **11**(1), 6208 (2020).
- Alimohamadi, Y. *et al.* Determine the most common clinical symptoms in COVID-19 patients: A systematic review and meta-analysis. *J. Prev. Med. Hyg.* **61**(3), E304–E312 (2020).
- Grant, M. C. *et al.* The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE* **15**(6), 234765 (2020).
- Liu, P. *et al.* Epidemiological and clinical features in patients with coronavirus disease 2019 outside of Wuhan, China: Special focus in asymptomatic patients. *PLoS Negl. Trop. Dis.* **15**(3), 0009248 (2021).
- Wise, J. Covid-19: Study reveals six clusters of symptoms that could be used as a clinical prediction tool. *BMJ* **370**, 2911 (2020).
- Yan, L. *et al.* An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* **2**(5), 283–288 (2020).
- Parohan, M. *et al.* Liver injury is associated with severe coronavirus disease 2019 (COVID-19) infection: A systematic review and meta-analysis of retrospective studies. *Hepatol. Res.* **50**(8), 924–935 (2020).
- Perrotta, F. *et al.* COVID-19 and the elderly: Insights into pathogenesis and clinical decision-making [published correction appears in *Aging Clin Exp Res*; **32**(9):1909]. *Aging Clin Exp Res.* **32**(8), 1599–1608 (2020).
- Nanda, A. *et al.* COVID-19 in older adults. *Aging Clin. Exp. Res.* **32**(7), 1199–1202 (2020).
- Guo, T. *et al.* Clinical characteristics of elderly patients with COVID-19 in Hunan Province, China: A multicenter, retrospective study. *Gerontology* **66**(5), 467–475 (2020).
- Faust, J. S. *et al.* All-cause excess mortality and COVID-19-Related mortality among US adults aged 25–44 years. *JAMA* **23**(8), 785–787 (2020).
- Lagunas-Rangel, F. A. Neutrophil-to-lymphocyte ratio and lymphocyte-to-C-reactive protein ratio in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *J. Med. Virol.* **92**(10), 1733–1734 (2020).
- Cavezzi, A. *et al.* COVID-19: Hemoglobin, iron, and hypoxia beyond inflammation. A narrative review. *Clin. Pract.* **10**(2), 1271 (2020).
- Deng, X. *et al.* Blood biochemical characteristics of patients with coronavirus disease 2019 (COVID-19): A systemic review and meta-analysis. *Clin. Chem. Lab. Med.* **58**(8), 1172–1181 (2020).
- Zheng, Z. *et al.* Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J. Infect.* **81**(2), e16–e25 (2020).
- Cheng, L. *et al.* Ferritin in the coronavirus disease 2019 (COVID-19): A systematic review and meta-analysis. *J. Clin. Lab. Anal.* **34**(10), 23618 (2020).
- Zhang, L. *et al.* D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19. *J. Thromb. Haemost.* **18**(6), 1324–1329 (2020).
- Velavan, T. P. & Meyer, C. G. Mild versus severe COVID-19: Laboratory markers. *Int. J. Infect. Dis.* **95**, 304–307 (2020).
- Chen, Y. *et al.* The utility of lasso-based models for real time forecasts of endemic infectious diseases: A cross country comparison. *J. Biomed. Inf.* **81**, 16–30 (2018).
- Liu, S. *et al.* Discriminating between adaptive and carcinogenic liver hypertrophy in rat studies using logistic ridge regression analysis of toxicogenomic data: The mode of action and predictive models. *Toxicol. Appl. Pharmacol.* **318**, 79–87 (2017).

## Acknowledgements

We would like to thank all the members who participated in this study, including those in Saitama City Hospital, Tokyo Saiseikai Central Hospital, Kawasaki Municipal Ida Hospital, Federation of National Public Service Personnel Mutual Aid Associations Tachikawa Hospital, National Hospital Organization Tokyo Medical Center, Saitama Medical Center, Tokyo Dental College Ichikawa General Hospital, Kitasato Institute Hospital, Keiyu Hospital, Nihon Koukan Hospital, Hino Municipal Hospital, Sano Kousei General Hospital, and Saiseikai Utsunomiya Hospital.

## Author contributions

Conceptualization: Y.M., B.S. Methodology: J.Q., B.S., H.L., H.T., M.I., K.F., Y.M., T.N. Investigation: H.L., H.T., K.F., M.I. Visualization: J.Q., B.S., Y.M., T.N. Funding acquisition: M.I., K.F., Y.M. Project administration: Y.M., T.N. Supervision: Y.M., T.N. Writing – original draft: J.Q., B.S., Y.M., T.N. Writing – review & editing: B.S., Y.M., M.I., T.N.

## Funding

The study was funded by Keio Research Global Institute (grant number: UMIN000041186).

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Y.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021