

Model Prediction for In-Hospital Mortality in Patients with COVID-19: A Case–Control Study in Isfahan, Iran

Ibrahim Abdollahpour,¹ Isabel Aguilar–Palacio,² Juan Gonzalez–Garcia,³ Golnaz Vaseghi,⁴ Zahra Otraj,⁵ Amirreza Manteghinejad,⁵ Azam Mosayebi,⁵ Yahya Salimi,⁶ and Shaghayegh Haghjooy Javanmard^{5*}

¹Child Growth and Development Research Center, Research Institute for Primordial Prevention of Non-Communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran; ²Preventive Medicine and Public Health Department, Research Group of Health Services in Aragon (GRISSA), IIS, Zaragoza University, Aragon, Spain; ³Biocomputing Unit/Data Science In Health Services and Policy Research Group/Health Services Research Network on Chronic Patients (REDISSEC), Instituto Aragonés de Ciencias de La Salud, Zaragoza, Spain; ⁴Cardiovascular Research Centre, Isfahan Cardiovascular Research Institute, Isfahan University of Medical Science, Isfahan, Iran; ⁵Applied Physiology Research Center, Isfahan Cardiovascular Research Institute, Isfahan University of Medical Science, Isfahan, Iran; ⁶Social Development and Health Promotion Research Center, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

Abstract. The COVID-19 pandemic has now imposed an enormous global burden as well as a large mortality in a short time period. Although there is no promising treatment, identification of early predictors of in-hospital mortality would be critically important in reducing its worldwide mortality. We aimed to suggest a prediction model for in-hospital mortality of COVID-19. In this case–control study, we recruited 513 confirmed patients with COVID-19 from February 18 to March 26, 2020 from Isfahan COVID-19 registry. Based on extracted laboratory, clinical, and demographic data, we created an in-hospital mortality predictive model using gradient boosting. We also determined the diagnostic performance of the proposed model including sensitivity, specificity, and area under the curve (AUC) as well as their 95% CIs. Of 513 patients, there were 60 (11.7%) in-hospital deaths during the study period. The diagnostic values of the suggested model based on the gradient boosting method with oversampling techniques using all of the original data were specificity of 98.5% (95% CI: 96.8–99.4), sensitivity of 100% (95% CI: 94–100), negative predictive value of 100% (95% CI: 99.2–100), positive predictive value of 89.6% (95% CI: 79.7–95.7), and an AUC of 98.6%. The suggested model may be useful in making decision to patient’s hospitalization where the probability of mortality may be more obvious based on the final variable. However, moderate gaps in our knowledge of the predictors of in-hospital mortality suggest further studies aiming at predicting models for in-hospital mortality in patients with COVID-19.

INTRODUCTION

The COVID-19 with a dramatic worldwide spreading trend has been caused by the seventh member of the coronavirus family called SARS-CoV-2¹ with the ability for human infection.² With R_0 of 1.4–3.58^{3,4} and a mortality rate of 5.6% for China and 15.2% for outside of China,⁵ COVID-19 is now the third most fatal virus, after SARS-CoV⁶ and Middle East respiratory syndrome coronavirus (MERS-CoV),⁷ in the coronavirus family. The WHO has declared its pandemic,⁸ and as of October 4, 2020, almost 35 million people had confirmed diagnoses of COVID-19, resulting in 1,039,298 deaths worldwide.⁹ The epidemic curve is still in its ascending stage in Iran, and the official number of Iranian patients with COVID-19 and its resulting deaths were 471,772 and 26,957, respectively. Isfahan is considered one of the COVID-19 outbreak zones in Iran.

A number of demographic, clinical, and laboratory variables including age,^{10,11} smoking history,¹² hypertension,¹¹ diabetes, fever,^{11–13} respiratory rate,¹³ baseline neutrophil count,¹¹ albumin,¹³ lymphocyte count,^{11,13} white blood cells (WBC),¹¹ platelet count,¹¹ higher sequential organ failure assessment score,¹⁴ dyspnea,¹³ and blood *D*-dimer greater than 1 $\mu\text{g/mL}$ ¹⁴ have been collected and suggested to play a role in the primary and secondary outcomes of COVID-19.

In the current COVID-19 pandemic, there are several issues that still remain controversial and need to be further

investigated. For example, data on the prediction of disease prognosis are still inconclusive, and health authorities are concerned about the most important predictors of mortality from COVID-19. Although there is no cure for the patients identified with COVID-19 up to now, early finding of patients with poor prognosis based on a number of characteristics could have an important role in the COVID-19 outcomes. This may also assist the health system in more effective resource-allocating and also in providing more supportive, appropriate, and on time supportive treatment. Here in, using a case–control study in Isfahan COVID-19 registry (I-CORE), we aimed to investigate the most important risk or protective factors of in-hospital mortality in Isfahan, Iran.

METHOD

This was a case–control study of 513 Iranian patients with COVID-19 aged 1–97 years registered in the I-CORE,¹⁵ Isfahan, Iran. Data on all patients registered from February 18, the date of the first patient admission, to March 26, 2020 were obtained.

Participants. *Confirmation of patients with COVID-19.* We collected respiratory samples, that is, throat swabs for outcome ascertainment. The WHO interim guidance^{16,17} was used for detecting all patients with COVID-19 pneumonia. We also used real-time PCR to confirm the COVID-19 infection in all of the study blood samples.

Study cases. Cases were those patients with COVID-19 who died at the end of hospitalization.

Study controls. We classified those patients with COVID-19 who were alive at the time of discharge as the study controls. The following criteria were defined as discharge conditions: 1) return of body temperature to the normal values for ≥ 3 days, 2)

*Address correspondence to Shaghayegh Haghjooy Javanmard, Department of Physiology, Applied Physiology Research Center, Cardiovascular Research Institute, Isfahan University of Medical Sciences, Hezar Jerib Street, Isfahan 81745, Iran. E-mail: shaghayegh.haghjooy@gmail.com

observation of apparent inflammatory decrease in the pulmonary imaging, and 3) noticing significant improvement in the respiratory symptoms.

The data were anonymous, and the study was approved by the Research Ethics Committee of Isfahan University of medical science (IR.MUI.MED.REC.1398.733). The Ethics Committee of Isfahan University of medical science waived the prerequisite for patients' informed consent.

Measurements. A trained team of physicians and nurses as well as medical students reviewed and registered demographic, clinical, laboratory findings, and finally the outcome data. Using a standardized data collection checklist, we extracted data on aforementioned variables. Admission data and outcome (death or discharge) were obtained for all of the study individuals.

Clinical data. We obtained clinical data at hospital admission date. Data on medical comorbidities, that is, diabetes, cancers, cardiovascular disease, neurological disorder, asthma, HIV/AIDS, chronic blood disease, chronic liver disease, chronic kidney disease (CKD), other lung disease, and immunodeficiency, were collected. We also recorded clinical signs and symptoms including cough, fever, muscle pain, decreased consciousness, and respiratory distress. Using digital thermometers, fever was defined as axillary temperature $> 37^{\circ}\text{C}$ or oral temperature $> 37.5^{\circ}\text{C}$.

Laboratory findings. Routine blood examinations, that is, complete blood count, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, activated partial thromboplastin time (aPTT), prothrombin time (PT), international normalized ratio (INR), blood glucose, blood urea nitrogen (BUN), creatinine, C-reactive protein (CRP), hematocrit, hemoglobin (HB), lymphocytes percentage, neutrophils percentage, Na, K, lactic acid dehydrogenase, platelets, and WBC, were obtained. We also calculated the AST/ALT ratio as well as categories for all of the aforementioned blood indices (normal, $<$ normal, and $>$ normal values).

Demographic variables. The checklist also included demographic data, that is, age and gender.

Statistical analysis. Meeting the required assumptions, that is, normal test using the Kolmogorov–Smirnov test ($P > 0.05$), we used the independent t -test for continuous variables and chi-square test for categorical covariates in the bivariate analysis to compare survivors' and non-survivors' demographic, clinical, and laboratory features at the time of admission. We analyzed the laboratory variables as both continuous and categorized (less-than-normal, normal, and more-than-normal values). Where appropriate, by replacing the binary predictors with a single predictor and taking category rank scores, we perform chi-squared tests for trend of ordinal categorical variables. Logistic regression model was fitted for determining factors related to in-hospital mortality. The significance levels in all study tests used in the bivariable analysis was 0.05. A gradient boosting model was used to predict mortality and estimate the diagnostic power of the logistic regression model as well as the relative importance of each contributor. For the logistic model, we used Akaike information criteria as well as the Hosmer–Lemeshow test for model fitting. Furthermore, the contribution and importance of each variable was estimated. The area under the curve (AUC) of the suggested model was also calculated. Diagnostic performance of the model, that is, its sensitivity, specificity,

positive predictive value (PPV), and negative predictive value (NPV), and their 95% CIs were estimated.

Gradient boosting model. The gradient boosting is a machine learning technique suited to create predictive models by an ensemble of weak predictive models typically logistic regressions or decision trees.¹⁸ It is a boosting method that allows for optimization of a differentiable loss function. As any other boosting method, it belongs to supervised learning algorithms. To apply this model, we first created two sets from the input data for training and testing. All of the study variables were selected based on the literature review as well as the experts' comments. Based on the fitting criteria as well as diagnostic values, the final prediction model including a combination of clinical and laboratory variables was proposed. The default value of 0.5 was used for the value at which predicted probabilities are assigned to binary outcomes in both of the gradient boosting and logistic model. The relative importance of each contributor in the suggested model was estimated using the "varImp" function from the "caret" library in R.¹⁸

Balancing techniques. Using both oversampling and under-sampling methods, we repeated the whole of the analysis steps. We used the "upSample" and "downSample" functions, both from the "caret" package, to obtain an over-sampled and under-sampled datasets, respectively. Then, we used the "createDataPartition" to obtain a train (80%) and test (20%) on each dataset.¹⁹ The gradient boost implementation used, that is "xgbtree," was the one included in the "caret" package and had receiver operating characteristic (ROC) as the optimization metric (for more information, see Appendix). We trained the gradient boosting model using the train data. We also validated the models in two ways: first, using the test set created for each dataset (test set of oversampled and test set of under-sampled), and second, we used all the original data. The specificity, sensitivity, PPV, and NPV figures reported correspond to the oversampling experiment model, and has been validated using all the original data. While under-sampling approach might be an option, because of the small number of death cases, the under-sampled training set had only 96 individuals, which is very poor to train a gradient boosting model. On the other hand, the oversampling approach led to a train set of 726 individuals which was large enough. This approach takes into account the missing values, and for this reason, it is robust enough to capture the sensitivity of 100%. Although the oversampled test set might not be representative enough, validation was performed using both test set and all of the original data. For more information, see supplemental material.

Sample size calculation. For developing multivariable prediction models, the ratio of events per variable (EPV) would possibly be a practical criterion for sample size calculation. Simulation studies have suggested EPV values of between five and 20.²⁰ Considering the rarity of event, that is, death in COVID-19 and the 12 included predictors in the final model and a minimum of five EPV, we employed 513 patients with 60 events minimum in this study.

RESULTS

We enrolled 513 patients with COVID-19 between February 18 and March 26, 2020. All patients had well-known final hospital outcomes, and, of them, 60 (11.7%) died during

hospitalization. The median age was 61 years, and the mean (SD) age of survivors was 57.1 (16.9) years and was significantly lower than that of non-survivors, that is, 71.3 (13.4), $P < 0.001$. Gender was not significantly different between the study groups ($P = 0.399$). The mean (SD) intensive care unit (ICU) days were significantly higher in dead subjects (5.4 [5.0] versus 0.6 [2.4], $P < 0.001$). Among comorbidities, there was a statistically significant difference between the study groups in the case of chronic kidney disease (8.3% in dead versus 2.9% in survivors, $P = 0.039$). For clinical signs and symptoms, we also found that a higher proportion of survivors were hospitalized with cough as the signs and symptoms of COVID-19 ($P = 0.033$). Moreover, the proportion of those with ≥ 2 primary signs of COVID-19 was significantly higher in the survivors ($P = 0.038$) (Table 1).

Table 2 demonstrated the unadjusted associations between laboratory variables and final in-hospital mortality. Except for blood glucose, platelets, CRP, and PT, we observed a significant association between the other laboratory variables and in-hospital mortality in the bivariable analysis ($P < 0.05$). Less-than-normal values of lymphocytes percentage had the strongest association with mortality (odds ratio [OR] = 7.98; 95% CI: 3.52–18.09, $P < 0.001$), followed by more-than-normal values of BUN (OR = 7.11; 95% CI: 3.95–12.81, $P < 0.001$), more-than-normal values of creatinine (OR = 6.49;

95% CI: 3.63–11.60, $P < 0.001$), and more-than-normal values of HB (OR = 5.30, 95% CI: 1.32–21.28, $P = 0.019$). Interestingly, apparent dose–response pattern was detected for categories of hematocrit, HB, WBC, lymphocytes, and neutrophils (Table 2).

There was complete laboratory data on 252 patients. Moreover, clinical and demographic variables were also complete for these 252 patients. The suggested model ($n = 252$) was based on laboratory variables, that is, neutrophils' percentage, PT, hematocrit, CRP, aPTT, BUN, creatinine, and AST/ALT along with a number of clinical variables, that is, respiratory distress and age. In the multivariable logistic regression analysis, neutrophil percentage (OR = 1.37; 95% CI: 1.02–1.83, $P = 0.035$), PT (OR = 1.09; 95% CI: 1.02–1.17, $P = 0.012$), and less-than-normal values of hematocrit (OR = 6.03; 95% CI: 1.02–35.75, $P = 0.048$) significantly increased the risk of in-hospital mortality. The other remaining variables significantly increased the diagnostic power of the final suggested model. The relative importance of variables for prediction of in-hospital mortality in the final model is demonstrated in Figure 1. Blood urea nitrogen was the most important predictor of mortality followed by AST/ALT ratio, neutrophil percentage, and aPTT. The values of the suggested model based on the gradient boosting method with oversampling techniques using all of the

TABLE 1

Demographic and clinical characteristics of patients with COVID-19 by in-hospital mortality in total sample and age-groups, Isfahan, Iran, 2020

	Total sample			Age-groups	
	Survivors	Non-survivors	Unadjusted OR, 95% CI*, P-value	≤ 60	> 60
Mean age (SD)	57.08 (16.96)	71.3 (16.37)	1.06 (1.04, 1.08), < 0.001	0.99 (0.94, 1.05), 0.926	1.07 (1.03, 1.11), < 0.001
Median age	58	74	–	–	–
Gender (male), N (%)	261 (57.62)	38 (63.33)	1.27 (0.73, 2.22), 0.399	1.24 (0.37, 4.09), 0.728	1.24 (0.66, 2.35), 0.501
Medical history variables, n (%)					
Cancer (yes)	13 (2.87)	2 (3.33)	1.17 (0.26, 5.30), 0.841	1.88 (0.09, 36.05), 0.676	1.26 (0.30, 5.34), 0.754
HIV/AIDS (yes)	2 (0.44)	0 (0)	1.49 (0.07, 31.45), 0.797	7.00 (0.27, 181.40), 0.241	1.42 (0.06, 35.32), 0.832
Asthma (yes)	13 (2.87)	1 (1.67)	0.57 (0.07, 4.46), 0.596	5.20 (0.79, 34.06), 0.086	0.27 (0.01, 4.90), 0.380
Chronic neurologic disorder (yes)	9 (1.99)	0 (0)	0.39 (0.02, 6.73), 0.514	2.30 (0.12, 45.41), 0.583	0.38 (0.02, 6.97), 0.514
Cardiovascular disease (yes)	69 (15.23)	9 (15.00)	0.98 (0.46, 2.09), 0.963	1.47 (0.25, 8.61), 0.668	0.69 (0.31, 1.54), 0.367
Blood chronic disease (yes)	5 (2.02)	0 (0)	0.51 (0.027, 9.33), 0.647	7.48 (0.27, 205.51), 0.234	0.32 (0.02, 6.11), 0.450
Chronic liver disease (yes)	4 (0.88)	0 (0)	0.82 (0.04, 15.52), 0.898	2.97 (0.14, 61.07), 0.479	1.42 (0.57, 35.32), 0.832
Diabetes (yes)	61 (13.47)	11 (18.33)	1.44 (0.71, 2.93), 0.355	1.27 (0.22, 7.41), 0.787	1.24 (0.57, 2.66), 0.587
Chronic kidney disease (yes)	13 (2.87)	5 (8.33)	3.08 (1.06, 8.96), 0.039	6.17 (0.91, 41.59), 0.062	2.37 (0.72, 7.56), 0.155
Other chronic disease (yes)	65 (14.35)	10 (16.67)	1.19 (0.58, 2.47), 0.633	2.98 (0.81, 10.97), 0.100	0.85 (0.36, 1.99), 0.706
Immune deficiency (yes)	4 (0.88)	0 (0)	0.84 (0.045, 15.79), 0.907	2.97 (0.14, 61.07), 0.479	1.45 (0.06, 36.06), 0.822
Total comorbidities, n (%)					
0	286 (63.13)	31 (52.54)	1	1	1
1	104 (22.96)	19 (32.20)	1.68 (0.91, 3.11), 0.095	1.47 (0.34, 6.40), 0.608	1.20 (0.60, 2.39), 0.600
≥ 2	63 (13.91)	9 (15.25)	1.32 (0.60, 2.90), 0.494	2.43 (0.55, 10.80), 0.244	0.84 (0.34, 2.06), 0.704
Signs and symptoms at admission date, n (%)					
Total signs and symptoms	1.75 (1.07)	1.63 (1.25)	0.90 (0.70, 1.16), 0.425	1.18 (0.66, 2.13), 0.571	0.84 (0.64, 1.11), 0.228
Categorized total signs and symptoms					
None	74 (16.34)	17 (28.33)	1	1	1
1	81 (17.88)	8 (13.33)	0.43 (0.17, 1.05), 0.065	0.20 (0.01, 4.40), 0.311	0.44 (0.17, 1.15), 0.094
≥ 2	298 (65.78)	35 (58.33)	0.51 (0.27, 0.96), 0.038	0.98 (0.23, 4.10), 0.974	0.42 (0.20, 0.86), 0.019
Fever	238 (52.54)	26 (43.33)	0.69 (0.40, 1.19), 0.182	1.11 (0.34, 3.54), 0.865	0.60 (0.32, 1.13), 0.113
Muscle pain	115 (25.39)	13 (21.67)	0.81 (0.42, 1.56), 0.532	0.71 (0.17, 2.95), 0.641	0.95 (0.46, 1.98), 0.902
Cough	248 (54.75)	24 (40.00)	0.55 (0.32, 0.95), 0.033	1.86 (0.52, 6.64), 0.337	0.44 (0.23, 0.65), 0.014
Decreased consciousness	8 (1.77)	3 (5.00)	2.93 (0.75, 11.35), 0.120	7.00 (0.27, 181.40), 0.241	2.05 (0.55, 7.59), 0.281
Respiratory distress	185 (40.84)	32 (53.33)	1.65 (0.96, 2.84), 0.068	1.48 (0.46, 4.74), 0.512	1.43 (0.77, 2.66), 0.253

OR = odds ratio. The bold values are significant at 0.05 level.

* In the case of sparse data, the OR (95% CI) was calculated using the penalized likelihood estimator.

TABLE 2
Laboratory characteristics of patients with COVID-19 by in-hospital mortality, Isfahan, Iran, 2020

Laboratory variable*	Survivors, n (%)	Non-survivors, n (%)	Unadjusted OR, 95% CI†	P-value
Mean ALT (SD)	34.60 (36.50)	39.88 (43.00)	1.00 (0.99, 1.01)	0.393
Categorized ALT (IU/L)				
7–55	250 (86.21)	34 (79.07)	1	–
< 7	2 (0.69)	0 (0)	1.45 (0.07, 30.88)	0.811
> 55	38 (13.10)	9 (20.93)	1.79 (0.81, 3.96)	0.150
<i>Test for trend</i>	–	–	1.33 (0.89, 1.98)	0.163
Mean AST (SD)	49.83 (60.68)	67.86 (58.80)	1.00 (0.99, 1.01)	0.130
Categorized AST (U/L)				
8–48	211 (86.21)	21 (79.07)	1	–
< 8	0 (0)	0 (0)	–	–
> 48	79 (0.69)	21 (0)	2.67 (1.38, 5.16)	0.003
AST/ALT ratio (IU/L)	1.75 (1.28)	2.07 (1.12)	1.15 (0.94, 1.40)	0.167
Neutrophils, n %	70.43 (12.18)	80.60 (12.31)	1.09 (1.06, 1.13)	< 0.001
Categorized neutrophils %				
55–70	161 (37.18)	8 (14.04)	1	–
< 50	42 (9.70)	2 (3.51)	0.96 (0.20, 4.68)	0.958
> 70	230 (53.12)	47 (82.46)	4.11 (1.89, 8.93)	< 0.001
<i>Test for trend</i>	–	–	2.13 (1.44, 3.16)	< 0.001
Lymphocytes, n %	22.67 (10.62)	14.54 (11.11)	0.91 (0.87, 0.94)	< 0.001
Lymphocytes %				
20–40	220 (50.57)	7 (12.28)	1	–
< 20	185 (42.53)	47 (82.46)	7.98 (3.52, 18.09)	< 0.001
> 40	30 (6.90)	3 (5.26)	3.14 (0.77, 12.81)	0.110
<i>Test for trend</i>	–	–	2.56 (1.64, 4.01)	< 0.001
Mean blood urea nitrogen (SD)	17.63 (10.42)	33.75 (24.20)	1.08 (1.05–1.10)	< 0.001
Categorized blood urea nitrogen (mg/dL)				
6–25	358 (84.83)	26 (45.61)	1	–
< 6	4 (0.95)	0 (0)	1.50 (0.08, 28.70)	0.786
> 25	60 (14.22)	31 (54.39)	7.11 (3.95, 12.81)	< 0.001
Mean creatinine (SD) (mg/dL)	1.12 (0.50)	1.84 (1.32)	2.59 (1.82, 3.68)	< 0.001
Categorized creatinine (mg/dL)				
For male: 0.7–1.3	365 (83.52)	25 (43.86)	1	–
For female: 0.4–1.1				
For male: < 0.7	0 (0)	0 (0)	–	–
For female: < 0.4				
For male: > 1.3	72 (16.48)	32 (56.14)	6.49 (3.63, 11.60)	< 0.001
For female: > 1.1				
Mean WBC (SD) (cells/L),	5,816.05 (2,692.70)	8,403 (4,303.5)	1.0002 (1.0001, 1.0003)	< 0.001
Categorized WBC (cells/L)				
4,000–10000	310 (68.43)	39 (66.10)	1	–
< 4,000	111 (24.50)	3 (5.08)	0.21 (0.06, 0.71)	0.012
> 10000	32 (7.06)	17 (28.81)	4.22 (2.15, 8.30)	< 0.001
<i>Test for trend</i>	–	–	1.62 (1.12, 2.34)	0.009
Mean aPTT (SD) (sec)	35.77 (12.91)	42.59 (18.11)	1.02 (1.01, 1.04)	0.003
Categorized aPTT (sec)				
30–45	198 (58.41)	32 (65.31)	1	–
< 30	104 (30.68)	6 (12.24)	0.36 (0.14, 0.88)	0.025
> 45	37 (10.91)	11 (22.45)	1.84 (0.85, 3.97)	0.121
<i>Test for trend</i>	–	–	1.09 (0.71, 1.66)	0.697
Mean CRP (SD) (mg/L)	46.12 (72.66)	59.81 (57.55)	1.002 (0.998, 1.005)	0.333
Categorized CRP (mg/L)				
< 10	52 (17.99)	4 (12.50)	1	–
> 10	237 (82.01)	28 (87.50)	1.53 (0.51, 4.57)	0.440
Mean Hematocrit (SD)	41.81 (5.83)	39.21 (7.06)	0.92 (0.88, 0.97)	0.001
Categorized hematocrit				
For male: 42–54 (% L/L)	271 (60.63)	18 (30.00)	1	–
For female: 38–46%				
For male: < 42%	159 (35.57)	37 (61.67)	3.50 (1.93, 6.36)	< 0.001
For female: < 38%				
For male: > 54%	17 (3.80)	5 (8.33)	4.43 (1.47, 13.37)	0.008
For female: > 46%				
<i>Test for trend</i>	–	–	2.58 (1.67, 4.01)	< 0.001
Mean HB (SD)	13.47 (1.81)	12.58 (2.50)	0.79 (0.70, 0.91)	0.001
Categorized HB (mg/dL)				
For male: 14–17.4	286 (63.13)	18 (30.00)	1	–
For female: 12–16				
For male: < 14	158 (34.88)	39 (65.00)	3.92 (2.17, 7.08)	< 0.001
For female: < 12				
For male: > 17.4	9 (1.99)	3 (5.00)	5.30 (1.32, 21.28)	0.019

(continued)

TABLE 2
Continued

Laboratory variable*	Survivors, n (%)	Non-survivors, n (%)	Unadjusted OR, 95% CI†	P-value
For female: > 16				
<i>Test for trend</i>	–	–	3.12 (1.93, 5.05)	< 0.001
AST/ALT ratio	1.75 (1.28)	2.07 (1.12)	1.15 (0.94–1.40)	0.167
Mean platelets (SD)	181,000.15 (71,000.66)	168,000.61 (72,000.31)	0.997 (0.993, 1.001)	0.204
Categorized platelets				
140,000–400,000/dL	304 (67.71)	37 (61.67)	1	–
< 140,000/dL	140 (31.18)	23 (38.33)	1.35 (0.78, 2.36)	0.292
> 400,000/dL	5 (1.11)	0 (0)	0.74 (0.04, 13.62)	0.838
Mean blood glucose (SD) (mg/dL)	143.30 (64.07)	149.95 (79.87)	1.001 (0.995, 1.007)	0.644
Categorized blood glucose (mg/dL)				
90–130	104 (50.00)	12 (52.17)	1	–
< 90	18 (8.65)	1 (4.35)	0.48 (0.059, 3.93)	0.495
> 130	86 (41.35)	10 (43.48)	1.01 (0.41, 2.44)	0.986
<i>Test for trend</i>	–	–	0.999 (0.64, 1.57)	0.998
Mean PT (SD) (sec)	13.27 (3.79)	16.53 (16.71)	1.04 (1.001, 1.09)	0.047
Categorized PT (sec)				
10–12	97 (30.99)	16 (31.37)	1	–
< 10	2 (0.64)	0 (0)	1.18 (0.05, 25.73)	0.915
> 12	214 (68.37)	35 (68.63)	0.99 (0.524, 1.88)	0.979

ALT = alanine aminotransferase; aPTT = activated partial thromboplastin time; AST = aspartate aminotransferase; CRP = C-reactive protein; HB = hemoglobin; OR = odds ratio; PT = prothrombin time. The bold values are significant at 0.05 level.

* The categories were created based on specific values for each variable, that is, normal, less than normal, and more than normal.

† In the case of sparse data, the OR (95% CI) was calculated using the penalized likelihood estimator.

original data were AUC of 98.6, sensitivity of 100 (95% CI: 94–100), specificity of 98.5% (95% CI: 96.8–99.4), an NPV of 100% (95% CI: 99.2–100), and a PPV of 89.6 (95% CI: 21.2–86.3). The values of the suggested model based on the gradient boosting method with oversampling techniques using test set data were AUC of 98.9, sensitivity of 100 (95% CI: 96.0–100), specificity of 93.3% (95% CI: 86.1–97.5), an NPV of 100% (95% CI: 95.7–100), and a PPV of 93.8 (95% CI: 86.9–97.7). However, using logistic regression and “estat classification” command in STATA v. 12 (StataCorp.,

College Station, TX), the obtained diagnostic values were AUC of 94.1, sensitivity of 64.7 (95% CI: 57.3–72.2), specificity of 97.9% (95% CI: 95.6–100), NPV of 95.8% (95% CI: 92.7–98.9), and PPV of 78.6 (95% CI: 72.2–85.00) (Table 3).

DISCUSSION

Using data of I-CORE system, we found an in-hospital mortality of 11.7%. Furthermore, based on early-available clinical and laboratory data, we suggested a prediction model

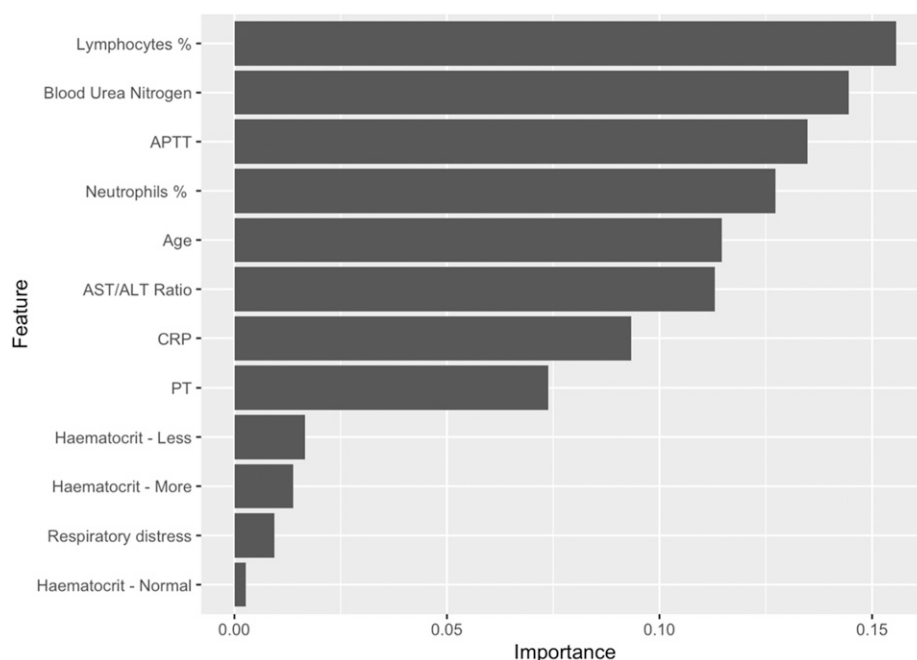


FIGURE 1. Variable importance for gradient boosting model using the “varImp” function.

TABLE 3
Multivariable logistic regression models of in-hospital mortality in patients with COVID-19, Isfahan, Iran, 2020

Predictor	Odds ratio (95% CI)		P-value
Neutrophils percentage	1.37 (1.02, 1.83)		0.035
Prothrombin time	1.09 (1.02, 1.17)		0.012
Categorized hematocrit			
Less than normal	6.03 (1.02, 35.75)		0.048
More than normal	3.86 (0.19, 77.18)		0.377
C-reactive protein	1.01 (0.99, 1.03)		0.244
Activated partial thromboplastin time	1.01 (0.96, 1.06)		0.711
Blood urea nitrogen	1.03 (0.99, 1.07)		0.185
Lymphocyte percentage	1.29 (0.92–1.83)		0.141
Creatinine	0.89 (0.43, 1.81)		0.745
Aspartate aminotransferase/alanine aminotransferase	1.42 (0.43, 6.49)		0.453
Respiratory distress	2.89 (0.49, 17.17)		0.243
Hemoglobin	1.11 (0.66, 1.86)		0.704
Age	1.03 (0.97, 1.09)		0.292
Measures of model performance	Logistic model	Gradient boosting model using all of the original data	Gradient boosting model using test set data
Specificity	97.9% (95% CI: 95.6–100)	98.5% (95% CI: 96.8–99.4)	93.3% (95% CI: 86.1–97.5)
Negative predictive value	95.8% (95% CI: 92.7–98.9)	100% (95% CI: 99.2–100)	100% (95% CI: 95.7–100)
Sensitivity	64.7% (95% CI: 57.3–72.2)	100% (95% CI: 94–100)	100% (95% CI: 96.0–100)
Positive predictive value	78.6% (95% CI: 72.2–85)	89.6% (95% CI: 79.7–95.7)	93.8% (95% CI: 86.9–97.7)
ROC	94.1	98.6	98.9
Akaike information criteria	0.502	–	–

for in-hospital mortality with an AUC of 98.6%. In term of relative importance, lymphocyte percentage, higher BUN, aPTT, neutrophil percentage, and age were the most important variables for prediction of in-hospital mortality in patients with COVID-19. Moreover, a higher AST/ALT ratio, CRP, PT, hematocrit, and respiratory distress were all among the top 10 important predictors of mortality.

The in-hospital mortality reported in this study is moderately higher than that (7.2%) reported by Kun Wang¹¹ in Wuhan, China. Several factors including the hospital referral time or the disease severity at the time of admission may be responsible for this difference. Among 10 reported important factors of in-hospital mortality in our study, there are five in common with Kun Wang study, that is, neutrophil percentage, aPTT, CRP, age, and lymphocyte percentage in. Although neutrophil percentage was reported as the most important predictor of mortality in Wuhan, here, with relative importance of around 13%, it is the fourth most important predictor in our study. We reported lymphocyte percentage and higher BUN as the first two most important predictors in our analysis. However, these two factors were not reported before.¹¹ C-reactive protein level is described in both studies as the fifth most important predictor of mortality.

The prevalence of CKD was significantly higher in non-survivors. Consistent with our findings, Cheng et al.²¹ reported a statistically significant association between kidney disease and elevated risk of in-hospital death in a prospective cohort study of 701 patients with COVID-19 in Wuhan. Acute renal failure has been reported in 0.5–7% of COVID-19 cases and in 2.9–23% of ICU patients.^{12,21,22} There is little evidence of

direct kidney injury by SARS-CoV-2; however, kidney involvement is common in complications of acute respiratory distress syndrome (ARDS). Pan et al.²³ have shown potential vulnerability of podocytes and proximal tubule cells through SARS-CoV-2–angiotensin-converting enzyme-2 interaction by single-cell RNA-seq analysis.²⁴ Cytokine damage and kidney failure secondarily to lung injury could be the other possible mechanisms.²⁵ Liver injury and failure following critical illnesses such as ARDS have been reported in up to 20% of ICU patients and are associated with significantly increased morbidity and mortality. The underlying mechanisms may be hypoxic, inflammatory, or drug induced.²⁶

In line with the results of a hospital-based case-cohort study on 200 COVID-19 patients in Wuhan, China,²⁷ we found a positive but statistically insignificant association between higher values of AST/ALT ratio and in-hospital mortality. Furthermore, we found that the baseline AST/ALT ratio, a marker of hepatic injury,²⁸ would be a strong predictor of mortality in the gradient boosting model. Although the elevated level of AST and ALT in patients with COVID-19 has been formerly reported,^{29,30} here, we also highlighted the prognostic value of AST/ALT in COVID-19 mortality.

Higher values of neutrophil percentage were associated with increased risk of mortality. This finding supports the previously reported results of a study on 61 patients with 2019-nCoV at Beijing Ditan Hospital.³¹ Although neutrophils have an important role at the time of infection, neutrophil infiltration might be a proxy for harmful inflammatory processes. We found that elevated aPTT may play a role in poor outcomes in patients with COVID-19. This is consistent

with the findings of Songping Cui et al.³² that reported significant prolongation of PT and aPTT as a reliable factor for identifying high-risk groups of patients with severe COVID-19. COVID-19 coagulopathy reflected by thrombocytopenia, prolongation of the PT/INR, PTT, elevation of *D*-dimer, and decreased fibrinogen levels could lead to disseminated intravascular coagulation (DIC).³³ Tang et al.³⁴ have shown that DIC in COVID-19 patients is associated with a higher mortality rate. A strong inflammatory index, CRP, was identified as one of the predictors of mortality in this study. It is now well accepted that CRP levels generally increase in patients with severe COVID-19 even at its initial stages. Moreover, it is positively correlated with lung damage and could predict disease severity.^{35,36}

In line with the findings of previous studies,^{11,13,14} older age was associated with an increase in the odds of mortality. Finally, among signs and symptoms on admission, respiratory distress was one of the 10 predictors of mortality.

Wynants et al.³⁷ in a systematic review highlighted that the currently suggested models for mortality from COVID-19 are poorly reported and had been suffered from some degree of bias.

The lack of differences in comorbidities and also the number of comorbidities between the survivor and non-survivor groups, apart from CKD, are a surprising result, given the existing literature. However, the OR (95% CI) shown in Table 1 are unadjusted one, and it would be statistically possible to find different or even a significant *P*-value if one would investigate the adjusted associations. Although the study's main objective was to propose a prediction model not adjusted ORs, we have not investigated such associations. Moreover, it is likely that hospital admission policy as well as referral patterns of patients with COVID-19 in Iran has been suffered from some degree of bias, that is, a subgroup of patients with high prevalence and history of illnesses more likely referred to hospitals and had a higher chance of hospital admission. Therefore, both survivors and non-survivors with a higher history of comorbidities were more likely hospitalized during the pandemic. We only reported the mortality in those hospitalized patients. Moreover, as the outcome of this study was in-hospital mortality, the real long-term risk of mortality would be different than the estimated risk in this study. The possibility and magnitude of imposed bias on the estimated risk of mortality cannot be predictable without sufficient information about the referral pattern of patients and its associated factors. Because of the retrospective nature of the study design, we only analyzed laboratory data of 252 patients because of the lack of complete information for all of the study subjects. This might decrease the precision of the study estimates. Moreover, data on *D*-dimer were not available for analysis, and the estimated results have not been adjusted for its effect in the final suggested model. The obtained sensitivity of the suggested model was low. One of the possible reasons is the high imbalance of dead and alive individuals in our sample, with an extremely higher proportion of alive discharged patients. Rare event bias could explain the low obtained sensitivity of the proposed model. The lower the proportion of the study events, the lower the predictive accuracy for events and the higher the predictive accuracy for nonevents.³⁸ Nonetheless, we consider that this is a useful model for clinical practice as it has a high specificity of 95.6% and a

high NPV of 92.5%. This means that our model correctly classifies those patients who survived from COVID-19. This is especially important to prioritize resources in health services, especially when the financial and human resources are limited. However, the registry needs to be evaluated and modified in term of its low sensitivity. Another limitation is that these results are not validated against another dataset. However, the gradient boosting method is a supervised learning method and the data used were divided into two sets, train and test, to avoid potential overfittings. Thus, it focuses on increasing the external validity of the results. Several determinants might be different between Iran and the other countries including the pattern of referral to the hospital, the average time from primary signs to hospital referral as well as the public educational programs, and health policy-maker decisions. This might be even considered as explanation for a slight different reported model performance of the suggested prediction model in this sample.

Conclusion. In summary, lymphocyte percentage, elevated BUN, aPTT, neutrophil percentage, and older age were the most important predictors of in-hospital mortality in patients with COVID-19 in Isfahan, Iran. The suggested model may help identify patients with poor prognosis and may be useful in making decision to patient's hospitalization where the probability of mortality may be more obvious based on the final variable.

Received August 18, 2020. Accepted for publication February 1, 2021.

Published online February 16, 2021.

Note: Supplemental material appear at www.ajtmh.org.

Acknowledgments: We would like to thank the Deputy of Research and Technology at the Isfahan University of Medical Sciences for all their support during the design and implementation of this study. Publication charges for this article were waived due to the ongoing pandemic of COVID-19.

Financial support: This work was supported by Isfahan University of medical science, "(grant number 198285)."

Disclosure: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. All Stata and *R* codes are available.

Authors' addresses: Ibrahim Abdollahpour, Child Growth and Development Research Center, Research Institute for Primordial Prevention of Non-Communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran, E-mail: ebrahemen@gmail.com. Isabel Aguilar-Palacio, Preventive Medicine and Public Health Department, University of Zaragoza, Zaragoza, Spain, E-mail: iaguilar@unizar.es. Juan Gonzalez-Garcia, Instituto Aragones de Ciencias de la Salud, Biocomputing Unit/Data Science in Health Services and Policy Research Group/Health Services Research Network on Chronic Patients (REDISSEC), Zaragoza, Spain, E-mail: jgonzalezgarc.iaacs@aragon.es. Golnaz Vaseghi, Cardiovascular Research Centre, Isfahan University of Medical Sciences, Isfahan, Iran, E-mail: golnazvaseghi@yahoo.com. Zahra Otraj, Amirreza Manteghinejad, and Azam Mosayebi, Applied Physiology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran, E-mails: zahraotroj@gmail.com, manteghinejad@edc.mui.ac.ir, and a.mosayebi98@gmail.com. Yahya Salimi, Department of Iran, Epidemiology, Kermanshah University of Medical Sciences, Kermanshah, Iran, E-mail: yahya.salimi@kums.ac.ir. Shaghayegh Haghjooy Javanmard, Department of Physiology, Isfahan University of Medical Sciences, Isfahan, Iran, E-mail: shaghayegh.haghjoo@gmail.com.

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC-BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

REFERENCES

1. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S, 2020. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect Genet Evol* 79: 104212.
2. Chan JFW, Yuan S, Kok KH, To KKW, Chu H, Yang J, Xing F, Liu J, Yip CCY, Poon RWS, 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395: 514–523.
3. WHO, 2020. *Statement on the Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV)*. Geneva, Switzerland: World Health Organization. Available at: [https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)). Accessed January 20, 2020.
4. Zhu N et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. *New Engl J Med* 382: 727–733.
5. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G, 2020. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis* 20: 773.
6. Jiang Y, Xu J, Zhou C, Wu Z, Zhong S, Liu J, Luo W, Chen T, Qin Q, Deng P, 2005. Characterization of cytokine/chemokine profiles of severe acute respiratory syndrome. *Am J Respir Crit Care Med* 171: 850–857.
7. Niu P, Zhao G, Deng Y, Sun S, Wang W, Zhou Y, Tan W, 2018. A novel human mAb (MERS-GD27) provides prophylactic and postexposure efficacy in MERS-CoV susceptible mice. *Sci China* 61: 1280–1282.
8. WHO, 2020. *WHO Announces COVID-19 Outbreak a Pandemic*. Geneva, Switzerland: World Health Organization. Available at: <http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>. Accessed January 20, 2020.
9. Worldometers.info A, 2020, Dover, Delaware. Available at: <https://www.worldometers.info/coronavirus/>. Accessed August 10, 2020.
10. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, 2020. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 323: 1061–1069.
11. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, Zhang H, Chen X, Liu C, 2020. Clinical and laboratory predictors of in-hospital mortality in 305 patients with COVID-19: a Cohort Study in Wuhan, China. *Clin Infect Dis* 71: 2079–2088.
12. Liu W et al., 2020. Analysis of factors associated with disease outcomes in hospitalized patients with 2019 novel coronavirus disease. *Chin Med J (Engl)* 133: 1032–1038.
13. Huang C et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395: 497–506.
14. Zhou F et al., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395: 1054–1062.
15. Javanmard SH, Nasirian M, Ataei B, Vaseghi G, Vaezi A, Changiz T, 2020. Isfahan COVID-19 Registry (I-CORE): design and methodology. *J Res Med Sci* 25: 32.
16. WHO, 2020. *Clinical Management of Severe Acute Respiratory Infection when 27 Novel Coronavirus (nCoV) Infection Is Suspected: Interim Guidance*. Geneva, Switzerland: World Health Organization. Available at: [https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-\(ncov\)-infection-is-suspected](https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected). Accessed January 20, 2020.
17. WHO, 2020. *Clinical Management of Severe Acute Respiratory Infection when Novel Coronavirus (nCoV) Infection Is Suspected: Interim Guidance, 25 January 2020*. Geneva, Switzerland: World Health Organization.
18. Boehmke B, Greenwell BM, 2019. *Hands-on Machine Learning with R*. Boca Raton, FL: CRC Press.
19. Kuhn M, 2008. Building predictive models in R using the caret package. *J Stat Software* 28: 1–26.
20. Ogundimu EO, Altman DG, Collins GS, 2016. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol* 76: 175–182.
21. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, Li J, Yao Y, Ge S, Xu G, 2020. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 97: 829–838.
22. Guan WJ et al., 2020. Clinical characteristics of coronavirus disease 2019 in China. *New Engl J Med* 382: 1708–1720.
23. Pan XW, Xu D, Zhang H, Zhou W, Wang LH, Cui XG, 2020. Identification of a potential mechanism of acute kidney injury during the COVID-19 outbreak: a study based on single-cell transcriptome analysis. *Intensive Care Med* 46: 1114–1116.
24. Bernheim A et al., 2020. Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection. *Radiology* 295: 200463.
25. Ronco C, Reis T, 2020. Kidney involvement in COVID-19 and rationale for extracorporeal therapies. *Nat Rev Nephrol* 16: 308–310.
26. Horvatits T, Drolz A, Trauner M, Fuhrmann V, 2019. Liver injury and failure in critical illness. *Hepatology* 70: 2204–2215.
27. Fu L et al., 2020. Analysis of death risk factors among 200 COVID-19 patients in Wuhan, China: a hospital-based case-cohort study. *Lancet* 25.
28. Nyblom H, Berggren U, Balldin J, Olsson R, 2004. High AST/ALT ratio may indicate advanced alcoholic liver disease rather than heavy drinking. *Alcohol Alcohol* 39: 336–339.
29. Xie H, Zhao J, Lian N, Lin S, Xie Q, Zhuo H, 2020. Clinical characteristics of non-ICU hospitalized patients with coronavirus disease 2019 and liver injury: a retrospective study. *Liver Int* 40: 1321–1326.
30. Zhao D, Yao F, Wang L, Zheng L, Gao Y, Ye J, Guo F, Zhao H, Gao R, 2020. A comparative study on the clinical features of COVID-19 pneumonia to other pneumonias. *Clin Infect Dis* 71: 756–761.
31. Liu J, Liu Y, Xiang P, Pu L, Xiong H, Li C, Zhang M, Tan J, Xu Y, Song R, 2020. Neutrophil-to-lymphocyte ratio predicts severe illness patients with 2019 novel coronavirus in the early stage. *J Transl Med* 18: 1–12.
32. Cui S, Chen S, Li X, Liu S, Wang F, 2020. Prevalence of venous thromboembolism in patients with severe novel coronavirus pneumonia. *J Thromb Haemost* 18: 1421–1424.
33. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, Psaltopoulou T, Gerotziapas G, Dimopoulos MA, 2020. Hematological findings and complications of COVID-19. *Am J Hematol* 95: 834–847.
34. Tang N, Li D, Wang X, Sun Z, 2020. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost* 18: 844–847.
35. Ling W, 2020. C-reactive protein levels in the early stage of COVID-19. *Méd Mal Infect* 50: 332–334.
36. Tan C, Huang Y, Shi F, Tan K, Ma Q, Chen Y, Jiang X, Li X, 2020. C-reactive protein correlates with CT findings and predicts severe COVID-19 early. *J Med Virol* 92: 856–862.
37. Wynants L, Van Calster B, Bonten MM, Collins GS, Debray TP, De Vos M, Haller MC, Heinze G, Moons KG, Riley RD, 2020. Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection. *medRxiv*.
38. Lusa L, 2017. Gradient boosting for high-dimensional prediction of rare events. *Comput Stat Data Anal* 113: 19–37.