Value of laboratory tests in COVID-19 hospitalized patients for clinical decision-makers: a predictive model, using data mining approach

Atefeh Mousavi

Shahid Beheshti University of Medical Sciences <https://orcid.org/0000-0001-7606-8237>

Soheyla Rezaei

Shahid Beheshti University of Medical Sciences

Jamshid Salamzadeh

Shahid Beheshti University of Medical Sciences

Ali Mirzazadeh

University of California, San Francisco

Farzad Peiravian

Shahid Beheshti University of Medical Sciences

Nazila Yousefi (✉ n.yousefi@sbmu.ac.ir)

Shahid Beheshti University of Medical Sciences <https://orcid.org/0000-0001-8079-7424>

Research Article

Keywords: COVID-19, Datamining, Laboratory test, Risk factor, Mortality rate

Posted Date: August 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-56252/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose: Because of the rapid increase in confirmed cases of COVID-19, in particular those with severe or critical status, overwhelming of health systems is a worldwide concern. Therefore, identifying high-risk COVID-19 patients, can help service providers for priority setting and hospital resource allocation.

Methods: 4542 adult patients with confirmed COVID-19 admitted in 15 hospitals in Tehran, Iran, from Feb 20 to April 18, 2020 were included in this retrospective cohort study with final outcomes of survived and died patients. Demographic features including age and sex, and laboratory data measured at admission were extracted and compared between recovered and died patients. Data analysis was performed applying SPSS modeler software using a logistic regression method.

Results: Of 4542 hospitalized adult patients, 822 patients (18.09%) died during hospitalization, and 3720 (81.90%) recovered and discharged. Based on logistic regression model, older age, 40-49 (RR= 1.80, CI: 1.13-2.87), 50-59 (RR=2.63, CI: 1.71-4.02), 60-69 (RR= 4.40, CI: 2.92-6.63), 70-79 (RR=7.49, CI: 5.01-11.19), Above 80 (RR=13.85, CI: 9.23-2.77), ALT $\geq$ 55 IU/ (RR=2.20, CI: 1.69-2.86), AST $\geq$ 100 IU/L (RR=5.93, CI: 4.75-7.39), ALP $\geq$ 200 IU/L (RR=2.46, CI: 1.80-3.37), sodium < 135 mEq/l (RR=1.69, CI: 1.35-2.11) or more than 145 mEq/l (RR=7.24, CI: 5.07-10.33), potassium > 5.50 mEq/l (RR=7.53, CI: 4.15-13.64), and calcium < 8.50 mEq/l (RR=3.39, CI: 2.81-4.09), CPK between 307-600 IU/L (RR=2.73, CI: 2.12-3.53) and above 600 IU/L (RR=4.41, CI: 3.40-5.71) in men, and 192-400 IU/L (RR=2.73, CI: 2.12-3.53), and above 400 (RR=4.41, CI: 3.40-5.71) in women, CRP > 3 mg/l (RR=3.22, CI: 1.99-5.20), and creatinine > 1.5 mg/l (RR=6.37, CI: 5.30-7.66) were significantly associated with COVID-19 mortality.

Conclusion: Our findings suggested less than one in five hospitalized patients with COVID-19 die mostly due to electrolyte disbalance, liver, and renal dysfunctions. Better supportive care is needed to improve outcomes for patients with COVID-19.

Introduction

In December 2019, an outbreak of Coronavirus disease 2019 (COVID-19) began in Wuhan, China, and has continued to spread globally [17]. COVID-19 can cause fever, cough, fatigue, shortness of breath, and high mortality due to severe respiratory symptoms [5].

The first official announcement of deaths caused by COVID-19 in Iran, was made on Feb 19, 2020[30]. As of July 15, 2020, 264,561 Iranians have been infected with COVID-19, of whom 13,410 are deceased. Due to the rapidly growing number of confirmed cases of COVID-19, there is a worldwide concern that overwhelmed health systems may face shortages of hospital beds, ICU beds, and ventilators [4], as well as burnout and fatigue [24] of healthcare professionals. Lack of resources in the health sector can lead to discrimination in the optimal distribution of medical services during pandemic diseases.

Prediction of disease severity at the time of a pandemic outbreak is one of the critical issues that influence physicians' decisions [34]. Previous experience with other outbreaks such as MERS and pandemic influenza has shown that identifying high-risk patients for hospitalization can help healthcare providers and emergency staffs in finding patients who would benefit the most from early, available treatments. In addition, policymakers can use this information to forecast healthcare needs [4].

There is a dearth of literature on assessing lab tests as risk factors associated with COVID-19 prognosis. Although, an elevated amount of alanine aminotransferase (ALT), aspartate aminotransferase (AST), C-reactive protein (CRP), lactate dehydrogenase, d-dimer, and low serum concentrations of albumin, sodium, potassium and calcium were reported to be associated with disease severity, unfavorable prognosis and mortality rate, most of these articles were limited to a specific population or a small sample size. Moreover, there are still other potential explanatory variables that require further assessment [28,33,37].

To achieve a robust estimation of related risk factors, having masses of data from different countries with different health-system settings is crucial.

In this study, we analyzed the available clinical and laboratory data of 4542 patients with confirmed COVID-19, obtained from healthcare records of multiple hospitals in Tehran, Iran. Using machine learning approaches that can learn algorithms through modeling can improve the predictive models, such as prediction of disease risk factors and mortality rate [21].

Methods

Study design

The present research is a retrospective cohort study of 4542 hospitalized patients with severe symptoms of COVID-19 in 15 hospitals in Tehran, Iran. These patients were later divided into two groups: recovered and discharged patients, and deceased patients. The predictors were demographic characteristics of sex and age, along with different laboratory tests. Also, the oxygen therapy and mechanical ventilation were set as intervention variables. All data analyses were performed using IBM's SPSS modeler 18.0 software. The SPSS modeler performed data preparation before primary analysis by replacing null values with the mean for continuous fields followed by z-transformation re-scaling. It increases the performance and accuracy of the model through machine learning and artificial intelligence techniques [14].

The data was then separated into two sections of training (80%) and testing (20%) which provided an infrastructure for the next steps. Due to the large number of model predictors, using the training and testing dataset helps to track overfitting. This is done based on the fitness criteria of the testing dataset. Additionally, before the principal analysis, a feature selection or variable selection was performed to choose the most relevant model predictors [21]. Next, multivariable and univariable logistic regressions were executed to identify the risk factors of COVID-19. In final stage of the analysis, a decision model was developed to identify high-risk patients.

This study was approved in the ethics committee of the Shahid Beheshti University of Medical Sciences (IR.SBMU.RETECH.REC.1399.007).

Data collection

Data was collected from health records of 4542 hospitalized adult patients (above 18 years old) with severe symptoms of COVID-19 in 15 hospitals in Tehran, Iran. Patients were hospitalized in the period of 20 February 2020 to 18 April 202. Their data was routinely registered on Hospital Information System (HIS) and the centralized access to them is possible through Hospital Intelligent Management system (HIM). Patients' clinical features including fever (above 38⁰ C), shortness of breathing, hypoxia, chest pain, severe cough, and loss of consciousness. Patients were divided into two groups of recovered and discharged, and deceased. The definitive diagnosis of COVID-19 was made according to Iranian Ministry of Health protocol using throat-swab specimens and the RT-PCR test. Available data included age, sex, inpatient wards, supportive interventions of oxygen therapy and mechanical ventilation, duration of hospitalization, and on admission laboratory test results. These laboratory tests were:

- 1) Complete Blood Count (CBC) tests including red blood cells (RBC) count, white blood cells (WBC) count, hemoglobin, hematocrit, platelets count, mean corpuscular volume (M.C.V), mean corpuscular hemoglobin (M.C.H), mean corpuscular hemoglobin concentration (M.C.H.C), red cell distribution width (RDW), mean platelet volume (MPV), Lymphocytes and Neutrophils count, MIX (the ratio of eosinophil, basophil, and monocyte to total white blood cells).
- 2) Liver function enzymes including ALT, AST, alkaline phosphatase (ALP).
- 3) Heart and muscles injury enzymes including troponin, creatine kinase myocardial band (CK-MB), creatine phosphokinase (CPK).
- 4) Arterial blood gases (ABG) including PO₂, O₂Sat, PCO₂, HCO₃, blood PH, base excess (BE).
- 5) Other tests: creatinine, uric acid, serum Ca, serum K, serum Na, 25-hydroxyvitamin D, fasting blood sugar (FBS), D-Dimer, CRP.

Statistics

Three main steps were taken to perform data analysis. These steps were as follows:

STEP 1) A Multivariable Logistic Regression (MLR) for calculating the odds ratio (OR) of risk factors

A Multivariable Logistic Regression (MLR) was performed to predict any relationship between predictors and the survival outcome (recovered or deceased).

STEP 2) A univariable logistic regression for calculating the relative risk (RR) of the risk factors

Predictors with a significant association (p -value < .05) with the MLR outcome were selected and transformed into a set of appropriate categorical variables. This step is done to identify the relationship between different levels (normal or abnormal levels) of laboratory tests and mortality rate due to COVID-19, and calculating RR by univariable logistic regression. Accordingly, some of the laboratory tests were divided into two or three categories (normal range, above, and below the normal range), and some were divided into more than three categories for their extra-outlier values.

STEP 3) Identifying high-risk patients using a decision tree model

The predictors were entered into the Chi-square automatic interaction detector (CHAID) model to develop a decision tree model and identify the high-risk and low-risk patients.

Results

Descriptive Statistics

From 20 February to 18 April 2020, 4791 adult patients with definitive diagnosis of COVID-19 were admitted to 15 university hospitals in Tehran, Iran, 249 patients were excluded due to their missing data. Of the 4542 hospitalized patients, who were between 18 and 97 years of age and had severe symptoms, 822 (18.09%) died during hospitalization, and 3720 (81.90%) patients recovered and discharged. The mean±SD and median [interquartile range] age of patients were, respectively, 55.55±16.84 and 56.00 [27.00] for the survived group, and 68.72±14.93 and 71.00 [21.25] for the deceased. 2473 (54.44%) patients were male. The demographic characteristics of the patients are displayed in Table 1.

The recruited patients had severe symptoms such as fever above 38⁰ C, severe cough, shortness of breath, respiratory rate > 30 time/minute, hypoxia (PO2 < 93%), chest pain, and loss of consciousness. Criteria for discharge from the hospital was 72 hours without a fever and no need for an antipyretic drug, PO2 > 93%, and improvement in the clinical, and respiratory symptoms.

The mean±SD of hospitalization days for the survived and deceased groups were 4.76±4.77 and 6.45±7.21 days, respectively (Table 1). The treatment program was identical in all hospitals, and according to the COVID-19 treatment guidelines of Iran’s Ministry of Health, unless there were concurrent diseases requiring specific treatments.

Table 1: Demographic and hospitalization characteristics of the COVID-19 patients

As mentioned in the method section, before entering all the independent variables in the MLR model, a feature selection algorithm was used to define the most important predictors associated with the target variable. According to the feature selection results, apart from RBC, platelets, hemoglobin, hematocrit, PDW, and PCO2, all laboratory tests, age, gender, and the need for special care such as oxygen therapy and ventilation for survived patients entered the model as potential variables correlated with disease prognosis (Table 2).

Table 2: Predictors statistical characteristics in patients with COVID-19

MLR results

In this step, the SPSS modeler ignored predictors with a lot of missing value. Thus the amount of the missing value in the datasets reached under 15%. Accuracy and AUC index of the testing dataset as fitness criteria was 85.87% and 85.00, respectively. So the model has been fitted well, without overfitting problems [14].

As shown in Table 3 and based on the MLR model, age, ventilation, oxygen therapy, CRP, AST, ALT, ALP, CPK, serum calcium, serum sodium, serum potassium, and creatinine, have a significant relationship with the survival of COVID-19 patients.

According to the results, the group in which ventilation was required had a higher mortality rate (OR=.45, 95% CI: .38-.52). From the 222 patients that received this intervention, only 16.67% (37 people) survived. Regarding oxygen therapy, the results are the same (OR=.93, 95% CI: .88-.98), with a weaker correlation (Beta = -.08). Regarding laboratory test predictors, the results are as follows:

Negative association of levels of CRP (OR=.77, 95% CI: .71-.84), AST (OR=.54, 95% CI: .45-.64), ALP (OR=.79, 95% CI: .71-.87), CPK (OR=.78, 95% CI: .71-.86), sodium (OR=.82, 95% CI: .74-.91), potassium (OR=.87, 95% CI: .78-.98) and creatinine (OR=.92, 95% CI: .87-.99) as well as a positive relationship with ALT (OR=1.23, 95% CI: 1.07-1.42), and calcium (OR=1.38, 95% CI: 1.24-1.54), have been shown in the MLR model for surviving patients.

Table 3: Multiple logistic regression results (target variable is the final outcome (survived or deceased))

ULR results

In the following stage, significant predictors were altered into categorical variables in order to enter them in the ULR model and calculate the Relative Risk. Thresholds used for categorization of these predictors were their standard ranges. This transformation enabled us to identify the critical values of laboratory tests, attributed to the patients' survival status (See appendix 1).

According to the ULR results, the age-related RR increased in older ages. The results showed that ages 40-49, 50-59, 60-69, 70-79 and above 80 are associated with elevated RR levels of 1.80 (CI: 1.13-2.87), 2.63 (CI: 1.71-4.02), 4.40 (CI: 2.92-6.63), 7.49 (CI: 5.01-11.19), and 13.85 (CI: 9.23-2.77) respectively. So, age is an essential COVID-19 risk factor that is not to be overlooked. Other significant risk factors are related to liver injury. These include ALT higher than the reference range 55-100 (RR=2.20, CI: 1.69-2.86) and above 100 (RR=2.67, CI = 2.86-5.34), AST range 40-100 IU/L (RR=2.51, CI: 2.04-3.10) and above 100 IU/L (RR=5.93, CI: 4.75-7.39), and ALP higher than 200 IU/L (RR=2.46, CI: 1.80-3.37). Serum electrolytes included sodium under 135 mEq/l (RR=1.69, CI: 1.35-2.11), and over 145 mEq/l (RR=7.24, CI: 5.07-1.33), potassium above 5.50 mEq/l (RR=7.53, CI: 4.15-13.64), and calcium under 8.50 mEq/l (RR=3.39, CI: 2.81-4.09) are related to increased mortality rate due to COVID-19. Also, elevated CPK range 307-600 IU/L

(male) and 192-400 IU/L (female) (RR=2.73, CI: 2.12-3.53) and above 600 IU/L (male), and above 400 (female) (RR=4.41, CI: 3.40-5.71), CRP between 8-100 mg/l (RR=1.40, CI: .94-2.08) and more than 100 mg/l are (RR=3.22, CI: 1.99-5.20), and creatinine more than 1.5 (RR=6.37, CI: 5.30-7.66) were associated with COVID-19 mortality rate (MR) (Table 4).

Table 4: RR related to different ranges of the studied predictors

CHAID model: Identifying high risk patients

Based on the MLR analysis, of the 35 laboratory test parameters that were analyzed, 9 had a significant relationship with patients' survival rate. These 9 parameters were then converted into categorical variables and entered into the CHAID node to develop a decision tree model.

According to the CHAID model, the patients were classified into five categories based their risk factors: low risk (under 20% MR), moderate risk 1 (20-30% MR), moderate risk 2 (30-40% MR), high risk (40-50% MR), and very high risk (more than 50% MR). These risk factors were separately defined for each age group (Figure 1).

In the ages under 40, sodium, CPK, AST, and creatinine (accuracy=96.25%, AUC=.63), in the age group of 40-59, AST, calcium, creatinine, and sodium (accuracy=88.57%, AUC=.75), within 60-69 year-old patients, AST, creatinine, CPK, ALP, and CRP (accuracy=82.70%, AUC=.71), in 70-79 age group, AST, creatinine, calcium, ALP, and CPK (accuracy=8.00%, AUC=.83), and for those above 80, creatinine, AST and sodium (accuracy=62.16%, AUC=.63) were associated with high-risk COVID-19 patients.

In the age under 40 (4.64% MR), with normal creatinine, an increase of $AST \geq 100$ IU/L, with $CPK \geq 307$ IU/L in male patients and $CPK \geq 192$ IU/L in female patients were associated with 41.67% MR. So, this group is considered high-risk patients. Also, in this age group, we see 75.00% MR correlated with elevated creatinine levels ≥ 1.5 mg/l and an abnormal sodium level.

In the age group 40-59 (11.84% MR), high-risk group patients have been seen:

Patients with creatinine ≥ 1.5 mg/l that have 41.52% MR, patients with hypocalcemia, and $AST \geq 100$ IU/L that have 56.08% MR, patients with creatinine levels ≥ 1.5 mg/l and hypocalcemia that have 46.67% MR, and patients with creatinine levels of ≥ 1.5 mg/l and $AST \geq 40$ IU/L that have 59.55% MR, hence regarded as high-risk COVID-19 patients.

In the age group 60-69 (18.69% MR), elevated $AST \geq 100$ IU/L with $CPK \geq 307$ IU/L in male patients and $CPK \geq 192$ IU/L in female patients were associated with 67.57% MR.

In the age group 70-79 (31.47% MR), six groups of high-risk patients have been seen:

Male patients with $CPK \geq 307$ IU/L and female patients with $CPK \geq 192$ IU/L have 5.00% MR.

In patients with AST level 40-99 IU/L and elevated ALP ≥ 200 IU/L, MR is 48.75%, and with elevated CPK ≥ 600 IU/L (male) and CPK ≥ 400 IU/L (female), is 61.54%. Also, in this age group, there is 52.63% MR when AST ≥ 100 IU/L, 67.19% MR when AST ≥ 100 IU/L and creatinine ≥ 1.5 mg/l, and 66.67% MR when AST ≥ 100 IU/L and there is hypocalcemia too.

In the age above 80 years (4.40% MR), MR, due to COVID-19, is 49.09% when AST ≥ 100 IU/L, 59.90% when creatinine ≥ 1.5 mg/l (43.59% when AST < 40 IU/L, and 71.56% when AST ≥ 40 IU/L), and 79.59% when creatinine ≥ 1.5 mg/l and there are abnormal sodium levels too. More information on this is presented in Figure 1.

Discussion

This study is the first attempt to investigate the risk factors of COVID-19 in 4542 adult patients in Iran. Using data mining methods, we identified the relationship between a large number of laboratory tests and the risk factors without any overfitting problems. According to the results of MLR, and ULR analysis, AST, ALP, ALT, calcium, sodium, potassium, creatinine, CPK, and CRP were correlated with the risk of death in COVID-19 patients. Furthermore, in the decision tree model, high risk patients were identified.

Age

In this study, we found a strong relationship between age and COVID-19 Mortality Rate. Results of the present study confirm previous studies that older age is an important risk factor of death in patients with COVID-19 [27]. It can be assumed that age-dependent defects in humoral and cellular immune function could lead to a decrease in number of T cells, phagocytosis and levels of interferon [25].

Liver enzymes

According to the results, elevated AST, ALT, and ALP of above the normal range were associated with increased risk of death in COVID-19 patients in all age groups. Liver injury and subsequently abnormal levels of AST, ALT, and ALP are prevalent in patients with COVID-19 during disease progression and longer hospital stays [6,11]. A serum marker of mitochondrial damage in hepatocytes is high levels of AST. In severe cases of patients with COVID-19, liver damage is associated with liver hypoxia/reperfusion injury and this could induce mitochondrial apoptosis, triggering cell injury, necrosis, and finally elevated AST levels [2]. Abnormal levels of liver enzymes can be caused by viral infections such as COVID-19 [35], consumption of hepatotoxic drug or by an immune-mediated inflammatory response during treatment. For this reason, more intensive care and specific therapeutic approaches are needed for severe patients with COVID-19 who have pre-existing liver diseases, and more especially, older patients [36].

Serum electrolytes

Abnormal serum electrolytes (calcium, potassium, and sodium) were also among the risk factors associated with COVID-19 MR. Concerning serum calcium, the herein results are consistent with previous reports on COVID-19 which have stated that patients with hypocalcemia have more severe symptoms and

mortality rates than others [28]. This is attributable to the role of calcium in immune response development [12]. It has been demonstrated that calcium homeostasis alteration within the cell could promote the activation of inflammatory pathways leading to increased levels of IL-1b, TNF and IL-6, linked to lung cell damage and edema accumulation [22]. On the other hand, hypocalcemia is associated with higher incidence of organ injury, hyperproteinemia, unbalanced vitamin D and parathyroid hormone. Therefore, early detection and correction of Ca levels of patients with hypocalcemia may lead to better outcomes as well as a reduction in the severity of symptoms [10].

Moreover, the results showed that both hyponatremia and hypernatremia were correlated with high COVID-19 mortality rate. Lippi *et al.* reported that in patients with COVID-19, serum electrolytes levels including sodium, is decreased [19]. In viral infections, low sodium levels may cause tissue damage and increase the severity of the viral disease [3]. Previous studies show that sodium is an essential factor regulating the expression of angiotensin-converting enzyme-2 (ACE2) in the body. Also, COVID-19 enters the host cell by connecting to this receptor, and hyponatremia may cause an overexpression of the ACE2 receptor in the long-term. Consequently, hyponatremia may cause more severe conditions of COVID-19 [20]. In addition, one of the clinical symptoms of COVID-19 is diarrhea and water loss [13], which can cause hypernatremia in the body [9]. It has seen that hypernatremia and hyponatremia had some adverse effects on various organ functions such as the central nervous system and, in turn, an increased mortality rate [16].

High-level serum potassium is another risk factor which is found to be associated with COVID-19 mortality rate. Hyperkalemia may be a result of other underlying diseases such as kidney dysfunction, and can lead to cardiac arrest. Evidence also suggests that immunodeficiency against viral infections could lead to hyperkalemia [23]. However, hyperkalemia, for whatever reason, is dangerous in itself and must be carefully monitored.

These findings emphasize electrolyte monitoring of patients with COVID-19 and their role in appropriate management of severely ill patients.

CPK

Elevated CPK is another risk factor that is shown to be associated with a high mortality rate in the patients. Increased CPK is caused by damage to muscles (mainly skeletal muscles) [8]. Previous studies show that patients with influenza type A have increased CPK levels due to skeletal muscle involvement [15]. Also, it has been reported that rhabdomyolysis and elevated CPK can occur in COVID-19 patients, too [29]. The pathogenesis of viral-induced rhabdomyolysis has several possible mechanisms: i) direct viral invasion; ii) cytokine storms and damaging muscle tissue following immune response; and iii) destroying muscle cell membrane due to direct action of viral toxins. The mechanism of COVID-19-induced rhabdomyolysis, however, has not yet been understood [1]. The only known fact is that elevated CPK $\geq$ 150 IU/L is a common finding in patients with COVID-19 [38]. Therefore, increased CPK which may be due to muscle involvement in the severe stages of disease, becomes a predictor for higher mortality rate in patients with COVID-19.

CRP

Finally, the results show that elevated CRP of above 8 mg/dl, is associated with an increase in COVID-19 mortality rate. Hence, and in parallel with previous studies, elevated CRP, as an inflammatory biomarker, could indicate lung damage and therefore represents the severity of the respiratory tract involvement in patients with COVID-19 [31].

Creatinine

Aligned with the results of this study, recent reports have shown that an elevated level of creatinine in patients with COVID-19 is seen and associated with mortality rate. These results demonstrated that creatinine levels can be a predictor of renal impairment, one of the most prominent causes of death among patients [32]. Furthermore, Cheng *et al.* demonstrated that a high level of creatinine increases the probability of admission to the intensive care unit, and these patients have a higher risk of deterioration [7]. ACE2 receptors expression in human kidneys and bladder could be a potential binding route of Coronavirus [18]. Since researchers have isolated Coronavirus from urine samples of some patients, the association of kidney impairment with COVID-19 appears to be possible [7].

In this study, the risk factors of COVID-19 were analyzed through logistic regression, and final predictors were identified displaying robust results. According to the results, abnormal levels of some laboratory tests appeared in patients with COVID-19, which can be due to non-respiratory multi-organ involvements seen during the course of COVID-19. These include the liver, heart, gastrointestinal tract and kidney [6]. These abnormalities are not specific to the novel Coronavirus infection [26], nevertheless prompt action for their management as potential risk-factors for COVID-19 severity may reduce its morbidity and mortality rate.

Considering the fact that there is currently no approved treatment for COVID-19, our findings support the idea that in patients with COVID-19, identifying high-risk patients and providing supportive care to prevent organ damage and homeostasis abnormalities have a crucial role in the final outcome of the therapeutic interventions.

Limitations

As a retrospective cohort study, this research had some inherent limitations. Some patients had incomplete data in their medical records. Also, slightly different in-hospital procedures led to a lack of order or records of some laboratory tests. A large sample size of the study, however, may overcome these drawbacks.

Declarations

Funding: This study has no funder.

Conflicts of interest/Competing interests: The authors declare no conflict of interest in this study.

Ethics approval: This study was approved in the ethics committee of the Shahid Beheshti University of Medical Sciences (IR.SBMU.RETECH.REC.1399.007).

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: All data and materials were used in this study are available.

Code availability: This is a cohort study without trial registration number.

Authors' contributions: A.M. contributed to the conception and design of the study, the analysis and drafting the article. S.R. contributed to the analysis, interpretation and drafting the article. J.S. and A.Mi. contributed to the interpretation and critical revision of the article. N.Y. contributed to the conception, data acquisition, and provided critical revision of the article. F.P. contributed to data acquisition, and provided a critical revision of the article.

Acknowledgments

The authors would like to thank Mohammad Hossein Shoushatri for his guidance in data mining, and comments that greatly improved the manuscript. They would also like to acknowledge Mahtab Moradi and Elham Salehian for their help with data acquisition. The authors would also like to express their gratitude to Sajad Kamali for data preparation, and to Zahra Ghalkhani for her expert advice on laboratory sciences.

References

1. Ayala, E., Kagawa, F.T., and Wehner, J.H. Rhabdomyolysis Associated With 2009 Influenza A(H1N1) To. *JAMA*, 302, 17 (2009), 1863–4.
2. Cannistrà, M., Ruggiero, M., Zullo, A., et al. Hepatic ischemia reperfusion injury: A systematic review of literature and the role of current drugs and biomarkers. *International Journal of Surgery*, 33, (2016), S57–S70.
3. Cao, Y., Deng, Q., and Dai, S. Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information. January (2020).
4. Castro, M.C., Carvalho, L.R. de, Chin, T., et al. Demand for hospitalization services for COVID-19 patients in Brazil. *medRxiv*, (2020), 2020.03.30.20047662.
5. Chen, G., Wu, D., Guo, W., et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *Journal of Clinical Investigation*, 130, 5 (2020), 2620–2629.
6. Chen, N., Zhou, M., Dong, X., et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, 395, 10223 (2020), 507–513.

7. Cheng, Y., Luo, R., Wang, K., et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney International*, 97, 5 (2020), 829–838.
8. Do, T.T. What is the role of creatine phosphokinase (CPK) determination in the evaluation of muscular dystrophy? <https://emedicine.medscape.com/>, 2018.
<https://www.medscape.com/answers/1259041-96568/what-is-the-role-of-creatine-phosphokinase-cpk-determination-in-the-evaluation-of-muscular-dystrophy>.
9. Drogué, H.J. and Adias, N.E. Hyponatremia. *The New England Journal of Medicine*, 342, (2000), 1493–9.
10. El-kurdi, B., Khatua, B., Rood, C., et al. Mortality from covid-19 increases with unsaturated fat, and may be reduced by early calcium and albumin supplementation. *Gastroenterology*, (2020).
11. Fan, Z., Chen, L., Li, J., et al. Clinical Features of COVID-19-Related Liver Damage. *Clinical Gastroenterology and Hepatology*, 18, 7 (2020), 1561–1566.
12. Grinstein, S. and Klip, A. Calcium homeostasis and the activation of calcium channels in cells of the immune system. *Bulletin of the New York Academy of Medicine: Journal of Urban Health*, 65, 1 (1989), 69–79.
13. Huang, C., Wang, Y., Li, X., et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395, 10223 (2020), 497–506.
14. IBM. IBM SPSS Modeler 18.1.1 Modeler Modeling Nodes. (2017), 422.
15. Kessler, H.A., Trenholme, G.M., Vogelzang, N.J., et al. Elevated Creatine Phosphokinase Levels Associated with Influenza A / Texas / 1 / 77 Infection Elevated Creatine Phosphokinase Levels Associated with Influenza A / Texas / 1 / 77 Infection. 5548, (2015), 3–7.
16. Kovesdy, C.P., Lott, E.H., Lu, J.L., et al. Hyponatremia, hypernatremia, and mortality in patients with chronic kidney disease with and without congestive heart failure. *Circulation*, 125, 5 (2012), 677–684.
17. Lillie, P.J., Samson, A., Li, A., et al. Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. *Journal of Infection*, 80, 5 (2020), 578–606.
18. Lin, W., Hu, L., Zhang, Y., et al. Single-cell Analysis of ACE2 Expression in Human Kidneys and Bladders Reveals a Potential Route of 2019-nCoV Infection. *bioRxiv*, (2020), 2020.02.08.939892.
19. Lippi, G., South, A.M., and Henry, B.M. ANNALS EXPRESS: Electrolyte Imbalances in Patients with Severe Coronavirus Disease 2019 (COVID-19). *Annals of clinical biochemistry*, 57, 3 (2020), 262–265.
20. Luo, Y., Li, Y., and Dai, J. Low blood sodium increases risk and severity of COVID-19 : a systematic review , meta-analysis and retrospective cohort study. *BMJ*, (2020).
21. Maniruzzaman, M., Rahman, M.J., Ahammed, B., and Abedin, M.M. Classification and prediction of diabetes disease using machine learning paradigm. *Health Information Science and Systems*, 8, 1 (2020), 1–14.

22. Nieto-Torres, J.L., Verdiá-Báguena, C., Jimenez-Guardeño, J.M., et al. Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology*, 485, (2015), 330–339.
23. Parham, W.A., Mehdirad, A.A., Biermann, K.M., and Fredman, C.S. Hyperkalemia revisited. *Texas Heart Institute Journal*, 33, 1 (2006), 40–47.
24. Sasangohar, F., Jones, S.L., Masud, F.N., Vahidy, F.S., and Kash, B.A. Provider Burnout and Fatigue During the COVID-19 Pandemic: Lessons Learned from a High-Volume Intensive Care Unit. *Anesthesia and analgesia*, 1, Xxx (2020), 1–6.
25. Sharma, G. and Goodwin, J. Effect of aging on respiratory system physiology and immunology. *Clinical interventions in aging*, 1, 3 (2006), 253–260.
26. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian Journal of Pediatrics*, 87, 4 (2020), 281–286.
27. Sun, H., Ning, R., Tao, Y., et al. Risk Factors for Mortality in 244 Older Adults With COVID-19 in Wuhan, China: A Retrospective Study. *Journal of the American Geriatrics Society*, (2020), 1–5.
28. Sun, J.-K., Zhang, W.-H., Zou, L., et al. Serum calcium as a biomarker of clinical severity and prognosis in patients with coronavirus disease 2019: a retrospective cross-sectional study. *Critical Care & Emergency Medicine*, (2020), 1–18.
29. Suwanwongse, K. and Shabarek, N. Rhabdomyolysis as a Presentation of 2019 Novel Coronavirus Disease. *Cureus*, 12, 4 (2020), 10–13.
30. Takian, A., Raoofi, A., and Kazempour-Ardebili, S. COVID-19 battle during the toughest sanctions against Iran. *The Lancet*, 395, 10229 (2020), 1035–1036.
31. Wang, L. C-reactive protein levels in the early stage of COVID-19. *Medecine et Maladies Infectieuses*, 50, (2020), 332–334.
32. Wang, L., Li, X., Chen, H., et al. Coronavirus Disease 19 Infection Does Not Result in Acute Kidney Injury: An Analysis of 116 Hospitalized Patients from Wuhan, China. *American Journal of Nephrology*, 430060, (2020), 343–348.
33. Wu, C., Chen, X., Cai, Y., et al. Risk Factors Associated with Acute Respiratory Distress Syndrome and Death in Patients with Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Internal Medicine*, (2020), 1–10.
34. Wu, J.T., Leung, K., Bushman, M., et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine*, 26, 4 (2020), 506–510.
35. Xu, L., Liu, J., Lu, M., Yang, D., and Zheng, X. Liver injury during highly pathogenic human coronavirus infections. *Liver International*, February (2020), 998–1004.
36. Zhang, C., Shi, L., and Wang, F.S. Liver injury in COVID-19: management and challenges. *The Lancet Gastroenterology and Hepatology*, 5, 5 (2020), 428–430.
37. 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. *The Lancet*, 6736, 20 (2020), 1–9.

38. Laboratory features associated with severe COVID-19. *www.uptodate.com*, 2020.

[https://www.uptodate.com/contents/image/print?](https://www.uptodate.com/contents/image/print?imageKey=ID%2F127820&topicKey=ID%2F127429&source=see_link)

[imageKey=ID%2F127820&topicKey=ID%2F127429&source=see_link](https://www.uptodate.com/contents/image/print?imageKey=ID%2F127820&topicKey=ID%2F127429&source=see_link).

Tables

Table 1: Demographic and hospitalization characteristics of the COVID-19 patients

	Total N=4542)	Discharged from hospital (N=3720)	Died in hospital (N=822)
Male	2676 (58.92%)	2169 (81.05%)	507 (18.94%)
Age groups			
18-40	785 (17.28%)	723 (92.10%)	34 (4.33%)
40-50	705 (15.52%)	671 (95.18%)	59 (8.37%)
50-60	885 (19.48%)	767 (86.67%)	118 (13.33%)
60-70	843 (18.56%)	689 (81.73%)	154 (18.27%)
70-80	760 (16.73%)	535 (70.39%)	225 (29.60%)
Above 80	564 (12.42%)	333 (59.04%)	231 (40.96%)
MCV (mean + SD)	57.94±17.27	55.55±16.84	68.72±14.93
Inpatient ward			
ICU/CCU	1029 (22.65%)	517 (50.24%)	512 (49.76%)
Others	3513 (77.34%)	3203 (91.17%)	310 (88.24%)

Table 2: Predictors statistical characteristics in patients with COVID-19

	Predictors	Target Groups	Mean	SD	Importance	Value of importance
	CBC					
1	MCV	Died	87.14	7.42	Important	1.00
		Survived	86.15	5.91		
2	MCH	Died	31.02	2.20	Important	1.00
		Survived	31.40	1.75		
3	MCHC	Died	33.13	1.60	Important	1.00
		Survived	33.73	1.35		
4	MIX	Died	4.97	2.66	Important	1.00
		Survived	6.49	3.10		
5	WBC	Died	11.34	6.52	Important	1.00
		Survived	7.39	3.77		
6	RBC	Died	4.19	0.85	Unimportant	0.85
		Survived	4.26	0.63		
7	Platelets	Died	196.97	82.38	Unimportant	0.75
		Survived	207.82	82.20		
8	Hemoglobin	Died	12.06	2.46	Unimportant	0.47
		Survived	12.33	1.80		
9	Hematocrit	Died	36.33	6.91	Unimportant	0.18
		Survived	36.53	4.94		
10	RDW-CV	Died	15.14	2.35	Important	1.00
		Survived	14.11	2.52		
11	PDW	Died	13.40	2.63	Unimportant	0.76
		Survived	13.53	2.41		
12	MPV	Died	9.78	1.06	Important	1.00
		Survived	9.32	1.09		
13	Neutrophils	Died	82.64	9.08	Important	1.00
		Survived	72.58	10.98		
14	Lymphocyte	Died	11.84	8.74	Important	1.00
		Survived	21.02	10.15		

	LFT					
15	AST	Died	126.32	188.05	Important	1.00
		Survived	64.26	71.75		
16	ALT	Died	73.28	155.15	Important	1.00
		Survived	41.94	72.75		
17	ALP	Died	248.74	189.23	Important	1.00
		Survived	196.83	139.49		
	Muscles injury related enzymes					
18	CPK	Died	500.33	1099.24	Important	1.00
		Survived	239.33	547.87		
19	CKMB	Died	45.61	62.64	Important	1.00
		Survived	23.56	30.91		
20	Troponin	Died	2.33	12.30	Important	1.00
		Survived	0.62	4.54		
	ABG					
21	PCO2	Died	43.97	14.54	Unimportant	0.80
		Survived	43.96	10.58		
22	PO2	Died	47.19	24.47	Important	1.00
		Survived	38.23	17.46		
23	Blood PH	Died	7.35	0.12	Important	1.00
		Survived	7.40	0.20		
24	HCO3	Died	25.71	5.23	Important	1.00
		Survived	28.29	4.41		
25	O2Sat	Died	63.67	18.72	Important	1.00
		Survived	55.84	18.43		
26	BE	Died	-2.30	6.25	Important	1.00
		Survived	1.96	4.24		
	Other test results					
27	VitaminD3	Died	35.68	27.24	Marginal	0.913

		Survived	30.06	18.19		
28	Serum Ca	Died	8.56	0.97	Important	1.00
		Survived	8.85	0.79		
29	Serum K	Died	4.29	0.64	Important	1.00
		Survived	4.07	0.89		
30	Serum Na	Died	138.79	5.47	Important	1.00
		Survived	138.04	4.22		
31	D-Dimer	Died	3026.40	3377.46	Important	1.00
		Survived	1729.43	3041.07		
32	CRP	Died	65.80	52.98	Important	1.00
		Survived	44.23	42.83		
33	Creatinine	Died	2.11	1.81	Important	0.99
		Survived	1.41	5.35		
34	FBS	Died	160.92	72.76	Important	1.00
		Survived	135.59	56.40		
35	Uric Acid	Died	8.06	3.46	Important	1.00
		Survived	5.60	2.25		
	Interventions					
36	Ventilation	Died	1.49	5.86	Important	1.00
		Survived	0.03	0.49		
37	Oxygen therapy	Died	0.82	2.01	Important	1.00
		Survived	0.54	1.41		

Table 3: Multiple logistic regression results (target variable is the final outcome (survived or deceased))

	Parameter Estimates						
Group		Beta coefficient	Std. Error	P-value	Odds ratio	95% Confidence Interval for odd ratio	
Survived						Lower Bound	Upper Bound
	Intercept	4.75		0.24	0.00		
	Age	-0.05	0.00	0.00	0.95	0.95	0.96
	[sex=1.000]	-0.01	0.11	0.93	0.99	0.81	1.22
	[sex=2.000]	0 ^b					
	Ventilation	-0.81	0.08	0.00	0.45	0.38	0.52
	Oxygen therapy	-0.08	0.03	0.00	0.93	0.88	0.98
	CRP	-0.26	0.05	0.00	0.77	0.71	0.84
	Creatinine	-0.08	0.03	0.02	0.92	0.87	0.99
	AST	-0.62	0.09	0.00	0.54	0.45	0.64
	ALT	0.21	0.07	0.00	1.23	1.07	1.42
	ALP	-0.24	0.05	0.00	0.79	0.71	0.87
	CPK	-0.24	0.05	0.00	0.78	0.71	0.86
	Serum Calcium	0.32	0.06	0.00	1.38	1.24	1.54
	Serum Potassium	-0.14	0.06	0.02	0.87	0.78	0.98
	Serum Sodium	-0.20	0.06	0.00	0.82	0.74	0.91
		a The reference category is: dead.					
		b This parameter is set to zero because it is redundant.					

Table 4: RR related to different ranges of the studied predictors

	Predictors	P-value	RR with 95% CI
1	age (years)		
	under 40		1
	40-49	0.01	1.80 (CI: 1.13-2.87)
	50-59	0.00	2.63 (CI: 1.71-4.02)
	60-69	0.00	4.40 (CI: 2.92-6.63)
	70-79	0.00	7.49 (CI: 5.01-11.19)
	Above 80	0.00	13.85 (CI: 9.23-20.77)
2	sex		
	male	0.06	1.18 (CI: 0.99-1.41)
	female		1
3	ALT (IU/L)		
	Up to 55		1
	55 -100	0.00	2.20 (CI: 1.69-2.86)
	Above 100	0.00	2.67 (CI: 2.86-5.34)
4	AST (IU/L)		
	Up to 40		1
	40-100	0.00	2.51 (CI: 2.04-3.10)
	Above 100	0.00	5.93 (CI: 4.75-7.39)
5	ALP (IU/L)		
	40-129		1
	129-200	0.58	0.91 (CI: 0.68-1.24)
	Above 200	0.00	2.46 (CI: 1.80-3.37)
6	CPK (IU/L)		
	Under 39 (male), under 26 (female)	0.21	0.55 (CI: 0.22-1.39)
	39-307 (male), 26-191 (female)		1
	307-600 (male), 192-400 (female)	0.00	2.73 (CI: 2.12-3.53)
	Above 600 (male), Above 400 (female)	0.00	4.41 (CI: 3.40-5.71)
7	Serum Na (mEq/l)		
	Under 135	0.00	1.69 (CI: 1.35-2.11)
	135 - 145		1

	Above 145	0.00	7.24 (CI: 5.07-10.33)
8	Serum Ca (mEq/l)		
	Under 8.50	0.00	3.39 (CI: 2.81-4.09)
	Above 8.50		1
9	Serum K (mEq/l)		
	Under 3.50	0.40	0.85 (CI: 0.59-1.23)
	3.50 - 5.49		1
	Above 5.50	0.00	7.53 (CI: 4.15-13.64)
10	Creatinine (mg/l)		
	Under 0.5		
	0.5-1.5		1
	Above 1.5	0.00	6.37 (CI: 5.30-7.66)
11	CRP (mg/l)		
	Under 8		1
	8 - 100	0.10	1.40 (CI: 0.94-2.08)
	Above 100	0.00	3.22 (CI: 1.99-5.20)

Figures

Figure 1

The results of decision tree model (CHAID model) in patients under 70 (A), and over 70 (B)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementfile.docx](#)