

Predictors of Mortality in Hospitalized Patients with COVID-19: A One-Year Case-Control Study

Running title: COVID-19 and Prediction of Mortality

Laura Camacho-Domínguez^{1,2}, Manuel Rojas¹, María Herrán¹, Yhojan Rodríguez^{1,2}, Santiago Beltrán¹, Paola Saboya Galindo¹, Nicolas Aguirre-Correal¹, María Espitia¹, Santiago García¹, Valeria Bejarano¹, Victoria Morales-González¹, Jaime Enrique Covalada-Vargas¹, Mónica Rodríguez-Jiménez¹, Elizabeth Zapata^{1,2}, Diana M. Monsalve¹, Yeny Acosta-Ampudia¹, Juan-Manuel Anaya³, Carolina Ramírez-Santana^{1*}

¹Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Bogota, Colombia.

²Clínica del Occidente, Bogota, Colombia.

³ Health Research and Innovation Center at Coosalud, Cartagena 130001, Colombia.

Corresponding author

Carolina Ramírez-Santana, Ph.D.*

Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario.

Carrera 24 # 63-C- 69, 110010, Bogota, Colombia.

E-mail address: heily.ramirez@urosario.edu.co

Abstract (281 words)

Objective: To determine the associated factors with mortality, in addition to age and sex, in a high-complexity hospital in Bogota, Colombia, during the first year of the pandemic.

Design: A case-control study.

Setting: High-complexity center above 2,640 meters above sea level (masl) in Colombia.

Methods: A case-control study was conducted on 564 patients admitted to the hospital with confirmed COVID-19. Deceased patients (n: 282) and a control group (n: 282), matched by age, sex, and month of admission, were included. Clinical and paraclinical variables were retrospectively obtained by systematic revision of clinical records. Multiple imputations by chained equation (MICE) were implemented to account for missing variables. Classification and regression trees (CART) were estimated to evaluate the interaction of associated factors on admission and their role in predicting mortality during hospitalization.

Results: Most of the patients included were males in the seventh decade of life. Most of the admissions occurred between July and August 2021. Surprisingly, recovered patients reported heterogeneous symptomatology, whereas deceased patients were most likely to present respiratory distress, dyspnea, and seizures on admission. In addition, the latter group exhibited a higher burden of comorbidities and alterations in laboratory parameters. After the imputation of datasets, CART analysis estimated 14 clinical profiles based on respiratory distress, LDH, dyspnea, hemoglobin, D-dimer, ferritin, blood urea nitrogen, C-reactive protein, $\text{PaO}_2/\text{FiO}_2$,

dysgeusia, total bilirubin, platelets, and gastroesophageal reflux disease. The accuracy model for prediction was 85.6% ($P < 0.0001$).

Conclusion: Multivariate analysis yielded a reliable model to predict mortality in COVID-19. This analysis revealed new interactions between clinical and paraclinical features in addition to age and sex. Furthermore, this predictive model could offer new clues for the personalized management of this condition in clinical settings.

Keywords: SARS-CoV-2, COVID-19, Mortality, Predictors, Risk Factors.

70

71 **List of Abbreviations**

72 ARBs: Angiotensin II receptor blockers.

73 ARDS: Acute respiratory distress syndrome.

74 BUN: Blood urea nitrogen.

75 CART: Classification and regression trees.

76 COVID-19: Coronavirus disease 2019.

77 CRP: C-reactive protein.

78 FiO_2 : Fraction of inspired oxygen.

79 GERD: Gastroesophageal reflux disease

80 HIV: Human immunodeficiency.

81 ICU: Intensive care unit.

82 IL-1: Interleukin 1.

83 IL-6: Interleukin 6.

84 IQR: Interquartile range.

85 LDH: Lactate dehydrogenase.

86 LOS: Length of stay.

87 Masl: Meters above sea level.

88 MICE: Multiple imputations by chained equation.

89 MV: Mechanical ventilation.

90 NSAIDs: Non-steroidal anti-inflammatory drugs.

91 $\text{PaO}_2/\text{FiO}_2$: partial pressure of oxygen (PaO_2) to inspired (FiO_2) partial pressure of
92 oxygen ratio.

93 RT-PCR: Reverse transcriptase-polymerase chain reaction.

94 S/F: Ratio of oxygen saturation to fraction-inspired of oxygen.

95 SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2.

96 TBC: Tuberculosis.

97 TBIL: Total bilirubin levels.

98 TNF α : Tumor necrosis factor- α .

99 WHO: World health organization.

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

Introduction (Words 2857)

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first reported as an outbreak of new viral pneumonia in Wuhan, China which was quickly distributed worldwide, generating a remarkable impact. Approximately 520,000,000 cases have been reported, from which about 6,300,000 have died (1). This infection has a wide range of clinical manifestations, ranging from asymptomatic disease to individuals who develop acute respiratory failure (2). However, most patients have mild symptoms of cough, headache, myalgia, fever, and diarrhea; a smaller proportion presents severe disease symptoms (1,2). The most common manifestation of severity is dyspnea, which can be seen in up to 40% of patients and is usually accompanied by hypoxemia (2).

The United States harbors almost a fifth of the infections worldwide and more than 1,000,000 deaths (1). It is followed by Brazil and India, with 665,000 and 524,000 deaths, respectively (1). Colombia occupies the twelfth position with approximately 140,000 deaths (1,3). Fatality, defined as the ratio between the cases of mortality and the cases confirmed with SARS-CoV-2, has progressively decreased from 3.7% in March 2020 to 1.22% in the present day (4). Colombia presents a fatality rate of 2.5%, higher than reported worldwide (3). Thus, further characterization of mortality in Colombia is required to develop personalized risk profiles to better respond to the pandemic, especially in the current transition to endemic disease.

Patients with severe diseases are more likely to suffer complications associated with higher mortality. One of the causes of this clinical deterioration is the cytokine

storm that generates a systemic inflammatory syndrome in which high cytokine levels are associated with hyperactivation of the cellular response (5–8). Molecular mimicry between host and viral proteins has also been described, as well as direct damage of the microorganism to tissues that present high expression of the angiotensin-converting enzyme-2 receptor, as other possible causes (5–8). These processes will result, in most cases, in a respiratory distress syndrome characterized by impaired gas exchange (6,8). They could also generate a hypercoagulability state responsible for the increase in thrombotic events in patients with this condition (5,7,9).

Multiple studies have found several factors associated with developing severe disease. Age, sex, ethnicity, socioeconomic status, inflammatory markers, and comorbid conditions have been widely associated with mortality. However, other factors include ancestry, environmental exposures, viral mutations, and geographic diversity (10–12). A recent study in Bogota, a city above 2,640 meters above sea level (masl) in Colombia, found that older age, low S/F: Ratio of oxygen saturation to fraction-inspired of oxygen (S/F), and high lactate dehydrogenase (LDH) on admission were predictors of mortality (13). However, observational studies have shown that other variables, in addition to age and sex, could exhibit interactions among them, which have been poorly characterized in our population (14).

Herein, we conducted a one-year case-control study to characterize the clinical and paraclinical factors associated with mortality during the first year of the pandemic. We implemented supervised machine learning algorithms to evaluate the interactions among variables and their potential to predict mortality on

admission. We adjusted the analysis by monthly progression, which may account for changes in the mortality rates due to viral mutations during the pandemic.

Methods

Study design

This case-control study was conducted in Clínica de Occidente, a tertiary referral center in Bogota, Colombia. The study screening was from February 29th, 2020, to March 1st, 2021. Included patients were selected from 4,163 cases positive for SARS-CoV-2 by reverse transcriptase-polymerase chain reaction (RT-PCR) and who attended the emergency room during the first year of the pandemic with complete clinical records. A non-probabilistic convenience sampling included a total of 282 deceased patients. An additional group of 282 patients, who were hospitalized but recovered from COVID-19, was included as the control group. These controls were matched to deceased patients by age, sex, and month of consultation (to account for viral mutations). None of the patients received vaccinations during the study timeframe, and the included patients attended the emergency room between April 3rd, 2020, to January 30th, 2021. This was a low-risk study according to the resolution 8430 of 1993 from the Ministry of Health of Colombia. The institutional review board of Clínica de Occidente approved the study design.

Inclusion and exclusion criteria

Inclusion criteria comprised the following: (1) aged at least 18 years; (2) COVID-19 diagnosis based on RT-PCR testing; (3) hospitalized patients; and (4) death

registered in death certificate adequately filled out by medical doctors who certified COVID-19 as the cause of death. In the case of controls, they were hospitalized but recovered during the hospitalization and received ambulatory management. Patients referred to another clinical setting were excluded from the analysis.

Clinical variables and data extraction

Medical records were reviewed through access to the server assigned by the hospital. Information considered relevant for this investigation was extracted, including sociodemographic variables, clinical features (signs, symptoms, laboratories on admission), past medical history, in-hospital management (i.e., requirement of ventilatory support, intensive care unit (ICU) admission, vasopressor support), medications administered on admission, organ-specific and systemic outcomes during hospitalization (Table 1, 2 and 3). Research participants remained anonymous at all times. All data were collected in an electronic and secure database as described elsewhere (15).

Patient and Public Involvement

Patients or the public were not involved in the research's design, conduct, reporting, or dissemination plans.

Statistical analyses

Univariate descriptive statistics were performed. Categorical variables were analyzed using frequencies, and continuous quantitative variables were expressed as the median and interquartile range (IQR). The Mann–Whitney U-test or Fisher

exact tests were used based on the results. None of the included parameters were subjected to statistical transformation or normalization.

The missing data rates for each variable in our study are shown in Tables 1,2,3 and Figure 1. Most of the missingness was secondary to the lack of standardization at the beginning of the pandemic for laboratory values required in the follow-up and management of the patients. We used multiple imputations by chained equation (MICE) to create and analyze five multiply imputed datasets for variables with less than 80% of data missingness (Table 3). Multiple imputations are considered cutting-edge by methodologists since they enhance accuracy and statistical power when compared to other missing data strategies. Incomplete variables were imputed under wholly conditional specification using the default settings of the MICE 3.14 package (16).

Then, we built classification and regression trees (CART) to evaluate the relationship between clinical variables and mortality in each imputed dataset on admission. This strategy aims to identify, at each partitioning step, the best predictive variable and corresponding splitting value while optimizing a statistical criterion. Variables with a p-value ≤ 0.25 in the bivariate analysis were included in the model. Each model's confusion matrix was built to determine MICE's best predictive model from the five imputed datasets and was reported accordingly. The significance level of the study was set to 0.05. Statistical analyses were done using R software version 4.1.2.

Results

General characteristics

The general characteristics of patients are shown in Table 1. The median age of the patients and sex were similar between groups. On admission, fever, odynophagia, anosmia, dysgeusia, chest pain, myalgias, arthralgias, headache, and diarrhea were most commonly reported in recovered patients. General discomfort, dyspnea, respiratory distress, and seizures were most common in deceased patients (Table 1). Most of the included patients visited the hospital between July and September 2020.

Treatment, comorbidities, and outcomes

Heart failure, hypertension, obesity, type II diabetes, and active smoking were most common in deceased patients. On the other hand, gastroesophageal reflux disease (GERD) was mainly reported in recovered patients (Table 1). Deceased patients were more likely to receive corticosteroids and antibiotics upon admission (Table 2). Few patients received antimalarials or antivirals. On admission, deceased patients exhibited paraclinical alterations in inflammatory markers related to hematological, liver, and pulmonary function (Table 3). This was further confirmed during the follow-up, given the higher rates of systemic and organic compromise presented during the hospitalization (Table 3).

Missing data imputation

The missing values across the variables ranged between 18.79% and 98.94%. The high missingness rate was related to laboratory variables, mainly albumin, creatine kinase, procalcitonin, and the erythrocyte sedimentation rate (Table 3 and Figure 1 A-C), whereas clinical characteristics, in-hospital admission management, and

comorbidities did not have missed values. Since the objective of our study was to evaluate the interaction between clinical and paraclinical factors on admission in predicting mortality, we conducted a MICE imputation strategy to include all the cases in multivariate models.

We created and analyzed five multiply imputed datasets for variables with less than 80% of data missingness (Figure 1C). The sensitivity analysis yielded no significant differences between the primary and the five imputed datasets (Table 4). This confirmed that the distribution of imputed data was similar to the original dataset, as well as the stability of the imputation models.

A multivariate predictive model for mortality

After imputation, we aimed to evaluate the interaction of multiple variables in predicting mortality on admission. We estimated CART models using the variables with $p\text{-values} \leq 0.25$ from the bivariate analysis for each imputed dataset. We constructed a confusion matrix for each CART model to estimate the best-fitted model to the data. After this analysis, we selected the best model based on estimated accuracy (Figure 2).

The analysis revealed that multiple variables interacted in the prediction of mortality. Respiratory distress on admission was the first splitting variable from the tree. Then, the second node was determined by LDH and dyspnea. The former interacted with hemoglobin, D-dimer, ferritin, blood urea nitrogen (BUN), C-reactive protein (CRP), partial pressure of oxygen (PaO_2) to inspired (FiO_2) partial pressure of oxygen ratio ($\text{PaO}_2/\text{FiO}_2$), dysgeusia, total bilirubin levels (TBIL), and platelets. The latter interacted with GERD (Figure 2).

Discussion

A few published papers have previously characterized the risk factors associated with COVID-19 mortality in Latin American patients (17,18). Therefore, the present study highlights multiple variables, including laboratory abnormalities and clinical features associated with COVID-19 mortality. The main findings in our study were significant associations between TBIL, ferritin, D-Dimer levels, dyspnea, GERD, and increased risk of mortality in COVID-19 patients. In contrast, dysgeusia was associated with a better prognosis. Other trends were found between BUN, respiratory distress, platelet count, PaO₂/FiO₂, and CRP.

Increased TBIL was associated with poor COVID-19 outcomes. This is consistent with previous findings (19) which suggest that high TBIL may reflect a severe level of hepatic injury among severely ill COVID-19 patients, possibly due to direct cytopathic effect, immune-mediated effects, hypoxia-induced changes, microvascular thrombosis, among others (20,21). It is not surprising that some authors have also considered that TBIL at admission is directly correlated with the hospital progression of COVID-19 (20).

Serum ferritin has been cited as one of the mortality indicators in COVID-19 patients due to its ability to assess intracellular iron status (22,23). These events are common in the pathogenesis of uncontrolled inflammation and massive cytokine release, supporting the hypothesis that hyper-inflammation is a possible pathogenic mechanism in COVID-19 and, therefore, the rise of serum ferritin level findings (23). Multiple studies have also suggested that higher levels of CRP on admission are linked to disease progression, severity, and death (24–26).

However, we found that CRP interacts with other variables related to inflammation and organ damage implicated in mortality.

The D-dimer antigen is a unique marker of fibrin degradation that may indicate infection-related coagulation effects (27,28). Furthermore, previous studies have found that critically ill patients with COVID-19 have extremely high D-dimer levels, which can lead to clotting disorders and peripheral microthrombi formation (29). In this study, we found that higher levels of D-dimer at admission were associated with increased mortality risk in COVID-19 patients. This is consistent with other studies that have found D-dimer as another crucial prognostic factor in estimating mortality in COVID-19 patients (30).

Similar results from other studies also elicited dyspnea as a significant clinical variable for mortality prediction among COVID-19 patients (31). A meta-analysis study by He et al. showed that dyspnea was the main difference between mild and severe COVID-19 [40], and another study confirmed this observation (32). On the other hand, in patients with gastrointestinal symptoms during COVID-19, GERD appeared to be a protective factor for acute respiratory distress syndrome (ARDS), and mortality possible due to a conversely increased acidic environment associated with GERD that suppresses COVID-19 viral load at the gastrointestinal point of entry, favoring a milder disease course (33). In our study, similar results were obtained; COVID-19 patients without GERD had an increased mortality risk.

Interestingly, dysgeusia was associated with lower odds of death in our study. This might be related to a different inflammatory profile with a better local immune response, which could limit the spread of the virus in the body, resulting in less severe disease and a strong local inflammatory response that could mainly affect

taste receptors. However, there is still a lack of information regarding dysgeusia, perhaps because of the heterogeneity in how it has been assessed and defined (34). On the other hand, previous studies have focused on anosmia as a protective factor for mortality (35,36).

A significant association was found between BUN and mortality risk among COVID-19 patients. Other studies also supported our results since BUN has been previously considered a death-related feature (37). Although COVID-19 impacts mainly the lung, it can also affect the kidney, and the increase in BUN may reflect kidney injury along with other biomarkers. Moreover, kidney involvement in severe COVID-19 patients has been frequently observed (37,38).

The glycolytic enzyme LDH has long been identified as an inflammation biomarker that plays a crucial role in the anaerobic glycolysis pathway and increases in the bloodstream under conditions of membrane instability (39). In most studies, authors have concluded that LDH is a deleterious prognostic biomarker with high accuracy for predicting in-hospital mortality in severe and critically ill patients with COVID-19 (40,41). In addition, thrombocytopenia has also been reported in COVID-19 patients and is considered a potential risk factor for mortality in this group of patients (42,43). The PaO₂/FiO₂ partial pressure of oxygen has also been widely used to diagnose and assess the severity of patients with ARDS (44,45).

Since several factors have been related to the incidence of mortality in COVID-19, the interaction among them may provide better insights into predicting deleterious outcomes. A similar study in Bogota found that older age, low S/F, and high LDH on admission were predictors of mortality (13). This study assessed the interaction of variables by CART analysis, yielding five plausible groups. However, it is unclear

whether the multivariate model included all the patients since information on missing data is unavailable, and the model accuracy is unknown.

In contrast, Afrash et al.(46) developed hybrid machine-learning algorithms to predict mortality. Authors found that the mixture of variables such as length of stay (LOS), age, cough, respiratory intubation, dyspnea, cardiovascular disease, leukocytosis, BUN, CRP, and pleural effusion yielded a high accuracy (90%), specificity (83%), and sensitivity (97%). However, it includes variables that depend on the patient follow-up (i.e., LOS), thus hindering its applicability during admission to the emergency room, that is, during the early stages of the disease.

Several manuscripts have developed similar approaches involving different variables in the prediction of mortality; some of them include dyspnea (46,47), BUN, platelet count (47,48), sex (47), age (46,48–51), cough (46), weight (49), cardiovascular disease (46), orotracheal intubation (46), and pleural effusion (46,49), respiratory rate (47), fraction of inspired oxygen (47), blood oxygen saturation (47,48) pH (47), aspartate aminotransferase levels (47), estimated glomerular filtration rate (47), lymphocyte count (49,52), white blood cell count (46,48), creatine (51) lactic acid (52) and serum calcium (52). However, most of them do not focus on the interaction of such variables on hospital admission, and their interpretability by clinicians is difficult. In this line, our study provides a novel, biologically plausible, and reliable model with clinical applicability in the emergency room during patient admission. Further studies are necessary to evaluate whether our model may help to predict the efficacy of therapeutic strategies or outpatient mortality (during post-COVID syndrome).

Study limitations must be acknowledged. It was a single-center retrospective study based on clinical records. This could have prone our study to reporting bias, for example, on time from symptoms onset to consultation or clinical features. In addition, the therapeutic strategies changed during the pandemic; however, our study was matched by month of consultation, making this variable an unlikely confounding/interaction factor for our results. The objective of our study was to uncover other risk factors for mortality besides sex and age. In this line, the adjustment for age, sex, and month of consultation allowed the discovery of new interactions for mortality prediction. The lack of follow-up beyond the hospitalization in recovered patients could have offered new insights into mortality beyond the clinical settings. Another potential shortcoming of the present study is that latent autoimmunity, especially the positivity for anti-IFN- α antibodies, was not evaluated. Further analysis involving such variables could improve our model's reliability and predictive accuracy.

Conclusions

Our study demonstrates the highly complex interactions among different risk factors to predict mortality in COVID-19, in addition to age and sex. This predictive approach may also provide new insights into the tailored management of this illness in several clinical settings, specifically in Colombian clinical settings. This study should encourage the follow-up of recovered patients beyond hospitalization and despite their clinical status during acute COVID-19.

Role of the Funder/Sponsor

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Funding

This study was supported by a grant from Universidad del Rosario (ABN011) Bogota, Colombia.

Ethics approval and consent to participate

The institutional review board of Clínica de Occidente approved the study design. Written informed consent was not obtained from participants since the study was based on a retrospective review of clinical records. It was a low-risk study according to the resolution 8430 of 1993 from the Ministry of Health of Colombia. The study was conducted following the principles stated in the Declaration of Helsinki and Good Clinical Practice guidelines.

Consent to publish

None of the patients were required to consent to publish, given the study's retrospective nature.

Declaration of competing interest

None.

Author Contributions Statement

Conceptualization: JMA; Acquisition of data: LCD, MR, YR, MH, SB, PSG, NAC, ME, SG, VB, VM, JEC, EZ; Methodology: JMA, LCD, MR; Statistical Analysis: MR; Funding acquisition: JMA, CRS; Center Coordination: YR, MR; Writing & editing: LCD, MR, CRS, YAA, DMM. All authors read and approved the final version of this manuscript.

Acknowledgments

The authors would like to thank the medical staff of Clinical del Occidente, Dr. Ivan Torres, and Maria Claudia Fuentes, for their assistance at the beginning of the study.

Availability of data and materials

Data will be available upon request to the corresponding author.

About the Author

Dr. Camacho-Domínguez is a research assistant in autoimmune diseases at Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Bogota, Colombia. Her primary research interests include infectious diseases, immunology, and rheumatology.

References

1. World Health Organization. WHO COVID-19 Dashboard. Geneva; 2020.
2. Gandhi RT, Lynch JB, del Rio C. Mild or Moderate Covid-19. N Engl J Med [Internet]. 2020 Oct 29 [cited 2022 Sep 23];383(18):1757–66. Available from: <https://pubmed.ncbi.nlm.nih.gov/32329974/>
3. Instituto Nacional de Salud. COVID-19 en Colombia . 2020.
4. Ritchie H, Mathieu E. Coronavirus Pandemic (COVID-19). 2020.
5. Chams N, Chams S, Badran R, Shams A, Araj A, Raad M, et al. COVID-19: A Multidisciplinary Review. Front Public Health [Internet]. 2020 Jul 29 [cited 2022 Sep 23];8. Available from: <https://pubmed.ncbi.nlm.nih.gov/32850602/>
6. Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. N Engl J Med [Internet]. 2020 Dec 17 [cited 2022 Sep 23];383(25):2451–60. Available from: <https://pubmed.ncbi.nlm.nih.gov/32412710/>
7. Parasher A. COVID-19: Current understanding of its Pathophysiology, Clinical presentation and Treatment. Postgrad Med J [Internet]. 2021 May 1 [cited 2022 Sep 23];97(1147):312–20. Available from: <https://pubmed.ncbi.nlm.nih.gov/32978337/>

8. Fajgenbaum DC, June CH. Cytokine Storm. *N Engl J Med* [Internet]. 2020 Dec 3 [cited 2022 Sep 23];383(23):2255–73. Available from: <https://pubmed.ncbi.nlm.nih.gov/33264547/>
9. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and Antiphospholipid Antibodies in Patients with Covid-19. *N Engl J Med* [Internet]. 2020 Apr 23 [cited 2022 Sep 23];382(17):e38. Available from: <https://pubmed.ncbi.nlm.nih.gov/32268022/>
10. Hu J, Li C, Wang S, Li T, Zhang H. Genetic variants are identified to increase risk of COVID-19 related mortality from UK Biobank data. *Hum Genomics* [Internet]. 2021 Dec 1 [cited 2022 Sep 23];15(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/33536081/>
11. Pansini R, Fornacca D. COVID-19 Higher Mortality in Chinese Regions With Chronic Exposure to Lower Air Quality. *Front Public Health* [Internet]. 2021 Jan 22 [cited 2022 Sep 23];8. Available from: <https://pubmed.ncbi.nlm.nih.gov/33585383/>
12. Anaele BI, Doran C, McIntire R. Visualizing COVID-19 Mortality Rates and African-American Populations in the USA and Pennsylvania. *J Racial Ethn Health Disparities* [Internet]. 2021 Dec 1 [cited 2022 Sep 23];8(6):1356–63. Available from: <https://pubmed.ncbi.nlm.nih.gov/33565050/>
13. Rodriguez Lima DR, Pinzón Rondón ÁM, Rubio Ramos C, Pinilla Rojas DI, Niño Orrego MJ, Díaz Quiroz MA, et al. Clinical characteristics and mortality associated with COVID-19 at high altitude: a cohort of 5161 patients in Bogotá, Colombia. *Int J Emerg Med* [Internet]. 2022 Dec 1 [cited 2022 Sep 23];15(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/35597911/>
14. Parohan M, Yaghoubi S, Seraji A, Javanbakht MH, Sarraf P, Djalali M. Risk factors for mortality in patients with Coronavirus disease 2019 (COVID-19) infection: a systematic review and meta-analysis of observational studies. *Aging Male* [Internet]. 2020 [cited 2022 Sep 23];23(5):1416–24. Available from: <https://pubmed.ncbi.nlm.nih.gov/32508193/>
15. Rojas M, Rodriguez Y, Pacheco Y, Zapata E, Monsalve DM, Mantilla RD, et al. Resilience in women with autoimmune rheumatic diseases. *Joint Bone Spine* [Internet]. 2018 Dec 1 [cited 2022 Sep 23];85(6):715–20. Available from: <https://pubmed.ncbi.nlm.nih.gov/29289647/>
16. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011 Dec;45(3):1–67.
17. Araujo M, Ossandón P, Abarca AM, Menjiba AM, Muñoz AM. Pronóstico de pacientes hospitalizados por COVID-19 en un centro terciario en Chile: estudio de cohorte. *Medwave*. 2020 Nov 17;20(10):e8066.

18. Alonso R, Camon AM, Cardozo C, Albiach L, Agüero D, Marcos MA, et al. Clinical Presentation and Outcome of COVID-19 in a Latin American Versus Spanish Population: Matched Case-Control Study. *Infect Dis Ther* [Internet]. 2022 Jun 1 [cited 2022 Sep 23];11(3):1243–51. Available from: <https://pubmed.ncbi.nlm.nih.gov/35476212/>
19. Lei F, Liu YM, Zhou F, Qin JJ, Zhang P, Zhu L, et al. Longitudinal Association Between Markers of Liver Injury and Mortality in COVID-19 in China. *Hepatology* [Internet]. 2020 Aug 1 [cited 2022 Sep 23];72(2):389–98. Available from: <https://pubmed.ncbi.nlm.nih.gov/32359177/>
20. Russo A, Pisaturo M, Palladino R, Maggi P, Numis FG, Gentile I, et al. Prognostic Value of Transaminases and Bilirubin Levels at Admission to Hospital on Disease Progression and Mortality in Patients with COVID-19-An Observational Retrospective Study. *Pathogens* [Internet]. 2022 Jun 1 [cited 2022 Sep 23];11(6). Available from: <https://pubmed.ncbi.nlm.nih.gov/35745506/>
21. Dawood DRM, Salum GM, El-Meguid MA. The Impact of COVID-19 on Liver Injury. *Am J Med Sci* [Internet]. 2022 Feb 1 [cited 2022 Sep 23];363(2):94–103. Available from: <https://pubmed.ncbi.nlm.nih.gov/34752738/>
22. Zhou S, Li H, Li S. The Associations of Iron Related Biomarkers with Risk, Clinical Severity and Mortality in SARS-CoV-2 Patients: A Meta-Analysis. *Nutrients* [Internet]. 2022 Aug 1 [cited 2022 Sep 23];14(16). Available from: <https://pubmed.ncbi.nlm.nih.gov/36014912/>
23. Dahan S, Segal G, Katz I, Hellou T, Tietel M, Bryk G, et al. Ferritin as a Marker of Severity in COVID-19 Patients: A Fatal Correlation. *Isr Med Assoc J* [Internet]. 2020 Aug 1 [cited 2022 Sep 23];22(8):494–500. Available from: <https://pubmed.ncbi.nlm.nih.gov/33236582/>
24. Chen W, Zheng KI, Liu S, Yan Z, Xu C, Qiao Z. Plasma CRP level is positively associated with the severity of COVID-19. *Ann Clin Microbiol Antimicrob* [Internet]. 2020 [cited 2022 Sep 24];19(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/32414383/>
25. Lavillegrand JR, Garnier M, Spaeth A, Mario N, Hariri G, Pilon A, et al. Elevated plasma IL-6 and CRP levels are associated with adverse clinical outcomes and death in critically ill SARS-CoV-2 patients: inflammatory response of SARS-CoV-2 patients. *Ann Intensive Care* [Internet]. 2021 Dec 1 [cited 2022 Sep 24];11(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/33439360/>
26. Pál K, Molnar AA, Huǎnu A, Szederjesi J, Branea I, Timár Á, et al. Inflammatory Biomarkers Associated with In-Hospital Mortality in Critical COVID-19 Patients. *Int J Mol Sci* [Internet]. 2022 Sep 9 [cited 2022 Sep

24];23(18):10423. Available from:

<https://pubmed.ncbi.nlm.nih.gov/36142336/>

27. Adam SS, Key NS, Greenberg CS. D-dimer antigen: current concepts and future prospects. *Blood* [Internet]. 2009 Mar 26 [cited 2022 Sep

24];113(13):2878–87. Available from:

<https://pubmed.ncbi.nlm.nih.gov/19008457/>

28. Ye W, Chen G, Li X, Lan X, Ji C, Hou M, et al. Dynamic changes of D-dimer and neutrophil-lymphocyte count ratio as prognostic biomarkers in COVID-19. *Respir Res* [Internet]. 2020 Jul 3 [cited 2022 Sep 24];21(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/32620118/>

29. Hilda F, Liana P, Nurtjahyo A, Hudari H, Sari NP, Umar TP, et al. D-Dimer as a Sensitive Biomarker of Survival Rate in Patients with COVID-19. *Eurasian J Med* [Internet]. 2022 Aug 11 [cited 2022 Sep 24]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/35950823>

30. Dessie ZG, Zewotir T. Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients. *BMC Infect Dis*. 2021 Dec 21;21(1):855.

31. Lee JH, Ahn JS, Chung MJ, Jeong YJ, Kim JH, Lim JK, et al. Development and Validation of a Multimodal-Based Prognosis and Intervention Prediction Model for COVID-19 Patients in a Multicenter Cohort. *Sensors (Basel)* [Internet]. 2022 Jul 1 [cited 2022 Sep 24];22(13). Available from: <https://pubmed.ncbi.nlm.nih.gov/35808502/>

32. Yang L, Jin J, Luo W, Gan Y, Chen B, Li W. Risk factors for predicting mortality of COVID-19 patients: A systematic review and meta-analysis. *PLoS One* [Internet]. 2020 Nov 1 [cited 2022 Sep 24];15(11). Available from: <https://pubmed.ncbi.nlm.nih.gov/33253244/>

33. Patil N, Kalgotra P, Sundaram S, Melquist S, Parasa S, Desai M, et al. Factors associated with poor outcomes among COVID-19 patients with gastrointestinal symptoms. *Gastro hep advances* [Internet]. 2022 Aug [cited 2022 Sep 24]; Available from: <https://pubmed.ncbi.nlm.nih.gov/36043056/>

34. Carrillo-Larco RM, Altez-Fernandez C. Anosmia and dysgeusia in COVID-19: A systematic review. *Wellcome Open Res* [Internet]. 2020 [cited 2022 Sep 24];5. Available from: <https://pubmed.ncbi.nlm.nih.gov/32587902/>

35. Talavera B, García-Azorín D, Martínez-Pías E, Trigo J, Hernández-Pérez I, Valle-Peñacoba G, et al. Anosmia is associated with lower in-hospital mortality in COVID-19. *J Neurol Sci* [Internet]. 2020 Dec 15 [cited 2022 Sep 24];419. Available from: <https://pubmed.ncbi.nlm.nih.gov/33035870/>

36. Lechien JR, Chiesa-Estomba CM, Beckers E, Mustin V, Ducarme M, Journe F, et al. Prevalence and 6-month recovery of olfactory dysfunction: a multicentre study of 1363 COVID-19 patients. *J Intern Med* [Internet]. 2021 Aug 1 [cited 2022 Sep 24];290(2):451–61. Available from: <https://pubmed.ncbi.nlm.nih.gov/33403772/>
37. Zadeh Hosseingholi E, Maddahi S, Jabbari S, Molavi G. Identification of High Death Risk Coronavirus Disease-19 Patients using Blood Tests. *Adv Biomed Res* [Internet]. 2022 [cited 2022 Sep 24];11(1):58. Available from: <https://pubmed.ncbi.nlm.nih.gov/36124024/>
38. Liu YM, Xie J, Chen MM, Zhang X, Cheng X, Li H, et al. Kidney Function Indicators Predict Adverse Outcomes of COVID-19. *Med (N Y)* [Internet]. 2021 Jan 15 [cited 2022 Sep 24];2(1):38-48.e2. Available from: <https://pubmed.ncbi.nlm.nih.gov/33043313/>
39. Gupta GS. The Lactate and the Lactate Dehydrogenase in Inflammatory Diseases and Major Risk Factors in COVID-19 Patients. *Inflammation* [Internet]. 2022 [cited 2022 Sep 24];1. Available from: <https://pubmed.ncbi.nlm.nih.gov/36124024/>
40. Dong X, Sun L, Li Y. Prognostic value of lactate dehydrogenase for in-hospital mortality in severe and critically ill patients with COVID-19. *Int J Med Sci* [Internet]. 2020 [cited 2022 Sep 24];17(14):2225–31. Available from: <https://pubmed.ncbi.nlm.nih.gov/32922185/>
41. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* [Internet]. 2020 Jul 1 [cited 2022 Sep 24];180(7):934–43. Available from: <https://pubmed.ncbi.nlm.nih.gov/32167524/>
42. Amgala A, Othman M. Hemostatic laboratory derangements in COVID-19 with a focus on platelet count. *Platelets* [Internet]. 2020 Aug 17 [cited 2022 Sep 24];31(6):740–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/32456506/>
43. Yuan Y, Wang G, Chen X, Ye XL, Li XK, Li R, et al. Thrombocytopenia and increased risk of adverse outcome in COVID-19 patients. *PeerJ* [Internet]. 2022 Jun 30 [cited 2022 Sep 24];10. Available from: <https://pubmed.ncbi.nlm.nih.gov/35791362/>
44. Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA* [Internet]. 2012 Jun 13 [cited 2022 Sep 24];307(23):2526–33. Available from: <https://pubmed.ncbi.nlm.nih.gov/22797452/>

45. Villar J, Blanco J, del Campo R, Andaluz-Ojeda D, Díaz-Domínguez FJ, Muriel A, et al. Assessment of PaO₂/FiO₂ for stratification of patients with moderate and severe acute respiratory distress syndrome. *BMJ Open* [Internet]. 2015 Mar 27 [cited 2022 Sep 24];5(3). Available from: <https://pubmed.ncbi.nlm.nih.gov/25818272/>
46. Afrash MR, Shanbehzadeh M, Kazemi-Arpanahi H. Predicting Risk of Mortality in COVID-19 Hospitalized Patients using Hybrid Machine Learning Algorithms. *J Biomed Phys Eng* [Internet]. 2022 Dec 1 [cited 2023 Jan 1];12(6):611–26. Available from: <https://pubmed.ncbi.nlm.nih.gov/36569564/>
47. Weizman O, Duceau B, Trimaille A, Pommier T, Cellier J, Geneste L, et al. Machine learning-based scoring system to predict in-hospital outcomes in patients hospitalized with COVID-19. *Arch Cardiovasc Dis* [Internet]. 2022 Dec [cited 2023 Jan 1];115(12). Available from: <https://pubmed.ncbi.nlm.nih.gov/36376208/>
48. Xu Y, Trivedi A, Becker N, Blazes M, Ferres JL, Lee A, et al. Machine learning-based derivation and external validation of a tool to predict death and development of organ failure in hospitalized patients with COVID-19. *Sci Rep* [Internet]. 2022 Dec 1 [cited 2023 Jan 1];12(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/36209335/>
49. Padilha DMH, Garcia GR, Liveraro GSS, Mendes MCS, Takahashi MES, Lascala F, et al. Construction of a nomogram for predicting COVID-19 in-hospital mortality: A machine learning analysis. *Inform Med Unlocked* [Internet]. 2023 Jan 1 [cited 2023 Jan 1];36. Available from: <https://pubmed.ncbi.nlm.nih.gov/36474601/>
50. Yu Z, Li X, Zhao J, Sun S. Identification of hospitalized mortality of patients with COVID-19 by machine learning models based on blood inflammatory cytokines. *Front Public Health* [Internet]. 2022 Nov 17 [cited 2023 Jan 1];10. Available from: <https://pubmed.ncbi.nlm.nih.gov/36466533/>
51. González-Cebrián A, Borràs-Ferrís J, Ordovás-Baines JP, Hermenegildo-Caudevilla M, Climente-Marti M, Tarazona S, et al. Machine-learning-derived predictive score for early estimation of COVID-19 mortality risk in hospitalized patients. *PLoS One* [Internet]. 2022 Sep 1 [cited 2023 Jan 1];17(9). Available from: <https://pubmed.ncbi.nlm.nih.gov/36137106/>
52. Ustebay S, Sarmis A, Kaya GK, Suhan M. A comparison of machine learning algorithms in predicting COVID-19 prognostics. *Intern Emerg Med* [Internet]. 2022 [cited 2023 Jan 1]; Available from: <https://pubmed.ncbi.nlm.nih.gov/36116079/>

Table 1. General characteristics of patients diagnosed with COVID-19.

Variable	Recovered patients (n=282)	Deceased patients (n=282)	P-value ^a	Missingness (n: 564)
Sociodemographics (%)				
Sex			1.0000	0 (0%)
Female	110 (39.0%)	110 (39.0%)		
Male	172 (61%)	172 (61%)		
Age (Median - IQR)	68 (60 - 77)	69 (60 - 78)	0.4843	0 (0%)
The month of admission – (2020 to 2021)			1.0000	0 (0%)
April	4 (1.4%)	4 (1.4%)		
May	8 (2.8%)	8 (2.8%)		
June	17 (6.0%)	17 (6.0%)		
July	77 (27.3%)	77 (27.3%)		
August	49 (17.4%)	49 (17.4%)		
September	20 (7.1%)	20 (7.1%)		
October	16 (5.7%)	16 (5.7%)		
November	9 (3.2%)	9 (3.2%)		
December	37 (13.1%)	37 (13.1%)		
January	45 (16.0%)	45 (16.0%)		
Clinical characteristics on admission (%)				
Fever	167 (59.2%)	138 (48.9%)	0.0179	0 (0%)
Cough with sputum	41 (14.5%)	45 (16.0%)	0.7255	0 (0%)
Hemoptysis	1 (0.4%)	1 (0.4%)	1.0000	0 (0%)
Dry cough	177 (62.8%)	162 (57.4%)	0.2286	0 (0%)
Odynophagia	66 (23.4%)	35 (12.4%)	0.0009	0 (0%)

Anosmia	32 (11.3%)	7 (2.5%)	< 1e-04	0 (0%)
Dysgeusia	34 (12.1%)	7 (2.5%)	< 1e-04	0 (0%)
Rhinorrhea	27 (9.6%)	18 (6.4%)	0.2134	0 (0%)
Wheezing	4 (1.4%)	4 (1.4%)	1.0000	0 (0%)
Chest Pain	55 (19.5%)	35 (12.4%)	0.0285	0 (0%)
Myalgias	70 (24.8%)	40 (14.2%)	0.0020	0 (0%)
Arthralgias	31 (11.0%)	14 (5.0%)	0.0122	0 (0%)
General discomfort	165 (58.5%)	226 (80.1%)	< 1e-04	0 (0%)
Dyspnea	137 (48.6%)	227 (80.5%)	< 1e-04	0 (0%)
Inability to walk	1 (0.4%)	0 (0.0%)	1.0000	0 (0%)
Respiratory distress	22 (7.8%)	149 (52.8%)	< 1e-04	0 (0%)
Headache	66 (23.4%)	29 (10.3%)	< 1e-04	0 (0%)
Seizure	0 (0.0%)	5 (1.8%)	0.0614	0 (0%)
Abdominal pain	21 (7.4%)	21 (7.4%)	1.0000	0 (0%)
Nausea/Vomiting	27 (9.6%)	15 (5.3%)	0.0766	0 (0%)
Diarrhea	45 (16.0%)	22 (7.8%)	0.0039	0 (0%)
Bleeding	8 (2.8%)	7 (2.5%)	1.0000	0 (0%)

^a p values for categorical variables obtained by Fisher's exact test. Quantitative variables were analyzed by Mann–Whitney U-test. Abbreviations;; IQR: Interquartile range.

Table 2. Comorbidities and hospital management of patients diagnosed with COVID-19.

Variable	Recovered patients (n=282)	Deceased patients (n=282)	P-value ^a	Misssingness (n: 564)
Comorbidities (%)				
Coronary heart disease	22 (7.8%)	25 (8.9%)	0.7609	0 (0%)
Heart failure	9 (3.2%)	21 (7.4%)	0.0374	0 (0%)
Hypertension	131 (46.5%)	162 (57.4%)	0.0114	0 (0%)
Thromboembolic disease	9 (3.2%)	10 (3.5%)	1.0000	0 (0%)
Dyslipidemia	17 (6.0%)	23 (8.2%)	0.4124	0 (0%)
Chronic obstructive pulmonary disease	43 (15.2%)	38 (13.5%)	0.6312	0 (0%)
Asthma	3 (1.1%)	2 (0.7%)	1.0000	0 (0%)
Chronic kidney disease	39 (13.8%)	51 (18.1%)	0.2058	0 (0%)
Chronic liver disease	2 (0.7%)	4 (1.4%)	0.6858	0 (0%)
Stroke	5 (1.8%)	13 (4.6%)	0.0910	0 (0%)
Gastroesophageal reflux disease	17 (6.0%)	6 (2.1%)	0.0311	0 (0%)
Hepatitis C	1 (0.4%)	0 (0.0%)	1.0000	0 (0%)
Hepatitis B	0 (0.0%)	0 (0.0%)	-	0 (0%)
HIV	1 (0.4%)	2 (0.7%)	1.0000	0 (0%)
TBC	2 (0.7%)	1 (0.4%)	1.0000	0 (0%)
Type II diabetes	63 (22.3%)	89 (31.6%)	0.0175	0 (0%)
Cancer	14 (5.0%)	23 (8.2%)	0.1729	0 (0%)
Obesity	16 (5.7%)	58 (20.6%)	< 1e-04	0 (0%)
Hypothyroidism	47 (16.7%)	51 (18.1%)	0.7390	0 (0%)
Autoimmune disease	10 (3.5%)	11 (3.9%)	1.0000	0 (0%)
Former smoking	32 (11.3%)	36 (12.8%)	0.6983	0 (0%)

Active smoking	3 (1.1%)	12 (4.3%)	0.0328	0 (0%)
In-hospital management on admission (%)				
Corticosteroids	130 (46.1%)	163 (57.8%)	0.0069	0 (0%)
Azithromycin	2 (0.7%)	5 (1.8%)	0.4502	0 (0%)
Antibiotics	124 (44.0%)	186 (66.0%)	< 1e-04	0 (0%)
Ivermectin	0 (0.0%)	0 (0.0%)	-	0 (0%)
NSAIDs	112 (39.7%)	35 (12.4%)	< 1e-04	0 (0%)
ARBs	56 (19.9%)	26 (9.2%)	0.0005	0 (0%)
Bronchodilators	86 (30.5%)	88 (31.2%)	0.9274	0 (0%)
Anticoagulation	30 (10.6%)	33 (11.7%)	0.7894	0 (0%)
Antimalarials	1 (0.4%)	4 (1.4%)	0.3728	0 (0%)
Antiviral drugs	1 (0.4%)	0 (0.0%)	1.0000	0 (0%)
In-hospital management follow-up (%)				
Pronation therapy	26 (9.2%)	121 (42.9%)	< 1e-04	0 (0%)
Nasal cannula	154 (54.6%)	127 (45.0%)	0.0285	0 (0%)
Non-rebreather mask	42 (14.9%)	208 (73.8%)	< 1e-04	0 (0%)
High flow cannula	9 (3.2%)	38 (13.5%)	< 1e-04	0 (0%)
Orotracheal intubation	14 (5.0%)	169 (59.9%)	< 1e-04	0 (0%)
Dialysis	9 (3.2%)	51 (18.1%)	< 1e-04	0 (0%)
ICU admission	20 (7.1%)	126 (44.7%)	< 1e-04	0 (0%)
Inotropic or vasopressor support	14 (5.0%)	141 (50.0%)	< 1e-04	0 (0%)

^a p values for categorical variables obtained by Fisher's exact test. Quantitative variables were analyzed by Mann-Whitney U-test. Abbreviations: NSAIDs: Non-steroidal anti-inflammatory drugs; IQR: Interquartile range; ARBs: Angiotensin II receptor blockers; HIV: Human immunodeficiency virus; TBC: Tuberculosis; ICU: Intensive care unit; PaO₂/FiO₂: partial pressure of oxygen (PaO₂) to inspired (FiO₂) partial pressure of oxygen ratio.

Table 3. Clinical outcomes and paraclinics on admission of patients diagnosed with COVID-19.

Variable	Recovered patients (n=282)	Deceased patients (n=282)	P-value ^a	Misssingness (n: 564)
Clinical outcomes (%)				
Renal alterations	37 (13.1%)	130 (46.1%)	< 1e-04	0 (0%)
Co-infection	20 (7.1%)	57 (20.2%)	< 1e-04	0 (0%)
Hematological alterations	146 (51.8%)	242 (85.8%)	< 1e-04	0 (0%)
Thrombotic events	6 (2.1%)	18 (6.4%)	0.0199	0 (0%)
Neurological alterations	22 (7.8%)	31 (11.0%)	0.2481	0 (0%)
Cardiac alterations	18 (6.4%)	95 (33.7%)	< 1e-04	0 (0%)
Paraclinics on admission (Median - IQR)				
Hemoglobin (g/dL)	14.8 (13.5-16.2)	14.4 (12.4-15.8)	0.0086	106 (18.79%)
Platelets (Cells/uL)	206,000 (170,000 - 268,000)	217,000 (157,000 - 274,000)	0.8015	106 (18.79%)
Leukocytes (Cells/uL)	7,860 (5,620 - 10,285)	10,670 (6,730 - 14,720)	< 1e-04	106 (18.79%)
Lymphocytes (Cells/uL)	1,130 (740 - 1,650)	860 (585-1200)	< 1e-04	106 (18.79%)
Neutrophils (Cells/uL)	5,700 (3,630 - 8,196)	8,915 (5,267 - 13,112)	< 1e-04	107 (18.97%)
C-reactive protein (mg/L)	102.3 (40.2 - 171.2)	174.7 (85.8 - 267.5)	< 1e-04	247 (43.79%)
Erythrocyte sedimentation rate (mm/hr)	22 (22 - 22)	18 (12 - 20)	0.3798	558 (98.94%)
International normalized ratio	1.06 (1 - 1.17)	1.07 (0.99 - 1.2)	0.9196	453 (80.32%)
Aspartate aminotransferase (U/L)	37 (27 - 59)	64.5 (41.8 - 97.5)	0.0002	474 (84.04%)
Alanine aminotransferase (U/L)	30 (21.5 - 54.5)	46 (30 - 93.3)	0.0268	473 (83.87%)
Albumin (gr/dL)	3.32 (2.97 - 3.48)	3.21 (2.43 - 3.43)	0.3918	543 (96.28%)
Total bilirubin (mg/dL)	0.64 (0.45 - 0.87)	0.68 (0.49 - 1.09)	0.4640	422 (74.82%)
Blood urea nitrogen (mg/dL)	18.4 (13.8 - 28.8)	26.3 (18.4 - 43.2)	< 1e-04	166 (29.43%)

Creatinine (mg/dL)	0.99 (0.78 - 1.24)	1.16 (0.9 - 1.85)	< 1e-04	138 (24.47%)
Creatine kinase (U/L)	6.18 (6.18 - 6.18)	105 (78 - 135)	0.1432	558 (98.94%)
D-dimer (mg/dL)	0.775 (0.41 - 1.74)	1.52 (0.67 - 4.08)	< 1e-04	186 (32.98%)
Ferritin (ng/mL)	878.6 (495.1 - 1,501)	1,378 (731.2 - 2,475.5)	< 1e-04	211 (37.41%)
Lactic acid (mmol/L)	1.44 (1.14 - 1.99)	2.11 (1.5 – 3.37)	0.0180	443 (78.55%)
Lactate dehydrogenase (U/L)	298 (225.5 - 377.5)	466 (335.5 - 652)	< 1e-04	190 (33.69%)
Procalcitonin (ng/mL)	0.778 (0.55 - 1)	0.14 (0.09 – 0.19)	0.0641	558 (89.94%)
PaO ₂ /FiO ₂ (mmHg)	252.6 (211 - 297.80)	115 (71.5 - 222.15)	< 1e-04	155 (27.48%)
Oxygen saturation (%)	92.9 (89.15 - 95.85)	90.2 (86 - 93,78)	0.0002	131 (23.23%)

^a p values for categorical variables obtained by Fisher's exact test. Quantitative variables were analyzed by Mann–Whitney U-test. Abbreviations: NSAIDs: Non-steroidal anti-inflammatory drugs; IQR: Interquartile range; ARBs: Angiotensin II receptor blockers; HIV: Human immunodeficiency virus; TBC: Tuberculosis; ICU: Intensive care unit; PaO₂/FiO₂: partial pressure of oxygen (PaO₂) to inspired (FiO₂) partial pressure of oxygen ratio.

Table 4. P-values from sensitivity analysis between the primary dataset and the imputed datasets.

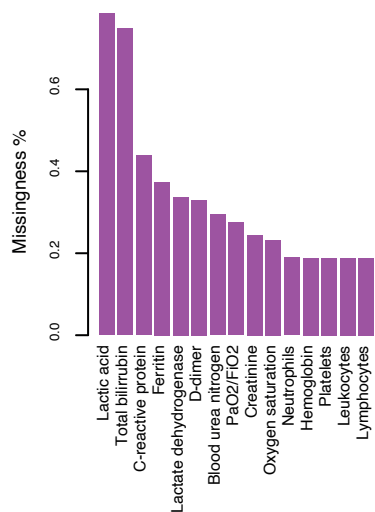
Variable	Main Vs. Imputed Dataset 1	Main Vs. Imputed Dataset 2	Main Vs. Imputed Dataset 3	Main Vs. Imputed Dataset 4	Main Vs. Imputed Dataset 5
Paraclinics on admission					
Hemoglobin (g/dL)	0.4503	0.8956	0.7301	0.8204	0.7955
Platelets (Cells/uL)	0.9705	0.7900	0.4258	0.5854	0.8971
Leukocytes (Cells/uL)	0.9079	0.7048	0.5637	0.4226	0.8877
Lymphocytes (Cells/uL)	0.9808	0.7245	0.8874	0.9945	0.8053
Neutrophils (Cells/uL)	0.8364	0.4069	0.6002	0.2971	0.6169
C-reactive protein (mg/L)	0.7267	0.9855	0.6387	0.8920	0.4734
Total bilirubin (mg/dL)	0.1539	0.4270	0.7628	0.8548	0.7584
Blood urea nitrogen (mg/dL)	0.9875	0.5288	0.6392	0.3416	0.9320
Creatinine (mg/dL)	0.8553	0.7842	0.9174	0.6214	0.7608
D-dimer (mg/dL)	0.9702	0.9092	0.6611	0.9079	0.7723
Ferritin (ng/mL)	0.7442	0.5230	0.1959	0.8348	0.4637
Lactic acid (mmol/L)	0.2466	0.2657	0.3806	0.4587	0.2240
Lactate dehydrogenase (U/L)	0.7523	0.4675	0.9154	0.7937	0.5737
PaO ₂ /FiO ₂ (mmHg)	0.9339	0.9456	0.6239	0.7643	0.5535
Oxygen saturation (%)	0.6625	0.9327	0.4655	0.6727	0.8556

^a p values for quantitative variables were analyzed by Mann–Whitney U-test. PaO₂/FiO₂: partial pressure of oxygen (PaO₂) to inspired (FiO₂) partial pressure of oxygen ratio.

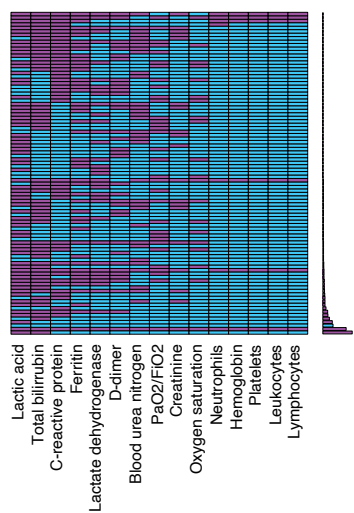
Fig 1. Missing data and imputation. **A.** Histogram of frequency of missing variables in the total of patients included (n: 584). **B.** Heatmap for the distribution of missing data. **C.** Distribution of imputed variables by MICE. The red lines correspond to the five imputed datasets, whereas the blue line corresponds to the original dataset. MICE: Multiple imputations by chained equation.

Fig 2. Classification and regression trees (CART). This strategy estimated a predictive model and 14 clinical profiles, including respiratory distress, LDH, dyspnea, hemoglobin, D-dimer, ferritin, blood urea nitrogen, C-reactive protein, $\text{PaO}_2/\text{FiO}_2$, dysgeusia, total bilirubin, platelets, and gastroesophageal reflux disease. LDH: lactate dehydrogenase; $\text{PaO}_2/\text{FiO}_2$: partial pressure of oxygen (PaO_2) to inspired (FiO_2) partial pressure of oxygen ratio.

A.



B.



C.

