

Letter to the Editor

Alejandro Santos-Lozano, Fernando Calvo-Boyero, Ana López-Jiménez, Cecilia Cueto-Felgueroso, Adrián Castillo-García, Pedro L. Valenzuela, Joaquín Arenas, Alejandro Lucia*, Miguel A. Martín and COVID-19 Hospital '12 Octubre' Clinical Biochemistry Study Group

Can routine laboratory variables predict survival in COVID-19? An artificial neural network-based approach

<https://doi.org/10.1515/cclm-2020-0730>

Received May 15, 2020; accepted September 9, 2020;
published online October 2, 2020

Keywords: biochemistry; coronavirus; hematology; mortality.

Alejandro Santos-Lozano and Fernando Calvo-Boyero contributed equally to this work.

Alejandro Lucia and Miguel A. Martín share senior authorship.

COVID19 Hospital '12 Octubre' Clinical Biochemistry Study Group: Miguel Pedrera-Jiménez, Alvar Blanco-Martínez, Jaime Cruz-Rojo, Aitor Delmiro, Alba Fernández-del Pozo, Jose L. Egiburu.

***Corresponding author: Alejandro Lucia**, MD, PhD, Research Institute of the Hospital 12 de Octubre ('imas12'), Madrid, Spain; Faculty of Sport Sciences, Universidad Europea de Madrid, Madrid, Spain; Biomedical Research Networking Center on Frailty and Healthy Aging (CIBERFES), Madrid, Spain; and European University of Madrid, Calle Tajo, s/n. 28670. Villaviciosa de Odón, Madrid, Spain, Phone: +3010 916647800, E-mail: alejandro.lucia@universidadeuropea.es

Alejandro Santos-Lozano, i+HeALTH, European University Miguel de Cervantes, Valladolid, Spain; Research Institute of the Hospital 12 de Octubre ('imas12'), Madrid, Spain

Fernando Calvo-Boyero, Ana López-Jiménez and Cecilia Cueto-Felgueroso, Clinical Biochemistry Department, Hospital 12 de Octubre, Madrid, Spain

Adrián Castillo-García, Fissac – Physiology, Health and Physical Activity, Madrid, Spain

Pedro L. Valenzuela, Systems Biology Department, University of Alcalá, Madrid, Spain

Joaquín Arenas, Research Institute of the Hospital 12 de Octubre ('imas12'), Madrid, Spain

Miguel A. Martín, Research Institute of the Hospital 12 de Octubre ('imas12'), Madrid, Spain; Mitochondrial Disorders Laboratory, "Hospital 12 de Octubre" and Spanish Biomedical Network for Rare Disorders (CIBERER), Madrid, Spain

To the Editor,

As of August 23, 2020, a cumulative total of over 23 million cases of coronavirus disease 2019 (COVID-19) infections and 800,000 related deaths has been reported [1]. Although most infected people present with mild-to-moderate symptoms, about one-third require hospitalization [2] (Last accessed 27 Aug 2020). Identification of valid prognostic factors for patients with COVID-19 might be helpful in the early diagnosis of "high-risk" individuals [3]. Some demographic and clinical variables – notably age, male sex, smoking or comorbidities such as cardiovascular disease, obesity or diabetes – have been associated with a worse prognosis [4]. By contrast, while some potential blood biomarkers (e.g., lactate dehydrogenase [LDH], C-reactive protein, coagulation parameters or lymphopenia) are emerging [4, 5], the evidence remains scarce and validation using advanced analyses in different cohorts is needed. The use of artificial intelligence (e.g., artificial neural network [ANN]) as a form of predictive analysis could help in this regard, and its combination with standard observation at triage might help to correctly identify those patients at a higher risk [6]. We have studied the prognostic value (in terms of survival) of potential "early" routine biochemistry and hematological biomarkers in patients with COVID-19.

This is a retrospective study of all admitted patients diagnosed with COVID-19 (by polymerase chain reaction) in a large public Hospital of Madrid, Spain (*Hospital 12 de Octubre*) from February 28 to March 30. The protocol was approved by the Ethics Committee of the aforementioned institution (reference #20/222) and adhered to the Declaration of Helsinki. The predictive value (i.e., odds of dying in the hospital versus discharge) of routine serum biochemistry (Cobas 8000 platform; Roche Diagnostics, Risch-Rotkreuz, Switzerland) and hematological parameters (DxH 900 hematology analyzer, Beckman Coulter,

Brea, CA) was tested in blood samples obtained at hospital admission.

Statistical analyses were performed in three stages. First, classical inference analyses (unpaired Student's and chi-square tests) and univariate logistic regression were used to assess differences between groups (survivors vs. non-survivors) and the association of each variable with the odds of COVID-19-related mortality, respectively. Secondly, we developed ANN models for the prediction of COVID-19 mortality using the hyperbolic tangent function as the activation function and with normalization of covariates through feature scaling. In the third stage, subjects were randomly divided into a training or testing set (7:3 ratio) to evaluate the performance of the ANN model [7]. We calculated the accuracy, sensitivity, specificity, and area under the receiver operating curve (AUC) of the ANN model. All analyses were performed using SPSS 24.0 (SPSS Inc., Chicago, IL) and STATA 13 (StataCorp LP.,

College Station, TX). The significance level was set at 0.05 for the univariate model and was Bonferroni-corrected for inference analyses (i.e., dividing 0.05 by the number of comparisons [thus, threshold p-value=0.05/35=0.0014]).

We studied 1,369 patients (1,090 survivors [median time before discharge, 40 days] and 279 non-survivors [median time before death, 6 days]). The ANN model in the training phase yielded 85% of correct predictions (specificity, 92%; sensitivity, 56%). According to the quality of classification [7], a final three-layer ANN model (36 input + bias units; 8 units in the hidden layer and 2 output units) was constructed for all patients. In the testing phase the model correctly defined 88% of the sample.

Together with a more advanced age, higher levels of serum lactate dehydrogenase (LDH) and lower levels of glomerular filtration rate (GFR), albumin and hemoglobin were the main predictors of mortality by sensitivity

Table 1: Results by group.

	Survivors (n=1,090)	Non-survivors (n=279)	p-Value between groups	OR (95% CI)	ANN results: Sensitivity analysis of the input variables	
					Importance	Relative importance, %
Age, years	57 ± 17	78 ± 12	<0.001	1.09 (1.08, 1.10)	0.102	100
LDH, U/L	315 ± 105	414 ± 161	<0.001	1.01 (1.00, 1.01)	0.099	97.4
GFR, mL/min	83 ± 21	55 ± 26	<0.001	0.96 (0.95, 0.96)	0.078	76.9
Albumin, g/L	42 ± 5	37 ± 5	<0.001	0.13 (0.09, 0.18)	0.074	72.1
Hemoglobin, g/L	139 ± 18	129 ± 20	<0.001	0.75 (0.70, 0.81)	0.045	44.1
Leukocytes, ×10 ⁹ /L	5.9 ± 2.6	7.5 ± 4.1	<0.001	1.16 (1.11, 1.21)	0.039	38.3
AST, U/L	39 ± 26	47 ± 39	<0.001	1.01 (1.00, 1.01)	0.038	37.3
RDW, %	14.0 ± 1.4	15.2 ± 2.1	<0.001	1.47 (1.36, 1.61)	0.033	32.6
Platelets, ×10 ⁹ /L	213 ± 76	195 ± 73	<0.001	1.00 (0.99, 1.00)	0.033	32.6
Monocytes, %	8.4 ± 3.7	6.6 ± 3.9	<0.001	0.86 (0.82, 0.90)	0.033	32.6
CRP, mg/L	6.6 ± 7.4	13.6 ± 10.3	<0.001	1.09 (1.07, 1.11)	0.033	32.2
MCH, pg/cell	30 ± 2	30 ± 2	0.217	1.04 (0.98, 1.10)	0.031	30.3
AST/ALT	1.3 ± 0.0	1.9 ± 0.8	<0.001	4.11 (3.18, 5.31)	0.031	30.2
GGT, U/L	62 ± 83	62 ± 79	0.988	1.00 (1.00, 1.01)	0.026	25.8
MPV, fL	8.8 ± 1.0	8.9 ± 1.0	0.016	1.18 (1.03, 1.36)	0.024	24.0
RBC, ×10 ¹² /L	4.7 ± 0.6	4.3 ± 0.7	<0.001	0.38 (0.30, 0.47)	0.022	21.6
Basophils, %	0.4 ± 0.2	0.3 ± 0.3	0.006	0.42 (0.22, 0.79)	0.020	19.8
aPTT, s	31 ± 5	31 ± 6	0.563	1.01 (0.98, 1.04)	0.020	19.6
Prothrombin time, s	14 ± 7	17 ± 13	<0.001	1.03 (1.01, 1.04)	0.019	18.4
Total protein, g/L	75 ± 7	71 ± 7	<0.001	0.44 (0.35, 0.55)	0.019	18.2
Potassium, mmol/L	4.0 ± 0.5	4.3 ± 0.7	<0.001	1.96 (1.52, 2.54)	0.018	17.8
Creatinine, μmol/L	59.6 ± 61.9	115 ± 88.4	<0.001	2.01 (1.59, 2.55)	0.017	16.6
Lymphocytes, %	20 ± 10	13 ± 11	<0.001	0.92 (0.91, 0.94)	0.016	15.8
Neutrophils, %	68 ± 12	76 ± 12	<0.001	1.07 (1.05, 1.08)	0.016	15.6
Eosinophils, %	0.4 ± 0.9	0.3 ± 0.7	0.121	0.85 (0.70, 1.04)	0.016	15.6
Sex, % male	52%	63%	<0.001	1.55 (1.18, 2.04)	0.014	14.0
Glucose, mmol/L	6.27 ± 1.94	7.44 ± 3.39	<0.001	1.01 (1.01, 1.01)	0.013	12.6
Total bilirubin, μmol/L	6.9 ± 3.5	8.6 ± 8.6	0.002	1.93 (1.21, 3.06)	0.011	11.1
ALT, U/L	36 ± 31	29 ± 30	0.002	0.99 (0.98, 1.00)	0.011	10.5
Chloride, mmol/L	96 ± 4	96 ± 5	0.489	0.99 (0.96, 1.02)	0.010	10.0
MDW, fL	25 ± 4	26 ± 5	<0.001	1.08 (1.05, 1.12)	0.010	9.4

Table 1: (continued)

	Survivors (n=1,090)	Non-survivors (n=279)	p-Value between groups	OR (95% CI)	ANN results: Sensitivity analysis of the input variables	
					Importance	Relative importance, %
ALP, U/L	76 ± 43	81 ± 44	0.063	1.00 (1.00, 1.01)	0.010	9.3
MCV, fl	89 ± 6	90 ± 6	<0.001	1.05 (1.03, 1.08)	0.009	9.3
Sodium, mmol/L	137 ± 4	136 ± 5	0.025	0.96 (0.93, 1.00)	0.007	6.9
Hematocrit, %	41 ± 5	39 ± 6	< 0.001	0.91 (0.89, 0.94)	0.007	6.7

Significant Bonferroni-corrected p-values for between-group comparisons (<0.0014) and significant associations with the risk of not surviving are in bold. ALP, alkaline phosphatase; ALT, alanine transaminase; ANN, artificial neural network; aPTT, activated partial thromboplastin time; AST, aspartate aminotransferase; CI, confidence interval; CRP, c-reactive protein; GFR, glomerular filtration rate; GGT, gamma-glutamyltransferase; LDH, lactate dehydrogenase; MCH; mean corpuscular hemoglobin; MCV, mean corpuscular volume; MDW, monocyte volume distribution width; OR, odds ratio; RBC, red blood cell count; RDW, red blood cell distribution width.

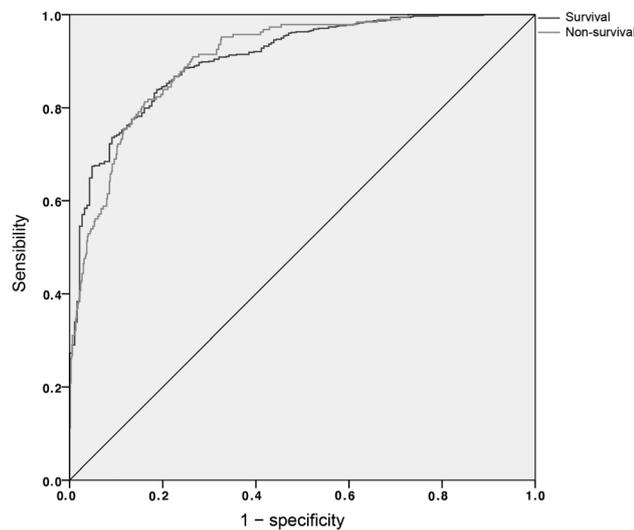


Figure 1: Results of receiver operating characteristic (ROC) curves of the artificial neural network model.

analysis (Table 1). The specificity, sensitivity, and AUC of the ANN model was 95%, 61% and 0.91 (95% confidence interval 0.71–0.96, $p < 0.001$), respectively (Figure 1).

Our study provides useful information on which routine laboratory variables determined at an early stage can predict a fatal outcome in patients with COVID-19. These findings are consistent with prior research conducted in smaller cohorts. For instance, Chen et al. also reported that higher levels of LDH and creatinine (the latter reflective of a lower GFR), leukocyte counts and aspartate aminotransferase were associated with higher mortality risk in a cohort of 274 patients with COVID-19 [8]. In line with our findings, Zhou et al. also observed that higher LDH and aspartate aminotransferase levels, as well as the presence of leukocytosis – which had a relatively large importance in our model (38.3%) – were associated with a

greater mortality risk [9]. However, these variables were not entered in their multivariate analysis. Less evidence is available on the predictive role of hemoglobin, which had high (44.1%) relative importance in our model but did not differ between deceased and surviving patients in the aforementioned study. In this regard, Zhou et al. also observed a greater prevalence of anemia in non-survivors (26%) compared to survivors (11%) [9], which would support our findings.

An important result of our study was that lower levels of albumin predicted a fatal outcome (with a high [72%] relative importance in the ANN model). This result is in agreement with the abovementioned Chen et al. study [8], as well as with recent findings in patients with COVID-19-associated nephritis [10] and with a study indicating that higher serum albumin might be associated with improved respiratory function [11]. There is in fact pre-clinical evidence that albumin can potentially down-regulate the expression of angiotensin-converting enzyme 2 (ACE2) [12] – the host receptor for COVID-19 – and albumin has been proposed as an outcome of future COVID-19 therapy trials [13].

While considering all the preanalytical, analytical and postanalytical issues that can affect interpretation [14], the present findings support the predictive role of some routine laboratory parameters (particularly LDH, GFR, albumin, and hemoglobin) for the early identification of “high-risk” patients with COVID-19. It must be however noted that in order to improve predictive accuracy these parameters should be considered in combination with a thorough examination of patients’ symptoms and clinical history (notably, comorbidities).

Research funding: None declared.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: The protocol was approved by the Ethics Committee of the aforementioned institution (reference #20/222) and adhered to the Declaration of Helsinki.

References

1. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200824-weekly-epi-update.pdf?sfvrsn=806986d1_4 [Accessed 27 Aug 2020].
2. <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-eighth-update>.
3. Lippi G, Plebani M. The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks. *Clin Chem Lab Med* 2020;58:1063–9.
4. Zhao X, Zhang B, Li P, Ma C, Gu J, Hou P, et al. Incidence, clinical characteristics and prognostic factor of patients with COVID-19: a systematic review and meta-analysis. *medRxiv* 2020:1–13. February 2019.
5. Lippi G, Plebani M. Laboratory abnormalities in patients with COVID-2019 infection. *Clin Chem Lab Med* 2020;58:1131–4.
6. Fong SJ, Dey N, Chaki J, editors. *Artificial intelligence for coronavirus outbreak*. Singapore: Springer Singapore; 2020.
7. Lezoray O. A neural network architecture for data classification. *Int J Neural Syst* 2001;11:33–42.
8. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ* 2020;368:m1091.
9. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
10. Gross O, Moerer O, Weber M, Huber TB, Scheithauer S. COVID-19-associated nephritis: early warning for disease severity and complications? *Lancet* 2020;395:e87–8.
11. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool to early predict severe corona virus disease 2019 (COVID-19) : a multicenter study using the risk nomogram in Wuhan and Guangdong, China. *Clin Infect Dis* 2020;71:833–40.
12. Liu BC, Gao J, Li Q, Xu LM. Albumin caused the increasing production of angiotensin II due to the dysregulation of ACE/ACE2 expression in HK2 cells. *Clin Chim Acta* 2009;403:23–30.
13. Mani Mishra P, Uversky VN, Nandi CK. Serum albumin-mediated strategy for the effective targeting of SARS-CoV-2. *Med Hypotheses* 2020;140:109790.
14. Kavsak PA, De Wit K, Worster A. Clinical chemistry tests for patients with COVID-19-important caveats for interpretation. *Clin Chem Lab Med* 2020;58:1142–3.