

The factor analysis approach to mortality prediction in COVID-19 severe disease using laboratory values: a retrospective study

Umut Kara¹, Fatih Şimşek¹, Mehmet Özgür Özhan², Mehmet Emin İnce¹, Gökhan Özkan¹, Serkan Şenkal¹, Ahmet Coşar¹

¹University of Health Sciences Turkey, Gülhane Training and Research Hospital, Clinic of Anesthesiology and Reanimation, Ankara, Turkey

²Çankaya Hospital, Clinic of Anesthesiology and Reanimation, Ankara, Turkey

Cite this article as: Kara U, Şimşek F, Özhan MÖ, et al. The factor analysis approach to mortality prediction in COVID-19 severe disease using laboratory values: a retrospective study. J Health Sci Med 2022; 5(2): 528-533.

ABSTRACT

Aim: Factor analysis is a statistical approach used mainly in social science scale development systems. The aim of this study was to evaluate the performance of factorial structures formed by laboratory values in predicting mortality in severe COVID-19 patients.

Material and Method: The study included 281 patients diagnosed with “severe coronavirus infection” according to the WHO COVID-19 clinical management guideline who were treated in a 13-bed adult tertiary-level critical care unit of a tertiary level hospital. For a total of 23 variables (ALT, AST, BUN, creatinine, Na, K, LDH, CRP, CK, ferritin, D-dimer, INR, TB, Glu, NLR, WBC, fibrinogen, % NEU, PLT, HTC, % LYM, TLC, Alb), laboratory values were collected. A two-step method was used to determine if exploratory factors might be used in place of laboratory variables. First, the ability of individual laboratory variables to predict mortality was obtained by analysis of the receiver operating characteristic (ROC) analysis. Then, the ability of factors created from these variables to predict mortality was measured using ROC analysis. The area under curve (AUC) values were compared between the two conditions.

Results: The Kaiser-Meyer-Olkin (KMO) value calculated using factor analysis on the variables was found to be 0.661. The significance level of the Bartlett's Test was <0.001. The correlation matrix determinant was found to be 0.001. CRP, ferritin, LDH, D-dimer, PLT, and TLC all had AUC values >0.6. A five-factor structure was created based on the Scree Plot. The fifth factor, which included CRP, fibrinogen, and ferritin, was the highest for predicting mortality (AUC: 0.677). According to the individual laboratory variables, the first factor comprising TLC, CK, and NLR, had the most remarkable success (AUC: 0,642).

Conclusions: The factor analysis approach can be used to present an alternative perspective for predicting mortality in COVID-19 critical disease. The factor including CRP, fibrinogen, and ferritin predicted mortality at the highest rate in this study.

Keywords: Coronavirus disease 2019, Covid-19, factor analysis, severe illness, mortality

INTRODUCTION

The mortality rate from coronavirus-19 (COVID-19) disease continues to increase (1). The etiology, prognostic factors, prevention, and treatment of the disease are all ongoing processes. The majority of research includes laboratory indicators, and particularly in COVID-19 disease, they are important studies performed using similar scientific procedures, in which each institution presents its own experience (2-4). Laboratory findings, such as severe lymphopenia and elevated D-dimer and ferritin levels; high C-reactive protein (CRP), lactate dehydrogenase (LDH), troponin and creatine kinase (CK) values have been related to severe COVID-19 illness (5-8).

It can be considered that the measurement methodologies of other scientific specialties might be beneficial in the COVID-19 pandemic. The factor analysis method is a multivariate data reduction method used mostly in the social sciences to identify fewer and unrelated variables (factors) by combining related variables (9). The goal of this analytic approach is to minimize the data set while keeping as much original data as reasonable.

The aim of the study was to evaluate the performance of factorial structures formed by laboratory values in predicting mortality in severe COVID-19 patients. A secondary aim of the study was to compare the performance of the factors for predicting mortality with the standard laboratory tests.

MATERIAL AND METHOD

This single center study was ethically approved by the University of Health Sciences Turkey, Gülhane Non-interventional Clinical Researches Ethics Committee (Project No: 2021/19, Date: 14.01.2021). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. As this was a retrospective study, informed consent was not required.

All patients aged ≥ 18 years admitted to the COVID-19 tertiary-level intensive care unit, from 29th March 2020 to 06th July 2021 were included in the study. Since March 2020, this 13-bed adult tertiary-level critical care unit has been dedicated to the admission of COVID-19 patients with "severe illness" who present to the hospital.

Patients who did not have "laboratory confirmed diagnosis of SARS-CoV-2 infection"; did not stay in the intensive care unit for at least 24 hours, had missing data in the hospital records, and were transferred to other intensive care units for whatever reason; were excluded from the study.

For every patient, demographic data (age, gender), laboratory tests and patient outcomes (length of ICU stay (days), and 28-day mortality) were recorded. On a total of 23 variables, laboratory values were collected. The laboratory tests on the first day of admission to the hospital were analysed. Medical data of the patients were retrieved from the digital medical records. The laboratory tests included: alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), creatinine, sodium (Na), potassium (K), lactate dehydrogenase (LDH), C-reactive protein (CRP), creatine kinase (CK), ferritin, D-dimer, international normalized ratio (INR), total bilirubin (TB), glucose (Glu), neutrophil/lymphocyte ratio (NLR), white blood cells (WBC), fibrinogen, % neutrophils (NEU), platelet count (PLT), hematocrit (HTC), % lymphocyte (LYM), total lymphocyte count (TLC), and albumin (Alb).

According to the WHO guidance, "laboratory confirmed diagnosis of SARS-CoV-2 infection" was defined as a positive result of RT-PCR assay of nasal and pharyngeal swabs. "Severe illness" is characterized by clinical signs of pneumonia (fever, cough, dyspnoea, fast breathing) plus one of the following: respiratory rate > 30 breaths/min; severe respiratory distress; or SpO₂ $< 90\%$ on room air (10).

Statistical Analysis

Categorical variables are reported as number (%). Continuous data were reported as mean \pm standard deviation (SD). Categorical data were compared using the χ^2 test. The Mann-Whitney U-test was used to compare non-normally distributed continuous data. The

hypothesis of a sample fit coefficient Kaiser-Meyer-Olkin (KMO) > 0.60 was used to measure the sample size fit to see if the data were acceptable for exploratory factor analysis. To evaluate whether or not there was a linear relationship between laboratory values, a correlation matrix was created. The oblimin technique, which is one of the factor rotation methods, was used in the analysis to achieve a homogenous equilibrium by ranking the independent and created factors.

To determine if exploratory factors could be employed instead of laboratory variables, the following method was performed. First, the performance of individual laboratory variables to predict mortality was determined using a receiver operating characteristic curve (ROC) analysis. Scree plot and the total variance explained table were used to determine the number of factors. The loads of laboratory variables in the factor model were measured using a rotated component matrix. Depending on the sample size, the lower power limit of the factor was determined to be 0.4. The performance of the newly developed factors to predict mortality was then determined using ROC analysis. Statistical analysis of the collected data was conducted using IBM SPSS software version 25 (SPSS Inc., Chicago, IL, USA) (11). The level of statistical significance was accepted as 0.05 in all analyses.

RESULTS

During the study period, 451 severe ill patients were admitted to the ICU, and 281 patients with laboratory confirmed diagnosis of SARS-CoV-2 infection results were analysed. 168 patients had died, corresponding to a mortality rate of 59.8%. The majority of critically ill patients with COVID-19 were males (63%). The mean age was 67.6 years. The median length of ICU stay day was 9. There was a statistically significant difference between survivors and non-survivors in terms of age and length of ICU stay ($p < 0.005$) (Table 1).

Table 1. Demographic characteristics of patients (N=281)

Characteristics	Survivors N=113	Nonsurvivors N=168	P value
Female, n (%)	47 (45.2%)	57 (54.8%)	0.192*
Male, n (%)	66 (37.3%)	111 (62.7%)	0.192*
Age (years), (mean \pm SD)	62.26 \pm 14.25	71.2 \pm 12.03	$< 0.001^{**}$
ICU LOS (days), (mean \pm SD)	10.56 \pm 9.24	11.71 \pm 8.27	0.009**

Abbreviations: ICU: Intensive Care Unit, LOS: Length of stay, SD: Standard deviation, *Pearson Chi-Square test, **Mann-Whitney U test

In the factor analysis applied to the variables, the KMO value was found to be 0.661. The significance level of Bartlett's test was < 0.001 . These findings indicated that

there were strong correlations between some variables, indicating that the data was suitable for the factor analysis approach. The majority of the correlation coefficients between laboratory variables in the correlation matrix ranged from 0.30 to 0.80. The binary combination with the highest positive correlation coefficient in the correlation matrix was ALT and AST (correlation coefficient:0.844). Other strongly positive related binary combinations were: BUN and creatinine (coefficient: 0.746); PLT and WBC (coefficient: 0.609); % NEU and NLR (coefficient: 0.516), CK and AST (coefficient: 0.535). The binary combinations with the highest negative correlation coefficient in the correlation matrix were: % NEU and % LYM (coefficient:-0.711) and % LYM and NLR (coefficient: -0.582). The correlation matrix determinant was found to be 0.001.

Table 2 presents the area under the curve (AUC) values and the related confidence intervals demonstrating the performance of the standard laboratory variables by the ROC analysis in predicting mortality. The variables were found to predict mortality at varying rates. The AUC values of CRP, ferritin, LDH, D-dimer, PLT and TLC were found to be > 0.6.

Table 2. Area under curve (AUC) of variables						
Variables	AUC	%95 CI		P value	Sensitivity (%)	Spesificitiy (%)
		Lower limit	Upper limit			
ALT	0.466	0.396	0.536	0.352	47	34
AST	0.496	0.425	0.566	0.902	49	30
BUN	0.502	0.428	0.576	0.000	63	45
Creatinine	0.539	0.469	0.610	0.000	58	50
Na	0.500	0.429	0.572	0.256	49	39
K	0.460	0.388	0.531	0.159	26	46
CRP	0.620	0.581	0.659	0.024	63	58
CK	0.455	0.385	0.525	0.215	46	38
Ferritin	0.605	0.564	0.646	0.009	62	59
LDH	0.603	0.534	0.673	0.017	58	55
D-dimer	0.616	0.578	0.654	0.008	63	53
INR	0.579	0.509	0.649	0.031	58	56
TB	0.559	0.489	0.630	0.001	55	54
Glu	0.482	0.411	0.553	0.630	59	50
NLR	0.582	0.530	0.634	0.622	53	43
WBC	0.432	0.362	0.503	0.064	40	45
Fibrinogen	0.497	0.429	0.566	0.006	58	48
% NEU	0.463	0.393	0.533	0.430	54	52
PLT*	0.601	0.552	0.650	0.023	61	58
HTC*	0.534	0.465	0.603	0.111	30	45
% LYM*	0.558	0.488	0.628	0.152	51	39
TLC*	0.617	0.549	0.686	0.031	63	60
Alb*	0.577	0.507	0.646	0.306	54	42

* Low values are associated with mortality, ALT:alanine aminotransferase, AST:aspartate aminotransferase, BUN:blood urea nitrogen, Na:sodium, K:potassium, LDH:lactate dehydrogenase, CRP:C-reactive protein, CK:creatine kinase, INR:international normalized ratio, TB:total bilirubin, Glu:glucose, NLR:neutrophil/lymphocyte ratio, WBC:white blood cells, NEU:% neutrophils, PLT:platelet count, HTC:hematocrit, LYM:% lymphocyte, TLC:total lymphocyte count, ALB:albumin, AUC:Area under curve, CI:Confidence Interval

The number of factors that may be created using the variables when factor analysis was applied to continuous variables in the data set is shown in the Scree plot diagram in **Figure 1**. When the slope of the Scree plot generated for seventeen variables was evaluated, the slope changes up to the fifth factor, after which it forms a stable plateau. Consequently, the factors to be considered after the fifth factor will be meaningless.

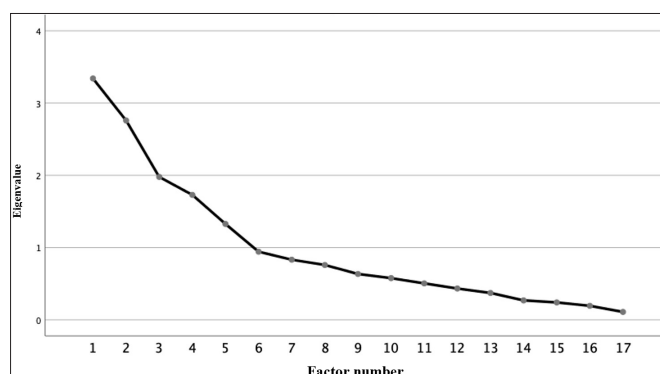


Figure 1. Scree plot diagram

Table 3 shows the number of components that may be constructed, their eigenvalues, and the total variance explained percentages. The five-factor structure accounted for 65.484% of the total variation. With a 19.648% explanatory rate among the total variance explained, the first factor is the most influencing factor structure. In the five-factor structure, there were problematic laboratory results for albumin, INR, glucose, and D-dimer, so these were removed from the analysis. **Table 4** shows the rotated component matrix obtained after applying the transformation, as well as the loads of variables.

Table 5 shows the ROC analysis of the factors and the variables, as well as the values of the AUC and related confidence intervals. The performance of the classification of the first and fifth factors showed the highest increase in respect of the individual performance of the variables in the ROC analysis. ROC analysis curve of fifth factor was presented in **Figure 2**. While the second factor's performance increased as the levels of AST, ALT, and % NEU increased, it performed less than the LDH value. The third factor had the lowest performance. The fourth factor's AUC was not considered successful because it was <0.6.

Table 3. Total variance explained			
Factor	Total	% of Variance	% Cumulative
1	3.340	19.648	19.648
2	2.759	16.229	35.877
3	1.978	11.635	47.512
4	1.729	10.169	57.681
5	1.327	7.803	65.484

*Values obtained by rotation of oblimin technique

Table 4. Rotated component matrix

	Factors				
	1	2	3	4	5
TLC	-.882	-.044	.061	-.016	-.014
CK	.856	-.009	.023	.030	-.062
NLR	.747	-.088	-.041	.008	.183
AST	-.085	.934	-.053	-.082	.015
ALT	-.109	.849	-.031	-.039	-.035
% NEU	.029	.653	-.052	-.022	-.034
LDH	.115	.624	.056	.168	.111
Creatinine	.035	.086	-.873	.018	.044
BUN	.102	.106	-.863	-.007	-.073
HTC	.229	.207	.505	.177	-.263
K	.117	.041	-.440	.248	-.332
PLT	.096	-.001	.022	.798	.098
% LYM	-.397	-.021	.006	.786	-.080
WBC	.317	.000	-.046	.729	.182
CRP	.174	.036	-.086	.097	.795
Fibrinogen	.119	-.106	.076	.301	.619
Ferritin	-.091	-.398	.012	-.150	.574

ALT:alanine aminotransferase, AST:aspartate aminotransferase, BUN:blood urea nitrogen, Na:sodium, K:potassium, LDH:lactate dehydrogenase, CRP:C-reactive protein, CK:creatinine kinase, INR:international normalized ratio, TB:total bilirubin, Glu:glucose, NLR:neutrophil/lymphocyte ratio, WBC:white blood cells, NEU:% neutrophils, PLT:platelet count, HTC:hematocrit, LYM:% lymphocyte, TLC:total lymphocyte count, ALB:albumin, AUC:Area under curve, CI:Confidence Interval

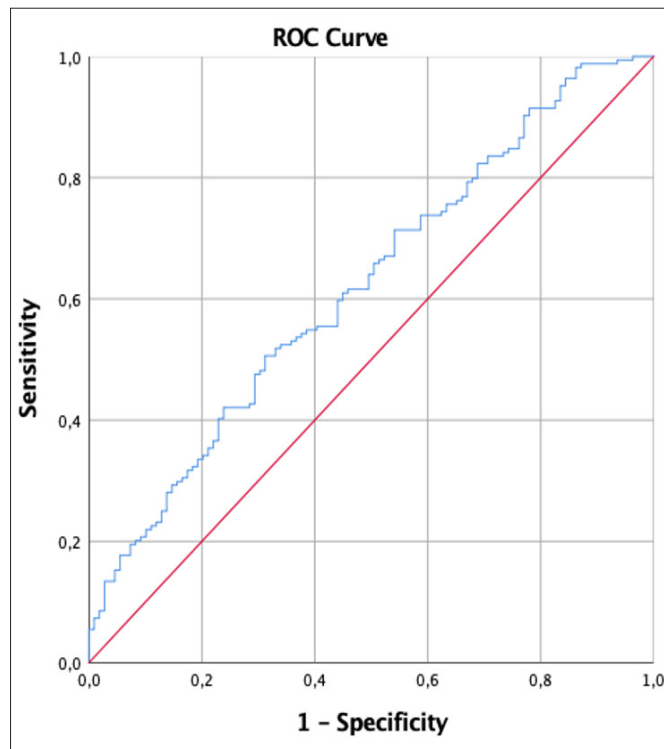


Figure 2. ROC analysis curve of 5th factor

Table 5. Area under curve (AUC) of Factors

Factors	Variables	AUC of variables (%95 CI)	AUC of factors (%95 CI)	P value (factors)	Sensitivity (%)	Spesificitiy (%)
1	TLC	0.617 (0.549-0.686)	0.642 (0.601-0.683)	0.032	58	56
	CK	0.455 (0.385-0.525)				
	NLR	0.582 (0.530-0.634)				
2	AST	0.496 (0.425-0.566)	0.554 (0.503-0.605)	0.068	48	45
	ALT	0.466 (0.396-0.536)				
	% NEU	0.463 (0.393-0.533)				
	LDH	0.603 (0.534-0.673)				
3	Creatinine	0.539 (0.469-0.610)	0.397 (0.329-0.465)	0.296	43	37
	BUN	0.502 (0.428-0.576)				
	HTC	0.534 (0.465-0.603)				
	K	0.460 (0.388-0.531)				
4	PLT	0.601 (0.552-0.650)	0.497 (0.426-0.567)	0.305	45	41
	% LYM	0.558 (0.488-0.628)				
	WBC	0.432 (0.362-0.503)				
5	CRP	0.620 (0.581-0.659)	0.677 (0.640-0.814)	0.007	61	59
	Fibrinogen	0.497 (0.429-0.566)				
	Ferritin	0.605 (0.564-0.646)				

ALT:alanine aminotransferase, AST:aspartate aminotransferase, BUN:blood urea nitrogen, Na:sodium, K:potassium, LDH:lactate dehydrogenase, CRP:C-reactive protein, CK:creatinine kinase, INR:international normalized ratio, TB:total bilirubin, Glu:glucose, NLR:neutrophil/lymphocyte ratio, WBC:white blood cells, NEU:% neutrophils, PLT:platelet count, HTC:hematocrit, LYM:% lymphocyte, TLC:total lymphocyte count, ALB:albumin, AUC:Area under curve, CI:Confidence Interval

DISCUSSION

In this study, the performance of the factorial structure derived by laboratory values at the time of hospital admission was researched to predict the mortality of patients with COVID-19 severe disease. The factor, which included CRP, fibrinogen, and ferritin, had the highest performance level of all the factors in predicting mortality (AUC:0.677).

In the individual assessment of laboratory parameters, CRP, ferritin, D-dimer, PLT, and TLC were found to be more successful in predicting mortality than other variables, but were usually weaker indicators. The performance of these variants is in accordance with the literature on COVID-19 (12-15). However, the AUC of the variables in this and other similar studies in the literature varies. The severity of disease at the time of

hospitalization and the variety of statistical methods used are two possible explanations for these variations. In our study, severe illness was defined according to WHO COVID 19 clinical management guideline.

TLC, CK, and NLR all contributed the same factor (1st factor) structure to the five-factor structure. The AUC value of this factor was also found to be >0.6. Individually, deep and prolonged lymphopenia and high NLR have been shown to be poor prognostic indicators in severe COVID-19 disease (16,17). The factors can also be analyzed in mortality prediction models, according to their structure.

In severe COVID-19 disease, the immune response is not controlled and severe inflammation results in the release of cytokines and ARDS (18). Inflammatory cytokines such as CRP, ferritin, and IL-6 are released in greater quantities. In factor 5 in this study, CRP showed the best mortality prediction performance when combined with fibrinogen and ferritin. A consistent result is the presence of inflammatory cytokines in the same factor structure.

The concept underlying factor analysis is that there are many inter-relationships between variables (19). The data and dimensions are reduced using factor analysis by first decreasing the interdependent structures. In this regard, the relationship between the variables is investigated prior to regression analysis in the COVID-19 mortality prediction models, and if many inter-relationships are identified, a factor analysis method can also be used. In this study, a mortality prediction model was established using only laboratory variables. Many classic models, including medical history, demographics, scores, and other radiological and clinical factors, are available in the literature (20,21). The fact that other non-laboratory variables were not included in the model may be one of the reasons why the AUC values of the first and fifth factorial structures were higher than 0.6 but still low.

With the exception of scale development, only one study on COVID-19 and factor analysis was found in the literature, and that study also focused on symptom classification (22). Laboratory data were employed to predict mortality in COVID-19 severe disease in this study to determine if a different application of factor analysis might improve diagnostic classification performance.

Limitations

Limitations of this study were single center and retrospective study design, the use of only laboratory measurements, the inclusion of only patients treated in the critical care unit, and the use of laboratory variables at the time of hospital admission.

CONCLUSION

Factor analysis using laboratory and other indications may be a better predictor of mortality in severe COVID-19 disease than using these markers alone. The factor analysis method might be utilized to provide a different viewpoint on mortality prediction. The highest mortality rate was predicted by a factor structure that included CRP, fibrinogen, and ferritin. Our study is a first in this area in terms of design, because the model includes all laboratory markers, not only those found in the literature related with a poor prognosis in severe COVID-19 disease. In the next stage, an improved model incorporating categorical data linked to a poor prognosis can be established.

ETHICAL DECLARATIONS

Ethics Committee Approval: This single center study was ethically approved by the University of Health Sciences Turkey, Gülhane Noninterventional Clinical Researchs Ethics Committee (Project No: 2021/19, Date: 14.01.2021).

Informed Consent: Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process: Externally peer-reviewed.

Conflict of Interest Statement: The authors have no conflicts of interest to declare.

Financial Disclosure: There is no financial source of this study

Author Contributions: All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. World Health Organization (WHO). Coronavirus Disease (COVID-19) Dashboard. WHO. Accessed date: September 26 2020; Available from: <https://covid19.who.int/>, (n.d.).
2. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; 369: m1328
3. Doganci S, Ince ME, Ors N, et al. A new COVID-19 prediction scoring model for in-hospital mortality: experiences from Turkey, single center retrospective cohort analysis. *Eur Rev Med Pharmacol Sci* 2020; 24: 10247-57.
4. Sperrin M, Grant SW, Peek N. Prediction models for diagnosis and prognosis in Covid-19. *BMJ* 2020; 369: m1464.
5. 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 Intern Med* 2020; 180: 934-43.
6. Güngörer B. Baseline demographic, clinical and laboratory risk factors for predicting admission to intensive care unit in patients diagnosed with COVID-19 in the emergency department. *Anatolian Curr Med J* 2021; 3: 279-83.

7. Yıldırım Ö, Bayram M, Özmen RS, et al. Evaluation of hematological indices in terms of COVID-19 related mortality and ICU admission. *J Health Sci Med* 2021; 4: 666-9.
8. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020; 323: 1061-9.
9. Tavşançıl E. Tutumların ölçülmesi ve SPSS ile veri analizi. Ankara, Nobel Yayın 2002
10. World Health Organization. Clinical management of COVID-19: living guidance. 25.01. 2021. WHO/2019-nCoV/clinical/2021.1
11. IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.
12. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. *Clinica Chimica Acta* 2020; 506: 145-8.
13. Kırıcı Berber N, Ulutaş Ö, Sarıcı A, et al. Clinical features and laboratory values associated with disease severity in Covid 19 patients: a single center experience. *J Health Sci Med* 2021; 4: 18-22.
14. Antunez Muiños PJ, López Otero D, Amat-Santos IJ, et al. The COVID-19 lab score: an accurate dynamic tool to predict in-hospital outcomes in COVID-19 patients. *Scientific Reports* 2021; 11: 9361.
15. Arslan K, Baş S. Frequency of troponin elevations in patients with COVID-19 and clinical course in these patients. *Anatolian Curr Med J* 2022; 4: 95-102.
16. Huang I, Pranata R. Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis. *J Intensive Care* 2020; 8: 36.
17. Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. *J Transl Med* 2020; 18: 206.
18. García LF. Immune response, inflammation, and the clinical spectrum of COVID-19. *Frontiers in immunology* 2020; 11: 1441.
19. Büyükoztürk Ş. Faktör analizi: temel kavramlar ve ölçek geliştirmede kullanımı. kuram ve uygulamada eğitim yönetimi. 2001; 32: 470-83.
20. Leoni MLG, Lombardelli L, Colombi D, et al. Prediction of 28-day mortality in critically ill patients with COVID-19: development and internal validation of a clinical prediction model. *PLoS One* 2021; 16: e0254550.
21. Altschul DJ, Unda SR, Benton J, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. *Scientific Reports* 2020; 10: 16726.
22. Luo Y, Wu J, Lu J, et al. Investigation of COVID-19-related symptoms based on factor analysis. *Ann Palliat Med* 2020; 9: 1851-8.