Research article

Predicting outcomes of the acute phase of COVID-19. High sensitive prognostic model, based on the results of the international registry "analysis of chronic non-infectious diseases dynamics after COVID-19 infection in adult patients" (ACTIV)

Gregory P. Arutyunov [a,b], Ekaterina I. Tarlovskaya [a,c], Dmitry S. Polyakov [c], Tatiana I. Batluk [a,*], Alexander G. Arutyunov [a,d]

[a] Eurasian Association of Internal Medicine, Moscow, Russia
[b] Department of Propaedeutics of Internal Diseases (Pediatric School), Pirogov Russian National Research Medical University, Moscow, Russia
[c] Department of Therapy and Cardiology, Privolzhsky Research Medical University, Nizhny Novgorod, Russia
[d] Department of Cardiology and Internal Medicine, National Institute of Health named after Academician S. Avdalbekyan, Yerevan, Armenia

ABSTRACT

The **aim** of this study is to investigate the course of the acute period of COVID-19 and devise a prognostic scale for patients hospitalized.
Materials and methods: The ACTIV registry encompassed both male and female patients aged 18 years and above, who were diagnosed with COVID-19 and subsequently hospitalized. Between June 2020 and March 2021, a total of 9364 patients were enrolled across 26 medical centers in seven countries. Data collected during the patients' hospital stay were subjected to multivariate analysis within the R computational environment. A predictive mathematical model, utilizing the "Random Forest" machine learning algorithm, was established to assess the risk of reaching the endpoint (defined as in-hospital death from any cause). This model was constructed using a training subsample (70% of patients), and subsequently tested using a control subsample (30% of patients).
Results: Out of the 9364 hospitalized COVID-19 patients, 545 (5.8%) died. Multivariate analysis resulted in the selection of eleven variables for the final model: minimum oxygen saturation, glomerular filtration rate, age, hemoglobin level, lymphocyte percentage, white blood cell count, platelet count, aspartate aminotransferase, glucose, heart rate, and respiratory rate. Receiver operating characteristic analysis yielded an area under the curve of 89.2%, a sensitivity of 86.2%, and a specificity of 76.0%. Utilizing the final model, a predictive equation and nomogram (termed the ACTIV scale) were devised for estimating in-hospital mortality amongst COVID-19 patients.
Conclusion: The ACTIV scale provides a valuable tool for practicing clinicians to predict the risk of in-hospital death in patients hospitalized with COVID-19.

* Corresponding author
  E-mail address: tatianabatluk92@gmail.com (T.I. Batluk).

Abbreviations

ACS	acute coronary syndrome
AF	atrial fibrillation
AST	aspartate aminotransferase
AUC	area under the curve
CAD	coronary artery disease
CHF	chronic heart failure
COPD	chronic obstructive pulmonary disease
CKD	chronic kidney disease
CRF	electronic case report forms
CRP	*C*-reactive protein
CT	computed tomography
FC	functional classes
GFR	glomerular filtration rate
HR	heart rate
ICU	intensive care unit
NYHA	New York Heart Association
OR	odds ratio
RR	respiratory rate
ROC	receiver operating characteristic
SpO2	blood oxygen saturation
T2DM	type 2 diabetes mellitus
WBC	white blood cell

1. Introduction

The COVID-19 pandemic spread worldwide, resulting in billions of patients requiring hospitalization, with high mortality rates, especially among vulnerable subgroups. Despite new viral mutations with higher transmission rates and antibody evasion capacity but lower symptom severity, some cohorts of initially multimorbid patients remain at higher risk for hospitalization and death during hospital stay.

In the last two years, a number of risk scales for the new infection have been developed or adopted, but most of them still lack real clinical evidence or have low specificity or sensitivity.

This work addresses the needs of the medical community involved in the management of hospitalized COVID-19 patients. A multivariate analysis of a large international database of hospitalized patients with confirmed SARS-CoV-2 infection resulted in a highly sensitive prognostic scale for the acute phase of COVID-19 infection. The scale has been tested on various strains of SARS-CoV-2 and although the covid-19 pandemic is over, new mutations of the virus cannot be ruled out, so the scale can be used worldwide in decision-making. The aim of this study is to investigate the course of the acute period of COVID-19 and devise a prognostic scale for patients hospitalized.

2. Materials and methods

The study received ethical approval from the Ethics Committee of the Pirogov Russian National Research Medical University of the Ministry of Health of the Russian Federation and was registered in the ClinicalTrials.gov database as "Analysis of Chronic Non-infectious Diseases Dynamics After COVID-19 Infection in Adult Patients (ACTIV)", NCT04492384 and ACTIV 2, NCT04709120. More information about the registry can be found on the Eurasian Association of Internal Medicine's website: https://ACTIV.euat.ru. The register was organized and controlled by 3 committees: organizational, supervisory and endpoint analysis and individual registration cards (IRC) committees. The IRCs and document flow are electronic only.

The ACTIV registry encompassed hospitalized male and female patients aged 18 years or older diagnosed with COVID-19. The diagnosis was confirmed either by a positive smear test (PCR), significant immunoglobulin M titer to the SARS-CoV-2 virus, or characteristic computed tomography findings. All data were anonymized. Between June 29, 2020, and March 30, 2021, a total of 9364 patients were enrolled in the ACTIV registry from 26 medical centers in seven countries, including the Republic of Armenia, the Republic of Belarus, the Republic of Kazakhstan, the Kyrgyz Republic, the Republic of Moldova, Russian Federation, and the Republic of Uzbekistan [1]. Data were collected by the researchers' physicians, who were the attending physicians of these patients, throughout their hospital stay using electronic forms of medical histories (CRFs). Each CRF was monitored by clinical research associates, with diagnoses based on the International Classification of Diseases 10 criteria.

2.1. Statistical analysis

Descriptive statistics for quantitative variables were calculated after testing for normal distribution using the Shapiro-Wilk or Kolmogorov-Smirnov tests. If the distribution did not significantly deviate from normal, the mean value and standard deviation (M ± σ) were used to describe central tendency and dispersion. Conversely, if the distribution significantly deviated from normal, the median and quartiles (Me [Q1; Q3]) were employed.

The odds ratio (OR) and its 95% confidence interval were calculated using the univariate binary logit regression method. Multivariate analysis was conducted in the R environment [2].

During the first stage of the analysis, a mathematical model was constructed using the machine learning algorithm, Random Forrester [3], to predict the risk of reaching the endpoint (death in the hospital from any cause). The model resulted in erroneous classification in only 0.09% of cases in the control sample ("out-of-bag sample").

The importance of each variable incorporated in the analysis was assessed by the mean decrease in the Gini index during the implementation of the Random Forest algorithm, utilizing permutation tests. The patient's clinical status at the time of hospitalization significantly influences the short-term prognosis of patients with COVID-19, and this status can diverge considerably among patients presenting with the same comorbidities. To enhance the precision of the prognostic model, a subsequent stage of multivariate analysis was performed, which included all patient characteristics: gender, age, comorbidities, degree of multimorbidity, and laboratory and instrumental examination data associated with an increased risk of mortality, based on the outcomes of the univariate analysis. To refine the model, the second stage of multivariate analysis incorporated all patient characteristics. Eleven variables were selected (minimum SpO2, GFR, age, hemoglobin level, percentage of lymphocyte count, WBC count, platelet count, AST level, glucose, HR, RR) by implementing the "random forest" algorithm alongside permutation tests. These variables exhibited the most significant influence on the variability of the dependent variable (hospital mortality from any cause).

For the second stage, the registry sample was randomly divided into two subsamples: a training subsample (70% of observations) and a control subsample (30% of observations). A logit model was constructed using the training subsample, with hospital death as the dependent variable and the variables selected during the first stage as predictors.

The area under the curve (AUC), along with the model's sensitivity and specificity, were estimated through receiver operating characteristic (ROC) analysis using both the training and control subsamples. A nomogram for the final multivariate model was created using the Regression Modeling Strategies library [4,5]. A variable's effect on prognosis was considered statistically significant if the p-value was less than 0.05.

3. Results

3.1. Baseline characteristics of patients

Among the 9364 hospitalized patients with COVID-19, 53% were women. The mean age was 59.0 [48.0; 68.0] years (Supplementary Table 1). Patients' conditions were assessed based on the extent of lung tissue damage on computed tomography (CT) scans, graded on a scale of 1–4, with each grade representing a 25% increment of damage. Patients were distributed as follows: 7.4% had grade 4 (over 75% damage), 38.9% had grade 3 (50–75% damage), 33.7% had grade 2 (25–50% damage), 15.7% had grade 1 (under 25% damage), and 4.4% were at grade 0. The majority of patients (69%) had comorbidities, with 11.6% having four or more concurrent conditions, 32.3% having 2–3, and 25.1% having only one concurrent disease (Supplementary Table 2). The most common diagnoses were hypertension (56.6% of patients), obesity (34.8%), and coronary artery disease (CAD) (23.0%) (Supplementary Table 1).

The most common comorbidity combinations were hypertension with obesity (22.8%); hypertension, CAD, and chronic heart failure (CHF) (10.8%); and hypertension, CAD, obesity, and CHF (3.3%) (Supplementary Table 2).

Of the patients, 48% developed complications (Supplementary Table 3), which increased the risk of death. The most frequent complications included cytokine storm – 3312 (36.0%), bacterial pneumonia - 1165 (12.7%), and acute kidney injury - 1090 (11.8%). Less common complications included sepsis - 92 (1.0%) patients, myocarditis - 76 (0.8%) patients, pulmonary embolism - 52 (0.6%), stroke - 47 (0.5%), acute coronary syndrome (ACS) – 40 (0.4%), deep vein thrombosis - 28 (0.3%) patients. Thrombotic complications were diagnosed in 167 (1.8%) patients.

Out of the 9364 hospitalized patients with COVID-19, 545 (5.8%) died. The primary causes of death were acute respiratory distress syndrome - 165 (30.3%), acute heart failure - 115 (21.1%), multiple organ failure - 71 (13.0%), septic shock - 71 (13.0%), ACS - 40 (7.34%), pulmonary embolism - 26 (4.77%), stroke - 25 (4.59%), fatal heart rhythm disturbances −11 (2.02%), acute renal failure - 6 (1.10%), disseminated intravascular coagulation syndrome - 6 (1.10%), and myocarditis - 1 (0.18%).

3.2. Development of a nomogram

A comparative analysis was conducted between demographic, instrumental, and clinical data of surviving and deceased patients. The comparison results are tabulated in Supplementary Table 1 and Supplementary Table 2. Deceased patients were found to be significantly older than survivors (71.0 [63.0; 80.0] vs. 59.0 [47.0; 67.0] years, p < 0.001). A one-year increase in age corresponded to an 8% increase in death risk from the age of 40 (Supplementary Table 1). Deceased patients were also observed to have higher incidences of grade 3 and 4 lung injury according to computed tomography (CT) findings, this association statistically significant with an increased risk of death.

Laboratory parameters from surviving and deceased COVID-19 patients showed that a decrease in blood oxygen saturation (SpO2), lymphocytes, platelets, total cholesterol, low density lipoprotein cholesterol levels, and glomerular filtration rate (GFR) increased the risk of death. Additionally, an increase in heart rate (HR), respiratory rate (RR), white blood cell (WBC) count, *C*-reactive protein (CRP), D-dimer, and aspartate aminotransferase (AST) levels, and glucose levels in both type 2 diabetes mellitus (T2DM) and non-T2DM patients were associated with increased mortality risk (Supplementary Table 4).

Comparative analysis of surviving and deceased hospitalized patients with COVID-19 depending on the drugs prescribed before hospitalization for concomitant diseases showed that the use of the following drugs immediately before hospitalization was associated with a decrease in the risk of death: statin treatment for CAD, angiotensin receptor blocker for CAD, hypertension, CHF, ticagrelol/prasugrel/clopidogrel for CAD, oral anticoagulants for atrial fibrillation (AF) (mainly direct oral anticoagulants), and oral glucose-lowering medications for T2DM (Supplementary Table 5).

Considering the impact of comorbidities on death risk, the OR for a fatal outcome was highest for New York Heart Association (NYHA) class III-IV CHF, AF, and a history of stroke. Other factors associated with an increased risk of death (in descending order of significance) were CAD, chronic kidney disease (CKD), NYHA class I-II CHF, hypertension, anemia, chronic obstructive pulmonary disease (COPD), T2DM, cancer, and obesity (Supplementary Table 1).

In comparison of the combinations of comorbidities in deceased and surviving patients, the combination of hypertension and CAD (15.2% of patients) was most associated with death risk (OR 4.19 [3.49; 5.03] p = 0.0001). Other significant combinations were hypertension + CAD + CHF (10.8% of patients, OR 4.45 [3.65; 5.40] p = 0.0001) and hypertension + CAD + CHF + T2DM (3.5% of patients, OR 4.91 [3.68; 6.48] p = 0.0001) (Supplementary Table 2). A univariate binary logistic regression demonstrated that various comorbidities, in numerous combinations, are associated with an increased risk of mortality in hospitalized patients, in addition to factors such as age and gender. The initial phase of the multivariate analysis elucidated the comorbidities that are most strongly correlated with an increased mortality risk.

In the first stage of multivariate analysis, a comprehensive review of demographic characteristics, concurrent diseases, and the extent of comorbidity (Table 1) revealed a statistically significant association between increased risk of mortality and factors such as advanced age, male gender, a history of stroke, NYHA class III-IV CHF, AF, prior cancer, and T2DM. A predictive model was established using the entirety of the dataset that encapsulated all these variables. A ROC analysis was executed to ascertain the predictive prowess of the model, incorporating age, gender, and comorbidities, and exhibited a substantial predictive value for this model (Fig. 1A). The model, constructed using the training subset (comprising 70% of patients), displayed an area AUC of 81.03%, with sensitivity and specificity equating to 73.6% and 76.1% respectively at the optimal cut-off point (Youden point) with a linear predictor value of −2.82. The model was subsequently evaluated using the control subset (consisting of 30% of the patients) (Fig. 1 B). The model derived from the control subset also demonstrated a substantial predictive value: the area under the curve was 81.0%, the sensitivity was 87.5%, and the specificity was 58.4%. Therefore, the multivariate analysis concluded that the presence of concurrent diseases such as T2DM, cancer, NYHA class III-IV CHF, and a history of stroke adversely influenced patient prognosis.

Fig. 2 visually delineates the importance of each variable as measured by the mean decrease in the Gini index. A final logistic regression model was then constructed, utilizing the selected 11 variables. Hospital mortality served as the dependent variable within this model, with the selected 11 variables functioning as predictors (Table 2). A ROC analysis was subsequently conducted to evaluate the predictive efficacy of this final model. The model, established using the training subsample (comprising 70% of patients), displayed

Table 1
Analysis of multivariate model variables that included comorbidities and gender data to predict the risk of in-hospital mortality in patients with COVID-19.

Variable	Exp (β)	2.5 %	97.5 %	p
Hypertension	0.936	0.613	1.448	0.763
History of CVA	**2.189**	**1.469**	**3.229**	**0.000**
T2DM	**1.410**	**1.022**	**1.939**	**0.035**
CKD	1.392	0.965	1.992	0.074
CAD	1.217	0.831	1.778	0.311
History of MI	0.936	0.627	1.388	0.744
AF	**1.858**	**1.305**	**2.632**	**0.001**
History of cancer	**1.572**	**1.004**	**2.411**	**0.042**
COPD	1.087	0.648	1.762	0.744
Asthma	0.960	0.466	1.809	0.906
Obesity	1.110	0.799	1.542	0.533
Age	**1.059**	**1.046**	**1.073**	**0.000**
Men	**1.587**	**1.229**	**2.052**	**0.000**
NYHA class I-II CHF	1.043	0.739	1.461	0.808
NYHA class III-IV CHF	**2.148**	**1.459**	**3.141**	**0.000**
1 condition	1.197	0.626	2.353	0.593
2-3 conditions	1.368	0.612	3.108	0.449
≥4 conditions	1.668	0.548	5.116	0.369

AF, atrial fibrillation; CAD, coronary artery disease; CHF, chronic heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; MI, myocardial infarction; NYHA, New York Heart Association; T2DM, type 2 diabetes mellitus.
β, coefficient of the regression equation; Exp (β), exponent of the coefficient of the regression equation; p, level of statistical significance for the coefficient of the regression equation; margins of 95% confidence interval (the lower is 2.5% and the upper is 97.5%).
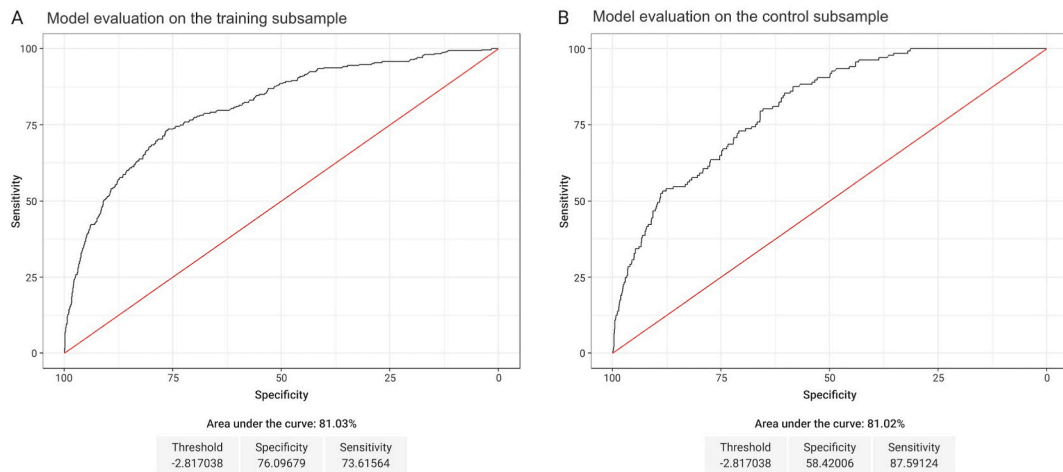
Fig. 1. Analysis of multivariate model that included comorbidities and gender data to predict the risk of in-hospital mortality in patients with COVID-19. ROC analysis of multivariate model that included comorbidities and gender data to predict the risk of in-hospital mortality in patients with COVID-19. The model included the following variables: age, male gender, history of strike, NYHA class III-IV CHF, AF, history of cancer, T2DM. A – Model evaluation on the training subsample
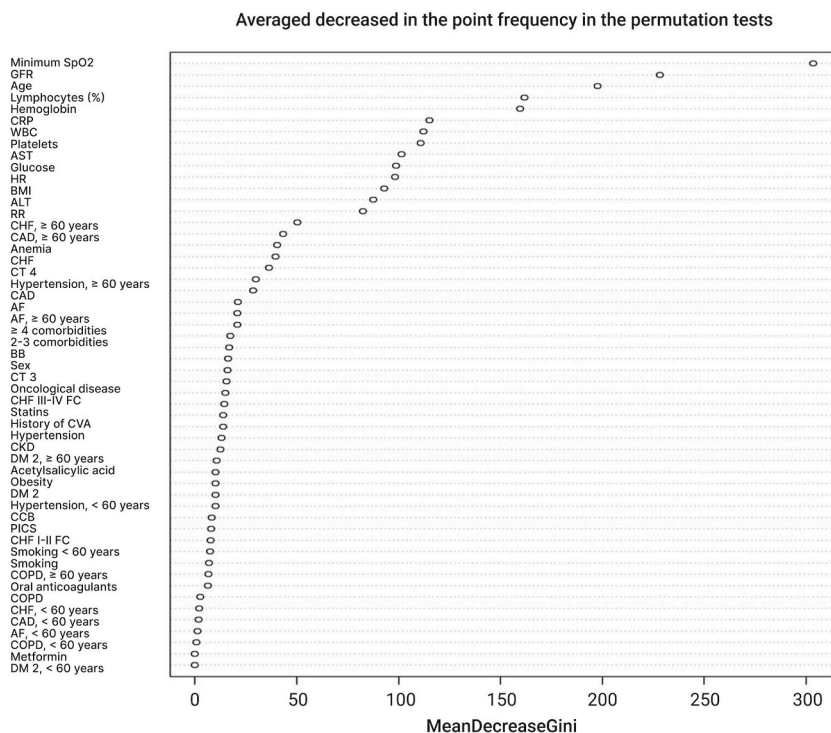B - Model evaluation on the control subsample.

Fig. 2. Ranking of variables by the average decrease in the Gini index. Importance of each included variable was based on the averaged value of Gini index. AF, atrial fibrillation; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BB, beta-blockers (administered before hospitalization); BMI, body mass index; CAD, coronary artery disease; CCB, calcium channel blockers (administered before hospitalization); CHF, chronic heart failure; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CRP, *C*-reactive protein; CT, computed tomography; CVA, cerebrovascular accident; GFR, glomerular filtration rate; HR, heart rate; MI, myocardial infarction; RR, respiratory rate; T2DM, type 2 diabetes mellitus; WBC, white blood cells.

superior characteristics relative to the model generated in the first stage of multivariate analysis. The area under the curve was 89.2%, with its sensitivity and specificity measuring 86.2% and 76.0%, respectively, at the optimal cut-off point (Youden point) with a linear predictor value of −2.95 (Fig. 3 A).

Table 2
Analysis of multivariate model to predict the risk of in-hospital mortality in patients with COVID-19.

Variable	Exp (β)	2.5 %	97.5 %	p
SpO2, %	0.902	0.879	0.924	**0.000**
GFR,	0.983	0.977	0.989	**0.000**
Age, years	1.061	1.047	1.076	**0.000**
Hemoglobin, g/L	0.987	0.981	0.994	**0.000**
Lymphocytes, %	0.970	0.958	0.982	**0.000**
WBC, $*10^9$/L	1.062	1.035	1.090	**0.000**
Platelets, $*10^9$/L	0.994	0.992	0.996	**0.000**
AST, units/L	1.004	1.002	1.006	**0.000**
Glucose, mmol/L	1.071	1.029	1.113	**0.001**
HR/1 min	1.013	1.004	1.022	**0.004**
RR/1 min	1.048	1.010	1.086	**0.012**

AST, aspartate aminotransferase; GFR, glomerular filtration rate; Hb, hemoglobin; HR, heart rate; RR, respiratory rate; SpO2, blood oxygen saturation; WBC, white blood cells.

β, coefficient of the regression equation; Exp (β), exponent of the coefficient of the regression equation; p, level of statistical significance for the coefficient of the regression equation; margins of 95% confidence interval (the lower is 2.5% and the upper is 97.5%).
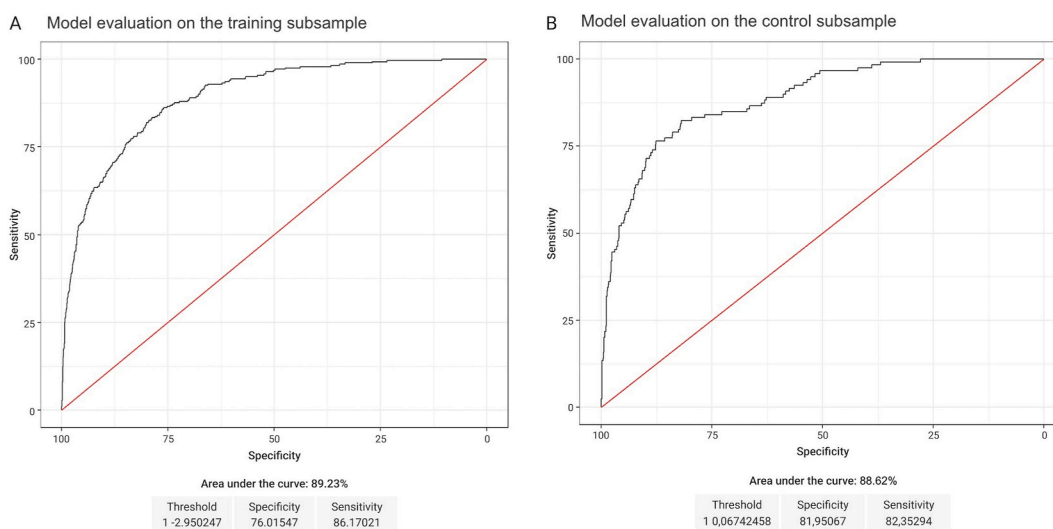

Fig. 3. Analysis of the final multivariate model that included comorbidities and gender data, results of the laboratory and instrumental tests to predict the risk of in-hospital mortality in patients with COVID-19. ROC analysis of the final multivariate model that included comorbidities and gender data, results of the laboratory and instrumental tests to predict the risk of in-hospital mortality in patients with COVID-19. The model included the following variables: age, HR, RR, SpO2, hemoglobin level, WBC count, platelet count, % of lymphocytes, GFR, glucose and AST levels.
A – Model evaluation on the training subsample
B - Model evaluation on the control subsample.

The model was tested using the control subsample (30% of patients). The model based on the control subsample also had better characteristics than the model obtained at the 1st stage of the analysis: the area under the curve was 88.6%, the sensitivity was 82.3 %, and the specificity was 81.9% (Fig. 3 B).

Therefore, according to the final multivariate model derived, predictors of an elevated mortality risk in hospitalized COVID-19 patients include increasing age of the patient, HR, RR, increased WBC count, AST level, glucose level, and a reduction in the blood concentrations of lymphocytes, platelets, hemoglobin, as well as decreased GFR and SpO2 levels.

The impact of elevated glucose levels on prognosis was separately investigated in patients with and without T2DM. For this, we independently evaluated the effect of hyperglycemia on patient prognosis as part of the multivariate analysis and found that the impact of glucose levels on patient prognosis does not rely on the presence of T2DM.

A predictive equation for hospital mortality was devised based on the final model. This equation incorporates the 11 parameters incorporated into the model:

LP (linear predictor value, logit odds ratio) = 4.3276783 - (0.1035048 * minimum SpO2, %) - (0.0171382 * GFR, ml/min/1.73 m^2) + (0.0594187 * age, years) – (0.0128326 * hemoglobin, g/l) – (0.0305012 * lymphocytes, %) + (0.0602048 * WBC, 10^9/L) – (0.0059304 * platelets, 10^9/L) + (0.0039620 * AST, U/L) + (0.0685963 * glucose, mmol/L) + (0.0129957 * HR, per minute) + (0.0467970 * RR, per minute).

A nomogram (ACTIV scale) was subsequently crafted utilizing this equation to forecast hospital mortality in COVID-19 patients in a

real-world clinical setting (Fig. 4). The personalized application of the ACTIV scale entails several steps. Initially, each of the 11 parameters utilized in the nomogram is scored individually (for instance, a 62-year-old patient, when plotted against the scale via a vertical line, corresponds to 46 points). To use the scale, draw a vertical line from each parameter to the SCORE line, sequentially, and summarize the points. After scoring each of the 11 parameters and aggregating them, mark this point on the "total scores" line at the bottom of the nomogram. From the point obtained on the "total scores" line, draw a vertical line downward to the "probability of death" line. For simplification, Supplementary Table 6 presents scores corresponding to specific parameter values.

Therefore, a multivariate analysis incorporating all variables revealed that patient age, clinical criteria for disease severity (HR, RR, SpO2), and straightforward laboratory markers such as hemoglobin level, percentage of lymphocytes, WBC count, platelet count, AST level, glucose, and GFR have the strongest impact on short-term prognosis (hospital mortality).

4. Discussion

The ability to predict the risk of death in COVID-19 patients using a straightforward and readily available method is of paramount importance in real clinical practice. This ability enables the earliest possible identification of patients at greatest risk, based on objective criteria. At the onset of the pandemic, no specialized scales for predicting the infection's course were available, and existing scales for sepsis patients (Systemic Inflammatory Response Syndrome (SIRS), Sequential [Sepsis-related] Organ Failure Assessment (SOFA), Quick Sequential [Sepsis-related] Organ Failure Assessment (qSOFA), Logistic Organ Dysfunction System (LODS) [6], as well as those for community-acquired pneumonia (CURB-65) [7] and viral pneumonia (MuLBSTA) [8], were used. Unfortunately, these models demonstrated insufficient predictive accuracy. The key objective of our study was to construct a multivariate model that predicts the risk of death for hospitalized COVID-19 patients with adequate sensitivity and specificity.

The scales utilized in the early phase of the pandemic failed to meet the criteria of high specificity and sensitivity. The SOFA and qSOFA scores have been validated as predictive tools for suspected sepsis in a typical infectious hospital. Nevertheless, it became apparent during the initial months of the pandemic that the state of septic shock is not equivalent to a severe course of COVID-19.

Indeed, many local guidelines across various countries have recommended the use of the SOFA scale for assessing the severity of COVID-19 in patients admitted to the intensive care unit (ICU). However, the SOFA scale includes criteria that are only ascertainable in the intensive care unit and cannot be determined for all consecutive hospitalized COVID-19 patients, such as the PaO2/FiO2 ratio and the administration of vasopressors. This fact undoubtedly restricts the use of the SOFA scale in real-world practice and renders it infeasible for use outside the ICU. Consequently, the SOFA scale is not a universal prognostic tool for hospitalized COVID-19 patients. Since the pandemic began, there have been attempts to apply other predictive models, such as those designed to assess the severity of community-acquired pneumonia (for instance, CURB-65) [7] or to predict mortality risk among patients with viral pneumonia (MuLBSTA) [8].

However, given that the pathogenesis and clinical course of the novel coronavirus pneumonia significantly differ from other viral pneumonias, the application of these scales for COVID-19 was not adequately justified. This gap in predicting COVID-19 outcomes during the pandemic has driven interest in creating predictive models. To date, several dozen models and scales, designed to predict the disease course and mortality risk, have emerged. The most noteworthy among these are two systematic reviews [9,10], as well as
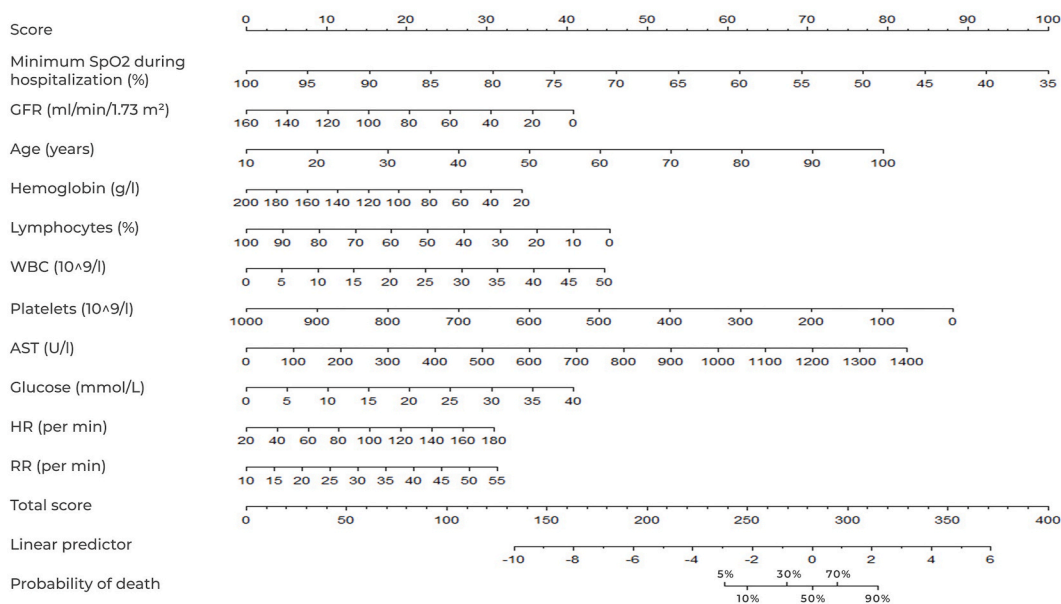

Fig. 4. Nomogram to calculate a risk of death in a hospitalized COVID-19 patient (ACTIV scale). AST, aspartate aminotransferase; GFR, glomerular filtration rate; HR, heart rate; RR, respiratory rate; SpO2, oxygen saturation.

studies conducted by Webb BJ [11], Jiang M [12], and Wang Z [13].

We analyzed the following predictive scales and models in comparison with the ACTIV risk scale. In a systematic review, Gupta RK attempted to assess the performance of 22 published predictive models for COVID-19 relative to models employed in routine clinical practice prior to the pandemic [9]. The primary conclusion of the review posited that all models proposed for COVID-19 patients, which calculated the likelihood of disease progression or mortality risk, exhibited low predictive performance. This performance was even lower than the predictive capacity of parameters commonly used in routine practice, such as blood oxygen saturation and patient age.

Contrary to these findings, the multivariate model in our study demonstrated superior predictive performance compared to individual factors, including age and SpO2 (Supplementary Table 1, Figs. 2, Fig 3).

It is crucial to compare the results of the ROC analysis performed in Gupta RK's systematic review with those of our study, that is, to evaluate the predictive value of different models. To facilitate this, we compared the areas under the curves, which reflect the predictive capacity of the model. The areas ranged from 56% to 78% in the models analyzed by Gupta RK. In contrast, our final model boasted the most robust predictive characteristics of ROC analysis: the area under the curve was 89.2%, and the sensitivity and specificity were 86.2% and 76.0%, respectively, at the optimal cutoff point (Youden point) with a linear predictor value of -2.95.

In a systematic review, Wynants L [10] examined 16 predictive models, encompassing 8 models to estimate mortality risk in COVID-19 patients, 5 models to predict progression from baseline to severe or critical illness, and 3 models to predict the duration of hospital stay.

Predictors for any outcome included age (incorporated in 7 models), CT findings (7 models), lactate dehydrogenase level (4 models), sex (3 models), CRP level (3 models), comorbidities (including hypertension, T2DM, cardiovascular and respiratory disease, included in 3 models), and lymphocyte count (3 models). All models demonstrated low predictive performance, which, according to the authors of the review, is predominantly due to inappropriate patient selection for the control group, exclusion of patients who did not experience an index event by the study's conclusion, and poor reporting. Consequently, the 16 analyzed models seemingly cannot be recommended for application in real-world clinical practice.

In contrast to the cited work by Wynants L, our study included all consecutively admitted patients and accounted for all outcomes (at the time of the registry data analysis, all patients had either been discharged or had passed away).

Webb BJ [11] proposed a model for predicting the risk of hospitalization and 28-day mortality in outpatients with COVID-19. This study employed predictors drawn from the published works of other researchers and, typically, derived from various population studies. It is noteworthy that this methodological approach precluded a comprehensive analysis of a specific patient population with COVID-19. In contrast, our study identified predictors based on a series of univariate regression models with data from a specific patient population in the Eurasian region.

Variables in the Webb BJ model included: age (0.5 points per decade); high-risk comorbidities (2 points each): diabetes mellitus, severe immunodeficiency, and obesity; non-White/Hispanic or Hispanic ethnicity (2 points), and 1 point each for male gender, dyspnea, hypertension, CAD, heart rhythm disturbances, congestive CHF, CKD, COPD, chronic liver disease, cerebrovascular disease, and chronic neurological disease. As evident from this list of predictors, the analysis predominantly covered the influence of comorbidities. Despite several methodological assumptions, the resulting models exhibited high predictive value: the AUC was 0.82 for hospitalization and 0.91 for 28-day mortality (n = 16,030).

In contrast to the aforementioned study, our investigation focused on mortality prediction for patients already hospitalized. Our findings demonstrated that the most crucial predictors of mortality for this group were not necessarily diagnoses of concurrent diseases, but rather age, along with clinical and laboratory characteristics at the time of hospitalization.

Jiang M's model for predicting 30-day mortality in hospitalized COVID-19 patients [12] bears the closest resemblance to our proposed nomogram. This model, derived from a study of 1905 patients, offers a scale for determining mortality risk (Age, Biomarkers, Clinical history, Sex, or ABCS). It incorporates ten variables strongly associated with mortality risk: age, sex, COPD, AST, highly sensitive CRP, highly sensitive Troponin I, WBC count, lymphocyte count, D-dimer, and procalcitonin. Interestingly, our model also employs such mortality risk markers as age, AST levels, WBC count, and lymphocyte counts.

However, Jiang M's scale is derived from data from a relatively small patient cohort from a single region in China. Additionally, the scale includes parameters not routinely tested in all COVID-19 hospitalized patients (for instance, procalcitonin and highly sensitive Troponin I).

Wang Z [13] incorporated entirely novel parameters to construct a predictive model. This study used a panel of 50 peptide markers derived from 30 proteins, whose functions were associated with COVID-19, to predict survival in hospitalized COVID-19 patients. Although the authors propose this prognostic method could outperform current systems, its implementation in clinical practice remains technically impractical.

Therefore, a significant distinctive aspect of our study is the principle of prognostic factor selection via a machine learning algorithm ("random forest") for constructing a model predicting mortality in COVID-19 patients. The model utilizes data from a substantial sample (n = 9364) of patients from the Eurasian region, gathered through the international ACTIV registry. The model's predictive performance, as evidenced by the ROC analysis, was substantiated using the control subsample.

In conclusion, it is noteworthy that the mortality rate among hospitalized patients in the ACTIV registry was 5.8%. According to a study conducted by the World Heart Federation, encompassing 5313 patients from 23 countries, the in-hospital mortality rate for COVID-19 was 6% among Caucasians, 13% among Asians [14], 10% for low-income countries, and 4% for high-income countries. Therefore, the severity of patients' conditions included in the ACTIV registry aligns with global trends and can be employed to validate the proposed risk model.

The ACTIV risk scale comprises 11 risk factors and serves as an intuitive visual tool, making it a helpful tool in clinical practice.

Personalized prognostic assessment derived from the ACTIV scale enables an objective approach to medical decision-making regarding individual treatment regimens and subsequent patient follow-up in the context of COVID-19.

4.1. Study limitations

1. The model's robustness could be further improved through external validation on larger datasets, as our validation process was restricted to a control subsample of respondents (30% of patients).
2. The patient cohort primarily represents individuals with moderate to severe infection, as those with asymptomatic or mild symptoms treated at home were not included in the study.
3. The predictive model does not account for the influence of specific COVID-19 treatments.
4. Enrollment into the ACTIV, ACTIV 2 register occurred from June 29, 2020, to March 30, 2021, spanning the 1st, 2nd, and 3rd waves of the pandemic.
5. The model does not consider the ethnicity of the respondents.
6. In this registry, as in most registries, obtaining accurate data in its entirety is challenging. Some patients have missing data, but the overall amount of such lost information is small relative to the number of patients and is unlikely to significantly distort descriptive statistics.

5. Conclusion

This study presents a straightforward predictive model, based on the clinical characteristics of COVID-19 patients (age, HR, RR, SpO2), and seven biomarkers (hemoglobin, white blood cell count, platelet count, % lymphocyte count, GFR, glucose, and AST). This model provides a practical tool for stratifying COVID-19 patients by death risk. The ACTIV scale, employed in risk assessment, demonstrated good results when testing the model on the control subsample (30% of patients).

Practitioners can use this tool to predict the risk of death in COVID-19 patients upon admission and to plan treatment strategies. This could help improve optimal distribution of limited medical resources within national healthcare systems and potentially reduce COVID-19 patient mortality.

Ethical approval statement

This study was reviewed and approved by the Ethics Committee of the Pirogov Russian National Research Medical University of the Ministry of Health of the Russian Federation, with the approval number: №199, №201.

All participants/patients (or their proxies/legal guardians) provided informed consent to participate in the study.

The study was registered in the ClinicalTrials.gov database as "Analysis of Chronic Non-infectious Diseases Dynamics After COVID-19 Infection in Adult Patients (ACTIV)", NCT04492384 and ACTIV 2, NCT04709120. More information about the registry can be found on the Eurasian Association of Internal Medicine's website: https://ACTIV.euat.ru.

Further information on the principles and practical aspects of informed consent in the ACTIV registry is available at https://activ.euat.ru/documents (in Russian only).

Data availability statement

Data associated with the study has been deposited into a publicly available repository - www.clinicaltrials.gov (ACTIV NCT04492384, ACTIV 2 NCT04709120) and supplementary data will be made available on request.

CRediT authorship contribution statement

Gregory P. Arutyunov: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Ekaterina I. Tarlovskaya:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Dmitry S. Polyakov:** Writing – review & editing, Software, Investigation, Formal analysis, Data curation. **Tatiana I. Batluk:** Writing – review & editing, Validation, Resources, Investigation. **Alexander G. Arutyunov:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Demko, Angelika V. Evzerikhina, Aleksandr V. Zharkov, Anna Zateishchikova, Umida K. Kamilova, Zulfiya F. Kim, Tatiana Y. Kuznetsova, Alexandr Kulikov, Natalia V. Lareva, Ekaterina V. Makarova, Svetlana V. Malchikova, Sergey V. Nedogoda, Marina M. Petrova, Ilya G. Pochinka, Konstantin V. Protasov, Denis N. Protsenko, Dmitry Y. Ruzanav, Sergey A. Sayganov, Akpay Sh. Sarybaev, Natalia M. Selezneva, Akhmetzhan B. Surgaliev, Igor V. Fomin, Olga V. Khlynova, Olga Y. Chizhova, Igor I. Shaposhnik, Dmitry A. Sh'ukarev, Augyl K. Abdrakhmanova, Susanna A. Avetisian, Hovhannes G. Avoyan, Karine K. Azarian, Galiya T. Aimakhanova, Dinara A. Ayipova, Almazbek Ch. Akunov, Marianna K. Alieva, Asel R. Almukhambedova, Alyona V. Aparkina, Olga R. Aruslanova, Ekaterina Y. Ashina, Olga Y. Badina, Olga Y. Barysheva, Alina S. Batchayeva, Roman A. Bashkinov, Anna M. Bitieva, Ismail U. Bikhteyev, Nataliya A. Borodulina, Maksim V. Bragin, Viktoria Brazhnik, Angela M. Budu, Galina A. Bykova, Kristina R. Vagapova, Darina D. Varlamova, Natalia N. Vezikova, Elena A. Verbitskaya, Olga E. Vilkova, Elena A. Vinnikova, Vera V. Vustina, Elena A. Galova, Vadim V. Genkel, Dmitry B. Giller, Elena I. Gorshenina, Elena V. Grigorieva, Ekaterina Y. Gubareva, Gaukhar M. Dabylova, Anastasiya I. Demchenko, Oleg Y. Dolgikh, Melis Y. Duyshobayev, Dmitry S. Evdokimov, Ksenia E. Egorova, Aliya E. Zheldybayeva, Nataliya V. Zarechnova, Yuliya D. Zimina, Svetlana Y. Ivanova, Elena Y. Ivanchenko, Mariya V. Ilina, Mariya V. Kazakovtseva, Elena V. Kazymova, Yuliya S. Kalinina, Nadezhda A. Kamardina, Anastasiya M. Karachenova, Igor A. Karetnikov, Nina A. Karoli, Magomed Kh. Karsiev, Daria S. Kaskaeva, Karina F. Kasymova, Zhainagul B. Kerimbekova, Evgeniy S. Kim, Nina V. Kiseleva, Daria A. Klimenko, Olga V. Kovalishena, Sergey Kozlov, Elena V. Kolmakova, Tatyana P. Kolchinskaya, Maria I. Koliadich, Olga V. Kondriakova, Marina P. Konoval, Dmitriy Y. Konstantinov, Elena A. Konstantinova, Vera A. Kordukova, Ekaterina V. Koroleva, Angelina Y. Kraposhina, Tamara V. Kryukova, Alla S. Kuznetsova, Tatyana Y. Kuzmina, Kirill V. Kuzmichev, Cholpon K. Kulchoroeva, Tatyana V. Kuprina, Irina M. Kuranova, Liliya V. Kurenkova, Natalia Y. Kurchugina, Nadira A. Kushubakova, Valeriya I. Levankova, Alla A. Ledyaeva, Tatyana Lisun, Victoria Lisyanskaya, Natalia A. Lyubavina, Nadezhda A. Magdeyeva, Konstantin V. Mazalov, Viktoria I. Maiseenko, Aleksandra S. Makarova, Abdirashit M. Maripov, Nikita Markov, Anastasiya A. Marusina, Evgeniy S. Melnikov, Anna Metlinskaya, Nikita B. Moiseenko, Fasliniso N. Muradova, Rimma G. Muradyan, Shagane N. Musaelian, Ekaterina S. Nekaeva, Nataliya M. Nikitina, Sergey Nifontov, Ekaterina Obolentseva, Anna Obukhova, Bela B. Ogurlieva, Alla A. Odegova, Yuliya V. Omarova, Nazgul A. Omurzakova, Shynar O. Ospanova, Victoria Pavlova, Ekaterina V. Pahomova, Liviu D. Petrov, Svetlana S. Plastinina, Dmitry Platonov, Vera A. Pogrebetskaya, Dmitry V. Polyakov, Ekaterina V. Ponomarenko, Larisa L. Popova, Artem Potanin, Natalia A. Prokofeva, Juliya Rabik, Nikita A. Rakov, Abdurahmon N. Rakhimov, Nadezhda A. Rozanova, Irina V. Samus, Saltanat Serikbolkyzy, Yana Sidorkina, Andrey A. Simonov, Valeriya V. Skachkova, Ruth Skvortcova, Daniil Skuridin, Daria V. Solovieva, Irina A. Solovieva, Irina M. Sukhomlinova, Anastasiya G. Sushilova, Dilnoza R. Tagayeva, Elena P. Tikhonova, Danil S. Tokmin, Anastasiya A. Tolmacheva, Mariya S. Torgunakova, Ksenia V. Trenogina, Natalia A. Trostianetckaia, Dmitriy A. Trofimov, Marina A. Trubnikova, Alexander A. Tulichev, Assiya T. Tursunova, Nina D. Ulanova, Oleg V. Fatenkov, Olga V. Fedorishina, Tatyana S. Fil, Irina Y. Fomina, Irina S. Fominova, Irina A. Frolova, Svetlana M. Tsvinger, Vera V. Tsoma, Meerim B. Cholponbaeva, Tatyana I. Chudinovskikh, Igor Shavrin, Olga A. Shevchenko, Dzhoshgun Shikhaliev, Ekaterina A. Shishkina, Konstantin Y. Shishkov, Stanislav Y. Sherbakov, Galina V. Shcherbakova, Ekaterina A. Yausheva.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28892.

References

[1] Database registration certificate No. 2021622728. Database of the register "ACTIV 1 & ACTIV 2" Analysis of Chronic Non-infectious Diseases Dynamics After COVID-19 Infection in Adult Patients (ACTIV SARS-CoV-2). Rights holder: Association "Eurasian Association of Therapists" (RU), 01.12.2021 (Russian). Available at: https://fips.ru/EGD/f72ff393-3299-41d3-b446-3ae9c0a74575 (accessed March 23, 2024).

[2] R Core Team, R: the R Project for statistical computing, R-project.org (2022). Available at: https://www.r-project.org/. (Accessed 23 March 2024).

[3] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[4] RMS. hbiostat.org. Available at: https://hbiostat.org/R/rms/(accessed March 23, 2024).

[5] F. Harrell, rms. GitHub (2022). Available at: https://github.com/harrelfe/rms. (Accessed 23 March 2024).

[6] C.W. Seymour, V.X. Liu, T.J. Iwashyna, F.M. Brunkhorst, T.D. Rea, A. Scherag, et al., Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3), JAMA 315 (8) (2016) 762–774, https://doi.org/10.1001/jama.2016.0288. Erratum in: JAMA 2016; 315(20):2237.

[7] W.S. Lim, M.M. van der Eerden, R. Laing, W.G. Boersma, N. Karalus, G.I. Town, et al., Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study, Thorax 58 (5) (2003) 377–382, https://doi.org/10.1136/thorax.58.5.377.

[8] L. Guo, D. Wei, X. Zhang, Y. Wu, Q. Li, M. Zhou, et al., Clinical features predicting mortality risk in patients with viral pneumonia: the MuLBSTA score, Front. Microbiol. 10 (2019) 2752, https://doi.org/10.3389/fmicb.2019.02752. Erratum in: Front Microbiol 2020;11:1304.

[9] R.K. Gupta, M. Marks, T.H.A. Samuels, A. Luintel, T. Rampling, H. Chowdhury, et al., Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study, Eur. Respir. J. 56 (6) (2020) 2003498, https://doi.org/10.1183/13993003.03498-20202020.

[10] L. Wynants, B. Van Calster, M.M.J. Bonten, G.S. Collins, T.P.A. Debray, M. De Vos, et al., Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal, BMJ 369 (2020) m1328, https://doi.org/10.1136/bmj.m1328. Update in: BMJ 2021;372:n236. Erratum in: BMJ 2020; 369:m2204.

[11] B.J. Webb, N.M. Levin, N. Grisel, S.M. Brown, I.D. Peltan, E.S. Spivak, et al., Simple scoring tool to estimate risk of hospitalization and mortality in ambulatory and emergency department patients with COVID-19, PLoS One 17 (3) (2022) e0261508, https://doi.org/10.1371/journal.pone.0261508.

[12] M. Jiang, C. Li, L. Zheng, W. Lv, Z. He, X. Cui, et al., A biomarker-based age, biomarkers, clinical history, sex (ABCS)-mortality risk score for patients with coronavirus disease 2019, Ann. Transl. Med. 9 (3) (2021) 230, https://doi.org/10.21037/atm-20-6205.

[13] Z. Wang, A. Cryar, O. Lemke, P. Tober-Lau, D. Ludwig, E.T. Helbig, et al., A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: an observational multi-cohort study, EClinicalMedicine 49 (2022) 101495, https://doi.org/10.1016/j.eclinm.2022.101495.

[14] D. Prabhakaran, K. Singh, D. Kondal, L. Raspail, B. Mohan, T. Kato, et al., Cardiovascular risk factors and clinical outcomes among patients hospitalized with COVID-19: findings from the world heart federation COVID-19 study, Glob Heart 17 (1) (2022) 40, https://doi.org/10.5334/gh.1128. Erratum in: Glob Heart 2022;17(1):79.