Logistic Regression Analysis to Predict Mortality Risk in COVID-19 Patients from Routine Hematologic Parameters

Sudhir Bhandari¹, Ajit Singh Shaktawat¹, Amit Tak², Bhoopendra Patel³, Jyotsna Shukla², Sanjay Singhal², Kapil Gupta², Jitendra Gupta², Shivankan Kakkar⁴, Amitabh Dube²

¹Department of Medicine, SMS Medical College and Attached Hospitals, Jaipur, ²Department of Physiology, SMS Medical College and Attached Hospitals, Jaipur, ³Department of Physiology, Government Medical College, Barmer, ⁴Department of Pharmacology, SMS Medical College and Attached Hospitals, Jaipur, Rajasthan, India

Abstract

Background: The triage of coronavirus-19 patients into various strata based on some prognostic indicator might prove a utilitarian strategy in the management of epidemic. The goal of health-care facilities is optimization of the use of medical resources. The present study aimed to develop a predictor model of mortality risk from routine hematologic parameters. Patients and Methods: In this retrospective case-control study, seventy survivors (n = 47) and nonsurvivors (n = 23) were enrolled who were laboratory-confirmed coronavirus disease 2019 (COVID-19) cases from SMS Medical College, Jaipur (Rajasthan, India). The clinical and routine blood profile of the survivors and nonsurvivors was recorded. A logistic regression model was fitted with step-wise method to the above dataset with dependent variable such as survivor or nonsurvivor and independent variables such as age, sex, symptoms, random blood glucose, and complete blood count. The best model was selected on the basis of Akaike information criterion. Results: It was observed that differential neutrophil count (%) and random blood sugar (RBS in mg/dL) are the statistically significant regressors (P < 0.05). The performance metrics of the model with 5-fold cross-validation showed area under the receiver operating characteristic curve, sensitivity, specificity, and validation accuracy to be 0.95, 90%, 92%, and 70%, respectively. The cutoff probability comes out at 0.30 for the outcome (nonsurvivor as success). Conclusion: The study concludes that differential neutrophil count and RBS levels can be used as early screening tools of mortality risk in COVID-19 patients and they assist in further patient management.

Keywords: Coronavirus disease 2019, differential neutrophil count, logistic regression model, random blood sugar, receiver operating characteristic curve

Address for correspondence: Dr. Amit Tak, 4, Pushpa Path, Uniara Garden, Moti Dungri Road, Jaipur - 302 004, Rajasthan, India. E-mail: dramittak@gmail.com

 Submitted:
 25-May-2020
 Revised:
 31-May-2020

 Accepted:
 17-Jun-2020
 Published:
 27-Jun-2020

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Bhandari S, Shaktawat AS, Tak A, Patel B, Shukla J, Singhal S, *et al.* Logistic regression analysis to predict mortality risk in COVID-19 patients from routine hematologic parameters. Ibnosina J Med Biomed Sci 2020;12:123-9.



Access this article online Website:

www.ijmbs.org

DOI: 10.4103/ijmbs.ijmbs_58_20

123

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a disease caused by severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). The emergence of the disease occurred in more than 200 countries of the world.^[1] The demographic profile of the SARS-CoV-2 infection varies widely across age, gender, and socioeconomic strata. Similarly, the clinical spectrum of the disease encompasses asymptomatic infection, mild upper respiratory tract illness, severe viral pneumonia with respiratory failure, and even death.^[2] There is a continuous search for efficient indicators of disease diagnosis, disease severity, therapeutic response, and disease outcome. According to the 5th edition of National Treatment Guidelines, the disease severity of COVID-19 is classified into four stages on the basis of pulmonary imaging. The present study was undertaken to develop a predictive model of disease mortality using easily available, cost-effective blood indicators that include random blood sugar (RBS) and complete blood count. This can be used as a quick screening tool, and patients with high risk would be evaluated for more precise indictors of mortality. During the epidemic, this strategy could be beneficial to triage patients and provide adequate management to these patients.

PATIENTS AND METHODS

A hospital-based, retrospective, case-control study was designed in the SMS Medical College and Hospital, Jaipur, to develop a prediction model for mortality risk using logistic regression analysis in COVID-19 patients. The patients were managed as per standard protocol of the institute. The study included case records of 23 nonsurvivors (33%) and 47 survivors (67%) with laboratory-confirmed SARS-CoV-2 infection. The demographic and laboratory details at the time of admission were collected to create database. In case of missing data, the whole observation has been removed from the analysis. The dependent variable was qualitative, either survivor or nonsurvivor. The regressors (or predictors) included age, gender, presence of symptoms, RBS, and complete blood count.

The aim of the present study is to develop a predictor model, where the choice of regressors is

not important in the sense that two models based on different regressors can be equally good in prediction. There are no implications of causality or even association, which are in the purview of explanatory and causal models.^[3]

The current scenario necessitates the early development of predictor model. In order to develop model with small sample, multiple analyses were performed to extract a subset of regressors from a bunch. The general rule is to have at least ten participants for each category of regressor.^[4] The preliminary analysis of the data includes univariate logistic regression analysis and comparison of means of all regressors in survivor and nonsurvivor groups. The regressors that show significant difference of means in two groups were selected and correlations were found among them. Between the two regressors that are significantly correlated, one with higher odds ratio (OR) was selected. The top five regressors (as the sample has 57 patients) having the highest ORs were selected to fit a multivariate logistic regression model using a step-wise method [Figure 1]. The best model was chosen with Akaike information criterion (AIC). The best model is one with minimum AIC value. The regressors contributing significantly in the prediction model were used to train a 5-fold cross-validation logistic regression model, and cutoff probability, area under the receiver operating characteristic (ROC) curve, sensitivity, specificity, and accuracy were calculated.^[5]

In the present logistic regression model, the success was defined when dependent variable took the value nonsurvival (or mortality).

The present study has included 21 regressors that are age (in years), gender (male or female), presence or absence of symptoms (symptomatic or asymptomatic), RBS in milligrams per deciliter, hemoglobin (Hb) in grams%, total leukocyte count (TLC) in 10³ cells per cubic millimeter, total red blood cell count in million cells per cubic millimeter, mean corpuscular volume in femtoliters per cell, mean corpuscular hemoglobin (MCH) in picograms per cell, mean corpuscular Hb concentration in grams per deciliter, red blood cell distribution width-coefficient of variation,

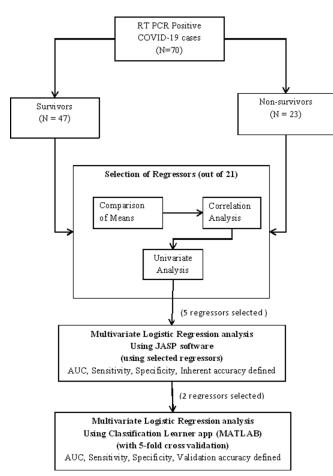


Figure 1: Flowchart of the study

platelet count (PLT) in 10⁵ per cubic millimeters, packed cell volume in percent, differential neutrophil count (NPHIL) in percent, differential lymphocyte count (LYMP) in percent, absolute neutrophil count (ANC) in 10³ cells per cubic millimeter, absolute lymphocyte count (ALC) in 10³ cells per cubic millimeter, differential monocyte count (MONO) in percent, absolute monocyte count in 10³ cells per cubic millimeter, differential neutrophil count-to-differential lymphocyte count ratio (NLR), and ANC-to-absolute lymphocyte count ratio (ANLR).

As the outcome (mortality) in the study was <5% (case fatality rate 2%-3%), we used OR as a fairly good approximation of relative risk (RR) of death.^[4,6] To predict the outcome, the test values of regressor were put in the prediction model to obtain a probability. The probability value obtained from the prediction model should be compared with the cutoff probability of the ROC curve. The prediction probability greater than the cutoff probability favors the outcome (mortality).

Statistical analysis

The continuous random variables were expressed as mean (standard deviation) or median (interquartile range [IQR]) and compared with Mann–Whitney test. The categorical variables were expressed as proportions and compared with Chi-square test or Z test for proportions. Before applying the above tests, the assumptions of normality or variances were checked.

The significance of logistic coefficients was tested using Wald test. The best model was selected using AIC and tested with Chi-square test. The level of statistical significance was considered at 5%.

The logistic regression model analysis was performed on JASP Team (2020). JASP (Version 0.11.1.0) [Computer software]. University of Amsterdam, Netherlands. MATLAB Team 2016a, MathWorks, MATLAB (Version 9.0.0341360) [Mathematical Computing software] Natick, Massachusetts, USA. software platforms.

RESULTS

Seventy patients were enrolled in the study with a median age of 50 years (IQR 30-60 years). The number of males was two times that of females. Symptomatic and asymptomatic cases were equal. The comparison of regressors between survivor and nonsurvivor groups was performed [Supplementary Tables 1 and 2]. There was a statistically significant difference in the mean age of survivors and nonsurvivors (P = 0.01), but gender differences were not statistically significant in the two groups (P = 0.81). The RBS, TLC, PLT, NPHIL, ANC, LYMP, ALC, NLR, and ANLR showed statistically significant differences between the survivor and nonsurvivor groups. A significant correlation was observed in various regressors [Supplementary Table 3].

The univariate analysis of regressors showed significant ORs for age (OR = 1.035), RBS (OR = 1.025), TLC (OR = 1.265), NPHIL (OR = 1.22), ANC (OR = 1.4), LYMP (OR = 0.838), ALC (OR = 0.223), NLR (OR = 1.218), and

Variable	Univariate analysis				Multivariate analysis			
	OR	95% CI		Р	OR	95% CI		Р
		LL	UL			LL	UL	
Age	1.035	1.004	1.066	0.025				
Sex (male)	0.879	0.306	2.523	0.81				
Symptoms	6×10 ⁸	0	00	0.991				
RBS	1.025	1.008	1.041	0.003	1.041	1.006	1.078	0.022
Hb	0.918	0.741	1.136	0.43				
TLC	1.265	1.069	1.498	0.006				
TRBC	0.686	0.407	1.159	0.159				
MCV	1.019	0.967	1.074	0.487				
MCH	1.042	0.927	1.172	0.487				
MCHC	1.084	0.907	1.296	0.375				
RDW-CV	1.056	0.818	1.364	0.676				
PLT	0.582	0.307	1.105	0.098				
PCV	0.97	0.905	1.039	0.384				
NPHIL	1.22	1.109	1.342	< 0.001	1.373	1.086	1.736	0.008
ANC	1.4	1.123	1.746	0.003				
LYMP	0.838	0.771	0.911	< 0.001				
ALC	0.223	0.092	0.539	< 0.001				
NLR	1.218	1.095	1.355	< 0.001				
ANLR	1.198	1.082	1.326	< 0.001				
MONO	0.702	0.485	1.017	0.061				
AMC	2.873	0.208	39.65	0.431				

The OR of significant regressors selected by step-wise method in multivariate model is shown for random blood sugar and differential neutrophil count. The full description of abbreviation of variables is given in the materials and methods section. OR: Odds ratio; CI: Confidence interval, LL: Lower limit of CI, UL: Upper limit of CI

ANLR (OR = 1.198) [Table 1]. The multivariate analysis was performed using ANC, NPHIL, ANLR, age, and RBS regressors using step-wise method. The best model was chosen with minimum AIC value as 29.9 (P < 0.001). It was found that two regressors that contribute significantly in the prediction model are differential neutrophil count and RBS [Table 1]. The ROC curves for differential neutrophil count and RBS were shown separately with AUC of 0.932 and 0.837, respectively [Figures 2 and 3]. The combined effect of regressors, NPHIL, and RBS in the prediction of nonsurvivors had AUC, sensitivity, specificity, and inherent accuracy of 0.96, 90.5%, 88.9%, and 89.5%, respectively [Table 2]. These regressors and their estimates of logistic coefficients are given below:

$\hat{\lambda} = -32.77 + 0.33 \times differential neutrophil count$ +0.04 ×random blood sugar

where $\hat{\lambda}$ is log odds of outcome. The 5-fold cross-validation logistic regression model was trained with differential neutrophil count and RBS regressors to calculate model performance metrics. The AUC, sensitivity, specificity, and accuracy were 0.95, 90%, 92%, and 70%, respectively. The cutoff probability was 0.30 for the mortality risk.

Sensitivity =
$$\frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{19}{21} = 0.905$$

Specificity
$$=\frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{32}{36} = 0.889$$

Inherent accuracy =
$$\frac{\text{TP} + \text{TN}}{\text{N}} = \frac{51}{57} = 0.895$$

DISCUSSION

The outbreak of the novel COVID-19 forces the medical fraternity around the world to discover the many unfamiliar facets of the disease. The risk factors of mortality is one of the important dimensions of the clinical research. The risk factors also direct the health authorities to utilize medical

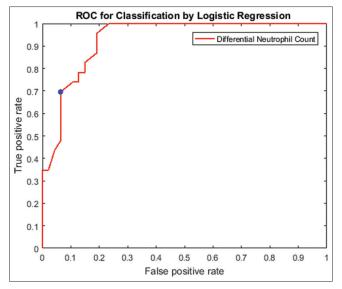


Figure 2: Receiver operating characteristic curve (red line) for differential neutrophil count is shown. The area under the curve is 0.932

Table 2: Confusion matrix							
Outcomes	Predic	Total					
	Nonsurvivor	Survivor					
Observed							
Nonsurvivor	19 (TP)	2 (FN)	21				
Survivor	4 (FP)	32 (TN)	36				
Total	23	34	<i>n</i> =57				

*TP: True positive, FP: False positive, TN: True negative, FN: False negative

infrastructure and human resource optimally to reduce the number of deaths during the epidemic. The present study showed that differential neutrophil count and RBS have significant contribution as indicators of mortality in COVID-19 patients. The above two hematologic parameters have 70% validation accuracy in predicting the outcome. Ruan et al. suggested age, the underlying diseases, and increased inflammatory indicators as mortality indicators.^[8] They showed significant differences in white blood cell counts, absolute lymphocyte counts, PLTs, albumin, total bilirubin, blood urea nitrogen, blood creatinine, myoglobin, cardiac troponin, C-reactive protein (CRP), and interleukin-6 in death and recovered groups. In the present study, differential lymphocyte count is highly negatively correlated with differential neutrophil count [Supplementary Table 3]. Green in a retrospective study of 150 COVID-19 patients found significant differences in ferritin and interleukin-6

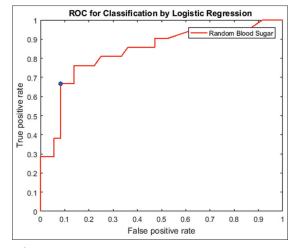


Figure 3: Receiver operating characteristic curve (red line) for random blood sugar is shown with area under the curve of 0.837

levels between nonsurvivors and survivors, suggesting that the cause of mortality may be hyperinflammation due to viral infection.^[9] Tan et al. retrospect the time course reports of complete blood count of dead and recovered cases. They suggest time lymphocyte (%) model as the prognostic factor of disease severity. The disease severity is proportional to decrease in lymphocyte count across timeline.^[10] Zhou et al. carried out a multivariable regression analysis and found increasing odds of in-hospital death associated with older age (OR 1.10, 95% confidence interval 1.03–1.17, per year increase; P = 0.0043), higher Sequential Organ Failure Assessment score (5.65, 2.61-12.23; P<0.0001), and d-dimer >1 μ g/mL (18.42, 2.64–128.55; P = 0.0033) on hospital admission.[11] Du et al. identified four risk factors including age ≥ 65 years, cardiovascular comorbidity, CD3 + CD8 + T cells \leq 75 cell/µL, and cardiac troponin I levels ≥ 0.05 ng/mL as predictors of mortality. They specifically mentioned the last two factors are more specific.^[12] Gupta et al. reported preventive measures for patients with diabetes and mentioned diabetes as an important risk factor for mortality in patients with other influenza epidemics. The present study also mentioned RBS as an important predictor of mortality risk.^[13] Singh reported that ACE-2 receptors are expressed on pancreatic islets and infection with SARS-CoV-1 cause hyperglycemia in people without existing diabetes mellitus. The hyperglycemia was seen to persist 3 years after

recovery from SARS, indicating damage to the beta cells of pancreas. Similar effects may be shown by SARS-CoV-2, which leads to increase in blood sugar levels.^[14]Henry (2020) emphasized that extracorporeal membrane oxygenation (ECMO) therapy can be an additional risk factor in COVID-19 patients, as it causes additional decrease in lymphocyte population in these patients.^[14] Li et al. designed a meta-analysis involving six studies that include the prevalence of cardiovascular disease in COVID-19, and they compared the incidence in non-intensive care unit (ICU)/severe and ICU/severe groups. The proportions of hypertension, cardio-cerebrovascular disease, and diabetes in patients with COVID-19 were 17.1%, 16.4%, and 9.7%, respectively.^[15] Vaduganathan et al. hypothesized the beneficial role of angiotensin-converting enzyme-2 (ACE-2) inhibitors in COVID-19; according to them, SARS-CoV-2 may cause activation of ACE-2 receptors and hypertension, which may be the risk factor for mortality in COVID-19.^[16] Vincent and Taccone emphasized the role of specific cause in COVID-19 deaths and also stressed that therapeutic limitations are contributing factors in case fatality rates.^[17] Lippi *et al.* assessed the relationship between COVID-19 and hypertension in a pooled analysis of COVID-19 patients and found 2.5-fold increased risk of severity and mortality in patients above 60 years of age.[18] Pal and Bhansali discussed the role of ACE2 inhibitors as contributing factors in mortality in diabetes mellitus patients. The use of ACE2 inhibitors causes overexpression of ACE receptors, which is the entry port of SARS-CoV-2.[19] Liu et al. demonstrated that NLR is an independent risk factor of mortality in COVID-19. They showed 8% higher risk of in-hospital mortality for each unit increase in NLR (OR = 1.08).^[17] The results corroborate with those of our study as NLR is highly correlated with differential neutrophil count.^[20] Du et al. in a retrospective observational study observed that the median age of the patients was 65.8 years and 72.9% were male. Hypertension, diabetes, and coronary heart disease were the most common comorbidities.^[21] Zhao et al. in a meta-analysis reported predictors of disease severity as old age (\geq 50 years, OR = 2.61), male

gender (OR = 1.348), smoking (OR = 1.734), and any comorbidity (OR = 2.635), especially chronic kidney disease (OR = 6.017), chronic obstructive pulmonary disease (OR = 5.323), and cerebrovascular disease (OR = 3.219). In terms of laboratory results, increased lactate dehydrogenase, CRP, and D-dimer and decreased blood platelet and lymphocyte count were highly associated with severe COVID-19 (all for P < 0.001).^[22] Muniyappa et al. from their perspectives also mentioned older age, diabetes mellitus, hypertension, and obesity as significant risk factors for hospitalization and death in COVID-19 patients.^[23] Most of the studies opine diabetes mellitus as one of the important risk factors for risk of mortality, which corresponds to the RBS of the present study. Second, decrease in differential lymphocyte count was observed in most of the studies, which also correlates negatively well with the differential neutrophil count of the present study. Although the validation accuracy of the present study is 70%, it could be a good screening tool. To increase the accuracy of the predictive model, new regressors which are mentioned in the above studies can be used, but they should be titrated against the ease of availability and cost.

CONCLUSION

The management of COVID-19 patients during the epidemic is a challenge due to limited medical resources. The present study is an effort to extract more information from routine laboratory investigations and thus develop a screening tool that guide caregivers to utilize specialized diagnostic and therapeutic procedures for a subset of patients on higher mortality risk.

Limitations of the study

The present study is a retrospective case–control study aimed to predict the mortality risk, though prospective studies are more accurate for prediction of risk. The sample size of the study is not large enough and, therefore, may affect the performance metrics of the predicted model.

Authors' contributions

All authors contributed equally.

Financial support and sponsorship Nil.

Conflicts of interest

There are no conflicts of interest.

Compliance with ethical principles

The reporting quality, formatting, and reproducibility guidelines of the present study are set forth by the EQUATOR Network. As the study was retrospective in nature, de-identified data were used, and consent of the patients was presumed, the approval of the institutional review board/ethics committee has been taken.

REFERENCES

- Worldometers. Available from: http://www.worldometers.info/ coronavirus/. [Last accessed on 2020 May 22].
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497-506.
- Indrayan A, Malhotra RK. Medical Biostatistics. 4th ed. Florida, USA: In Relative Risk and Odds Ratio CRC Press, Taylor & Francis Group; 2018. p. 358.
- Indrayan A, Malhotra RK. Medical Biostatistics. 4th ed. Florida, USA: In Relationships: Qualitative Dependent, CRC Press, Taylor & Francis Group; 2018. p. 472-3.
- MATLAB Team, Statistics and Machine Learning Toolbox 10.2, Classification Learner App, MATLAB. Version 9.0.0.341360 (R 2016a). Natick, Massachusetts: The Mathworks Inc.; R2015a.
- Singhal T. A review of coronavirus disease-2019 (COVID-19). Indian J Pediatr 2020;87:281-6.
- JASP Team, JASP version 0.12.2 [Computer software] University of Amsterdam, Neherlands; Copyright 2013-2019.
- Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020;46:846-8.
- 9. Green MS. Did the hesitancy in declaring COVID-19 a pandemic reflect a need to redefine the term? Lancet 2020;395:1034-5.
- Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang YQ, et al. Lymphopenia predicts disease severity of COVID-19: A descriptive and predictive

study. Signal Transduct Target Ther 2020;5:33.

- 11. 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.
- Du R-H, Liang L-R, Yang C-Q, Wang W, Cao T-Z, Li M, *et al.* Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study. European Respiratory Journal [Internet]. 2020;55:2000524. Available from: http://dx.doi. org/10.1183/13993003.00524-2020.
- Gupta R, Ghosh A, Singh AK, Misra A. Clinical considerations for patients with diabetes in times of COVID-19 epidemic. Diabetes Metab Syndr 2020;14:211-2.
- Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. Diabetes Metab Syndr 2020;14:303-10.
- Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, *et al.* Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol 2020;109:531-8.
- Vaduganathan M, Vardeny O, Michel T, McMurray JJ, Pfeffer MA, Solomon SD. Renin-angiotensin-aldosterone system inhibitors in patients with Covid-19. N Engl J Med 2020;382:1653-9.
- 17. Vincent JL, Taccone FS. Understanding pathways to death in patients with COVID-19. Lancet Respir Med 2020;8:430-2.
- Lippi G, Wong J, Henry BM. Hypertension in patients with coronavirus disease 2019 (COVID-19): A pooled analysis. Pol Arch Intern Med 2020;130:304-9.
- Pal R, Bhansali A. COVID-19, diabetes mellitus and ACE2: The conundrum. Diabetes Research and Clinical Practice [Internet]. 2020;162:108132. Available from: http://dx.doi.org/10.1016/j. diabres.2020.108132.
- Liu Y, Du X, Chen J, Jin Y, Peng L, Wang HH, et al. Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19. J Infect 2020;81:e6-12.
- Du Y, Tu L, Zhu P, Mu M, Wang R, Yang P, et al. Clinical features of 85 fatal cases of COVID-19 from Wuhan. A retrospective observational study. Am J Respir Crit Care Med 2020;201:1372-9.
- 22. 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 [Internet]. Cold Spring Harbor Laboratory; 2020. Available from: http://dx.doi.org/10.1101/2020.03.1 7.20037572.
- Muniyappa R, Gubbi S. COVID-19 pandemic, coronaviruses, and diabetes mellitus. Am J Physiol Endocrinol Metab 2020;318:E736-41.

Reviewers:

Khalid B Akkari (Dammam, Saudi Arabia) Khadija Hajidh (Dubai, UAE) Ahmed Elhassi (Benghazi, Libya) Nazeer Khan (Karachi, Pakistan) Editors: Salem A Beshyah (Abu Dhabi, UAE) Elmahdi A Elkhammas (Columbus OH, USA)

129