Trans R Soc Trop Med Hyg 2021; **0**: 1–8
https://doi.org/10.1093/trstmh/trab108 Advance Access publication 0 2021

FULLENGTH

A novel epidemiological scoring system for the prediction of mortality in COVID-19 patients

Prabhaker Mishra [a], Ratender Kumar Singh [b], Alok Nath [c], Shantanu Pande [d], Anil Agarwal [e], Om Prakash Sanjeev [b], Devendra Gupta [e], Prateek Singh [e], Tanmoy Ghatak [b], Zia Hashim [c], Vansh Khare [e], Sandeep Khuba [e], Amit Rastogi [e], and Radha K. Dhiman [f,*]

[a]*Biostatistics and Health Informatics;* [b]*Emergency Medicine;* [c]*Pulmonary Medicine;* [d]*Cardiovascular and Thoracic Surgery;* [e]*Anaesthesiology;* [f]*Hepatology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, , Lucknow-226014, Uttar Pradesh, India*

*Corresponding author: Tel: +91-7087009337; E-mail: rkpsdhiman@gmail.com

Received 6 March 2021; revised 18 May 2021; editorial decision 1 July 2021; accepted 11 August 2021

Background: Most of the reported risk score models for coronavirus disease 2019 (COVID-19) mortality are based on the levels of inflammatory markers, comorbidities or various treatment modalities, and there is a paucity of risk score models based on clinical symptoms and comorbidities.

Methods: To address this need, age, clinical symptoms and comorbidities were used to develop a COVID-19 scoring system (CSS) for early prediction of mortality in severe COVID-19 patients. The CSS was developed with scores ranging from 0 to 9. A higher score indicates higher risk with good discrimination quality presented by Mann Whitney U test and area under receiver operating characteristic curve (AUROC).

Results: Patient age of ≥60 y, cough, breathlessness, diabetes and any other comorbidity (with or without diabetes) are significant and independent risk factors for non-survival among COVID-19 patients. The CSS showed good sensitivity and specificity (i.e. 74.1% and 78.5% at CSS≥5, respectively), with an overall diagnostic accuracy of 82.8%, which was close to the diagnostic accuracy detected in the validation cohort (81.9%). In the validation cohort, high (8–9), medium (5–7) and low (0–4) CSS groups had 54.80%, 28.60% and 6.5% observed mortality, respectively, which was very close to the predicted mortality (62.40%, 27.60% and 5.2%, respectively, by scoring cohort).

Conclusions: The CSS shows a positive relationship between a higher score and proportion of mortality and, as its validation showed, it is useful for the prediction of risk of mortality in COVID-19 patients at an early stage, so that referral for triage and admission can be predetermined even before admission to hospital.

Keywords: clinical symptoms, co-morbidities, COVID-19 scoring system, early prediction of mortality, risk factors, Scoring and validation cohort

Introduction

The first human case of coronavirus disease 2019 (COVID-19), subsequently named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was reported by officials in Wuhan City, China, in December 2019. The COVID-19 virus has varied manifestation within the human population.[1] Infected people develop a wide spectrum of disease ranging from asymptomatic, mild to moderate and sometimes severe illness. However, over a period of time, the case fatality rate has reduced, but it remains high in older age groups, especially in patients with associated comorbidities.[2] Various studies have reported the common symptoms of this disease as fever, dry cough and tiredness, whereas less common symptoms are body aches, sore throat, diarrhoea, conjunctivitis, headache, loss of taste or smell, a rash on the skin or discolouration of fingers or toes, whereas more serious symptoms often include difficulty breathing or shortness of breath.[3] There was no proven effective therapy or vaccine available for COVID-19 patients up to the end of 2020. Similarly, in 2021, up to 15 May, various treatments and vaccines have been introduced and administered to patients, but their long-term acceptance and side effects are still unclear.[4,5] It has been reported that death is mainly observed in severely symptomatic patients, while many infected patients who presented with mild to moderate symptoms quickly recovered.[6] However, as per the literature, as of now there are significant differences in clinical symptoms and comorbidities between survivors (Ss) and non-survivors (NSs).[7]

Comorbid conditions have been associated with the adverse outcomes of COVID-19 infections.[8]

To date, most of the reported risk score models for COVID-19 mortality are based on the levels of inflammatory markers, comorbidities or various treatment modalities, and there is a paucity of risk score models based on clinical symptoms and comorbidities.[9–15] A robust risk score based on clinical symptoms and comorbidities will not only help to categorise patients at an early stage, but will also give guidance regarding an appropriate level of care and treatment.

In this study, we aim to identify the clinical symptoms and risk factors leading to mortality in COVID-19 patients and establish a scoring system (risk score) model to predict the mortality associated with this novel disease.

Materials and Methods

This is a hospital-based, analytical, retrospective observational study performed in a dedicated Level III COVID-19 care facility in a leading tertiary care hospital of North India. For the development of a COVID-19 scoring system (CSS), all consecutive COVID-19 patients (n=1349) confirmed by RT-PCR and discharged after treatment (either outcome) at this facility from 21 March to 15 October 2020 were enrolled for the study (study cohort or scoring cohort). To validate the result of the CSS, another 703 prospective consecutive COVID-19 patients confirmed by RT-PCR and discharged after treatment (either outcome) at this facility from 16 October to 14 December 2020 were enrolled for validation (validation cohort). The primary clinical outcome was defined as clinical recovery or death at the time of hospital discharge.

All the patients' data (retrospective data) were extracted from the discharge summaries of NSs and Ss available in the hospital information system, which is an electronic database of patients available via the institute's network. The extracted data were entered and reviewed by a team of clinicians and biostatisticians. In case the same patient was admitted more than once, only the latest admission was considered for the study. Demographic data of the patients, including age (y) upon admission, gender and clinical symptoms—fever, dry cough, sore throat, breathlessness or shortness of breath—were recorded for the study. Similarly, all associated comorbidities, such as diabetes, hypertension, renal, cardiac, lung, liver, malignancy and any other significant comorbidities, were also recorded.

The criteria for discharge were at least two RT-PCR negative test results and, in case of divergent test results, a third RT-PCR negative result was relied upon to discharge patients.

Clinical symptoms and comorbidity

Each patient's detailed history, including clinical symptoms and comorbidities, were recorded at the time of hospital admission through their available records and clinical observations. Symptoms were recorded as per Centers for Disease Control and Prevention (CDC) definitions.[16] Fever was defined as a temperature of $\geq$100.4°F. A cough was defined as a forceful expulsion of air from the lungs that helps to clear secretions, foreign bodies and irritants from the airways and is productive (with sputum/mucus expectoration) or dry. Breathlessness was defined as an intense tightening in the chest, air hunger, difficulty breathing, breathlessness or a feeling of suffocation. A sore throat was defined as consisting of pain, scratchiness or irritation of the throat that often worsens when you swallow. Diabetes was defined as a diagnosed case of diabetes mellitus or being on treatment with antidiabetic drugs. Diagnosed hypertension was defined as an individual undergoing treatment with antihypertensive medicines. A diagnosed case of chronic renal disease or end stage renal disease was defined as requiring renal replacement therapy. Defined heart disease was diagnosed as structural heart disease or a history of treatment for cardiac disease. Lung disease was defined as a diagnosed case of chronic lung disease with treatment for those conditions. Cancer was defined as a diagnosed case of malignancy confirmed by tissue diagnosis performed prior to admission.

Statistical analysis

Normality of continuous variables was assessed, and a variable was considered normally distributed when skewness was within $\pm$2. The Mann–Whitney U test was used to compare the distribution of age between NSs and Ss, whereas a χ^2 test was used to test the association between a patient's outcomes and demographic and clinical variables. Univariate analysis was used to identify the variables associated with mortality. From the significant factors in univariate analysis, diabetes was taken separately, whereas from the other comorbidities, one new variable, namely, 'any other co-morbidity (with or without diabetes)', was created. Furthermore, the significant variables were included in multivariate analysis (stepwise) to identify the independent risk factors for mortality. The regression coefficients (β)—validated by its 95% CI using bias-corrected and accelerated (BCa) in bootstrapping methods—and adjusted ORs were calculated for the independent risk factors evaluated through multivariate analyses. The calculated regression coefficients were multiplied by 2 (for increasing the size of the score) and the nearest integers ($<$0.5 was considered to be 0 whereas $\geq$0.5 was considered to be 1) were taken for the scoring system. The CSS was significantly higher in NSs compared with Ss, as evidenced from the Mann–Whitney U test, and the discrimination capability of the risk score between NSs and Ss and its appropriate cut-off value was assessed through the receiver operating characteristic (ROC) curve with corresponding sensitivity and specificity. Hierarchical cluster analysis was used to divide the risk score among three clusters in terms of their severity of risk score. The Hosmer–Lemeshow test was used to assess the goodness of fit between observed and predicted deaths. A nomogram chart was used to present the scoring model with predicted probability of mortality. $p<0.05$ was considered as statistically significant. Statistical package for social sciences, version 23 (SPSS-23, IBM, Chicago, USA), R Software version 3.3.0 (R Foundation for Statistical Computing, Vienna, Austria) and MedCalc Statistical Software version 20 (MedCalc Software Ltd, Ostend, Belgium) were used for data analysis.

Table 1. Distribution of demographic and clinical variables between survivors and non-survivors in COVID-19 patients (n=1349)

Variable	Total (n=1349)	Survivors (n=1156, 85.7%)	Non-survivors (n=193, 14.3%)	p
Age (y) (median, IQR)#	51 (36–61)	49 (35–59)	61 (52–68)	**<0.001**
Age (≥60 y)	383 (28.4%)	276 (23.9%)	107 (55.4%)	**<0.001**
Gender (female)	968 (71.8%)	820 (70.9%)	148 (76.7%)	0.100
Fever	819 (60.7%)	667 (57.7%)	152 (78.8%)	**<0.001**
Breathlessness	412 (30.5%)	275 (23.8%)	137 (71%)	**<0.001**
Cough	508 (37.7%)	377 (32.6%)	131 (67.9%)	**<0.001**
Sore throat	180 (13.3%)	157 (13.6%)	23 (11.9%)	0.529
Diabetes	428 (31.7%)	322 (27.9%)	106 (54.9%)	**<0.001**
Hypertension	440 (32.6%)	340 (29.4%)	100 (51.8%)	**<0.001**
Renal disease	194 (14.4%)	138 (11.9%)	56 (29%)	**<0.001**
Heart disease	109 (8.1%)	71 (6.1%)	38 (19.7%)	**<0.001**
Lung disease	101 (7.5%)	68 (5.9%)	33 (17.1%)	**<0.001**
Cancer	30 (2.2%)	29 (2.5%)	1 (0.5%)	0.083
Any other comorbidities	165 (12.2%)	149 (12.9%)	16 (8.3%)	**<0.001**
Any other comorbidity (with or without diabetes)	717 (53.2%)	568 (49.1%)	149 (77.2%)	**<0.001**

Median (Q1, Q3) compared by Mann–Whitney U test.
Frequency (%) compared by χ^2 test.
$p<0.05$ significant values in bold.

Results

From 21 March to 15 October 2020, a total of 1349 patients with COVID-19 were discharged from hospital. Of these, 1156 (85.7%) survived and 193 (14.3%) died. Mean and median age (y) of the patients were 48.24 and 51 y (IQR: 36–61 y, range of 3 mo to 92 y). Of these, 25.1% (n=338) were in the age group of ≥60 y. Most of the study patients (n=968, 71.8%) were male. The demographic and clinical characteristics of the study patients are given in Table 1. The median age (61 vs 49 y, $p<0.001$) and age of >60 y (55.4% vs 23.9%, $p<0.001$) were statistically significant whereas male gender (76.7% vs 70.9%, $p = 0.100$) was insignificant when comparing NSs with Ss. Similarly, the proportions of patients presenting with fever (78.8% vs 57.7%, $p<0.001$), breathlessness (71% vs 23.8%, $p<0.001$), cough (67.9% vs 32.6%, $p<0.001$) or any other symptoms (8.8% vs 3.7%, p=0.002) were significantly higher, whereas sore throat (11.9% vs 13.5%, p=0.340) was insignificantly lower in NSs compared with Ss. The proportions of diabetes (54.9% vs 27.9%, $p<0.001$), hypertension (51.8% vs 29.4%, $p<0.001$), renal disease (29% vs 11.9%, $p<0.001$), heart disease (19.7% vs 6.1%, $p<0.001$) and lung disease (17.1% vs 5.9%, $p<0.001$) were significantly higher in NSs compared with Ss. An interesting result was observed for cancer and other comorbidities, where the presence of cancer (0.5% vs 2.5%, p=0.083) was insignificant and any other comorbidities (8.3% vs 12.9%, $p<0.001$) were significantly lower in NSs compared with Ss (Table 1).

Risk factors and the prediction model for death in COVID-19 patients

In total, 13 variables were analysed by univariate binary logistic regression and, of those, all except gender (p=0.110), sore throat (p=0.529) and cancer (p=0.083) were significantly associated with mortality. From these 10 variables, fever was significantly associated with cough ($p<0.001$). Similarly, hypertension and diabetes ($p<0.001$) were highly associated with each other, and hypertension was also significantly associated with other comorbidities (each $p<0.05$). To overcome these collinearities and with the objective of developing a community friendly scoring system that could easily be calculated without any complexity by healthcare providers, six variables (hypertension, renal disease, heart disease, lung disease, cancer and any other comorbidities) were replaced by a newly created variable. This variable, 'any other comorbidity (with or without diabetes)', was also significant in the univariate analysis. In the multivariate analysis, all the significant variables—age of >60 y, diabetes, at least one comorbidity with or without diabetes, fever, breathlessness and cough—were included. By using lasso binary logistic regression, in the first step, fever was excluded from the model (highly associated with cough, $p<0.001$), while the remaining five variables were statistically significant (each $p<0.05$) and considered as independent predictors of the mortality and were included in the final CSS. The weight for each factor associated with mortality was obtained by regression coefficients observed in a multivariate binary logistic regression model. The coefficients for each variable and results of the multivariate analysis are given in Table 2.

All the regression coefficients were further validated by bootstrapping 95% CIs. All the CIs (lower and upper limit) were >0, indicating a significant value of the regression coefficient (Figure 1). The CSS was generated as shown in Table 3. The score distributions were as follows: age of ≥60 y, 2 points; presenting with cough, 2 points; having breathlessness, 3 points; presence of diabetes, 1 point; and any other comorbidity with or without diabetes, 1 point (Table 3). Predicted score for the risk of mortality of COVID-19 patients ranged from 0 to 9; 20% of patients had

Table 2. Independent predictors of mortality in COVID-19 patients (n=1349)

Variable	Regression coefficient (β)	Adjusted OR			p
		Value	95% CI		
			Lower	Upper	
Age (≥60 y)	0.898	2.46	1.72	3.51	**<0.001**
Cough	0.862	2.37	1.65	3.41	**<0.001**
Breathlessness	1.62	5.06	3.51	7.29	**<0.001**
Diabetes	0.537	1.71	1.20	2.45	**0.003**
Any other comorbidity (with or without diabetes)	0.724	2.06	1.38	3.08	**<0.001**

Multivariate binary logistic regression analysis was used.
p<0.05 significant values in bold.

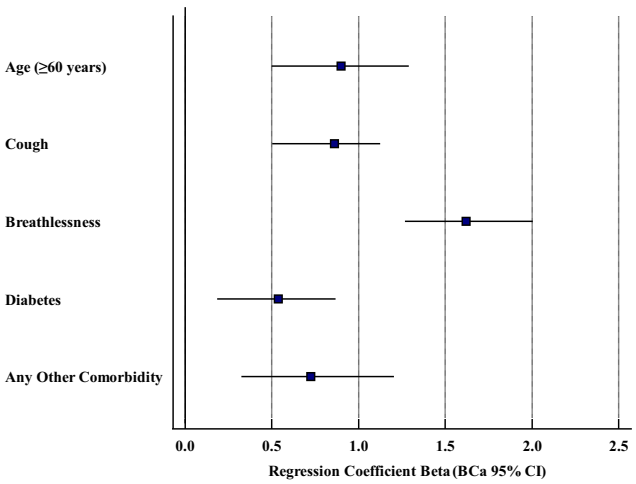

Figure 1. Forest plot showing the regression coefficient and its 95% Confidence Interval computed through bootstrapping 1000 samples. (n=1349).

a score of 0 whereas 4.1% had a score of 9. Contributions of the individual risk factors in the CSS and the corresponding predicted probability are presented in the nomogram (Figure 2).

Accuracy of the CSS

The area under the ROC (AUROC) curve indicates that the accuracy of the scoring system (from the scoring cohort, n=1349) was 82.8% (95% CI 79.6 to 85.9%; p<0.001) (Figure 3). The ROC curve revealed that for an at-risk score of ≥3, sensitivity and specificity were 88.1% and 55.1%, respectively. Similarly, sensitivity and specificity were observed as 74.1% and 78.5% (at ≥5) and 58.5% and 86.9% (at ≥6), respectively (Table 4). The scores were divided into three groups using hierarchical cluster analysis and indicated that score group 1 (0 to 4, n=958/1349, 71.01%), score group 2(5 to 7, n=290/1349, 21.5%) and score group 3(8 to 9, n=101/1349, 7.5%) had a mortality of 5.2%, 27.6% and 62.4%, respectively, demonstrating significant associations between the score groups (clusters) and patients' mortality (p<0.001).

The CSS developed from the study cohort (n=1349) showed great prognostic accuracy, as evidenced by the ROC analysis. The performance of this model was validated by the Hosmer–Lemeshow goodness of fit test, which demonstrated good agreement (87.3%) between predicted and observed mortality with non-significant difference (p<*i*=0>0.398) in the study cohort, indicating that there was no departure from perfect fit. For depiction, the corelation coefficient between the probability of observed deaths and probability of predicted deaths was calculated, which indicated a very strong correlation in the scoring cohort (ρ=0.989, p<0.001) (Figure 4).

Table 3. Final COVID-19 risk score (n=1349)

Variable	Regression coefficient (β)		β*2	Final score
	Value	BCa 95% CI		
Age (≥60 y)	0.898	0.498 to 1.290	1.796	2
Cough	0.862	0.503 to 1.125	1.724	2
Breathlessness	1.62	1.269 to 2.005	3.24	3
Diabetes	0.537	0.185 to 0.867	1.074	1
Any other comorbidities with or without diabetes	0.724	0.323 to 1.204	1.448	1

BCa 95% CI=bias-corrected and accelerated (BCa) 95% CI.
Bootstrap results are based on 1000 bootstrap samples.

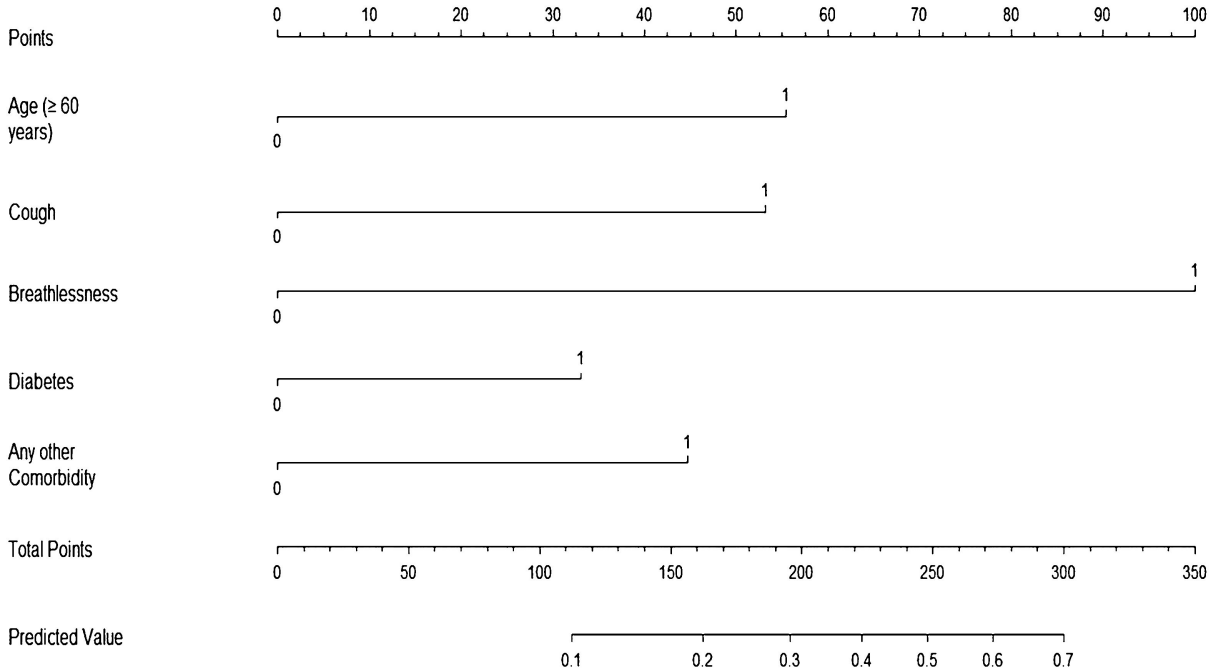

Figure 2. A nomogram describing the relationship between the calculated score and the probability of COVID-19 death. (n=1349).

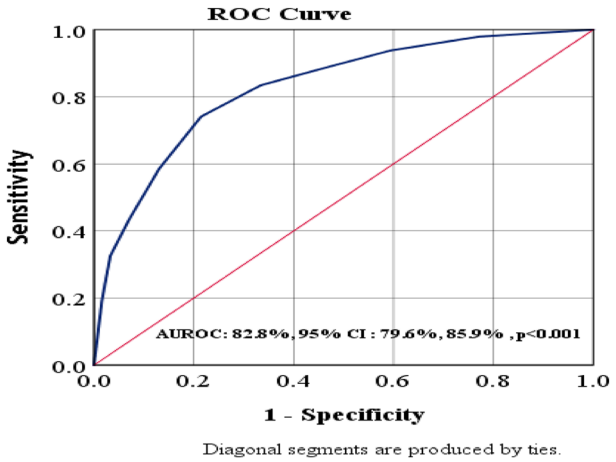

Figure 3. Area under the ROC curve showing the diagnostic accuracy of the COVID-19 risk score model (n=1349).

Table 4. Diagnostic accuracy of observed COVID-19 risk scores in the study cohort (n=1349)

No. of patients	Risk sccore	Sensitivity	Specificity
269	0	100	0
213	1	97.9	22.9
178	2	93.8	40.7
142	3	88.1	55.1
156	4	83.4	66.6
127	5	74.1	78.5
95	6	58.5	86.9
68	7	44	92.7
46	8	32.6	96.7
55	9	19.2	98.4

A larger risk score indicates an increased risk of mortality. AUROC curve: 82.8% (95% CI 79.6 to 85.9%; p<0.001).

Performance of the prediction model through independent validation

The CSS, which was calculated using a cohort of 1349 COVID-19 patients, was validated through an independent prospective cohort (n=703) of COVID-19 patients admitted to the same hospital and receiving the same treatment protocol. The AUROC curve of the independent validation cohort of 703 COVID-19 patients was 81.9% (95% CI 78.2 to 85.6%; p<0.001) (Figure 5), which was slightly less than the risk-scoring cohort of 1349 patients (AUROC: 82.8%; p<0.001). In the validation cohort (n=703), score group 1 (0 to 4, n=449/703, 63.9%), score group 2 (5 to 7, n=192/703, 27.3%) and score group 3 (8 to 9, n=62/703, 8.8%) had a mortality of 6.5%, 28.60% and 54.80%, respectively, which were almost the same as the scoring cohort (5.2%, 27.6% and 62.4%, respectively), indicating the usefulness of the CSS in predicting the mortality of COVID-19 patients (Figure 6).

Discussion

In this study, a total of 1349 COVID-19 patients (study or scoring cohort) were analysed for development of the CSS and another 703 prospective and consecutive COVID-19 patients (validation

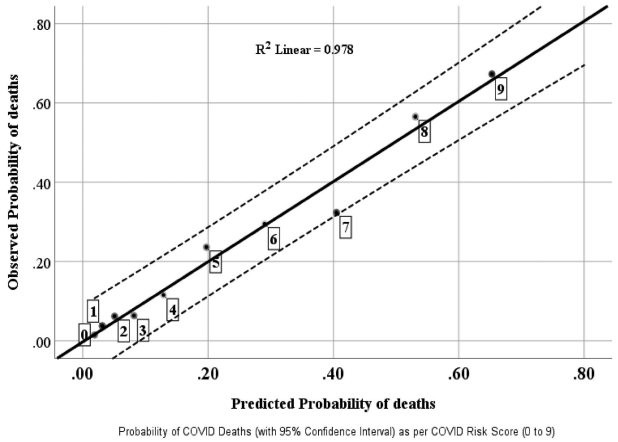

Figure 4. Calibration curve showing the agreement in the observed and predicted probability of death for individual COVID-19 risk scores in the scoring cohort (n=1349).

cohort), whose data were retrospectively collected from the discharged summary, were used for validation of the CSS model. In the scoring cohort (n=1349), the proportions of NSs and Ss were 14.3% (n=193) and 85.7% (n=1156), respectively. Age of ≥60 y, cough, breathlessness, diabetes and any other comorbidity (with or without diabetes) were significant and independent risk factors in NSs and were included in the CSS model. In the CSS model, the score ranged from 0 to 9 and, out of the variables, breathlessness (score 3) contributed the most, followed by age of ≥60 y (score 2) and cough (score 2), whereas diabetes (score 1) and any other comorbidity with or without diabetes (score 1) contributed the least to the scoring system. Diagnostic accuracy in terms of the AUROC curve of the model was 82.8% (95% CI 79.6 to 85.9%; p<0.001) and no significant difference was observed between observed and predicted mortality in the scoring cohort revealed by the Hosmer–Lemeshow goodness of fit test (p<>0.05i>). Similarly, the AUROC curve of the independent validation cohort of 703 COVID-19 patients was 81.9% (95% CI 78.2 to 85.6%; p<0.001), which was close to the scoring cohort. Proportions of predicted NSs (probability of NSs) from the scoring cohort and observed NSs in the validation cohort were similar in each of the three scoring groups, indicating the usefulness of the CSS for the prediction of mortality in COVID-19 patients.

The identified risk factors for mortality in COVID-19 patients in the current study were almost similar to those presented in other available studies.[9–15] To date, only a few studies have been conducted to develop a COVID-19 risk-scoring model and all of them have been based on demographics, inflammatory markers, treatment and investigations as variables.[9–15] The CSS developed in our study is different to that reported in other studies conducted to calculate COVID-19 risk scores (i.e. scoring system), as it is based on clinical symptoms along with demographic measurements and comorbidities, which makes it easier and faster to calculate.

Zhao et al. developed a scoring model to predict mortality in severe COVID-19 patients. They found that the presence of heart disease, chronic obstructive pulmonary disease, elevated procalcitonin and LDH levels, elevated heart rate, lower peripheral oxygen saturation and increasing age were independent risk factors. The risk score for predicting mortality ranged from 0 to 7. The risk score model yielded good accuracy with an AUROC curve of 83% (95% CI 73 to 92%).[9] Altschul et al.'s model for a risk score to predict mortality in COVID-19 patients ranged from 0 to 10. The variables identified as independent predictors of mortality comprised a patient's age, oxygen saturation, mean arterial pressure, blood urea nitrogen, C-reactive protein and the international normalised ratio. An ROC curve analysis performed in the derivation cohort achieved an AUROC curve of 82.4% (95% CI 81.4 to 85.1%) and an AUROC curve of 79.8% (95% CI 78.9 to 81.8%) in the validation cohort.[10] A study conducted by Shang et al. developed a scoring model to predict mortality in severe COVID-19 patients. This study concluded that old age, coronary heart disease, lymphocyte percentage, elevated procalcitonin and D-dimer were independently related to mortality. Based on regression coefficients, a risk score was developed that showed the usefulness of the model in predicting mortality and complications in independent validation data.[11] Fumagalli et al. conducted a study to develop a risk score to predict in-hospital mortality in COVID-19 patients using a retrospective cohort of patients after considering hospital stay days. Cox regression analysis showed that increasing age (≥75 y), number of chronic diseases (≥4), respiratory rate, PaO_2/FiO_2 ratio, serum creatinine and platelet count were independent predictors of mortality and were used to build the COVID-19 risk score, which proved to be highly accurate in stratifying patients to a low, intermediate or high risk of in-hospital death (p<0.001).[12] Gue et al. conducted a study to develop a risk score for prediction of 30-d hospital mortality. The independent predictors were age, gender, platelet count, international normalised ratio, quick sequential organ failure and assessment score. The predictive risk score ranged from 0 to 10 with an AUROC curve of 79% (95% CI 75 to 84%).[13] Wynants et al. performed a systemic review of all the available models for diagnosis and prognosis in COVID-19-infected patients. They identified 50 models for COVID-19 prognosis. Of these, 22 models predicted mortality, while others used different endpoints of prognosis. The most common factors were age, comorbidities, gender, lymphocyte counts, C-reactive protein and imaging features along with other clinical variables. Moreover, these studies do not reveal the clear intent of using these models.[14] Barda et al. reported their clinical model for prediction of mortality using baseline respiratory infections as variables, with an AUROC curve of 82%.[15]

Besides mortality risk score models, there are other studies available that developed a risk score model for severity of COVID-19 disease,[17,18] while other studies have identified the risk factors for severity of COVID-19 disease, but did not develop a risk score model.[19–21]

CDC has provided guidelines for the population at risk.[22] These guidelines define those patients who can contract infection easily and are at risk of the disease. However, they do not comment on the course of disease with these factors.

In the available risk-scoring models, a patient requires to be admitted to hospital and investigated before a score can be calculated.[9–15] However, in our scoring system, we can assess the possible mortality risk even before patients are admitted to hospital. Hence, this score is proposed for determining an appropriate level of care upon a patient's admission, as a higher score indicates a requirement for more intense and specialised care, thus prompting allocation to an appropriate hospital. However,

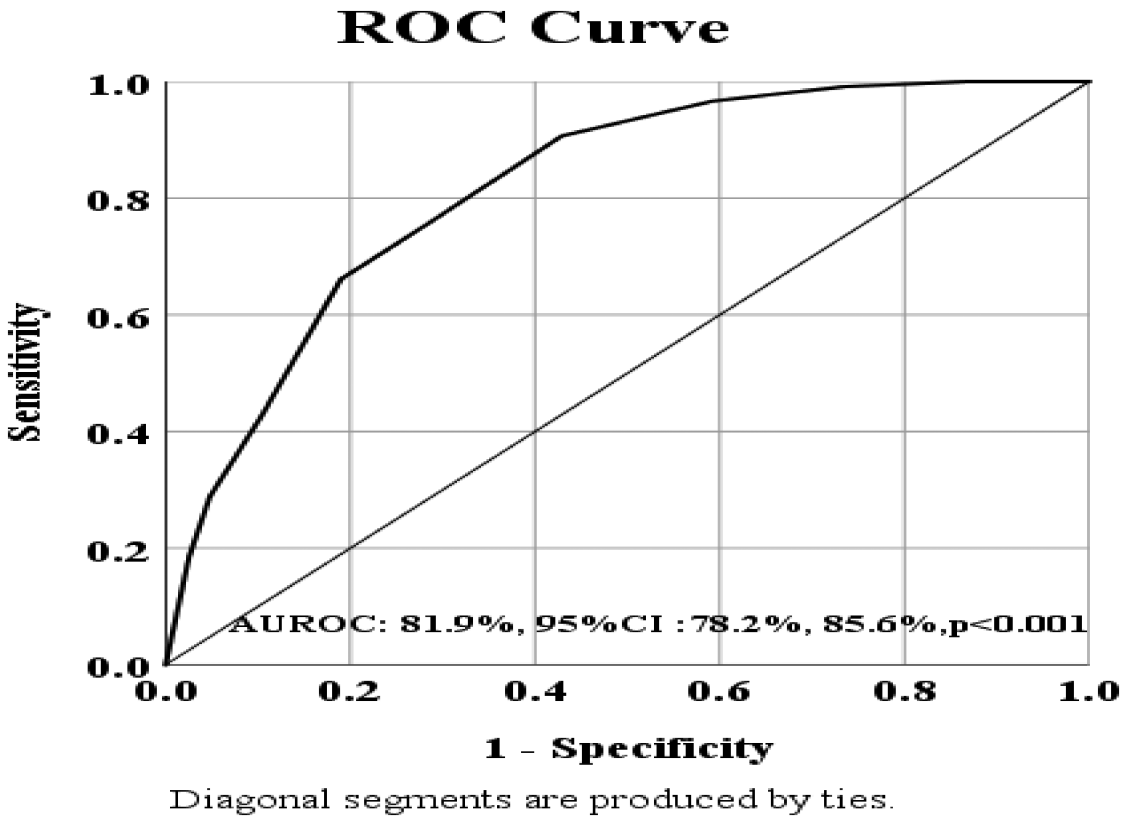

Figure 5. Area under the ROC curve showing the diagnostic accuracy of the COVID-19 risk score model in the validation cohort (n=703).

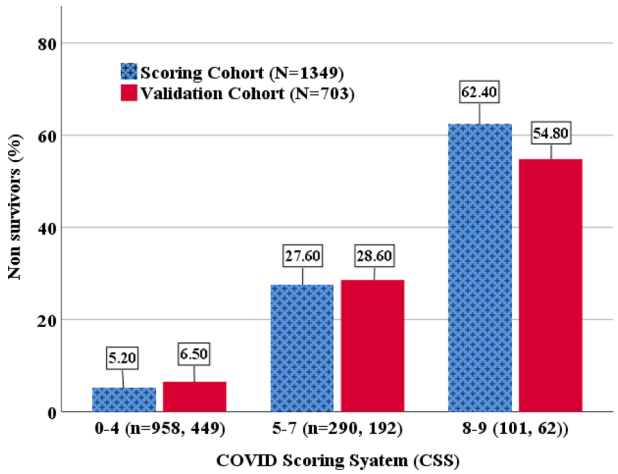

Figure 6. Comparison of the deaths between the scoring (n=1349) and validation (n=703) cohorts for three COVID-19 scoring system (CSS) groups.

it is clear that the final outcome for a patient depends on progression of the disease and its response to treatment modalities that are still changing as more evidence is rolled out.

Although the model developed in our study is robust and can be used even at the primary level by a basic healthcare worker as an important tool for triage, there remain a few limitations. Our study was conducted in a cohort of referred patients already admitted to a speciality hospital for treatment and this a major limitation of our study. We recommend that our unique model should be tested in the general population with people who are diagnosed with COVID-19, for optimal testing of the strength of our scoring model in future studies. This would achieve an important goal of early care-based prediction in COVID-19 patients, which is desirable as there are substantial yet variable mortality rates in a subpopulation of these patients.

Conclusions

Patient age of ≥60 y, cough, breathlessness, diabetes and any other comorbidity (with or without diabetes) are significant and independent risk factors for non-survival among COVID-19 patients. The CSS, ranging from 0 to 9 and designed for early prediction of mortality in severe COVID-19 patients, showed good sensitivity and specificity (i.e. 74.1% and 78.5% at CSS≥5, respectively), with an overall diagnostic accuracy of 82.8%, which was close to the diagnostic accuracy detected in the validation cohort (81.9%). In the validation cohort, high (8–9), medium (5–7) and low (0–4) CSS groups had 54.80%, 28.60% and 6.5% observed mortality, respectively, which was very close to predicted mortality (62.40%, 27.60% and 5.2%, respectively, by scoring cohort). The CSS shows a positive relationship between a higher score and proportion of mortality and, as its validation showed, it is useful for the prediction of risk of mortality in COVID-19 patients at

an early stage, so that referral for triage and admission can be predetermined even before admission to hospital.

Authors' contributions: RKD, RKS and SP conceived the study. PM, AN, AA, OPS, DG, PS, TG, GH, VK, SK and AR performed the study, analysed the data, wrote and revised the manuscript. All the authors read and approved the final version of the manuscript.

Funding: None.

Competing interests: None declared.

Ethical approval: The current study received ethical clearance from the Sanjay Gandhi Post Graduate Institute of Medical Sciences Lucknow Ethical Committee [2021–54-IP-EXP-36].

Data availability:

Data used in this article is available at Hospital Information system (HIS), which is an electronic database of patients available via the institute's network. http://www.sgpgi.ac.in.

References

1 Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.

2 Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507–13.

3 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(11):1061–9.

4 COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov [last accessed May 15, 2021].

5 Coronavirus disease (COVID-19): Vaccines. Available at https://www.who.int/covid-19/vaccines [last accessed May 15, 2021].

6 Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020: NEJMoa2002032.

7 Sun Y, Koh V, Marimuthu K, et al. Epidemiological and clinical predictors of COVID-19. Clin Infect Dis. 2020;71(15):786–92.

8 Atkins JL, Masoli JAH, Delgado J, et al. Preexisting comorbidities predicting COVID-19 and mortality in the UK Biobank Community Cohort. J Gerontol A Biol Sci Med Sci. 2020;75(11):2224–30.

9 Zhao Z, Chen A, Hou W, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One. 2020;15(7):e0236618.

10 Altschul DJ, Unda SR, Benton J, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. Sci Rep. 2020;10(1):16726.

11 Shang Y, Liu T, Wei Y, et al. Scoring systems for predicting mortality for severe patients with COVID-19. Clin Med Lancet. 2020;24:100426.

12 Fumagalli C, Rozzini R, Vannini M, et al. Clinical risk score to predict in-hospital mortality in COVID-19 patients: a retrospective cohort study. BMJ Open. 2020;10(9):e040729.

13 Gue YX, Tennyson M, Gao J, et al. Author correction: development of a novel risk score to predict mortality in patients admitted to hospital with COVID-19. Sci Rep. 2021;11(1):8011.

14 Wynants L, Calster BV, Collins GS, et al. Prediction models for diagnosis and prognosis of Covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328.

15 Barda N, Riesel D, Akriv A, et al. Developing a COVID-19 mortality risk prediction model when individual-level data are not available. Nat Commun. 2020;11(1):4439.

16 Centers for Disease Control and Prevention. Available at https://www.cdc.gov [last accessed June 15, 2021].

17 Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. Clin Infect Dis. 2020;71(6):1393–9.

18 Xu W, Huang C, Fei L, et al. A novel prediction model of COVID-19 progression: a retrospective cohort study. Infect Dis Ther. 2021;10(3):1491–504.

19 Zadori N, Vancsa S, Farkas N, et al. The negative impact of comorbidities on the disease course of COVID-19. Intensive Care Med. 2020;46(9):1784–6.

20 Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA. 2020;323(20):2052–9.

21 Wu JT, Leung K, Bushman M, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nat Med. 2020;26(4):506–10.

22 Centres for Disease Control and Prevention. People Who Are at Higher Risk for Severe Illness, Vol. 2020 (Centres for Disease Control and Prevention, 2020). Available at https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html [last accessed June 24, 2021].