

## ORIGINAL PAPER

## Infectious diseases

# Performance of the CURB-65, ISARIC-4C and COVID-GRAM scores in terms of severity for COVID-19 patients

Fatih Doğanay<sup>1</sup>  | Rohat Ak<sup>2</sup> <sup>1</sup>Department of Emergency Medicine, Edremit State Hospital, Balıkesir, Turkey<sup>2</sup>Department of Emergency Medicine, Dr. Lütfi Kırdar Kartal Eğitim ve Araştırma Hastanesi, İstanbul, Turkey**Correspondence**

Fatih Doğanay, Department of Emergency Medicine, Edremit Devlet Hastanesi, Acil tıp kliniği, Cennet Ayağı Mahallesi, 781. Sokak No: 7, PC: 10300, Edremit/Balıkesir, Turkey. Email: drdoganay@gmail.com

**Funding information**

The authors received no financial support for the research, authorship, and/or publication of this article.

**Abstract**

**Background:** In the COVID-19 pandemic, difficulties have been experienced in the provision of healthcare services because of excessive patient admissions to hospitals and emergency departments. It has become important to use clear and objective criteria for the early diagnosis of patients with high-risk classification and clinical worsening risk.

**Objective:** The aim of this study was to assess the prognostic accuracy of CURB-65, ISARIC-4C and COVID-GRAM scores in patients hospitalised for COVID-19 and to compare the scoring systems in terms of predicting in-hospital mortality and intensive care unit requirement.

**Methods:** The files of all COVID-19 patients over the age of 18 who were admitted to the emergency department and hospitalised between September 1, 2020 and December 1, 2020 were retrospectively scanned. The area under the receiver operating characteristic curve and Youden J Index were used to compare scoring systems for predicting in-hospital mortality and intensive care requirement.

**Results:** There were 481 patients included in this study. The median age of the patients was 67 (52-79). In terms of in-hospital mortality, the AUC of CURB-65, ISARIC-4C and COVID-GRAM were 0.846, 0.784 and 0.701 respectively. In terms of intensive care requirement, the AUC of CURB-65, ISARIC-4C and COVID-GRAM were 0.898, 0.797 and 0.684 respectively. In our study, Youden's J indexes of CURB-65, ISARIC-4C and COVID-GRAM scores were found to be 0.59, 0.27 and 0.01 respectively, for mortality prediction of COVID-19 patients. Whereas Youden's J indexes were found to be 0.63, 0.26 and 0.01 respectively for determining intensive care requirement.

**Conclusions:** Among the scoring systems assessed, CURB-65 score had better performance in predicting in-hospital mortality and ICU requirement in COVID-19 patients. ISARIC-4C has been found successful in identifying low-risk patients and the use of the ISARIC-4C score with CURB-65 increases the accuracy of risk assessment.

## 1 | INTRODUCTION

The existence of a new Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was first reported in China and later spread throughout the world straining the health systems of many countries.<sup>1</sup> The viral pneumonia associated with SARS-CoV-2 has

been officially named Coronavirus disease 2019 (COVID-19).<sup>2</sup> During the pandemic period, difficulties were experienced in the provision of health services because of excessive patient admissions in hospitals and emergency departments (EDs).

In the first reports, it was stated that ~25% of patients required an intensive care unit (ICU).<sup>3</sup> Mortality in hospitalised COVID-19

patients is related to acute respiratory failure, especially in those with comorbidities.<sup>4-6</sup>

It is important to use clear and objective criteria for risk stratification and for early diagnosis of patients who are at a high risk of clinical worsening. Scoring systems used for this purpose are valuable tools that help clinicians predict outcomes and guide treatment-related decisions.<sup>7</sup> Among these scoring systems, the CURB-65 score has been developed as a clinical prediction rule suitable for use in the ED. In addition, this score is used to predict the prognosis of patients using variables that can be easily measured during the initial evaluation.<sup>8</sup> It has been reported that this score, which is widely used for pneumonia patients, also has a strong predictive value in COVID-19 patients.<sup>9</sup> Scoring systems that are specific to COVID-19 patients have also been developed. Among these, the Acute Respiratory Infection Consortium Clinical Characterization Protocol-Coronavirus Clinical Characterization Consortium (ISARIC-4C) score is derived from a prospective observational cohort study based on COVID-19 patients admitted to 260 hospitals in England, Scotland and Wales and provides information about the prognosis of patients.<sup>10</sup> According to this scoring system, it has been reported that patients in the low risk group (mortality rate 1%) can be followed on an outpatient basis, while patients in the medium risk group (mortality rate 10%) should be admitted to the hospital.<sup>10</sup> Similarly, the purpose of the COVID-GRAM score is to help prediction of the COVID-19 patients' risk rate for critical illness.<sup>11</sup> It has been suggested that monitored follow-up of patients in the low-risk group for critical illness will be sufficient, and that patients in the high-risk group should receive more aggressive treatment.<sup>11</sup>

The aim of this study was to determine the prognostic accuracy of CURB-65, ISARIC-4C and COVID-GRAM scores in patients hospitalised for COVID-19 and to compare the scores with each other in this regard.

## 2 | MATERIALS AND METHODS

### 2.1 | Study design and settings

This retrospective observational study was carried out in the ED of Kartal Dr Lutfi Kırdar City Hospital between September 1, 2020 and December 1, 2020. The institutional review board approved the analysis and issued a waiver of consent (Ethics Committee Ruling number: 514/196/24).

The Hospital, which is in the east of the Istanbul, is a centre that diagnoses and treats ~3 million patients annually with a capacity of 971 inpatient units (IU) and 224 intensive care units (ICU). From March 11, 2020—the first seen case of COVID-19 in the country—until March 2021, 322 390 outpatients, 7015 IU patients and 1286 ICU patients of COVID-19 have diagnosed and treated in this hospital. Throughout the pandemic, 38 of the 86 beds in the ED of the hospital were reserved for COVID-19 patients. During the period included in the study, 45 428 patients were diagnosed with COVID-19 by reverse transcriptase-polymerase chain reaction (RT-PCR).

### What's known

The prognostic value of COVID-GRAM and ISARIC-4C scores in COVID-19 patients have been studied separately in the literature. However, when examining the prognostic values of these scores in studies, generally only comparisons were made with the AUC value. In addition, false positive and false negative cases were not perscrutated.

### What's new

When calculating AUC with ROC analysis, converting the risk ranking of scoring systems to 3-stage (low-medium-high risk) and then comparing their AUC values will give more accurate results. We think that comparing these COVID-19 scoring systems with a widely used and validated scoring system such as CURB-65 will contribute to the literature.

During this period, 359 patients were hospitalised in ICU and 1958 patients were hospitalised in IU because of COVID-19.

The guideline used in determining the ICU requirement of patients in the hospital where the study was conducted is the "COVID-19 Diagnosis and Treatment Guide" published by the Ministry of Health (Table S1).<sup>12</sup>

Typical thoracic computed tomography (CT) findings for COVID-19 were recorded according to CT features previously described for COVID-19.<sup>13</sup>

### 2.2 | Selection of participants

All COVID-19 patients over the age of 18 who were hospitalised between September 1, 2020 and December 1, 2020 were included in this study. The diagnosis of COVID-19 was determined based on the World Health Organization (WHO) guidelines. This study includes only patients who had positive results in the real-time RT-PCR test of nasal and pharyngeal swab samples.<sup>14</sup> The digital records of the Hospital Information Management System (HIMS) was used to collect data. Vital parameters and symptoms at the time of first admission and laboratory tests performed in the ED were also taken into consideration. The patients, whose variables for the CURB-65, ISARIC-4C and COVID-GRAM scores that could not be reached, were not included in this study (Table S2). Additionally, patients who needed CPR in the ED, who died in the ED and who were pregnant were not included in this study.

### 2.3 | Measurements

For the patients who were included in this study, their age; gender; CT findings; symptoms; Glasgow coma scale (GCS) score; chronic

diseases; vital signs including body temperature (Temp), heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate (RR), peripheral oxygen saturation (spO<sub>2</sub>); and laboratory tests including white blood cell (WBC), neutrophil (Neu), lymphocyte (Lym), C-reactive protein (CRP), lactate dehydrogenase (LDH), urea, direct bilirubin (D Bil.) were recorded on a form. All parameters evaluated for the calculation of points for the scores were obtained from HIMS electronic records. A trained physician, blind to the purpose and outcomes of the study, recorded the demographic characteristics and parameters used in the study with using a standard template. The second author recorded the patients' outcomes. Before data analysis, the first author performed merging and

checking of datasets. Patients with missing data were excluded by the first author.

## 2.4 | Outcomes

The patient's hospitalisation outcome, ie survivor and non-survivor groups, represent in-hospital mortality. The primary outcome was to determine the diagnostic accuracy of each scoring system for in-hospital mortality. The patient's ED outcome, that is, IU and ICU groups, represents the first unit where the patient is hospitalised from the ED. The secondary outcome was to determine the

TABLE 1 Comorbidities and categorical descriptives of the study population

Variable	Category	Mortality groups					Intensive Care requirement groups				
		Survivors (n = 361)		Non-Survivors (n = 120)		Sig.	IU (n = 396)		ICU (n = 85)		Sig.
		n	% <sup>a</sup>	n	% <sup>a</sup>	P	n	% <sup>a</sup>	n	% <sup>a</sup>	P
Sex	Male	189	74.1	66	25.9	.615	204	80.0	51	20.0	.155
	Female	172	76.1	54	23.9		192	85.0	34	15.0	
COPD	No	344	75.8	110	24.2	.135	377	83.0	77	17.0	.094
	Yes	17	63.0	10	37.0		19	70.4	8	29.6	
Astma	No	341	74.6	116	25.4	.336	374	81.8	83	18.2	.219
	Yes	20	83.3	4	16.7		22	91.7	2	8.3	
DM	No	270	75.0	90	25.0	.964	295	81.9	65	18.1	.703
	Yes	91	75.2	30	24.8		101	83.5	20	16.5	
HT	No	248	75.8	79	24.2	.560	266	81.3	61	18.7	.410
	Yes	113	73.4	41	26.6		130	84.4	24	15.6	
CHF	No	346	76.7	105	23.3	.001	378	83.8	73	16.2	.001
	Yes	15	50	15	50		18	60	12	40	
CAD	No	330	75.7	106	24.3	.316	364	83.5	72	16.5	.038
	Yes	31	68.9	14	31.1		32	71.1	13	28.9	
CRF	No	348	77.7	100	22.3	<.001	372	83.0	76	17.0	.134
	Yes	13	39.4	20	60.6		24	72.7	9	27.3	
Malignancy	No	350	76.3	109	23.7	.005	378	82.4	81	17.6	.949
	Yes	11	50.0	11	50.0		18	81.8	4	18.2	
CND	No	350	76.1	110	23.9	.014	382	83.0	78	17.0	.054
	Yes	11	52.4	10	47.6		14	66.7	7	33.3	
A & D	No	351	76.8	106	23.2	<.001	379	82.9	78	17.1	.130
	Yes	10	41.7	14	58.3		17	70.8	7	29.2	
Dyspnea	No	225	85.9	37	14.1	<.001	240	91.6	22	8.4	<.001
	Yes	136	62.1	83	37.9		156	71.2	63	28.8	
CT finding	No	116	96.7	4	3.3	<.001	120	100	0	0.0	<.001
	Yes	245	67.9	116	32.1		276	76.5	85	23.5	
UNC	Normal	354	79.4	92	20.6	<.001	388	87	58	13	<.001
	Altered	7	20	28	80		8	22.9	27	77.1	

Abbreviations: A & D, alzheimer and/or demencia; CAD, coronary artery disease; CHF, congestive heart failure; CND, chronic neurological disease; COPD, chronic obstructive pulmonary disease; CRF, chronic renal failure; CT, computed tomography; DM, diabetes mellitus; HT, hypertension; ICU, intensive care unit; IU, inpatient unit; Sig, asymptotic 2-sided significance between groups with chi-square test; UNC, unconsciousness.

<sup>a</sup>The percentages of patients falling into groups according to the variables were calculated as the percentage of the rows.

diagnostic accuracy of each scoring system for ICU requirement. Outcomes were retrospectively assessed by reviewing the hospital medical database.

## 2.5 | Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics (IBM Corp. Released 2020. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY) and MedCalc Statistical Software version 19.0.6 (MedCalc Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2019). To examine the relationship between the groups and the scoring systems, normality was tested with Kolmogorov-Smirnov test firstly. None of the age, scoring system points, vital parameters or laboratory data could meet the normality assumption. The population distribution of the in-hospital mortality and ICU requirement groups for the points of the scoring systems is presented in Figure 3.

CURB-65 and ISARIC-4C scores of the patients were determined according to the points they received from each category (Table S1).

First, the COVID-GRAM risk score was determined using the following formula, " $\alpha = (\text{CT abnormality} \times 1.2205) + (\text{Age} \times 0.0276) + (\text{Hemoptysis} \times 1.5116) + (\text{Dyspnea} \times 0.632) + (\text{Unconsciousness} \times 1.5494) + (\text{Number of comorbidities} \times 0.4668) + (\text{Cancer history} \times 1.4037) + (\text{Neutrophil-to-lymphocyte ratio} \times 0.0562) + (\text{Lactate dehydrogenase} \times 0.0024) + (\text{Direct bilirubin} \times 0.1376) - 6.6127$ ." Then the risk rate was calculated using the formula, "COVID-GRAM risk rate (%) =  $e^{\alpha} / (1 + e^{\alpha})$ ."<sup>11</sup>

Comparisons of mortality and ICU groups were analysed using Mann-Whitney *U* test for numerical data and chi-square test for categorical data. Numerical data were reported as medians and interquartile ranges (25th-75th), while categorical data were reported as frequencies and percentages (Tables 1 and 2).

Among the scoring systems, the COVID-GRAM score is divided into three risk groups, ISARIC-4C into four risk groups and the CURB-65 into five risk groups. When comparing the scores, all three were standardised into three risk groups as low, medium and high. Based on the literature, this was done to avoid being unstandardised because of the different number of groupings.<sup>10,11,15,16</sup> They were organised with three risk thresholds: for COVID-GRAM low (<1.7%),

TABLE 2 Descriptive statistics for age, vital parameter, laboratory measurements and critically illness prediction scores of the groups

Variable	All patients (n = 481)	Mortality groups		P	ICU requirement groups		P
		Survivor (n = 361)	Non-survivor (n = 120)		IU (n = 396)	ICU (n = 85)	
Age, median (25th-75th)							
Age	67 (52-79)	63 (49-76)	75 (66-82)	<.001	64 (50-78)	75 (65.50-81)	<.001
Critically illness prediction scores, median (25th-75th)							
CURB-65 (0-5)	1.00 (0.00-2.00)	1.00 (0.00-2.00)	2.00 (2.00-3.00)	<.001	1.00 (0.00-2.00)	3.00 (2.00-3.00)	<.001
ISARIC-4C (0-21)	8.00 (4.00-11.00)	6.00 (3.00-9.50)	13.00 (11.00-15.75)	<.001	7.00 (3.00-10.00)	14.00 (11.50-16.00)	<.001
COVID-GRAM (0-100)	66.02 (30.11-94.60)	47.2 (22.39-80.99)	97.20 (84.05-99.81)	<.001	51.25 (24.29-83.74)	98.39 (86.15-99.94)	<.001
Vital signs and GCS, median (25th-75th)							
SBP (mmHg)	120 (110-138)	120 (110-138)	120.5 (110-137)	.437	120 (110-138)	126 (106-139)	.404
DBP (mmHg)	72 (66-80)	73 (68-80)	71 (65-80)	.114	71.5 (67-80)	76 (63-80)	.672
HR (bpm)	87 (77-98.50)	85 (75-96)	95.50 (81-106)	<.001	85 (75-96)	97 (85-110)	<.001
RR (bpm)	20 (16-26)	18 (16-22)	34 (24-42)	<.001	18 (16-22)	40 (34-44)	<.001
Temp (°C)	36.70 (36.2-37.2)	36.70 (36.20-37.20)	36.70 (36.33-37.40)	.452	36.70 (36.20-37.20)	37.00 (36.45-37.45)	.063
spO <sub>2</sub> (%)	95 (90-97)	96 (93-98)	87 (81-93.75)	<.001	96 (93-98)	85 (80-87.50)	<.001
GCS	15 (15-15)	15 (15-15)	15 (13.25-15)	<.001	15 (15-15)	15 (13-15)	<.001
Laboratory measurements, median (25th-75th)							
WBC (10 <sup>3</sup> /μL)	6.50 (4.90-9.35)	6.10 (4.70-8.20)	8.45 (6.23-12.83)	<.001	6.20 (4.70-8.30)	8.90 (6.45-13.30)	<.001
CRP (mg/L)	46.90 (14.3-112.4)	30.5 (9.91-86.55)	101 (57.1-177.5)	<.001	33.85 (11.53-94.58)	106 (59.60-181.0)	<.001
Urea (mg/dL)	37 (27-57)	33 (25-46)	58 (38-88)	<.001	35 (26-49.75)	58 (37-89)	<.001
LDH (U/L)	264 (193-391)	256 (190-356)	347 (220.3-543.3)	<.001	259 (190-376.75)	364 (220.50-567.50)	<.001
D Bil. (μmol/L)	16.80 (11.49-30.06)	14.14 (10.61-23.87)	28.73 (16.80-45.97)	<.001	15.02 (10.61-25.64)	30.06 (16.35-51.71)	<.001
NLR	4.50 (2.43-8.57)	3.67 (2.13-7.19)	8.13 (4.28-13.72)	<.001	3.85 (2.18-7.41)	8.57 (5.72-16.79)	<.001

Abbreviations: CRP, c-reactive protein; D Bil., direct bilirubin; DBP, diastolic blood pressure; GCS, Glasgow coma scale; HR, heart rate; ICU, intensive care unit; IU, inpatient unit; LDH, lactate dehydrogenase; NLR, neutrophile-to-lymphocyte ratio; P, asymptotic 2-sided significance between groups with Mann Whitney *U* test; RR, respiratory rate; SBP, systolic blood pressure; spO<sub>2</sub>, blood oxygen saturation; Temp, body temperature; WBC, white blood cell.

medium (1.7%-40.4%), high (40.4%-100%); for CURB-65 low (0 and 1), medium (2), high (3-5); and for ISARIC-4C low (0-3), medium (4-8), high (9-21).

To examine the accuracy of the diagnostic accuracy of the scoring systems in detail, a receiver operating characteristic (ROC) analysis was performed and area under the curve (AUC) was calculated. Sensitivities, specificities, PPV, NPV and Youden's J indexes (YJI) were calculated at a criterion  $>1$ , as has been suggested with guidelines.<sup>15,16</sup> YJI was calculated as well as AUC to assess the predictive accuracy of the scoring systems. Comparisons between the AUC values of the scoring systems were analyzed by the DeLong method.<sup>17</sup> A  $P < .05$  was considered statistically significant.

### 3 | RESULTS

This study was conducted with data from 481 patients after applying the inclusion and exclusion criteria. In the groups compared in the study, there were 361 patients in the surviving group, 120 patients

in the non-surviving group, 396 patients in the IU group, and 85 patients in the ICU group. When the in-hospital mortality outcome was evaluated by gender, 54 women (45%) and 66 men (55%) died (Table 1).

The median age of the population included in the study was 67 (52-79). The median age of the survivor group was 63 (49-76) and the non-survivor group was 75 (66-82) (Table 2). Among the chronic diseases, there was a significant difference between the mortality groups for CHF ( $P = .001$ ), CRF ( $<0.001$ ), CND ( $P = .014$ ), Alzheimer-dementia ( $P < .001$ ) (Table 1). There was a significant difference between the ICU requirement groups for CHF ( $P = .001$ ) and CAD ( $P = .038$ ) (Table 1). A significant difference was observed between both mortality and ICU requirement groups for dyspnea and typical CT findings (Table 1). Among vital signs, HR, RR, and  $spO_2$  were significantly different ( $P < .001$ ) for both grouping methods, while there was no significant difference for SBP, DBP and Temp (Table 2). When the groups were compared according to the laboratory parameters, WBC, Neutrophil to lymphocyte ratio (NLR), Urea, LDH, Direct Bilirubin, CRP measurements used in scoring systems, there

**TABLE 3** Prediction accuracy for Mortality and ICU requirement groups with CURB-65, ISARIC-4C and COVID-GRAM scores in COVID-19 patients

Scores	Risk levels	Survivor (n = 361)	Non-survivor (n = 120)	Sens.	Spec.	PPV	NPV	AUC	YJI
Prediction accuracy for mortality									
CURB 65	Low	267 (93.7%)	18 (6.3%)	85	73.96	52.0	93.7	<b>0.846</b> (0.810-0.877) <sup>a</sup>	<b>0.59</b>
	Med.	83 (65.9%)	43 (34.1%)						
	High	11 (15.7%)	59 (84.3%)						
ISARIC 4C	Low	102 (98.1%)	2 (1.9%)	98.33	28.25	31.3	98.1	<b>0.784</b> (0.744-0.820) <sup>a</sup>	<b>0.27</b>
	Med.	137 (91.9%)	12 (8.1%)						
	High	122 (53.5%)	106 (46.5%)						
COVID GRAM	Low	3 (100%)	0 (0.0%)	100.00	0.83	25.1	100.0	<b>0.701</b> (0.658-0.742) <sup>a</sup>	<b>0.01</b>
	Med.	151 (98.1%)	3 (1.9%)						
	High	207 (63.9%)	117 (36.1%)						
Scores	Risk levels	IU (n = 396)	ICU (n = 85)	Sens.	Spec.	PPV	NPV	AUC	YJI
Prediction accuracy for ICU requirement									
CURB 65	Low	281 (97.9%)	6 (2.1%)	92.94	70.45	40.3	97.9	<b>0.898</b> (0.867-0.923) <sup>a</sup>	<b>0.63</b>
	Med.	103 (81.7%)	23 (18.3%)						
	High	14 (20.0%)	56 (80.0%)						
ISARIC 4C	Low	104 (100%)	0 (0.0%)	100.0	26.26	22.5	100.0	<b>0.797</b> (0.758-0.832) <sup>a</sup>	<b>0.26</b>
	Med.	147 (97.4%)	4 (2.6%)						
	High	147 (64.5%)	81 (35.5%)						
COVID GRAM	Low	3 (100.0%)	0 (0.0%)	100.0	0.76	17.8	100.0	<b>0.684</b> (0.640-0.725) <sup>a</sup>	<b>0.01</b>
	Med.	154 (98.7%)	2 (1.3%)						
	High	241 (74.4%)	83 (25.6%)						

Abbreviations: AUC, area under the curve; ICU, intensive care unit; IU, inpatient unit; Med., medium; NPV, negative predictive value; PPV, positive predictive value; Sens., sensitivity; Spec., specificity; YJI, Youden's J Index.

<sup>a</sup>95% confidence interval.

\*The  $p$  values of the AUC for both mortality and ICU requirement groups of all three scoring systems were calculated as " $<.001$ ". The Youden J Index is a significance measurement and does not generate a  $p$ -value.

were significant differences ( $P < .001$ ) for both grouping methods (Table 2). The data describing the study population are presented in Tables 1 and 2.

When the scoring systems were compared according to the groups with the Man Whitney  $U$  test, a statistically significant difference was observed between both the survivor-nonsurvivor groups and the IU-ICU groups for all three scoring systems ( $P < .001$ ).

ROC analysis was performed to examine the diagnostic accuracy in predicting in-hospital mortality and ICU requirement. AUC values for in-hospital mortality were calculated as 0.846 (0.810-0.877), 0.784 (0.744-0.820) and 0.701 (0.658-0.742) for CURB-65, ISARIC-4C and COVID-GRAM respectively (Table 3). In terms of in-hospital mortality groups, a statistically significant difference was found between AUC values for all three possibilities of the paired comparisons of the COVID severity scores ( $P < .001$ ). AUC values for ICU requirement were calculated as 0.898 (0.867-0.923), 0.797 (0.758-0.832) and 0.684 (0.640-0.725) for CURB-65, ISARIC-4C and COVID-GRAM respectively (Table 3). In terms of ICU requirement groups, a statistically significant difference was found between AUC values for all three possibilities of the paired comparisons of the COVID severity scores ( $P < .001$ ).

Sensitivity, specificity, PPV, NPV and YJI values for in-hospital mortality were calculated as 85%, 73.96%, 52%, 93.7%, 0.59 for CURB-65%, 98.3%, 28.25%, 31.3%, 98.1%, 0.27 for ISARIC-4C, and 100%, 83%, 25.1%, 100%, 0.01 for COVID-GRAM respectively (Figure 1, Table 3). When comparing ICU requirement groups, sensitivity, specificity, PPV, NPV and YJI values were calculated as 92.94%, 70.45%, 40.3%, 97.9%, 0.63 for CURB-65%, 100%, 26.26%, 22.5%, 100%, 0.26%, 0.27 for ISARIC-4C, and 100%, 0.76%, 17.8%, 100%, 0.01 for COVID-GRAM respectively (Figure 2, Table 3).

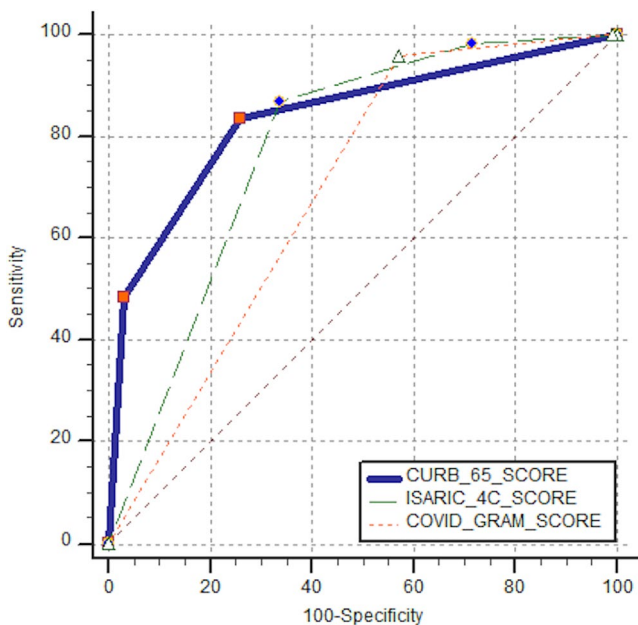


FIGURE 1 ROC curves of CURB-65, ISARIC-4C and COVID-GRAM scores for mortality prediction in COVID-19 patients

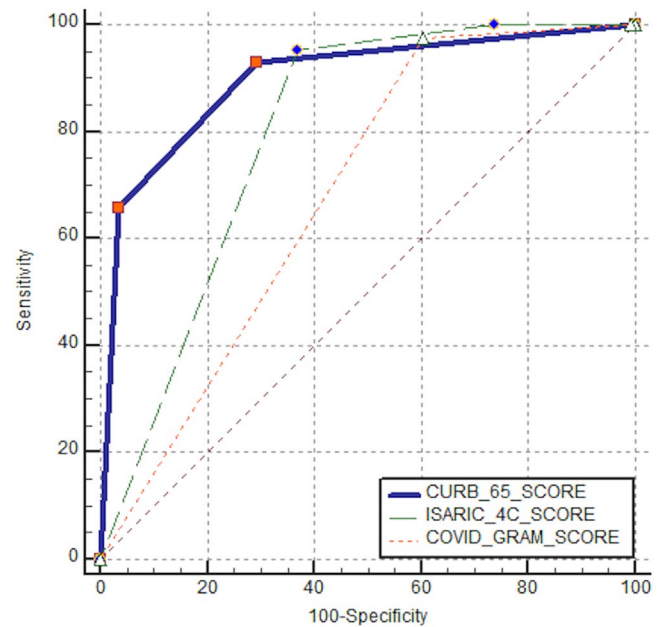


FIGURE 2 ROC curves of CURB-65, ISARIC-4C and COVID-GRAM scores for ICU requirement in COVID-19 patients

#### 4 | DISCUSSION

Our study showed that the CURB-65 score assessed on admission to the ED in COVID-19 patients is more accurate than the COVID-GRAM and ISARIC-4C scores in predicting ICU admission and in-hospital mortality risk. Although the sensitivity of the COVID-GRAM score was higher than the CURB-65 and ISARIC-4C scores, the diagnostic accuracy of the CURB-65 score was better than the other two scores.

Quick and accurate identification of critically ill patients ensures the appropriate and correct use of medical resources. Implementing scoring systems could facilitate more effective evaluation by ED physicians and ICU physicians in identifying critically ill patients.

In the early stages of the pandemic, there was no specific score for COVID-19, for this reason, known scores such as CURB-65 were utilised.<sup>18</sup> The CURB-65 score has been used as a safe predictor of 30-day mortality in patients with pneumonia for many years.<sup>8</sup> It also helps clinicians make the decision to admit or discharge such patients. Demir et al used the Pandemic Medical Early Warning Score (PMEWS), Simple Triage Scoring System (STSS) and CURB-65 score in ED triage for patients with COVID-19 pneumonia; it was reported that these 3 scores were successful in determining mortality, ICU admission and the need for mechanical ventilation.<sup>19</sup> Recently, a study involving COVID-19 patients in a hospital in Wuhan reported that CURB-65 had good performance (AUC 0.81) in determining in-hospital mortality.<sup>20</sup> In our study, the diagnostic performance of CURB-65 (AUC 0.84) for mortality shows similarity with the study by Su et al (AUC 0.85), which was included 116 patients with COVID-19 and whose primary outcome criteria were respiratory or vasopressor support.<sup>21</sup> These results are also consistent with other COVID-19 studies where mortality is the primary outcome.<sup>22,23</sup>

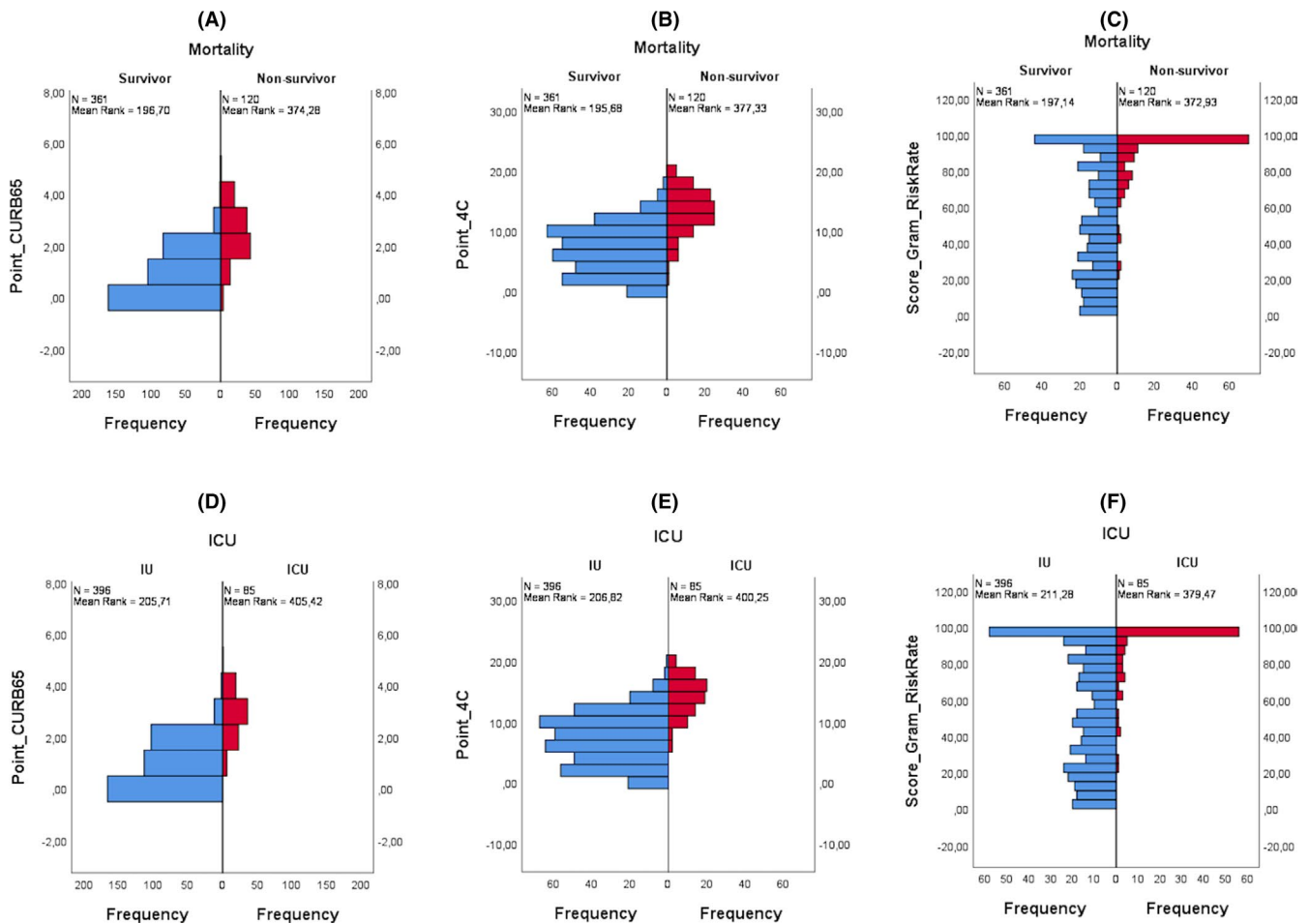


FIGURE 3 Distribution bar graphs of COVID-19 severity scores classified according to mortality and ICU requirement-

CURB-65 score has been successful in predicting low-risk as well as high-risk, as can be seen in Figure 3A,D and in Table 3. For the low-risk patients, the numbers of false-negative cases are more than ISARIC-4C, but the numbers of true positive cases are ~2.7 times more than the ISARIC-4C (Table 3). The majority of patients defined as the medium risk (65.9%-81.7%) with CURB-65 are in the surviving and IU groups. The CURB-65 score giving the lowest numbers of false-negative cases in the high-risk patients is the reason that it is the most successful score in classifying the high-risk patients correctly (Figure 3, Table 3).

Recently, more than 22 specific scoring systems have been developed for COVID-19.<sup>24</sup> Liang et al developed and validated a complex scoring system called COVID-GRAM, a clinical risk score for predicting the prognosis among hospitalised patients infected with COVID-19. The performance of this risk score was satisfactory in the development and validation cohorts with certainty based on an AUC of 0.88. However, the modest sample size for establishing the risk score and the relatively small sample for verification, constitute the limitations of this scoring system.<sup>11</sup> In a retrospective study involving geriatric COVID-19 patients, the National Early Warning Score (NEWS), COVID-GRAM, ISARIC-4C and quick COVID-19 Severity Index (qCSI) scores were compared with each other in terms of

in-hospital mortality, and superiority of the scores to each other was not determined.<sup>25</sup> In another study, NLR and COVID-GRAM were compared with each other in mortality prediction for COVID-19 patients, and no difference was found between these two predictors (AUC of 0.65 and 0.66 respectively).<sup>26</sup>

As can be seen in Tables 3 and in Figure 3C,F, mortality and ICU requirement are very low in patients whose COVID-GRAM score is defined as low- and medium risk (Figure 3, Table 3). In the patients defined as high risk, those without mortality or ICU outcome were predominated (63.9%-74.4%). The COVID-GRAM score classified only three patients at low risk in our study population. These three patients did not require ICU, and their outcome was not death. Therefore, the sensitivity of the COVID-GRAM score was found to be 100% (Figure 3, Table 3). For our study population, we can say that the COVID-GRAM score is insufficient for identifying low-risk patients. One of the reasons for this could be the high median age (67) of our study population. This is because in the COVID-GRAM score calculation—as can be seen from the COVID-GRAM risk calculation formula—age is contributed to the score by multiplying with a coefficient. For example, a 70-year-old person who does not have thoracic CT findings, is asymptomatic, does not have a comorbidity, and has normal laboratory values falls into the medium-risk class with a risk coefficient of 3.3%.

Al Hassan et al reported that the CALL score and COVID-GRAM score were poor in the prediction of ICU admission or death, and these scores might not be useful in different populations.<sup>27</sup> The median age of the patients in the development and validation cohorts of COVID-GRAM score was 48.2, while the median age of the patients in the study of Al Hassan et al was 73. This may explain the poor prediction accuracy of the COVID-GRAM score in the study population of Al Hassan et al, and in ours.<sup>11,27</sup>

As can be seen in Table 3 and in Figure 3B,E, ISARIC-4C score points below 8 have very low mortality or ICU outcome. The fact that only 4 (1.6%) out of 251 patients defined as low and medium risk by the ISARIC-4C score who were hospitalised ICU from the ED, shows the success of ISARIC-4C in determining low risk (Figure 3, Table 3). ISARIC-4C did not show the same success in the patient population that it describes as high risk. The patients defined as high risk by ISARIC-4C, who were 35.5% of the total, are in the IU group in terms of ICU requirement. When the same rate is evaluated in terms of in-hospital mortality, 53.5% of the patients are in the survivor group.

Using YJI in addition to sensitivity, specificity and AUC provides more reliable results when comparing the accuracy of the scoring systems.<sup>28</sup> In our study, for the mortality prediction of COVID-19 patients, the YJI of CURB-65, ISARIC-4C and COVID-GRAM scores were found to be 0.58, 0.27 and 0.01 respectively; for determining the ICU requirement, the YJI were found to be 0.64, 0.26 and 0.01 respectively (Table 3).

CURB-65 is a most well-known, simple to use and widely validated scoring system that has proven its prognostic accuracy. The CURB-65 score had the highest predictive accuracy also in our study population.

## 5 | LIMITATIONS

The sample size of this single centre study was relatively small. In retrospective studies, the study population is formed by convenience sampling methods, so it does not represent the general population and may lead to selection bias. Therefore, more studies with a larger sample size are needed to confirm these results.

## 6 | CONCLUSIONS

Among the scores evaluated in our study, the CURB-65 score had better performance than ISARIC-4C and COVID-GRAM scores in predicting in-hospital mortality and ICU requirement in COVID-19 patients. Instead of COVID-GRAM score (which is more complex to calculate) we suggest that the CURB-65 score (which has been validated and can be easily calculated) can also be used for COVID-19 patients. Based on the success of ISARIC-4C in the low-risk group, we believe that using CURB-65 and ISARIC-4C scores together will positively affect the decision-making process of clinicians.

## ACKNOWLEDGEMENTS

The authors thank the employees of Kartal Dr Lütfi Kırdar City Hospital for their dedication during the pandemic process.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## ETHICAL APPROVAL

This study was approved by the ethics committee of Kartal Dr Lütfi Kırdar City Hospital (Ethics Committee Ruling number: 514/196/24).

## DATA AVAILABILITY STATEMENT

Data available on request due to privacy/ethical restrictions.

## ORCID

Fatih Doğanay  <https://orcid.org/0000-0003-4720-787X>

Rohat Ak  <https://orcid.org/0000-0002-8324-3264>

## REFERENCES

1. Emanuel EJ, Persad G, Upshur R, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N Engl J Med*. 2020;382(21):2049-2055. 10.1056/NEJMs2005114
2. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. 2020.
3. 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):507-513.
4. Bhatraju PK, Ghassemieh BJ, Nichols M, et al. Covid-19 in critically ill patients in the Seattle region—case series. *N Engl J Med*. 2020;382(21):2012-2022.
5. Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA*. 2020;323:1574-1581.
6. Ruan Q, Yang K, Wang W, et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med*. 2020;46:846-848.
7. Jeong B-H, Koh W-J, Yoo H, et al. Performances of prognostic scoring systems in patients with healthcare-associated pneumonia. *Clin Infect Dis*. 2013;56(5):625-632.
8. Lim W, Van der Eerden M, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;58(5):377-382.
9. García Clemente MM, Herrero Huertas J, Fernández Fernández A, et al. Assessment of risk scores in Covid-19. *Int J Clin Pract*. 2020;e13705. <https://doi.org/10.1111/ijcp.13705>
10. Knight SR, Ho A, Pius R, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ*. 2020;370:m3339.
11. Liang W, Liang H, Ou L, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Int Med*. 2020;180(8):1081-1089.
12. BAKANLIĞI, TC SAĞLIK. COVID-19 (SARS-CoV2 ENFEKSİYONU) REHBERİ. Erişim (Erişim Tarihi: July 16, 2020). 2020. [https://covid19bilgi.saglik.gov.tr/depo/rehberler/COVID-19\\_Rehberi.pdf](https://covid19bilgi.saglik.gov.tr/depo/rehberler/COVID-19_Rehberi.pdf).
13. Pan F, Ye T, Sun P, et al. Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. 2020.



14. Seyhan AU, Doğanay F, Yılmaz E, et al. The comparison of chest CT and RT-PCR during the diagnosis of COVID-19. *J Clin Med Kazakhstan*. 2021;18(1):53-56.
15. Lim WS, Baudouin S, George R, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax*. 2009;64(Suppl 3):iii1-iii55.
16. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007;44(Supplement\_2):S27-S72.
17. DeLong ER, DeLong DM, Clarke-Pearson DLJB. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.
18. Nguyen Y, Corre F, Honsel V, et al. Applicability of the CURB-65 pneumonia severity score for outpatient treatment of COVID-19. *J Infect*. 2020;81(3):e96-e98.
19. Demir MC, İlhan BJSPMJ. Performance of the Pandemic Medical Early Warning Score (PMEWS), Simple Triage Scoring System (STSS) and Confusion, Uremia, Respiratory rate, Blood pressure and age  $\geq$  65 (CURB-65) score among patients with COVID-19 pneumonia in an emergency department triage setting: a retrospective study. *Sao Paulo Med J*. 2021;139(2):170-177.
20. Guo J, Zhou B, Zhu M, et al. CURB-65 may serve as a useful prognostic marker in COVID-19 patients within Wuhan, China: a retrospective cohort study. *Epidemiol Infect*. 2020;148:e241.
21. Su Y, Tu G-W, Ju M-J, et al. Comparison of CRB-65 and quick sepsis-related organ failure assessment for predicting the need for intensive respiratory or vasopressor support in patients with COVID-19. *J Infect*. 2020;81(4):647-679.
22. Artero A, Madrazo M, Fernández-Garcés M, et al. Severity scores in COVID-19 pneumonia: a multicenter, retrospective, cohort study. *J Gen Intern Med*. 2021;36(5):1338-1345.
23. Fan G, Tu C, Zhou F, et al. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. *Eur Respir J*. 2020;56(3):2002113.
24. Martinez-Lacalzada M, Viteri-Noël LA, Manzano L, et al. Predicting critical illness on initial diagnosis of COVID-19: development and validation of the PRIORITY model for outpatient applicability. *Clin Microbiol Infect*. 2020.
25. Covino M, De Matteis G, Burzo ML, et al. Predicting in-hospital mortality in COVID-19 older patients with specifically developed scores. *J Am Geriatr Soc*. 2021;69:37-43.
26. Yildiz H, Yombi JC, Castanares-Zapatero DJID. Validation of a risk score to predict patients at risk of critical illness with COVID-19. *Infect Dis*. 2021;53(1):78-80.
27. Al Hassan H, Cocks E, Jesani L, et al. Clinical risk prediction scores in coronavirus disease 2019: beware of low validity and clinical utility. *Crit Care Explor*. 2020;2(10):e0253.
28. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J*. 2005;47(4):458-472.

### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

**How to cite this article:** Doğanay F, Ak R. Performance of the CURB-65, ISARIC-4C and COVID-GRAM scores in terms of severity for COVID-19 patients. *Int J Clin Pract*. 2021;75:e14759. <https://doi.org/10.1111/ijcp.14759>