Diagnostic utility of the Covichem score in predicting COVID-19 disease

Cigdem Ozpolat, MD *, Erhan Altunbas, MD

Department of Emergency Medicine, Marmara University Pendik Research and Training Hospital, 34890 Istanbul, Turkey

ARTICLE INFO

ABSTRACT

Background: Identifying which patients with COVİD-19 have a high risk of severe illness is essential to optimizing management and resource utilization strategies.
Objectives: The aim of this study was to externally validate the diagnostic utility of the Covichem score for predicting COVID-19 disease severity, and secondarily to evaluate its utility in predicting intensive care unit (ICU) admission, and in-hospital mortality.
Methods: All consecutive COVID-19 patients who presented to the emergency department (ED) were included, and patients' demographic data, comorbidities, vital signs, oxygen requirement, and laboratory results were recorded. We calculated patients' Covichem scores and estimates (using a threshold of 0.5) and evaluated the utility of the Covichem score for predicting disease severity, ICU admission, and mortality.
Results: The median Covichem score was significantly higher for patients with severe illness (Covichem score: 0.170, IQR: 0.298, $n = 300$ vs. Covichem score: 0.026, IQR: 0.065, n: 191; $p < 0.001$). Based on their Covichem scores, 12.4% (61/491) of the patients were predicted to experience severe illness (threshold: 0.5), the accuracy of the Covichem score was poor, as the area under curve (AUC) was 48.5% (18.1% sensitivity and 93.8% specificity). When we calculated a new ideal threshold, the AUC reached 82%, but the sensitivity was 79.9% and the specificity was 71.2%.
Conclusion: In this external validation of the Covichem score, we found that it performed worse than in the original derivation and validation study, even with the assistance of a new cutoff.

© 2022 Elsevier Inc. All rights reserved.

1. Introduction

The Coronavirus disease 2019 (COVID-19) epidemic is caused by the SARS-CoV-2 virus, which emerged in Wuhan in December 2019 and quickly spread around the world, developing into a worldwide pandemic. According to the World Health Organization (WHO), as of February 2022, approximately 425 million cases and 5,880,000 deaths have been reported worldwide [1]. The high number of cases and deaths has led to various investigations of the use of vital signs, comorbidities, and laboratory and radiological findings to determine the severity of the disease and reduce morbidity and mortality [2,3].

The COVID-19 pandemic places a serious burden on healthcare facilities. Especially during acute surges, when hospital resources are scarce, it is essential to triage patients accurately, direct them to appropriate centers, and initiate specific treatments as early as possible. Therefore, an algorithm that can accurately and reliably predict the risk of severe illness in patients with COVID-19 would help healthcare professionals.

Early risk assessment allows physicians to triage patients and prioritize resources in a highly congested system. Bats et al. developed a score, called the Covichem score, for predicting a patient's risk of severe illness with COVID-19. This score can be calculated at hospital admission [4]. It uses clinical parameters and commonly available laboratory results and does not require imaging results or advanced testing.

The primary aim of the present study was to externally validate the diagnostic utility of the Covichem score for predicting COVID-19 disease severity. The secondary aims were to evaluate the score's utility for predicting admission to the intensive care unit (ICU) and in-hospital mortality.

2. Material and methods

2.1. Design and setting

This prospective, single-center, observational diagnostic accuracy study was conducted in the pandemic wing of a university hospital. Our hospital is one of two referral COVID-19 hospitals located on the Anatolian side of Istanbul, which has six million inhabitants. All patients diagnosed with COVID-19 at another hospital with an indication for hospitalization are referred to our hospital. Our facility also evaluates

* Corresponding author at: Marmara University PendikResearchand Training Hospital, Department of Emergency Medicine, Üst Kaynarca / Pendik, 34890 İstanbul, Turkey.
E-mail address: cumed17@gmail.com (C. Ozpolat).

patients' COVID-19 status on an outpatient basis and admits them if necessary. The local committees granted ethics and institutional board approval for the study. Informed written consent was obtained from each participant or next of kin.

2.2. Study population and sample

All consecutive patients who presented at the pandemic hospital's emergency department (ED) constituted the study population. All consecutive patients who presented to the hospital over a period of three months, from October 1, 2021, to January 1, 2022, and who met the inclusion criteria were defined as the study sample.

2.3. Inclusion and exclusion criteria

All consecutive patients who were admitted to our ED were included in the study if they (1) were aged 18 years or older and (2) had a positive SARS-CoV-2 polymerase chain reaction (PCR, with a nasopharyngeal swab) test result or computed tomography (CT) images associated with a high clinical probability of COVID-19 with the following symptoms: dry cough, fever, chills, fatigue, dyspnea, chest pain, myalgia, diarrhea, anosmia, or ageusia. Asymptomatic patients who presented for a PCR test for work, travel, or social requirements, symptomatic patients without a positive CT or PCR, and patients who were directly admitted to the ICU or ward without ED evaluation were excluded. We also excluded patients who withdrew their consent or for whom data included in the analysis were missing.

2.4. Management

Emergency physicians (EP) evaluated all patients presenting to the ED. The institutional protocol for treating any patient suspected of having or diagnosed with COVID-19 was as follows: evaluate the vital signs, perform a thorough physical examination, order a set of laboratory tests (CBC, electrolytes, liver and renal function tests, CRP, ferritin, D-Dimer, CK, cardiac enzymes and venous blood gases) and a low dose thorax CT, if no contraindication was present.

Some included patients were referred from other hospitals and transferred to our facility via ambulance for admission. Those patients were re-evaluated in our ED using the approach described above and then admitted to the ward or ICU. No patients referred by other hospitals were discharged from the ED.

The EP decided to admit or discharge patients after evaluating the patient using the institutional protocol. Patients admitted to ward beds were treated by the ward staff (a specialist and resident were assigned to that ward monthly from the whole roster of the hospital). An infectious disease specialist evaluated all ward patients daily and determined when they should be discharged. An intensive care specialist evaluated all patients in case of clinical deterioration and decided whether they should be admitted to ICU.

Patients who were discharged from the ED after the initial evaluation were ordered to check their peripheral oxygen saturation (%, SpO2) using a fingertip pulse oximeter and to take their temperature twice daily and to return to the ED if any SpO2 reading was below 90% or their temperature was >38C for at least three measurements over a period of two hours.

2.5. Study process and data collection

2.5.1. Data collection

EPs prospectively filled out patient data collection forms for all patients who were admitted to ED and met the inclusion criteria. Progress notes, treatment orders, and nursing charts for admitted patients were recorded in the electronic hospital information system (HIS) by the ward staff.

Demographic data (age, gender, weight, and height), the presence of comorbidities, measured vital signs at admission to ED (blood pressure [mmHg], pulse rate [bpm], SaO2 [%], and temperature [C]), the results of the physical examination, and the level of oxygen requirement were noted on the form. Also, we asked the EP to record their assessment if the patient would go on to develop severe disease (yes or no) at time of admission before any lab values were available. For the present study, laboratory test results and the official radiology results from the initial thorax CT scan (Philips Ingenuity 128 CT Scanner, 5 mm slice thickness reconstruction) were retrieved from the HIS. Each patient's CORADS (COVID-19 Reporting and Data System) class was charted on the data collection form; for patients whose CORADS class was not recorded, CT images were evaluated to determine the CORADS class [5].

2.5.2. Researchers and blinding

All data charted on forms and in the HIS were prospectively evaluated, analyzed, and collated on a separate electronic chart by the head researchers, both of whom are EM specialists with at least ten years of work experience. The researchers were not responsible for treatment, intervention, admission, or discharge decisions. The EPs and radiologists who reported the CT exams were blinded to the participants' Covichem scores and to the outcomes of the study. The Covichem scores and estimates were not available to the researchers until all outcome data had been collected and recorded.

2.5.3. Measurements and calculations

Drawn blood was collected using BD Vacutainer® SST™ II Advance tubes. Biochemical parameters were measured using a Roche Cobas 8000 analyzer and the following analytical methods: Indirect potentiometry was used to analyze plasma sodium; an enzymatic method was used to evaluate CK and LDH; an electro-chemiluminescence immunoassay was used to measure ferritin; and a colorimetric assay was used to measure albumin.

2.5.4. Index test

The Covichem score is an estimate ranging from 0 to 1. It is calculated using a regression function that includes two comorbidities (obesity (defined as a BMI ≥30) and cardiovascular disease (CVD; coronary artery diseases such as angina and myocardial infarction, heart failure, cardiomyopathy, abnormal heart rhythms, valvular heart disease, aortic aneurysms, heart transplant, peripheral artery disease, thromboembolic disease, venous thrombosis and stroke)) and five laboratory test results (Na, albumin, ferritin, LDH, and CK; Supplement 1). It was derived and externally validated by Bats et al., who used the 0.5 (50%) value of this estimate as a threshold to assign patients to positive and negative Covichem groups [4]. We collected all data and then calculated the estimate and Covichem groups using the regression function described in Bats et al.

2.5.5. Reference standard

We reviewed the admission forms, HIS records, and charts of all patients to determine disease severity. The reference standard test was the disease severity defined by Bats et al. as prognostically severe and non-severe [4]. Patients who met one of the following criteria at admission or during hospitalization were categorized as severe: (1) SpO_2 <90% on room air or (2) at least 4 L/min oxygen needed to obtain a SpO_2 above 94%. Patients who presented with acute respiratory distress syndrome or who were directly admitted to the ICU on admission were also categorized as severe. All other patients were categorized as non-severe.

We also evaluated the utility of the Covichem score for predicting ICU admission and mortality. We assessed mortality outcomes using a follow-up phone call made on the 15th day after admission for discharged patients; in-hospital mortality based on HIS records was used for all patients admitted to a ward or ICU.

2.5.6. Sample size estimation

We estimated our sample size using the data (89% sensitivity and 95% specificity) obtained in an external validation cohort of 100 patients using the formula suggested by Hajian-Tilaki [4,6]. We calculated that we needed at least 160 and 73 patients at the defined sensitivity and specificities, if our severe disease prevalence would be 48% as in the target study, with a marginal error of 7%, Type 1 error of 5%, power of 80% and confidence interval of 95%. We downgraded our estimate based on a severe disease prevalence of 30% and sensitivity of 80%, which resulted in an estimated sample size of 420 patients. Finally, we added a safety margin of 20% and thus calculated a final sample size of 500 patients.

2.6. Statistical analysis

For continuous variables, either the mean, standard deviation, and 95% confidence interval (CI) or the median and interquartile range (IQR) were calculated based on the data distribution patterns. Categorical variables were expressed in counts and frequencies. We calculated predictive utility measures, sensitivity, specificity, and positive and negative likelihood ratios using contingency tables. We excluded patients whose Covichem scores could not be calculated due to missing data. There were no patients with missing data regarding prognostic severity, mortality, or ICU admission.

3. Results

The final study population was 507 patients (Fig. 1). The comparison of the study groups according to disease severity for vital signs, comorbidities, demographics (Table 1), and laboratory and radiology test results (Table 2) were presented in tables.

Three hundred and thirteen (61.7%) patients were classified as having severe COVID-19. While the median age for all patients was 63 years (IQR: 28), the median age of those in the severe group was significantly higher (67 years, IQR: 25 vs. 55.5 years, IQR:28, $p < 0.001$). There was also a higher percentage of male patients in the severe group than in the non-severe group (171 (55%) vs. 75 (%39); p < 0.001). The prevalence of diabetes mellitus (DM) (37.7% vs. 25.3), hypertension (HT) (52.7% vs. 34.5%) and CVD (33.5% vs. 14.4%) were significantly higher in the severe group as well.

The prevalence of patients without vaccination or with an unknown vaccination status was significantly higher in the severe group (n/N: 161/313, 51.4% vs. 80/194, 41.2%; $p = 0.026$). The relative risk of severe prognosis was 1.17 (95% CI: 1.02 to 1.34; p = 0.026) for patients without vaccination or with an unknown vaccination status; the number needed to harm (NNH) was 10.4 (95% CI: 81.2 to 5.5).

All laboratory parameters included in the Covichem score differed significantly between the severe and non-severe prognostic groups (sodium, albumin, ferritin, LDH, and CK; see Table 2). D-dimer, CRP, and AST levels also differed clinically and statistically for the two groups (Table 2).

Seventy-six percent of the study population (n/N: 385/507) presented with high-risk radiological features for COVID-19 on a thorax CT and had a CORADS of 4 or 5. High-risk features were present in 58.2% of the non-severe and 86.9% of the severe prognosis group; this difference was statistically significant ($p < 0.001$). The relative risk of severe prognosis was 2.1 (95% CI: 1.6 to 2.7; p < 0.001) for patients with a CORADS classification of 4 or 5 based on their CT, with an NNH of 2.7 (95% CI: 3.6 to 2.2).

The median Covichem score (0–1) for all patients was 0.093 (IQR: 0.238, n/N: 491/507, 96.8%). The median Covichem was significantly
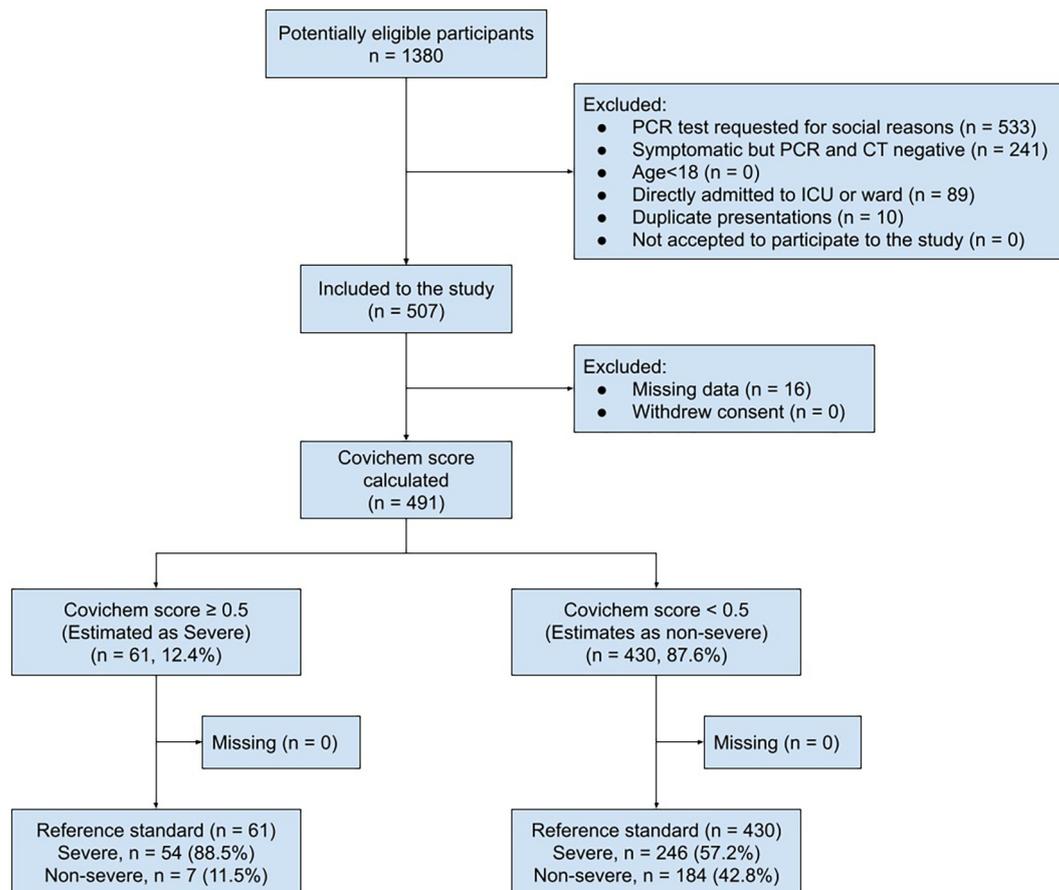

Fig. 1. Patient flowchart.

Table 1
Comparison of the study groups according to disease severity for vital signs, comorbidities, and demographics.

Variable	Total	Outcome (Severity)		P
		Non-severe	Severe	
	($n = 507$)	($n = 194$)	($n = 313$)	
Demographics				
Age, years, median (IQR)	63 (28)	55.5 (28)	67 (25)	**<0.001***
Male, n (%)	246 (48.5)	75 (39.0)	171 (55.0)	**<0.001****
Vital Signs and Measurement				
RR, /min	30 (8)	24 (12)	30 (7)	**<0.001***
SpO$_2$, %	90 (7.5)	95 (4)	88 (9)	**<0.001***
Supp. O$_2$, L/min	3 (3)	0 (2)	4 (5)	**<0.001***
BMI, n = 480, 184, 296	27.6 (5.8)	27.5 (6.0)	27.7 (5.7)	**<0.001***
History				
DM, n (%)	167 (32.9)	49 (25.3)	118 (37.7)	**0.004****
HL, n (%)	29 (5.7)	15 (7.7)	14 (4.5)	0.120**
HT, n (%)	232 (45.8)	67 (34.5)	165 (52.7)	**<0.001****
Obesity, n (%)	166 (32.7)	62 (32.0)	104 (33.2)	0.770**
CVD, n (%)	133 (26.2)	28 (14.4)	105 (33.5)	**<0.001****
Malignancy, n (%)	38 (7.5)	10 (5.2)	28 (8.9)	0.120**
Liver Disease, n (%)	3 (0.6)	1 (0.5)	2 (0.6)	NA
Renal Disease, n (%)	40 (7.9)	12 (6.2)	28 (8.9)	0.260**
Lung Disease, n (%)	93 (18.3)	33 (17.0)	60 (19.2)	0.540**
Covid Hx and Vaccination				
Covid Hx, n = 494	8 (1.6)	3 (1.6)	5 (1.7)	NA
Vaccination Status, n (%, column percentages)				
Unknown	70 (13.8)	19 (9.8)	51 (16.3)	**0.026****
No vaccination	171 (33.7)	61 (31.4)	110 (35.2)	
One dose	23 (4.5)	12 (6.2)	11 (3.5)	
Two doses	148 (29.2)	69 (35.6)	79 (25.2)	
3 doses	94 (18.5)	32 (16.5)	62 (19.8)	
4 doses	1 (0.1)	1 (0.5)	0 (0.0)	

n: number, IQR: interquartile range, RR: respiratory rate, pSO$_2$: peripheral oxygen saturation, BMI: bady mass index, DM: diabetes mellitus, HL: hyperlipidemia, HT: hypertension, CVD: cardiovascular disease.

* Mann-Whitney *U* test.

** Chi-square test. Statistically significant *p* values are written in bold. NA: Cannot be calculated.

Table 2
Comparison of the study groups according to disease severity for lab and radiology results.

Variable	Total	Outcome (Severity)		P
		Non-severe	Severe	
	(n = 507)	(n = 194)	(n = 313)	
Laboratory Values, median (IQR)				
Na, mEq/L	137 (5)	138 (5)	136 (5)	**<0.001***
K, mEq/L	4.3 (0.7)	4.3 (0.6)	4.3 (0.8)	0.294*
Total Protein, g/L, n = 506, 193, 313	67 (8)	69 (7)	66 (8)	**<0.001***
Albumin, g/L, n = 505, 193, 312	38 (6)	40 (5)	36 (6)	**<0.001***
D-Dimer, mg/L	0.51 (0.88)	0.39 (0.55)	0.65 (1.12)	**<0.001***
CRP, mg/L	70.8 (106.9)	39.6 (70.4)	94.6 (113.8)	**<0.001***
AST, U/L	32.9 (21.3)	27.5 (18.0)	36.4 (26.6)	**<0.001***
ALT, U/L	21.7 (18.9)	20.8 (15.5)	22.5 (20.0)	0.184
ALP, U/L; n = 505, 192, 313	69 (33.0)	71.5 (27.5)	67 (37.0)	0.436
Ferritin, mcg/L	380 (576)	203 (305)	519 (675)	**<0.001***
LDH, U/L	347 (199)	270 (130)	396 (220)	**<0.001***
CK, U/L n = 491, 191, 300	103 (120)	85 (73)	113 (175)	**<0.001***
CO-RADS Class, n (%, column percentages)				
1	50 (9.9)	44 (22.7)	6 (1.9)	**<0.001****
2	26 (5.0)	14 (7.2)	12 (3.8)	
3	46 (9.1)	23 (11.9)	23 (7.4)	
4	50 (9.9)	21 (10.8)	29 (9.3)	
5	335 (66.1)	92 (47.4)	243 (77.6)	

* Mann-Whitney *U* test.

** Chi-square test. Statistically significant p values are written in bold.

higher for patients in the severe group (Covichem score: 0.170, IQR: 0.298, $n = 300$ vs. Covichem score: 0.026, IQR: 0.065, n: 191; $p < 0.001$ Table 3). According to the Covichem score, with a threshold of 0.5, 12.4% (61/491) of the patients were estimated to develop severe disease. In fact, 61.7% (313/507) of the patients developed severe disease. The true positive rate of the Covichem score for predicting severe disease was only 18% (54/313), while the false positive rate was 3.7% (7/194).

Bats et al. used 0.50 as the threshold Covichem score. At this threshold, in our study, the accuracy of the Covichem score was poor, with an AUC of 48.5%. Bats et al. excluded patients who were discharged from the ED. When we calculated the accuracy of the Covichem score as described by Bats et al. (outpatients excluded, cutoff 0.5), it was quite low with an AUC of 38.8 (34.1,−43,7), sensitivity of 18.1 (13.9–22.9) and specificity of 93.8 (87.7–97.5). We therefore recalculated the accuracy of the Covichem estimate for this subset and found a statistically significant AUC of 74.3% (SE: 2.8%, $p < 0.001$) using a threshold value of 0.087 at the highest Youden J index (Table 4). Other prognostic utility metrics are presented in Table 4.

We calculated the accuracy of the Covichem score using an area under the curve (AUC) of 82% (SE: 2%; p < 0.001) in our study population. The Youden J Index revealed that the Covichem score had the highest sensitivity (79.7%) and specificity (71.2%, Table 4) at a threshold value of 0.059.

Physicians predicted that 33.9% (167/493) of the patients would develop severe disease. The true positive rate for physician predictions was only 49% (148/313), while the false positive rate was 9.9% (19/194). The accuracy of physician assessment in distinguishing severe patients was low with AUCs of 64.9 (60.5–69.1) (Table 4).The median length of stay (LOS) was almost three times longer for the severe group (ten days versus three days; Table 3).

Admission to ICU and mortality were the defining criteria of the severity group. One hundred and two patients (20.3%) were admitted to the ICU. The total mortality rate was 17.3% (n/N: 86/497 patients). The Covichem score predicted ICU admission and mortality with AUCs of 72.3% (95% CI: 68.0% to 76.2%) and 75.0% (95%CI: 70.8% to 78.8%), respectively (Fig. 2). The threshold estimate with the highest Youden J Index was 0.0916 for ICU admission and 0.0963 for mortality.

4. Discussion

In this study, we utilized a prospective cohort to validate the use of the Covichem score to predict disease severity, ICU hospitalization, and mortality after 15 days in patients diagnosed with COVID-19. We found that the Covichem score can't predict COVID-19 severity and has a very low accuracy (AUC: 38.8; sensitivity: 18.1; specificity: 93.8) at its original criteria and threshold values. We added patients discharged from the ED to the sample, recalculated the threshold of the estimate for severe disease and found that this revised Covichem score has an AUC of 82%, and at the threshold of 0.059, the sensitivity and specificity were 79.7% and 71.2%, respectively.

Several studies have used clinical, laboratory, or radiological parameters to establish a prognostic model for COVID-19 patients to predict outcomes such as mortality, progression to severe disease, and ICU admission. However, most of these studies lack validation in multiple cohorts [7]. Few studies report scores associated with COVID-19 severity that use blood markers at hospital admission [8,9].

Bats et al. developed the Covichem severity score using seven parameters: obesity, cardiovascular condition, plasma sodium, albumin, ferritin, LDH, and CK. They found these parameters to be independent predictors of disease severity. They report that the Covichem score predicts disease severity with an AUC of 0.91, sensitivity of 0.85, and specificity of 0.88 at a threshold of 0.5 [4].

We found that Covichem score did not predict disease severity with the original criteria defined by Bats et al. at the Turkish population (AUC: 38.8 (34.1,−43,7); sensitivity: 18.1 (13.9–22.9); specificity:

Table 3

Comparison of the study groups according to disease severity for scores and outcomes.

Variable	Total	Outcome (Severity)		P
		Non-severe	Severe	
	(n = 507)	(n = 194)	(n = 313)	
Score / Index Test				
Covichem Score, n = 491, median (IQR)	0.093 (0.238), n = 491	0.026 (0.065), *n* = 191	0.170 (0.298), n = 300	**<0.001**[*]
Covichem Score > 0.5 = Severe, *n* = 491, n (%)	61 (12.4)	7 (3.7)	54 (18)	**<0.001**[**]
Physician Assessment = Severe, *n* = 493, n (%)	167 (33.9)	19 (9.9)	148 (49)	**<0.001**[**]
Outcome				
LOS, (day) *n* = 496, median (IQR)	6 (8)	3 (5) *n* = 192	10 (9.3) *n* = 304	**<0.001**[*]

n: number, IQR: interquartile range.

[*] Mann Whitney U test.

[**] Chi-squared test. Bold indicates statistical significance.

93.8 (87.7–97.5). Inclusion of the outpatients to overcome selection bias did not increase the predictive utility of the score as well (AUC: 48.5; sensitivity: 18.0; specificity: 96.3 respectively). We calculated the optimal threshold value with the ROC analysis and found that Covichem score achieves an AUC of 0.82 at a significantly lower estimate threshold value of 0.059 compared to the original value of 0.5 in the study at Bats et al. There may be several reasons for this disparity: First, serum albumin level, which has the highest coefficient in the Covichem regression function, is inversely proportional to the overall score, as is sodium. The measurement method influences the level of serum albumin, which can vary depending on whether the measurement method uses bromocresol purple or green dye. In cases of hypoalbuminemia, green dye overestimates the albumin level compared to purple dye [10]. Since we measured albumin using bromocresol green dye, our albumin results were higher than those of Bats et al. This may have resulted in lower Covichem scores and threshold levels. Second, Turkish physicians usually initiate O2 therapy at 4 L/min in hypoxic patients. Since this was a criterion for disease severity, some of our patients may have been categorized in the severe group, despite lower Covichem scores. However, the overall predictive value of the Covichem score was still high enough to predict severe disease with an AUC of 0.82. We therefore conclude that this discrepancy with Bats et al. reflects a bias regarding the threshold rather than the utility of the Covichem score.

Unlike Bats et al., we included both patients who were admitted to ward or ICU and patients who were discharged. When we excluded discharged patients, the accuracy of the Covichem score dropped to an AUC of 74.3 (95% CI: 69.8–78.4) with a threshold of 0.087. Although it has an effect, it is not possible to explain the five-fold change in the

optimal threshold with the albumin measurement method alone. The lower serum ferritin level of our cohort decreased the overall Covichem score of our sample, but not the threshold. In addition to this, the difference in the prevalence of patients with severe disease in our cohort may have resulted in a lower threshold value.

We found that all components of the Covichem score differed significantly between severe and non-severe patients and that these components are valuable predictors of disease severity. Gong et al. have shown that higher serum LDH and lower albumin on admission indicate the risk of progression to severe COVID-19 in patients admitted with non-severe COVID-19 [9]. Halalau et al. have shown that obesity and a history of cardiovascular disease predict hospital admission [11].

We also investigated the utility of the Covichem score for predicting ICU admission and in-hospital mortality and found the AUCs of 72.3 (95% CI: 68.0–76.2) and 75 (95% CI: 70.8–78.8), respectively. Several previous studies have examined various risk factors for mortality in COVID-19 patients. High serum ferritin, LDH, and CK; low albumin; and a history of cardiovascular disease have been reported as risk factors for mortality [3,11,12]. It seems that the combination of those risk factors in a predictive score also performs well.

4.1. Limitations

This study has several limitations. First, the study was conducted in a single referral hospital, where most of the patients were relatively severe and there were few outpatient admissions. Second, close monitoring of the exact amount of O2 flow/min needed was not possible in the ED. Therefore, physicians may have preferred to use higher O2 flow

Table 4

Predictive utility of covichem score and physician assessment for severe COVID-19 disease and mortality.

Variable	Threshold	AUC, % (95% CI)	Sensitivie, % (95% CI)	Specificity, % (95% CI)	+LR, (95% CI)	−LR, (95% CI)
Disease Severity						
Covichem Estimate, Severe/Non-severe (outpatients excluded), n = 412	Contingency table, 0.5	38.8 (34.1–43.7)	18.1 (13.9–22.9)	93.8 (87.7–97.5)	2.92 (1.37–6.22)	0.87 (0.81–0.93)
Covichem Estimate, Severe/Non-severe (outpatients included)	Contingency table, 0.5	48.5 (44.0, 53.0)	18.0 (13.8, 22.8)	96.3 (92.6, 98.5)	4.91 (2.28, 10.57)	0.85 (0.80, 0.90)
Covichem Score (outpatients included), n = 491	ROC curve, 0.059	82.0 (0.78, 0.85)	79.7 (74,7–84,1)	71.2 (64.2–77.5)	2.77 (2.20–3.48)	0.29 (0.22–0.36)
Covichem Score (outpatients excluded), n = 412	ROC curve, 0.087	74.3 (69.8–78.4)	70.9 (65.4–76.0)	70.8 (61.5–79)	2.43 (1.81–3.26)	0.41 (0.33–0.51)
Physician Assessment, Severe/Non-severe, n = 493	Contingency table, Yes/No	64.9 (60.5–69.1)	49.0 (43.2–54.8)	90.1 (84.9–93.9)	4.93 (3.17–7.67)	0.57 (0.50–0.64)
Mortality						
Covichem Score, n = 481	ROC curve, 0.096	75.0 (70.8–78.8)	78.05 (67.5–86.4)	58.15 (53.1–63.0)	1.86 (1.58–2.19)	0.38 (0.25–0.57)
ICU						
Covichem Score, n = 487	ROC curve, 0.092	72.3 (68.0–76.2)	78.35 (68.8–86.1)	56.67 (51.6–61.6)	1.81 (1.55–2.11)	0.38 (0.26–0.56)

n: number, IQR: interquartile range, AUC: area under curve, CI: confidence interval, LR: likelihood ratio, ICU: intensive care unit.
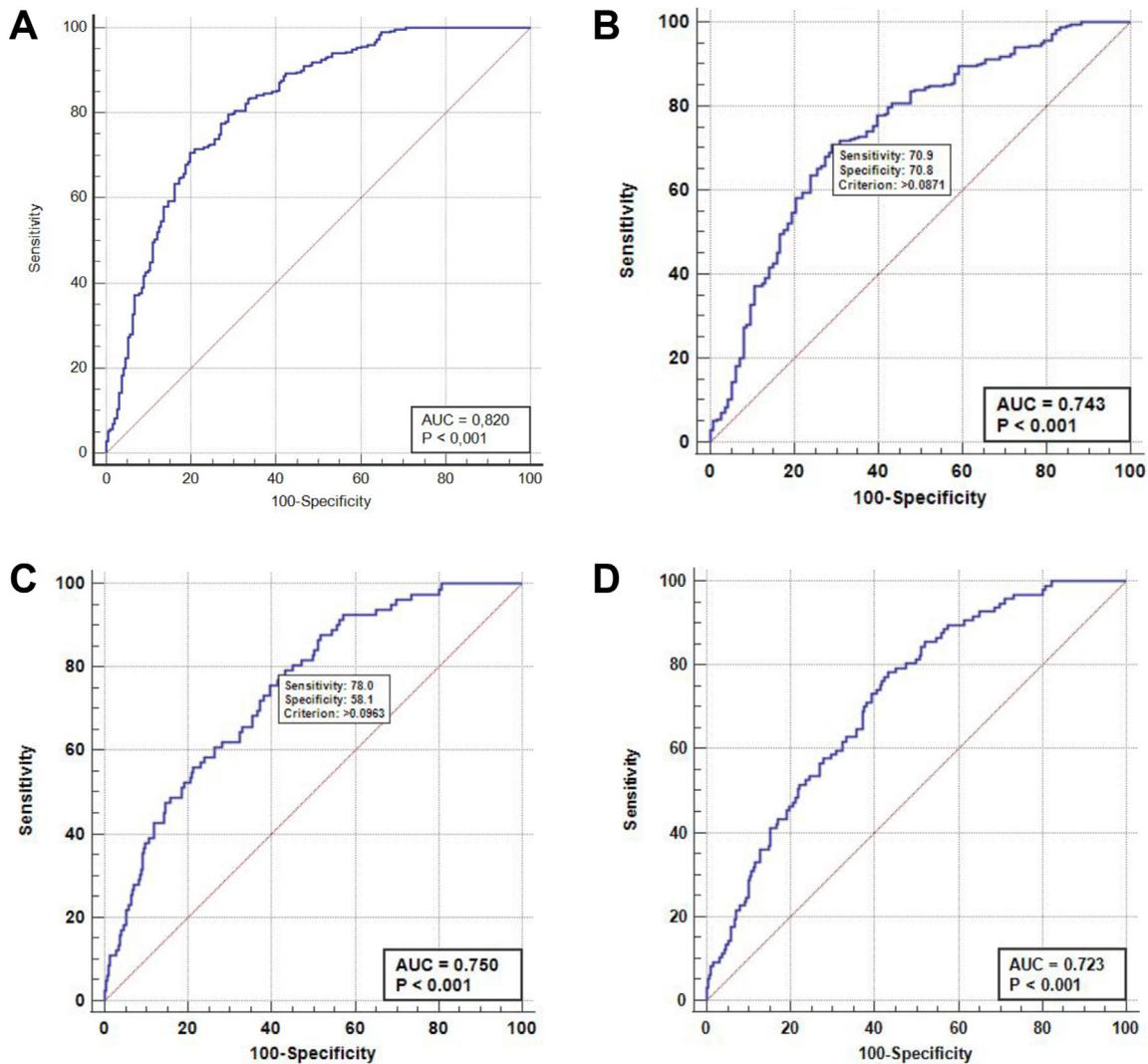
Fig. 2. ROC curves of the covichem score to predict (A) disease severity, (B) disease severity when outpatients were excluded, (C) mortality, (D) ICU admission.

rates to prevent desaturation. As a result, more patients may have been labeled as severe, leading to a lower threshold for severe disease. Third, different methods for measuring serum albumin can significantly impact the measured value, which significantly affects the Covichem score. We measured albumin levels using bromocresol green dye, which overestimates the amount at low albumin concentrations.

5. Conclusion

In this external validation of the Covichem score, we found that it performed worse than in the original derivation and validation study, even with the assistance of a new cutoff.

Authors' contributions

Cigdem Ozpolat performed conceptualization, methodology, formal analysis, investigation, writing-review and editing. Erhan Altunbas performed conceptualization, software, validation, investigation, data curation, writing original draft.

Funding

Source(s) of support

Credit authorship contribution statement

Cigdem Ozpolat: Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Erhan Altunbas:** Writing – original draft, Validation, Software, Investigation, Data curation, Conceptualization.

Declaration of Competing Interest

None.

Acknowledgments

Appendix A. Supplementary data

References

[1] WHO. Coronavirus (COVID-19) Dashboard [Internet]. Available from: https://covid19.who.int; 2022 Mar 10.

[2] Fang X, Li S, Yu H, Wang P, Zhang Y, Chen Z, et al. Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. Aging (Albany NY). 2020 Jul 13;12(13):12493–503.

[3] Izcovich A, Ragusa MA, Tortosa F, Lavena Marzio MA, Agnoletti C, Bengolea A, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: a systematic review. PLoS One. 2020;15(11):e0241955.

[4] Bats M-L, Rucheton B, Fleur T, Orieux A, Chemin C, Rubin S, et al. Covichem: a biochemical severity risk score of COVID-19 upon hospital admission. PLoS One. 2021;16(5):e0250956.

[5] Prokop M, van Everdingen W, van Rees Vellinga T, Quarles van Ufford H, Stöger L, Beenen L, et al. CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19-definition and evaluation. Radiology. 2020 Aug;296(2):E97–104.

[6] Hajian-Tilaki K. Sample size estimation in epidemiologic studies. Caspian J Intern Med. 2011;2(4):289–98.

[7] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020 Apr 7;369:m1328.

[8] Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. Clin Infect Dis. 2020 Sep 12;71(6):1393–9.

[9] Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong. China Clin Infect Dis. 2020 Jul 28;71(15):833–40.

[10] Garcia Moreira V, Beridze Vaktangova N, Martinez Gago MD, Laborda Gonzalez B, Garcia Alonso S, Fernandez Rodriguez E. Overestimation of albumin measured by Bromocresol green vs Bromocresol purple method: influence of acute-phase globulins. Lab Med. 2018 Oct 11;49(4):355–61.

[11] Halalau A, Imam Z, Karabon P, Mankuzhy N, Shaheen A, Tu J, et al. External validation of a clinical risk score to predict hospital admission and in-hospital mortality in COVID-19 patients. Ann Med. 2021 Dec;53(1):78–86.

[12] Wang Z, Wang Z. Identification of risk factors for in-hospital death of COVID - 19 pneumonia – lessions from the early outbreak. BMC Infect Dis. 2021 Jan 25;21(1):113.