External Validation and Recalibration of the CURB-65 and PSI for Predicting 30-Day Mortality and Critical Care Intervention in Multiethnic Patients with COVID-19

Amr Elmoheen [1,2,*], Ibrahim Abdelhafez [2,#], Waleed Salem [1,2,#], Mohamed Bahgat [1], Ali Elkandow [1], Amina Tarig [1], Nauman Arshad [1], Khoulod Mohamed [2], Maryam Al-Hitmi [2], Mona Saad [2], Fatima Emam [2], Samah Taha [2], Khalid Bashir [1,2], Aftab Azad [1,2]

[1] *Department of Emergency Medicine, Hamad Medical Corporation, Doha, Qatar*
[2] *College of Medicine, QU Health, Qatar University, Doha, Qatar*

A R T I C L E   I N F O

A B S T R A C T

Objectives: To validate and recalibrate the CURB-65 and pneumonia severity index (PSI) in predicting 30-day mortality and critical care intervention (CCI) in a multiethnic population with COVID-19, along with evaluating both models in predicting CCI.

Methods: Retrospective data was collected for 1181 patients admitted to the largest hospital in Qatar with COVID-19 pneumonia. The area under the curve (AUC), calibration curves, and other metrics were bootstrapped to examine the performance of the models. Variables constituting the CURB-65 and PSI scores underwent further analysis using the Least Absolute Shrinkage and Selection Operator (LASSO) along with logistic regression to develop a model predicting CCI. Complex machine learning models were built for comparative analysis.

Results: The PSI performed better than CURB-65 in predicting 30-day mortality (AUC 0.83, 0.78 respectively), while CURB-65 outperformed PSI in predicting CCI (AUC 0.78, 0.70 respectively). The modified PSI/CURB-65 model (respiratory rate, oxygen saturation, hematocrit, age, sodium, and glucose) predicting CCI had excellent accuracy (AUC 0.823) and good calibration.

Conclusions: Our study recalibrated, externally validated the PSI and CURB-65 for predicting 30-day mortality and CCI, and developed a model for predicting CCI. Our tool can potentially guide clinicians in Qatar to stratify patients with COVID-19 pneumonia.

Introduction

Illness from coronavirus first appeared almost 20 years ago through severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS). At the end of 2019, COVID-19 (SARS-CoV-2) took the world by storm, spreading even to the most remote places. Unfortunately, COVID-19 proved to be far more challenging than its predecessors. The World Health Organization (WHO) reported that as of April 8, 2021, the number

of confirmed cases of COVID-19 was 132,730,691, with staggering 2,880,726 deaths worldwide. As of April 7, 2021, a total of 650,382,819 vaccine doses have been administered (WHO, 2021).

The State of Qatar has a comprehensive public healthcare system of government-operated facilities of the Hamad Medical Corporation (HMC). All acute care cases, including adults and children, are seen. The emergency physicians (EPs) are usually involved in the initial assessments and the clinical management decisions to direct the patients to the most appropriate facility according to the severity of the disease.

Recently, there has been an urgent need to predict critical care intervention (CCI) and mortality in patients with COVID-19. In the State of Qatar, as of April 16, 2021, there have been 194,930 confirmed cases with 367 deaths and 1,209,648 vaccine doses administered (COVID19 - Qatari Ministry of Publich Health 2021). Qatar

* Correspondence author. Amr Elmoheen, Consultant Emergency Medicine, Department of Emergency Medicine, Hamad Medical Corporation, P.O. Box 3050, Doha, Qatar, +97433695722
E-mail address: aelmoheen@hamad.qa (A. Elmoheen).
# Both authors contributed equally to this manuscript.

is currently being hit by the second wave, with an average of 32 admissions to the intensive care unit (ICU) and an average of 327 ICU beds occupied per day. Respiratory illness has been the leading cause of death from COVID-19 worldwide (Ruan et al., 2020).

Different scoring systems can be used to assess the severity of pneumonia (Ranzani et al., 2018). Most of the EPs are familiar with the CURB-65 score (confusion, urea, respiratory rate, blood pressure, and age ≥65 years) (Lim et al., 2003) and pneumonia severity index (PSI) (Fine et al., 1997). While the PSI score is an extensive tool that provides excellent risk stratification of patients with community-acquired pneumonia (CAP), the CURB-65 is easier to use in a clinical setting with its fewer variables (Chalmers et al., 2010, Fine et al., 1997). PSI scores comprise five groups; I, II, III, IV, and V. Patients with scores (I–III) and (IV–V) are deemed low and high-risk groups, respectively. Similarly, the CURB-65 scores range from (0–4), with scores of 0–1 and >2 indicating low and high mortality risks, respectively.

Various prediction models have emerged during the COVID-19 pandemic (Brabrand et al., 2010), aiming to optimize patient stratification for potentially reducing morbidity and mortality. Such scoring systems can only be used after rigorous validation (i.e., external validation) in another population different from where they were developed for reliability testing (Brabrand et al., 2010).

Our first aim of the study was to externally validate and recalibrate the CURB-65 and PSI in predicting CCI and 30-day mortality in a highly multiethnic population. The second aim was to use variables of PSI and CURB-65 to evaluate predictors of CCI amongst COVID-19 patients.

Methods

Study design, setting, and data collection

Our study is retrospective cross-sectional, including 1,181 patients admitted to Hamad Medical Corporation (HMC), Doha, Qatar due to a confirmed diagnosis of COVID-19 pneumonia. Patients were admitted throughout the peak of the first wave of the COVID-19 pandemic in May 2020. HMC is one of the largest hospitals in the region, consisting of multiple facilities delivering medical care for Qatar's whole population. Testing for SARS-CoV-2 is performed for individuals presenting with COVID-19 symptoms, those with reported contact with confirmed COVID-19 cases, and all travelers returning to the country. SARS-CoV-2 infection was diagnosed by either TaqPath COVID-19 Combo Kit (Thermo Fisher Scientific, Waltham, Massachusetts) or Cobas SARS-CoV-2 Test (Roche Diagnostics, Rotkreuz, Switzerland) utilizing upper respiratory tract samples (throat and nasopharyngeal swabs). Testing for SARS-CoV-2 was carried out at the HMC laboratory.

Data was collected from CERNER Millennium electronic healthcare system (Cerner Millennium, Kansas, USA). Our study was conducted following the recommendations of STrengthening the Reporting of Observational Studies in Epidemiology (STROBE) (von Elm et al., 2014). Ethical approval was obtained from the Institutional Review Board (IRB) at HMC Medical Research office (MRC-01-20-461).

External validation of the PSI and CURB-65 scores

External validation refers to the use of new datasets that were not utilized to construct prediction models (Altman and Royston, 2000, Steyerberg, 2019). Although the target population is patients with COVID-19 rather than CAP, both groups share similar characteristics. We assessed the performance of the CURB-65 (Lim et al., 2003) and PSI (Fine et al., 1997) in predicting death within 30-days mortality as well as CCI in COVID-19 patients. These models have been used to predict death within 30-

days in CAP, with the former being recently tested in predicting CCI (Ilg et al., 2019). However, they were not designed to classify patients with COVID-19 pneumonia. For comparative analysis, we performed recalibration of the PSI and CURB-65 via estimating new beta coefficients for all variables in each model in the full dataset.

Potential predictive variables of CCI

We used the PSI and CURB-65 variables as potential predictors of severity. Data were collected for all the variables at admission, including age, data on living in nursing or long-term care residence, and history of comorbid conditions; cerebrovascular disease, renal disease, neoplastic disease, liver disease, congestive heart failure (CHF). Physical measurements included oxygen saturation < 90%, body temperature <35°C or >39.9°C, respiratory rate (RR) ≥ 30 b/m, pulse rate ≥125 b/m, systolic blood pressure (SBP) < 90, diastolic blood pressure (DBP) ≤ 60, in addition to confusion or altered mental status. Imaging results included the presence of pulmonary effusion on chest radiography (CXR). Laboratory measures were blood urea nitrogen (BUN) > 11, pH < 7.35, serum sodium < 130, serum glucose ≥ 14 mmol/l, hematocrit < 30%, partial arterial oxygen pressure (PO_2) < 60. All the continuous variables were dummy coded for ease of clinical use.

Definition of CCI

The severity of COVID-19 (1 vs. 0) was defined as any of the following CCI: invasive or non-invasive mechanical ventilation, extracorporeal membrane oxygenation (ECMO), and/or administration of vasopressor and/or ionotropic medications, commencing assisted ventilation, insertion of invasive catheters including central line and/or arterial line, and/or renal replacement therapy.

Statistical analysis

Categorical variables were reported as n (%) and continuous variables as median (IQR). Continuous variables were dummy coded as (0 vs. 1) in the regression analysis representing above and below cut-off values. We recognize that this approach is often criticized, however; we believe it aids with rapid patient stratification, especially during the second wave of the pandemic at the time of the study (April 2021).

Variable selection techniques

Data were analyzed after a random 80/20 splitting into development and validation datasets, respectively. All 938 patients in the development dataset were included for variable selection. Twenty one variables entered the Least Absolute Shrinkage and Selection Operator (LASSO) binary logistic regression using 10-fold cross-validation for internal validation and lambda within one standard error of the minimum (λ.1se). LASSO is a machine learning logistic regression that minimizes over-fitting and potential collinearity between predictor variables, while shrinking coefficients of the weakest variables to zero and therefore excluding them from further analysis. Lasso regression was performed using the R package "glmnet", followed by regular variable-selection using logistic regression. The β-coefficient of significant predictors was multiplied by 10 and rounded to the nearest integer. The optimal cut-off (15) was derived from the optimal Youden index. Subsequently, subjects with scores <15 and >15 were considered the low and high-risk groups, respectively.

Machine Learning Analysis

The other methodological approach used in this study was based on more complex machine learning (ML) models using the
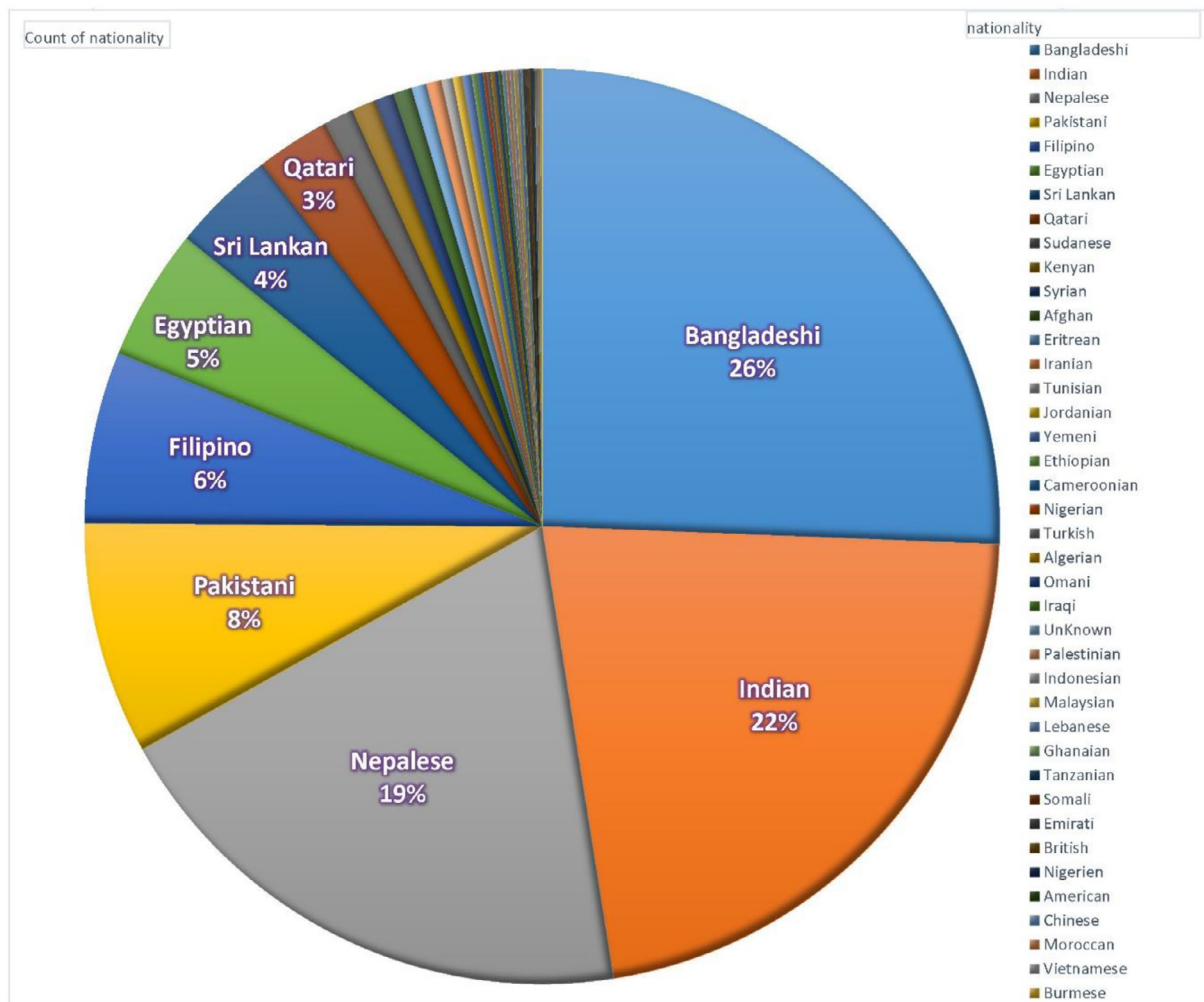
Figure 1. The patients' multiethnic nature.

same data splits and significant variables from the logistic regression model to construct four ML models. These models included random decision forest (RF), gradient boosting (XgBoost), along a deep learning neural network (DL). We used the same parameters utilized in LASSO for internal validation. The ML models were built using R package "h2o" (version 3.32.0.5).

Assessment of accuracy and precision

Assessment of the accuracy of all the models externally validated, recalibrated, and developed was conducted through the area under the curve (AUC). In addition, the AUCs of the four sophisticated ML models were compared with the AUC values of the logistic regression analyses. Calibration (precision) of all the models was evaluated using calibration curves from the "rms" package in R, and values of McKelvoy's R^2, Brier score, calibration slope, intercept were generated using the 200 bootstrap approach to account for optimism. In addition, we performed the DeLong test (DeLong et al., 1988) for AUC comparative analysis. Statistical analysis was conducted using R software (version 4.0.4), with significance being accepted at *p-value* <0.05.

Results

Patients' characteristics

A total of 1181 patients were included in this study, out of which, 45 (3.8%) died within 30 days, and 229 (19.3%) underwent CCI. Patients were 94.5% males between the ages of 19 and 87, median age 43 (IQR: 35-53) while being from diverse nationalities. Patients showed Multiethnic nature and the most common nationalities were Bangladeshi, Indian, Nepalese, Pakistani, Filipino, Egyptian, Sri Lankan, Qatari; 26%, 22%, 19%, 8%, 6%, 5%, 4%, 3% of patients in order (figure 1). The median length of hospital stay was 7 days (IQR: 4-15). Eighty-one (6.85%) patients had at least one comorbidity (Table 1 and Table 2). Deceased patients were more likely to be older, with CCI, with incident hypoxia, tachycardia, tachypnea, and hypotension at admission along with the prolonged length of hospital stay 20 days (IQR: 10-30) (Table 1).

Eight hundred eighty-nine (75.3%) patients had a CURB-65 score of 1, 207 (17.5%) had a score of 2, 57 (4.8%) had a score of 3, 20 (1.7%) had a score of 4, and only 8 (0.7%) scored 5. Out of which, 13 (1.4%) died within 30-days in score 1, 10 (4.83%) in score 2, 14 (24.56%) in score 3, 3 (15%) in score 4, and 5 (68.5%) in score 5. On the other hand, 944 patients (79.9%) were found to be in PSI class I, 44 (3.7%) were in class II, 94 (8.0%) in class III, 60 (5.1%) in class IV, and 39 (3.3%) in class V. The mortality rate was 10 (22.2%) in
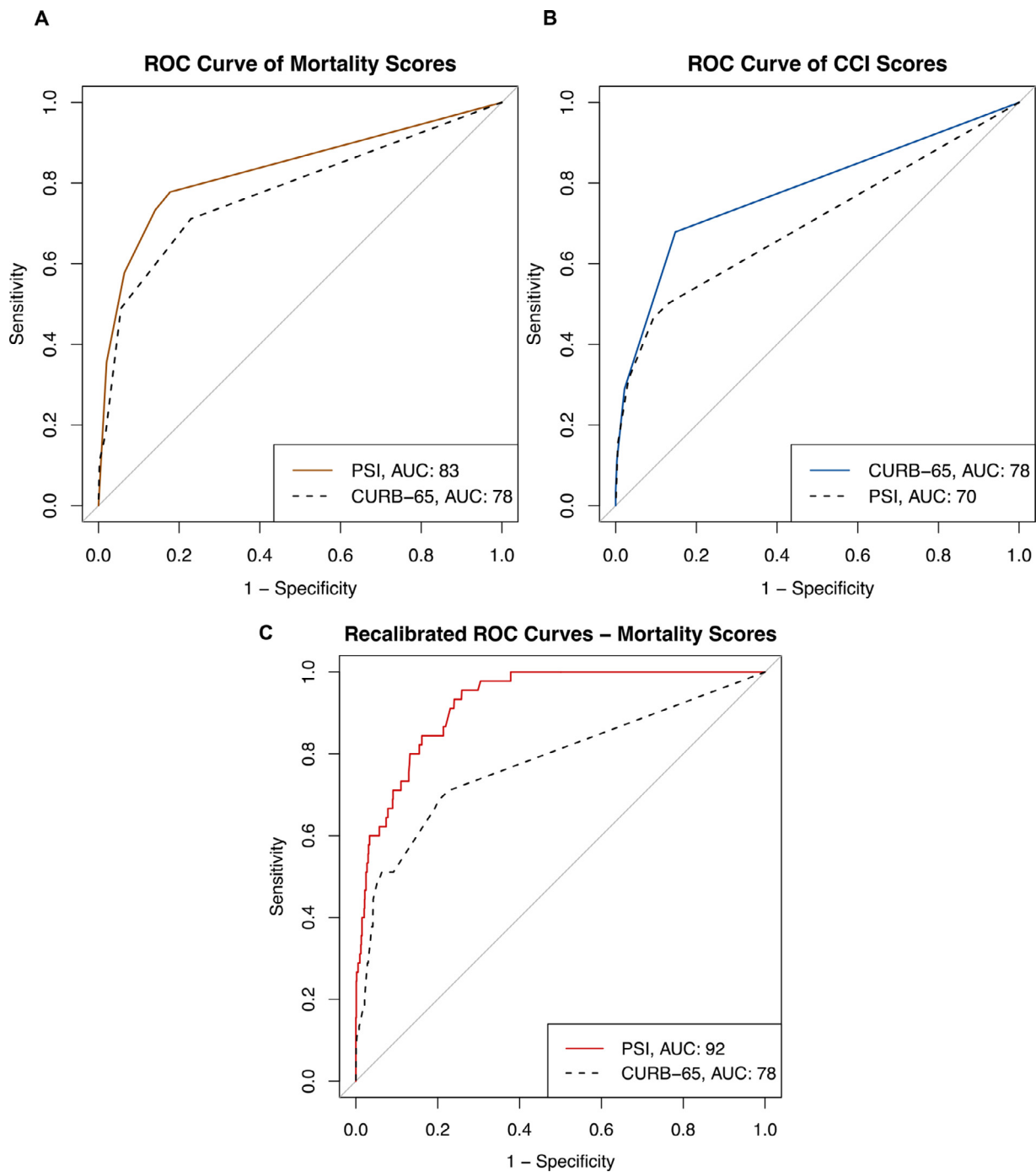
A

B

C

Figure 2. The receiver-operator characteristic (ROC) curve for PSI and CURB-65 in predicting 30-day mortality (A) and critical care intervention (CCI) (B). ROC curve for the recalibrated PSI and CURB-65 in predicting 30-day mortality (C).

class I, 4.5% in class II, 7.4% in class III, 16.6% in class IV, and 41% in class V (Table 1).

External validation of models predicting 30-day mortality

Using the original points, we were able to calculate mortality scores for the CURB-65 and PSI in 1181 (100%) patients. Both models had a comparable performance, with AUCs 0.83 (95% CI 0.765-0.901) and 0.78 (95% CI 0.703-0.855), respectively (Figure 2-A). The DeLong test was insignificant with differences between the AUCs

(0.05, p-value= 0.2232). However, the PSI showed a substantially better calibration in predicting 30-day mortality (Table S1).

External validation of models predicting CCI

Using the original points, we were able to calculate CCI scores for the CURB-65 and PSI models in 1181 (100%) COVID patients. The CURB-65 proved better calibration metrics (Table S1) and better overall accuracy, AUC 0.78 (95% CI 0.746-0.814) compared to the PSI 0.70 (95% CI 0.665-0.738) (Figure 2-B), with a significant
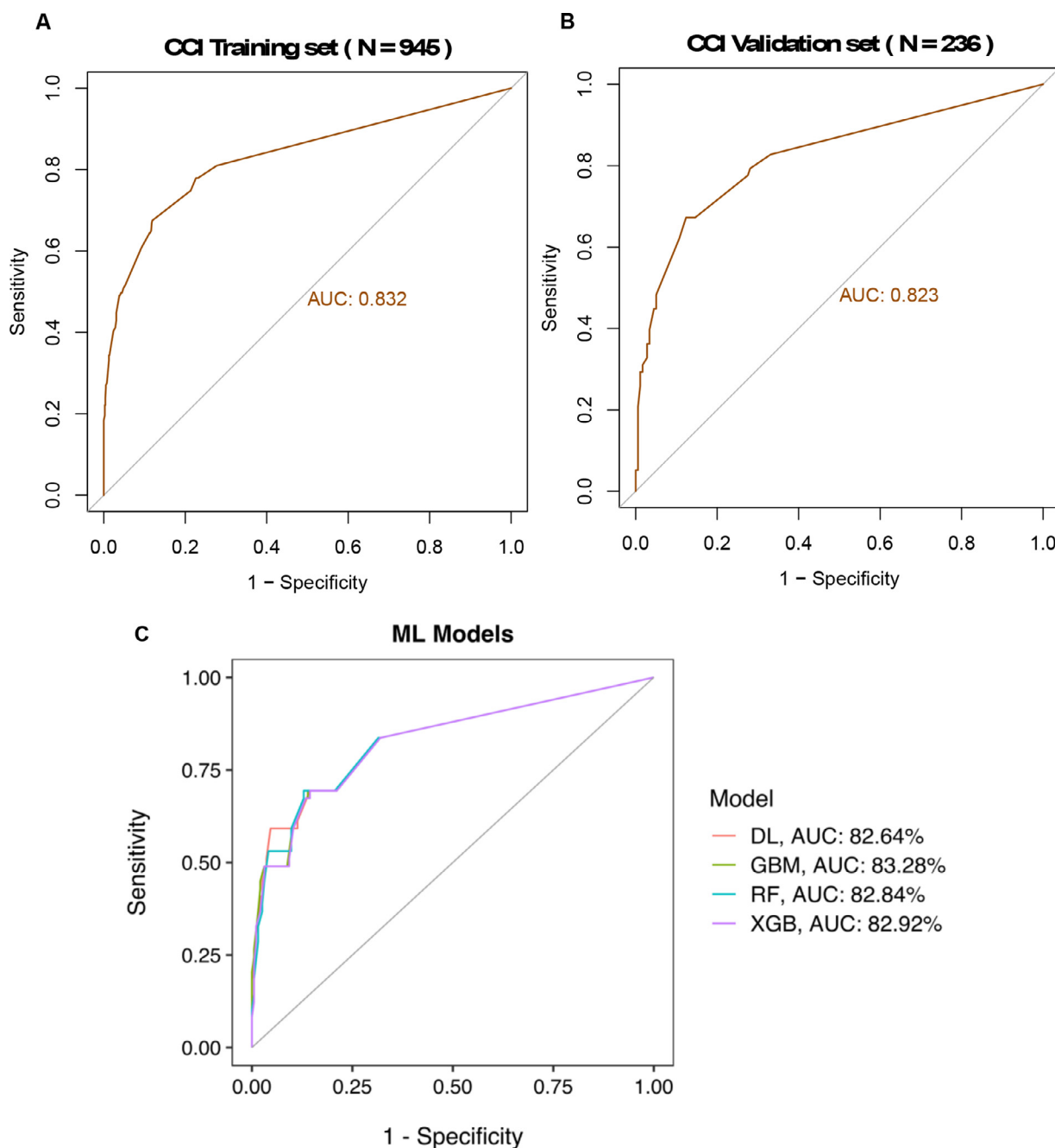
Figure 3. The receiver-operator characteristic (ROC) curve of predicting critical care intervention (CCI) among patients with COVID-19 in the training cohort (A) and the validation cohort (B). ROC curve for predicting T2DM using machine learning models. Deep learning (DL), gradient boosting machine (GBM), random forest (RF), XgBoost (XGB) (C).

DeLong test (differences between the AUCs: 0.08, p-value <0.001). Both models showed comparable calibration (Figure S1-C&D).

Recalibration of models predicting 30-day mortality

Recalibration of the CURB-65 and PSI models 30-day mortality resulted in significant improvement in the PSI accuracy with AUC 0.925 (95% 0.896-0.955) vs. 0.83 (95% CI 0.765-0.901) in the original model (Figure 2-C), with excellent calibration (Figure S2-E).

The DeLong test for the recalibrated models was significant (differences between the AUCs: 0.14, p-value <0.001). While the AUC remaining did not improve in the recalibrated CURB-65 model, we observed an overall better calibration (Figure S2-F). Based on the

recalibrated coefficients (Table S2), a formula was developed for the recalibrated PSI to calculate the risk of 30-day:

Recalibrated PSI risk score = $10.19 - 0.075 \times$ age (years) $+ 1.862 \times$ male $- 9.542 \times$ Nursing home resident (yes) $+ 2.818 \times$ Neoplastic disease (yes) $- 3.207 \times$ Liver disease (yes) $+ 1.747 \times$ History of CHF (yes) $+ 1.526 \times$ Cerebrovascular disease (yes) $+ 0.841 \times$ Renal disease (yes) $+ 1.247 \times$ Confusion/altered mental status (yes) $+ 0.557 \times$ Respiratory rate ≥ 30 (b/m) $+ 0.402 \times$ SBP <90 or DBP ≤ 60 (mmHg) $- 1.419 \times$ Pulse rate ≥ 125 (b/m) $- 0.386 \times$ Temperature $<35^{\circ}$C or >39.9 ($^{\circ}$C) $- 0.194 \times$ BUN ≥ 11 (mmol/L) $+ 0.588 \times$ pH <7.35 (yes) $+ 0.475 \times$ Sodium <130 (mmol/L) $+ 0.650 \times$ Glucose ≥ 14 (mmol/L) $+ 2.559 \times$ Hematocrit $<30 + 1.188 \times$ PO2 <60 (mmHg) $+ 0.554 \times$ Pleural effusion (yes).

Table 1

Demographic, clinical features and laboratory measurements in COVID-19 patients with or without critical care intervention.

Characteristic	N	Total N = 1,181[a]	Survivor N = 1,136[a]	Non-survivor N = 45[a]	p-value[b]
Critical Care Intervention	1,181				<0.0001
Yes		221 (18.7%)	177 (15.6%)	44 (97.8%)	
No		960 (81.3%)	959 (84.4%)	1 (2.2%)	
Length of stay (days)	1,181	7 (4–15)	6 (4–14)	20 (10–30)	<0.0001
Age (years)	1,181	43 (35–53)	42 (34–52.00)	59 (52–65)	<0.0001
Gender (%)	1,181				0.509
Male			1,072 (94.4%)	44 (97.8%)	
Female			64 (5.6%)	1 (2.2%)	
Nursing home resident	1,181		1 (0.1%)	0 (0.0%)	-
Neoplastic disease	1,181		7/1 (0.6%)	5 (11.1%)	<0.0001
Liver disease	1,181		4 (0.4%)	1 (2.2%)	0.176
History of CHF	1,181		7 (0.6%)	6 (13.3%)	<0.0001
Cerebrovascular disease	1,181		16 (1.4%)	3 (6.7%)	0.032
Renal disease	1,181		35 (3.1%)	12 (26.7%)	<0.0001
Confusion/altered mental status	1,181		18 (1.6%)	6 (13.3%)	<0.0001
Respiratory rate ≥30 b/m	1,181		166 (14.6%)	24 (53.3%)	<0.0001
SBP <90 mmHg or DBP ≤ 60 mmHg	1,181		66 (5.8%)	13 (28.9%)	<0.0001
Pulse rate ≥125 b/m	1,181		76 (6.7%)	8 (17.8%)	0.011
Temperature <35°C or >39.9°C	1,181		27 (2.4%)	7 (15.6%)	<0.0001
Oxygen saturation %	1,181	97 (94–98)	97 (95–98)	93 (89–96)	<0.0001
BUN ≥11 mmol/L	1,181		1,093 (96.2%)	31 (68.9%)	<0.0001
pH <7.35	1,181		146 (12.9%)	20 (44.4%)	<0.0001
Sodium <130 mmol/L	1,181		49 (4.3%)	8 (17.8%)	0.001
Glucose ≥14 mmol/L	1,181		138 (12.1%)	16 (35.6%)	<0.0001
Hematocrit <30%	1,181		19 (1.7%)	12 (26.7%)	<0.0001
PO$_2$ <60 mmHg	1,181		75 (6.6%)	19 (42.2%)	<0.0001
Pleural effusion	1,181		102 (9.0%)	7 (15.6%)	0.180
CURB-65 Groups	1,181				<0.0001
Score 0		0 (0%)	0 (0%)	0 (0%)	
Score 1		889 (75.3%)	876 (77.1%)	13 (28.9%)	
Score 2		207 (17.5%)	197 (17.3%)	10 (22.2%)	
Score 3		57 (4.8%)	43 (3.8%)	14 (31.1%)	
Score 4		20 (1.7%)	17 (1.5%)	3 (6.7%)	
Score 5		8 (0.7%)	3 (0.3%)	5 (11.1%)	
PSI Classes	1,181				<0.0001
Group I		944 (79.9%)	934 (82.2%)	10 (22.2%)	
Group II		44 (3.7%)	42 (3.7%)	2 (4.4%)	
Group III		94 (8.0%)	87 (7.7%)	7 (15.6%)	
Group IV		60 (5.1%)	50 (4.4%)	10 (22.2%)	
Group V		39 (3.3%)	23 (2.0%)	16 (35.6%)	

Abbreviations: CCI, critical care intervention; CHF, congestive heart failure; SBP, systolic blood pressure; DBP, diastolic blood pressure; BUN, blood urea nitrogen; PO$_2$, partial pressure of oxygen.

[a] Data is presented as n (%) or median (IQR).

[b] p-values were obtained from Fisher's exact test, Chi-square test, or Mann–Whitney U test.

Variable Selection for predicting CCI

The LASSO selected 7/21 variables (shown in Table 3). Inclusion of the selected variables in logistic regression yielded 6 significant predictive variables of COVID-19 CCI and therefore constituted the risk score. This included RR ≥ 30 bpm (20 points), oxygen saturation <90% (18 points), hematocrit < 30% (15 points), age > 55 years old (12 points), serum sodium < 130 mmol/l (10 points), and serum glucose ≥ 14 mmol/l (8 points) (Table 3).

Comparison of logistic regression and complex ML models

ROC curves of the 6 significant predictors of severe COVID-19 status were generated from the complex ML models. The ML models' performance in the validation cohorts was similar to the regression model (Figure 3-A&B) with AUC=0.826 for DL, AUC=0.832 for GBM, AUC=0.828 for RF, and AUC=0.829 for XGB (Figure 3-C).

Discussion

To our knowledge, the CURB-65 and PSI have not been used or recalibrated to assess mortality in COVID-19 patients among multicultural populations from different countries living in Qatar, Middle East. This study utilizes these common tools in risk stratify-ing patients with community-acquired pneumonia (CAP) in this diverse population group.

Qatar is an Arab state on the Qatar Peninsula with a heavy male predominance and females representing only 25% of the total population due to the high flow of male laborers. The population has tripled in the past 10 years up to 2011, with Qatari locals representing <15% of the whole population. This is followed by other Arab nationalities (13%), Indian (24%), Nepali (16%), Filipino (11%), Bangladeshi (5%) and Sri Lankans (5%) .

The CURB-65 and the PSI can reliably predict the mortality of in-patients with CAP. Previously, these tools have been used to assess CAP's mortality, but they have not been used to assess mortality in COVID19 patients amongst multicultural populations from different countries living in Qatar, Middle East. In this study, we assessed, recalibrated, and modified the PSI and CURB-65 as prognostic scoring mechanisms in the prediction of 30-day mortality and CCI.

Although 30-day mortality (3.8%) in our study was lower than recent reports (Artero et al., 2021), our findings have shown that 71% and 68.8% of our sample underwent CCI after admission to ICU with COVID-19 pneumonia. The CURB-65 and PSI scores were 1–2 and 1-3, respectively, with a relatively high need for CCI (18.7%), hence the need for clinical models for COVID-19 pneumonia based on other outcomes modified by in-patient care as compared to

Table 2

Demographic, clinical features and laboratory measurements among COVID-19 patients based on survival and 30-day mortality

Characteristic	N	Total N = 1181[a]	Non-CCI N = 960[a]	CCI N = 221[a]	p-value[b]
Survival	1,181				<0.0001
Survivor			959 (99.9%)	177 (80.1%)	
Non-survivor			1 (0.1%)	44 (19.9%)	
Length of stay (days)	1,181	7 (4–15)	5 (3–9)	22 (16–30)	<0.0001
Age (years)	1,181	43 (35–53)	41 (34–50)	52 (41–61)	<0.0001
Gender (%)	1,181				0.025
Male			914 (95.2%)	202 (91.4%)	
Female			46 (4.8%)	19 (8.6%)	
Nursing home resident	1,181				0.187
Neoplastic disease	1,181		5 (0.5%)	7 (3.2%)	0.002
Liver disease	1,181		4 (0.4%)	1 (0.5%)	1
History of CHF	1,181		3 (0.3%)	10 (4.5%)	<0.0001
Cerebrovascular disease	1,181		11 (1.1%)	8 (3.6%)	0.015
Renal disease	1,181		19 (2.0%)	28 (12.7%)	<0.0001
Confusion / altered mental status	1,181		10 (1.0%)	14 (6.3%)	<0.0001
Respiratory rate ≥30 b/m	1,181		72 (7.5%)	118 (53.4%)	<0.0001
SBP <90 mmHg or DBP ≤ 60 mmHg	1,181		38 (4.0%)	41 (18.6%)	<0.0001
Pulse rate ≥125 b/m	1,181		47 (4.9%)	37 (16.7%)	<0.0001
Temperature <35°C or >39.9°C	1,181		15 (1.6%)	19 (8.6%)	<0.0001
Oxygen saturation %	1,181	97 (94–98)	97 (95–98)	93 (89–96)	<0.0001
BUN ≥11 mmol/L	1,181		15 (1.6%)	42 (19.0%)	<0.0001
pH <7.35	1,181		105 (10.9%)	61 (27.6%)	<0.0001
Sodium <130 mmol/L	1,181		30 (3.1%)	27 (12.2%)	<0.0001
Glucose ≥14 mmol/L	1,181		87 (9.1%)	67 (30.3%)	<0.0001
Hematocrit <30%	1,181		11 (1.1%)	20 (9.0%)	<0.0001
PO$_2$ <60 mmHg	1,181		24 (2.5%)	70 (31.7%)	<0.0001
Pleural effusion	1,181		76 (7.9%)	33 (14.9%)	0.001
CURB-65 Classes	1,181				<0.0001
Score 0		0 (0%)	0 (0%)	0 (0%)	
Score 1		889 (75.3%)	818 (85.2%)	71 (32.1%)	
Score 2		207 (17.5%)	121 (12.6%)	86 (38.9%)	
Score 3		57 (4.8%)	18 (1.9%)	39 (17.6%)	
Score 4		20 (1.7%)	3 (0.3%)	17 (7.7%)	
Score 5		8 (0.7%)	0 (0%)	8 (3.6%)	
PSI Classes	1181				<0.0001
Group I		944 (79.9%)	834 (86.9)	110 (49.8%)	
Group II		44 (3.7%)	36 (3.8%)	8 (3.6%)	
Group III		94 (8.0%)	60 (6.2%)	34 (15.4%)	
Group IV		60 (5.1%)	25 (2.6%)	35 (15.8%)	
Group V		39 (3.3%)	5 (0.5%)	34 (15.4%)	

Abbreviations: CCI, critical care intervention; CHF, congestive heart failure; SBP, systolic blood pressure; DBP, diastolic blood pressure; BUN, blood urea nitrogen; PO$_2$, partial pressure of oxygen.

[a] Data is presented as n (%) or median (IQR)

[b] p-values were obtained from Fisher's exact test, Chi-square test or Mann–Whitney U test.

mortality. Therefore, we performed variable selection using LASSO and logistic regression on all the CURB-65 and PSI variables to predict CCI, with only 6/21 significant including RR ≥ 30 bpm, O2 saturation <90%, hematocrit < 30%, age> 55 years old, serum sodium < 130 mmol/l, and serum glucose ≥ 14 mmol/l. The model showed better accuracy and calibration compared to the PSI and CURB-65 scores.

While many studies showed the PSI and CURB-65 systems to be beneficial in the assessment of COVID-19 severity and good predictors of mortality, some found other scoring systems better. SCAP score (Severe Community-Acquired Pneumonia) was found to be a more accurate marker of 14-day mortality (Anurag and Preetam, 2021). Another cohort found the COVID-GRAM score as the preferred means to identify patients with higher mortality with pneumonia caused by SARS-CoV-2 (Esteban Ronda et al., 2021).

Another cohort used NEWS2 (National Early Warning Score 2) and found it to be more beneficial than SIRS and qSOFA but not as predictive as CURB-65 and PSI (Holten et al., 2020). Yet another study found A-DROP, a modified version of CURB-65, to be superior in predicting in-hospital death compared to other widely used CAP-specific tools (Fan et al., 2020).

External validation was essential in other cohorts as well as this one since it was expected that newer scoring systems outperform older ones (Neto et al., 2021). PSI and CURB-65 have been used for risk stratification and prognostic assessment in patients with COVID-19 previously (Wang et al., 2020).

In our cohort, the PSI and CURB-65 performed similarly to other studies on CAP (Ahnert et al., 2019, Asai et al., 2019, George et al., 2019). Notwithstanding its implications, two limitations in this study are important to point out, including its retrospective design and the heavily male predominance. However, the lack of missing data and the robust methodological approach, and Qatar's multiethnic and diverse population represented by a relatively large sample with a wide age group made this study unique, accounting for the lack of population diversity in other similar studies (Artero et al., 2021). External validation proved that well-respected scoring systems such as CURB-65 and PSI lend credence to predictability and prognostic value in treating patients with COVID-19. Further studies are needed to evaluate the model developed in larger cohorts along with testing the validity of the CURB-65 and PSI in predicting CCI. Authors worldwide are encouraged to share data and collaborate in order to develop robust clinical models that can aid with clinical decisions while avoiding issues arising from introducing bias and overfitting (Wynants et al., 2020).

Table 3

Variable selection using the Least Absolute Shrinkage and Selection Operator (LASSO) followed by logistic regression.

Variable	LASSO	Logistic regression		
		Odds ratio (95% CI)	P value	Score
Age >55	**selected**	3.18 (2.02–4.97)	<.001	12
Gender	-	-	-	-
Male	-	-	-	-
Female	-	-	-	-
Nursing home resident	-	-	-	-
Neoplastic disease	-	-	-	-
Liver disease	-	-	-	-
History of CHF	-	-	-	-
Cerebrovascular disease	-	-	-	-
Renal disease	-	-	-	-
Confusion/altered mental status	-	-	-	-
Respiratory rate ≥30 b/m	**selected**	7.41 (4.60–11.97)	<.001	20
SBP <90 mmHg or DBP ≤ 60 mmHg	-	-	-	-
Pulse rate ≥125 b/m	-	-	-	-
Temperature <35°C or >39.9°C	-	-	-	-
BUN ≥11 mmol/L	**selected**	-	-	-
pH <7.35	-	-	-	-
Sodium <130 mmol/L	**selected**	2.84 (1.23–6.41)	.01	10
Glucose ≥14 mmol/L	**selected**	2.18 (1.28–3.69)	.003	8
Hematocrit <30%	**selected**	4.68 (1.42–15.81)	.01	15
PO_2 <60 mmHg	-			
Oxygen saturation <90%	**selected**	5.95 (2.64–14.27)	<.001	18
Pleural effusion [a]	-			

[a] Pleural effusion excluded from entering the analysis, while oxygen saturation was used as a substitute to simplify the score.

Conflict of interest

The authors of this article declare no conflict of interest.

Author Contributions

Conceptualization, AE; Data collection AE, IA, WA, AT, NA, MB, AE, MA-H, MS, KM, FE and ST; Data curation and methodology, AE, and IA; Formal analysis, IA; Writing—original draft preparation, AE and IA; Writing—review and editing; AE and IA; Supervision, KB, AA. All authors approved the final version of this manuscript for submission.

Uncited References:

COVID19 - Qatari Ministry of Publich Health 2021, Qatar Open Data - Coronavirus Disease 2021, World Population Review 2021

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2021.08.027.

References

COVID19 - Qatari Ministry of Publich Health; Available from: https://covid19.moph.gov.qa/EN/Pages/default.aspx. 2021].

Qatar Open Data - Coronavirus Disease 2019 (COVID-19) Statistics; Available from: https://qatar.opendatasoft.com/explore/dataset/covid-19-cases-in-qatar/table/?sort=date. 2021].

World Population Review; Available from: https://worldpopulationreview.com/countries/qatar-population. 2021].

Ahnert P, Creutz P, Horn K, Schwarzenberger F, Kiehntopf M, Hossain H, et al. Sequential organ failure assessment score is an excellent operationalization of disease severity of adult patients with hospitalized community acquired pneumonia - results from the prospective observational PROGRESS study. Crit Care 2019;23(1):110.

Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19(4):453–73.

Anurag A, Preetam M. Validation of PSI/PORT, CURB-65 and SCAP scoring system in COVID-19 pneumonia for prediction of disease severity and 14-day mortality. Clin Respir J 2021.

Artero A, Madrazo M, Fernández-Garcés M, Muiño Miguez A, González García A, Crestelo Vieitez A,, et al. Severity Scores in COVID-19 Pneumonia: a Multicenter, Retrospective, Cohort Study. J Gen Intern Med 2021:1–8.

Asai N, Watanabe H, Shiota A, Kato H, Sakanashi D, Hagihara M, et al. Efficacy and accuracy of qSOFA and SOFA scores as prognostic tools for community-acquired and healthcare-associated pneumonia. Int J Infect Dis 2019;84:89–96.

Brabrand M, Folkestad L, Clausen NG, Knudsen T, Hallas J. Risk scoring systems for adults admitted to the emergency department: a systematic review. Scand J Trauma Resusc Emerg Med 2010;18:8.

Chalmers JD, Singanayagam A, Akram AR, Mandal P, Short PM, Choudhury G, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. Thorax 2010;65(10):878–83.

DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–45.

Esteban Ronda V, Ruiz Alcaraz S, Ruiz Torregrosa P, Giménez Suau M, Nofuentes Pérez E, León Ramírez JM, et al. Application of validated severity scores for pneumonia caused by SARS-CoV-2. Med Clin (Barc) 2021.

Fan G, Tu C, Zhou F, Liu Z, Wang Y, Song B, et al. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. Eur Respir J 2020;56(3).

Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med 1997;336(4):243–50.

George N, Elie-Turenne MC, Seethala RR, Baslanti TO, Bozorgmehri S, Mark K, et al. External Validation of the qSOFA Score in Emergency Department Patients With Pneumonia. J Emerg Med 2019;57(6):755–64.

Holten AR, Nore KG, Tveiten C, Olasveengen TM, Tonby K. Predicting severe COVID-19 in the Emergency Department. Resusc Plus 2020;4.

Ilg A, Moskowitz A, Konanki V, Patel PV, Chase M, Grossestreuer AV, et al. Performance of the CURB-65 Score in Predicting Critical Care Interventions in Patients Admitted With Community-Acquired Pneumonia. Ann Emerg Med 2019;74(1):60–8.

Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003;58(5):377–82.

Neto FL, Marino LO, Torres A, Cilloniz C, Meirelles Marchini JF, Garcia de Alencar JC, et al. Community-acquired Pneumonia Severity Assessment Tools in Patients Hospitalized with COVID-19: a Validation and Clinical Applicability Study. Clinical Microbiology and Infection 2021.

Ranzani OT, Taniguchi LU, Torres A. Severity scoring systems for pneumonia: current understanding and next steps. Curr Opin Pulm Med 2018;24(3):227–36.

Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to

COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020;46(5):846–8.

Steyerberg EW. Clinical prediction models. Springer; 2019.

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg 2014;12(12):1495–9.

Wang X, Hu ZW, Hu Y, Cheng Y, Zhang H, Li HC, et al. [Comparison of severity classification of Chinese protocol, pneumonia severity index and CURB-65 in risk stratification and prognostic assessment of coronavirus disease 2019]. Zhonghua Jie He He Hu Xi Za Zhi 2020;43(10):834–8.

WHO. WHO Coronavirus (COVID-19) Dashboard; 2021. Available from: https://covid19.who.int. 2021].

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