Received: 14 September 2020 | Accepted: 3 December 2020

DOI: 10.1111/ijcp.13926

SHORT REPORT

INFECTIOUS DISEASES

THE INTERNATIONAL JOURNAL OF CLINICAL PRACTICE WILEY

# Validation of pneumonia prognostic scores in a statewide cohort of hospitalised patients with COVID-19

Yiyun Shi[1] | Aakriti Pandita[2] | Anna Hardesty[1] | Meghan McCarthy[2] | Jad Aridi[2] | Zoe F. Weiss[3] | Curt G. Beckwith[2] | Dimitrios Farmakiotis[2]

[1]Department of Medicine, The Warren Alpert Medical School of Brown University, Providence, RI, USA

[2]Division of Infectious Diseases, The Warren Alpert Medical School of Brown University, Providence, RI, USA

[3]Division of Infectious Diseases, Brigham and Women's Hospital and Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

**Correspondence**
Yiyun Shi, 593 Eddy Street, Gerry House 111, Providence, RI, 02903, USA.
Email: yiyun_shi@brown.edu

**Funding information**
Brown Physicians Inc. Academic Assessment Grant

## Abstract

**Objective:** We aimed to externally validate the predictive performance of two recently developed COVID-19-specific prognostic tools, the COVID-GRAM and CALL scores, and prior prognostic scores for community-acquired pneumonia (CURB-65), viral pneumonia (MuBLSTA) and H1N1 influenza pneumonia (Influenza risk score) in a contemporary US cohort.

**Methods:** We included 257 hospitalised patients with laboratory-confirmed COVID-19 pneumonia from three teaching hospitals in Rhode Island. We extracted data from within the first 24 hours of admission. Variables were excluded if values were missing in >20% of cases, otherwise, missing values were imputed. One hundred and fifteen patients with complete data after imputation were used for the primary analysis. Sensitivity analysis was performed after the exclusion of one variable (LDH) in the complete dataset (n = 257). Primary and secondary outcomes were in-hospital mortality and critical illness (mechanical ventilation or death), respectively.

**Results:** Only the areas under the receiver-operating characteristic curves (RO-AUC) of COVID-GRAM (RO-AUC = 0.775, 95% CI 0.525-0.915) for in-hospital death, and CURB65 for in-hospital death (RO-AUC = 0.842, 95% CI 0.674-0.932) or critical illness (RO-AUC = 0.766, 95% CI 0.584-0.884) were significantly better than random. Sensitivity analysis yielded similar trends. Calibration plots showed better agreement between the estimated and observed probability of in-hospital death for CURB65, compared with COVID-GRAM. The negative predictive value (NPV) of CURB65 ≥2 was 97.2% for in-hospital death and 88.1% for critical illness.

**Conclusions:** The COVID-GRAM score demonstrated acceptable predictive performance for in-hospital death. The CURB65 score had better prognostic utility for in-hospital death and critical illness. The high NPV of CURB65 values ≥2 may be useful in triaging and allocation of resources.

## 1 | BACKGROUND

Risk stratification tools for patients with COVID-19 are needed to provide validated triaging decision support, especially in overwhelmed healthcare settings. Consequently, more than 20 prognostic models[1-3] have been proposed to predict progression to severe pneumonia and death. Many variables in these scores overlap with those in models developed to predict the severity of community-acquired pneumonia.

To this end, we conducted a contemporary study to compare two COVID19-specific prognostic models recently developed in China, the COVID-GRAM[3] and CALL scores,[4] and previous prognostic tools for community-acquired pneumonia (CAP) (CURB65[5]), viral

  

pneumonia (MuLBSTA[6]) or H1N1-influenza pneumonia (influenza risk score[7]) in a contemporary cohort in the United States (US).

## 2 | METHODS

We retrospectively studied 257 adult patients admitted for COVID-19, diagnosed by PCR in a nasopharyngeal sample, at three teaching hospitals (Newport Hospital, Newport; The Miriam Hospital, Rhode Island Hospital, Providence) in RI, USA, between 3/1/2020 and 5/18/2020. The study was approved by the Lifespan Institutional Review Board, with a waiver of informed consent.

The first 97 patients were enrolled consecutively from 3/1/2020 to 4/3/2020. During the surge in April and May, 160 additional patients were randomly enrolled because of the limitations in our abstraction capacity. To ensure that we had a relatively representative sample, we compared weekly case fatality rates between the study sample and all COVID-19 patients admitted to the three hospitals during the study period with the Wilcoxon test.

The COVID-GRAM,[3] CALL,[4] CURB65,[5] MuLBSTA[6] and influenza risk[7] scores were calculated using clinical information collected within the first 24 hours of admission. Values missing in <20% of patients were imputed using predictive mean matching for continuous variables and logistic regression for categorical variables, as in the previous reports.[2,3,6] Subsequent analyses were pooled per Rubin's rule[8] from five post-imputation subsets. Analyses involving variables missing in >20% of patients were conducted on the subset with complete values.

Our primary outcome was in-hospital death; the secondary outcome was critical illness, defined as mechanical ventilation or in-hospital death. Categorical variables were compared with $X^2$ or Fisher's exact (for expected frequencies <5) tests. Continuous variables were compared with Student's *t* test or the Mann-Whitney *U* criterion, for variables that had normal distribution (assessed by the Shapiro-Wilk test) or not, respectively. We built Receiver Operating Characteristic (ROC) curves to assess the predictive performance of all scores for the primary and secondary outcomes. We calculated pooled Areas Under the Curve (AUC) and 95% Confidence Intervals (CI).

In sensitivity analysis, we validated prognostic scores without LDH (CURB65, MuLBSTA, influenza risk score) in patients with available LDH levels (n = 115), and scores with LDH (COVID-GRAM, CALL), after the removal of LDH values from these models, in the entire cohort (n = 257). We assessed the fitness (agreement between estimated and observed probability) of scores with statistically significant RO-AUC (lowest 95% CI >0.5) for in-hospital death, by means of calibration plots with intercept adjustment. Last, we calculated sensitivity, specificity, positive (PPV) and negative (NPV) predictive value of the CURB65 score, given its good performance and simplicity, for in-hospital death and progression to critical COVID-19, using a cut-off value of 2, similar to CAP.[5] Data were analysed by R software (version 3.6.3, R Foundation).

## 3 | RESULTS

There were no significant differences in weekly in-hospital case-fatality rates between the study sample (n = 257) and the whole patient population (n = 817) of patients with COVID-19 admitted to the three hospitals during the study period (Wilcoxon $P = .412$).

The only parameter with >20% missing values in the first 24 hours was lactate dehydrogenase (LDH), in 142 patients (55.3%), who were excluded from the initial comparisons of scores with LDH as one of the parameters. Direct bilirubin levels were missing in 45 (17.5%), neutrophil or lymphocyte counts in 4 (1.6%), and blood urea nitrogen in 3 (1.5%) patients. These values were imputed.

Mortality was associated with advanced age, presence of certain comorbidities (hypertension, diabetes), admission from a nursing home, hypoxia or tachypnea on admission, thrombotic events during hospitalisation, higher LDH, BUN, bilirubin, white blood cell count, neutrophil-to-lymphocyte ratio, lower albumin and $tCO_2$ on admission. The comparisons between patients who developed critical illness vs those who did not follow a similar pattern, except that patients who developed critical illness had a higher percentage of unconsciousness, imaging abnormalities, ferritin and aspartate transaminase (AST) levels and incidence of other viral coinfections. There was no difference in the percentage of hypertension or the level of direct bilirubin (Table S1).

For in-hospital death, only the RO-AUC of the COVID-GRAM and CURB65 scores were significantly better than random (Table 1, Figure 1). For critical illness, only the RO-AUC of the CURB65 score was significantly better than random (Table 1, Figure S1). Validation of models without LDH in patients with available LDH levels (n = 115, Table S2), and validation of models developed with LDH in the entire cohort, after the removal of LDH (n = 257, Table S3), yielded similar results.

The CURB65 score showed better agreement between the estimated and observed probability of in-hospital death compared with the COVID-GRAM score (Figure , calibration slopes of 1.03 vs 0.62, respectively). In all patients (n = 257), sensitivity of CURB65 $\geq 2$[5] for predicting in-hospital death was 89.5%, specificity 63.5%, PPV 29.8%, NPV 97.2%. For critical illness, sensitivity was 71.2%, specificity 63.6%, PPV 36.8%, NPV 88.1%. In-hospital mortality for patients with CURB65 ≥2 was 29.9%.

## 4 | DISCUSSION

There has been a worldwide effort to develop COVID-19-specific prognostic tools. The variables used in such models are often similar to previously validated pneumonia prediction tools. The CURB65 model, likely the easiest score to calculate, has been widely used to compare the predictive value of new scoring systems. While its performance is usually inferior compared with novel scores in the derivative populations, its RO-AUC overall has been reproducible between 0.7 and 0.9 in the studies of community acquired,[9] viral[6] or

**TABLE 1** Prognostic performance of different pneumonia scores in hospitalised patients with COVID-19

| Score | Variables | n | Mortality: RO-AUC (95% CI) | Critical illness: RO-AUC (95% CI) |
|---|---|---|---|---|
| COVID-GRAM | Age<br>Abnormal chest X-ray<br>Hemoptysis<br>Dyspnea<br>Unconsciousness<br>Comorbidities<br>Cancer<br>ANC/ALC ratio<br>LDH<br>Direct bilirubin | 115 | **0.775 (0.525-0.915)** | 0.698 (0.436-0.874) |
| CALL | Any comorbidity<br>Age > 60 y<br>ALC ≤ $1.0 \times 10^9$/L<br>LDH<br>≤250 vs. 250-500 vs. >500 U/L | 115 | 0.640 (0.361-0.849) | 0.573 (0.318-0.794) |
| CURB65 | Confusion<br>BUN > 19 mg/dL<br>RR > 30 bpm<br>Hypotension:<br>SBP < 90 or<br>DBP < 60 mmHg<br>Age ≥ 65 y | 257 | **0.842 (0.674-0.932)** | **0.766 (0.584-0.884)** |
| MuLBSTA | Multilobar infiltrates on chest X-ray or CT<br>Bacterial infection<br>ALC < $0.8 \times 10^9$/Lt<br>Age ≥ 65 y<br>Hypertension<br>Smoking | 257 | 0.650 (0.425-0.823) | 0.614 (0.411-0.783) |
| Influenza risk | Age > 45 y<br>Male sex<br>≥3 comorbidities<br>Pneumonia<br>Confusion<br>Dyspnea | 257 | 0.616 (0.390-0.802) | 0.601 (0.397-0.774) |

Abbreviations: ALC, absolute lymphocyte count; ANC, absolute neutrophil count; bpm, breaths per minute; BUN, blood urea nitrogen; CI, confidence intervals; CT, computerised tomography; LDH, lactate dehydrogenase; RO-AUC, receiver-operating area under the curve; RR, respiratory rate; SBP, DBP, systolic, diastolic blood pressure.

COVID-19[2,3,10,11] pneumonia, in agreement with our results. The 30-day mortality rates of patients with COVID-19 and CURB65 values ≥2 were 30.5%[10] and 33.3%[11] in two other COVID-19 cohorts, similar to the 29.9% in-hospital mortality rate in our study. Additionally, CURB65 ≥2 had an NPV of >97% for inpatient death and >88% for critical illness. Therefore, the majority of such inpatients can be safely managed outside of the ICU, which could potentially save valuable resources.

Our findings may raise the question of which features of COVID-19 pneumonia are unique enough to warrant specific predictive tools. A higher age cutoff may be needed, compared with other viral pneumonia, as demonstrated by epidemiological studies[12] and better performance of scores with higher age cut-off in our study (Table 1, Table S1). Also, patients with COVID-19 experience more endothelial injury and thromboembolic events when compared with patients with influenza pneumonia,[13] and some demonstrate a hyperinflammatory phenotype with rapid progression.[14] Markers reflecting potential COVID-19-specific sequelae, such as inflammation or hypercoagulability, may enhance the predictive value of scores.

Our study showed that even scores with satisfactory performance in predicting mortality performed poorly in predicting critical illness, which is in agreement with another recent report from Italy, validating CALL score.[15] This may be the reflection of different thresholds of ICU transfer or mechanical ventilation between different countries and time periods.

Our report has limitations, mainly the small number of cases and imputation of missing values. Also, we did not have enough data on coagulation and inflammation to modify the above scores, as

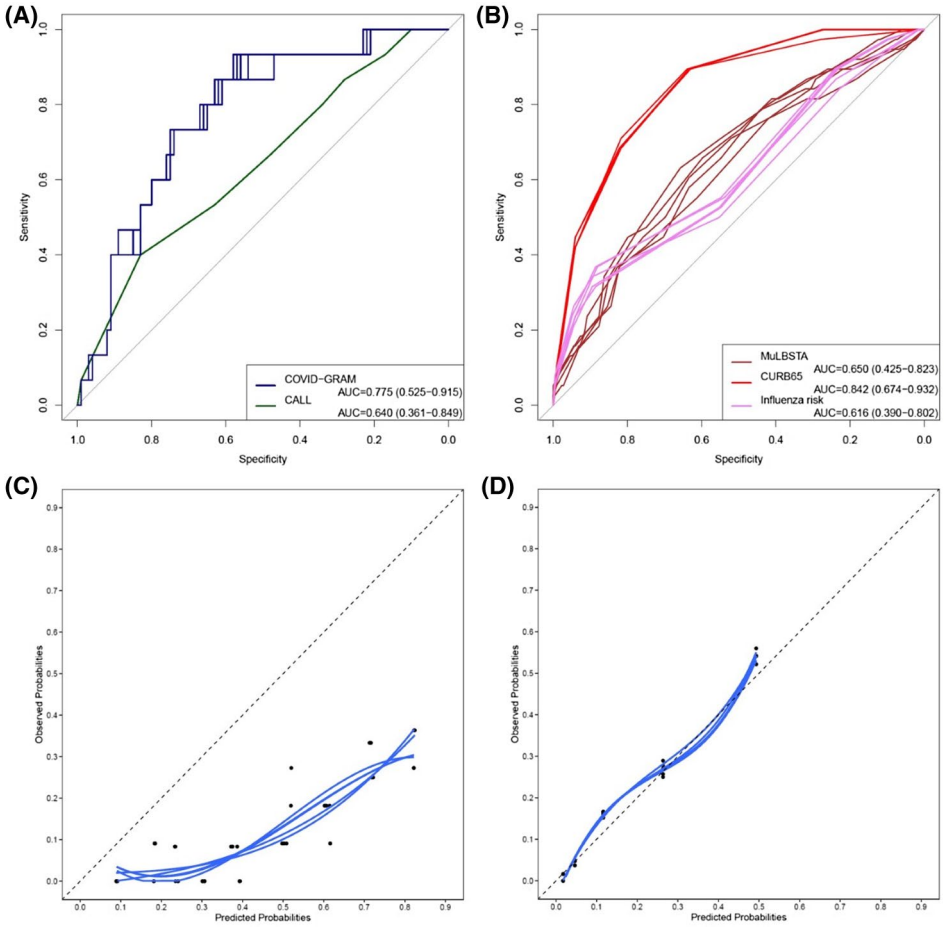


**FIGURE 1** Pooled ROC curves and calibration plots for in-hospital mortality: A, ROC curves for scores with LDH (COVID-GRAM, CALL, n = 115); B, ROC curves for scores without LDH (CURB65, MuLBSTA, Influenza Risk, n = 257); C, Calibration plot for COVID-GRAM; D, Calibration plot for CURB65

they were not routinely ordered during the early pandemic at our hospitals.

In summary, in this contemporary US cohort of inpatients with COVID-19, the easily calculated CURB65 score is potentially useful in predicting critical illness and death. Our findings highlight the value of coordinated efforts to validate and enhance existing scores. These efforts will help streamline patient triage, improve the allocation of resources, and aid in appropriate stratification for the design of future clinical trials.

## DISCLOSURES

DF has received research grants from Astellas and Viracor, and consultation fee from Viracor, outside of the submitted work. All other authors have nothing to disclose. The present work was partially supported by a Brown Physicians Inc Academic Assessment Grant.

## DATA AVAILABILITY STATEMENT

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## ORCID

*Yiyun Shi* https://orcid.org/0000-0002-7175-2125
*Anna Hardesty* https://orcid.org/0000-0002-5519-4477
*Meghan McCarthy* https://orcid.org/0000-0002-0384-030X
*Zoe F. Weiss* https://orcid.org/0000-0003-2931-3660
*Dimitrios Farmakiotis* https://orcid.org/0000-0001-8489-108X

## REFERENCES

1. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.
2. Gupta RK, Marks M, Samuels THA, et al. Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study. *Eur Respir J*. 2020. https://doi.org/10.1183/13993003.03498-2020
3. 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 Internal Med*. 2020;180:1081.
4. Ji D, Zhang D, Xu J, et al. Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clin Infect Dis*. 2020;71:1393-1399.
5. Lim WS, van der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to

hospital: an international derivation and validation study. *Thorax*. 2003;58:377-382.
6. Guo L, Wei D, Zhang X, et al. Clinical features predicting mortality risk in patients with viral pneumonia: the MuLBSTA score. *Front Microbiol*. 2019;10:2752.
7. Capelastegui A, Quintana JM, Bilbao A, et al. Score to identify the severity of adult patients with influenza A (H1N1) 2009 virus infection at hospital admission. *Eur J Clin Microbiol Infect Dis*. 2012;31:2693-2701.
8. Barnard J, Miscellanea RD. Small-sample degrees of freedom with multiple imputation. *Biometrika*. 1999;86:948-955.
9. Capelastegui A, España PP, Quintana JM, et al. Validation of a predictive rule for the management of community-acquired pneumonia. *Eur Respir J*. 2006;27:151-157.
10. Satici C, Demirkol MA, Sargin Altunok E, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. *Int J Infect Dis*. 2020;98:84-89.
11. 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:e96-e98.
12. Petersen E, Koopmans M, Go U, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis*. 2020;20:e238-e244.
13. Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in covid-19. *N Engl J Med*. 2020;383:120-128.
14. Siddiqi HK, Mehra MR. COVID-19 illness in native and immunosuppressed states: a clinical-therapeutic staging proposal. *J Heart Lung Transplant*. 2020;39:405-407.
15. Grifoni E, Valoriani A, Cei F, et al. The CALL score for predicting outcomes in patients with COVID-19. *Clin Infect Dis*. 2020. https://doi.org/10.1093/cid/ciaa686

## SUPPORTING INFORMATION

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

**How to cite this article:** Shi Y, Pandita A, Hardesty A, et al. Validation of pneumonia prognostic scores in a statewide cohort of hospitalised patients with COVID-19. *Int J Clin Pract*. 2021;75:e13926. https://doi.org/10.1111/ijcp.13926