Validation of clinical risk models for predicting COVID-19 severity

Liang and colleagues developed a risk prediction score, COVID-GRAM, to identify adults with COVID-19 at higher risk of intensive care stay, mechanical ventilation or death.[1] This score had strong performance in Chinese cohorts and has been validated in multiple non-US cohorts, although with variation in its performance (C-statistic ranging from 0.64 to 0.91).[1 2] It has yet to been studied in US populations.[1 2] Differences in the US hospital practices and patient population may affect the applicability of COVID-GRAM to this population. Additionally, clinical rationale and prior studies suggest that CURB-65 may predict severe disease in COVID-19.[3] We compare the performances of COVID-GRAM with CURB-65 for predicting critical illness in patients with COVID-19 in a US population.

This retrospective study included adult patients admitted to an academic medical centre in Boston Massachusetts with a diagnosis of COVID-19 between 1 January 2020 and 29 June 2020. Individuals with prior COVID-19 hospitalisations were excluded. Patients were followed until outcome occurrence or the end of hospitalisation (whichever came first). Demographic and clinical data, patient outcomes and variables used in COVID-GRAM and CURB-65 were obtained from the electronic health record. The primary outcome was critical illness—defined as a composite of mechanical ventilation or death. We used multivariable logistic regression to determine the association between predictors and the outcome of critical illness. Two models of critical illness were constructed: one with all predictors in COVID-GRAM and one with all predictors in CURB-65, with the model coefficients matching those in the original risk scores. The predictive ability of each scoring system was determined using C-statistic, and predictive performance between the two scores was assessed using the de-Long test. Missing data were accounted for using multiple imputations. We used five imputations each to determine the values of the missing data. We chose this approach due to <20% missingness for each predictor. In sensitivity analyses, we repeated the analytical approach but as a complete case analysis. A p value of 0.05 was considered statistically significant. Analyses were conducted with R V.3.5.2.

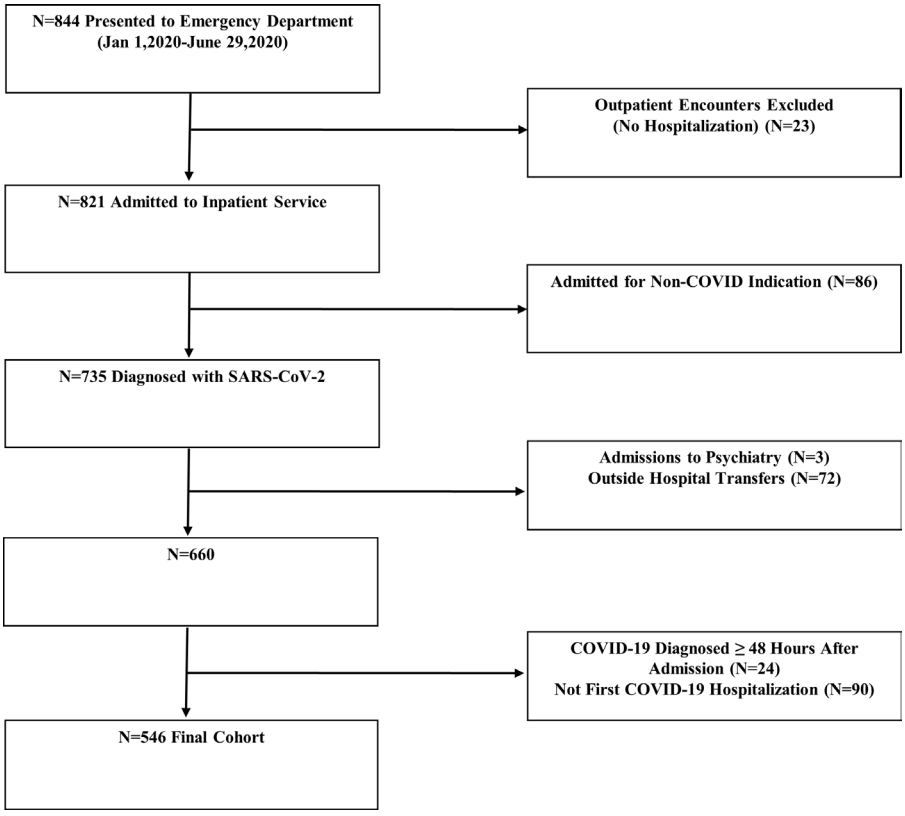

Figure 1 Flow chart of inclusion and exclusion criteria. Presented are the inclusion and exclusion criteria for the study cohort.

The study was approved by the Beth Israel Deaconess Medical Centre institutional review board and determined to be exempt. No patients were directly involved in our study.

Of 844 patients presenting to the ED, 546 were admitted (figure 1). The mean age was 66.8 years (SD: 16.9 years), of whom 48.5% were women (table 1). The primary composite outcome occurred in 170 individuals. Due to missing data for score calculation (131 individuals with missing data for COVID-GRAM calculation and 51 with missing data for CURB-65 calculation), 495 individuals were included in the primary analysis. Increasing score with COVID-GRAM was associated with critical illness (p<0.001). COVID-GRAM had modest discrimination (C-statistic=0.72 (95% CI: 0.67 to 0.76)) (figure 2). CURB-65 score was also associated with critical illness (p<0.001). However, discrimination of CURB-65 (C-statistic: 0.61 (95% CI: 0.56 to 0.66)) was lower than COVID-GRAM (p<0.001). The association of each predictor with the outcome of interest is demonstrated in online supplemental file 1. In our complete case sensitivity analyses, COVID-GRAM (C-statistic: 0.70 (95% CI: 0.65 to 0.75), p<0.001) and CURB-65 (C-statistic: 0.58 (95% CI: 0.53 to 0.64), p=0.003) had similar performance to the primary analysis. Discrimination of CURB-65 was lower than COVID-GRAM (p<0.001).

There are several limitations to our study. Small cohort size may have prevented the identification of certain associations with severe disease. Total bilirubin was used instead of direct bilirubin. Our study was conducted at a single academic medical centre and could not assess the effects of different COVID-19 strains. Because we focused on outcomes occurring during the index hospitalisation, we may have failed to capture rehospitalisation with critical illness.

In this US hospital, COVID-GRAM had modest accuracy in identifying patients who were likely to require mechanical ventilation or expire and substantially outperformed CURB-65. Although COVID-GRAM incorporates predictors routinely obtained in clinical settings and is superior to CURB-65, we believe that the moderate discrimination of COVID-GRAM means that it should not be used in isolation for risk

Table 1 Baseline characteristics by the presence or absence of critical illness

	Critical illness (n=546)	
	No	Yes
Characteristics	**376 (68.9%)**	**170 (31.1%)**
Mean age (SD), years	66.0 (17.3)	68.6 (15.9)
Female	50.70%	41.50%
Race/ethnicity		
White	36.10%	30.00%
Black	33.70%	30.80%
Hispanic	16.30%	13.10%
Other	13.90%	26.20%
Mean length of stay (SD), days	7.7 (7.1)	22.8 (14.4)
Obesity	17.30%	37.50%
Renal failure	32.90%	22.70%
Liver disease	7.20%	7.80%
Presented from nursing home	24.30%	25.40%
Chest X-ray abnormality	67.60%	87.60%
Haemoptysis	2.10%	2.90%
Altered mental status	21.40%	32.50%
Comorbidity count (SD)	4.8 (0.9)	4.9 (0.7)
Cancer history	12.20%	17.10%
Neutrophil to lymphocyte ratio (SD)	5.3 (4.4)	8.7 (8.2)
Lactate dehydrogenase (SD)	357.2 (166.0)	595.6 (1242.2)
Total bilirubin (SD)	0.03 (0.03)	0.03 (0.05)
Uraemia	63.40%	47.50%
Elevated respiratory rate	9.10%	3.60%
Critical blood pressure	18.20%	14.30%
Mean COVID-GRAM score (SD)	7.1 (1.5)	8.2 (3.1)
CURB-65 score		
0	28.80%	19.70%
1	19.50%	31.60%
2	29.80%	23.10%
3	16.90%	17.10%
4	4.70%	8.50%
5	<1%	0%

Critical illness was defined as an individual requiring mechanical ventilation or death. Uraemia was defined as a blood urea nitrogen >19 mg/dL. Elevated respiratory rate was defined as a rate of ≥30 breaths per minute. Critical blood pressure was defined as a systolic blood pressure of <90 mm Hg and/or a diastolic blood pressure ≤60 mm Hg. Chest X-ray abnormalities were based on chart review by medical professionals. All risk factors were collected within 48 hours of presentation to the hospital. Predictors in COVID-GRAM included chest X-ray abnormality, age, haemoptysis, dyspnoea, altered mental status, comorbidity count, cancer history, neutrophil–lymphocyte ratio, lactate dehydrogenase and bilirubin. Predictors in CURB-65 included altered mental status, uraemia, respiratory rate, critical blood pressure and age of 65 years or older.

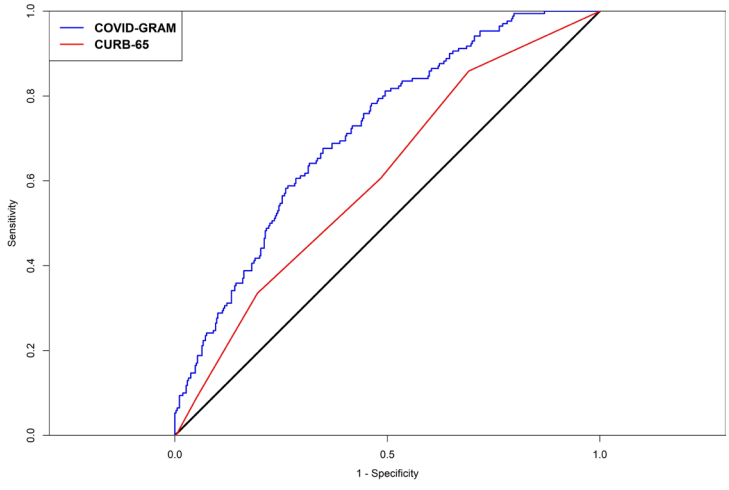

Figure 2 Sensitivity versus 1-specificity for COVID-GRAM and CURB-65. Presented are the C-statistic graphs for the predictive performance of both COVID-GRAM (blue) and CURB-65 (red). X-axis represents 1-specificity, and y-axis represents sensitivity. C-statistic for COVID-GRAM was 0.72 (95% CI: 0.67 to 0.76) and for CURB-65 was 0.61 (95% CI: 0.56 to 0.66).

prediction, but rather as an adjunct to clinical reasoning.

Rahul Aggarwal ,[1] **Timothy S Anderson,**[1,2] **Aditya Mohanty,**[1,2] **Adlin Pinheiro,**[2] **Long Ngo,**[1,2] **Andrew Ahn,**[2] **Neal Peterson,**[2] **Mark Dunlop,**[2] **Thomas Mawson,**[2] **Taliya Lantsman,**[2] **Natalia Forbath,**[2] **Jennifer P Stevens,**[1,2] **Shoshana J Herzig**[1,2]

[1]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[2]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

Correspondence to Dr Shoshana J Herzig, Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; sherzig@bidmc.harvard.edu

Twitter Rahul Aggarwal @RahulAggarwalMD

Contributors RA, TSA, JPS, SJH conceptualised the idea. AM, AP, AA, NP, MD, TM, TL, NF were involved in data collection. RA was involved in data analytics. LN was involved in analytical planning and statistical guidance. JPS and SJH supervised the study. All authors were involved in drafting the manuscript, intellectual design and critical revision of the manuscript for intellectual content.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests TSA discloses consulting fees from Alosa Health.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the Beth Israel Deaconess Medical Centre institutional review board and determined to be exempt.

Provenance and peer review Not commissioned; externally peer reviewed.

Handling editor Richard Body

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/emermed-2021-211821).

Check for updates

To cite Aggarwal R, Anderson TS, Mohanty A, *et al*. *Emerg Med J* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/emermed-2021-211821

Accepted 12 August 2022

Emerg Med J 2022;**0**:1–3. doi:10.1136/emermed-2021-211821

ORCID iD
Rahul Aggarwal http://orcid.org/0000-0002-4746-8511

REFERENCES

1 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 Intern Med* 2020;180:1081–9.

2 Armiñanzas C, Arnaiz de Las Revillas F, Gutiérrez Cuadra M, *et al*. Usefulness of the COVID-GRAM and CURB-65 scores for predicting severity in patients with COVID-19. *Int J Infect Dis* 2021;108:282–8.

3 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;58:377–82.