



OPEN ACCESS

EDITED BY

Soheil Ebrahimpour,
Babol University of Medical Sciences,
Iran

REVIEWED BY

Farzaneh Hamidi,
Tabriz University of Medical Sciences,
Iran

Jesus Ruiz,
Sant Pau Institute for Biomedical
Research, Spain

*CORRESPONDENCE

Sandra Liliانا Valderrama-Beltrán
✉ slvalderrama.med@javeriana.edu.co

RECEIVED 04 November 2025

REVISED 07 January 2026

ACCEPTED 26 January 2026

PUBLISHED 18 February 2026

CITATION

Valderrama-Beltrán SL, Cuervo-Rojas J, Rondón M, Martínez-Vernaza S, De La Hoz Siegler IH, Rodríguez-Morales AJ, Cañas-Arboleda A, Muñoz O, Padilla M, Mesa-Rubio ML, Rojas JA, Bravo Ojeda JS, Villa J, Chacón Sarmiento JA, Patiño S, Tarud Ayub R, Aristizábal C, Rengifo P, Tambini G, Bertagnolio S, Díaz J, Reveiz L and Álvarez-Moreno C (2026) Validation of CURB-65, CRB-65, NEWS2, qSOFA, and 4C scores for predicting mortality in COVID-19 patients across seven emergency Departments in Colombia. *Front. Med.* 13:1738978. doi: 10.3389/fmed.2026.1738978

COPYRIGHT

© 2026 Valderrama-Beltrán, Cuervo-Rojas, Rondón, Martínez-Vernaza, De La Hoz Siegler, Rodríguez-Morales, Cañas-Arboleda, Muñoz, Padilla, Mesa-Rubio, Rojas, Bravo Ojeda, Villa, Chacón Sarmiento, Patiño, Tarud Ayub, Aristizábal, Rengifo, Tambini, Bertagnolio, Díaz, Reveiz and Álvarez-Moreno. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Validation of CURB-65, CRB-65, NEWS2, qSOFA, and 4C scores for predicting mortality in COVID-19 patients across seven emergency Departments in Colombia

Sandra Liliانا Valderrama-Beltrán^{1,2*}, Juliana Cuervo-Rojas³, Martín Rondón³, Samuel Martínez-Vernaza^{4,5}, Ilich Herbert De La Hoz Siegler⁶, Alfonso J. Rodríguez-Morales^{7,8}, Alejandra Cañas-Arboleda^{4,5}, Oscar Muñoz^{4,5}, Mónica Padilla⁹, María Lucía Mesa-Rubio¹⁰, José Antonio Rojas¹¹, Juan Sebastián Bravo Ojeda¹², Jaime Villa¹³, Julio Alberto Chacón Sarmiento¹⁴, Sandra Patiño¹⁵, Roberto Tarud Ayub¹⁶, Claudia Aristizábal¹⁷, Paola Rengifo⁶, Ginna Tambini¹⁸, Silvia Bertagnolio⁹, Janet Díaz⁹, Ludovic Reveiz⁹ and Carlos Álvarez-Moreno¹⁹

¹PhD Program in Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia, ²Division of Infectious Diseases, Department of Internal Medicine, Infectious Diseases Research Group, Hospital Universitario San Ignacio, Bogotá, Colombia, ³Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia, ⁴Department of Internal Medicine, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia, ⁵Department of Internal Medicine, Hospital Universitario San Ignacio, Bogotá, Colombia, ⁶Clinica Colsanitas, Bogotá, Colombia, ⁷Faculty of Health Sciences, Universidad Científica del Sur, Lima, Peru, ⁸Grupo de Investigación Biomedicina, Faculty of Medicine, Fundación Universitaria Autónoma de las Américas-Institución Universitaria Visión de las Américas, Pereira, Colombia, ⁹Pan American Health Organization, Bogotá, Colombia, ¹⁰Universidad de Los Andes, Bogotá, Colombia, ¹¹Fundación Universitaria Sanitas, Clínica Universitaria Colombia, Bogotá, Colombia, ¹²Clínica Infantil Santa María del Lago, Bogotá, Colombia, ¹³Clinica el Carmen, Barranquilla, Colombia, ¹⁴Clinica Reina Sofia, Bogotá, Colombia, ¹⁵Clinica Sebastián de Belalcázar, Cali, Colombia, ¹⁶Clinica Iberoamérica, Barranquilla, Colombia, ¹⁷Unidad de Investigación Clínica, Fundación Universitaria Sanitas, Bogotá, Colombia, ¹⁸Department of Evidence and Intelligence for Action in Health (EIH), Pan American Health Organization, Washington, DC, United States, ¹⁹Clinica Colsanitas and Facultad de Medicina, Universidad Nacional de Colombia, Bogotá, Colombia

Background: Accurate risk stratification is essential for guiding hospitalization decisions in COVID-19. We evaluated the performance of CURB-65, CRB-65, NEWS2, qSOFA, and the 4C Mortality Score in predicting 30-day mortality among patients presenting to the emergency department with COVID-19.

Methods: We conducted an external validation study using an ambispective cohort of COVID-19 patients who presented to emergency departments of seven high-complexity hospitals in Colombia between March 2020 and September 2021. We assessed discrimination using the area under the receiver operating characteristic curve (AUC) and calibration using the GiViTi belt, the observed-to-expected (O/E) ratio, and the calibration intercept and slope. Decision curve analysis and net benefit were used to evaluate clinical utility. The 4C model underwent logistic recalibration.

Results: Among 7,973 patients included, 30-day mortality was 11.3%. The 4C model showed the highest discrimination (AUC 0.71, 95% CI 0.70–0.73) and clinical utility, but poor calibration. NEWS2, CURB-65, CRB-65, and qSOFA performed poorly across all performance metrics. After recalibration, the 4C model achieved an O/E ratio of 1 and showed a modest improvement in discrimination. Decision curve analysis confirmed its utility for guiding hospitalization decisions at a $\geq 4\%$ mortality risk threshold.

Conclusion: The 4C Mortality Score outperformed other models in predicting COVID-19 mortality. Its use in emergency settings alongside clinical judgment can enhance risk stratification, guide hospitalization decisions, and optimize resource allocation. Recalibration and decision analysis are essential for its clinical applicability. Further validation with contemporary data is essential to ensure its transportability across epidemiological settings.

KEYWORDS

COVID-19, emergency medicine, mortality prediction, prognostic models, risk assessment

1 Introduction

Although the COVID-19 pandemic has officially ended, SARS-CoV-2 remains a global public health concern (1). It is hazardous for immunosuppressed individuals and those with comorbidities such as diabetes, cardiovascular disease, chronic kidney disease, and obesity, who face higher risks of severe complications and mortality (2, 3). Despite widespread vaccination, COVID-19 has become endemic, necessitating practical tools to identify patients at high risk of unfavorable outcomes (4).

Given the broad clinical spectrum, accurate triage in emergency settings is crucial. Predictive models assist clinicians in determining which patients require outpatient management versus hospitalization, optimizing resource allocation, and improving outcomes (5, 6). Over 50 prognostic models have been developed or validated for COVID-19 (6, 7), among them CURB-65, NEWS2, qSOFA, and the 4C mortality score, which have been widely studied for assessing prognosis and guiding hospitalization decisions (6, 7).

These models are valued for their simplicity, minimal laboratory requirements, and clinical applicability in conditions such as pneumonia and sepsis (8). The 4C model, developed explicitly for COVID-19, demonstrated adequate validity, discrimination, and calibration in predicting mortality (9). However, despite numerous prognostic models, external validation studies remain limited, and few have comprehensively evaluated calibration performance and clinical utility—particularly through decision curve analysis—to inform real-world clinical decision-making (6, 10). Given the potential for model overfitting and changes in disease severity, circulating variants, and healthcare delivery over time, external validation is essential to assess whether a prediction model maintains its accuracy and clinical utility across diverse populations, settings, and epidemiological conditions. This study aimed to validate the predictive performance of CURB-65, CRB-65, NEWS-2, qSOFA, and 4C in patients with laboratory-confirmed COVID-19 for predicting 30-day mortality following an emergency department visit, using data from the Colombian cohort of the World Health Organization's Global Clinical Platform (11).

2 Methods

2.1 Study design, setting, and participants

This study validates prognostic models using data from an ambispective cohort of COVID-19 patients within the WHO Global Clinical Platform (11), to which Colombia contributed anonymized, individual-level data. The platform uses standardized ISARIC case report forms and a predefined analytic framework, with harmonized variable domains including demographics, baseline comorbidities, clinical presentation at admission, laboratory parameters, treatments, and clinical outcomes (11).

The cohort included patients aged 18 years or older who presented to the emergency department with suspected SARS-CoV-2 infection, later confirmed by RT-PCR or antigen testing, between March 2020 and September 2021. Data were collected from seven high-complexity hospitals: four in Bogotá, two in Barranquilla, and one in Cali.

Patients were included if they had confirmed SARS-CoV-2 infection and at least one COVID-19-related symptom (see [Supplementary material M1](#)). Exclusion criteria included transfer to another institution within 24 h or teleconsultation. To avoid duplicate observations, only the first emergency department visit per patient during the study period was included in the analysis. Cohort entry began at the time of the first emergency department consultation, and all patients were followed for up to 30 days after admission.

Model validation followed the TRIPOD (transparent reporting of a multivariable prediction model for individual prediction or diagnosis) guidelines (12, 13).

2.2 Variables

The variables were those originally included in each model. They encompassed demographic variables, comorbidities (modified Charlson Comorbidity Index) (14), vital signs at admission, clinical parameters (use of supplemental oxygen, and presence of confusion), and admission laboratory tests (urea, C-reactive protein). The recorded comorbidities were chronic cardiac disease, chronic respiratory disease

(excluding asthma), chronic renal disease (estimated glomerular filtration rate ≤ 30), mild to severe liver disease, dementia, chronic neurological conditions, connective tissue disease, diabetes mellitus, HIV/AIDS, malignancy, and clinician-defined obesity.

The primary outcome was 30-day mortality, defined as death from any cause within 30 days from the emergency department visit. For discharged patients, survival status was confirmed via the Registro Único de Afiliados (RUIAF) of the Colombian Ministry of Health (15). A blinded investigator verified death certificates. Additional data on hospitalization, intensive care unit (ICU) admission, and hospital stay length were also collected.

All seven participating hospitals used electronic health record (EHR) systems, which allowed standardized extraction of clinical variables using definitions aligned with the WHO/ISARIC case report forms (11, 16). Data entry followed the standardized procedures of the WHO Global Clinical Platform and was performed either through direct entry into the platform or via upload of a harmonized WHO Excel template. Laboratory data were transferred electronically from institutional laboratory information systems to the ISARIC database in all centers, ensuring uniform units of measurement and minimizing transcription errors. Data collection was performed by trained physician-investigators at each site. To ensure data quality and completeness, 10% of the sample was independently reviewed in duplicate. In addition, two investigators conducted monthly reviews of the Colombian National Public Health Surveillance System to verify consecutive patient inclusion and data accuracy.

2.3 Sample size

Sample size calculation followed the Riley et al. methodology (17), using the 4C model ($n = 20$ parameters, C-statistic 0.78) and an estimated 11.2% mortality rate observed in a previous study based on the Colombian WHO Global Clinical Platform cohort (18). The minimum required sample size was 1,624 patients with 82 mortality events. Given our cohort of 7,973 patients and 901 deaths, the sample was sufficient for model validation.

Because the 4C model includes the largest number of predictors and represents the most statistically demanding validation scenario, meeting its sample size requirements also ensured adequate sample size for the validation of CURB-65, NEWS2, and qSOFA models.

2.4 Statistical analysis

Exploratory data analysis identified outliers, entry errors, and missing values, and adjustments were made as needed. No missing values were found in demographic or clinical variables. However, C-reactive protein and urea were missing in 2.0% ($n = 159$) and 2.7% ($n = 216$) of patients, respectively. The only model that incorporated both variables was the 4C, with 4.6% ($n = 369$) missing data. Since this percentage was below 5%, a complete-case analysis was conducted. There was no missing data for the outcome variable.

A descriptive analysis was conducted, summarizing continuous variables as medians and interquartile ranges (IQRs), and categorical variables as absolute and relative frequencies. The Kaplan–Meier method was used to estimate overall survival, with comparisons across predefined risk groups performed using log-rank tests for each model. Administrative censoring was applied at 30 days post-admission.

For model validation, mortality predictions were generated using CURB-65, CRB-65, NEWS2, qSOFA, and 4C models (see

Supplementary material M2), applying the original scoring criteria (9, 19–21). Discrimination was assessed using the receiver operating characteristic (ROC) curves and the area under the curve (AUC) with a 95% confidence interval (95% CI), with AUC > 0.7 considered acceptable (13).

Calibration was evaluated using the GiViTI calibration belt. This method incorporates a confidence band to quantify prediction uncertainty and employs a likelihood ratio test, and to capture non-linear deviations from perfect calibration (22). Additionally, observed-to-expected (predicted) events were analyzed by risk group to provide a detailed performance assessment.

In addition, for the 4C and NEWS2 models, we conducted analyses to evaluate potential changes in model performance across key phases of the pandemic in Colombia. Discrimination and calibration were assessed separately for two relevant time periods: (a) before and after the implementation of corticosteroid therapy in routine clinical practice (July 2020) (23), and (b) before and after the launch of the national COVID-19 vaccination program in Colombia (April 2021).

For the 4C model, which provided intercept and regression coefficients (9). We calculated additional overall performance measures such as the Cox-Snell R-squared, Brier Score, and calibration metrics such as the calibration intercept and calibration slope.

Despite its good discriminatory performance, the 4C model exhibited poor calibration. As initial recalibration-in-the-large (adjusting only the intercept) was insufficient to correct this miscalibration. Therefore, logistic recalibration, involving adjustments to both the intercept and slope, was performed, resulting in improved discrimination and calibration. The choice of recalibration method was supported by a likelihood ratio test, which identified this approach as the most conservative and statistically appropriate (24, 25).

Finally, to assess the clinical utility of the 4C model, we conducted decision curve analysis (DCA) and evaluated net benefit, which balances intervention benefits against potential harms. To determine an appropriate risk threshold for death to guide hospitalization decisions for COVID-19, we selected a risk threshold range between 2 and 10% based on expert clinical consultation and previous prediction models (9, 19–21). The recalibrated 4C model was compared to the original model, and the strategies of hospitalizing all or none (10, 26).

Data analysis was performed in R v4.1 using the packages pmsampsize, predRupdate, pROC, givitiR, rms, and rmda.

2.5 Ethical aspects

The study was approved by the Research and Ethics Committee of HUSI and Pontificia Universidad Javeriana (FM-CIEI-0686-21). The Institutional Review Board (IRB) approved each participating institution's participation in the WHO Global Clinical Platform (11), whose data serve as the basis for this study. Informed consent was waived due to the retrospective design. Data were anonymized, with access restricted to designated investigators.

3 Results

3.1 Population characteristics

A total of 7,973 out of 9,050 patients from the Colombian cohort of the WHO Global Clinical Platform for COVID-19 met the

inclusion criteria (see [Supplementary material F1](#)). In the study population, 58.3% were men, and the median age was 60.7 years (IQR: 49–80). At 30 days, the median age was 59.6 years (IQR: 48.3–80.1) and 70.6 years (IQR: 60.1–81.6) among survivors and non-survivors, respectively.

The most frequent underlying condition was hypertension (29.5%, $n = 2,353$), followed by obesity (27.6%, $n = 2,202$) and diabetes mellitus (16.9%, $n = 1,346$). Among non-survivors, having two or more comorbidities was common (39.3%, $n = 354$) (see [Table 1](#)). Overall,

95.7% ($n = 7,627$) of patients required hospitalization for more than 24 h, whereas 13.4% ($n = 1,072$) were admitted to the ICU. The median hospital stay was 6 days (IQR: 3.2–11.4).

At 30 days post-admission to the emergency department, the cumulative incidence of death was 11.3% ($n = 901$). The Kaplan–Meier survival curves, stratified by NEWS2, CURB-65, qSOFA, and 4C score categories, demonstrate significant differences in survival probability across risk groups (log-rank test, $p < 0.0001$). For all models, patients in the highest-risk category experience a decline in

TABLE 1 Characteristics of a cohort of patients with SARS-CoV-2 infection presenting to seven high-complexity emergency departments in Colombia between March 2020 and September 2021.

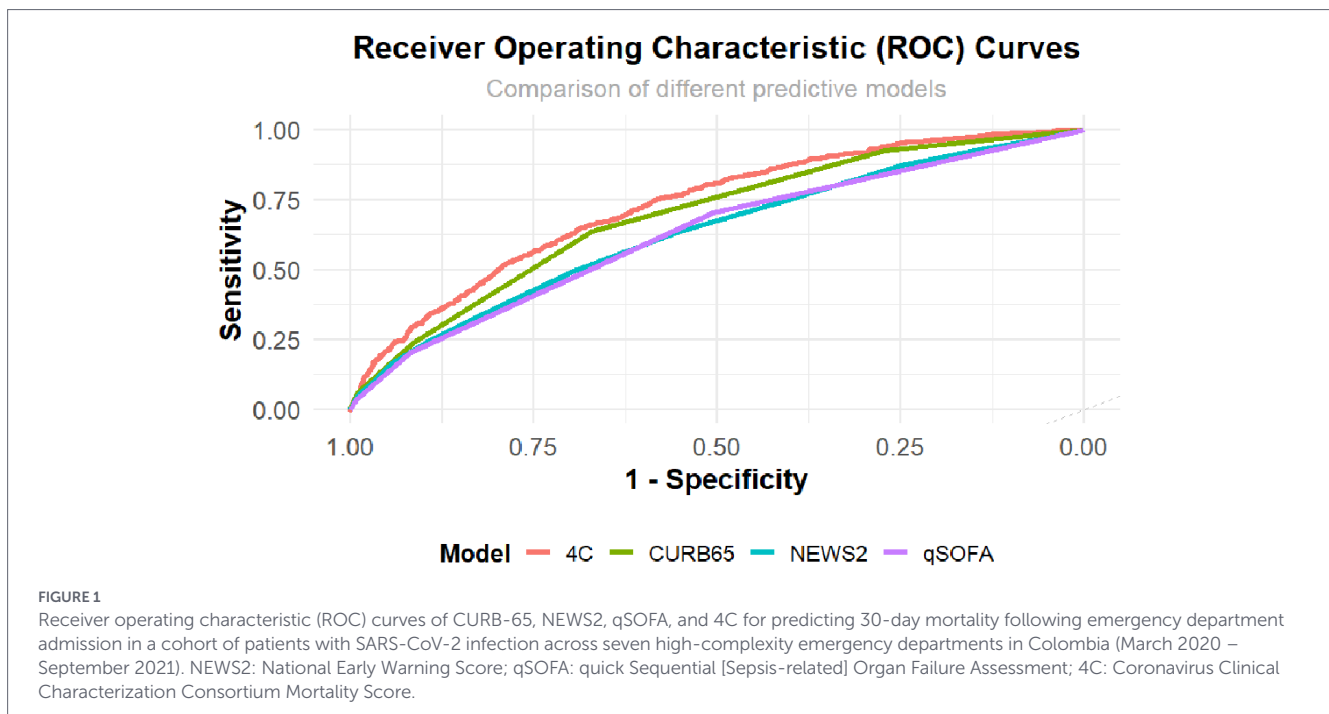
Variable	Total (N = 7,973) ¹	Survivor (N = 7,072) ¹	Non-survivor (N = 901) ¹
Age (years), n (%)			
<50	2,118 (26.6%)	2,007 (28.4%)	111 (12.3%)
50–59	1,716 (21.5%)	1,595 (22.6%)	121 (13.4%)
60–69	1,865 (23.4%)	1,663 (23.5%)	202 (22.4%)
70–79	1,286 (16.1%)	1,056 (14.9%)	230 (25.5%)
≥80	988 (12.4%)	751 (10.6%)	237 (26.3%)
Sex at birth, n (%)			
Male	4,652 (58.3%)	4,102 (58.0%)	550 (61.0%)
Female	3,321 (41.7%)	2,970 (42.0%)	351 (39.0%)
Temperature (°C) ² , median (IQR ³)	36.5 (36.0–37.0)	36.5 (35.5–37.5)	36.5 (35.5–37.5)
Temperature (°C)², n (%)			
≤35	91 (1.1%)	73 (1.0%)	18 (2.0%)
35.1–36	2,007 (25.2%)	1,777 (25.1%)	230 (25.5%)
36.1–38	5,300 (66.5%)	4,712 (66.6%)	588 (65.3%)
38.1–39	482 (6.0%)	429 (6.1%)	53 (5.9%)
>39	93 (1.2%)	81 (1.1%)	12 (1.3%)
Heart rate ⁴ , median (IQR ³)	85 (76–99)	85 (61–109)	86 (63–109)
Heart rate⁴, n (%)			
≤40	11 (0.1%)	3 (0.0%)	8 (0.9%)
41–50	48 (0.6%)	44 (0.6%)	4 (0.4%)
51–90	4,915 (61.6%)	4,364 (61.7%)	551 (61.2%)
91–110	2,012 (25.2%)	1,796 (25.4%)	216 (24.0%)
111–130	861 (10.8%)	770 (10.9%)	91 (10.1%)
>130	126 (1.6%)	95 (1.3%)	31 (3.4%)
Respiratory rate ⁵ , median (IQR ³)	20 (18–22)	20 (16–24)	21 (15–27)
Respiratory rate⁵, n (%)			
<22	5,288 (66.3%)	4,825 (68.2%)	463 (51.4%)
≥22	2,685 (33.7%)	2,247 (31.8%)	438 (48.6%)
Systolic blood pressure ⁶ , median (IQR ³)	123 (112–135)	123 (100–146)	124 (98–152)
Systolic blood pressure⁶, n (%)			
<90	204 (2.6%)	149 (2.1%)	55 (6.1%)
≥90	7,769 (97.4%)	6,923 (97.9%)	846 (93.9%)
Diastolic blood pressure ⁶ , median (IQR ³)	75 (69–75)	75 (62–88)	73 (57–89)
Diastolic blood pressure⁶, n (%)			
≤60	1,002 (12.6%)	821 (11.6%)	181 (20.1%)
>60	6,971 (87.4%)	6,251 (88.4%)	720 (79.9%)

(Continued)

TABLE 1 (Continued)

Variable	Total (N = 7,973) ¹	Survivor (N = 7,072) ¹	Non-survivor (N = 901) ¹
Glasgow Coma Scale ⁷ , median (IQR ³)	15 (15–15)	15 (15–15)	15 (14–16)
Glasgow Coma Scale⁷, n (%)			
<15	1,500 (18.8%)	1,251 (17.7%)	249 (27.6%)
15	6,473 (81.2%)	5,821 (82.3%)	652 (72.4%)
Oxygen saturation ⁸ , median (IQR ³)	92 (88–94)	92 (89–95)	90 (85–94)
Oxygen saturation⁸, n (%)			
≤91%	3,852 (48.3%)	3,330 (47.1%)	522 (57.9%)
92–93%	1,502 (18.8%)	1,353 (19.1%)	149 (16.5%)
94–95%	146 (15.6%)	1,129 (16.0%)	117 (13.0%)
≥96%	1,373 (17.2%)	1,260 (17.8%)	113 (12.5%)
Confusion ⁹ , n (%)	1,500 (18.8%)	1,251 (17.7%)	249 (27.6%)
Underlying conditions¹⁰, n (%)			
Hypertension	2,353 (29.5%)	2,008 (28.4%)	345 (38.3%)
Chronic cardiac disease	280 (3.5%)	226 (3.2%)	54 (6.0%)
Diabetes mellitus	1,346 (16.9%)	1,145 (16.2%)	201 (22.3%)
Chronic renal disease	1,284 (16.1%)	1,067 (15.1%)	217 (24.1%)
Clinician-defined obesity	2,202 (27.6%)	1,974 (27.9%)	228 (25.3%)
Chronic respiratory disease	1,085 (13.6%)	923 (13.1%)	162 (18.0%)
Asthma	118 (1.5%)	101 (1.4%)	17 (1.9%)
Connective tissue disease	171 (2.1%)	143 (2.0%)	28 (3.1%)
Liver disease	259 (3.2%)	220 (3.1%)	39 (4.3%)
Chronic neurological conditions	150 (1.9%)	112 (1.6%)	38 (4.2%)
Malignancy	1,125 (14.1%)	941 (13.3%)	184 (20.4%)
HIV/AIDS ¹¹	212 (2.7%)	198 (2.8%)	14 (1.6%)
Underlying conditions number, n (%)			
0	4,593 (57.6%)	4,163 (58.9%)	430 (47.7%)
1	952 (11.9%)	835 (11.8%)	117 (13.0%)
2 or more	2,428 (30.5%)	2,074 (29.3%)	354 (39.3%)
C-reactive protein, mg/L (n = 7,757), median (IQR ³)	69.2 (20.4–146.2)	66.5 (37.6–95.3)	102.1 (57.5–146.6)
C-reactive protein (n = 7,757), n (%)			
<50 mg/L	3,161 (40.8%)	2,905 (42.0%)	256 (30.2%)
50–99 mg/L	1,692 (21.8%)	1,529 (22.1%)	163 (19.2%)
≥100 mg/L	2,904 (37.4%)	2,475 (35.8%)	429 (50.6%)
Urea, mmol/L (n = 7,816), median (IQR ³)	6.8 (5.5–8.2)	6.7 (5.5–8.0)	7.7 (5.5–9.8)
Urea (n = 7,816), n (%)			
≤7 mmol/L	4,073 (52.1%)	3,731 (53.8%)	342 (38.7%)
>7 mmol/L	3,743 (47.9%)	3,201 (46.2%)	542 (61.3%)
Hospitalization, n (%)	7,627 (95.7%)	6,771 (95.7%)	856 (95.0%)
Dexamethasone use ¹² , n (%)	6,933 (87.0%)	6,173 (87.3%)	760 (84.3%)
ICU admission ¹³ , n (%)	1,072 (13.4%)	675 (9.5%)	397 (44.1%)
Length of hospital stay, median (IQR ³)	6.0 (3.2–11.4)	5.6 (2.0–9.2)	11.0 (4.8–17)

1. Column-based percentages. 2. Body temperature in degrees Celsius (°C) at admission. 3. IQR: interquartile range. 4. Heart rate in beats per minute (bpm) at admission. 5. Respiratory rate in breaths per minute (bpm) at admission. 6. Blood pressure in millimeters of mercury (mmHg) at admission. 7. Glasgow Coma Scale score at admission. 8. Oxygen saturation at admission. 9. Confusion at admission. 10. Patients may have multiple underlying conditions; thus, categories are not mutually exclusive. 11. Human immunodeficiency virus (HIV) infection. 12. Dexamethasone use was defined according to the RECOVERY trial protocol for COVID-19, which recommends a dose of 6 mg once daily for up to 10 days in hospitalized patients requiring oxygen therapy or an equivalent dose of another corticosteroid (34). 13. ICU: Intensive care unit.



survival within the first days of follow-up. In contrast, those in the lowest-risk category maintain high survival probabilities over time (see [Supplementary material F2](#)).

3.2 Performance of prognostic models

The discriminative performance of the models was assessed using the AUC, yielding values of 0.69 (95% CI: 0.67–0.70) for CURB-65, 0.67 (95% CI: 0.66–0.69) for CRB-65, 0.63 (95% CI: 0.61–0.65) for NEWS2, 0.62 (95% CI: 0.61–0.64) for qSOFA, and 0.71 (95% CI: 0.70–0.73) for the 4C mortality score. [Figure 1](#) presents the ROC curves for the evaluated models.

In terms of calibration, the GiViTI calibration belt analysis indicated that none of the models achieved adequate calibration ([Figure 2](#)). Additionally, the likelihood ratio test ($p < 0.05$) rejected the null hypothesis of no differences between predicted and observed probabilities for all models.

An observed-to-expected event plot stratified by predefined risk groups was also generated ([Figure 2](#)). The results showed that the CURB-65 and CRB-65 models underestimated mortality risk across all predefined risk groups. The NEWS2 model underestimated mortality risk in the low-to-intermediate risk groups. The qSOFA model underestimated the risk in the low-risk group. Similarly, the 4C model underestimated mortality risk in the lower-risk groups and consistently overestimated it in the remaining groups. Detailed information on the comparison of observed and predicted event probabilities for all models is available in [Supplementary Table 1R](#).

When comparing the performance of the 4C and NEWS2 models before and after the introduction of corticosteroid therapy, or before and after the implementation of COVID-19 vaccination in Colombia, we did not observe clinically meaningful differences in overall discriminative ability. For the NEWS2, no significant differences in calibration were found between the periods. However, the 4C model did show better calibration during the period prior to than

during the period following the introduction of corticosteroids. Once there was an effective therapy, the model overestimated the risk of death, especially among those at high risk. Detailed performance metrics for each period are presented in [Supplementary Figures F3–F6](#).

The overall performance of the 4C model is presented in [Table 2](#). Given that the 4C model showed the best discrimination among the evaluated models, we recalibrated it and reassessed its performance.

3.3 Recalibrated 4C model

The recalibrated 4C model is of the form: $\text{New model} = \text{intercept}_{\text{new}} (-1.049) + \text{overall slope} (0.781) * \text{original 4C model logit}$. This was used to adjust the coefficients of all variables in the logistic model. The original and recalibrated coefficients, along with the performance of the 4C Mortality Score, are presented in [Table 2](#). An online calculator based on the recalibrated model is available at: <https://github.com/Mars1971/Recalibrated-4C-Mortality-Score-for-COVID-19-in-Colombia>.

The updated model achieved an observed-to-expected event ratio of 1.00 (95% CI, 0.94–1.07), an intercept of 0 (95% CI, -0.07 to 0.07), and a calibration slope of 1 (95% CI, 0.91–1.09) ([Table 2](#)). Calibration was also confirmed via a calibration belt plot, and a non-significant likelihood ratio test ($p = 0.644$), indicating no meaningful difference between observed and predicted events ([Supplementary Figures F7–F9](#)). Discrimination ability slightly improved (DeLong's test < 0.001), as did overall model performance.

3.4 Clinical utility

The original and recalibrated 4C models demonstrated a net benefit for guiding hospitalization decisions at a mortality risk threshold of approximately 4% and maintained this benefit across the proposed threshold range. At a 4% mortality risk threshold, the cost-benefit

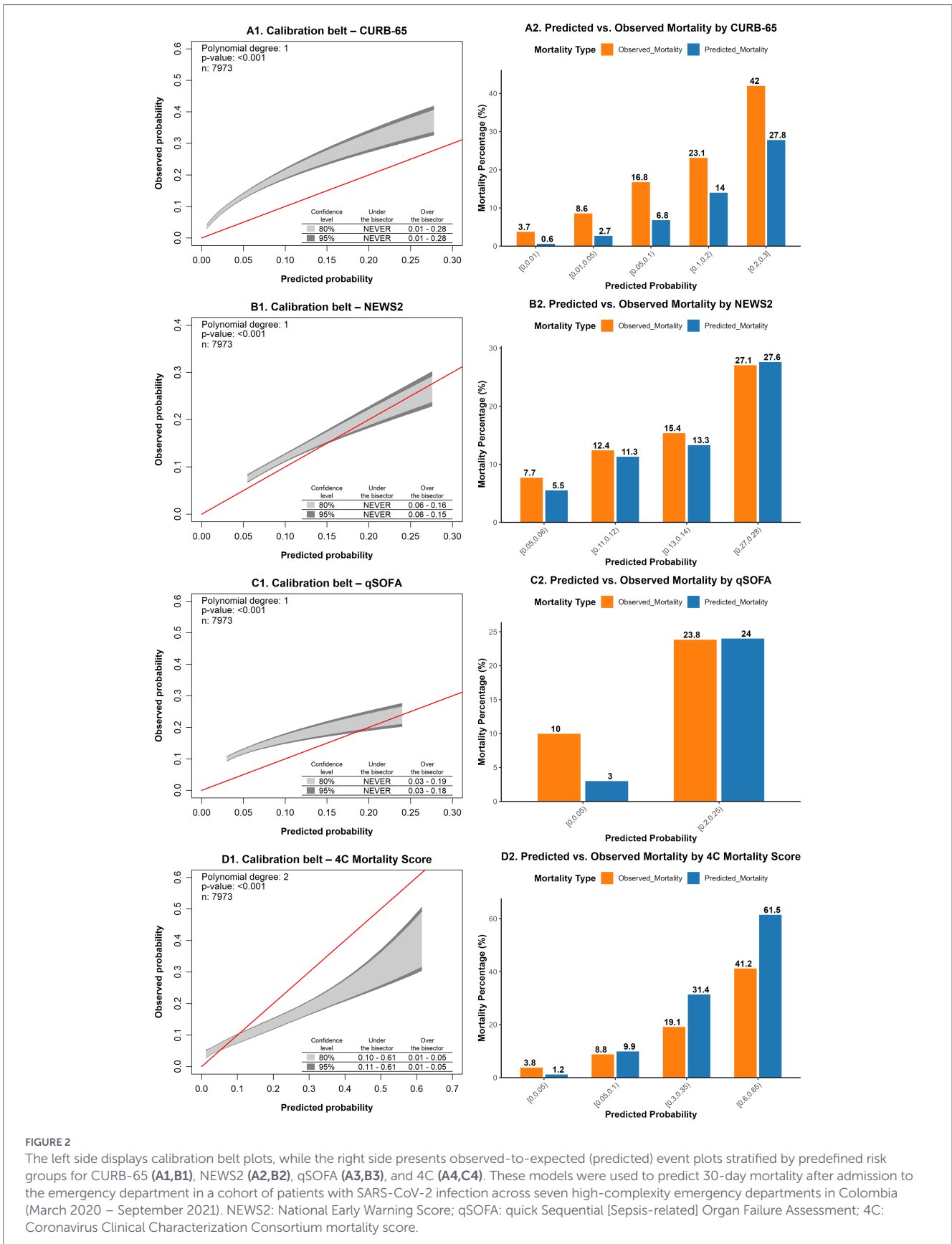


FIGURE 2

The left side displays calibration belt plots, while the right side presents observed-to-expected (predicted) event plots stratified by predefined risk groups for CURB-65 (A1,B1), NEWS2 (A2,B2), qSOFA (A3,B3), and 4C (A4,C4). These models were used to predict 30-day mortality after admission to the emergency department in a cohort of patients with SARS-CoV-2 infection across seven high-complexity emergency departments in Colombia (March 2020 – September 2021). NEWS2: National Early Warning Score; qSOFA: quick Sequential [Sepsis-related] Organ Failure Assessment; 4C: Coronavirus Clinical Characterization Consortium mortality score.

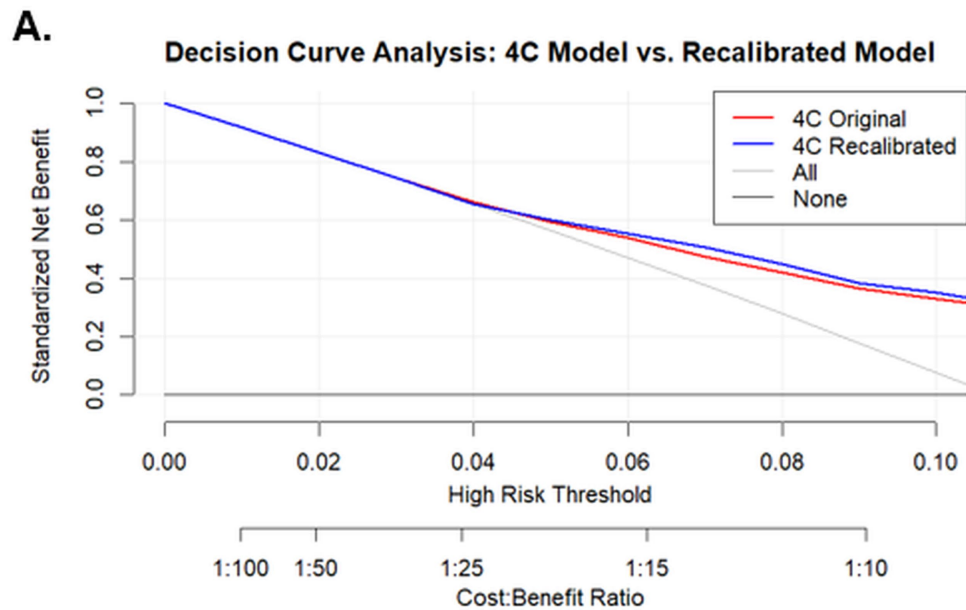
ratio was 1:25, indicating that accepting approximately 25 potentially unnecessary hospital admissions may be justified to avoid missing one patient at high risk of death (Figure 3). A detailed table presenting net

benefit values across different mortality risk thresholds is provided in Figure 3. The curves with confidence intervals are shown in Supplementary Figure F10.

TABLE 2 Regression coefficients and predictive performance of the original and recalibrated 4C prediction models for 30-day mortality after emergency department admission in patients with SARS-CoV-2 infection.

	Original 4C mortality model ¹	Recalibrated 4C mortality model ²
Model coefficients		
Intercept	−4.203	−4.047
Age (years)		
<50 (Ref) ³	−	−
50–59	0.687	0.536
60–69	1.337	1.044
70–79	1.842	1.438
≥80	2.252	1.758
Sex at birth		
Female (Ref) ³	−	−
Male	0.172	0.134
Number of comorbidities		
0 (Ref) ³	−	−
1	0.3	0.234
≥2	0.532	0.415
Respiratory rate⁴		
<20 (Ref) ³	−	−
20–29	0.232	0.181
≥30	0.649	0.507
Peripheral oxygen saturation⁵		
≥92% (Ref) ³	−	−
<92%	0.577	0.450
Glasgow coma scale <15		
15 (Ref) ³	−	−
<15	0.558	0.436
Urea (mmol/L)		
<7 (Ref) ³	−	−
7–14	0.439	0.343
>14	1.011	0.789
C-reactive protein (mg/L)		
<50 (Ref) ³	−	−
50–99	0.363	0.283
≥100	0.747	0.578
Performance measures⁶		
Calibration (95% CI) ⁷		
Observed/expected ratio	0.57 (0.54–0.61)	1.00 (0.94–1.07)
Calibration intercept	−0.77 (−0.85–0.70)	0.00 (−0.07–0.07)
Calibration slope	0.77 (0.70–0.84)	1.00 (0.91–1.09)
Discrimination (95% CI) ⁷		
C-statistic (AUROC) ⁸	0.71 (0.70–0.73)	0.73 (0.71–0.75)
Overall model fit		
Nagelkerke R ²	0.06	0.23
Cox-Snell R ²	0.03	0.16
Brier Score (95% CI) ⁷	0.16 (0.15–0.16)	0.09 (0.08–0.09)

1. The coefficients of the original model correspond to those published by the developers of the 4C model in Knight SR et al. (9). 2. The coefficients of the recalibrated model are based on logistic recalibration. 3. Ref. reference category in the regression model. 4. Respiratory rate (breaths/min). 5. Oxygen saturation on air room (%). 6. Evaluation of the performance in predicting 30-day mortality after admission to the emergency department in a cohort of patients with SARS-CoV-2 infection across seven high-complexity emergency departments in Colombia (March 2020 – September 2021). 7. 95% CI: 95% confidence interval. 8. AUROC: area under the receiver operating characteristic curve.



B.

Pt (%)	Risk threshold	Net Benefit		Advantage of Model	
		Treat all	Prediction Model	Net Benefit	Cost:Benefit Ratio
1	0.01	0.916	0.916	0	1:100
2	0.04	0.629	0.671	0.042	1:25
3	0.06	0.471	0.555	0.084	1:15
4	0.11	-0.025	0.307	0.332	1:9
5	0.16	-0.58	0.164	0.744	1:5
6	0.21	-1.204	0.092	1.296	1:4
7	0.26	-1.914	0.072	1.986	1:3
8	0.31	-2.726	0.04	2.766	1:2
9	0.36	-3.665	0.021	3.686	1:1.7
10	0.41	-4.763	-0.004	4.759	1:1.4

FIGURE 3 Decision curve analysis (A) and net benefit table (B) comparing the original and recalibrated 4C mortality scores with both the “all” and “none” hospitalization approaches in a cohort of patients with SARS-CoV-2 infection across seven high-complexity emergency departments in Colombia (March 2020 – September 2021). In the decision curve analysis (A), the x-axis represents the selected probability of death threshold for deciding hospitalization, and its corresponding cost–benefit ratio, while the y-axis displays the standardized net benefit.

4 Discussion

We conducted a comprehensive evaluation of five models used in clinical practice to predict COVID-19 mortality. Among these, the 4C Mortality Score demonstrated the best overall performance, with acceptable discriminative ability, adequate clinical utility, and initially poor calibration, especially after the introduction of corticosteroid therapy. In contrast, in our population, CURB-65, CRB-65, NEWS2, and qSOFA performed poorly in both

discrimination and calibration, limiting their clinical usefulness in our setting.

In our cohort, the 4C model outperformed the other evaluated models, exhibiting the highest discriminative ability, consistent with findings from other validation studies (27–31). However, our validation showed a slightly lower AUC than the original study (9).

Regarding calibration, we observed an overestimation of mortality risk at intermediate and high scores and an underestimation at lower scores—a pattern also reported in previous studies (32, 33). These

differences may be explained by temporal changes in clinical management and population-level risk over the course of the pandemic. The 4C model was derived during the early pandemic, a period characterized by higher baseline mortality and limited availability of effective therapies, before the routine use of corticosteroids in hypoxemic patients and the implementation of large-scale vaccination programs (34, 35).

In addition to corticosteroid use, other time-varying, treatment-related factors may have contributed to shifts in observed mortality risk, including evolving oxygen therapy protocols, selective use of antivirals, changes in ICU capacity and resource availability during regional surges, and the circulation of different SARS-CoV-2 variants. This is further supported by our stratified analysis of model performance before and after the introduction of corticosteroid therapy, in which the 4C model demonstrated better calibration during the period prior to its introduction (see [Supplementary material F6](#)), underscoring the impact of changing treatment contexts on model performance.

Despite the calibration being suboptimal, the 4C model remains a valuable tool, as it effectively stratifies patients based on disease severity, enabling the identification of those at higher or lower risk of mortality (7). To address this issue, we recalibrated the model for our cohort, which is mainly composed of patients receiving corticosteroid therapy, achieving near-perfect calibration while preserving discriminative ability. The recalibrated model demonstrated clinical utility in guiding in-hospital monitorization or hospitalization decisions, particularly at a mortality probability threshold of 4% or higher.

Our validation showed that the NEWS-2 model exhibits inadequate discriminative ability and underestimated mortality risk at low scores, as observed in multiple studies (9, 28, 30, 36). We believe these performance limitations stem from the exclusion of key variables such as age, comorbidities, and critical laboratory parameters—a design choice intended to facilitate rapid bedside assessment—all of which have been previously identified as significant risk factors for unfavorable outcomes (3, 37, 38).

We found that both CURB-65 and CRB-65 consistently underestimated mortality risk across all score values, a limitation previously reported in two extensive validation studies (9, 28). These tools, developed initially for community-acquired pneumonia (CAP), may not be suitable for COVID-19-related CAP, which often presents with silent hypoxemia and may rapidly progress to acute respiratory distress syndrome (39, 40). In line with current American Thoracic Society and the Infectious Diseases Society of America ATS/IDSA guidelines (41), CURB-65 and CRB-65 are no longer recommended for guiding hospitalization decisions or level of care in CAP.

Similarly, our data suggest that qSOFA may underestimate mortality risk, particularly at low scores that classify patients as not at high risk of in-hospital mortality. We observed that nearly 10% of patients who died had a qSOFA score of 0 or 1 upon admission. This limited predictive accuracy has also been reported in other validation studies (7, 9, 40, 42) and may be attributed to the phenomenon of silent hypoxemia, as previously described (39, 40). Therefore, we consider that qSOFA should not be used as a triage tool in patients with suspected COVID-19-related CAP.

To our knowledge, this study represents one of the largest multi-center external validation efforts of prognostic models for COVID-19 mortality conducted in Latin America to date. Although COVID-19 mortality has declined, the disease has become endemic, making the

identification of optimal triage strategies essential for clinical decision-making in the emergency setting, as these patients will continue to present for care.

To improve the calibration of the 4C model, it was necessary to adjust both the intercept and the calibration slope. This was possible because the 4C model was developed with methodological rigor, and its original intercept and beta coefficients were fully reported (9, 10). In contrast, recalibration of the NEWS2, CURB-65, and qSOFA models is limited, as their coefficients are not publicly available.

Furthermore, recalibration is a recommended strategy to enhance model performance and clinical usefulness across diverse settings (32). In our study, this process optimized the predictive accuracy of the 4C model in a high-complexity hospital care context in Latin America. To support implementation in similar settings, we developed an online calculator incorporating the recalibrated 4C Mortality Score: <https://github.com/Mars1971/Recalibrated-4C-Mortality-Score-for-COVID-19-in-Colombia>.

Nonetheless, our study has certain limitations. First, the validation was conducted during a phase of the pandemic when most hospitalized patients had not yet received the COVID-19 vaccination, and before the widespread circulation of the Omicron variant (35). These factors may have influenced the model's performance and limited the generalizability of our findings to current clinical settings.

However, our results are consistent with more recent validation studies of the 4C model, which have demonstrated stable predictive performance and support its use in clinical practice (6, 31), further reinforcing the robustness of our findings. Moreover, many resource-limited settings continue to face limited access to vaccination and effective therapies, underscoring the continued relevance of our findings in such contexts.

To further strengthen the model's clinical applicability, temporal validation and, if necessary, recalibration using contemporary data are essential, especially in the context of evolving epidemiological conditions. Incorporating individual-level data on vaccination status and therapeutic interventions may also enhance its accuracy and relevance for ongoing and future use in diverse clinical settings.

A second limitation is that our study was conducted in high-complexity emergency care settings; thus, to use the model in lower-complexity healthcare settings, further validation may be required. Third, we did not include in this validation study other models developed for COVID-19 or for CAP, such as the CCEDRRN COVID-19 Mortality Score, the SEIMC score, and PSI, because some of them were not thoroughly developed and validated, or available, or usable in our context, given the need for laboratory assessments not readily available in our emergency settings (6).

Despite its clinical utility, the 4C model is not exempt from the risk of misclassification. When applying a low-risk threshold ($\geq 4\%$) to guide hospitalization decision, some patients with low predicted mortality may be admitted unnecessarily. In these cases, clinical reassessment during hospitalization is essential to ensure timely discharge and avoid unnecessary use of healthcare resources. Conversely, patients classified as low risk ($< 4\%$) may still present atypical signs of severity not captured by the model or may deteriorate rapidly after emergency department discharge. Given the

potential consequences of false-negative risk assessments, the 4C model should always be used alongside clinical judgment and appropriate follow-up to minimize decision-making errors and ensure patient safety.

In conclusion, the 4C Mortality Score outperformed NEWS2, CRB-65, CURB-65, and qSOFA in predicting 30-day mortality among patients presenting to the emergency department with suspected COVID-19. Given its ability to stratify patients into lower- and higher-risk-of-death categories, we recommend its use as a complementary tool to clinical judgment.

Integrating recalibrated 4C models into routine clinical practice may support more accurate triage and better resource allocation. However, it should not be used as a standalone criterion. Clinicians must remain aware of its limitations, apply it within a comprehensive and individualized clinical assessment, and ensure it is adapted to current clinical and epidemiological conditions through recalibration when necessary.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Research and Ethics Committee of HUSI and Pontificia Universidad Javeriana (FM-CIEI-0686-21). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

SV-B: Data curation, Visualization, Resources, Conceptualization, Writing – review & editing, Writing – original draft, Investigation, Project administration, Validation, Methodology, Funding acquisition, Formal analysis, Supervision. JC-R: Writing – review & editing, Methodology, Formal analysis, Validation, Supervision, Investigation, Writing – original draft, Conceptualization, Visualization. MR: Visualization, Formal analysis, Writing – original draft, Conceptualization, Methodology, Investigation, Writing – review & editing, Validation, Supervision. SM-V: Formal analysis, Writing – original draft, Writing – review & editing. ID: Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Supervision, Data curation. AR-M: Writing – original draft, Supervision, Formal analysis, Writing – review & editing. AC-A: Writing – review & editing, Formal analysis, Writing – original draft, Data curation. OM: Formal analysis, Data curation, Writing – original draft, Writing – review & editing. MP: Formal analysis, Writing – review & editing, Writing – original draft. MM-R: Formal analysis,

Writing – review & editing, Writing – original draft. JR: Writing – review & editing, Writing – original draft, Formal analysis, Data curation. JB: Formal analysis, Writing – original draft, Data curation, Writing – review & editing. JV: Writing – original draft, Formal analysis, Writing – review & editing, Data curation. JC: Writing – review & editing, Formal analysis, Writing – original draft, Data curation. SP: Data curation, Writing – review & editing, Formal analysis, Writing – original draft. RT: Formal analysis, Writing – original draft, Data curation, Writing – review & editing. CA: Writing – review & editing, Data curation, Formal analysis, Writing – original draft. PR: Writing – review & editing, Data curation, Writing – original draft, Formal analysis. GT: Formal analysis, Data curation, Writing – original draft, Writing – review & editing. SB: Writing – review & editing, Formal analysis, Writing – original draft. JD: Formal analysis, Writing – original draft, Data curation, Writing – review & editing. LR: Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration, Supervision, Methodology. CÁ-M: Supervision, Methodology, Conceptualization, Writing – review & editing, Writing – original draft, Formal analysis, Funding acquisition, Project administration.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by The World Health Organization, Hospital Universitario San Ignacio, and Keralty supported this work.

Acknowledgments

We extend our sincere appreciation to the Ministry of Health of Colombia at the time of the study for its support in developing this research. We also acknowledge the contributions of the clinical and laboratory staff from participating hospitals, whose efforts facilitated the collection of anonymized clinical data from hospitalized patients with suspected or confirmed COVID-19, subsequently integrated into the WHO Global Clinical Platform for COVID-19. Lastly, we express our profound gratitude to the patients and their families, whose anonymized data will help advance understanding of this disease and refine early prediction models to identify individuals who require hospitalization. We would also like to acknowledge the valuable support of Ronaldo Silva and Soe-Soe Thwin, whose guidance and technical assistance were instrumental in developing and structuring the national database for Colombia as part of the World Health Organization's efforts. Their contribution was essential to ensuring the comprehensiveness and quality of the data collected for this initiative. The authors would like to express their sincere gratitude to Hospital Universitario San Ignacio's Registry Program, Oficina de Investigaciones, for the technical assistance. We also appreciate the contribution and guidance of the Doctoral (PhD) Program in Clinical Epidemiology of the Faculty of Medicine of Pontificia Universidad Javeriana, in Bogotá, Colombia, in which Sandra Liliانا Valderrama-Beltrán is enrolled as a PhD student/PhD candidate. This article has

been registered in the Research Proposal Registration of the Coordination of Scientific Integrity and Surveillance of Universidad Científica del Sur, Lima, Peru.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors alone are responsible for the views expressed in this paper, and they do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

The author AR-M declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

References

- World Health Organization 2023. Infection prevention and control guideline for coronavirus disease 2019 (COVID-19): Executive summary. Geneva: WHO.
- Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. (2020) 395:1054–62. doi: 10.1016/S0140-6736(20)30566-3
- Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L, et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *Lancet*. (2020) 395:1715–25. doi: 10.1016/S0140-6736(20)30854-0
- World Health Organization. WHO roadmap on uses of COVID-19 vaccines in the context of omicron and high population immunity. Geneva: World Health Organization. 2023:1–22.
- Rogers P, Boussina AE, Shashikumar SP, Wardi G, Longhurst CA, Nemati S. Optimizing the implementation of clinical predictive models to minimize National Costs: Sepsis case study. *J Med Internet Res*. (2023) 25:e43486. doi: 10.2196/43486
- Appel KS, Geisler R, Maier D, Miljukov O, Hopff SM, Vehreschild JJ. A systematic review of predictor composition, outcomes, risk of bias, and validation of COVID-19 prognostic scores. *Clin Infect Dis*. (2024) 78:889–99. doi: 10.1093/cid/ciad618
- 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) 369:m1328. doi: 10.1136/bmj.m1328
- Lazar Neto F, Marino LO, Torres A, Cilloniz C, Meirelles Marchini JF, de Garcia Alencar JC, et al. Community-acquired pneumonia severity assessment tools in patients hospitalized with COVID-19: a validation and clinical applicability study. *Clin Microbiol Infect*. (2021) 27:1037.e1–8. doi: 10.1016/j.cmi.2021.03.002
- Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: development and validation of the 4C mortality score. *BMJ*. (2020) 371:m4334. doi: 10.1136/bmj.m3339
- Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ*. (2024) 384:e074820. doi: 10.1136/bmj-2023-074820
- World Health Organization. WHO COVID-19 global clinical platform [internet]. (2021). Available online at: <https://www.who.int/teams/health-care-readiness-clinical-unit/covid-19/data-platform> (Accessed July 4, 2025).
- Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation*. (2015) 131:211–9. doi: 10.1161/CIRCULATIONAHA.114.014508
- Steyerberg EW. Statistics for biology and health clinical prediction models a practical approach to development, validation, and updating second edition. (2019). Available online at: <http://www.springer.com/series/2848> (Accessed September 14, 2025).
- Charlson ME, Pompei P, Ales KL, Mackenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. (1987) 40:373–83. doi: 10.1016/0021-9681(87)90171-8
- Ministerio de Salud y Protección Social. Bogotá: SISPRO - Sistema Integral de Información de la Protección Social. (2011). Available online at: <https://www.sispro.gov.co/central-prestadores-de-servicios/Pages/RUAF-Registro-Unico-de-Afiliados.aspx> (Accessed May 10, 2025).
- World Health Organization. 2021. WHO global clinical platform for the clinical characterization of COVID-19 statistical analysis plan [internet]. Available online at: <https://www.who.int/publications/i/item/WHO-2019-nCoV-Clinical-Analytic-plan-2021.1>
- Riley RD, Snell KIE, Archer L, Ensor J, Debray TPA, Van Calster B, et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. *BMJ*. (2024) 384:e074821. doi: 10.1136/bmj-2023-074821
- Alvarez-Moreno C, Valderrama-Beltrán SL, Silva R, De La Hoz Siegler IH, Varón Vega FA, Martínez-Vernaza S, et al. Associated factors for mortality in a COVID-19 colombian cohort: is the third wave relevant when mu variant was predominant epidemiologically? *Travel Med Infect Dis*. (2023) 53:102579. doi: 10.1016/j.tmaid.2023.102579
- 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) 58:377–82. doi: 10.1136/thorax.58.5.377
- Royal College of Physicians. National Early Warning Score (NEWS) 2. (2017). Available online at: <https://www.rcp.ac.uk/improving-care/resources/national-early-warning-score-news-2/> (Accessed September 30, 2025).
- Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of clinical criteria for sepsis for the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. (2016) 315:762–74. doi: 10.1001/jama.2016.0288
- Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. *PLoS One*. (2011) 6:e16110. doi: 10.1371/journal.pone.0016110
- University of Oxford - News and Events 2020 Low-cost dexamethasone reduces death by up to one third in hospitalised patients with severe respiratory complications of COVID-19. Available online at: <https://www.ox.ac.uk/>

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2026.1738978/full#supplementary-material>

- news/2020-06-16-low-cost-dexamethasone-reduces-death-one-third-hospitalised-patients-severe (Accessed November 1, 2025).
24. Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW, et al. A closed testing procedure to select an appropriate method for updating prediction models. *Stat Med.* (2017) 36:4529–39. doi: 10.1002/sim.7179
 25. Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol.* (2022) 22:1801. doi: 10.1186/s12874-022-01801-8
 26. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak.* (2006) 26:565–74. doi: 10.1177/0272989x06295361
 27. Kuroda S, Matsumoto S, Sano T, Kitai T, Yonetsu T, Kohsaka S, et al. External validation of the 4C mortality score for patients with COVID-19 and pre-existing cardiovascular diseases/risk factors. *BMJ Open.* (2021) 11:e052708. doi: 10.1136/bmjopen-2021-052708
 28. Zahra A, van Smeden M, Abbink EJ, van den Berg JM, Blom MT, van den Dries CJ, et al. External validation of six COVID-19 prognostic models for predicting mortality risk in older populations in a hospital, primary care, and nursing home setting. *J Clin Epidemiol.* (2024) 168:111270. doi: 10.1016/j.jclinepi.2024.111270
 29. Miller JL, Tada M, Goto M, Chen H, Dang E, Mohr NM, et al. Prediction models for severe manifestations and mortality due to COVID-19: a systematic review. *Acad Emerg Med.* (2022) 29:206–16. doi: 10.1111/acem.14447
 30. Andrea Gallego Aristizabal P, Paola Lujan Chavarría T, Isabel Vergara Hernández S, Rincón Acosta F, Paula Sánchez Carmona M, Andrea Salazar Ospina P, et al. External validation of two clinical prediction models for mortality in COVID-19 patients (4C and NEWS2), in three centers in Medellín, Colombia: assessing the impact of vaccination over time. *Infect Dis Now.* (2024) 54:1–7. doi: 10.1016/j.idnow.2024.10492
 31. Vallipuram T, Schwartz BC, Yang SS, Jayaraman D, Dial S. External validation of the ISARIC 4C mortality score to predict in-hospital mortality among patients with COVID-19 in a Canadian intensive care unit: a single-Centre historical cohort study. *Can J Anesth.* (2023) 70:1362–70. doi: 10.1007/s12630-023-02512-4
 32. De Jong VMT, Rousset RZ, Antonio-Villa NE, Buenen AG, Van Calster B, Bello-Chavolla OY, et al. Clinical prediction models for mortality in patients with covid-19: external validation and individual participant data meta-analysis. *The. BMJ.* (2022) 378:1–11. doi: 10.1136/bmj-2021-069881
 33. Maze MJ, Williman J, Anstey R, Best E, Bhally H, Bryce A, et al. Evaluation of risk prediction scores for adults hospitalized with COVID-19 in a highly-vaccinated population, Aotearoa New Zealand 2022. *IJID Regions.* (2024) 12:100424. doi: 10.1016/j.ijregi.2024.100424
 34. RECOVERY Collaborative Group Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with Covid-19 - preliminary report. *N Engl J Med.* (2021) 384:693–704. doi: 10.1056/NEJMoa2021436
 35. Beladiya J, Kumar A, Vasava Y, Parmar K, Patel D, Patel S, et al. Safety and efficacy of COVID-19 vaccines: a systematic review and meta-analysis of controlled and randomized clinical trials. *Rev Med Virol.* (2024) 34:e2507. doi: 10.1002/rmv.2507
 36. Marincowitz C, Saffi L, Hasan M, Hodkinson P, McAlpine D, Fuller G, et al. External validation of triage tools for adults with suspected COVID-19 in a middle-income setting: an observational cohort study. *Emerg Med J.* (2023) 40:509–17. doi: 10.1136/emmermed-2022-212827
 37. Dessie ZG, Zewotir T. Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients. *BMC Infect Dis.* (2021) 21:855. doi: 10.1186/s12879-021-06536-3
 38. Smilowitz N, Kunichoff D, Garshick M, Shah B, Pillinger M, Hochman J, et al. C-reactive protein and clinical outcomes in patients with COVID-19. *Eur Heart J Oxford University Press.* (2021) 42:2270–9. doi: 10.1093/eurheartj/ehaa1103
 39. Akdur G, Daş M, Bardakci O, Akman C, Siddıkoğlu D, Akdur O, et al. Prediction of mortality in COVID-19 through combining CT severity score with NEWS, qSOFA, or peripheral perfusion index. *Am J Emerg Med.* (2021) 50:546–52. doi: 10.1016/j.ajem.2021.08.079
 40. 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) 56:10–3. doi: 10.1183/13993003.02113-2020
 41. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and treatment of adults with community-acquired pneumonia. *Am J Respir Crit Care Med.* (2019) 200:E45–67. doi: 10.1164/rccm.201908-1581ST
 42. Kwok SWH, Wang G, Soheli F, Kashani KB, Zhu Y, Wang Z, et al. An artificial intelligence approach for predicting death or organ failure after hospitalization for COVID-19: development of a novel risk prediction tool and comparisons with ISARIC-4C, CURB-65, qSOFA, and MEWS scoring systems. *Respir Res.* (2023) 24:1–16. doi: 10.1186/s12931-023-02386-6