Infectious diseases now 54 (2024) 104921

Contents lists available at ScienceDirect

Infectious Diseases Now

journal homepage: www.sciencedirect.com/journal/infectious-diseases-now

Original article

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

Paola Andrea Gallego Aristizabal [a,c,1], Tania Paola Lujan Chavarría [a,c,*,1], Sara Isabel Vergara Hernández [b], Federico Rincón Acosta [b], María Paula Sánchez Carmona [b], Paula Andrea Salazar Ospina [b], Carlos Jose Atencia Florez [a,b,c], Carlos Mario Barros Liñán [a,b,d], Fabián Jaimes [a,b,c]

[a] Department of Internal Medicine, Faculty of Medicine, Universidad de Antioquia, Medellín, Colombia
[b] Faculty of Medicine, Universidad de Antioquia, Medellín, Colombia
[c] Hospital Universitario San Vicente Fundación, Medellín, Colombia
[d] Hospital Alma Mater, Medellín, Colombia

ARTICLE INFO

Keywords:
COVID-19
The National Early Warning Score 2 (NEWS2)
4C Mortality Score
Validation studies
Mortality

ABSTRACT

Objectives: External validation of the 4C and NEWS2 scores for the prediction of in-hospital mortality in COVID-19 patients, and evaluation of its operational performance in two time periods: before and after the start of the vaccination program in Colombia.
Methods: Retrospective cohort in three high complexity hospitals in the city of Medellín, Colombia, between June 2020 and April 2022.
Results: The areas under the ROC curve (AUC) for the 4C mortality risk score and the NEWS2 were 0.75 (95% CI 0.73–0.78) and 0.68 (95% CI 0.66–0.71), respectively. For the 4C score, the AUC for the first and second periods was 0.77 (95% CI 0.74–0.80) and 0.75 (95% CI 0.71–0.78); whilst for the NEWS2 score, it was 0.68 (95% CI 0.65–0.71) and 0.69 (95% CI 0.64–0.73). The calibration for both scores was adequate, albeit with reduced performance during the second period.
Conclusions: The 4C mortality risk score proved to be the more adequate predictor of in-hospital mortality in COVID-19 patients in this Latin American population. The operational performance during both time periods remained similar, which shows its utility notwithstanding major changes, including vaccination, as the pandemic evolved.

1. Introduction

SARS-CoV2 infection expanded to every continent and it was declared as a pandemic by the World Health Organization (WHO) the 11th of March 2020, as it jeopardized public health with thousands of daily fatalities worldwide, which represented a substantial challenge for governments, individuals, and societies [1,2]. Three years after, the end of the health emergency was declared. However, it is important to recognize that complete eradication of the virus is highly unlikely, since SARS-CoV2 maintains a significant virulence and continues to circulate. This could lead to daily cases and outbreak periods [3].

The clinical presentation and course of the disease is highly variable amongst patients, due to its heterogeneous immunological response, genetically determined individual expression, associated comorbidities, and the implicated SARS-CoV2 variant [4,5]. These different characteristics hamper classification and adverse event risk assessment by health professionals. Mortality risk prediction is essential to efficient management, which includes opportunity of care and the selection of patients likely to benefit from interventions that change the course of the disease [6]. During the pandemic, numerous prediction models were developed for prognostic purposes, with extensive heterogeneity in their performance in external validation cohorts [7]. The systematic review

* Corresponding author at: Calle 66a # 55ª51, 060568 Medellín, Antioquia, Colombia.
E-mail address: tania.lujan@udea.edu.co (T. Paola Lujan Chavarría).
[1] These authors contributed equally to the production of this research and to the writing of this manuscript.

https://doi.org/10.1016/j.idnow.2024.104921
Received 7 October 2023; Received in revised form 19 December 2023; Accepted 29 April 2024
Available online 3 May 2024

carried out by the COVID-PRECISE group found 107 prediction models with high (97 %) or unclear (3 %) risk of bias, according to the PROBAST tool [8]. Two prediction models (one for prognosis) were identified as having a superior quality compared to others, and it was suggested that an effort be made to validate them with other datasets. The 4C Mortality Score prognostic model [9], deemed to be of good quality, was constructed with a wide range of data, and ranked as having a low risk of bias in most domains [10]. On the other hand, NEWS2 was originally developed in 2012 as an early warning score to monitor hospitalized patients and to predict in-hospital mortality within the next 24 h [11,12]. In the context of the COVID-19 pandemic, NEWS2 has been studied and its use recommended as a means of identifying and treating patients at risk of falling severely ill due to the disease. Agencies such as the WHO and the National Institute for Health and Care Excellence (NICE) in the United Kingdom have supported its use [13]. Furthermore, within the local context, Colombia's health ministry recommended the NEWS2 score to evaluate and direct COVID-19 patients to the appropriate care setting, taking into consideration different risk levels indicated by the score [14].

Both models were developed in the United Kingdom and have had widespread, even worldwide acceptance and validation. However, neither has been validated in a Latin American population. In addition, the strategies used for the prevention and management of COVID-19 have changed with the emergence of vaccines and new treatment options [15], which have had an impact in mortality, creating a need to evaluate the changes in different prediction scores over time [6]. Research and publication of new findings on COVID-19 are fundamental to evidence-based decision making, which should be inclusive of all populations.

Given these considerations, the aim of this study was to locally validate the 4C and NEWS2 scores for mortality risk prediction in hospitalized patients with a confirmed COVID-19 infection test taken on their admission to the emergency department, and to compare the performance of these scores in two time-periods, marked by the start of the vaccination program in Colombia.

2. Methods

Study design, setting and participants: Retrospective multicentric validation cohort, in three high complexity hospitals in Medellín, Colombia. A consecutive review of medical records was done, using the *CIE-10* code: *U07.1 COVID-19*, between June of 2020 and April of 2022. Patients were included if they were 18 years or older, with a confirmed diagnosis of COVID-19 by antigen test for SARS-CoV2 or RT-PCR, and had been admitted to the emergency department. The exclusion criteria included individuals who had passed away within 24 h of admission, as well as those transferred from another center where they had been located for more than 24 h.

Variables and information sources: The data to calculate the 4C and NEWS2 scores were obtained from the medical admission records, with the first registered vital signs and laboratory test taken within the first 24 h. Each model has different stratification variables, to which a score is assigned and a risk group is determined (Tables S1 and S2 in the Supplementary Appendix).

Primary outcome: In-hospital mortality.

Sample size: Determination was made following the recommendations of a minimum of 100 outcomes in the validation cohort [16].

Statistical analysis and missing data: Quantitative variables with normal distribution were presented as mean and standard deviations, and variables with non-normal distribution were presented as medians with interquartile ranges. Qualitative variables were presented with absolute and relative frequencies expressed in percentages. For each score, performance metrics were calculated, including sensitivity, specificity, positive predictive value, and negative predictive value. Additionally, likelihood ratios (LR) were calculated using the Bayes theorem, with in-hospital mortality outcomes assumed as the 'gold standard.' Discrimination capacity for each prediction score was determined through the area under the receiver-operating curve (AUC), while calibration was established using calibration plots and the goodness-of-fit Hosmer-Lemeshow test (H-L). The model's discrimination refers to its capacity to accurately separate patients in high and low-risk groups. Calibration, on the other hand, refers to how close the model's predictions are to reality, namely, how accurate are the risk estimations.

A complete case analysis was performed, as well as simple data imputation, assuming missing variables as normal. For the 4C score, inpatients whose oxygen saturation on admission to the emergency department was registered while they were on supplemental oxygen therapy, oxygen saturation was presumed to be less than 92 %. The NEWS2 score was calculated in terms of registered oxygen saturation, regardless of a patient's history of chronic obstructive pulmonary disease. The analysis was made in two time periods (before and after May 1st, 2021) determined by the initiation of the vaccination program in Colombia and the findings of the RECOVERY study, leading to the widespread use of steroids. The study was reported according to the TRIPOD recommendations [17]. Statistical analysis was performed using STATA software (StataCorp, 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC).

Ethical aspects: The study was performed in alignment with the 2013 Declaration of Helsinki, the research protocol was evaluated and approved by each of the three participating centers' ethics committees, before data collection started.

3. Results

Participants: A total of 3013 medical records were reviewed, with 1789 meeting the inclusion criteria (Fig. 1)**.** The in-hospital mortality rate was 42.1 % (n = 754). The mean age of included patients was 64.9 years (standard deviation (SD): 16.1), with 53.6 % of the population being male, and 36.2 % having at least one comorbidity. Initial severity of the patients was determined by the requirement of advanced vital support, including norepinephrine (n = 56, 3 %) and non-invasive mechanical ventilation (n = 470, 26 %) upon admission. Table 1 shows the clinical characteristics at hospital entry according to their vital status upon discharge, and Table 2 presents the main complications observed during hospitalization. The percentages of missing data for each variable of the prediction models are available in **Table S3 of the** supplementary appendix**.** The highest frequencies were observed in temperature (16 %) and Glasgow Coma Scale (10.6 %).

Performance of the models: The AUC for the 4C score was 0.75 (CI 95 % 0.73 – 0.78) (Fig. 2**A**) and for the NEWS2, it was 0.68 (CI 95 % 0.66 – 0.71) (Fig. 2**B**). The calibration plot showed adequate capacity for the 4C score to predict mortality, even though it was lower regarding patients at very high or very low risk, for whom the risk of death was slightly underestimated (Fig. 3**A**). Measured by the H-L test, calibration resulted in a p-value of 0.13. The NEWS2 score exhibited lower accuracy for prediction of mortality in low-risk patients, overestimating the occurrence of death (Fig. 3**B**).

Impact of changes over time on discrimination and calibration of the scores: The AUC for the 4C scores over the first and second proposed study periods were 0.77 (CI 95 % 0.74 – 0.80) and 0.75 (CI 95 % 0.71 – 0.78), respectively. The AUC for the NEWS2 for the first study period was 0.68 (CI 95 % 0.65 – 0.71) comparable to the second period, which was 0.69 (CI 95 % 0.64 – 0.73). For the first period, the calibration plot for the 4C score exhibited an adequate capacity to predict risk of death, but registered a lower performance in the second study period. The H-L test had a p-value of 0.51 and 0.13 for the first and second periods, respectively. For the NEWS2 score, the calibration plot showed an overestimation of the risk of death in the low and very high-risk groups, mainly during the second period. When measured with H-L test, the p-value was 0.81 for the first period and 0.33 for the second period (see Table 3 **and Figure S1 in the** supplementary appendix).

Other performance measures: The 4C mortality risk score

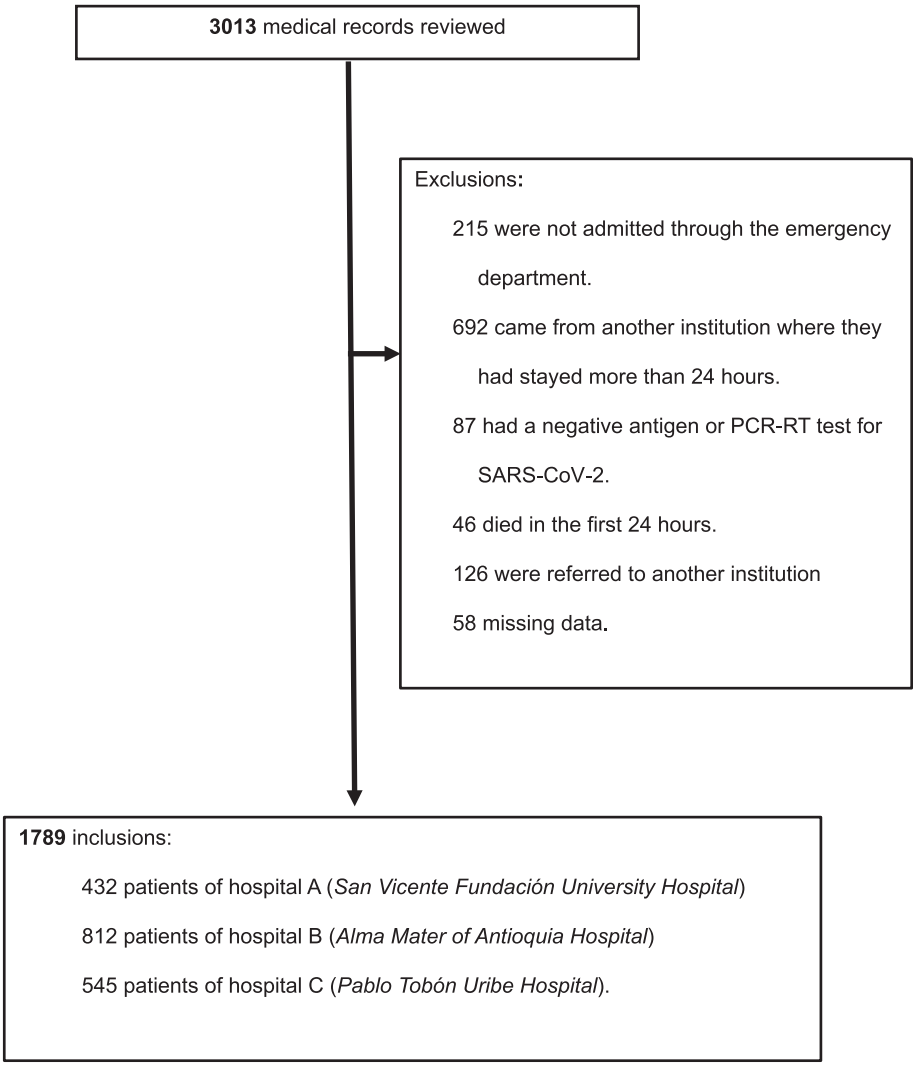


Fig. 1. Study inclusion and exclusion flowchart.

presented good performance in clinically relevant metrics, with a negative predictive value of 95.6 % (CI 95 % 84 – 99 %) for a score of 4 or less, and a positive predictive value of 71.6 % (CI 95 % 65.9 – 76.9) for a score of 15 or higher (Table 4). On the other hand, for the NEWS2 score, a punctuation of 7 or more points had a positive predictive value of 52.7 % (CI 95 % 49.3 – 56) while a punctuation of 5 or less demonstrated a negative predictive value of 72.1 % (CI 95 % 67.6 – 76.2 %) (Table 5).

Total mortality was greater during the first period of the present study (42.15 %), in comparison with the second period (31.3 %), and the mortality observed throughout the different risk groups for the 4C score was greater than reported in the original derivation cohort (Table 6). Likewise, in the different NEWS2 score risk groups, mortality was greater than reported in the original cohort (Table 7). When stratified by risk groups, differences were observed mainly in low-risk (27.93 %) and high-risk categories (52.66 %). The association between the predictor variables of each model and mortality are presented in **Table S4 and Table S5 of the** supplementary appendix.

Sensitivity analysis: In the sensitivity analysis, which involved data imputation, the operational performance of both scores remained the same (**Table S6, Figure S2 and Figure S3 of the** supplementary appendix).

4. Discussion

4.1. Major findings

Performed in the city of Medellín, Colombia, the present study is the first validation concerning Colombian patients of the 4C and NEWS2 scores for in-hospital mortality risk prediction,. The operational performance of both prediction scores was evaluated over time, considering the changes in vaccination and treatment that occurred during the pandemic. The 4C mortality risk score showed adequate discrimination and calibration to predict in-hospital death risk, although a slight underestimation was observed in the low and high-risk groups. These results are similar to those presented in the derivation cohort in the United Kingdom in 2020, where the internal validation cohort had an AUC of 0.76 (CI 95 % 0.760 – 0.773) [9]. Furthermore, the external validation done by De Jong et al. in 18 countries showed that the 4C score exhibited the highest calibration, with an AUC of 0.80 (95 % CI 0.75 – 0.84) [18]. Mortality across risk groups was higher in the present study than in the original derivation cohort, likely due to the features of the studied population.

The 4C score also showed good performance in clinically relevant parameters: a cut-off of four points had high sensitivity, very similar to the original development cohort (99.2 %), while a cut-off of 15 or more showed high specificity (92.5 %), even greater than the derivation cohort (89.9 %). Predicting which patients have a very high probability

Table 1
Baseline clinical and laboratory characteristics according to vital status at discharge.

Variable	Total, n: 1789	Alive n: 1035	Dead n: 754
Age (mean ± SD)	64.9 ± 16.1	61.03 ± 16.4	70.26 ± 13.9
Sex, male (n, %)	959 (5.6)	540 (53.4)	419 (53.9)
Cancer* (n, %)	150 (8.4)	78 (7.7)	72 (9.3)
HIV* (n, %)	16 (0.9)	13 (1.3)	3 (0.4)
Diabetes mellitus* (n, %)	462 (25.9)	252 (24.8)	210 (27.0)
Connective tissue disease* (n, %)	91 (5.1)	47 (4.6)	44 (5.7)
Neurological disorder* (n, %)	93 (5.2)	53 (5.2)	40 (5.2)
Dementia* (n, %)	107 (6.0)	50 (4.9)	57 (7.3)
Liver disease* (n, %)	34 (1.9)	16 (1.6)	16 (2.1)
Renal disease* (n, %)	123 (6.9)	61 (6.0)	62 (8.0)
Lung disease* (n, %)	223 (12.5)	110 (10.8)	113 (14.5)
Heart disease* (n, %)	193 (10.8)	96 (9.5)	97 (12.5)
Obesity** (n, %)	505 (28.2)	266 (26.2)	239 (30.8)
BMI (kg/m^2) (mean ± SD)	28.7 ± 12.8	28.5 ± 10.5	28.9 ± 14.6
Temperature C° (mean ± SD)	36,9 ± 2.13	36.9 ± 0.9	36.9 ± 0.9
Oxygen Saturation % (mean ± SD)	84 ± 12.9	86.8 ± 10	81.1 ± 15.1
DBP (mmHg) (mean ± SD)	92 ± 13.6	77 ± 12	73 ± 14.3
SBP (mmHg) (mean ± SD)	127 ± 23.7	127 ± 22.9	127 ± 24.8
RR (/min) (mean ± SD)	25.1 ± 8.2	23.7 ± 7.2	27
HR (/min) (mean ± SD)	95.3 ± 18.6	95.5 ± 18.5	94.9 ± 18.9
Glasgow Scale < 15 (n (%))	160 (8.9)	32 (3.1)	64 (8.5)
BUN (mg/dL) (mean ± SD)	25.5 ± 18.6	25.5 ± 18.6	30.6 ± 20.8
CRP (mg/dL) (mean ± SD)	12,.4 ± 13.7	12.9 ± 14.1	11.7 ± 12.9
PO2 (mmHg) (mean ± SD)	73.1 ± 26.9	73.1 ± 26.9	66.3 ± 23.6
PCO2 (mmHg) (mean ± SD)	32.8 ± 10.1	32,.8 ± 10.1	32.7 ± 10.6
Ferritin ng/mL (IQR)	1313 (752 – 2199)	1100 (154–1984)	1526 (238–2694)
Troponin pg/ml (IQR)	302 (2.5–396)	151 (2.5–140)	485 (4.77–752)
D-dimer ng/ml (IQR)	2247 (0.65 – 2446)	2631 (0/6 – 2192)	2102 (0.69–679)
LDH U/L (RIQ)	440 (289 – 703)	363 (202–540)	537 (265–840)

HIV: Human immunodeficiency virus, BMI: body mass index, DBP: diastolic blood pressure, SBP: systolic blood pressure, RR: respiratory rate, HR: heart rate, BUN: blood urea nitrogen, CRP: C reactive protein, PO2: partial arterial oxygen pressure, PCO2: partial arterial carbon dioxide pressure, LDH: lactate dehydrogenase, SD: standard deviation, IQR: interquartile range.
* Comorbidities defined by the Charlson index.
** Clinically defined obesity.

Table 2
Complications during hospitalization according to vital status at discharge.

Variable	Total, n: 1789	Alive n: 1035	Dead n: 754
IMV (n, %)	553 (30.9)	122 (11.7)	430 (57)
DVT (n, %)	41 (2.3)	40 (3.9)	17 (2.2)
PE (n, %)	69 (3.8)	24 (2.3)	29 (3.8)
Bacteremia (n, %)	572 (31.9)	257 (41.7)	315 (2.3)

IMV: invasive mechanical ventilation, DVT: deep vein thrombosis, PE: pulmonary embolism.

of dying during hospitalization, this score could contribute to selection of in-hospital care settings, including the need for monitoring in high-dependency units. In other external validation cohorts, similar findings have been reported for the 4C mortality risk score. Another study from the United Kingdom undertaken during the second outbreak of the pandemic (August 2020 to February 2021), found an AUC of 0.76 with

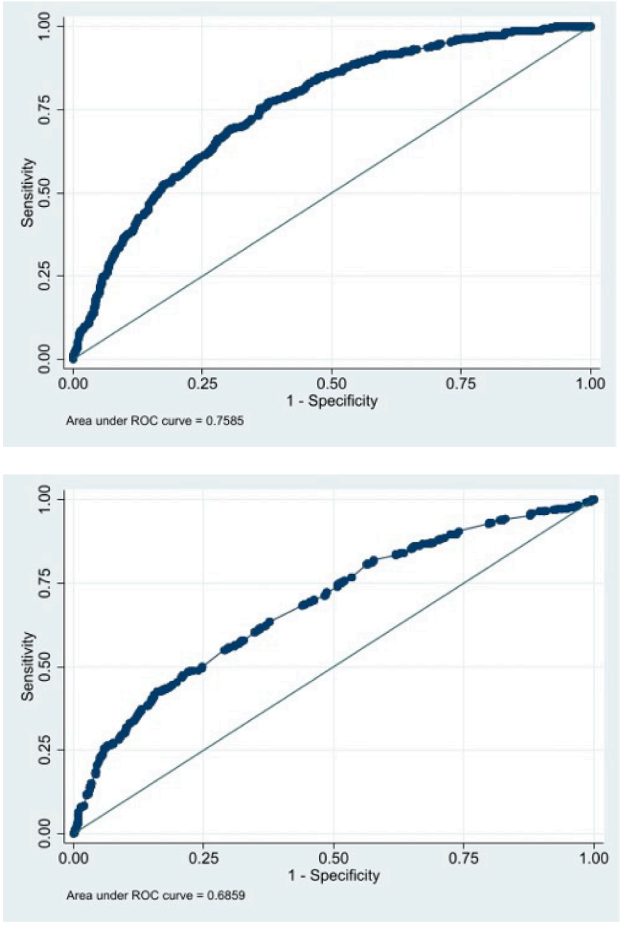
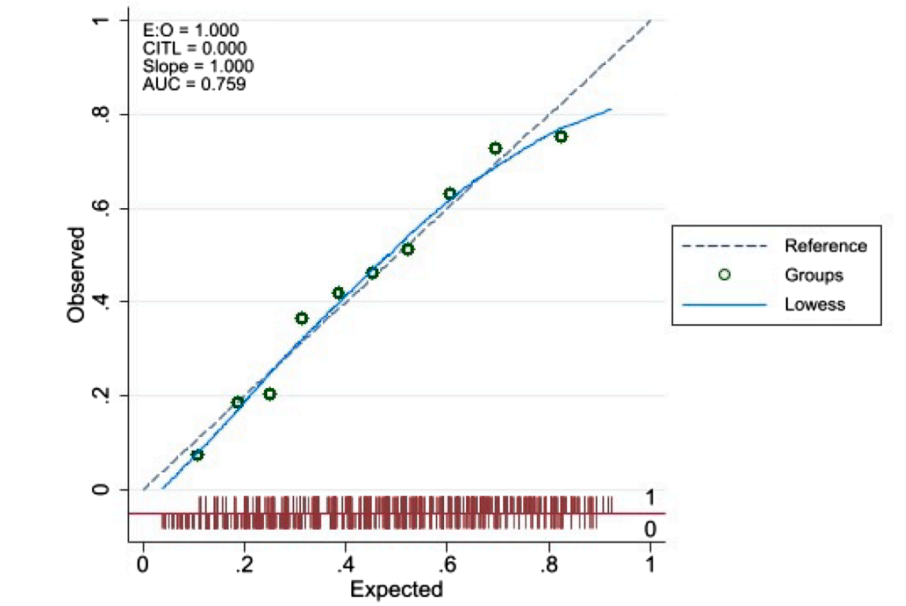


Fig. 2. Receiver Operating Characteristics Curve for mortality in COVID-19 patients. A. 4C mortality score, B. NEWS2 score.

good calibration. In the four risk groups, reported mortality was very similar to the one reported on the inception cohort [19]. In Paris, in a retrospective multicenter cohort of adult patients hospitalized for COVID-19 from January 2020 to April 2021, the 4C mortality risk score outperformed 32 other scores, with an AUC of 0.784 (CI 95 % 0.77–0.795) [20]. In the USA, external validation was performed using the dataset from the RECOVERY trial, which had comparable discrimination, with an AUC of 0.786 (CI95% 0.773–0.799) and adequate calibration [21].

The NEWS2 score displayed acceptable discrimination, albeit slightly lower than that of the 4C score, but the calibration showed overestimation of the probability of death in both the low and high-risk categories. Similar findings were reported in an external validation study o, COVID-19 patients in England at the start of the pandemic, finding that the model underestimated mortality risk in all risk categories, evidencing poor discrimination and an AUC of 0.65 (CI 95 % 0.61 – 0.68) for prediction of in-hospital mortality [22]. Another cohort from the United Kingdom found similar performance, with an AUC of 0.65 (CI 95 % 0.61 – 0.68) [12]. In this study, the cut-off of 5 points displayed sensitivity of 83 % and negative predictive value of 79 % [22], lower than reported in a study from Norway, which reported a sensitivity of 86.7 % (CI 95 % 59.5 – 98.3), and negative predictive value of 94.7 (CI 95 % 83.0 – 98.5).

The findings suggest that the 4C mortality risk score has greater discriminatory power and better calibration than the NEWS2 score, consistent with its performance in the development cohort, where it outperformed 15 other scores, including NEWS2 (CI 95 % 0.774 Vs 0.654) [9]. In Belgium, comparable results were achieved. An AUC of



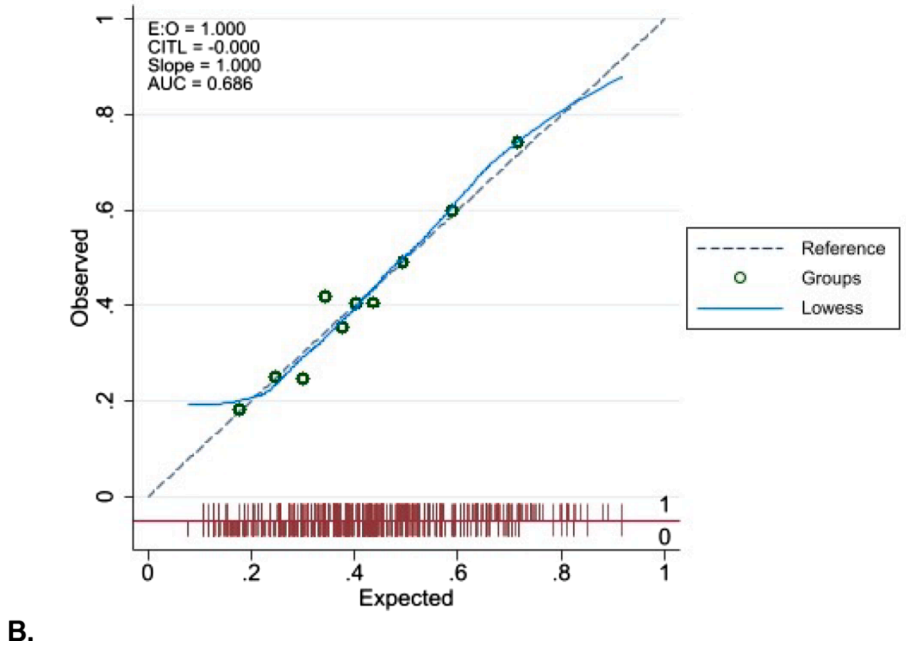


Fig. 3. Probability of the expected result compared to the observed result is shown as calibration plot. A. 4C mortality score, B. NEWS2 score. Groups: Deciles of risk; Lowess: Locally Weighted Scatterplot Smoothing; E:O: Expected:Observed ratio; CITL: calibration-in-the-large; AUC: Area Under the Curve.

Table 3
Area under the curve (AUC) and Hosmer-Lemeshow test for mortality in patients with COVID-19 by study period.

Score	Total	First period*	Second period*
AUC (IC 95 %)			
Mortality 4C	0.75 (0.73–0.78)	0.77 (0.74–0.80)	0.75 (0.71–0.78)
NEWS2	0.68 (0.66–0.71)	0.68 (0.65–0.71)	0.69 (0.64–0.73)
Hosmer-Lemeshow test (*p*-statistic)			
Mortality 4C	0.13	0.51	0.13
NEWS2	0.68	0.82	0.33

* First period: Before May 1st, 2021. Second period: After May 1st, 2021.

0.71 (CI 95 % 0.58 – 0.83) was obtained for NEWS2, and an AUC of 0.80 (CI 95 % 0.69 – 0.91) for the 4C score, for predicting mortality risk at 30 days [23]. The potentially lower discrimination observed in the NEWS2 score may be attributed to its non-specific development for COVID-19 patients, while its criteria primarily focus on hemodynamic variables, which may be less pertinent in the initial presentation of the disease [24].

The operational characteristics of the 4C and NEWS2 risk scores were maintained throughout the evaluated periods, determined mainly by the start of vaccination in Colombia, increased circulation of variant B.1 at the onset of the second period, and publication of the RECOVERY trial findings, which on February 25th, 2021confirmed the efficacy of steroids. Similar findings have been reported for the 4C risk score in a study conducted in Canada, where evaluation of the performance of the score

Table 4
Performance metrics for the 4C mortality risk score to confirm the mortality in different cut-off points.

Cut-off	No (%) of patients	Sensitivity (%) (CI 95 %)	Specificity (%) (IC 95 %)	PPV (%) (CI 95 %)	NPV (%) (CI 95 %)	LR (+) (CI 95 %)	LR (−) (CI 95 %)
≥15	275(15.3)	**26.1** (23–29.4)	**92.5** (90.7–94)	**71.6** (65.9–76.9)	**63.2** (60.7–65.6)	**3.4** (2.7–4.4)	**0.79** (0.7–0.8)
≥9	1205(67)	**86.9** (84.2–89.2)	**46.9** (43.8–50)	**54.4** (51.5–57.2)	**83** (79.8–86)	**1.63** (1.5–1.7)	**0.28** (0.2–0.3)
≥4	1744(97)	**99.7** (99–100)	**4.15** (3.02–5.56)	**43.1** (40.8–45.5)	**95.6** (84–99)	**1.04** (1.0–1.0)	**0.063** (0.01–0.26)

LR (+): Positive likelihood ratio, LR (−): Negative likelihood ratio, PPV: Positive predictive value, NPV: Negative predictive value No: Number.

Table 5
Performance metrics for the NEWS2 mortality risk score to confirm the mortality in different cut-off points.

Cut-off	No (%) of patients	Sensitivity (%) (CI 95 %)	Specificity (%) (CI 95 %)	PPV (%) (CI 95 %)	NPV (%) (CI 95 %)	LR (+) (CI 95 %)	LR (−) (CI 95 %)
≥7	886(48,41)	**60.5** (56.9–64)	**60.4** (57.3–63.4)	**52.7** (49.3–56)	**67.7** (64.6–70.7)	**1.53** (1.3–1.6)	**0.65** (0.5–0.72)
≥5	1345(75)	**83.6** (80.7–86.1)	**30.9** (28.1–33.8)	**46.8** (44.1–49.5)	**72.1** (67.6–76.2)	**1.21** (1.1–1.27)	**0.53** (0.4–0.64)

LR (+): Positive likelihood ratio, LR (−): Negative likelihood ratio, PPV: Positive predictive value, NPV: Negative predictive value No: Number.

Table 6
Mortality in patients with COVID-19 by period, stratified by 4C mortality risk groups, and compared to the derivation cohort.

Risk groups, 4C mortality risk score	Cohorts			
	First period	Second period	Total	United Kingdom (Derivation cohort) *
All groups	**39.91 %** (439/ 1100)	**45.72 %** (315/ 689)	**42.15 %** (754/ 1789)	**31.3 %** (18155/ 57924)
Low (0–3)	**0 %**(0/ 27)	**11.11 %** (2/18)	**4.44 %** (2/45)	**1.5 %**(65/4224)
Intermediate (4–8)	**15.65 %** (54/ 345)	**22.16 %** (43/194)	**18 %** (97/ 539)	**9.4 %**(1237/ 13166)
High (9–14)	**46.99 %** (265/ 564)	**52.73 %** (193/366)	**49.25 %** (458/ 930)	**34 %**(9976/ 29755)
Very high (15–21)	**73.17 %** (120/ 164)	**69.37 %** (77/111)	**71.64 %** (197/ 275)	**64 %**(6877/ 10679)

* Data reproduced from [9].

Table 7
Mortality in patients with COVID-19 by period, stratified by NEWS2 mortality risk groups, and compared to the derivation cohort.

Risk groups, NEWS2 mortality risk score	Cohorts			
	First period	Second period	Total	Derivation cohort *
All groups	**39.91 %** (439/ 1100)	**45.72 %** (315/ 689)	**42.15 %** (754/ 1,789)	**14,8%**
Low (0–4)	**25.87 %** (74/ 286)	**31.65 %** (50/158)	**27.93 %** (124/444)	**5.5 %**
Medium (5–6)	**35.93 %** (106/295)	**36.96 %** (68/184)	**36.33 %** (174/479)	**11.3 %**
High (≥7)	**49.90 %** (259/519)	**56.77 %** (197/ 347)	**52.66 %** (456/866)	**27.03 %**

* Data reproduced from [10,19].

during outbreaks was conducted across three periods, concluding in June 2021 [6]. The AUC values during outbreaks were 0.81 (CI 95 % 0.76 – 0.86), 0.74 (CI 95 % 0.69 – 0.80) and 0.76 (CI 95 % 0.69 – 0.83), respectively. However, few studies have evaluated performances after the start of vaccinations.

To date, there have been no published studies evaluating the clinical impact of the 4C and NEWS2 scores, based on the available information. This underscores the need for future research to explore the practical implications of these scores, which might identify potential barriers in their application. Such investigations will provide insights into whether these scores results in changes in medical decisions, ultimately contributing to improved clinical outcomes [25].

4.2. Strengths and limitations

The strengths of the present study lie in the clinical setting, where no prior validation data had existed for this Colombian population. This holds significance in the global public health context of diseases like COVID-19, contributing to the bridging of information and research gaps between high and low-income countries. Additionally, a significant number of outcomes per variable were achieved, and the potential impact of changing factors during the pandemic, which modified the performance of prediction models, was taken into consideration.

Among the limitations, missing data in the variables composing the prediction models, particularly the Glasgow Scale and temperature, could impact performance of the models. To address this limitation, simple data imputation was employed, assuming missing variables as normal. The operational performance of the scores remained consistent in this sensitivity analysis. Another limitation is the retrospective character of the study, which limited control over registration and quality of predictor variables. Finally, the scores are likely to change in settings with lower mortality, thereby limiting the external validity of the findings.

4.3. Conclusions and implications

In this study, the mortality risk prediction tool 4C was shown to have acceptable characteristics, justifying further evaluation in comparison with the NEWS2 score for prediction of in-hospital mortality in patients with COVID-19 in Medellín, Colombia. In addition, it was found that the operational characteristics of the prognostic model remained intact, despite changes that arose during the SARS-CoV2 pandemic, such as vaccination and therapeutic interventions. Implementation of these models in clinical practice could contribute to better risk management, selection of care settings, and evidence-based decision-making.

The impact on decision-making and patient outcomes remains to be explored in future investigations, along with the contribution of risk

scores to the selection of care settings, and the frequency and depth of medical follow-up.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express their sincere appreciation to all the people and institutions who contributed in a significant manner to the realization of the present study. We would especially like to thank Ana Maria Londoño Giraldo, Sara Botero Bolivar, Isabella Rocha Giraldo, Cristian Fernando Marcillo Calderón and Esteban Castrillon Martínez, for their valuable contributions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idnow.2024.104921.

References

[1] Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, et al. Virology, epidemiology, pathogenesis, and control of COVID-19. Viruses 2020;12(4):372.
[2] Shi Y, Wang G, Cai XP, Deng JW, Zheng L, Zhu HH, et al. An overview of COVID-19. J Zhejiang Univ Sci B 2020;21(5):343–60.
[3] Callaway E. Beyond omicron- what's next for COVID's viral evolution. Nature 2021;600(7888):204–7.
[4] Pereira NL, Ahmad F, Byku M, Cummins NW, Morris AA, Owens A, et al. COVID-19: understanding inter-individual variability and implications for precision medicine. Mayo Clin Proc 2021;96(2):446.
[5] Flisiak R, Rzymski P, Zarębska-Michaluk D, Ciechanowski P, Dobrowolska K, Rogalska M, et al. Variability in the clinical course of COVID-19 in a retrospective analysis of a large real-world database. Viruses 2023;15(1):149.
[6] Jones A, Pitre T, Junek M, Kapralik J, Patel R, Feng E, et al. External validation of the 4C mortality score among COVID-19 patients admitted to hospital in Ontario, Canada: a retrospective study. Sci Rep 2021;11(1):18638.
[7] Harish K, Zhang B, Stella P, Hauck K, Moussa MM, Adler NM, et al. Validation of parsimonious prognostic models for patients infected with COVID-19. BMJ Heal Care Informatics 2021;28(1):e100267.
[8] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170(1):51–8.
[9] 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;370:m3339.
[10] 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.
[11] Royal College of Physicians. *National Early Warning Score (NEWS) 2 Standardising the Assessment of Acute-Illness Severity in the NHS*.; 2017. www.rcplondon.ac.uk. Accessed October 23, 2023.
[12] Scott LJ, Tavaré A, Hill EM, Jordan L, Juniper M, Srivastava S, et al. Prognostic value of national early warning scores (NEWS2) and component physiology in hospitalised patients with COVID-19: A multicentre study. Emerg Med J 2022;39(8):589–94.
[13] World Health Organization. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected. 2020. https://www.who.int/publications/i/item/10665-332299 Accessed October 23, 2023.
[14] Ministerio de Salud y Protección Social. Lineamientos para la evaluacion del riesgo y el tratamiento domiciliario, según la valoración médica así lo determine, en pacientes con sospecha o confirmación de infección por SARS-CoV-2 / COVID-19. 2020:1-20.
[15] Baba H, Ikumi S, Aoyama S, Ishikawa T, Asai Y, Matsunaga N, et al. Statistical analysis of mortality rates of coronavirus disease 2019 (COVID-19) patients in japan across the 4C mortality score risk groups, age groups, and epidemiological waves: A report from the nationwide COVID-19 Cohort. Open Forum Infect Dis 2022;10(1):ofac638.
[16] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016;35(2):214–26.
[17] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med 2015;162(1):W1–.
[18] 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. BMJ 2022;378:e069881.
[19] Knight SR, Gupta RK, Ho A, Pius R, Buchan I, Carson G, et al. Prospective validation of the 4C prognostic models for adults hospitalised with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol. Thorax 2022;77(6):606–15.
[20] Lombardi Y, Azoyan L, Szychowiak P, Bellamine A, Lemaitre G, Bernaux M, et al. External validation of prognostic scores for COVID-19: a multicenter cohort study of patients hospitalized in Greater Paris University Hospitals. Intensive Care Med 2021;47(12):1426–39.
[21] Gordon AJ, Govindarajan P, Bennett CL, Matheson L, Kohn MA, Camargo C, et al. External validation of the 4C Mortality Score for hospitalised patients with COVID-19 in the RECOVER network. BMJ Open 2022;12(4):e054700.
[22] Bradley P, Frost F, Tharmaratnam K, Wootton DG. NW Collaborative Organisation for Respiratory Research. Utility of established prognostic scores in COVID-19 hospital admissions: Multicentre prospective evaluation of CURB-65, NEWS2 and qSOFA. BMJ Open Respir Res 2020;7(1). e000729.
[23] Yildiz H, Castanares-Zapatero D, Hannesse C, Vandermeersch D, Pothen L, Yombi JC. Prospective validation and comparison of COVID-GRAM, NEWS2, 4C mortality score, CURB-65 for the prediction of critical illness in COVID-19 patients. Infect Dis (Auckl) 2021;53(8):640–2.
[24] Joosten SA, Smeets MJR, Arbous MS, Manniën J, Laverman S, Driessen MMG, et al. Daily disease severity in patients with COVID-19 admitted to the hospital: The SCODA (severity of coronavirus disease assessment) score. PLoS One 2023;18(9).e0291212.
[25] Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed. JAMAevidence. McGraw Hill Medical. https://jamaevidence.mhmedical.com/Book.aspx?bookId=847. Accessed November 8, 2023.