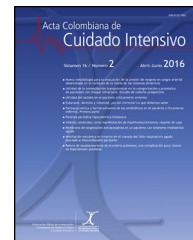




# Acta Colombiana de Cuidado Intensivo

[www.elsevier.es/accii](http://www.elsevier.es/accii)



## ORIGINAL ARTICLE

### Performance of APACHE II, SOFA, and CURB-65 for death prognosis in COVID-19 critical patients: A prospective cohort study

John J. Sprockel<sup>a,b,\*</sup>, Anggie L. Murcia<sup>a,c</sup>, María C. Díaz<sup>c</sup>, Luisa F. Rios<sup>c</sup>, Oscar I. Quirós<sup>c</sup>, Jhon E. Parra<sup>a</sup>

<sup>a</sup> Intensive Care Unit of Hospital El Tunal – Integrated Subnet of Health Services of the South, Bogotá, Colombia

<sup>b</sup> Institute of Research of the Fundación Universitaria de Ciencias de la Salud, Bogotá, Colombia

<sup>c</sup> Specialization in Epidemiology of the Universidad del Rosario and Universidad CES, Colombia

Received 2 June 2023; accepted 8 December 2023

#### KEYWORDS

Clinical prediction;  
COVID-19;  
Death;  
Intensive care;  
Prognosis

#### Abstract

**Rationale:** The identification of patients with high risk of death makes individual decision making more efficient, optimizing resources and improving the quality of the medical attention. The prognostic utility of APACHE II, SOFA, and CURB-65 in critical COVID-19 has not yet been determined.

**Objective:** The present work intends to validate these scoring systems for the prediction of death within 60 days in patients hospitalized in intensive care with COVID-19.

**Methods:** A prospective cohort was conducted which included adults with confirmed COVID-19 hospitalized in ICU. The scores were calculated by building ROC curves and calculating the areas under the curve and the curves of decision analysis.

**Measurements:** The operating characteristics and Kaplan Meier curves were calculated.

**Results:** 320 patients between July and December 2020 were included, mortality within 60 days was 49.7%. CURB-65 had an AUC of 0.68 (CI 0.62–0.74), sensitivity 73.6%, and specificity 55.9%; APACHE-II had an AUC of 0.65 (CI 0.60–0.71), sensitivity 51.6%, and specificity 70.2%; and SOFA had an AUC of 0.70 (CI 0.64–0.75), sensitivity 83.6%, and specificity 52.2%. The three scoring systems obtained values of  $p < 0.001$  for the LongRank test in the survival curves, offering moderate increments in the net benefit.

**Conclusion:** The scoring systems for clinical prediction CURB-65, APACHE II, and SOFA exhibited moderate discriminatory ability for death within 60 days in patients with COVID-19 hospitalized in intensive care; for the optimal cut-off level, there was an adequate power of discrimination.

© 2024 Asociación Colombiana de Medicina Crítica y Cuidado Intensivo. Published by Elsevier España, S.L.U. All rights reserved.

\* Corresponding author.

E-mail address: [jjsprockel@fucsalud.edu.co](mailto:jjsprockel@fucsalud.edu.co) (J.J. Sprockel).

<https://doi.org/10.1016/j.accii.2023.12.006>

0122-7262/© 2024 Asociación Colombiana de Medicina Crítica y Cuidado Intensivo. Published by Elsevier España, S.L.U. All rights reserved.

## PALABRAS CLAVE

Predicción clínica;  
COVID-19;  
Muerte;  
Cuidados intensivos;  
Pronóstico

## Rendimiento de APACHE II, SOFA y CURB-65 para el pronóstico de muerte en pacientes críticos con COVID-19: un estudio prospectivo de cohortes

### Resumen

**Introducción:** La identificación de pacientes con alto riesgo de muerte hace más eficiente la toma de decisiones individuales, optimizando recursos y mejorando la calidad de la atención médica. Aún no se ha determinado la utilidad pronóstica de las escalas *Acute Physiology And Chronic Health Evaluation II* (APACHE II), *Sepsis-related Organ Failure Assessment* (SOFA) y *Confusion, Urea Nitrogen, Respiratory Rate, Blood Pressure, 65 years of age and older* (CURB-65) en la COVID-19 crítica.

**Objetivo:** El presente trabajo pretende validar estos sistemas de puntuación para la predicción de muerte en 60 días en pacientes hospitalizados en la Unidad de Cuidados Intensivos (UCI) con COVID-19.

**Métodos:** Se realizó una cohorte prospectiva que incluyó adultos con COVID-19 confirmada hospitalizados en la UCI. Las puntuaciones se calcularon construyendo curvas *Receiver-Operating-Characteristic* (ROC) y estimando las áreas bajo la curva (AUC), así como las curvas de análisis de decisión.

**Mediciones:** Se calcularon las características operativas y las curvas de Kaplan Meier.

**Resultados:** Se incluyeron 320 pacientes entre julio y diciembre de 2020, la mortalidad en 60 días fue de 49,7%. La escala CURB-65 tuvo un AUC de 0,68 (intervalo de confianza [IC] 0,62 - 0,74), sensibilidad de 73,6% y especificidad de 55,9%; APACHE-II mostró un AUC de 0,65 (IC 0,60 - 0,71), sensibilidad de 51,6% y especificidad de 70,2%; y SOFA exhibió un AUC de 0,70 (IC 0,64 - 0,75), sensibilidad de 83,6% y especificidad de 52,2%. Los tres sistemas de puntuación obtuvieron valores de  $p < 0,001$  para el test de *LongRank* en las curvas de supervivencia, ofreciendo incrementos moderados en el beneficio neto.

**Conclusiones:** Los sistemas de puntuación para la predicción clínica CURB-65, APACHE II y SOFA mostraron una capacidad discriminatoria moderada para la muerte en un plazo de 60 días en pacientes con COVID-19 hospitalizados en UCI; para el nivel de corte óptimo, hubo un poder de discriminación adecuado.

© 2024 Asociación Colombiana de Medicina Crítica y Cuidado Intensivo. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

## Introduction

Since the COVID-19 pandemic was declared on 11 March 2020, the intensive care units (ICU) worldwide experienced periods of sudden increase in cases of acute respiratory insufficiency following the infection with the virus, which caused a deficit in the capacity of response of the hospital system.<sup>1</sup> Between 5% and 32% of the patients hospitalized with COVID-19 worldwide, have required attention in ICU.<sup>2</sup> Mechanical ventilation is a support measure for critical patients in intensive care units; however, the patients with this condition register a mortality rate of 27%, which can vary according to the presence of comorbidities and other clinical conditions, such as hypoxic respiratory failure, where the mortality rate reaches up to 50%.<sup>3</sup>

The APACHE II<sup>4</sup> and SOFA<sup>5</sup> scoring systems have been adapted and used to assess the risk of death in patients hospitalized in intensive care, while the CURB-65 was designed for the assessment of risk of death in patients with community-acquired pneumonia.<sup>6</sup> A meta-analysis of 23 studies performed in 2010, which included 22753 participants, found that CURB-65  $\geq 2$  had an odds ratio (OR)

of 6.4 for mortality in patients with community-acquired pneumonia, the data grouped found sensitivity of 0.62%, specificity of 0.79%, PPV of 24%, and NPP of 95%<sup>7</sup>; another meta-analysis of 40 studies reported aggregated AUC of 0.80 for CURB-65 for mortality within 30 days.<sup>8</sup> The PROWESS study, an analysis of 275 patients with pneumonia hospitalized in ICU, reported a C-statistic for CURB-65  $\geq 2$  of 0.66 and for APACHE II  $\geq 25$  of 0.64 for death within 28 days.<sup>9</sup> A study that included 406 Japanese patients with pneumonia found that SOFA registered an AUC of 0.769 for patients with community-acquired pneumonia.<sup>10</sup>

The identification of patients with high risk of death in ICU makes support systems more efficient in accordance with the individual requirements, optimizing resources and improving the quality of the medical attention. Currently, there is insufficient information on the utility of APACHE II, SOFA, and CURB65 for the prognosis of death in critical COVID-19 patients hospitalized in ICU in the Colombian population. The purpose of the present work was to determine the utility of these scoring systems for the prognosis of death, including adjustments by other variables of clinical importance, in patients hospitalized in ICU in a hospital of Bogotá.

## Materials and methods

An analytic prospective cohort study was conducted, which included patients aged 18 years or older hospitalized in intensive care by diagnosed severe COVID-19, confirmed via polymerase chain reaction test in real time for SARS-CoV-2 between July and December 2020 in one of nine intensive care units of the Integrated Subnet of Health Services of the South – Hospital El Tunal of Bogotá. The patients who stayed for over 72 h in ICU in a different institution, pregnant women, patients with conditions determining a low life expectation, and the patients who died within the first 24 h upon admission, were excluded.

Upon screening the patients through the census at each department, the patients admitted were identified. The laboratory tests and imaging required for the first 72 h upon admission were verified by entering the medical history data in a virtual form containing the demographic data, clinical presentation, antecedents, physical examination, laboratory tests, and imaging. Subsequently, we computed the APACHE II, SOFA, and CURB-65 scores for each patient upon admission.

## Statistical analysis

A convenience sample was selected based on the patients admitted during the period of time defined for the study. A descriptive analysis and frequency tables were made to characterize and summarize the clinical data for the target population. ROC (Receiver Operating Characteristic) curves were built in order to assess the performance of each scoring system, the areas under the curve (AUC) were analyzed and the cut-off points were set so that sensitivity and specificity were the best, obtaining the following values: for CURB-65 1.5, SOFA 5.5, and APACHE II 14.5; in addition, the Youden Index, the positive and negative predictive values, and the positive and negative likelihood ratio (LR) were calculated. A Kaplan–Meier curve was built for each scale using the cut-off point obtained by Youden,  $p$  was calculated through the LogRank test.

A decision curve analysis framework was used to compare the models with the strategy of treating all or none of the patients. The net benefit for each strategy was calculated by subtracting the false positives from the true positives weighted by the relative damage of a false positive result and a false negative. The analyses were made using the statistical software R version 4.0.2 (R Foundation, Vienna, Austria) using the packages “pROC”, “ROCit”, “rmda”, “survival” and “survminer”.

The present study was approved by the committee of ethics of the Integrated Subnet of Health Services of the South, register 138 of 2020 and the Universidad CES in accordance with act 167 of July 2021, it was determined that an informed consent was not necessary. No funding was received.

## Results

Fig. 1 shows the flowchart of the patients indicating how the screening was made, 320 patients with COVID-19 admitted in ICU at Hospital El Tunal between July and December

2020 were included. 206 (64.4%) patients were male, 52.8% of the patients were between 60 and 79 years old, and 7.8% were older than 80. 134 (46.2%) registered a history of obesity, 79 (24.7%) smoking, 70 (21.9%) diabetes mellitus type 2, 128 (40.0%) arterial hypertension, and 66 (20.6%) a chronic pulmonary condition. Mortality within 60 days was 49.7% (159 patients). The patients who did not survive were elderly, with higher levels of ferritin, lactate dehydrogenase, relation Pa/Fi O<sub>2</sub> and troponin, and developed shock, renal lesion, liver disease, and coagulopathy on a frequently basis (Table 1). Fig. 2 shows the ROC curve of each clinical prediction rule.

CURB-65 had an average result of 1.76 and a standard deviation of 1.03, with an AUC of 0.68 (CI 0.62–0.74). Out of the 188 patients with high risk of death (score > 1.5 by Youden) 117 died (62.2%), while 42 out of 132 (31.81%) patients with low risk died. The difference of ratios was 30.4% (CI 19.2–41.6). The patients with high risk registered a RR of 1.95 (CI 1.48–2.57) with  $p < 0.001$ . The sensitivity for CURB-65 at a cut-off point of 1.5 was 73.6% (CI 95% 68.8–78.4%) and a had a specificity of 55.9% (CI 50.5–61.3%). The predictive positive and negative values were 62.2% and 68.2%, respectively.

APACHE-II reported a minimum value of 2 and a maximum of 44, with a median of 13.9 and a standard deviation of 6.7. It had an AUC of 0.65 (CI 0.60–0.71). Out of the 130 patients classified as high risk of death (score > 14.5 identified by Youden), 82 died (63.1%) compared with 77 (40.5%) out of 190 of low risk, with a difference of ratios of 22.55% (CI 11.5–34). The subjects with high risk of death had a RR of 1.55 (CI 1.25–1.93) and  $p < 0.001$ , a sensitivity of 51.6% (CI 46.1–57%), and a specificity of 70.2% (CI 65.2–75.2).

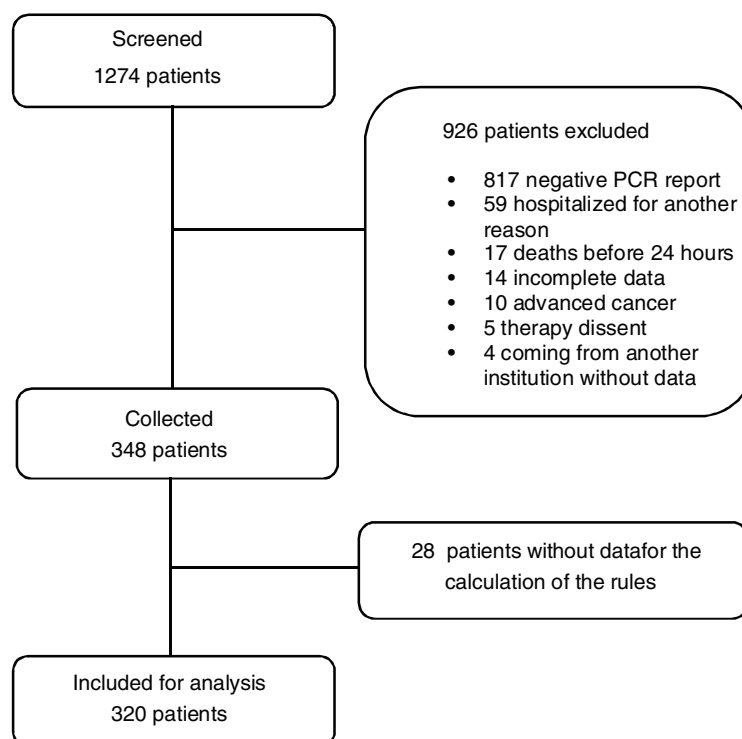
SOFA had a minimum value of 2 and a maximum of 17 with an average of 6.8 and a standard deviation of 3.3. The AUC was 0.70 (CI 0.64–0.75) (Fig. 2c). Out of the 210 patients classified as high risk of death, 133 died (63.3%) compared with 26 (23.6%) out of 110 in the low risk of death group, with a difference of ratios of 39.7% (CI 28.7–50.7). The subjects with high risk of death (score > 5.5 by Youden) had a RR of 2.67 (CI 2.67–1.88) and  $p < 0.001$ , a sensitivity of 83.6% (CI 79.6–87.7%), and a specificity of 52.2% (CI 46.7–57.6%).

Fig. 3 shows the Kaplan and Meier survival curves for the three scoring systems for clinical prediction assessed in accordance with the cut-off values obtained through the Youden test (1.5 for CURB-65, 14.5 for APACHE II, and 5.5 for SOFA), a significant difference was found for survival within 60 days in cases with values for  $p$  smaller than 0.001 in all the cases.

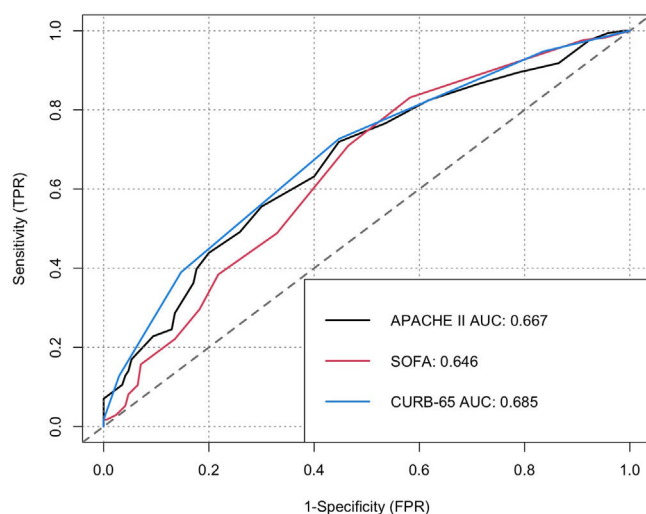
For the threshold range between 1% and 50% in the analysis of the decision curve (Fig. 4), CURB-65 might prevent a larger number of unnecessary interventions; however, the three scoring systems offer moderate increments in the net benefit compared with the risk strategies of ‘treating all’ and ‘treating none’.

## Discussion

Almost three years after the COVID-19 pandemic was declared, more data about the prognostic markers for decision making in patients with critical disease are still necessary. Even with the proliferation of tools designed



**Figure 1** Flowchart of patients included in the study.



**Figure 2** ROC curve for death within 60 days calculated for CURB-65, APACHE II, and SOFA.

specifically for COVID-19, it is remarkable that the traditional tools for intensive care continue to be valid as comparators and in many cases these tools have demonstrated their superiority.

The present work has demonstrated that the average of the three scoring systems assessed was significantly higher for non survivors (APACHE II 15.6 vs 12.2; SOFA 7.8 vs 5.9; and CURB-65 2.1 vs 1.4) (Table 1). Other studies show similar findings, such as a multi-center cohort of ICU in Spain and Andorra that included 663 patients with critical COVID-19, in which the average for APACHE II was 17 vs 11 ( $p < 0.001$ )

and for SOFA 7 vs 4 ( $p < 0.001$ ) including non survivors and survivors.<sup>1</sup> Another study in Argentina, SATICOVID, with a cohort of 1909 patients undergoing invasive ventilatory support, registered a difference for APACHE II of 16 vs 13 ( $p < 0.0001$ ) and for SOFA 6 vs 4 ( $p < 0.0001$ ).<sup>11</sup>

A moderate discriminatory ability for death within 60 days was found in each of the systems, with an AUC for APACHE II of 0.70 (CI 0.64–0.75), SOFA 0.65 (CI 95% 0.60–0.71), and CURB-65 0.68 (CI 0.62–0.74). A cross-sectional study that included 1990 patients for a period of two years hospitalized in ICU in India showed that APACHE II had sensitivity of 89.9%, specificity of 97.6% and an AUC of 0.983 for mortality; while the results for SOFA were 90.1%, 96.6%, and 0.986, respectively. With regard to the necessity of mechanical ventilation, APACHE-II had sensitivity of 93.4%, specificity of 89.7% and an AUC of 0.966; SOFA had sensitivity of 90.5%, specificity of 95.8% and an AUC 0.976.<sup>12</sup> The report for a cohort of 204 patients hospitalized for COVID-19 in ICU at the hospital complex Imam Khomeini in Iran found that the average scores for APACHE II and SOFA were significantly higher for non survivors than for survivors (14.4 vs 9.5  $p \leq 0.001$ , and 7.3 vs 3.1  $p \leq 0.001$ , respectively); the AUC was 0.895 for SOFA and 0.730 for APACHE II. An APACHE II  $> 13$  had an adjusted OR of 10.7 (CI 5.0–22.9;  $p \leq 0.001$ ) while the SOFA  $> 5$  had 32.4 (CI 14.4–72.9;  $p \leq 0.001$ ).<sup>13</sup>

A study that included 249 patients hospitalized in ICU in Lithuania reported that the most accurate scoring system was APACHE II, with an AUC of 0.772 (CI 0.714–0.830;  $p < 0.001$ ), finding that a one point increase in the score increased the risk of mortality by 1155 (OR 1155, CI 95% 1085–1.229;  $p < 0.001$ ); SOFA had an AUC of 0.679 (CI 0.611–0.747;  $p < 0.001$ ).<sup>14</sup> Another cohort that included 140

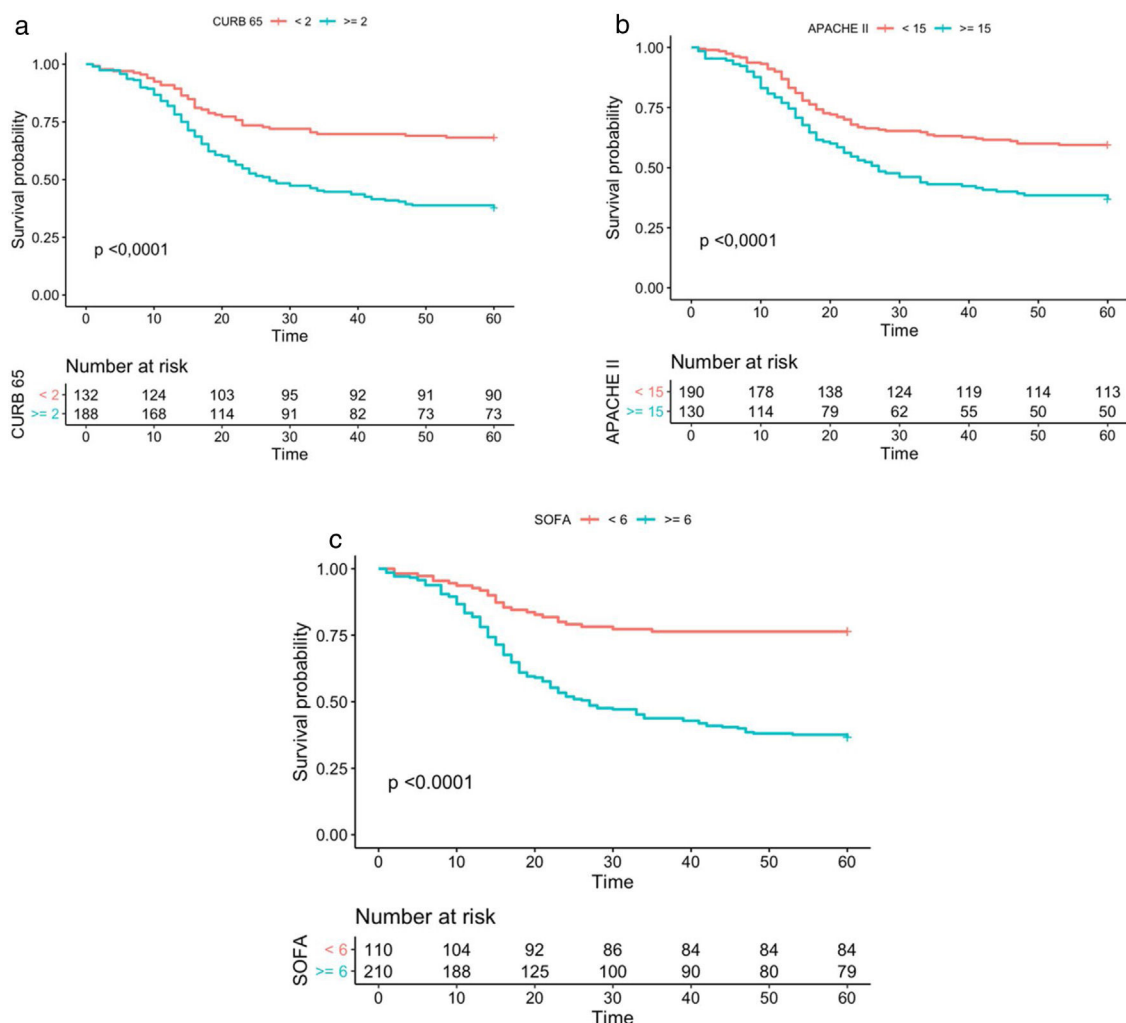
**Table 1** Clinical characteristics of the patients included in the study.

| Characteristic                                                                 | All patients<br>(n = 320) | Survivors<br>(n = 161) | Non survivors<br>(n = 159) | Value p          |
|--------------------------------------------------------------------------------|---------------------------|------------------------|----------------------------|------------------|
| <i>Females, n (%)</i>                                                          | 114 (35.6%)               | 64 (39.8%)             | 50 (31.4%)                 | 0.152            |
| <i>Age (years), average (SD)</i>                                               | 59.4 (14.9)               | 55.0 (14.4)            | 64.0 (14.1)                | <b>&lt;0.001</b> |
| <i>Obesity, number of patients with data (%)</i>                               | 134/290<br>(46.2%)        | 74/149 (49.7%)         | 60/141 (42.6%)             | 0.273            |
| <i>Comorbidities, n (%)</i>                                                    |                           |                        |                            |                  |
| Hypertension                                                                   | 128 (40.0%)               | 63 (39.1%)             | 65 (40.9%)                 | 0.908            |
| Diabetes                                                                       | 70 (21.9%)                | 34 (21.1%)             | 36 (22.6%)                 | 0.846            |
| Chronic heart disease (except hypertension)                                    | 41 (12.8%)                | 14 (8.7%)              | 27 (17.0%)                 | <b>0.040</b>     |
| Chronic renal disease                                                          | 15 (4.7%)                 | 6 (3.7%)               | 9 (5.7%)                   | 0.580            |
| Smoking                                                                        | 79 (24.7%)                | 44 (27.3%)             | 35 (22.0%)                 | 0.330            |
| Chronic pulmonary disease                                                      | 66 (20.6%)                | 27 (16.8%)             | 39 (24.5%)                 | 0.115            |
| Chronic neurological disease                                                   | 14 (4.4%)                 | 5 (3.1%)               | 9 (5.7%)                   | 0.400            |
| <i>Duration of the disease before hospitalization (days), median (SD)</i>      | 8.4 (4.2)                 | 8.4 (4.0)              | 8.4 (4.3)                  | 0.892            |
| <i>Laboratories</i>                                                            |                           |                        |                            |                  |
| Creatinine (mg/dL), average (SD)                                               | 2.4 (4.0)                 | 2.1 (3.5)              | 2.7 (4.4)                  | 0.203            |
| High sensitivity C-reactive protein (mg/L), average (SD)                       | 17.8 (14.7)               | 16.3 (15.8)            | 19.3 (13.4)                | 0.072            |
| Ferritin (ng/mL), average (SD)                                                 | 1167 (639)                | 1053 (605)             | 1294 (654)                 | <b>0.002</b>     |
| D-dimer (μg/mL), average (SD)                                                  | 4.4 (6.6)                 | 3.8 (6.2)              | 5.1 (6.9)                  | 0.084            |
| Ratio PaO <sub>2</sub> /FiO <sub>2</sub> , average (SD)                        | 103 (65)                  | 114 (75)               | 92 (51)                    | <b>0.002</b>     |
| Lactate dehydrogenase (U/L), average (SD)                                      | 1083 (1084)               | 897 (374)              | 1272 (1472)                | <b>0.002</b>     |
| High positive sensitivity cardiac troponin I, number of patients with data (%) | 127/304<br>(41.8%)        | 45/137 (32.8%)         | 72/139 (51.8%)             | <b>&lt;0.001</b> |
| <i>Complications and supports, (n, %)</i>                                      |                           |                        |                            |                  |
| Shock                                                                          | 214 (66.9%)               | 77 (47.8%)             | 137 (86.2%)                | <b>&lt;0.001</b> |
| Invasive mechanical ventilation requirement                                    | 254 (79.4%)               | 110 (68.3%)            | 144 (90.6%)                | <b>&lt;0.001</b> |
| Acute renal lesion                                                             | 162 (50.6%)               | 49 (30.4%)             | 113 (71.1%)                | <b>&lt;0.001</b> |
| Renal replacement therapy                                                      | 81 (25.3%)                | 16 (9.9%)              | 65 (40.9%)                 | <b>&lt;0.001</b> |
| Liver disease                                                                  | 53 (16.6%)                | 14 (8.7%)              | 39 (24.5%)                 | <b>&lt;0.001</b> |
| Coagulopathy                                                                   | 43 (13.4%)                | 9 (5.6%)               | 34 (21.4%)                 | <b>&lt;0.001</b> |
| Nervous central system compromise                                              | 29 (9.1%)                 | 13 (8.1%)              | 16 (10.1%)                 | 0.67             |
| Venous thrombosis                                                              | 27 (8.4%)                 | 10 (6.2%)              | 17 (10.7%)                 | 0.215            |
| <i>Duration of hospital stay (days), median (SD)</i>                           | 23.4 (15.9)               | 28.2 (17.9)            | 18.6 (11.8)                | <b>&lt;0.001</b> |
| <i>Scoring systems for risk prediction upon admission</i>                      |                           |                        |                            |                  |
| APACHE II, average (SD)                                                        | 13.9 (6.7)                | 12.2 (5.9)             | 15.6 (7.1)                 | <b>&lt;0.001</b> |
| SOFA, average (SD)                                                             | 6.8 (3.3)                 | 5.9 (3.4)              | 7.8 (2.9)                  | <b>&lt;0.001</b> |
| CURB-65, average (SD)                                                          | 1.8 (1.0)                 | 1.4 (0.9)              | 2.1 (1.0)                  | <b>&lt;0.001</b> |

SD: standard deviation, APACHE II: Acute Physiology and Chronic Health Evaluation, SOFA: Sequential Organ Failure Assessment Score, CURB-65: Confusion, Blood urea nitrogen, Respiratory rate, Systolic BP and Age > or = 65.

patients with COVID-19 in critical condition found an AUC of 0.890 (CI 0.826–0.955)<sup>15</sup> for SOFA in the prediction of mortality, an analysis using the same data shows that a SOFA score  $\geq 3$  is a moderately good positive predictor (LR +5.35; CI 3.43–8.36) and a strong negative predictor (LR –0.12; CI 0.03–0.45) for in-hospital mortality.<sup>16</sup> These positive findings contrast the reports that show the opposite results; a retrospective study that included 675

patients with mechanical ventilation in ICU for COVID-19 found an AUC for SOFA of 0.59 (CI 0.55–0.63) for death,<sup>17</sup> similar to the result described in an extensive US cohort of 5122 patients with ventilation in 86 ICUs, which had an AUC of 0.66 (CI 0.65–0.67) showing that the SOFA score itself had deficient discriminatory accuracy for in-hospital mortality.<sup>18</sup> Furthermore, a study that modeled the performance of SOFA to guide the indication of



**Figure 3** Kaplan-Meier curves for the cut-off levels obtained for each scoring system (a) CURB-65, (b) APACHE II, and (c) SOFA.

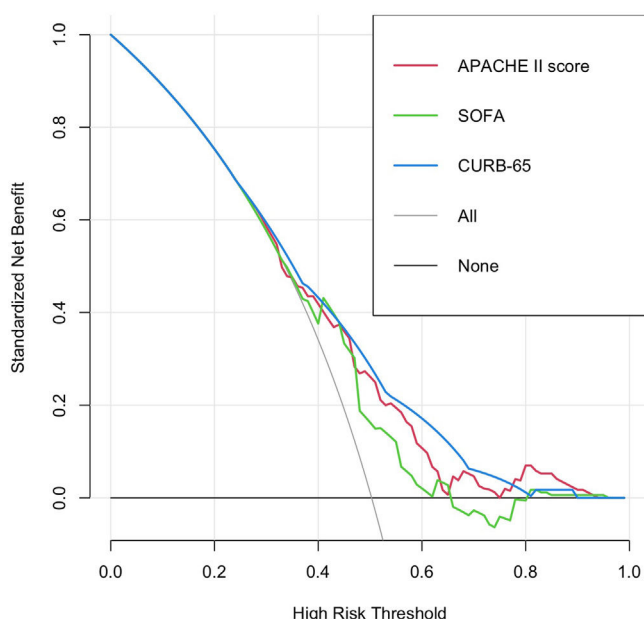
mechanical ventilation found a limited utility for decision making.<sup>19</sup>

Upon assessing cohorts of patients hospitalized in general ward, reports with similar success rates can be found. The Brazilian multi center cohort RECOVER-SUS, that included 1589 patients hospitalized for COVID-19, showed that a SOFA score  $\geq 10$  was one of the factors associated in an independent manner with in-hospital mortality (HR of 1.51; CI 1.08–2.10).<sup>20</sup> In a retrospective observational study that included 237 adults hospitalized for COVID-19 outside ICU, a SOFA score  $\geq 2$  had an AUC of 0.732 (0.67–0.79) with sensitivity of 83.33%, specificity of 65.43%, PPV of 38.1% and NPV of 93.84% to predict mortality within 28 days.<sup>21</sup>

A register that included 1363 patients from Sao Paulo and Barcelona hospitalized for COVID-19 found for CURB-65 an AUC of 0.74 (CI 0.72–0.77) for in-hospital death.<sup>22</sup> An extensive Spanish multi-center register called SEMI-COVID-19, which included 10238 patients hospitalized for COVID-19, reported an AUC of 0.825 for CURB-65 for in-hospital mortality, admission to ICU, and use of mechanical ventilation; there was sensitivity of 82.1% (CI 80.4–83.8), specificity of 70.6% (CI 69.6–71.6), PPV of 42.2% (CI 40.6–43.7) and NPV of 93.8% (CI 93.2–94.4) for a score  $> 2$ .<sup>23</sup> A study

that compared 682 patients hospitalized for COVID-19 in eight hospitals for adults with 7449 community-acquired pneumonia cases in Louisville, Kentucky, found that CURB-65 had an AUC of 0.79 (CI 95%, 0.75–0.84) and 0.75 (CI 95%, 0.73–0.77), respectively.<sup>24</sup> A study that included 481 patients from Turkey found that CURB-65 had better performance to predict in-hospital mortality and ICU requirement in patients hospitalized for COVID-19 (AUC of 0.846 and 0.898, respectively).<sup>25</sup>

A non-systematic review that assessed the scoring systems with respect to mortality within 90 days in patients hospitalized for COVID-19, which included 76 studies, found that the score for APACHE II had the highest AUC values (0.966 and 0.937 in two studies), followed by SOFA (with 0.915, 0.926 and 0.876 in three studies). This may be explained by the fact that both age and comorbidities are taken into consideration, which are components recognized as prognostic markers for COVID-19.<sup>26</sup> In a cohort of 247 patients infected with COVID-19 in Ecuador, the utility of the CURB-65 score  $\geq 2$  was assessed to predict mortality within 30 days, obtaining sensitivity of 84%, specificity of 54%, PPV of 56% and NPV of 83%, concluding that CURB-65 is adequate to predict mortality in this group of patients.<sup>27</sup>



**Figure 4** Decision analysis curves for the three scoring systems.

We can mention as limitations of the present study the size of the population since a calculation of the sample size was not carried out. We cannot rule out a selection bias when including only patients with complete data for the calculation of clinical rules on admission. A limitation of our study is that we have data for the SOFA score only at admission, and we lack daily or periodic measurements of its variables over the time. Although the present work has a unicenter character, it can be considered representative of a wide population since this reference center is located in the South of Bogotá and covers about two million people.

In conclusion, the scoring systems for clinical prediction CURB-65, APACHE II, and SOFA, used with patients hospitalized for COVID-19, showed a moderate discriminatory ability for death within 60 days (AUC 0.68, 0.70 and 0.65, respectively) and demonstrated that for a maximum discrimination cut-off value (via Youden test) these scoring systems have an adequate prediction ability, which was confirmed through the Kaplan-Meier curves and the clinical decision curves.

## Authors' contributions

JS, AM and JP conceived and designed the study. JS, AM and OQ performed the statistical analysis. All authors participated in the collection and analysis of information, as well as the writing, revision, and approval of the final manuscript.

## Availability of data and materials

All data generated and/or analyzed during this study are available from the corresponding author on reasonable request conditioned by its review by the institutional ethics and research committee.

## Ethics approval and consent to participate

The study was approved by the ethics and research committees of each of the institutions (act number 138 of the Health Services Unit from the El Tunal Hospital, 0498-2020 of the San José Hospital and SDM-026-20 of the University Children's Hospital San Jose) and informed consent was not required.

## Consent for publication

Not applicable.

## Funding

No funding was received for the creation, development, and analysis of the data.

## Conflict of interests

The authors declare that they have no competing interests.

## Acknowledgements

We thank the relatives of the participants for their support. We also want to thank all the health and non-health care personnel involved in the care of COVID-19 patients.

## References

1. Ferrando C, Mellado-Artigas R, Gea A, Arruti E, Aldecoa C, Bordell A, et al. Patient characteristics, clinical course and factors associated to ICU mortality in critically ill patients infected with SARS-CoV-2 in Spain: a prospective, cohort, multicentre study. *Rev Esp Anestesiología y Reanimación*. 2020;67:425–37.
2. Plotnikow GA, Matesa A, Nadur JM, Alonso M, Nuñez II, Vergara G, et al. Características y resultados de los pacientes infectados con nCoV19 con requerimiento de ventilación mecánica invasiva en la Argentina. *Rev Bras Ter Intensiva*. 2020;32:348–53.
3. Chica-Meza C, Peña-López LA, Villamarín-Guerrero HF, Moreno-Collazos JE, Rodríguez-Corredor LC, Lozano WM, et al. Cuidado respiratorio en COVID-19. *Acta Colomb Cuid Intensivo*. 2020;20:108–17.
4. Wa K, Ea D, Dp W, Je Z. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13. Available from: <https://pubmed.ncbi.nlm.nih.gov/3928249/> [cited 3.12.22].
5. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996;22:707–10.
6. 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.
7. Loke YK, Kwok CS, Niruban A, Myint PK. Value of severity scales in predicting mortality from community-acquired pneumonia: systematic review and meta-analysis. *Thorax*. 2010;65:884–90.
8. Chalmers JD, Singanayagam A, Akram AR, Mandal P, Short PM, Choudhury G, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired

- pneumonia. Systematic review and meta-analysis. *Thorax*. 2010;65:878–83.
9. Richards G, Levy H, Laterre PF, Feldman C, Woodward B, Bates BM, et al. CURB-65 PSI, and APACHE II to assess mortality risk in patients with severe sepsis and community acquired pneumonia in PROWESS. *J Intensive Care Med*. 2011;26:34–40.
  10. Asai N, Watanabe H, Shiota A, Kato H, Sakanashi D, Hagihara M, et al. Efficacy and accuracy of qSOFA and SOFA scores as prognostic tools for community-acquired and healthcare-associated pneumonia. *Int J Infect Dis*. 2019;84:89–96.
  11. Estenssoro E, Loudet CI, Rios FG, Kanoore Edul VS, Plotnikow G, Andrian M, et al. Clinical characteristics and outcomes of invasively ventilated patients with COVID-19 in Argentina (SATI-COVID): a prospective, multicentre cohort study. *Lancet Respir Med*. 2021;9:989–98.
  12. Kumar S, Gattani SC, Baheti AH, Dubey A. Comparison of the performance of APACHE II SOFA, and mNUTRIC scoring systems in critically ill patients: a 2-year cross-sectional study. *Indian J Crit Care Med*. 2020;24:1057–61.
  13. Beigmohammadi MT, Amoozadeh L, Rezaei Motlagh F, Rahimi M, Maghsoudloo M, Jafarnejad B, et al. Mortality predictive value of APACHE II and SOFA scores in COVID-19 patients in the intensive care unit. *Can Respir J*. 2022;2022:5129314.
  14. Vicka V, Januskeviciute E, Miskinyte S, Ringaitiene D, Serpytis M, Klimasauskas A, et al. Comparison of mortality risk evaluation tools efficacy in critically ill COVID-19 patients. *BMC Infect Dis*. 2021;21:1173.
  15. Liu S, Yao N, Qiu Y, He C. Predictive performance of SOFA and qSOFA for in-hospital mortality in severe novel coronavirus disease. *Am J Emerg Med*. 2020;38:2074–80.
  16. Wang DX, Wang MK, Rochweg B, Burns KEA. Predictive performance of SOFA & qSOFA for in-hospital mortality in patients with severe novel coronavirus disease. *Am J Emerg Med*. 2021;50:793–4.
  17. Raschke RA, Agarwal S, Rangan P, Heise CW, Curry SC. Discriminant accuracy of the SOFA score for determining the probable mortality of patients with COVID-19 pneumonia requiring mechanical ventilation. *JAMA*. 2021;325:1469–70.
  18. Keller MB, Wang J, Nason M, Warner S, Follmann D, Kadri SS. Preintubation sequential organ failure assessment score for predicting COVID-19 mortality: external validation using electronic health record from 86 U.S. healthcare systems to appraise current ventilator triage algorithms. *Crit Care Med*. 2022;50:1051–62.
  19. Cuartas PA, Tavares Santos H, Levy BM, Gong MN, Powell T, Chuang E. Modeling outcomes Using Sequential Organ Failure Assessment (SOFA) score-based ventilator triage guidelines during the COVID-19 pandemic. *Disaster Med Public Health Prep*. 2022;1–11.
  20. Perazzo H, Cardoso SW, Ribeiro MPD, Moreira R, Coelho LE, Jalil EM, et al. In-hospital mortality and severe outcomes after hospital discharge due to COVID-19: a prospective multicenter study from Brazil. *Lancet Reg Health Am*. 2022;11:100244.
  21. Lalueza A, Lora-Tamayo J, de la Calle C, Sayas-Catalán J, Arrieta E, Maestro G, et al. Utilidad de las escalas de sepsis para predecir el fallo respiratorio y la muerte en pacientes con COVID-19 fuera de las Unidades de Cuidados Intensivos. *Rev Clin Esp*. 2022;222:293–8.
  22. Lazar Neto F, Marino LO, Torres A, Cilloniz C, Meirelles Marchini JF, Garcia de Alencar JC, et al. Community-acquired pneumonia severity assessment tools in patients hospitalized with COVID-19: a validation and clinical applicability study. *Clin Microbiol Infect*. 2021;27:1037.e1–8, <http://dx.doi.org/10.1016/j.cmi.2021.03.002>.
  23. Artero A, Madrazo M, Fernández-Garcés M, Muño Miguez A, González García A, Crestelo Vieitez A, et al. Severity scores in COVID-19 pneumonia: a multicenter, retrospective, cohort study. *J Gen Intern Med*. 2021;36:1338–45.
  24. Bradley J, Sbaih N, Chandler TR, Furmanek S, Ramirez JA, Cavallazzi R. Pneumonia severity index and CURB-65 score are good predictors of mortality in hospitalized patients with SARS-CoV-2 community-acquired pneumonia. *Chest*. 2022;161:927–36.
  25. Doğanay F, Ak R. Performance of the CURB-65 ISARIC-4C and COVID-GRAM scores in terms of severity for COVID-19 patients. *Int J Clin Pract*. 2021;75:e14759.
  26. Chu K, Alharahsheh B, Garg N, Guha P. Evaluating risk stratification scoring systems to predict mortality in patients with COVID-19. *BMJ Health Care Inform*. 2021;28:e100389.
  27. Carriel J, Muñoz-Jaramillo R, Bolaños-Ladinez O, Heredia-Villacreses F, Menéndez-Sanchón J, Martín-Delgado J. CURB-65 como predictor de mortalidad a 30 días en pacientes hospitalizados con COVID-19 en Ecuador: estudio COVID-EC. *Rev Clin Esp*. 2022;222:37–41.