



# 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

## Validation externe du score de mortalité ISARIC 4C pour prédire la mortalité à l'hôpital chez les personnes atteintes de COVID-19 dans une unité de soins intensifs canadienne : une étude de cohorte historique monocentrique

Tharaniya Vallipuram, BScN · Blair C. Schwartz, MDCM, MHS, FRCPC · Stephen S. Yang, MDCM, MSc, FRCPC · Dev Jayaraman, MDCM, FRCPC · Sandra Dial, MD, MSc, FRCPC

Received: 3 August 2022/Revised: 19 December 2022/Accepted: 31 December 2022/Published online: 8 June 2023  
© Canadian Anesthesiologists' Society 2023

### Abstract

**Purpose** With uncertain prognostic utility of existing predictive scoring systems for COVID-19-related illness, the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) 4C Mortality Score was developed by the International Severe Acute Respiratory and Emerging Infection Consortium as a COVID-19 mortality prediction tool. We sought to externally validate this score among critically ill patients admitted to an intensive care unit (ICU) with COVID-19 and compare its discrimination characteristics to that of the Acute Physiology and Chronic Health Evaluation (APACHE) II and Sequential Organ Failure Assessment (SOFA) scores.

**Methods** We enrolled all consecutive patients admitted with COVID-19-associated respiratory failure between 5 March 2020 and 5 March 2022 to our university-affiliated and intensivist-staffed ICU (Jewish General Hospital, Montreal, QC, Canada). After data abstraction, our primary outcome of in-hospital mortality was evaluated with an objective of determining the discriminative properties of the ISARIC 4C Mortality Score, using the area under the curve of a logistic regression model.

**Results** A total of 429 patients were included, 102 (23.8%) of whom died in hospital. The receiver operator curve of the ISARIC 4C Mortality Score had an area under the curve of 0.762 (95% confidence interval [CI], 0.717 to 0.811), whereas those of the SOFA and APACHE II scores were 0.705 (95% CI, 0.648 to 0.761) and 0.722 (95% CI, 0.667 to 0.777), respectively.

**Conclusions** The ISARIC 4C Mortality Score is a tool that had a good predictive performance for in-hospital mortality in a cohort of patients with COVID-19 admitted to an ICU for respiratory failure. Our results suggest a good external validity of the 4C score when applied to a more severely ill population.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12630-023-02512-4>.

T. Vallipuram, BScN  
Faculty of Medicine, McGill University, Montreal, QC, Canada

B. C. Schwartz, MDCM, MHS, FRCPC (✉) ·  
S. S. Yang, MDCM, MSc, FRCPC ·  
D. Jayaraman, MDCM, FRCPC · S. Dial, MD, MSc, FRCPC  
Division of Critical Care, Jewish General Hospital, McGill  
University, Pavilion H-364.1, 3755 Chemin de la Côte-Sainte-  
Catherine, Montreal, QC H3T 1E2, Canada  
e-mail: blair.schwartz@mcgill.ca

### Résumé

**Objectif** Compte tenu de l'utilité pronostique incertaine des systèmes de notation prédictive existants pour les maladies liées à la COVID-19, le score de mortalité ISARIC 4C a été mis au point par l'International Severe

*Acute Respiratory and Emerging Infection Consortium en tant qu'outil de prédiction de la mortalité associée à la COVID-19. Nous avons cherché à valider en externe ce score chez les patient-es gravement malades atteint-es de COVID-19 admis-es dans une unité de soins intensifs (USI) et à comparer ses caractéristiques de discrimination à celles des scores APACHE II (Acute Physiology and Chronic Health Evaluation) et SOFA (Sequential Organ Failure Assessment).*

**Méthode** Nous avons recruté toutes les personnes consécutives admises pour insuffisance respiratoire associée à la COVID-19 entre le 5 mars 2020 et le 5 mars 2022 dans notre unité de soins intensifs affiliée à l'université et dotée d'intensivistes (Hôpital général juif, Montréal, QC, Canada). Après l'abstraction des données, notre critère d'évaluation principal de mortalité à l'hôpital a été évalué dans le but de déterminer les propriétés discriminatives du score de mortalité ISARIC 4C, en utilisant la surface sous la courbe d'un modèle de régression logistique.

**Résultats** Au total, 429 patient-es ont été inclus-es, dont 102 (23,8 %) sont décédés à l'hôpital. La fonction d'efficacité du récepteur (courbe ROC) du score de mortalité ISARIC 4C avait une surface sous la courbe de 0,762 (intervalle de confiance [IC] à 95 %, 0,717 à 0,811), tandis que celles des scores SOFA et APACHE II étaient de 0,705 (IC 95%, 0,648 à 0,761) et 0,722 (IC 95%, 0,667 à 0,777), respectivement.

**Conclusion** Le score de mortalité ISARIC 4C est un outil qui a affiché une bonne performance prédictive de la mortalité à l'hôpital dans une cohorte de patient-es atteint-es de COVID-19 admis-es dans une unité de soins intensifs pour insuffisance respiratoire. Nos résultats suggèrent une bonne validité externe du score 4C lorsqu'il est appliqué à une population plus gravement malade.

**Keywords** COVID-19 · external validation · intensive care units · ISARIC 4C Mortality Score · prediction model · respiratory failure

Since its emergence in December 2019, COVID-19, caused by SARS-CoV-2, has been responsible for over 763 million confirmed cases and at least 6.9 million cumulative deaths worldwide as of 25 April 2023.<sup>1</sup> The wide clinical spectrum of COVID-19 ranges from asymptomatic or mild infection to critical illness leading to severe respiratory failure, requiring admission to the intensive care unit (ICU).

Predictive scoring systems allow clinicians to estimate a variety of clinical outcomes, including mortality, that may

facilitate decision-making. Existing scores such as the Acute Physiology and Chronic Health Evaluation (APACHE II) and Sequential Organ Failure Assessment (SOFA) scoring systems are prognostic tools that are widely used within the ICU to assess the risk of mortality in critically ill patients.<sup>2-4</sup> These scores were initially developed for general critical care admission, so their prognostic utility for COVID-19-related critical illness is uncertain. The few studies that have attempted to explore the role of these scores in COVID-19 disease are limited by small sample sizes and variable conclusions and were conducted immediately at the onset of the pandemic, thus may not be reflective of current therapeutic options.<sup>5-8</sup>

The 4C Mortality Score was developed by the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) to predict mortality among patients admitted to the hospital with COVID-19.<sup>9</sup> The variables used in the ISARIC 4C Mortality Score include patient demographics, clinical observations such as vital signs, and common laboratory values available at hospital admission.<sup>9</sup> In contrast, the APACHE II score requires 12 physiologic variables in addition to the patient's age and chronic health status, while the SOFA score includes variables such as partial pressure of arterial oxygen/fraction of inhaled oxygen (PaO<sub>2</sub>/F<sub>I</sub>O<sub>2</sub>) that are not always available or easily attainable in this population.<sup>2-4</sup> While there are clear advantages in the ease of use of the 4C Mortality Score, there are currently limited data on whether this disease-specific risk stratification tool can predict mortality in populations with more severe COVID-19 who are being treated in the ICU, since it was developed and validated in a general hospital population.<sup>9</sup>

The primary objective of this historical cohort study was to externally validate the ISARIC 4C Mortality Score as an effective predictive scoring system among critically ill patients admitted to a Canadian ICU with COVID-19. The secondary objective was to compare the 4C Mortality Score's discriminative ability with that of the APACHE II and SOFA scoring systems for in-hospital mortality.

## Methods

### *Study design and setting*

This historical cohort study was conducted at the ICU of the Jewish General Hospital in Montreal, QC, Canada. The medical-surgical ICU is located within a university-affiliated tertiary care hospital and operates using the closed, intensivist-led model.

### Ethics

This study was approved by the institutional Research Ethics Committee of the Centre Intégré Universitaire en Santé et Services Sociaux (CIUSSS) West-Central Montreal, Jewish General Hospital; Study Identifier 2021-2362; approved on 3 February 2022. Due to the nature of the study and anonymous data collection, the need for informed consent was waived.

### Cohort assembly

At the onset of the pandemic, a clinical database was prospectively established for all patients with a positive nasopharyngeal polymerase chain reaction test result for SARS-CoV-2 admitted to the ICU at the Jewish General Hospital, either directly from our emergency department, COVID-19 wards, or another institution. Our institution was designated as the regional ICU for COVID-19 admissions in the first wave of the pandemic, and transfers occurred within 24 hr of arrival at the referring centre. All patients listed in this database between 5 March 2020 and 5 March 2022 were eligible for inclusion in this study. No patients were excluded from the cohort based on pre-existing limitations of care. Patients were excluded if their COVID-19 status was determined to be incidental to the indication for ICU admission (i.e., not causing respiratory failure), as adjudicated independently by two team members (S. D. and B. S.).

### Medical care

All elements of medical care were left to the discretion of the treating intensivist. Patients with pre-established care directives available at the time of ICU admission that limited life-sustaining therapies, specifically endotracheal intubation, were still recipients of other forms of respiratory support such as high-flow nasal oxygen therapy or noninvasive ventilation. They are referred to as “not for intubation” in this manuscript. Moreover, decisions regarding therapies, including corticosteroid dose and timing, were left to the discretion of the treating intensivist, though we did have an institutional protocol to use high-dose steroids (dexamethasone 20 mg *iv* daily for five days, followed by 10 mg *iv* daily for five days) for patients on greater than 70%  $F_{I}O_2$  during the first wave.<sup>10</sup>

### Data sources and study variables

Each patient’s complete electronic medical record (ChartMaxx<sup>®</sup> version 7.00; Quest Diagnostics<sup>®</sup> Incorporated, Secaucus, NJ, USA) was reviewed and relevant data were extracted into an encrypted

computerized spreadsheet. The data were extracted independently by two team members (S. D. and T. V.). Resulting data were compared, and discrepancies rectified using a consensus process, with a goal of ensuring valid and high-quality data extraction. Additionally, the ICU admission note was abstracted for demographic information, including age, sex, and patient comorbidities. Patient vital signs, supplemental oxygenation treatments, and laboratory values were extracted from the critical care flowsheet and electronic laboratory system (Open Architecture Clinical Information System [Oacis], Telus Health, Montreal, QC, Canada). Information regarding limitations to therapy was extracted from care directives available at the time of ICU admission. Treatments received were determined from the medication administration records, and other clinical outcomes, such as length of stay, intubation, and mortality were abstracted from the clinical progress notes. Extracted data were then used to calculate the 4C Mortality, APACHE II, and SOFA scores for each patient in our cohort. The variables within the 4C Mortality Score include age, sex, number of comorbidities from the Charlson index, respiratory rate, peripheral oxygen saturation on room air, Glasgow Coma Scale, serum urea, and level of C-reactive protein. The 4C Mortality Score was calculated using the worst value of the variables within the first 24-hr period of hospital admission. In cases of hospital transfer, data from the external institution was used.

The SOFA score and the APACHE II score were calculated for each patient based on the worst variables obtained within the first 24-hr period of ICU admission. The respiratory component of the SOFA score requires an arterial blood gas value. Nevertheless, many patients in our cohort did not have an arterial blood gas value available from the first 24-hr window of ICU admission, leading to missing  $PaO_2/F_{I}O_2$  variables. Therefore, the respiratory and total SOFA score for all cohort patients was calculated using a validated technique,<sup>11</sup> which includes using the imputed  $SpO_2/F_{I}O_2$  ratio in place of  $PaO_2/F_{I}O_2$ . The  $SpO_2/F_{I}O_2$  ratio for these patients was determined by averaging the three lowest  $SpO_2$  values in the first 24 hr of ICU admission and taking the highest delivered  $F_{I}O_2$  among these three  $SpO_2$  values to calculate the imputed  $SpO_2/F_{I}O_2$  ratio.

### Outcomes

The primary outcome of our study was in-hospital mortality, with the objective of evaluating the ISARIC 4C Mortality Score’s model calibration and discrimination for predicting in-hospital mortality in a cohort of intensive care patients.

### Missing data

In the event that a variable required for calculating the ISARIC 4C Mortality Score was missing, the median value for this variable from our cohort was imputed.

### Statistical analysis

Descriptive statistics were used to summarize baseline demographic and clinical characteristics. Continuous variables are presented as median [interquartile range (IQR)] values because of data distribution. Categorical variables are presented as count and percentage unless stated otherwise.

We conducted comparisons between groups using the Wilcoxon rank-sum test for nonnormally distributed continuous data and the Chi square test for categorical variables. All analyses were conducted in a two-tailed fashion with statistical significance set at  $P < 0.05$  and were performed using SAS 9.4 (SAS Institute Inc., Cary,

NC, USA), STATA/MP version 15 (StataCorp LLC, College Station, TX, USA), and R version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria) with the *-rms-* package used in R.

We constructed a univariate logistic regression analysis to examine each mortality prediction score and the outcome of interest, in-hospital mortality. We use the Wald test to estimate 95% confidence intervals (CIs).

To validate the ISARIC 4C Mortality Score model, we measured its discriminative properties and its calibration within our cohort. Calibration compares the proportion of observed patients who had in-hospital mortality against the expected proportion based on the 4C Mortality Score model. Discrimination assessed the ability of the model to correctly distinguish between hospital decedents vs survivors, as assessed by the C-statistic. Calibration of our cohort model to the predicted model was assessed using calibration at large, examining the slope and intercept of the calibration model.

**Table 1** Demographic and clinical characteristics of the overall cohort, and comparison between survivors and patients who died

Variable	Total N = 429	Survivors N = 327	Died N = 102
Age (yr)	67 [56–75]	64 [53–72]	75 [69–83]
Age > 80 yr	64/429 (14.9%)	28/327 (8.6%)	36/102 (35%)
Resides in assisted living facility	24/429 (5.6%)	13/327 (3.9%)	14/102 (14%)
Male	271/429 (63.2%)	205/327 (62.7%)	66/102 (65%)
Pregnant	14/429 (3.3%)	14/327 (4.3%)	0
Charlson Comorbidity Score	-	3 [1–4]	5 [4–6]
Not for intubation	50/429 (11.7%)	16/327 (4.9%)	34/102 (33%)
Vaccinated ( $\geq 2$ doses)	16/429 (3.7%)	7/327 (2.1%)	9/102 (9%)
Time from hospitalization to ICU admission (days)	1 [0–2]	1 [0–2]	1 [0–2]
Highest F <sub>I</sub> O <sub>2</sub> (%) received in first 24 hours	70 [50–100]	70 [50–90]	85 [60–100]
24-hr SpO <sub>2</sub> /F <sub>I</sub> O <sub>2</sub> ratio	125 [93–171]	131 [101–178]	103 [89–150]
Intubated in first 24 hours	100/429 (23.3%)	66/327 (20.2%)	34/102 (33%)
Total intubated	168/429 (39.2%)	102/327 (31.2%)	66/102 (65%)
HFNO	312/429 (72.7%)	238/327 (72.8%)	74/102 (73%)
ECMO	6/429 (1.4%)	5/327 (1.5%)	1/102 (1.0%)
Corticosteroids	378/429 (88.1%)	288/327 (88.1%)	90/102 (88%)
High dose*	295/429 (68.8%)	220/327 (67.3%)	75/102 (74%)
Remdesivir	93/429 (21.7%)	74/327 (22.6%)	19/102 (19%)
IL-6 antagonist	130/429 (30.3%)	110/327 (33.6%)	24/102 (24%)
ISARIC 4 C Mortality Score	12 [9–14]	11 [9–14]	14 [12–16]
APACHE II score	13 [10–19]	12 [9–16]	18 [13–23]
SOFA score	4 [3–6]	4.0 [3–5]	5.0 [4–8]

Values are *n*/total *N* (%) or median [interquartile range]

\*Dexamethasone 20 mg *iv* daily for 5 days, followed by 10 mg *iv* daily for 5 days

APACHE = Acute Physiology and Chronic Health Evaluation; CRP = C-reactive protein; ECMO = extracorporeal membrane oxygenation; F<sub>I</sub>O<sub>2</sub> = fraction of inspired oxygen; HFNO = high-flow nasal oxygen; IL-6 = interleukin 6; ISARIC = International Severe Acute Respiratory and Emerging Infection Consortium; SOFA = Sequential Organ Failure Assessment; SpO<sub>2</sub> = oxygen saturation by pulse oximetry

We conducted prespecified sensitivity analyses to explore the stability of the model's discrimination of the 4C Mortality Score with the exclusion of the following individuals: 1) patients > 80 yr old; 2) patients admitted 24 hr or more post calculation of their 4C Mortality Score; and 3) patients with "not for intubation" status.

Furthermore, we constructed a multivariable logistic regression model to adjust for the effect of "not for intubation" status on the predictive ability of the 4C Mortality Score for in-hospital mortality. Internal validation of this model was completed using a bootstrapping technique of 1,000 samples to determine its calibration and discrimination and correct for optimism.

Finally, we measured and compared the discriminative properties of the 4C Mortality Score, the APACHE score, and the SOFA score within our cohort.

## Results

During our study period, 459 patients were admitted to the ICU with a diagnosis of COVID-19, of whom 30 (6.5%) were excluded from this study because of incidental COVID-19 infection at the time of ICU admission. Of 429 patients in our study, 211 (49.2%) were admitted from our emergency department, 186 (43.4%) from our COVID-19 medical ward, and 32 (7.5%) were transferred from the emergency department of another hospital within 24 hr of

arrival. Patients spent a median of 1 day [IQR, 0 to 2] on the COVID-19 medical ward prior to ICU transfer.

The clinical and demographic characteristics of the overall cohort and a breakdown between survivors and patients who died is shown in Table 1.

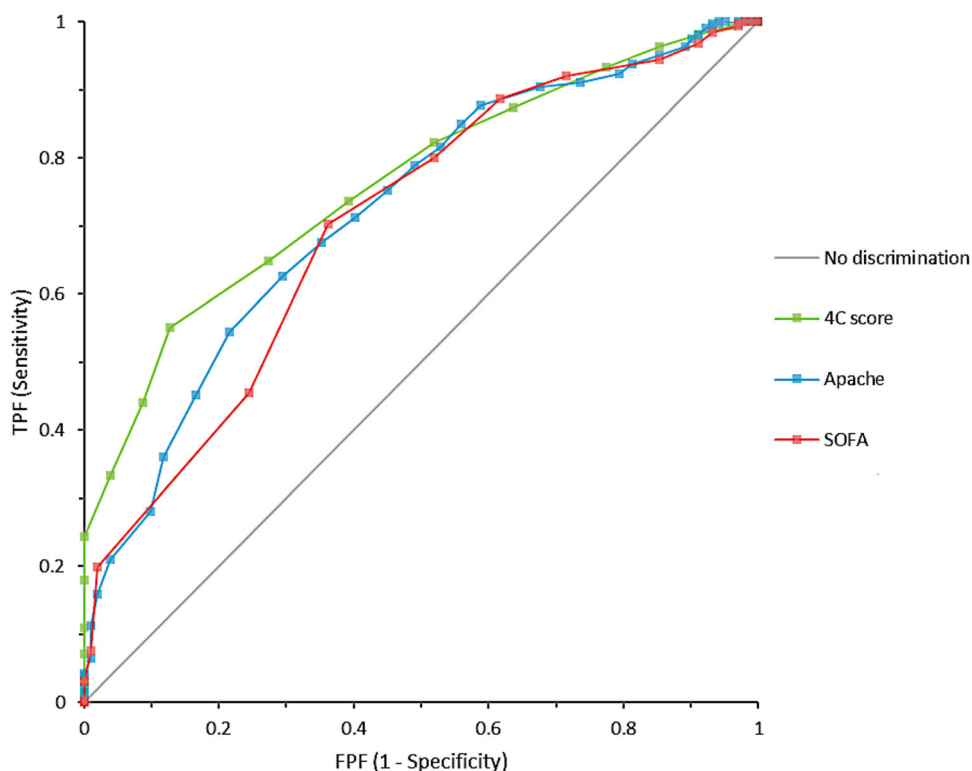
In two (0.5%) patients, C-reactive protein was not measured at admission. No other variables were missing.

The primary outcome of in-hospital mortality occurred in 102 (23.8%) patients in our cohort. Patients who died were noted to be older, be more hypoxemic on ICU admission, and have more underlying comorbidities compared with survivors.

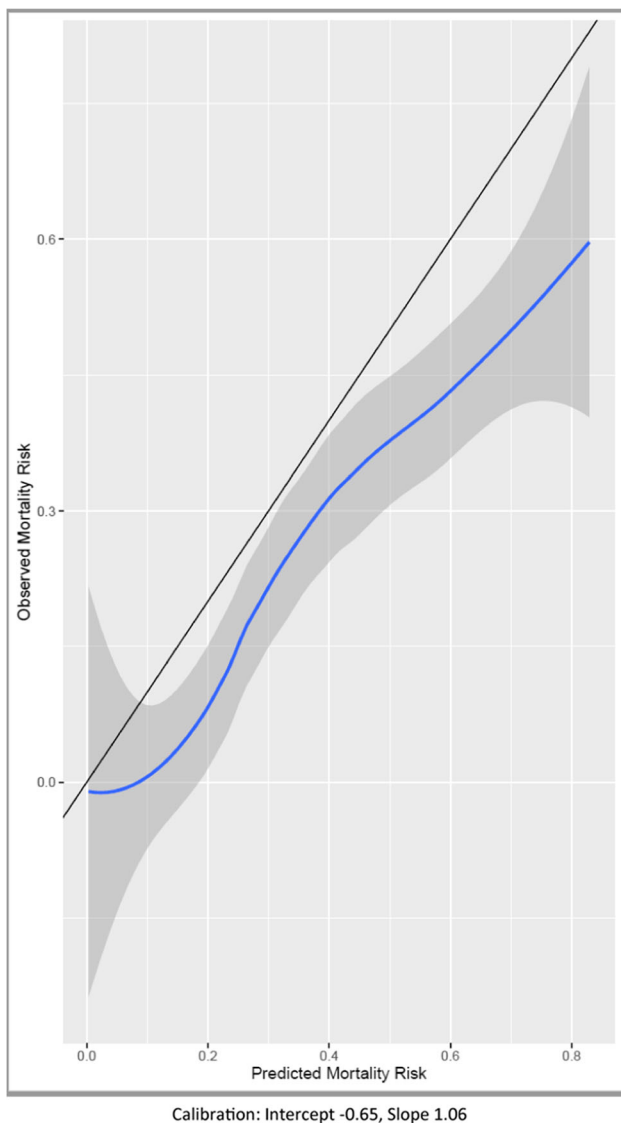
Five percent of our patients had a 4C Mortality Score between 0 and 3, and 17.4% had a score of 4–8; both these subgroups had 0 deaths. The 242 (56.4%) patients with 4C Mortality Scores between 9 and 14 had a mortality rate of 21.9% whereas 107 (24.9%) of our patients had scores  $\geq 15$  and a mortality rate of 45.8%. This is comparable with the original validation cohort in which 52.2% of patients had scores between 9 and 14 and 18.6% had scores  $\geq 15$ , with mortality rates of 31.4% and 61.5%, respectively.<sup>9</sup>

The fitted receiver operator characteristic curve of the ISARIC 4C Mortality Score had an area under the curve (AUC) of 0.762 (95% CI, 0.717 to 0.811) for in-hospital mortality in our cohort. The SOFA score had an AUC of 0.705 (95% CI, 0.648 to 0.761) and the APACHE score had an AUC of 0.722 (95% CI, 0.667 to 0.777) for the prediction of in-hospital mortality (Fig. 1). Calibration of

**Fig. 1** Model performance comparison for discrimination of in-hospital mortality using receiver operating characteristics curves for the ISARIC 4C Mortality Score, the APACHE II score, and the SOFA score. APACHE = Acute Physiology and Chronic Health Evaluation; ISARIC = International Severe Acute Respiratory and Emerging Infection Consortium; SOFA = Sequential Organ Failure Assessment







**Fig. 2** Calibration plot for cohort mortality (observed) vs ISARIC 4C Mortality Score (predicted). The black line represents ideal fit; the blue line represents logistic calibration with 95% confidence interval bands.

the ISARIC 4C Mortality Score appeared adequate, with a calibration slope of 1.06 and an intercept of -0.65 (Fig. 2).

Sensitivity analyses suggested that the 4C Mortality Score had stable discriminatory ability across subsets of data that excluded patients: > 80 yr old, admitted 24 hr or more post calculation of their 4C Mortality Score, and those with “not for intubation” status (Table 2).

The 50 (11.7%) patients that were “not for intubation” were noted to be significantly older with median age 83 [74–87] vs 65 [54–72] and had higher median 4C scores at 15 [12–17] vs 12 [9–14] compared with those with no limits to care. This group had a mortality of 68%, representing 33% of all patients who died.

The AUC of the 4C Mortality Score after adding “not for intubation” status as a variable into the logistic regression model was 0.802 [0.757–0.847, optimism-corrected 0.800]. The calibration of the new model showed good fit with our ICU cohort (Fig. 3).

## Discussion

The ISARIC 4C Mortality Score was derived and validated in a cohort that was comprised of mostly noncritically ill patients.<sup>9</sup> Therefore, we wanted to assess whether the ISARIC 4C Mortality Score was able to maintain its predictive calibration and discrimination in a cohort of solely critically ill patients within the ICU setting and compare this model with existing prognostic models such as the APACHE II and the SOFA scores, which are already critical illness specific.

Findings from our study provide additional validation and support for the use of the ISARIC 4C Mortality Score as a prognostic tool for mortality among patients admitted with COVID-19 to the ICU. The AUC for the 4C Mortality Score obtained in our study of 0.762 was similar to that obtained in the original validation study (0.767 [95% CI, 0.76 to 0.77]),<sup>9</sup> despite the lower incidence of mortality at 23.8% in our cohort compared with 30–32%. Furthermore, our cohort model showed no evidence of poor model fit as suggested by evaluation of the calibration slope and intercept of our calibration curve. In addition to showing the utility of this score in an exclusively critically ill population, given the timeframe of the cohort, our study also validated the use of the 4C Mortality Score across the spectrum of different SARS-CoV-2 variants, whereas the original ISARIC study was derived exclusively among the initial SARS-CoV-2 wild type strain.

We also found that this predictive ability remained robust across a wide variety of sensitivity analyses that are important to the critical care environment and were not initially explored in the validation cohort. Finally, we observed that this predictive ability is preserved, and possibly even slightly enhanced, when adjusting for patient’s preference to not undergo endotracheal intubation. This finding may be of significant relevance for practicing critical care providers who are often faced with sparse data regarding the outcomes of patients with such limitations to therapy.

Recently, another validation study of the ISARIC 4C Mortality Score in a critically ill cohort of 1,493 ICU patients in Saudi Arabia was conducted.<sup>12</sup> In this cohort, a similar AUC to ours was reported (0.81), providing additional support for the use of the 4C Mortality Score in a critical care setting.<sup>12</sup> In contrast to our study, they reported an overall higher mortality (38% vs 24%) despite

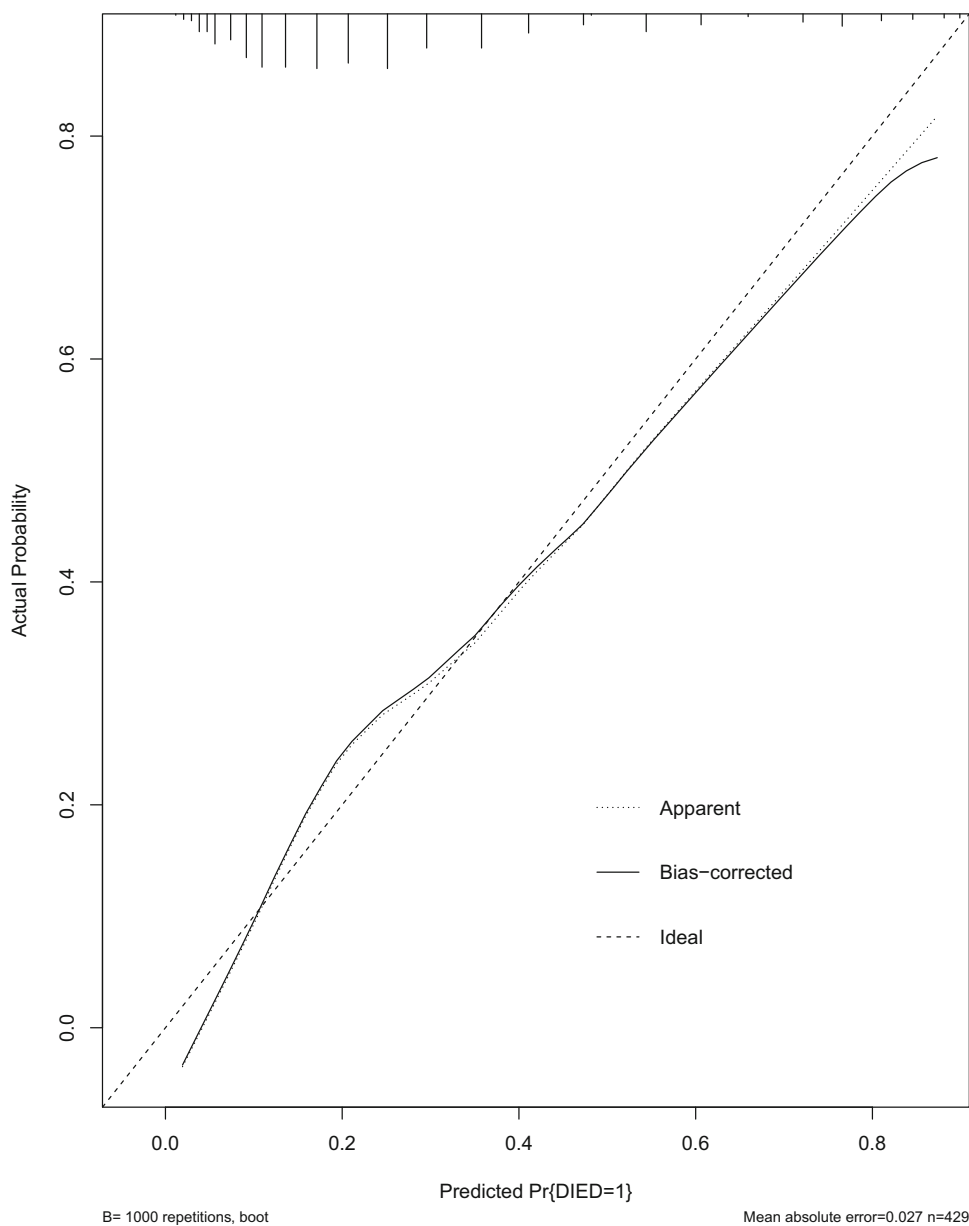
**Table 2** Prespecified sensitivity analyses

Score	Whole cohort <i>N</i> = 429	Age < 80 yr <i>N</i> = 365	No therapeutic limits <i>N</i> = 379	ICU admit on day 0 or 1 <i>N</i> = 275
ISARIC 4C Mortality Score	0.762 (0.717 to 0.811)	0.740 (0.682 to 0.797)	0.745 (0.690 to 0.800)	0.782 (0.722 to 0.842)
APACHE II	0.722 (0.667 to 0.777)	0.737 (0.672 to 0.803)	0.742 (0.680 to 0.808)	0.759 (0.687 to 0.831)
SOFA	0.705 (0.648 to 0.761)	0.720 (0.653 to 0.787)	0.744 (0.678 to 0.806)	0.734 (0.664 to 0.805)

Values represent area under receiver operating curve (95% confidence interval)

APACHE = Acute Physiology and Chronic Health Evaluation; ICU = intensive care unit; ISARIC = International Severe Acute Respiratory and Emerging Infection Consortium; SOFA = Sequential Organ Failure Assessment

**Fig. 3** Calibration plot for the new model including ISARIC 4C Mortality Score adjusted for “not for intubation” status. Perfect predicting accuracy is represented by the “ideal” line. The “apparent” line represents the predicting accuracy within our study. The “bias-corrected” line represents the prediction accuracy following bootstrap resampling (1,000 resamples).



a younger median age of their population (51.1 yr for survivors and 56.8 yr for nonsurvivors *vs* 64 and 75 yr old) and lower 4C Mortality Scores (6.3 and 12.4 *vs* 11 and

14).<sup>12</sup> This may reflect differences in patient demographics or variations in health care systems or practices. Nonetheless, the results of our study not only confirm the

utility of the 4C Mortality Score in critically ill cohorts, but also in different critical care settings, adding to the landscape by providing support for its use within Canadian critical care units.

The discrimination of the ISARIC 4C Mortality Score (0.762) showed good predictive ability in comparison with the APACHE II (0.722) and SOFA (0.705) scoring systems. In addition, we feel that the 4C Mortality Score has other advantages over the traditional scoring systems in the context of COVID-19-related admissions. It is an easy-to-use tool that requires only eight variables, all of which are commonly available at first assessment in the hospital. In contrast, many of the variables required for the APACHE II and SOFA require measurements that are available only after admission to a critical care setting and necessitate more invasive interventions.

Additionally, recent studies have shown that male sex, older age, a higher number of comorbidities, and more inflammatory changes within the respiratory system are key determinants of prognosis related to COVID-19.<sup>10–14</sup> The ISARIC 4C Mortality Score includes all these important variables thus reflecting COVID-19-related prognosis effectively. On the other hand, the SOFA score was created primarily to assess organ dysfunction related to sepsis and among the six variables it uses, studies have shown that only the respiratory, renal, and hepatobiliary organ systems were found to be associated with mortality in COVID-19.<sup>15,16</sup> The APACHE II score has some elements arguing against its use for COVID-19-related pneumonia. Firstly, all current evaluations of ICU performance are weighed against the original reference population of 5,815 patients from 13 American hospitals in 1985; therefore, its use for present-day populations is often associated with issues such as poor calibration.<sup>17,18</sup> Secondly, the inclusion of many complex variables meant to capture general critical illness does not capture the predominant involvement of the respiratory system present in our population.<sup>17,18</sup> Taking these factors into account, as well as the 4C Mortality Score's ease of use and ability to generate meaningful prognostic data prior to ICU admission, the 4C Mortality Score is, in our opinion, a more appropriate tool for use in cases of COVID-19 respiratory failure.

Our study is limited by its single-centre nature and mixed prospective and retrospective design, with a significant strength being the inclusion of all unselected cases of COVID-19-associated respiratory failure. As with any retrospective study, there is the risk of misclassification and missing data for certain variables in the primary record. These limitations were mitigated by having the data abstracted by trained researchers and possible discrepancies discussed among the authors. Our study had an extremely small number of 2 (0.5%) missing observations for C-reactive protein values, and thus a low

risk of any bias from missing data. Our use of imputation of the median value would not have been ideal in a situation with more missing data, but we are confident that a complete case analysis would have yielded similar results and any bias of our approach would be towards reduced discrimination. Finally, our single-centre design could potentially limit generalizability to other Canadian ICUs, but should be considered representative of units with similar populations and operational models.

In conclusion, the ISARIC 4C Mortality Score is an easy-to-use tool that showed a good predictive performance for in-hospital mortality in a cohort of COVID-19 patients admitted to an intensive care for respiratory failure. Our results suggest a good external validity of the score in a more severely ill population.

**Author contributions** Tharaniya Vallipuram, Blair C. Schwartz, Stephen S. Yang, Dev Jayaraman, and Sandra Dial contributed to study conception and design; analysis, and interpretation of data; and drafting and revising the article. Tharaniya Vallipuram and Sandra Dial contributed to data acquisition.

**Acknowledgements** The authors would like to thank all health care workers in our intensive care unit for their tireless, professional, and compassionate care of our patients during this pandemic and Dr. Andrea Benedetti for her assistance with statistical model calibration and diagnostics.

**Disclosures** The authors have no conflicts of interests to declare.

**Funding statement** There was no funding received for this study.

**Editorial responsibility** This submission was handled by Dr. Alexis F. Turgeon, Associate Editor, *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*.

## References

1. World Health Organization. WHO COVID-19 dashboard. Available from URL: <https://covid19.who.int> (accessed 25 April 2023).
2. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13: 818–29.
3. Lewandowski K, Lewandowski M. Scoring systems in the intensive care unit [German]. *Anaesthesist* 2003; 52: 965–87. <https://doi.org/10.1007/s00101-003-0550-7>
4. Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. *Crit Care* 2010; 14: 207. <https://doi.org/10.1186/cc8204>
5. Wilfong EM, Lovly CM, Gillaspie EA, et al. Severity of illness scores at presentation predict ICU admission and mortality in COVID-19. *J Emerg Crit Care Med* 2020; 5: 7. <https://doi.org/10.21037/jccm-20-92>
6. Lombardi Y, Azoyan L, Szychowiak P, 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: 1426–39. <https://doi.org/10.1007/s00134-021-06524-w>
7. Yang Z, Hu Q, Huang F, Xiong S, Sun Y. The prognostic value of the SOFA score in patients with COVID-19: a retrospective, observational study. *Medicine* 2021; 100: e26900. <https://doi.org/10.1097/md.00000000000026900>
  8. Vandenberghe J, Verbrugge L, Bruckers L, *et al.* Validation of the Acute Physiology and Chronic Health Evaluation (APACHE) II and IV Score in COVID-19 patients. *Crit Care Res Pract* 2021; 2021: 5443083. <https://doi.org/10.1155/2021/5443083>
  9. Knight SR, Ho A, Pius R, *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. <https://doi.org/10.1136/bmj.m3339>
  10. Villar J, Ferrando C, Martínez D, *et al.* Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *Lancet Respir Med* 2020; 8: 267–76. [https://doi.org/10.1016/s2213-2600\(19\)30417-5](https://doi.org/10.1016/s2213-2600(19)30417-5)
  11. Pandharipande PP, Shintani AK, Hagerman HE, *et al.* Derivation and validation of Spo<sub>2</sub>/Fio<sub>2</sub> ratio to impute for Pao<sub>2</sub>/Fio<sub>2</sub> ratio in the respiratory component of the Sequential Organ Failure Assessment score. *Crit Care Med* 2009; 37: 1317–21. <https://doi.org/10.1097/ccm.0b013e31819cefa9>
  12. Aletreby WT, Mumtaz SA, Shahzad SA, *et al.* External validation of 4C ISARIC mortality score in critically ill COVID-19 patients from Saudi Arabia. *Saudi J Med Med Sci* 2022; 10: 19–24. [https://doi.org/10.4103/sjms.sjms\\_480\\_21](https://doi.org/10.4103/sjms.sjms_480_21)
  13. Ocho K, Hagiya H, Hasegawa K, Fujita K, Otsuka F. Clinical utility of 4C mortality scores among Japanese COVID-19 patients: a multicenter study. *J Clin Med* 2022; 11: 821. <https://doi.org/10.3390/jcm11030821>
  14. Jones A, Pitre T, Junek M, *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: 18638. <https://doi.org/10.1038/s41598-021-97332-1>
  15. 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. <https://doi.org/10.1001/jama.2021.1545>
  16. Gupta S, Hayek SS, Wang W, *et al.* Factors associated with death in critically ill patients with coronavirus disease 2019 in the US. *JAMA Intern Med* 2020; 180: 1436–47. <https://doi.org/10.1001/jamainternmed.2020.3596>
  17. Chen J, Liu B, Du H, *et al.* Performance of CURB-65, PSI, and APACHE-II for predicting COVID-19 pneumonia severity and mortality. *Eur J Inflamm* 2021; 19: <https://doi.org/10.1177/20587392211027083>
  18. Lew CC, Wong GJ, Tan CK, Miller M. Performance of the Acute Physiology and Chronic Health Evaluation II (APACHE II) in the prediction of hospital mortality in a mixed ICU in Singapore. *Proc Singapore Healthc* 2019; 28: 147–52. <https://doi.org/10.1177/2010105818812896>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.