




MAIN TEXT

Evaluation of the discrimination and calibration of predictive scores of mortality in ECMO for patients with COVID-19

Ivan Alfredo Huespe^{1,2,3,4}  | Carolina Lockhart¹  | Rahul Kashyap^{4,5,6} | Fernando Palizas Jr⁷ | Malena Colombo⁸ | Maria del Pilar Romero⁸ | Eduardo Prado¹ | Christian A. Casabella García⁵ | Marcos Las Heras¹ | Indalecio Carboni Bisso¹ 

¹Unidad de Terapia Intensiva Adultos, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

²Área de investigación en medicina Interna, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

³Universidad de Buenos Aires, Buenos Aires, Argentina

⁴Global Clinical Scholars Research Trainee, Harvard Medical School, Boston, Massachusetts, USA

⁵Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, Minnesota, USA

⁶Department of Research, WellSpan Health, York, Pennsylvania, USA

⁷Unidad de Terapia Intensiva Adultos, Clínica Bazterrica, Buenos Aires, Argentina

⁸Instituto Universitario del Hospital Italiano, Buenos Aires, Argentina

Correspondence

Ivan Alfredo Huespe, Unidad de Cuidados Intensivos, Hospital Italiano de Buenos Aires, Tte. Gral. J. D. Perón 4190, Buenos Aires 1199, Argentina.
Email: ivan.huespe@hospitalitaliano.org.ar

Abstract

Background: The criteria for the selection of COVID-19 patients that could benefit most from ECMO organ support are yet to be defined. In this study, we evaluated the predictive performance of ECMO mortality predictive models in patients with COVID-19. We also performed a cost–benefit analysis depending on the mortality predicted probability. We conducted a retrospective cohort study in COVID-19 patients who received ECMO at two tertiary care hospitals between March 2020 to July 2021.

Materials and Methods: We evaluated the discrimination (C-statistic), calibration (Cox calibration), and accuracy of the prediction of death due to severe ARDS in V-V ECMO score (PRESERVE), the Respiratory Extracorporeal Membrane Oxygenation Survival Score (RESP) score, and the PREDiction of Survival on ECMO Therapy-Score (PRESET) score. In addition, we compared the RESP score with Plateau pressure instead of Peak pressure.

Results: We included a total of 36 patients, 29 (80%) of them male and with a median (IQR) APACHE of 10 (8–15). The PRESET score had the highest discrimination (AUROCs 0.81 [95%CI 0.67–0.94]) and calibration (calibration-in-the-large 0.5 [95%CI –1.4 to 0.3]; calibration slope 2.2 [95%CI 0.7/3.7]). The RESP score with Plateau pressure had higher discrimination than the conventional RESP score. The cost per QALY in the USA, adjusted to life expectancy, was higher than USD 100 000 in patients older than 45 years with a PRESET > 10.

Conclusion: The PRESET score had the highest predictive performance and could help in the selection of patients that benefit most from this resource-demanding and highly invasive organ support.

KEYWORDS

acute respiratory distress syndrome, COVID-19, extracorporeal life support, extracorporeal membrane oxygenation, mortality, prediction, survival

1 | INTRODUCTION

Acute respiratory distress syndrome (ARDS) caused by coronavirus disease 2019 (COVID-19) is a severe and life-threatening cause of respiratory failure.¹ Extracorporeal membrane oxygenation (ECMO) was largely used in patients with COVID-19 with ARDS.² However, the information regarding these patients is from retrospective observational studies. In these studies, the ECMO indication in COVID-19 patients was between 1% and 7%.^{2,3} This variability reflects the different availability of ECMO equipment, experience, and trained personnel between Intensive Care Units (ICU).

ELSO guidelines recommend evaluating ECMO support for COVID-19 patients with PaFi < 80 for 6 h or < 50 for 3 h, after other measures have been attempted, in particular, prone positioning.⁴ However, ECMO is a finite resource and requires other finite resources (such as ICU beds and staffing). The patient selection must be equitable and judicious, however, it becomes more stringent as ICU capacity diminishes.⁴ In this way, better survival prediction may improve resource utilization, allow risk-adjusted comparison of centers, and help clinicians target patients most likely to benefit from ECMO. This could be of utmost importance in a pandemic with unclear long-term outcomes accompanied by a shortage of resources.⁵

Three mortality predictive models for patients with respiratory failure treated with V-V ECMO have been designed: (1) the PREDiction of Survival on ECMO Therapy-Score (PRESET) score (published by Hilder et al.,⁶ developed with a cohort of 108 patients) (Table S1); (2) the Respiratory Extracorporeal Membrane Oxygenation Survival Score (RESP) (developed by Schmidt et al.,⁷ with a cohort of 2355 patients) (Table S2); and (3) Prediction of Death due to Severe ARDS in V-V ECMO score (PRESERVE) (published by Schmidt et al.,⁸ developed with a cohort of 140 patients) (Table S3). However, these scores were developed in a pre-pandemic scenario.

Additionally, during the external validation of the RESP score, the researchers changed one variable, they used the Plateau pressure (>30 cm H₂O) instead of the Peak pressure (Peak pressure > 42 cm H₂O) because the external validation database did not have the Peak pressure.⁷ In this way, a recent study reported good predictive performance of the RESP score for ECMO patients with COVID-19. However they used the modified RESP score (Plateau instead of Peak pressure) because they did not have Peak pressure.⁵ This modification could have improved the predictive performance of this well-known score in COVID-19 since Plateau pressure is more related to death in ARDS than Peak pressure.⁹ However, no study compares these two versions of the RESP score for COVID-19 patients.

During the COVID-19 surges, ECMO indications were under continuous revision.^{10,11} In this regard, predictive models for ECMO mortality could be a helpful tool in the decision-making process of whether a patient should be selected for ECMO support.

Additionally, the cost of this technology in developing countries has had significant repercussions,¹² especially during the COVID-19 pandemic. The extended use of ECMO requires highly specialized staff, and equipment which increases resource use and hospital costs.⁸ Therefore, early identification of mortality risk proportion could improve equitable access to this technology.

In this study we aimed to evaluate the discrimination and calibration of specific ECMO scores at the moment of the placement of V-V ECMO in patients with COVID-19-related ARDS. We evaluated the PRESERVE, the RESP with Peak, the RESP with Plateau, and the PRESET scores. We also evaluated other classic scoring systems such as the Sequential Organ Failure Assessment (SOFA),¹³ and Acute Physiology and Chronic Health disease Classification System II (APACHE II).¹⁴ Additionally, we studied the usefulness of including the probability of mortality in the calculation of QALYs.

2 | MATERIALS AND METHODS

2.1 | Study design

This retrospective study was performed in two tertiary care hospitals, and all data was collected from the Electronic Health Records (EHR). This study was approved by the Institutional Review Board (Ethics Committee of the Hospital Italiano de Buenos Aires #5563) and adheres to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline (Table S4). Due to the retrospective design of this study and anonymous data evaluation, the need for informed consent was waived.

2.2 | Setting and participants

This study was carried out in the Hospital Italiano de Buenos Aires (HIBA) and the Clinica Bazterrica, both from the Ciudad Autónoma de Buenos Aires, Argentina. The HIBA is a high-complexity third-level university hospital, with 750 beds and 38 Intensive Care Unit (ICU) beds for adult patients. Clinica Bazterrica is a high-complexity hospital with 160 beds and 19 ICU beds for adult patients. Both centers are Extracorporeal Life Support Organization (ELSO) centers (Hospital Italiano de Buenos Aires #352 and Clinica Bazterrica #347).



We included patients of over 18 years of age with COVID-19 infection, confirmed by reverse transcriptase polymerase chain reaction (rtPCR) from nasopharyngeal swabs, who received V-V ECMO due to COVID-19-related ARDS. The patients admitted between March 2020, and July 2021 were included.

2.3 | ECMO indication and management

The ECMO indications in both centers followed the recommendations of the ELSO for COVID-19.⁴ We considered ECMO in patients with ARDS and $\text{paO}_2/\text{FiO}_2$ ratio less than 80 for 6 h or less than 50 for 3 h, in which all the other treatment options, like lung protective mechanical ventilation (MV), prone positioning, and neuro-muscular blockade had been exhausted.

The decision to indicate ECMO required an evaluation by an ECMO team consisting of intensivists physicians and cardiothoracic surgeons. Percutaneous cannulation using the Seldinger technique¹⁵ was our technique of choice for VV ECMO. Both centers performed two-site cannulation (femoral-jugular or femoral-femoral). As recommended by ELSO,¹⁶ pre and post-membrane pressures were continuously monitored, and we conducted examinations of the ECMO circuit twice a day to detect white platelet/fibrin thrombi and clots.

Both centers used unfractionated heparin to achieve and maintain a targeted activated partial thromboplastin time of 1 to 1.5 times above the normal range (20–35 s) or a minimum anti-Xa heparin activity assay levels of 0.25 U/ml.¹⁷

2.4 | Variables and data sources

Included patients were treated, per protocol, as part of local standard care. Demographic data and medical history were collected at ICU admission from the Electronic Health Record. The PRESERVE, RESP, PRESET, SOFA, and APACHE II scores were measured at ECMO initiation. We also measured the Plateau pressure, and we evaluated the modified RESP Score (we used Plateau > 30 cm H₂O instead of Peak > 42 cm H₂O).^{5,7}

Laboratory, treatment regimes, MV parameters, and ECMO settings were evaluated during the whole ICU hospitalization. Finally, complications occurring post-ECMO implantation, including multi-organ failure, infections, bleeding, and thromboembolic events, were recorded and analyzed. The final outcome for all predictive models was ICU survival.

2.5 | Statistical analysis

Categorical variables were presented as proportions and percentages and compared by chi-square test or Fisher's exact test as appropriate. Quantitative variables were presented as mean and standard deviation (SD), or median and interquartile range (IQR), and compared by Student's *t*-test or Mann–Whitney *U* test according to distribution.

To evaluate the predictive performance of each score, we assessed Discrimination and Calibration. Discrimination addresses the extent to which a model predicts a higher probability of having an event among patients who will versus those who will not have an event.¹⁸ Usually, it is quantified with a concordance (*c*) statistic that is identical to the area under the receiver operating characteristic (ROC) curve for a binary outcome.¹⁹ We used the *c* statistic with a 95% confidence interval (CI) as the discrimination measurement and the Hanley and McNeil test to compare them.

Calibration reflects the extent to which a model correctly estimates the absolute risk. Poorly calibrated models will underestimate or overestimate the outcome of interest.¹⁸ Calibration can be assessed graphically in a plot with predictions on the *x*-axis and the observed endpoint on the *y*-axis, dividing groups of patients with similar predicted risk by deciles of predictions.¹⁹ In our study, to assess the graphical evaluation, we divided groups into quartiles of predictions because of our sample size. In this way, the Hosmer–Lemeshow goodness-of-fit test compares observed predictions against ideal predictions. However, the Hosmer–Lemeshow test does not indicate the direction of any miscalibration and only provides a *p*-value that, like any other hypothesis test, depends on the sample size. Thus, since a *p* > 0.05 in the Hosmer–Lemeshow test represents good calibration, if the database has a small sample size, all predictive models would have good calibration. Therefore, we used Cox's approach to evaluate calibration.¹⁹ With this approach the perfect prediction must be on the ideal line, and described with an intercept alpha (calibration-in-the-large [CITL]) of 0 and slope beta (calibration slope) of 1.

Additionally, we performed the Brier score to test the predictive performance of each score, and to compare our observations with Moyon et al.⁵ The Brier score is influenced by discrimination and calibration simultaneously, and can range from 0 for a perfect model to 0.25 for a non-informative model with a 50% incidence of the outcome. However, this score does not evaluate the clinical utility of diagnostic tests or prediction models.²⁰ Therefore, to assess the clinical usefulness of the predictive model in



terms of net benefit (NB), we performed a decision curve analysis (DCA).^{21–23} For the whole range of decision threshold probabilities (pt), the net benefit of the model was compared to default strategies of treating all or no patients.^{21,22}

Additionally, we evaluated sensitivity, specificity, and likelihood ratios for the score with the highest calibration and discrimination. The optimal threshold was chosen based on the Youden index.²⁴ Finally, we used the predictive model with the best calibration and discrimination to divide the population into high and low risk of death according to the optimal threshold. We then compared time to death in both groups with a Cox proportional-hazards regression model. The regression model was used to estimate the Hazard Ratio (HR) for in-hospital death in low and high-risk groups. We also presented the Kaplan–Meier curve. We used STATA v.16 for the statistical analysis.

2.6 | Cost–benefit analysis

We performed an economic evaluation of both cost-effectiveness and cost-utility from the perspectives of the publicly funded health and social care sector in middle-income (Argentina) and high-income (United States [US]) countries. In the US, the mean total cost of ECMO organ support ranges from United States dollar (USD) 105 034 to USD 335 565, with an average of USD 200 000.²⁵ In non-US countries, the mean total cost of ECMO use is between USD 42 554 to USD 537 554.²⁵ In the participant centers, ECMO cost was estimated at 80 000 USD.

We selected the health-related quality of life reported by the CESAR Trial.²⁶ They measured the health utilities through the Visual Analog Scale (VAS). The overall VAS score six months after the ECMO weaning was 0.67 in the CESAR Trial.²⁶

In our study, the lifetime incremental cost-utility was estimated with several simplifying assumptions: (1) age-specific and sex-specific life expectancy for each surviving patient were calculable from US²⁷ and Argentinian²⁸ life tables (Table S5); (2) survival probability six months after ECMO was calculated with the predictive model with best calibration and discrimination; (3) after hospital discharge, the average health-service expenditure for surviving patients was the same as that of similar age groups in the US and Argentina. Therefore, the cost/QALY formula was:

$$\text{cost/QALY} = \frac{\text{cost}}{\text{life expectancy} \times \text{health utilities} \times \text{survival probability}}$$

For example, for a US male of 30 years with a survival probability of 60%, the calculated cost-utility for ECMO was USD 7922 per QALY.

$$\text{cost/QALY} = \frac{150\,000}{47.1 \times 0.67 \times 0.6} = 7922 \text{ USD/QALY}$$

3 | RESULTS

3.1 | Demographic data

During the study period, a total of 315 adult patients with severe COVID-19 were admitted to both ICUs. 239 patients required MV and 36 ECMO support (Figure S1). Eighty percent of patients were male with a median (IQR) age of 48 (41–59) years. The median(IQR) hospital length of stay (LOS) was 49 (30–66) days, and the in-hospital mortality of the series was 50% ($n = 18$). The clinical and demographic characteristics of the study patients are in Table 1.

Additionally, only one patient (2.7%) had a Peak pressure ≥ 42 , but 6 (16%) patients had a Plateau pressure ≥ 30 . Therefore, only one patient scored points for this variable in the RESP Score, but 6 patients scored points for this variable in the RESP score with Plateau pressure.

3.2 | Discrimination and calibration

The PRESET score and the RESP calculated with the Plateau score, with AUROCs of 0.81 (95%CI 0.67/0.94) and 0.80 (95%CI 0.65/0.94), were the models with higher discrimination (Table 2). The model with the worst AUROC was the APACHE II with 0.52 (95%CI 0.32/0.72). These differences were evaluated with the Hanley and McNeil test, and were statistically significant ($p < 0.03$). The AUROCs graph is in Figure 1.

Regarding the calibration, the PRESET score had the best calibration with a CITL of -0.54 (95%CI $-1.4/0.3$), and a Slope of 2.2 (95%CI 0.7/3.7), therefore, both parameters were non-significantly different from the perfect calibration (CITL of 0 and SLOPE of 1). The RESP score and the RESP score with Plateau had a CITL significantly higher than 0, and the PRESERVE score had a Slope significantly lower than 1 (Table 2). Also, the PRESET score had the lowest Brier score with 0.18 (Table 2). For the graphic evaluation of the predicted/observed probability, we divided the population into four groups based on the average predicted probability (Figure 2).

Additionally, decision curve analysis showed that the PRESET score had a higher net benefit across a wide range of threshold probabilities for death compared to RESP and PRESERVE scores (Figure 3). The PRESET score has a positive net benefit with threshold probabilities between 15% (PRESET score 2) and 85% (PRESET score of 10) (Figure 3).

TABLE 1 Clinical and demographic characteristics of study patients

	All participants (n = 36)	Survivors (n = 18)	Non-survivors (n = 18)	p-overall
Demographic characteristics				
Male gender, no. (%)	29 (80.6)	12 (66.7%)	17 (94.4%)	0.088
Age, median (IQR)	47.5 [40.8; 59.2]	44.0 [39.2; 49.5]	52.5 [43.0; 61.0]	0.064
Severity scoring, median (IQR)				
APACHE II	10.0 [8.00; 14.5]	10.0 [8.00; 12.0]	10.0 [8.00; 15.0]	0.817
SOFA day 1	5.50 [3.00; 7.00]	4.50 [2.25; 6.00]	6.00 [4.00; 7.75]	0.203
Charlson	0.00 [0.00; 1.00]	0.00 [0.00; 1.00]	0.50 [0.00; 2.00]	0.327
Comorbidities, no. (%)				
Obesity	21 (58.3%)	10 (55.6%)	11 (61.1%)	0.99
Active smoking	9 (25.0%)	5 (27.8%)	4 (22.2%)	0.99
Diabetes	6 (16.7%)	2 (11.1%)	4 (22.2%)	0.658
Coronary artery disease	4 (11.1%)	1 (5.56%)	3 (16.7%)	0.603
Asthma	3 (8.33%)	1 (5.56%)	2 (11.1%)	0.99
COPD	2 (5.56%)	0 (0.00%)	2 (11.1%)	0.486

Note: Missing data: none.

Abbreviations: APACHE II, Acute Physiology and Chronic Health disease Classification System II; COPD, chronic obstructive pulmonary disease; ICU, intensive care unit; IQR, interquartile range; LOS, length of stay; SOFA, sequential organ failure assessment.

TABLE 2 Calibration, discrimination and brier value of each score

	AUROC	CITL	Slope	Brier
SOFA Score	0.60 (95%CI 0.41–0.79)	0.43 (95%CI –0.2/1.14)	2.31 (95%CI 1.6/3.04)	0.38
APACHE II	0.52 (95%CI 0.32–0.72)	0.2 (95%CI –0.72/1.12)	1.92 (95%CI 1.22/2.62)	0.36
RESP score	0.78 (95%CI 0.62–0.93)	1.20 (95%CI 0.53/1.86)	3.02 (95%CI 0.78/5.25)	0.28
RESP Score with plateau	0.80 (95%CI 0.65–0.94)	1.15 (95%CI 0.48/1.82)	3.05 (95%CI 0.95/5.16)	0.27
PRESERVE score	0.64 (95%CI 0.45–0.83)	3.1 (95%CI 2.33/3.92)	0.39 (95%CI –0.05/0.84)	0.37
PRESET score	0.81 (95%CI 0.67–0.94)	–0.22 (95%CI –0.92/0.46)	2.2 (95%CI 0.7/3.7)	0.18

Abbreviations: AUROC, area under the ROC curve; CITL, calibration in the large; PRESERVE, prediction of death due to severe ARDS in V-V ECMO score; PRESET, PREDiction of survival on ECMO therapy-score; RESP, respiratory extracorporeal membrane oxygenation survival score.

3.3 | Evaluation of the accuracy

According to the Youden index, the optimal threshold for the PRESET score was 7 (sensitivity of 67%, specificity of 89%, and accuracy of 78%). The sensitivity, specificity, and likelihood ratios of the PRESET score are in [Table S6](#). Additionally, the Kernel density population distribution in patients who died and survived ECMO depending on the PRESET score are in [Figure S2](#).

According to the optimal threshold of the PRESET score, mortality in patients with a PRESET score equal to or higher than 7 was 86% (95%CI 57%/98%), and in patients with PRESET lower than 7 was 27% (95%CI 10/50%). The Hazard Ratio of mortality in patients with PRESET equal to or higher than 7 compared with patients with a lower

score was 4,16 (CI 95% 1.21/14.33). [Figure S3](#) shows the Kaplan–Meier curve.

3.4 | Cost-benefit analysis

The USD/QALY relation for the US towards age, sex, and survival probabilities are in [Table 3](#). Clinical situations with USD per QALY costs higher than USD 100 000 are in red because the Institute for Clinical and Economic Review (ICER) from the US uses a threshold of USD 100 000–150 000²⁹ as the maximum cost per QALY. In Argentina, as there is no established threshold per QALY, we did not determine these limits. The USD/QALY relation for Argentinians is in [Table S7](#) of supplementary material.

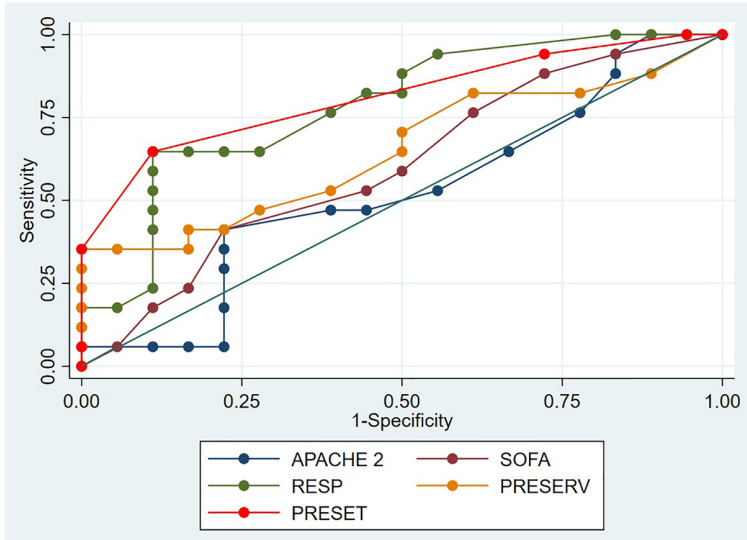


FIGURE 1 Area under the receiver operating characteristic (AUROC). APACHE II, Acute Physiology and Chronic Health disease Classification System II; PRESERVE: prediction of death due to severe ARDS in V-V ECMO score; PRESET: Prediction of survival on ECMO therapy-score; SOFA: sequential organ failure assessment; RESP: respiratory extracorporeal membrane oxygenation survival score.

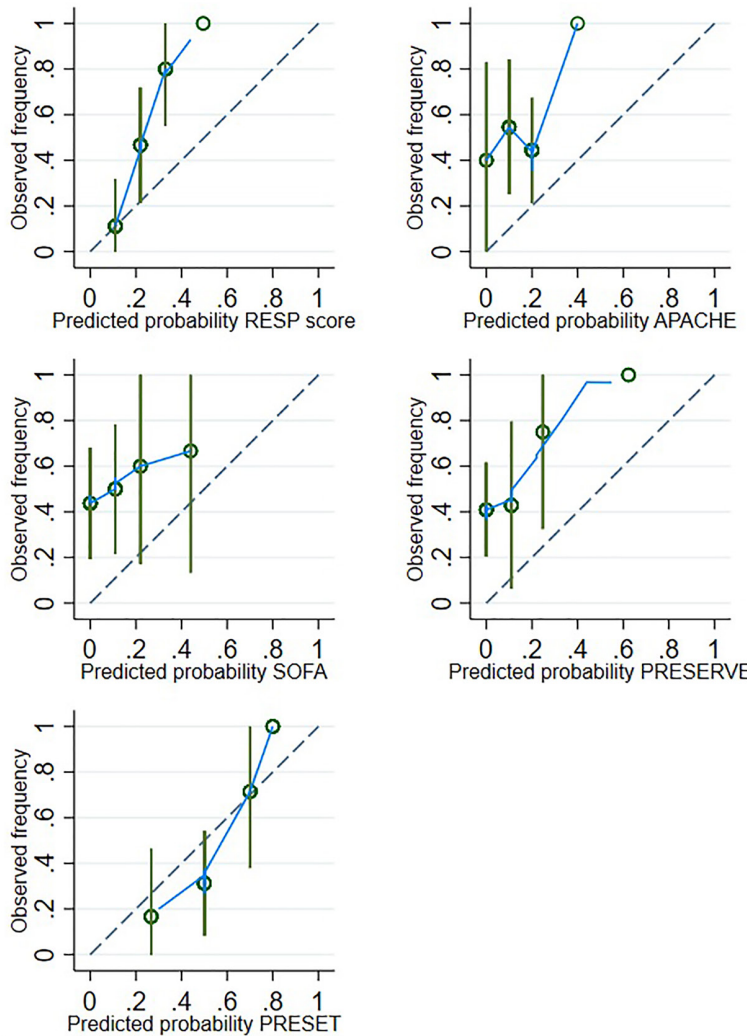


FIGURE 2 Calibration by quartiles of the RESP score, APACHE II, SOFA, PRESERVE score and PRESET for the prediction of outcome mortality. APACHE II, Acute Physiology and Chronic Health disease Classification System II; PRESET, Prediction of survival on ECMO therapy-score; PRESERVE: prediction of death due to severe ARDS in V-V ECMO score; RESP, respiratory extracorporeal membrane oxygenation survival score; SOFA, sequential organ failure assessment.

4 | DISCUSSION

In this retrospective multicenter research, we evaluated the external validity of the ECMO mortality predictive

models in COVID-19 patients. The PRESET score presented a better performance in terms of discrimination and calibration. On the other hand, we determined that SOFA and APACHE II scores do not seem useful for COVID-19 patients.

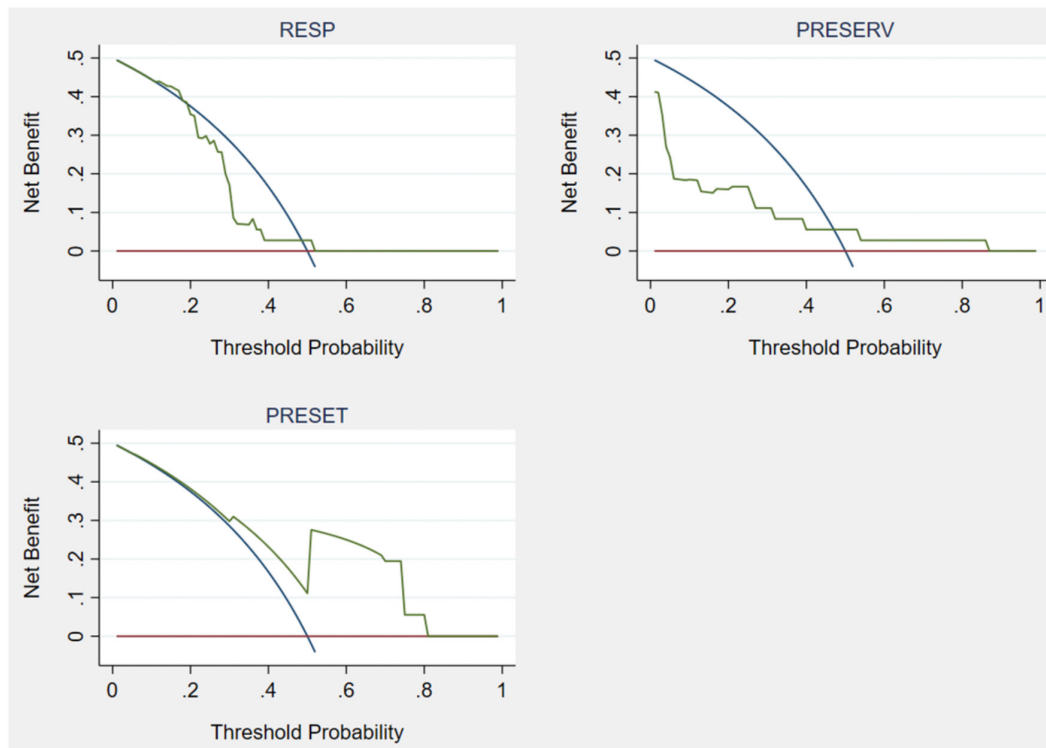


FIGURE 3 Decision curves of the predictive model for death in patients with ECMO. The x-axis represents threshold probabilities and the y-axis the net benefit. RESP, respiratory extracorporeal membrane oxygenation survival score; PRESERVE, prediction of death due to severe ARDS in V-V ECMO score; PRESET, Prediction of survival on ECMO therapy-score.

Broman et al. reported³⁰ that survival in COVID-19 ECMO patients was 53% (770) in the first wave and 44% (677; $p < 0.0001$) in the second wave³¹; also Lockhart et al. informed a 57% ECMO survival rate in Argentina.³² Therefore, to avoid futility, predictable complications, and therapeutic obstinacy, the decision to start ECMO should rely on the best available evidence (Table 3).

Four studies evaluated the discrimination of the main ECMO mortality predictive models in COVID-19 patients, and they had heterogeneous results. The first study was in Germany with 19 patients, and they observed that the RESP score had an AUROC of 0.79, equal to our observations.³³

In the second study, Moyon et al. reported that the RESP score was the best predictive model in ECMO COVID-19 patients.⁵ However, this study did not evaluate the PRESET score, and they used Plateau pressure instead of Peak pressure to calculate the RESP score. The original RESP score includes Peak pressure. However, Plateau pressure was also used in the external validation of the original RESP score study (as the external database did not have Peak pressure).^{5,7}

The RESP score has been developed for every patient with respiratory failure, including asthma or trauma. In these diseases, Peak pressure has predictive value.

However, in ARDS, Plateau pressure is more closely related to mortality than Peak pressure.⁹ In this sense, Joshi et al.³⁴ observed that the RESP Score (with Peak pressure) did not have good predictive performance for ECMO COVID-19 patients. They suggested that a recalibration of the RESP score may be necessary for different types of respiratory diseases and different treatment regimes.³⁴

Supady et al. also evaluated the RESP and PRESERVE scores in 127 patients with COVID-19 treated with V-V ECMO in 15 centers. They reported low discrimination for both scores and concluded that these scores should not have been recommended for treatment decisions.¹¹ Finally, Tabatabai et al. also evaluated the accuracy of the RESP score in 40 patients under V-V ECMO for COVID-19. They reported that the RESP score had a bad predictive performance for mortality prediction in COVID-19 patients.³⁵

The discrepancy between Moyon et al.⁵ (the only study that reported a good performance of RESP score for COVID-19) and Joshi et al.,³⁴ Supady et al.,¹¹ Tabatabai et al.³⁵ and our study may be because Moyon et al. improved the RESP score using Plateau pressure instead of Peak pressure. We compared both RESP scores (with Peak and with Plateau pressure) and we observed that the RESP Score with Plateau pressure has higher discrimination than the RESP score with Peak pressure. In turn, we

TABLE 3 Relation USD/QALYs based on age and death probability

Age	Death probability <10% (PRESET SCORE ≤2)	Death probability 10%–40% (PRESET SCORE 3,4)	Death probability 40%–60% (PRESET SCORE 5,6)	Death probability 60%–80% (PRESET SCORE 7,8)	Death probability 80%–90% (PRESET SCORE 9,10)	Death probability >90% (PRESET SCORE ≥11)
Relation USD/QALYs based on age and death probability in male for the USA						
20	5881	7561	10 585	17 642	35 285	66 159
25	6403	8232	11 525	19 209	38 418	72 034
30	7042	9054	12 675	21 126	42 252	79 222
35	7804	10 034	14 047	23 412	46 825	87 796
40	8774	11 281	15 794	26 323	52 647	98 713
45	9960	12 806	17 928	29 881	59 761	112 052
50	11 437	14 705	20 587	34 311	68 622	128 667
55	13 320	17 126	23 977	39 961	79 922	149 853
60	15 870	20 404	28 565	47 609	95 218	178 533
65	19 283	24 793	34 710	57 850	115 701	216 939
Relation USD/QALYs based on age and PRESET Score in female for the USA						
20	5420	6968	9755	16 259	32 517	60 970
25	5891	7574	10 604	17 674	35 347	66 276
30	6440	8280	11 593	19 321	38 642	72 453
35	7102	9131	12 784	21 307	42 613	79 900
40	7916	10 178	14 249	23 748	47 495	89 054
45	8916	11 463	16 049	26 748	53 496	100 305
50	10 143	13 041	18 257	30 429	60 858	114 108
55	11 762	15 122	21 171	35 285	70 569	132 317
60	13 878	17 843	24 980	41 633	83 266	156 123
65	16 667	21 429	30 001	50 001	100 003	187 505

Note: We calculated the health utilities from the CESAR trial and we adjusted the USD/QALY depending on the probability of death. We also include the PRESET score for each risk. This table is based on the USA life expectancy. Patients with Costs per QALY lower than USD 100 000 are in green and those with higher Costs per QALY in red.



hypothesized that this modification of the RESP score, using Plateau instead of Peak pressure, could have a better predictive performance for ARDS patients without COVID-19. However, more studies are needed to test this hypothesis.

The only study that evaluated the PRESET score for COVID-19 patients was Tabatabai et al. They compared PRESET against RESP in patients who had ECMO for COVID-19 in a prospective cohort. They found that all patients with a PRESET-Score equal to or lower than 6 survived, and patients with a score of 7 or higher had an increased risk of mortality. This was the same cutoff observed in our study, although they reported an accuracy of 100% for this threshold, and we observed an accuracy of 78%. They also reported that the RESP score (with Peak pressure) did not have any utility in predicting mortality.³⁵

Finally, of the previous studies, only Moyon et al. evaluated the calibration of the models with the Hosmer–Lemeshow goodness-of-fit test. However, this test depends on the sample size. In this way, they found that all evaluated predictive models had a $p > 0.05$, even the SOFA score (which has the worst discrimination).⁵ The Hosmer–Lemeshow goodness-of-fit test does not indicate the direction of any miscalibration and only provides a p -value for differences between observed and predicted endpoints per group of patients. Therefore, all predictive models are going to be well calibrated if the database is small.¹⁹ On the other hand, the older recalibration idea as proposed by Cox in 1958 sustained that perfect prediction should be on the ideal line, with a CITL (“A”) of 0 and slope beta (“B”) of 1.³⁶ Thus, if the 95%CI of the CITL does not include 0 and the 95%CI of the Slope does not include 1, the model is not well calibrated.¹⁹ Of course, this also depends on the sample size, but in our study with 36 patients, we have a large enough sample size to see that the only predictive model that achieved both 95%CI included in the perfect prediction values was the PRESET score. Additionally, we performed a decision curve analysis to evaluate which model had the highest Net benefit. Through this analysis, we observed that the PRESET score also has the highest net benefit.

Finally, we performed a cost-effectiveness analysis with the cost per QALY analysis. There is some controversy regarding the quality of life after ECMO.³⁷ Some studies reported a benefit of ECMO over conventional MV,^{26,38} while others reported worse outcomes for ECMO^{39,40} or no difference.⁴¹ For this study, we select the health-related quality of life reported by the CESAR Trial and by Kanji et al.³⁷ However, two critical patients of the same age who need ECMO, even if one has a 10% possibility of surviving and the other 90%, both have the same QALY. Therefore, we propose to add the output of the best predictive model to the calculation of QALYs. Through this analysis, we can

see in Table 2 that the cost per QALY has a huge variation depending on the survival probabilities of the patient. In addition, we were able to calculate at what PRESET score and age a patient exceeded the cost per QALY cutoff point of the Institute for Clinical and Economic Review (ICER) from the US.²⁹ However, the proposal to add mortality probabilities to the QALY cost analysis should be further evaluated by the scientific community.

There are several strengths in our study. It is the first that evaluates the calibration and discrimination of the PRESET score among COVID-19 patients with ECMO, whereas Tabatabai et al. only evaluated the accuracy of PRESET. Previously, the PRESET score has shown an excellent performance in predicting mortality in non-COVID-19 ARDS patients requiring ECMO.⁶ Another strength is that we performed a cost-effectiveness analysis, even though this was not the main objective of this study. It is important to include this type of analysis, especially regarding this high-cost technology in low and middle-income countries, such as Latin-American countries, to improve equitable access to this technology.

Our study has some limitations. First, the small sample size: due to the cost of this technology in our environment, its use is limited. However, the sample size was sufficient to find differences between the different predictive models evaluated. Second, our patients were treated in two high-complexity academic hospitals in Argentina, limiting the generalizability. Third, we used the health utilities calculated in the CESAR Trial (we did not calculate the utilities in our setting). Therefore, the USD/QALY relation for Argentinians lacks robustness. An analysis of the Health Utilities of ECMO survivors in the Latin-American population should be carried out. Fourth, the maximum cost per QALY was available only in the US (from the Institute for Clinical and Economic Review²⁹) because Argentina has not established a maximum cost per QALY. Finally, in this study, we did not compare the predictive performance of the ROCH score and other ECMO scores.

5 | CONCLUSION

The present study observed that the PRESET score had the highest calibration and discrimination for mortality prediction in ECMO organ support for COVID-19 patients. We also observed that the RESP score could be improved if Plateau pressure is used instead of Peak pressure. Additionally, we propose adding the outcome of the mortality prediction model to the ECMO's cost-benefit analysis. However, due to our sample size, further research is required to optimize the external validation of ECMO scores.



AUTHOR CONTRIBUTIONS

Iván Huespe: Concept/design, Data analysis/interpretation, Statistics and Drafting article. Carolina Lockhart: Data analysis/interpretation, Drafting article and Data collection. Rahul Kashyap: critical revision of article, drafting article, Statistics, English grammar check. Fernando Páizas Jr: Critical revision of article. Malena Colombo: Data collection. María del Pilar Romero: Data collection. Eduardo Prado: Data collection. Christian Casabella García: Critical revision of article. Marcos Las Heras: Critical revision of article, Approval of article. Indalecio Carboni Bisso: Data analysis/interpretation, Drafting article and Data collection.

ACKNOWLEDGMENTS

The research team would like to thank Instituto Universitario del Hospital Italiano for developing a participation program for undergraduate students in research projects, stimulating continuing education among investigation activities, and also contributing to the collection and analysis of data used in this work.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

None to declare.

ETHICS STATEMENT

The study was approved by the Institutional Ethics Committee (protocol number 5563).

ORCID

Ivan Alfredo Huespe  <https://orcid.org/0000-0003-3445-6981>

Carolina Lockhart  <https://orcid.org/0000-0002-3888-8725>

Indalecio Carboni Bisso  <https://orcid.org/0000-0002-4834-4676>

Indalecio Carboni Bisso  <https://orcid.org/0000-0002-4834-4676>

Indalecio Carboni Bisso  <https://orcid.org/0000-0002-4834-4676>

REFERENCES

1. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020;8:475–81.
2. Bertini P, Guarracino F, Falcone M, Nardelli P, Landoni G, Nocci M, et al. ECMO in COVID-19 patients: a systematic review and meta-analysis. *J Cardiothorac Vasc Anesth*. 2022;36:2700–6.
3. Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in new York City: a prospective cohort study. *Lancet*. 2020;395:1763–70.
4. Badulak J, Antonini MV, Stead CM, Shekerdemian L, Raman L, Paden ML, et al. Extracorporeal membrane oxygenation for COVID-19: updated 2021 guidelines from the extracorporeal life support organization. *ASAIO J*. 2021;67:485–95.
5. Moyon Q, Pineton de Chambrun M, Lebreton G, Chaieb H, Combes A, Schmidt M. Validation of survival prediction models for ECMO in Sars-CoV-2-related acute respiratory distress syndrome. *Crit Care*. 2022;26:1–4.
6. Hilder M, Herbstreit F, Adamzik M, Beiderlinden M, Bürschen M, Peters J, et al. Comparison of mortality prediction models in acute respiratory distress syndrome undergoing extracorporeal membrane oxygenation and development of a novel prediction score: the PREDiction of survival on ECMO therapy-score (PRESET-score). *Crit Care*. 2017;21:301.
7. Schmidt M, Bailey M, Sheldrake J, Hodgson C, Aubron C, Rycup PT, et al. Predicting survival after extracorporeal membrane oxygenation for severe acute respiratory failure. The respiratory extracorporeal membrane oxygenation survival prediction (RESP) score. *Am J Respir Crit Care Med*. 2014;189:1374–82.
8. Schmidt M, Zogheib E, Rozé H, Repesse X, Lebreton G, Luyt CE, et al. The PRESERVE mortality risk score and analysis of long-term outcomes after extracorporeal membrane oxygenation for severe acute respiratory distress syndrome. *Intensive Care Med*. 2013;39:1704–13.
9. Chan MC, Tseng JS, Chiu JT, Hsu KH, Shih SJ, Yi CY, et al. Prognostic value of plateau pressure below 30 cm H₂O in septic subjects with acute respiratory failure. *Respir Care*. 2015;60:12–20.
10. Li C, Hou X, Tong Z, Qiu H, Li Y, Li A, et al. Extracorporeal membrane oxygenation programs for COVID-19 in China. *Crit Care*. 2020;24:317.
11. Supady A, DellaVolpe J, Taccone FS, Scharpf D, Ulmer M, Lepper PM, et al. Outcome prediction in patients with severe COVID-19 requiring extracorporeal membrane oxygenation—a retrospective international multicenter study. *Membranes (Basel)*. 2021;11(3):170.
12. Park M, Mendes PV, Zampieri FG, Azevedo LCP, Costa ELV, Antoniali F, et al. The economic effect of extracorporeal membrane oxygenation to support adults with severe respiratory failure in Brazil: a hypothetical analysis. *Rev Bras Ter Intensiva*. 2014;26:253–62.
13. 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.
14. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9:591–7.
15. Pranikoff T, Hirschl RB, Remenapp R, Swaniker F, Bartlett RH. Venovenous extracorporeal life support via percutaneous cannulation in 94 patients. *Chest*. 1999;115:818–22.
16. Shekar K, Badulak J, Peek G, Boeken U, Dalton HJ, Arora L, et al. Extracorporeal life support organization coronavirus disease 2019 interim guidelines: a consensus document from an International Group of Interdisciplinary Extracorporeal Membrane Oxygenation Providers. *ASAIO J* [Internet].



- 2020;66:707–21. <https://doi.org/10.1097/mat.0000000000001193>
17. Figueroa Villalba CA, Brogan TV, McMullan DM, Yalon L, Jordan DI, Chandler WL. Conversion from activated clotting time to anti-Xa heparin activity assay for heparin monitoring during extracorporeal membrane oxygenation. *Crit Care Med*. 2020;48:e1179–84.
 18. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. *JAMA*. 2017;318:1377–84.
 19. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35:1925–31.
 20. Assel M, Sjöberg DD, Vickers AJ. The brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagn Progn Res*. 2017;1:19.
 21. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26:565–74.
 22. Martínez-Lacalzada M, Viteri-Noël A, Manzano L, Fabregate M, Rubio-Rivas M, Luis García S, et al. Predicting critical illness on initial diagnosis of COVID-19 based on easily obtained clinical variables: development and validation of the PRIORITY model. *Clin Microbiol Infect*. 2021;27:1838–44.
 23. Wang H, Chen C, Li B, Cheng Z, Wang Z, Huang X, et al. Nomogram to predict survival outcome of patients with venoarterial extracorporeal membrane oxygenation after refractory cardiogenic shock. *Postgrad Med*. 2022;134:37–46.
 24. Youden WJ. Index for rating diagnostic tests. *Cancer* [Internet]. 1950;3:32–5. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::aid-cnrcr2820030106>3.0.co;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cnrcr2820030106>3.0.co;2-3)
 25. Harvey MJ, Gaies MG, Prosser LA. U.S. and international in-hospital costs of extracorporeal membrane oxygenation: a systematic review. *Appl Health Econ Health Policy*. 2015;13:341–57.
 26. Peek GJ, Mugford M, Tiruvoipati R, Wilson A, Allen E, Thalanany MM, et al. Efficacy and economic assessment of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR): a multicentre randomised controlled trial. *Lancet*. 2009;374:1351–63.
 27. Statistics NCFH. National Center for Health Statistics. *Natl Vital Stat Rep* [Internet]. 2021;70(9):5–12. <https://doi.org/10.15620/cdc:107021>
 28. INDEC: Instituto Nacional de Estadística y Censos de la República Argentina [Internet]. [cited 2022 Apr 19]. Available from: <https://www.indec.gov.ar/indec/web/Nivel4-Tema-2-24-84>
 29. Institute For Clinical and Economic Review [Internet]. [cited 2022 May 19]. Available from: https://icer.org/wp-content/uploads/2020/10/ICER_2020_2023_VAF_102220.pdf
 30. Barbaro RP, MacLaren G, Boonstra PS, Iwashyna TJ, Slutsky AS, Fan E, et al. Extracorporeal membrane oxygenation support in COVID-19: an international cohort study of the extracorporeal life support organization registry. *Lancet*. 2020;396:1071–8.
 31. Broman LM, Eksborg S, Lo Coco V, De Piero ME, Belohlavek J, Lorusso R, et al. Extracorporeal membrane oxygenation for COVID-19 during first and second waves. *Lancet Respir Med*. 2021;9:e80–1.
 32. Lockhart C, Casabella García C, Las Heras M, Matarrese A, Espinosa L, Norese M, et al. Characteristics and outcomes in extracorporeal membrane oxygenation support in COVID-19 patients: a nationwide cohort-study in Argentina. *Acta Colomb Cuid Intensivo* [Internet]. 2022 [cited 2022 Nov 28]. <https://doi.org/10.1016/j.acci.2022.10.001>
 33. Zayat R, Kalverkamp S, Grottko O, Durak K, Dreher M, Autschbach R, et al. Role of extracorporeal membrane oxygenation in critically ill COVID-19 patients and predictors of mortality. *Artif Organs*. 2021;45:E158–70.
 34. Joshi H, Flanagan M, Subramanian R, Drouin M. Respiratory ECMO survival prediction (RESP) score for COVID-19 patients treated with ECMO. *ASAIO J*. 2022;68:486–91.
 35. Tabatabai A, Ghneim MH, Kaczorowski DJ, Shah A, Dave S, Haase DJ, et al. Mortality risk assessment in COVID-19 Venovenous extracorporeal membrane oxygenation. *Ann Thorac Surg*. 2021;112:1983–9.
 36. Cox DR. Two further applications of a model for binary regression. *Biometrika* [Internet]. 1958;45:562. <https://doi.org/10.2307/2333203>
 37. Kanji HD, Chouldechova A, Harris-Fox S, Ronco JJ, O'dea E, Harvey C, et al. Quality of life and functional status of patients treated with venovenous extracorporeal membrane oxygenation at 6 months. *J Crit Care*. 2021;66:26–30.
 38. Lindén VB, Lidegran MK, Frisén G, Dahlgren P, Frenckner BP, Larsen F. ECMO in ARDS: a long-term follow-up study regarding pulmonary morphology and function and health-related quality of life. *Acta Anaesthesiol Scand*. 2009;53:489–95.
 39. Hodgson CL, Hayes K, Everard T, Nichol A, Davies AR, Bailey MJ, et al. Long-term quality of life in patients with acute respiratory distress syndrome requiring extracorporeal membrane oxygenation for refractory hypoxaemia. *Crit Care*. 2012;16:R202.
 40. Stoll C, Haller M, Briegel J, Meier M, Manert W, Hummel T, et al. Health-related quality of life. Long-term survival in patients with ARDS following extracorporeal membrane oxygenation (ECMO). *Anaesthesist*. 1998;47:24–9.
 41. Luyt CE, Combes A, Becquemin MH, Beigelman-Aubry C, Hatem S, Brun AL, et al. Long-term outcomes of pandemic 2009 influenza a(H1N1)-associated severe ARDS. *Chest*. 2012;142:583–92.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Huespe IA, Lockhart C, Kashyap R, Palizas F Jr, Colombo M, Romero MdP, et al. Evaluation of the discrimination and calibration of predictive scores of mortality in ECMO for patients with COVID-19. *Artif. Organs*. 2023;00:1–11. <https://doi.org/10.1111/aor.14493>