Validation of a simplified risk prediction model using a cloud based critical care registry in a lower-middle income country

Bharath Kumar (✉ bharath@icuconsultants.com)

Apollo Hospitals Chennai   https://orcid.org/0000-0002-1801-0667

Dilanthi Gamage

Network for Improving Critical care Systems and Training

Aasiyah Rashan

Network for Improving Critical care Systems and Training

Abi Beane

Mahidol Oxford Tropical Medicine Research Unit

Ramesh Venkataraman

Apollo Hospitals Chennai

Nagarajan Ramakrishnan

Apollo Hospitals Chennai

Rashan Haniffa

Mahidol Oxford Tropical Medicine Research Unit

Abstract

Background:

The use of severity of illness scoring systems such as the Acute Physiology and Chronic Health Evaluation in lower-middle income settings comes with important limitations, primarily due to data burden, missingness of key variables and lack of resources. To overcome these challenges, in Asia, a simplified model, designated as e-TropICS was previously developed. We sought to externally validate this model using data from a multi-centre critical care registry in India.

Methods:

Seven ICUs from the Indian Registry of IntenSive care(IRIS) contributed data to this study. Patients > 18 years of age with an ICU length of stay > 6 hours were included. Data including age, gender, co-morbidity, diagnostic category, type of admission, vital signs, laboratory measurements and outcomes were collected for all admissions. e-TropICS was calculated as per original methods. The area under the receiver operator characteristic curve was used to express the model's power to discriminate between survivors and non-survivors. For all tests of significance, a 2-sided P less than or equal to 0.05 was considered to be significant. AUROC values were considered poor when ≤ to 0.70, adequate between 0.71 to 0.80, good between 0.81 to 0.90, and excellent at 0.91 or higher. Calibration was assessed using Hosmer-Lemeshow C -statistic.

Results:

We included data from 2094 consecutive patient episodes. The median age of the cohort was 60 and predominantly male (n=1350, 65.47%). Mechanical Ventilation and vasopressors were administered at admission in 504 (24.44%) and 423 (20.51%) patients respectively. Overall, mortality at ICU discharge was 10.28% (n=212). Discrimination (AUC) for the e-TropICS model was 0.81 (95% CI 0.78-0.84) with an HL C statistic p value of < 0.05. Brier score, used to evaluate overall model accuracy for prediction of mortality was 0.15 (0-14-0.16). The best sensitivity and specificity (72% and 74% respectively) were achieved with the model at an optimal cut-off for probability of 0.37.

Conclusion:

e-TropICS has utility in the care of critically unwell patients in the South Asia region with good discriminative capacity. Further refinement of calibration in larger datasets from India and across the South-East Asia region will help in improving model performance.

Introduction

Severity of illness scoring systems such as the Simple Acute Physiology Score (SAPS)[1] and the Acute Physiology and Chronic Health Evaluation (APACHE)[2] help in risk prediction, benchmarking, quality improvement and patient selection for research. Over the past three decades, several iterations of these

models have been developed and validated based on changes in the epidemiology of critical illness and substantial improvements in survival.[3,4] Most of these models have been evaluated in the context of high-income countries (HICs). There are several limitations to the use of these models in middle income and lower-middle income countries (LMICs) such as differences in epidemiology of critical illness, including the high burden of tropical infections, the lack of resources for data collection, burden of data collection due to the large number of variables, missing variables and the absence of electronic health records that would otherwise facilitate seamless data flow.[5] To overcome these problems, researchers have developed and tested newer simplified models in LMICs.[6,7]

Recently, in Asia, one such simplified model, the e-TropICS (described as 'model 1') has been developed and validated.[8] The researchers, cognizant of the limited availability of variables and consequent high proportion of missingness, which has hampered intensive care units(ICUs) in resource limited settings from being able to utilise existing prognostic scores, aimed to develop a comparable score based on a more universally obtainable set of variables. However, this score has not been externally validated. We, therefore, aimed to validate the performance of the e-TropICS model on a multi-centre Indian data set from the recently established Indian Registry of IntenSive care(IRIS).[9]

Methods

Study setting

Seven ICUs located in 6 private and 1 not-for-profit institution, part of the Indian Registry of IntenSive care(IRIS) contributed data to this study. Of these, 5 were general (mixed medical-surgical) ICUs and two were medical ICUs. The Indian Registry of IntenSive care a cloud-based registry of critical care units was established in Jan 2019.[9] Details of the implementation and preliminary results of the case-mix program have been previously published.[9]

Patients

All patients reported to the registry between January 2019 to May 2019, were considered. Patients > 18 years of age with an ICU length of stay > 6 hours were included in the study. Patients with missing outcomes and those not meeting the inclusion criteria were excluded

Data collection

This retrospective study used data collected as part of the IRIS dataset. Age, gender, pre-existing co-morbidity, diagnostic category, type of admission (planned, unplanned, medical or surgical), physiological vital signs and laboratory measurements were collected as per the definitions described for e-TropICS (Table 1) for all consecutive admissions. ICU outcomes rather than hospital outcomes were collected due to well-described logistical challenges in such settings.[5,8] Data was collected daily by either nursing staff or by data collectors appointed to the registry network, all of whom had been trained in the process of data acquisition. Daily telephone reminders encouraging data input and checks for consistency of the

number of admissions, discharges and outcomes from each ICU were undertaken by staff from the central coordinating centre. In-built measures in the data entry portal such as mandatory fields, range validations, drop down and checkboxes as opposed to free text entries were employed to promote fidelity of data recording.

Ethics:

The study was approved by the Institutional Ethics Committee centrally at the study coordinating centre (AMH-021/07-19). The informed consent model used in the registry has been described and published previously.[9]

Statistical analysis

Availability of physiological and laboratory measurements was described using descriptive statistics. e-TropICS was calculated as per the authors' original methods.[8] The area under the receiver operator characteristic curve (AUROC) was used to express each of the models' power to discriminate between survivors and non-survivors. For all tests of significance, a 2-sided P less than or equal to 0.05 was considered to be significant. AUROC values were considered poor when less than or equal to 0.70, adequate between 0.71 to 0.80, good between 0.81 to 0.90, and excellent at 0.91 or higher.[10] Calibration for each model was assessed using Hosmer-Lemeshow C -statistic. Overall model accuracy in measuring predictions was calculated using Brier Scores. Continuous variables and differences between the means of normally distributed variables were compared using Student's t-test. Chi-square test was employed to compare categorical variables and to compare between AUROC values. All analysis was performed using Stata software version 13.1[11]

Handling of missing data and analysis:

When faced with high proportions of missing data, one approach is to assume normality for a variable when not measured or unavailable, resulting in a score of "0" in weighted scoring systems. Such an approach may not be justified in LMICs where measurements may be unavailable due to lack of resource availability or to differing approaches in decision-making in critical illness. Assumptions of normality in the above manner can adversely impact model performance by underestimating severity scores. In this study, multiple Imputation (MI) with chained equations was employed to handle missing data. It was assumed that the missingness of a variable depends on some of the other observed variables i.e. Missing At Random (MAR). MI was performed using Predictive Mean Matching (PMM) of the "MI impute chained" command in Stata (version 13.1, Stata Corp, College Station, TX, USA). The number of imputations (M) was set at 20 and "k-nearest neighbours" (kNN#, Stata syntax) was set at 10. Multiple Imputation (MI) generates several values reflecting the uncertainty in the estimation of the imputed value. Both continuous and categorical variables were imputed, as PMM generates predicted values that have been drawn from data which has already been observed within the variable. This ensures that categorical variables which can only take specific values do not have predicted values which are not allowed for the variable. The scores (and their mortality probabilities) were then calculated individually for each of the 20

multiple imputed datasets. The mean of 20 probabilities was then calculated and used the MI mortality prediction.

Results

Characteristics of population and availability of variables for the e- TropICS model.

During the evaluation period, 2094 consecutive patient episodes were reported to the IRIS registry from the seven participating centres. Thirty-two patients were excluded; 19 patients were less than 18 years of age and 13 had no outcome information.

The characteristics, demographics and outcomes for these episodes is described in Table 1. The median age of the cohort was 60 and predominantly male (n=1350, 65.47%). Planned admissions accounted for 383 (18.57%) episodes. Mechanical Ventilation and vasopressors were administered at admission in 504 (24.44%) and 423 (20.51%) patients respectively. Overall, mortality at ICU discharge was 10.28% (n=212).

Availability of the variables for the e-TropICS model is described in Table 2. Availability was lowest for blood urea (88.60%) and highest for gender and admission type (100%). For all other variables, availability ranged from 95% to 100%.

Ability of models to predict mortality.

Discrimination (AUC) for the e-TropICS model was 0.81 (95% CI 0.776-0.836) (Table 3) with an HL C statistic p value of < 0.05 suggesting poor model fit. Brier score, used to evaluate overall model accuracy for prediction of mortality was 0.15 (0-14-0.16). The best sensitivity and specificity (72% and 74% respectively) were achieved with the model at an optimal cut-off for probability of 0.37.

Discussion

Our study validates the performance of a simplified prognostic model designed for use in critical care units, where information needed to calculate prognostic models from HIC may be absent or burdensome. In this multi-centre cohort from the recently established IRIS critical care registry, e-TropICS had good ability to discriminate death, but poor calibration. This suggests that in this cohort, the model can identify those patients at greatest risk of death, but has less ability to differentiate between degrees of severity of illness.

Several prognostic models for critically ill adults are validated and in use in HIC healthcare systems. However, their applicability in settings where due to limited point of care testing, data collection resources and even perhaps appropriate judicious use of laboratory tests, remains limited. Whilst missingness can be managed for the purposes of performance assessment and validation, a score which is not easily calculated - has limited application in the clinical setting. The e-TropICS model, developed from a South Asian dataset attempts to overcome several of these challenges By limiting model covariates to clinical

information that would likely be requested by clinicians in the care of acutely unwell patients and that are likely available in all health systems reporting to provide ICU care availability of data for validation was much higher than reported in previous literature from similar middle income settings. Our experience of applying this model to this registry data set shows high availability of the variables required for validation. Availability for most variables ranged from 95–100%. Only blood urea had an availability less than 95%. The decision to perform laboratory tests is also influenced by clinician preference, in addition to access to equipment, disposables, costs and expertise.

e-TropICS model had good discriminative ability (AUC of 0.81), but poor calibration. This is not surprising though, as several well-established prediction models, when validated externally have shown poor calibration.[12] Several reasons could explain this including the limitations of the HL test itself such as a high sensitivity to the sample size.[13,14] Other reasons could include differences in case-mix.[15] Another potential explanation is the lower mortality in our dataset as compared to the dataset from which e-TropICS was developed (10.2% versus 28.4%). Previous research has shown that even small differences in mortality can affect the calibration of a model.[16] Whilst both good discrimination and good calibration are desirable for prognostic scores to be deemed suitable for application in clinical practice, in reality, once an appropriate and implementable model is selected for use in a clinical setting, it can and perhaps should be regularly calibrated for the current population to which it is being applied. Our group is undertaking model refinement in the IRIS and in the recently established nine country critical care registry collaboration (Crit Care Asia) in South and Southeast Asia.

Our study has also demonstrated the feasibility of using registry data in a resource-limited setting to validate a locally relevant model. There remains limited investment in systems that enable routine data collection in LMICs and subsequently efforts to benchmark services and evaluate interventions to improve care remain hampered by low availability of information. The greater availability reported in this validation compared to earlier studies in the region suggests that investment in health system strengthening including the implementation of critical care registries can improve the availability of information during critical care admission.

Conclusion

e-TropICS has utility in the care of critically unwell patients in the South Asia region. e-TropICS thus offers a prediction model that is simplified with low data collection burden for resource-limited settings. Further refinement of calibration of this model in larger datasets from India and across the South-East Asia region will help in improving model performance.

Declarations

(bharath@icuconsultants.com) for such access; alternatively, IRIS can be contacted at info@irisicuregistry.org and joinus@irisicuregistry.org.

References

1. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D et al. A simplified acute physiology score for ICU patients. Crit Care Med 1984;12(11):975-7

2. 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(8):591-7

3. Gall JR, Lemeshow S, Saulnier F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. JAMA J Am Med Assoc. 1993;

4. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;

5. Haniffa R, Isaam I, De Silva AP, Dondorp AM, De Keizer NF. Performance of critical care prognostic scoring systems in low and middle-income countries: a systematic review.

6. Riviello ED, Kiviri W, Fowler RA, Mueller A, Novack V, Banner-Goodspeed VM et al. Predicting Mortality in Low-Income Country ICUs: The Rwanda Mortality Probability Model(MPM). PLOS ONE 2016; https://doi.org/10.1371/journal.pone.0155858

7. Houweling TA, van Klaveren D, Das S, Azad K, Tripathy P, Manandhar D et al. A prediction model for neonatal mortality in low- and middle-income countries: an analysis of data from population surveillance sites in India, Nepal and Bangladesh. Int J Epidemiol 2019;48(1):186-98

8. Haniffa R, Mukaka M, Munasinghe SB, De Silva AP, Saroj K, Jayasinghe A et al. Simplified prognostic model for critically ill patients in resource limited settings in South Asia. Crit Care 2017;21:250

9. Adhikari NKJ, Arali R, Attanayake U, Balasundaram S, Beane A, Chakravarthy V et al. Implementing an intensive care registry in India: preliminary results of the case-mix program and an opportunity for quality improvement and research[version 1; peer review: awaiting peer review]. Wellcome Open Res 2020;5:182

10. Hosmer DW, Lemeshow S. Applied Logistic Regression, 2nd Ed. Chapter 5, John Wiley and Sons, New York, NY (2000), pp. 160-164

11. 2013. Stata Statistical Software: Release 13. College Station , TX:StataCorp LP.

12. Nassar Jr AP, Mocelin AO, Baptiston Nunes AL, Giannini FP, Brauer L, Andrade FM et al. Caution when using prognostic models: A prospective comparison of 3 recent prognostic models. Journal of Critical Care 2012;27:423.e1-423.e7

13. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med 2007;35(9):2052-6

14. Serrano N. Calibration strategies to validate predictive models: is new always better? Intensive Care Med 2012;38:1246-48

15. Glance LG, Osler T, Shinozaki T. Effect of varying the case mix on the standardized mortality ratio and W statistic: a simulation study. Chest 2000;117:1112-7

16. Glance LG, Osler TM, Papadakos P. Effect of mortality rate on the performance of the Acute Physiology and Chronic Health Evaluation II: a simulation study. Crit Care Med 2000;28:3424-8.

Tables

Table 1: Patient characteristics at the time of ICU presentation

| Characteristic n=2062	All (n=2062)	Dead (n=212,10.3%)	Alive (n=1850, 89.7%)
Planned admission (n (%))	383(18.57)	31(14.62)	352(19.03)
Gender male (n (%))	1350(65.47)	136(64.15)	1214(65.62)
LoS$^\$$	3.41(4.20)	3.22(3.91)	5.34(6.34)
Age*	60(18-110)	64(22-92)	60 (18-110)
Heart rate$^\$$	94.35(22.95)	104.50(26.59)	93.18(22.21)
GCS*	14(3-14)	10(3-14)	14(3-14)
Respiratory rate$^\$$	22.83(5.58)	25.70(8.06)	22.50(5.12)
Systolic BP $^\$$	129.20(25.91)	120.53(29.14)	130.19(25.33)
Blood urea$^\$$	51.96(47.72)	72.64(52.79)	49.59(46.4)
Haemoglobin, g/dL$^\$$	11.46(2.77)	10.34(2.62)	11.59(2.76)
Vasopressor used on admission (n (%))	423(20.51)	124(58.49)	299(16.16)
Mechanical ventilation on admission (n (%))	504(24.44)	122(57.55)	382(20.65)

*median (range) $^\$$ mean (sd)

Table 02: Availability of the parameters of e-Tropics model

Parameters	Availability out of 2062, n(%)
Eye_opening	2056(99.71)
Motor_response	2057(99.76)
Verbal_response	2056(99.71)
Mechanically ventilated	2052(99.52)
Vasoactive drugs	2050(99.42)
Systolic BP	2057(99.76)
Respiratory rate	2056(99.71)
Haemoglobin	2013(97.62)
Blood urea	1827(88.60)

Table 03: Performance of the e-tropICS model with multiple imputation

Performance item	MI model N=2062
Probability, mean (SD)	0.29(0.003)
Optimal cut-off probability	0.23
Sensitivity (at optimum cut-off)	0.84
Specificity (at optimum cut-off)	0.72
AUC (95% CI)	0.83(0.812-0.839)
H/L C-statistic (p)	340.87(0.00)
Brier score (95% CI)	0.12(0.120-0.127)

Figures
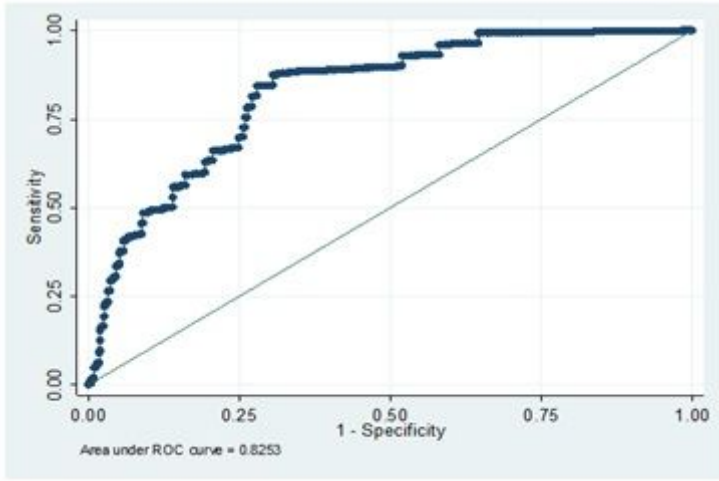
Figure 1

ROC curve for imputed model