Journal of Brain Disorders

Research Article   **Open Access**

Evaluation of Sequential Organ Failure Assessment (SOFA) Performance in Neurocritical Care Patients Overtime: A Retrospective Cohort Study

Mariana Rabelo de Brito[1], Alexandre Guimarães de Almeida Barros[2], Lenise Valler[1], Fabricio Buchdid Cardoso[1], Ana Paula Devite Cardoso Gasparotto[2], Luciana Tiziani Silva[2], Cristina Bueno Terzi Coelho[2], Desanka Dragosavac[2] and Antônio Eiras Falcão[2]*

[1]Department of Neurology, Faculty of Medical Sciences, State University of Campinas (Unicamp), Campinas, São Paulo, Brazil

[2]Department of Surgery, Faculty of Medical Sciences, State University of Campinas (Unicamp), Campinas, São Paulo, Brazil

Abstract

Prediction models are widespread used as surrogate markers of disease severity, benchmarking and resource allocation in critical care. After derivation, they should pass through validation in different settings before their general use in other populations. Moreover, constant assessment of score's performance is necessary to keep them reliable. In this study, we sought to evaluated Sequential Organ Failure Assessment (SOFA) score discrimination and calibration for mortality in a cohort of neurosurgical and neurological patients overtime. Although SOFA showed a good discrimination in all timepoints, our data suggest that its calibration may improve as time pass. Therefore, SOFA is a good model for mortality prediction in neurological and neurosurgical patients and may help for organ dysfunction objective evaluation and benchmarking in neurocritical care.

Keywords

Neurocritical care, Severity scores, SOFA, Critical care medicine

Introduction

Admission to intensive care units is a critical decision for many patients [1]. This is because of the great burden for the patient and family regarding the equilibrium among real benefits and harms cause by it [1,2]. Moreover, critical care is associated with elevated costs for healthcare systems that may increase as therapeutics evolve. Therefore, models that aid for an objective evaluation of endpoints in this setting, allowing to an evidence-based decision process and management, are important to pursue. In this context, predictive scores appear as tools for outcome prediction, severity of illness evaluation, comparison among reference standards, and resource allocation [3-7].

Several scores are available for general critical care patient's stratification. Among them, Acute Physiology and Chronic Health Disease Classification System (APACHE) II, Simplified Acute Physiology Score (SAPS) 3 and Sequential Organ Failure Assessment (SOFA) have been widely used and studies support their reliability for mortality prediction in general ICUs [6,8]. SOFA is a score initially proposed for organ dysfunction quantification in general critically ill patients, that was posteriorly validated for mortality prediction as well [4,9,10]. Studies have shown that values calculated using original SOFA score or composites of it,

like maximum SOFA score or delta of variation in SOFA score, are associated with increased mortality [4,11,12]. In opposition to APACHE II and SAPS 3, that have cumbersome equations and are validated to be calculated at the first 24 hours from admission, SOFA is easier to calculate and may be applied sequentially overtime [8,13,14]. Traditionally, this is applied to patients admitted in the ICU and gives the overall severity of organic dysfunction, based on six different organic system's parameters. It returns the probability of death and gauges illness severity objectively. As any score, SOFA has flaws that may disrupt its predictive accuracy in certain situations [4,15]. For instance, in cirrhotic patients its predictions may be biased by the chronic thrombocytopenia and hyperbilirubinemia observed in advanced stages of the disease [16,17]. Therefore, it is necessary to validate predictive scores in different clinical backgrounds to assure their accuracy. Our group have recently tested SOFA's accuracy at admission of neurological and neurosurgical patients to predict mortality in a neurocritical ICU and found a good performance in this scenario [18].

Critical care patients are subject to many conditions unique to the ICU environment. For example, ventilation associated pneumonia and critical illness polyneuropathy, which are associated with the use of invasive treatments and prolonged ICU stay, may impact morbidity and mortality in critical care [19,20]. Therefore, mortality predicted by severity scores measured at admission may not reflect the observed mortality of the subgroup of patients that remains in the ICU for longer periods. In general ICU patients, SOFA has been used sequentially to estimate how organ dysfunction progress overtime. To test whether SOFA remains an accurate model for mortality prediction in neurological and neurosurgical patients that request more than 48 hours of ICU hospitalization, we assessed the predictive features of the model in different timepoints in a neurocritical care unit.

Materials and Methods

Study design and setting

This was retrospective cohort study using data, collected in a single neurocritical ICU, from January 1th 2013 and December 31st 2016, in Unicamp's teaching hospital. This ICU has seven beds and a full time intensive care specialists, nurses, assistants and physiotherapists. Patients had their medical history of interest fed in an electronic database and their death data corresponding to the hospitalization period. This was an observational study and every clinical decision was at the discretion of the attending physician. Therefore, an informed consent was waived. Local and national ethics committee approved the study's protocol.

Selection of participants and data gathering

During the study period, all patients aged 18 years or older, admitted to the neurocritical ICU of Unicamp's teaching hospital were evaluated. Data were collected from an electronic database protected to hide patient's identification and fed with demographic, clinical and laboratory variables of interest.

SOFA, APACHE II and SAPS 3 were automatically calculated using standard equations with appropriate variables. APACHE II and SAPS 3 were used in descriptive evaluation of patients. Inclusion criteria were to have a primary neurological or neurosurgical diagnosis at ICU admission. Exclusion criteria were individuals transferred from other ICUs and those with missing information for analysis in the database.

Statistical analysis

To evaluate SOFA's accuracy in neurocritical patients, its discrimination and calibration features were assessed in the first 24 hours from patient's admission to the ICU (admission SOFA) and in the third and fifth days of ICU's hospitalization. Statistical analyses were performed using MedCalc version 17.8.1. Continuous variables were presented as median and range. Categorical variables were presented as absolute values and percentages. Score discrimination was evaluated by calculating the Area Under the Receiver Operating Characteristic Curve (AUROC), with 95% Confidence Intervals (CI) and compared with the DeLong method [21]. The discrimination was considered excellent, very good, good, moderate and poor with AUROC values of 0.9-0.99, 0.8-0.89, 0.7-0.79, 0.6-0.69 and < 0.6, respectively. The model's calibration for probability of death was evaluated by the Hosmer-Lemeshow goodness-of-fit test [22]. A p value < 0.05, two-tailed and 95% CI were used for logistic regressions.
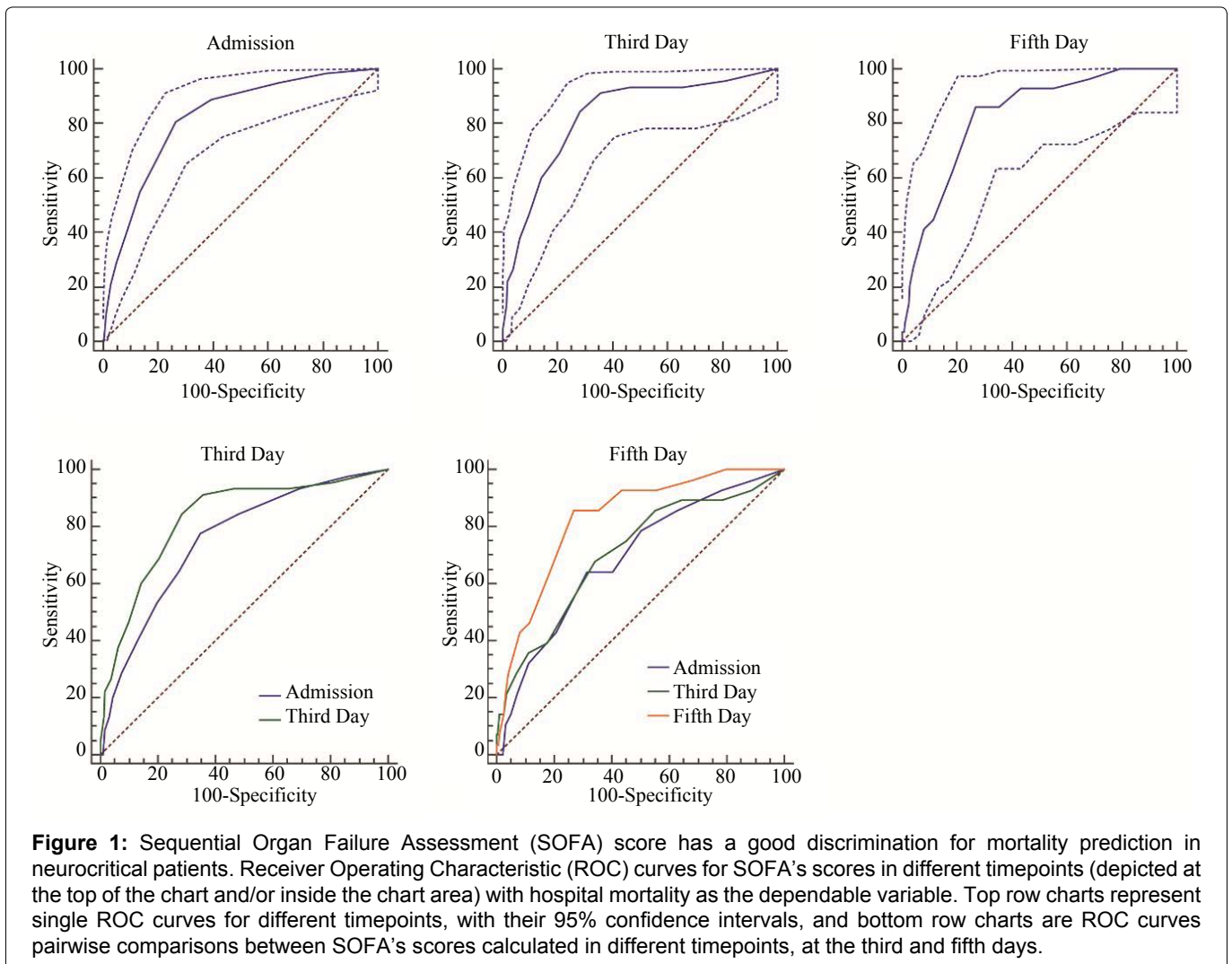
Results

A total of 978 patients fulfilled the inclusion criteria and were enrolled for further analyzes. Age and sex distribution are found in Table 1. Patients admitted for neurosurgical procedures include 70.1% of total and for neurological reasons 29.9%. Median ICU length of stay was four days and hospital length of stay 13 days. Observed in-hospital mortality was 6.3% and median prognostic scores values were 10, 35 and 2 for APACHE II, SAPS 3 and SOFA respectively. Baseline patient's characteristics are found in Table 1 and overall SOFA distribution in Supplementary Figure 1.

SOFA's discrimination was good in all timepoints. Area Under the ROC curves (AUROC) were 0.82 (95% CI: 0.795 to 0.844) for admission SOFA, 0.827 (95% CI: 0.795 to 0.856) for third day SOFA and 0.827 (95% CI: 0.779 to 0.869) for fifth day SOFA (Figure 1). Next, we sought to measure SOFA's values in different timepoints and compared them with each other. Results showed that

Table 1: Baseline patient's characteristics.

Characteristic	All	Admission (24 hours)	Third day	Fifth day
Sample size, n	978	978	624	289
Age, median (IQR)	51 (37-61)	51 (37-61)	51 (38-62)	52 (40-63)
Male, n (%)	471 (48.2%)	471 (48.2%)	297 (47.6%)	137 (47.4%)
Admission type, n (%)				
Neurological	292 (29.9%)	292 (29.9%)	235 (37.7%)	145 (50.2%)
Neurosurgical	686 (70.1%)	686 (70.1%)	389 (62.3%)	144 (49.8%)
Length of ICU stay, median (IQR)	4 (2-8)	4 (2-8)	6 (4-13)	11 (6-20)
Length of hospital stay, median (IQR)	13 (8-26)	13 (8-26)	17 (10-31)	27 (15-44)
Severity scores, median (IQR)				
APACHE II	10 (7-14)	10 (7-14)	12 (8-15)	13 (10-17)
SAPS 3	35 (25-46)	35 (25-46)	39 (28-52)	46 (32-56)
SOFA	2 (1-4)	2 (1-4)	2 (1-5)	3 (1-6)
Hospital mortality, n (%)	62 (6.3%)	62 (6.3%)	45 (7.2%)	29 (10%)

Figure 1: Sequential Organ Failure Assessment (SOFA) score has a good discrimination for mortality prediction in neurocritical patients. Receiver Operating Characteristic (ROC) curves for SOFA's scores in different timepoints (depicted at the top of the chart and/or inside the chart area) with hospital mortality as the dependable variable. Top row charts represent single ROC curves for different timepoints, with their 95% confidence intervals, and bottom row charts are ROC curves pairwise comparisons between SOFA's scores calculated in different timepoints, at the third and fifth days.

model's discrimination remained stable over time. For admission SOFA, AUROC were 0.753 (95% CI: 0.718 to 0.787) in the third day and 0.688 (95% CI: 0.631 to 0.741) in the fifth day. For third day SOFA, AUROC was 0.707 (95% CI: 0.650 to 0.759) in the fifth day (Figure 1). Pairwise evaluation achieved significance against comparisons with SOFA values of the day according to the

DeLong test (Third day: admission vs. third day SOFA p = 0.0243/Fifth day: admission vs. fifth day SOFA p = 0.0127; third day vs. fifth day SOFA p = 0.0041; admission vs. third day SOFA p = 0.6941). These observations suggest that SOFA measured daily has a better discrimination than values from the past, being able to separate patients most likely to evolve to a negative outcome. For

Table 2: Hosmer & Lemeshow goodness of fit test for Sequential Organ Failure Assessment (SOFA) score overtime. A p value equal to or more than 0.05 indicates insignificance and that the score is probably calibrated. If the p value is above 0.1, it is suggested a good calibration of the score.

SOFA	Hosmer & Lemeshow test, chi-squared	Hosmer & Lemeshow test, p value
Admission	96.840	0.0847
Third day	86.771	0.1926
Fifth day	97.029	0.2060

example, using a cutoff value of seven which has near 95% specificity and low sensitivity, SOFA has a positive likelihood for death of 6.04 (95% CI: 3.7 to 9.8) at admission, 4.74 (95% CI: 3.2 to 7.1) at third and 3.89 (95% CI: 2.3 to 6.6) at fifth day in our population.

A previous study from our group suggests that SOFA has a poor calibration compared to other severity scores [18]. This means that SOFA may predict a risk of death different than that observed. To evaluate for model's predictive agreement, we sought to assess SOFA's calibration in different timepoints. Observed p values suggest that model's calibration is acceptable at admission and may increase overtime (Table 2). Perhaps the number and features of patients analyzed changes overtime as they are discharged or died and may explain observed data. Our result suggests that SOFA may predict better in neurocritical patients who have a longer hospitalization in the ICU.

Discussion

In this study, we assessed SOFA's mortality prediction in different timepoints of a cohort of neurocritical patients. We observed that SOFA has a good discrimination and may increase its calibration as time pass. Moreover, our data suggests that SOFA should be measured daily to keep its calibration stable and trustworthy. Our group compares SOFA to APACHE II and SAPS 3 in a similar cohort of patients and observed a poor calibration compared to APACHE II for mortality in the ICU [18]. However, we showed that SOFA may be better calibrated for predictions in patients after the first 24 hours of ICU hospitalization. Perhaps changes in cohort characteristics overtime may fits better in SOFA's calibration. This may be justified by new organ dysfunctions appearance and resolution or as patients are discharged or die. For example, patients may develop ventilator associated pneumonia after prolonged periods in mechanical ventilation or other critical care related complication during ICU hospitalization that increases morbidity and mortality. Therefore, SOFA calculated in different timepoints can gauge those events better and predict more accurately than previous calculated values. Another explanation is that as the sample size reduces as time pass, Hosmer-Lemeshow goodness of fit statistics may lose precision and may not be able to indicate a poor calibration [23,24].

Severity scores are good surrogates for illness severity assessment and gives an objective measurement of patient's evolution [10,14]. However, studies have showed that prediction models should be validated prior to extrapolation to different populations [5,24-26]. This is because they are grounded on reference databases which are used to set the standards for comparison and assessment of outcomes. Studies suggest that those databases are subject to heterogeneity, regional variation and loss of representativeness over time [5,24-26]. Moreover, evolving clinical practice and standards of care may diminishes model's accuracy and restrict their applicability. SOFA is often used in general critical care and its prediction overtime has not been validated in a cohort of neurological and neurosurgical patients. Our data suggests that SOFA keeps its accuracy when measured sequentially and may be useful to stratify neurocritical patients.

It is well established that SOFA may be used for mortality prediction in general ICU patients. Several studies showed that as SOFA score increases its calculated values, mortality follows a similar pattern [4,10,11,13]. Our data support this trend in a specific subgroup of neurological and neurosurgical patients and adds perspectives about mortality prediction in this setting. However, several limitations should be stressed. This was a single-center retrospective study that must be reproduced in other centers and populations to increase reliability in the results. Moreover, although data was collected following a strict protocol, it was a snapshot of a moment and it is not possible to assure we were using the worse variable values in each day of score's calculation. This fact may reduce score's accuracy. One way to overcome it, is to use maximum values for variables measured three times daily for example, although this strategy makes score's calculation more costly and cumbersome.

Prognostic models are important for benchmark in critical care. Even though they are not used in an individual fashion, they may provide the attendant physician with an objective measurement of patient's organ dysfunction. Our results suggest that SOFA has good discrimination and calibration for mortality prediction in neurocritical patients and may be used in this setting.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

References

1. Nates JL, Nunnally M, Kleinpell R, et al. (2016) ICU admission, discharge, and triage guidelines: A framework to enhance clinical operations, development of institutional policies, and further research. Crit Care Med 44: 1553-1602.

2. Orsini J, Butala A, Ahmad N, et al. (2013) Factors influencing triage decisions in patients referred for ICU admission. J Clin Med Res 5: 343-349.

3. Minne L, Ludikhuize J, de Jonge E, et al. (2011) Prognostic models for predicting mortality in elderly ICU patients: A systematic review. Intensive Care Med 37: 1258-1268.

4. Minne L, Abu-Hanna A, de Jonge E (2008) Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. Crit Care 12: 161.

5. Salluh JIF, Soares M (2014) ICU severity of illness scores: APACHE, SAPS and MPM. Curr Opin Crit Care 20: 557-565.

6. Gartman EJ, Casserly BP, Martin D, et al. (2009) Using serial severity scores to predict death in ICU patients: A validation study and review of the literature. Curr Opin Crit Care 15: 578-582.

7. Keegan MT, Gajic O, Afessa B (2011) Severity of illness scoring systems in the intensive care unit. Crit Care Med 39: 163-169.

8. Sakr Y, Krauss C, Amaral AC, et al. (2008) Comparison of the performance of SAPS II, SAPS 3, APACHE II, and their customized prognostic models in a surgical intensive care unit. Br J Anaesth 101: 798-803.

9. Vincent JL, Moreno R, Takala J, et al. (1996) 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 22: 707-710.

10. Moreno R, Vincent J, Matos R, et al. (1999) The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working Group on Sepsis related Problems of the ESICM. Intensive Care Med 25: 686-696.

11. Ferreira FL, Bota DP, Bross A, et al. (2001) Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 286: 1754-1758.

12. Raith EP, Udy AA, Bailey M, et al. (2017) Prognostic accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. JAMA 317: 290-300.

13. Badreldin A, Elsobky S, Lehmann T, et al. (2012) Daily-mean-SOFA, a new derivative to increase accuracy of mortality prediction in cardiac surgical intensive care units. Thorac Cardiovasc Surg 60: 43-50.

14. Khwannimit B (2007) A comparison of three organ dysfunction scores: MODS, SOFA and LOD for predicting ICU mortality in critically ill patients. J Med Assoc Thai 90: 1074-1081.

15. Lamia B, Hellot MF, Girault C, et al. (2006) Changes in severity and organ failure scores as prognostic factors in onco-hematological malignancy patients admitted to the ICU. Intensive Care Med 32: 1560-1568.

16. Saliba F, Ichai P, Levesque E, et al. (2013) Cirrhotic patients in the ICU: Prognostic markers and outcome. Curr Opin Crit Care 19: 154-160.

17. Karvellas CJ, Bagshaw SM (2014) Advances in management and prognostication in critically ill cirrhotic patients. Curr Opin Crit Care 20: 210-217.

18. de Almeida Barros AG, Mescolotte GM, de Souza OA, et al. (2017) Severity scores performance in Neurocritical Care: A retrospective cohort study. Open Access J Neurol Neurosurg 5: 5-8.

19. Hermans G, De Jonghe B, Bruyninckx F, et al. (2008) Clinical review: Critical illness polyneuropathy and myopathy. Crit Care 12: 238.

20. Latronico N, Shehu I, Seghelini E (2005) Neuromuscular sequelae of critical illness. Curr Opin Crit Care 11: 381-390.

21. DeLong E, DeLong D, Clarke-Pearson D (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. JSTOR: Biometrics 44: 837-845.

22. Lemeshow S, Hosmer DW Jr (1982) A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 115: 92-106.

23. Paul P, Pennell ML, Lemeshow S (2013) Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. Stat Med 32: 67-80.

24. Nassar AP, Mocelin AO, Nunes AL, et al. (2012) Caution when using prognostic models: A prospective comparison of 3 recent prognostic models. J Crit Care 27: 423.e1-423.e7.

25. Breslow MJ, Badawi O (2012) Severity scoring in the critically ill: Part 1--interpretation and accuracy of outcome prediction scoring systems. Chest 141: 245-252.

26. Kramer AA, Zimmerman JE (2007) Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med 35: 2052-2056.
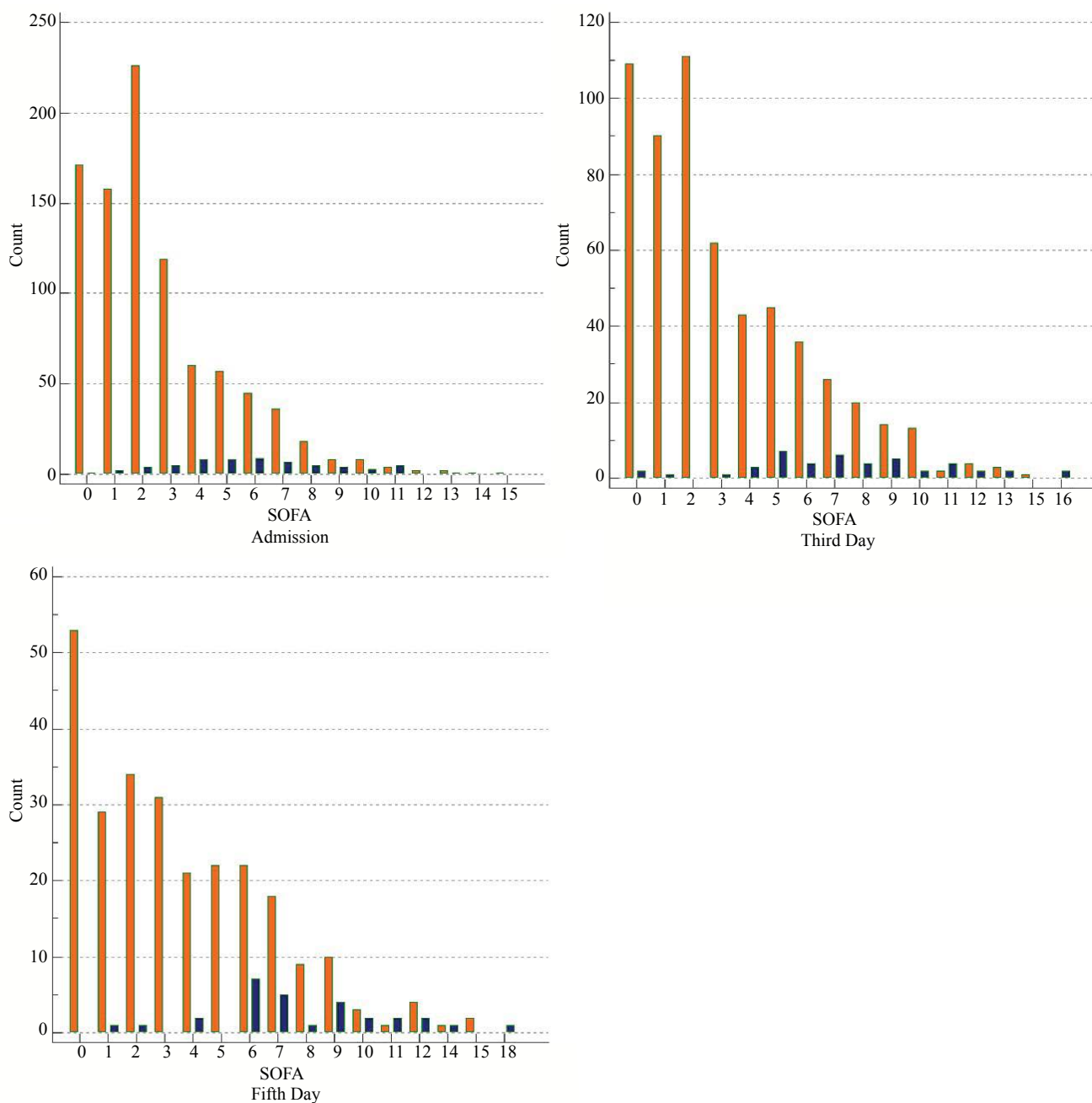
Supplementary Figure 1: Sequential Organ Failure Assessment (SOFA) score distribution and outcomes overtime. Charts show frequencies of survivors (orange bars) and non-survivors (blue bars) according to SOFA score's values (X-axis) in different timepoints. Noteworthy, y-axis scale has diverse ranges in each graph.

SCHOLARS.DIRECT

de Brito et al. J Brain Disord 2017, 1(1):38-43   ISSN: 2578-6571 | • Page 43 •