Original Article

The prognostic implication of the SOFA score on acutely admitted medical patients

L Holm & M Brabrand

Abstract

Objectives: *To investigate the prognostic implication of the SOFA score on all acutely admitted medical patients, to see how well it could predict 30-day mortality and ICU- admission.*

Main outcome measures: *Discriminatory power was calculated as AUROC. Calibration was assessed using Hosmer-Lemeshow goodness-of-fit.*

Results: *AUROC for 30-day mortality was 0.68, (95% CI, 0.64 to 0.71) and ICU admission 0.71, (95% CI, 0.66 to 0.76). Goodness of fit for 30-day mortality and ICU admission was acceptable.*

Conclusions: *The SOFA score showed average ability to predict 30-day mortality and ICU admission with acceptable calibration. When substituting GCS with AVPU the performance of the SOFA score was unacceptable.*

Keywords

Sepsis, Sequential Organ Failure Assessment, emergency department, mortality, intensive care unit

Abbreviations

AUROC, area under receiver characteristics curve; ED, emergency department; GCS, Glascow Coma Scale; ICU, intensiv care unit; IQR, interquartil range; OR, odds ratio; PF, PaO2/FiO2; SF, SpO2/FiO2; SOFA, Sequential Organ Failure Assessment

Highlights

- The Sequential Organ Failure Assessment (SOFA) score is normally applied in the ICU
- Septic patients often present in the ED needing quick assessment
- We tested the scores ability to predict mortality and ICU-admission in ED patients
- Patients with missing variables were included in the calculation
- The score showed average predictive abilities for both outcomes
- An important finding was that the performance of the SOFA score is highly dependent on the use of the Glascow coma scale

Louise Holm
MD, Department of Surgery, Vejle Hospital

Mikkel Brabrand
MD PhD clinical associate professor - University of Southern Denmark and Research Director - Department of Emergency Medicine, Hospital of South West Jutland

Correspondance
Mikkel Brabrand
MD PhD
Department of Emergency Medicine, Odense University Hospital
Postal and clinical address: Sdr. Boulevard 29, Entrance 64, ground floor, DK-5000 Odense C, Denmark
Department of Emergency Medicine, Hospital of South West Jutland,
Postal and clinical address: Finsensgade 35, DK-6700 Esbjerg, Denmark
Email: mikkel.brabrand@rsyd.dk

Introduction

The definition of sepsis was in February 2016 redefined by The Third International Consensus Definitions Task Force as life-threatening organ dysfunction caused by a dysregulated host response to infection.[1] Sepsis remains a major worldwide health problem with a high mortality rate, causing potentially 5.3 million deaths annually.[2, 3]

Numerous scoring systems have been developed to assess the level of organ dysfunction, such as the Sequential Organ Failure Assessment (SOFA) score. The SOFA score is used to determine the degree of organ dysfunction based on six different organ specific parameters (i.e. the cardiovascular, respiratory, hepatic, coagulation, renal and neurological systems). From these parameters both the number and severity of organ dysfunction is calculated (See table 1).[4, 5] A SOFA score of 2 points or more indicates organ dysfunction and is associated with a mortality risk exceeding 10% during admission for patients with presumed infection in the general hospital population.[1]

The SOFA score is most commonly used in the intensive care unit (ICU) where septic patients represent a large proportion of the admissions. However, since the majority of patients are initially assessed in the emergency department (ED) prior to admission in the ICU, the ED plays an essential role in identifying patients with sepsis.[6] Immediate diagnosis is very important, since early onset of treatment in the ED prior to admission to the ICU decreases mortality in patients with severe sepsis.[7] Nevertheless, previous studies have shown that many of the scoring systems routinely used in the ICU setting have poor predictive abilities when applied in the ED.[8]

The aim of this study was to examine how well the SOFA score, determined at the time of admission in the ED, predicted admission to ICU and the 30-day mortality of all acutely admitted medical patients.

© 2018 Rila Publications Ltd.

Table 1. Sequential Organ Failure Assessment (SOFA) score[a]

System	Score				
	0	1	2	3	4
Respiration (PaO_2/FiO_2, mmHg (kPa))	≥400 (53.3)	<400 (53.3)	<300 (40)	<200 (26.7) with respiratory support	<100 (13.3) with respiratory support
Coagulation (Platelets, x $10^3/\mu L$)	≥150	<150	<100	<50	<20
Liver (Bilirubin, mg/dl (µmol/L))	<1.2 (20)	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	>12.0 (204)
Cardiovascular	MAP ≥70 mmHg	MAP <70 mmHg	Dopamine <5 or dobutamine (any dose)[b]	Dopamine 5.1-15 or epi ≤0.1 or norepi ≤0.1	Dopamine >15 or epi >0.1 or norepi >0.1
Central nervous system (Glascow Coma Scale)	15	13-14	10-12	6-9	<6
Renal (Creatinin, mg/dl (µmol/L) Urine output, ml/d)	<1.2 (110)	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-400) <500	>5.0 (440) <200

Abbreviations: FiO_2, fraction of inspired oxygen; MAP, mean arterial pressure; PaO_2, partial pressure of oxygen; epi, epinephrine; norepi, norepinephrine

[a] Adapted from Vincent JL et al[5], [b]Cathecholamin doses are given as µg/kg/min for at least 1 hour

Methods

Study design and setting:

An observational study that was a post hoc analysis from a database of all adult patients admitted to the medical unit of a regional 460-bed teaching hospital in western Denmark.[9] The database was formed in the period from October 2008 through February 2009 and February through May 2010.[10] It was made over two time periods since two cohorts were needed to develop and validate a score for another study.[11]

The hospital serves a population of 220,000 and all subspecialties in internal medicine are represented. Patients can be admitted through their general practitioner, out-of-hours emergency medical service, outpatient clinics, emergency department and ambulance services.

Participants

All adult patients (age ≥15 years) admitted at the medical admission unit within the study's two periods were included in this study.

Data sources/measurements

Upon arrival at the medical admission unit a nurse registered vital signs, primary complaint and demographic information. After inclusion to the study initial blood test results were extracted from the electronic health records.

Through the unique Danish personal identification number we were able to retrieve data from the national databases.[12] Data on length of hospital admission as well as preceding and current discharge diagnoses were obtained from the Danish National Patient Register[13] and data on deaths were obtained from The Danish Civil registration System.[12]

Statistical analysis

The SOFA score was calculated upon arrival at the ED post hoc. The respiratory component in the SOFA score is based on the PaO_2/FiO_2 ratio (PF). However, it is not always possible to obtain arterial blood gasses from all of the patients in the ED and thus the SpO_2/FiO_2 ratio (SF) is used as an alternative respiratory score in all of our SOFA score estimations. The SF has shown to correlate with the PF and when compared the PF and SF predict outcomes similarly.[14]

To enable us to calculate the SOFA score on all 5035 patients, missing values were assigned a score of zero.

In our sensitivity analyses we calculated the SOFA score separately on 790 patients for which all the score's variables (except for the replacement of SpO_2/FiO_2) were known. We also calculated the SOFA score in 1462 using the equivalent AVPU scale instead of the Glasgow Coma Scale.[15] All patients with missing data besides the GCS were excluded.

© 2018 Rila Publications Ltd.

In parts of further analysis and comparison, the scores were divided into arbitrary subgroups by the number of SOFA score points; 0-1, 2-4, 5-7, 8+.

It was not possible to achieve data on the use of vasopressors from the database. Since the use of vasopressors is one of the parameters in the SOFA score a theoretical maximum of points could not be achieved in either of the three calculations of the score.

To assess the SOFA score's ability to identify patients at increased risk of dying (the discriminatory power) we calculated the area under receiver characteristics curve (AUROC). This represents different measures of sensitivity/specificity pairs corresponding to each possible cut-off and aims to be above 0.8 to represent acceptable discriminatory power.[9]

The calibration (precision) was assessed by applying the Hosmer-Lemeshow goodness-of-fit test. This was used to assess if the observed mortality rate matched the expected rate in our study. The calibration is considered acceptable if the p-value is above 0.05.[9]

We used a Kaplain-meier survival plot to estimate survival according to time depending on the score subgroup.

To assess the patient's comorbidity, we used the Charlson Comorbidity Score which is commonly used and a well validated method of measuring comorbidity.[16]

Data are presented in interquartile range (IQR) or number (proportion %) when applicable. Odds ratio (OR) are calculated for the three scores according to outcomes.

Stata 14.0 (Stata corp., Texas, USA) was used for statistical analysis.

Ethics

The study was approved by the Danish Data Protection Agency. According to Danish regulation approval from the Regional Committee on Health Ethics was not required. All patient information was anonymized prior to analysis and thus written consent from the patients to participate in the study was not necessary.

Results

Participants

A total of 5035 participants were included in the study cohort. Among these, 2561 (51%) were female and the median age was 65 years (48-77). The median length of stay was 1 day (1-6). Two hundred forty-nine patients (5%) died within 30 days of admission and 106 (2%) were admitted to the ICU. The median Carlson comorbidity score was 1 (0-3).

The SOFA score

Five thousand and thirty-five patients (100%) were included in the calculation of the SOFA score, where missing values were assigned a score of zero ($SOFA^{imputed}$). The median score was 0 (0-1). Patients with a SOFA score ≥ 2 had on average a 4.3-fold increase in 30-day mortality (95% CI, 3.3 to 5.7), P < 0.001, when compared with patients with a SOFA score of 0. The same group of patients had a 5.2-fold increase in ICU admission (95% CI, 3.5 to 7.6), P < 0.001.

The baseline characteristics of patients in the $SOFA^{imputed}$ are presented in table 2.

Table 2. Patient characteristics and $SOFA^{imputed}$ (total and subgroups)

	All patients N = 5035 (100%)	SOFA 0-1 N = 4284 (85%)	SOFA 2-4 N = 673 (13%)	SOFA 5-7 N = 74 (2%)	SOFA 8+ N = 4 (0.1%)
Age	65	64	72	73	70.5
(Median, IQR)	(48-77)	(47-76)	(60-82)	(61-81)	(56.5-73.5)
Sex, female	2561	2183	341	35	2
(Female, %)	(51%)	(51%)	(51%)	(47%)	(50%)
LOS	1	1	5	6	8
(Median, IQR)	(1-6)	(1-5)	(1-10)	(2-12)	(0.5-15.5)
Charlson score	1	1	2	3	3.5
(Median, IQR)	(0-3)	(0-2)	(1-4)	(1-5)	(1.5-6)
30-day mortality	249	148	77	21	3
(Number, %)	(5%)	(3%)	(11%)	(28%)	(75%)
95% CI	0.04 - 0.06	0.03 - 0.04	0.09 - 0.14	0.19 - 0.40	0.04 - 1.00
ICU	106	57	34	14	1
(Number, %)	(2%)	(1%)	(5%)	(19%)	(25%)
95% CI	0.02 - 0.03	0.01 - 0.02	0.04 - 0.07	0.11 - 0.30	0.01 - 0.96

CI: 95% confidence interval

© 2018 Rila Publications Ltd.

Table 3. Summary measures of the SOFA^{complete}, the SOFA^{imputed}, the SOFA^{AVPU} and sensitivity analysis.

	Discriminatory power		Calibration	
	AUROC for 30-day mortality	AUROC for ICU admission	GOF for 30-day mortality	GOF for ICU admission
SOFA^{complete} (N = 5035)	0.68 (0.64 to 0.71)	0.71 (0.66 to 0.76)	0.09	0.53
SOFA ^{Imputed} (N = 760)	0.71 (0.64 to 0.78)	0.84 (0.75 to 0.93)	0.02	0.07
SOFA^{AVPU} (N = 1462)	0.61 (0.54 to 0.68)	0.66 (0.56 to 0.76)	0.00	0.00

GOF = Goodness of fit test (Hosmer-Lemeshow^2 test)
Area under the receiver-operating curve (AUROC) above 0.8 represents exceptional discriminatory, p-value above 0.05 represents acceptable calibration.

Summary measures

The primary SOFA score with missing values adding 0 to the score, the SOFA^{imputed}, showed an AUROC for 30-day mortality of 0.68, (95% CI, 0.6 to 0.7) and for ICU admission an AUROC of 0.71, (95% CI, 0.7 to 0.8) (see table 3). The goodness of fit test for 30-day mortality, P = 0.092 and for ICU admission, P = 0.528. The SOFA score thereby showed an average ability to predict mortality and ICU admission with acceptable precision for 30-day mortality and good precision for ICU-admission.

Sensitivity analysis

The score with all missing data excluded, the SOFA^{complete}, was calculated on 760 patients (15.1% of the cohort). The median score in this category was 0 (0-1). Patients with a score ≥ 2 had in average a 1.16-fold increase in 30-day mortality when compared with patients with a score of 0 (95% CI, 0.761 to 1.772), P = 0.489. The same group has a 1.84-fold increase in ICU admission (95% CI, 0.845 to 3.976), P = 0.122.

The score showed good to exceptional ability to identify patients at risk of dying and an acceptable calibration for ICU admission, but very poor calibration in 30-day mortality, which could be caused by fewer deaths than expected (see table 3).

The score with the substitution of GCS for AVPU, the SOFA^{AVPU}, was calculated on 1462 patients (29.0% of the cohort). The median score was 1 (1-2). A score ≥ 2 indicated a 1.30-fold increase in 30-day mortality on average compared with a score of 0 (95% CI, 0.916 to 1.841), P = 0.142 and a 1.25-fold increase in ICU admission (95% CI, 0.742 to 2.115), P = 0.398.

The score failed in both discriminatory power and calibration (see table 3).

Missing data

Bilirubin (74% missing) and GCS (50% missing) were the major categories with incomplete data. PaO_2 (74% missing) had no impact on the calculations since this variable was replaced with SpO_2/FiO_2 ratio (3% missing).

Discussion

We performed a post-hoc study to investigate the prognostic ability of the SOFA score used on all acutely admitted medical patients in the ED. We found that the SOFA score showed an average ability to predict 30-day mortality or ICU admission when calculated on patients with incomplete data, where missing values were assigned a score of zero. We called it SOFA^{imputed}. Calibration was acceptable for 30-day mortality and good for ICU-admission.

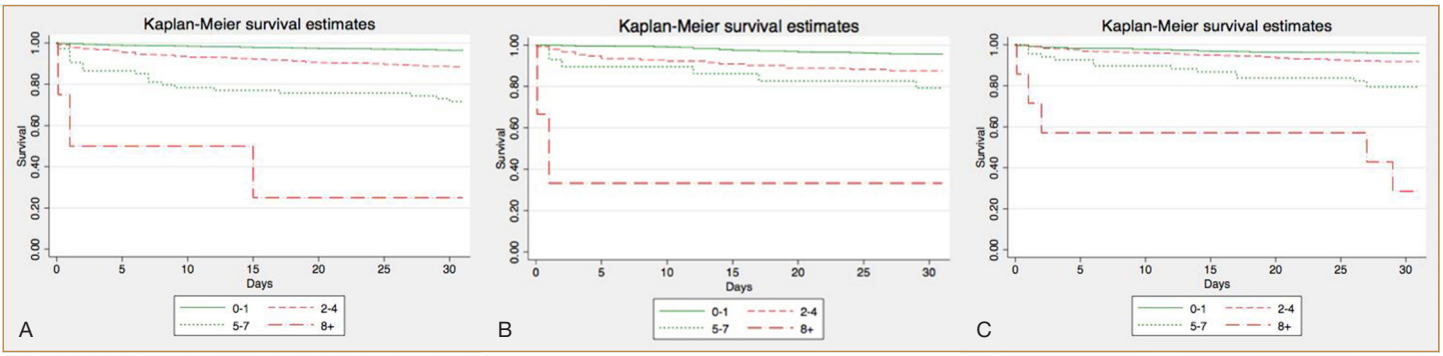

Figure 1. Kaplan Meier survival plot comparing 30-day mortality of the subgroups in the SOFA scores. A: SOFA^{imputed}, B: SOFA^{complete}, C: SOFA^{AVPU}

© 2018 Rila Publications Ltd.

The SOFA score is routinely used in the ICU, and there are challenges in converting it into an ED- based scoring system. First of all, not all patients have blood samples drawn when admitted to the ED, and the laboratory and clinical variables used in the SOFA score are not all routinely used in this setting. Secondly, the SOFA score cannot provide an immediate assessment of the patient since laboratory tests can delay calculation of the score. Thirdly, applying an ICU-based scoring system outside the ICU could alter the performance of the predictive models.

Previous studies investigating the use of the SOFA score applied in the ED have found a good ability to identify patients at risk of dying with AUROC's between 0.75 and 0.79,[1, 6] these studies were performed on patients with confirmed infection, severe sepsis, septic shock or organ failure. In comparison our study found an acceptable AUROC of 0.68 and was performed on all acutely admitted medical patients. However, the patients included in our study were probably less severely ill than those in the abovementioned studies. The comparison of these results could support the issues concerning the use of an ICU-developed scoring system on ED patients in a less critical state[8] and might explain the lower AUROC of our SOFA score. That case mix and cohort mortality can have an impact on the abilities of the predictive models[8, 17] only amplifies the importance of noticing the patient cohort on which the score has been applied. To our knowledge this study is the first to investigate the prognostic implication of the SOFA score applied on all acutely admitted medical patients in an ED. By including all medical patients admitted to the ED it is insured that all patients with infection are in fact in the cohort. Clinical deterioration is thereby not required and thus the full predictive ability of the SOFA score is investigated.

The $SOFA^{imputed}$ had an average ability to identify ICU-admission with an AUROC of 0.71. The best predictive ability was seen in the $SOFA^{complete}$ which showed an AUROC of 0.84. Previous studies have investigated the relation between the SOFA score and ICU-mortality, but not the SOFA scores ability to predict ICU-admission alone. To our knowledge this study is the first to describe this association.

However, the assumption must be made that the patients with all parameters measured must have been thought to be someone very sick and therefor had a complete SOFA score done. With this in mind it is hardly surprising that the SOFA score excluding missing data has the best predictive ability for both outcomes.

The three methods of calculating the SOFA score in this study had different technical advantages and disadvantages. The $SOFA^{imputed}$ and the $SOFA^{AVPU}$ were possible to calculate even with missing data. This could be perceived as an advantage, yet the $SOFA^{imputed}$ performed only with an average predictive ability and the $SOFA^{AVPU}$ failed in both discriminatory power and calibration (see table 3). This indicates that the calculation of the $SOFA^{imputed}$ can be used in the ED, but that none of the methods including missing data are ideal for calculating the score.

Complete data are preferable and the $SOFA^{complete}$ did in this study perform with the highest AUROC for both outcomes (see table 3).

A technical disadvantage in the score itself is the use of the GCS as one of the parameters. The GCS is highly operator dependent and not easy to perform[18]. Substituting the GCS with AVPU in our study made the score fail in both summary measures. This is a very important finding. That these factors seem depending on the GCS raises big concerns of the use of the SOFA score as a practical universal screening tool in settings like the ED and general wards where employees are not skilled in the use of the GCS.

One of the strengths in our study was our wide inclusion due to an unselected study cohort. We investigated the prognostic implication of the SOFA score on all medical patients and not only on patients with confirmed sepsis. This broadened our capacity to test the SOFA score's ability to identify patients at risk and thus widened the existing knowledge upon the use of the score in an ED setting.

Another strength was the use of the Danish databases and the Danish personal identification number to insure complete and precise follow-up.[12, 13]

There were limitations to this study. One of them was the substantial amount of missing data when calculating the score. Had we had all the missing data it is likely that the results of the SOFA score had been altered to higher scores, generating a higher maximum of SOFA score points than described in our results. In the calculations of the $SOFA^{complete}$ this was not the case though, since it did not alter the maximum of points. However, the possibility of the above-mentioned alteration of the SOFA score results cannot be ruled out.

A second limitation was the adaption of the respiratory score to the use of the $SpO_2/Fi0_2$ ratio instead of the original PaO_2/FiO_2 ratio. It is possible that our results had been altered if we had used the original ratio.

A third limitation was the lacking possibility to achieve a maximum of SOFA score points in our study, since we did not have data on the use of vasopressors or urine output in the cardiovascular and renal category respectively. Though, judged by the maximum of points in our study it would still not have been an option to reach 24 points in our case mix.

© 2018 Rila Publications Ltd.

A fourth limitation was the lack of knowledge upon pre-existing organ dysfunction causing an initial SOFA ≥ 2. This could result in a higher score without it being caused by infection. A final limitation was the use of post-hoc analysis.

Conclusion

The SOFA score demonstrated an average ability to predict 30-day mortality and ICU admission when applied to all acutely admitted medical patients in an ED. The score was calculated with inclusion of patients with incomplete data, where missing values added 0 to the SOFA score. We called it SOFA^{imputed}. Calibration was acceptable for 30-day mortality and good for ICU-admission. We also found that the performance of the SOFA score is highly dependent on the use of the Glascow coma scale.

Conflict of interest

None declared.

References

1. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, *et al.* Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Jama.* 2016;**315(8)**:762-74.
2. Vincent JL, Marshall JC, Namendys-Silva SA, Francois B, Martin-Loeches I, Lipman J, *et al.* Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. *The Lancet Respiratory medicine.* 2014;**2(5)**:380-6.
3. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, *et al.* Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations. *American journal of respiratory and critical care medicine.* 2016;**193(3)**:259-72.
4. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca 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 medicine.* 1996;**22(7)**:707-10.
5. Vincent JL, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, *et al.* Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Critical care medicine.* 1998;**26(11)**:1793-800.
6. Macdonald SP, Arendts G, Fatovich DM, Brown SG. Comparison of PIRO, SOFA, and MEDS scores for predicting mortality in emergency department patients with severe sepsis and septic shock. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine.* 2014;**21(11)**:1257-63.
7. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, *et al.* Early goal-directed therapy in the treatment of severe sepsis and septic shock. *The New England journal of medicine.* 2001;**345(19)**:1368-77.
8. Jones AE, Fitch MT, Kline JA. Operational performance of validated physiologic scoring systems for predicting in-hospital mortality among critically ill emergency department patients. *Critical care medicine.* 2005;**33(5)**:974-8.
9. Straede M, Brabrand M. External validation of the simple clinical score and the HOTEL score, two scores for predicting short-term mortality after admission to an acute medical unit. *PLoS One.* 2014;**9(8)**:e105695.
10. Brabrand M, Knudsen T, Hallas J. Identifying admitted patients at risk of dying: a prospective observational validation of four biochemical scoring systems. *BMJ open.* 2013;**3(6)**.
11. Brabrand M, Lassen AT, Knudsen T, Hallas J. Seven-day mortality can be predicted in medical patients by blood pressure, age, respiratory rate, loss of independence, and peripheral oxygen saturation (the PARIS score): a prospective cohort study with external validation. *PLoS One.* 2015;**10(4)**:e0122480.
12. Pedersen CB. The Danish Civil Registration System. *Scandinavian journal of public health.* 2011;**39(7 Suppl)**:22-5.
13. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scandinavian journal of public health. 2011;**39(7 Suppl)**:30-3.
14. Pandharipande PP, Shintani AK, Hagerman HE, St Jacques PJ, Rice TW, Sanders NW, *et al.* Derivation and validation of Spo2/Fio2 ratio to impute for Pao2/Fio2 ratio in the respiratory component of the Sequential Organ Failure Assessment score. *Critical care medicine.* 2009;**37(4)**:1317-21.
15. Rajabi Kheirabadi A, Tabeshpour J, Afshari R. Comparison of Three Consciousness Assessment Scales in Poisoned Patients and Recommendation of a New Scale: AVPU Plus. *Asia Pacific Journal of Medical Toxicology*. 2015;**4(2)**:58-63.
16. Frenkel WJ, Jongerius EJ, Mandjes-van Uitert MJ, van Munster BC, de Rooij SE. Validation of the Charlson Comorbidity Index in acutely hospitalized elderly adults: a prospective cohort study. Journal of the American Geriatrics Society. 2014;**62(2)**:342-6.
17. Williams JM, Greenslade JH, Chu K, Brown AF, Lipman J. Severity Scores in Emergency Department Patients With Presumed Infection: A Prospective Validation Study. *Critical care medicine.* 2016;**44(3)**:539-47.
18. Rowley G, Fielding K. Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet.* 1991;**337(8740)**:535-8.

© 2018 Rila Publications Ltd.