

## Risk of mortality assessment in pediatric heart surgery

### Pediyatrik kalp cerrahisinde mortalite riskinin değerlendirilmesi

Özge Köner,<sup>1</sup> Deniz Özsoy,<sup>2</sup> İsmail Haberal,<sup>2</sup> Ali Ekrem Köner,<sup>3</sup> Cenk Eray Yıldız,<sup>2</sup> Gürkan Çetin<sup>2</sup>

<sup>1</sup>Department of Anaesthesiology and Reanimation, Yeditepe University Hospital, İstanbul, Turkey

<sup>2</sup>Department of Cardiovascular Surgery, İstanbul University Institute of Cardiology, İstanbul, Turkey

<sup>3</sup>Department of Cardiovascular Surgery, Medical Faculty of Gaziosmanpaşa University, Tokat, Turkey

**Background:** This study aims to evaluate the validity of Pediatric Index of Mortality (PIM) 1, PIM 2, and modified Sequential Organ Failure Assessment (m-SOFA) scores for predicting mortality in pediatric heart surgery.

**Methods:** Between June 2003 and January 2009, medical files of 456 pediatric patients who were monitored in a 12-bed postoperative cardiac surgery care unit following heart surgery were retrospectively analyzed. A total of 373 files were included in the study. Age, gender, diagnosis, the length of stay in the intensive care unit and hospital, survival rates, PIM 1, PIM 2 scores and m-SOFA scores on admission, at 24 and 48 hours and peak m-SOFA scores were recorded. Student's t test was used to compare the normally distributed data, whereas Mann-Whitney-U test was used to compare non-parametric data. Calibration of the scores was performed using the Hosmer and Lemeshow Goodness of Fit test. Discrimination power of the scores was analyzed using the receiver operating characteristic (ROC) curves.

**Results:** Fifty patients (13.4%) died perioperatively. Peak and m-SOFA scores on admission were significantly higher in non-survivors ( $9.8 \pm 2$  and  $9.2 \pm 2$ , respectively) than survivors ( $5 \pm 2.5$  and  $4.6 \pm 2.5$ , respectively;  $p < 0.01$ ). Calibration with Hosmer-Lemeshow Goodness of Fit test was chi-square  $df(8) = 30.4$ ,  $p = 0.0002$  for PIM 1 and chi-square  $df(9) = 13.5$ ,  $p = 0.13$  for PIM 2. Discrimination power and calibration strength of PIM 2 score was good (ROC 0.82), whereas PIM 1 had a better value (ROC 0.87) of discrimination power with a poor calibration strength. The ROC values of peak and m-SOFA scores on admission were observed to have a good discrimination power (0.93 and 0.92, respectively).

**Conclusion:** Our study results demonstrate that peak and m-SOFA scores on admission are improved for the prediction of mortality in pediatric cardiac surgery, compared to PIM 1 and PIM 2 scores.

**Key words:** Mortality prediction; m-SOFA score; pediatric cardiac surgery; PIM 1 score; PIM 2 score.

**Amaç:** Bu çalışmada Pediyatrik Mortalite İndeksi (PIM) 1, PIM 2 ve modifiye Ardışık Organ Yetersizlik Değerlendirme (m-SOFA) skorlarının pediyatrik kalp cerrahisinde mortalite tahminindeki geçerliliği değerlendirildi.

**Çalışma planı:** 2003 Haziran - 2009 Ocak yılları arasında, 12 yataklı ameliyat sonrası kalp cerrahisi yoğun bakım ünitesinde kalp cerrahisi sonrasında izlenen 456 pediyatrik hastanın tıbbi dosyası retrospektif olarak incelendi. Bu dosyalardan 373'ü çalışmaya dahil edildi. Hastaların yaş, cinsiyet, tanı, yoğun bakımda ve hastanede kalış süreleri, sağkalım oranları, PIM 1, PIM 2 skorları ve başlangıçtaki, 24. ve 48. saatlerdeki m-SOFA skorları ile zirve m-SOFA skorları kaydedildi. Normal dağılım gösteren veriler Student t testi, nonparametrik veriler Mann-Whitney-U testi ile karşılaştırıldı. Skorların kalibrasyonu Hosmer ve Lemeshow Uyum İyiliği testi ile yapıldı. Skorların ayırım gücü, Alıcı İşletim Karakteristikleri (ROC) eğrisi kullanılarak analiz edildi.

**Bulgular:** Ameliyat sırası dönemde 50 hasta (13.4%) kaybedildi. Zirve ve başlangıç m-SOFA skorları, ölen hastalarda (sırasıyla  $9.8 \pm 2$  ve  $9.2 \pm 2$ ), sağ kalanlara (sırasıyla  $5 \pm 2.5$  ve  $4.6 \pm 2.5$ ;  $p < 0.01$ ) kıyasla, anlamlı düzeyde yüksekti. Hosmer-Lemeshow Uyum İyiliği testi ile kalibrasyon PIM 1 için ki-kare  $df(8) = 30.4$ ,  $p = 0.0002$  ve PIM 2 için ki-kare  $df(9) = 13.5$ ,  $p = 0.13$  bulundu. PIM 2'nin ayırım ve kalibrasyon gücü iyi iken (ROC 0.82), PIM 1'in daha iyi bir ayırım (ROC 0.87), ancak zayıf bir kalibrasyon gücü olduğu görüldü. Zirve ve başlangıç m-SOFA skoru ROC değerlerinin iyi bir ayırım gücü olduğu gözlemlendi (sırasıyla 0.93 ve 0.92).

**Sonuç:** Çalışma bulgularımız, pediyatrik kalp cerrahisinde zirve ve başlangıç m-SOFA skorlarının mortaliteyi belirleme gücünün, PIM 1 ve PIM 2 skorlarına kıyasla, gelişmiş olduğunu göstermektedir.

**Anahtar sözcükler:** Mortalite belirleme; m-SOFA skoru; pediyatrik kalp cerrahisi; PIM 1 skoru; PIM 2 skoru.



Mortality risk and outcome prediction are of great importance in the intensive care unit (ICU), and mortality indices are tools that aid in predicting patient outcome, especially in pediatric ICUs. The standard mortality prediction model in pediatric ICUs is the Paediatric Risk of Mortality (PRISM)<sup>[1]</sup> in which the scores are calculated using the most abnormal values of 14 physiological variables obtained within the first 24 hours of ICU stay.<sup>[2]</sup> However, it is difficult to collect the variables for the PRISM, and other problems exist with this model. For example, the score is less accurate than it appears, and the worst values obtained within 24 hours hide the differences between the various centers. Because of these issues with PRISM, another index, the Pediatric Index of Mortality (PIM), based on eight variables, was developed. The score for this model is based on the data used for admission to the ICU. It is simple and has a good predictive power.<sup>[3]</sup> In 2003, a revised version of the PIM score (PIM2) was developed,<sup>[4]</sup> and the categories of admission after cardiopulmonary bypass (CPB) and low risk diagnosis variables were added.

Modified sequential organ failure assessment (m-SOFA) is another assessment instrument that consists of the evaluation of five organ systems. It is designed to predict organ failure, but has also been successfully used to predict mortality after pediatric heart surgery.<sup>[5]</sup>

In this study, we retrospectively collected the data of pediatric patients who underwent heart surgery, calculated their PIM 1, PIM 2, and m-SOFA scores, and evaluated their mortality prediction rates.

**PATIENTS AND METHODS**

This study, which was approved by the ethics committee, was conducted in a 12-bed postoperative cardiac surgery ICU that admits 300 patients yearly who undergo heart surgery, 100-130 of whom are pediatric patients. For this retrospective study, data from those who underwent heart surgery between June 2003 and

January 2009 at our facility was collected, and a total of 456 files were evaluated. Only those files with complete records and laboratory tests were included in the study since that made them eligible to be evaluated by the PIM and SOFA scores. Seventy-seven were excluded for this reason. Furthermore, patients older than 16 years old were also excluded. Hence, 376 patients remained after applying the exclusion criteria; however, three patients died in the operating room, thus leaving 373 (212 boys, 161 girls). Their data were the assembled and recorded. Standard anesthesia and CPB methods were used during all of the operations.

The age, gender, diagnosis, ICU and hospital length of stay, mortality, and cardiopulmonary and aortic cross-clamp times along with the PIM 1 PIM 2, m-SOFA scores on admission and at 24 and 48 hours, and the peak m-SOFA score during the study period were recorded. The m-SOFA score had a maximum value of 20 (Table 1),<sup>[5]</sup> and the PIM 1 and PIM 2 scores were recorded according to the ICU referral data.

Parametric values were expressed as mean ± standard deviation (SD) where appropriate, and all variables were tested for normal distribution using the Kolmogorov-Smirnov test. Student’s t-test was used for comparison of the normally distributed parametric data while the m-SOFA scores, both among the survivors and non-survivors, were compared with the Mann-Whitney U test. The performance of the scoring systems was assessed by applying the standardized mortality ratio (SMR), which was calculated by dividing the observed deaths by the predicted number of deaths in the total group. Calibration of the PIM 1 and PIM 2 scores was performed with the Hosmer and Lemeshow goodness-of-fit test for deciles of mortality risk based on the ranked mortality risks of all patients for each scoring system. A *p* value <0.05 indicated a poor calibration or fit. The discrimination power of the scores was assessed with receiver operating characteristic (ROC) curves in which the sensitivity was plotted against the 1-specificity. The

**Table 1. Modified sequential organ failure assessment score**

	0	1	2	3	4
Respiration PaO <sub>2</sub> /FiO <sub>2</sub>	>400	400-301	300-201	200-101	≤100
Coagulation platelet count x10 <sup>3</sup> /mm <sup>3</sup>	>150	150-101	100-51	50-21	≤20
Liver total bilirubin (mg/dL)	<1.2	1.2-1.9	2-5.9	6-11.9	>12
Cardiovascular hemodynamic support	No agents	Any cardiac agent	Dopamine <5 µg/kg/min or any dose dobutamine	Dopamine 5-15 µg/kg/min or epi/norepi ≤0.1 µg/kg/min	Dopamine >15 µg/kg/min or epi/norepi >0.1 µg/kg/min
Renal creatinine (mg/dL)	<1.2	1.2-1.9	2 – 3.4	3.5-4.9	>5

**Table 2. Distribution and mortality rate according to the operative procedures**

Procedure	Numbers	Mortality (numbers)
Ventricular septal defect repair	111	11
Tetralogy of Fallot repair	57	6
Arterial switch	45	10
Atrial septal defect repair	37	0
Common atrioventricular canal repair	23	2
Double outlet right ventricle repair	17	4
Total anomalous pulmonary venous return repair	16	4
Aortic arch repair	11	2
Bidirectional Glenn shunt	10	0
Corrected transposition of the great arteries repair	9	1
Modified blalock-taussig shunt	9	0
Patent ductus arteriosus ligation	5	0
Truncus arteriosus	7	5
Others	19	5

area under the receiving operating characteristic (AUROC) curve was the overall summary measure of discriminatory performance. For example, an AUROC of 0.5 meant that there was no discriminative ability or that it was equal to random chance, whereas an AUROC of 1.0 indicated a perfect discrimination power. Statistical analysis was performed with the SPSS for Windows version 10.0 software program (SPSS Inc, Chicago, IL, USA), and a value of  $p < 0.05$  was considered to be significant.

## RESULTS

The distribution of the patients according to the operative procedures is shown in Table 2, and the demographic and operative data are shown in Table 3. Fifty of the patients died perioperatively (13.4%), and 15 of these were neonates (30% of all the non-survivors). The ages of these patients ranged from three days to 14 years old. The non-survivors were younger than the survivors (Table 3) as they were all younger than four years of age.

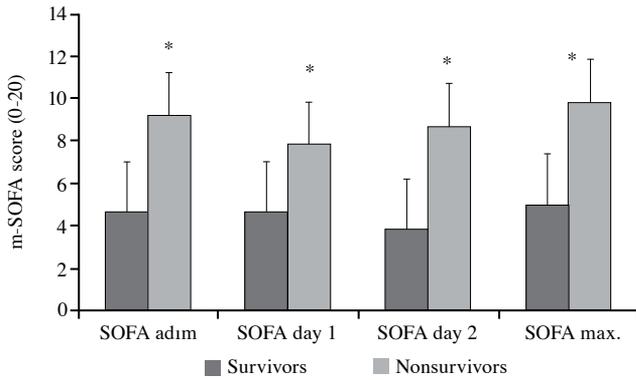
The length of ICU stay along with the aortic cross-clamp and CPB times were longer for the non-survivors (Table 3). Furthermore, the m-SOFA admission scores were higher for the non-survivors ( $9.2 \pm 2$ ) than the survivors ( $4.6 \pm 2.5$ ) ( $p < 0.01$ ), and the m-SOFA scores on day one and day two were higher for the non-survivors ( $7.8 \pm 2$  and  $8.7 \pm 2.5$ , respectively) than for the survivors ( $4.6 \pm 2.5$  and  $3.8 \pm 2.7$ ) ( $p < 0.01$  for both). In addition, the peak m-SOFA score was significantly higher for the non-survivors ( $9.8 \pm 2$ ) when compared against the survivors ( $5 \pm 2.5$ ) ( $p < 0.001$ ) (Figure 1). All the non-survivors had both peak m-SOFA and m-SOFA admission scores of  $\geq 6$  (range 6-14).

The m-SOFA admission score had a sensitivity of 97% and a specificity of 80%, which made for a cut-off value of 6.5, and the peak m-SOFA values revealed a similar sensitivity and specificity (96% and 74%, respectively) with the same cut-off value. The m-SOFA on day one had a sensitivity of 88% and specificity of 75%, yielding a cut-off value of 5.5. On

**Table 3. Demographic and operative data**

	All (n=373)		Survivors (n=323)		Non-survivors (n=50)		p
	n	Mean $\pm$ SD	n	Mean $\pm$ SD	n	Mean $\pm$ SD	
Gender							
Male	212		184		28		NS
Female	161		139		22		NS
Age (months)		32 $\pm$ 38		36 $\pm$ 39		11 $\pm$ 20	<0.001
Aortic cross clamp time (min)		68 $\pm$ 47		63 $\pm$ 43		100 $\pm$ 62	<0.001
Cardiopulmonary bypass time (min)		108 $\pm$ 75		100 $\pm$ 66		162 $\pm$ 104	<0.001
ICU LOS (hours)		134 $\pm$ 291		122 $\pm$ 278		223 $\pm$ 361	0.003
Hospital LOS (days)		16 $\pm$ 14		16 $\pm$ 14		14 $\pm$ 15	NS

SD: Standard deviation; ICU: Intensive care unit; LOS: Length of stay; NS: Not significant.



**Figure 1.** Comparison of sequential m-SOFA scores among the survivors and non-survivors. m-SOFA: Modified Sequential Organ Failure Assessment.

day two, the sensitivity was 86% and the specificity was 86%, making a cut-off value of 6.5. Of all of the non-survivors, the peak m-SOFA and m-SOFA admission scores of only two patients were below the cut-off value of 6.5. These results revealed that the peak m-SOFA score along with m-SOFA admission scores of 6.5 have reliable sensitivity and specificity for mortality.

The overall performance of the PIM 1 and PIM 2 scores was evaluated via SMR, which was calculated by comparing the expected and observed deaths in the whole group (Table 4). The PIM 1 score had a sensitivity of 83% and a specificity of 78% for a cut-off value of 2.85, and the PIM 2 score had a similar sensitivity and specificity (83% and 76%, respectively) for a cut-off value of 2.45. Of all the non-survivors, the PIM 1 scores of 10 patients were below the cut-off value of 2.85 and the PIM 2 scores of seven patients were below the cut-off value of 2.45. Therefore, the discriminative power of admission and the peak m-SOFA score of 6.5 is higher than both PIM scores.

Calibration with the Hosmer-Lemeshow goodness-of-fit test for PIM 1 revealed a chi-square value of degrees of freedom (df) (8)=30.4 and a 95% confidence interval (CI) of 1.2 (range 1.12-1.27) (p=0.0002). The results of the PIM 2 showed a chi-square value

of df (9)=13.5 and a 95% CI of 1.38 (range 1.2-1.5) (p=0.13). The PIM 2 mortality prediction model proved to be a better option compared to PIM 1 for predicting mortality in pediatric patients undergoing heart surgery as the Hosmer-Lemeshow goodness-of-fit test had a p value of >0.05 (p=0.13). However, both scores underestimated the mortality risk in this study (Table 5).

The discriminatory performance of the scores assessed with the ROC curves showed that the PIM 2 had a fair discrimination power (ROC score 0.82), while the PIM1 ROC (0.87) was better. However, there was poor calibration (p=0.0002). The m-SOFA scores have good discrimination power (Table 6). However, the m-SOFA admission and m-SOFA on day two had the greatest power of all (Table 6). The m-SOFA score is shown in Table 1.

**DISCUSSION**

In this study, we found that the peak m-SOFA and m-SOFA admission scores performed better than the PIM 1 and PIM 2 scores for predicting mortality after pediatric cardiac surgery. Additionally, peak m-SOFA and m-SOFA admission scores of 6.5 have a reliable sensitivity and specificity for predicting mortality.

A preliminary study evaluating the mortality predicting power of PIM over PRISM in pediatric cardiac surgery demonstrated that PIM is better than PRISM when applied to infants and children.<sup>[6]</sup> Furthermore, the authors declared that collecting data for the PIM is much easier than for the PRISM. However, in a recent study, the performance of the PIM 2 was found to be poor regarding the calibration and predictive ability in pediatric cardiac surgery.<sup>[7]</sup> The difference between these studies regarding the PIM scores might be attributable to differences among the centers and poor inter-rater reliability.<sup>[8]</sup> The Czaja et al.<sup>[7]</sup> study, comprised of 8,391 pediatric patients who underwent cardiac surgery, also had an AUROC of 0.80. The performance of the PIM 2 score in our study had a similar AUROC of 0.82.

**Table 4. Observed and expected mortality as standard mortality rate (n=373)**

	n	%	SMR (95% CI)
Observed mortality	50	13.4	
Pediatric index of mortality 1	42	11.3	1.19 (0.88-1.5)
Expected mortality			
Pediatric index of mortality 2	35.8	9.6	1.39 (1.01-1.77)
Expected mortality			

CI: Confidence interval.

**Table 5. Observed and expected mortality for pediatric index of mortality 1 and pediatric index of mortality 2 scores in pediatric patients undergoing heart surgery**

	Pm<1	1<Pm≤5	5<Pm≤15	15<Pm≤30	Pm>30	Total	
	very low	low	moderate	high	very high	n	%
<b>PIM 1</b>							
Number	26	305	26	10	6	373	–
Mean risk	0.76	2.1	8.5	19	44	–	–
Observed death	0	7.8	61.5	50	83	–	13.4
Expected death	0.78	2.2	9	19	46	–	11.3
<b>PIM 2</b>							
Number	16	321	24	8	4	373	–
Mean risk	0.7	2.1	7.7	18.8	42.5	–	–
Observed death	0	8.72	50	75	100	–	13.4
Expected death	1	2.3	7.5	12.7	35.4	–	9.6

Pm: Probability of mortality as %; PIM: Pediatric index of mortality.

In a study performed on 75 pediatric patients who underwent cardiac surgery, Barlas et al.<sup>[9]</sup> determined that the PRISM score had a poor mortality prediction power when compared with the modified Acute Physiology and Chronic Health Evaluation (APACHE) III score.

The SOFA is an organ failure assessment score. However, it is also valuable for predicting ICU mortality for both adult and pediatric ICU patients.<sup>[10,11]</sup> This score can also be used as an independent predictor of mortality in adult patients undergoing heart surgery, as shown in the study by Pätälä et al.<sup>[12]</sup> In that same study, the peak SOFA scores were measured during the first three days. A SOFA score of over 20 points in pediatric patients undergoing heart surgery was found to be reliable for predicting death within the first 36 hours.<sup>[5]</sup> The Shime et al.<sup>[5]</sup> study, with a smaller number of participants (n=142) than our study, had a very high neonatal mortality rate (7 out of 8). Furthermore, the authors did not report the AUROC or the sensitivity and specificity of the SOFA scores. They didn't perform a neurological evaluation based on the SOFA score, which we chose not to include, either. In our study, the AUROC plot values for admission regarding

the m-SOFA scores and m-SOFA scores at 48 hours were higher than all of the other m-SOFA scores. The m-SOFA admission and peak m-SOFA scores of 6.5 showed good discrimination power. In a systematic review of SOFA-based models for predicting mortality in the ICU, Minne et al.<sup>[13]</sup> concluded that SOFA admission scores were competitive with severity of illness scores limited to the first 24 hours of admission. They advocated for the use of a combination of sequential SOFA scores in conjunction with traditional models (e.g. APACHE). We agree that studies which combine SOFA with other severity scores, such as PIM, could also be valuable for predicting mortality after pediatric cardiac surgery.

We acknowledge that our study had several limitations. A retrospective analysis cannot provide as strong evidence in favor of the predictive power of a mortality scoring system as a prospective study would. Moreover, although statistically significant results have emerged in our limited patient set, multi-center studies with larger patient numbers are necessary for enhancing statistical strength in testing the universality and reliability of m-SOFA score as a mortality predictor.

**Table 6. Discriminatory performance of the scores assessed by receiver operating curve curves**

	AUROC	SE	Significance	95% CI	
				Lower	Upper bound
Pediatric index of mortality 1	0.87	0.03	0.000	0.81	0.93
Pediatric index of mortality 2	0.82	0.03	0.000	0.74	0.89
Sequential organ failure assessment admission	0.93	0.01	0.000	0.90	0.96
Sequential organ failure assessment day 1	0.84	0.02	0.000	0.79	0.89
Sequential organ failure assessment day 2	0.91	0.02	0.000	0.86	0.95
Sequential organ failure assessment peak	0.92	0.015	0.000	0.89	0.95

AUROC: Area under the receiver operating curve; SE: Standard error; CI: Confidence interval.

### Conclusion

The m-SOFA peak and admission scores performed better for pediatric patients who underwent heart surgery and for the prediction of mortality than the PIM 1 and 2 scores in our retrospective study. Larger prospective studies are needed to investigate the value of these scores as a better means of predicting surgical outcomes in pediatric heart patients.

### Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

### Funding

The authors received no financial support for the research and/or authorship of this article.

### REFERENCES

1. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med* 1988;16:1110-6.
2. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. *Crit Care Med* 1996;24:743-52.
3. Shann F, Pearson G, Slater A, Wilkinson K. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Med* 1997;23:201-7.
4. Slater A, Shann F, Pearson G; Paediatric Index of Mortality (PIM) Study Group. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive Care Med* 2003;29:278-85.
5. Shime N, Kageyama K, Ashida H, Tanaka Y. Application of modified sequential organ failure assessment score in children after cardiac surgery. *J Cardiothorac Vasc Anesth* 2001;15:463-8.
6. Jones GD, Thorburn K, Tigg A, Murdoch IA. Preliminary data: PIM vs PRISM in infants and children post cardiac surgery in a UK PICU. *Intensive Care Med* 2000;26:145.
7. Czaja AS, Scanlon MC, Kuhn EM, Jeffries HE. Performance of the Pediatric Index of Mortality 2 for pediatric cardiac surgery patients. *Pediatr Crit Care Med* 2011;12:184-9. doi: 10.1097/PCC.0b013e3181e89694.
8. van Keulen JG, Polderman KH, Gemke RJ. Reliability of PRISM and PIM scores in paediatric intensive care. *Arch Dis Child* 2005;90:211-4.
9. Barlas S, Tireli E, Tekinalp H, Karşlı M, Dağlı S, Dayıoğlu E ve ark. Apache III sisteminin kalp ameliyatı sonrasında kardiyopulmoner yetersizlikli çocuklardaki değeri. *Türk Gogus Kalp Dama* 1994;4:304-8.
10. Vincent JL, de Mendonça 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. *Crit Care Med* 1998;26:1793-800.
11. Vincent JL, Ferreira F, Moreno R. Scoring systems for assessing organ dysfunction and survival. *Crit Care Clin* 2000;16:353-66.
12. Pätälä T, Kukkonen S, Vento A, Pettilä V, Suojaranta-Ylinen R. Relation of the Sequential Organ Failure Assessment score to morbidity and mortality after cardiac surgery. *Ann Thorac Surg* 2006;82:2072-8.
13. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care* 2008;12:R161. doi: 10.1186/cc7160.