Original Article

Comparison of Infection Probability Score, APACHE II, and APACHE III Scoring Systems in Predicting Need for Ventilator and Ventilation Duration in Critically Ill Patients

Mohammadreza Safavi MD•*, Azim Honarmand MD*

Background: This study examines the efficacy of the predicting power for need for mechanical ventilation and duration of mechanical ventilation of three different scoring systems in a medical-surgical intensive care unit.

Methods: One-hundred eighty critically ill patients were included prospectively in our study in a consecutive period of seven months. On the day of admission, data were collected from patients to compute the Acute Physiology and Chronic Health Evaluation (APACHE) II and III, and Infection Probability Score (IPS). The sensitivity, specificity, and overall correctness of prediction were calculated, and the cut-off point giving the best likelihood ratio was determined. The area under receiver operating characteristic curve was computed.

Results: For prediction of need for mechanical ventilation the best cut-off points were 52 for APACHE III, 12 for APACHE II, and 12 for IPS. The area under the curve was 0.89 in APACHE III, 0.74 in APACHE II and 0.82 in IPS. There were statistical differences between APACHE III, APACHE II and IPS in terms of likelihood ratio and the area under the curve ($P < 0.05$). None of the three scoring systems provide good discrimination in prediction of more than 5 days respiratory support under mechanical ventilation.

Conclusion: For prediction of need for mechanical ventilation, the APACHE III has better accuracy than APACHE II or IPS.

Archives of Iranian Medicine, Volume 10, Number 3, 2007: 354 – 360.

Keywords: APACHE • critical care • mechanical ventilation

Introduction

Mechanical ventilation is associated with numerous life-threatening complications.[1] The major factor in successful management of patients undergoing mechanical ventilation is the resolution of precipitating illness and a stable low requirement for oxygen.[2] The recognition of risk factors that can stratify the critically ill patients under mechanical ventilation into subgroups with different outcomes is of great prognostic value for the clinician.[3] The Infection Probability Score (IPS) is a simple scoring system that helps assess the probability of infection in critically ill patients. IPS uses six simple and commonly used variables consisting of temperature, heart rate, respiratory rate, white blood cell count, C-reactive protein, and Sequential Organ Failure Assessment score.[4] The revised Acute Physiology and Chronic Health Evaluation system (APACHE II) has been frequently applied in many intensive care units (ICU) throughout the world. Twelve physiological variables including the respiratory rate (non-ventilated or ventilated) and oxygenation are used in the APACHE II system. Various results confirm the contribution of the APACHE II in ICU patients.[5, 6] In the previous studies, on admission scoring based on APACHE II system proved to be of the highest predictive value for the duration of mechanical ventilation and weaning success.[7] The APACHE III will include improved incorporation

Authors' affiliation: *Department of Anesthesia and Intensive Care, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.

•Corresponding author and reprints: Mohammadreza Safavi MD, Department of Anesthesia and Intensive Care, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. P.O. Box: 8174675731/81744.

Tel: +98-311-668-4510, E-mail: safavi@med.mui.ac.ir.

Accepted for publication: 15 November 2006

of physiological measures, chronic health measures, and disease classification.[8, 9] The APACHE II, APACHE III, and IPS scoring systems include some major respiratory-related modifications, such as taking into account the alveolar-arterial O2 difference (AaDO2), partial pressure of oxygen in arterial blood (Pao2), fraction of inspired oxygen (FIO2) as well as the acid-base status of patients. These scoring systems may thus be useful as a prognostic and comparative tool for patients' mechanical ventilation need. Although the predictive criteria for duration of mechanical ventilation may help to evaluate the suitability of disconnecting a patient from a ventilator, the role of the APACHE or IPS system in prediction of ventilator need or its duration needs further evaluation. This was our major reason for designing this study.

Materials and Methods

This prospective study was conducted at the medical-surgical intensive care unit (ICU) at Alzahra hospital affiliated to Esfahan University of Medical Sciences. The study sample consisted of 180 eligible critically ill patients. Exclusion criteria were; burn injuries, age less than 16 years, death within the first hour after admission, admission after cardiac surgery or for exclusion of myocardial infarction, patients who were resuscitated in the emergency room before admission and patients who were scheduled for organ donation. We evaluated the patients by using the APACHE II, APACHE III, and IPS systems on the day of admission to our ICU and recorded the findings for each patient. All data were collected by a single physician.

IPS uses six simple and commonly used variables and ranges from 0 to 26 points (0 – 2 for temperature, 0 – 12 for heart rate, 0 – 1 for respiratory rate, 0 – 3 for white blood cell count, 0 – 6 for C-reactive protein, 0-2 for Sequential Organ Failure Assessment score).[4] The value of 14 has been chosen as the cut-off.[10]

The APACHE III scores were gathered using the method presented by Knaus in 1991 with 18 variables in acute physiological scores ranging from 0 to 252, the age score from 0 to 24 and the chronic health evaluation from 0 to 23.[11] The total scores in the APACHE III, summarized by the three mentioned categories, ranged from 0 to 299.

The total score in the APACHE II is 71, which includes the sum of physiological score, age score, and chronic health evaluation.[12] The calculation of APACHE II, APACHE III, and IPS scores was based on the worst values taken during the first 24 hours after admission. Duration of ventilation was defined as number of days mechanical ventilation was used for the patient; no attempt was made to subdivide into hours. The procedure of weaning from mechanical ventilation started by 5 minutes spontaneous breathing trough a T-tube circuit, with the FiO2 set at the level used during mechanical ventilation. During a 2-hour trial, the patient had to meet the following objective criteria: spontaneous respiratory frequency <35/min, arterial blood oxygen saturation (SaO2) >90% at FiO2 <0.4, heart rate <140/min or <20% change from the baseline, systolic blood pressure <200 mm Hg or >80 mm Hg, PaO2 >60 mm Hg, pH <7.30, and stable clinical condition. The patients who fulfilled these criteria at the end of the 2-hour trial were extubated. The weaning procedure was considered successful if reintubation was not required within the 48 hours. The patients with weaning difficulties during a 2-hour spontaneous breathing trial were followed until ICU discharge or death.

Data analysis

Chi-square test was used to assess the differences of need for mechanical ventilation within the three scoring systems. Descriptive statistics were expressed as mean ± SD unless otherwise stated. Discrimination was tested using the receiver operating characteristic (ROC) curves and by comparing areas under the curve (AUC).[13] The ROC curve was calculated by the Systat software program (Systat Inc., Evanstan, IL.). AUCs between 0.7 and 0.8 were classified as "acceptable" and between 0.8 and 0.9 as "excellent" discrimination.[14] For the different scoring systems tested, the sensitivity, specificity, and overall correctness of prediction were calculated, and the cut-off point giving the best likelihood ratio was determined.[15] The higher the likelihood ratio the more accurate is the prediction for a positive test result at the cut-off point. This cut-off point was also used to calculate the predicted and observed outcome for patients. A P value less than 0.05 was considered statistically significant for all the above analyses. Calibration was assessed using the Hosmer-Lemeshow goodness of fit statistic, which divides subjects into deciles based on predicted probabilities of need for

Table 1. Clinical diagnosis of the patients.

Surgical	
Gastrointestinal surgery	58 (32.2)
Thoracic surgery	26 (13.4)
Trauma	28 (15.5)
Others	4 (2.2)
Total	**116 (64.4)**
Medical	
Cerebrovascular accident	5 (2.8)
Drug intoxication	6 (3.3)
Pulmonary thromboembolism	13 (7.2)
Sepsis	12 (6.6)
Respiratory failure	16 (8.9)
Others	12 (6.6)
Total	**64 (35.5)**

mechanical ventilation and then computes a Chi-square from observed and expected frequencies.[16] Values more than 15.5 represent poor agreement of calibration between the outcomes estimated from the model and the observed outcomes. A low value represents good agreement.[16] A good fit was defined as $P > 0.05$.

Results

One-hundred eighty critically ill adult patients were admitted into our ICU in a 7-month period. One-hundred sixteen (64.4%) of the patients had surgical problems (Table 1). Their age range was from 16 to 84 years with a mean of 39.4 years (Table 2). There were 103 males and 77 females. One-hundred fifty two (84.4%) patients were intubated in ICU; 146 (81.1%) patients needed mechanical ventilation. The distribution of scores on day 1 and probability of need for mechanical ventilation and duration of mechanical ventilation derived from each scoring system are shown in Figures 1 and 2. The sensitivity, specificity, correct prediction outcome, Youden index, and the ROC area at the best cut-off point for mechanical ventilation need are presented in Table 3. There were statistical differences in likelihood ratio and area under the ROC curve between APACHE III, which provided good results for mechanical ventilation need, and APACHE II and IPS (Table 3). By contrast, in other physiological variables, APACHE II and IPS, yielded poor results. Therefore, only APACHE III plays a crucial role in the prediction of need for mechanical ventilation. None of the three scoring systems provide good discrimination in prediction of more than 5 days respiratory support under mechanical ventilation (AUC < 0.5). The calibration of the APACHE III system is adequate, as shown in Figure 3.

Discussion

Comparing the APACHE III with the APACHE II and IPS, we found that the accuracy of the APACHE III was significantly better than that of the APACHE II and IPS for prediction of need for mechanical ventilation. The APACHE III provides more information on determining factors, such as age, underlying diseases, special respiratory parameter, and acute physiological condition than APACHE II or IPS. These informations are crucial

Table 2. Demographic characteristics of the 180 patients.

Variable	Cases	%	Mean ± SD	Range	*P* value
Sex					
Male (intubated)	103 (91)	57.2 (88.3)			0.102[a]
Female (intubated)	77 (61)	42.8 (79.2)			
Male (under MV)	103 (87)	57.2 (84.5)			0.248[a]
Female (under MV)	77 (59)	42.8 (76.6)			
Age					
Total	—	—	39.4 ± 20.0	16-84	0.436*
Need MV	—	—	40.0 ± 18.7	16-84	
No MV	—	—	37.2 ± 18.4	16-80	
APACHE II					
Total	—	—	15.5 ± 4.1	8-30	0.000*
Need MV	—	—	16.1 ± 4.2	8-30	
No MV	—	—	13.0 ± 2.7	9-20	
APACHE III					
Total	—	—	64.5 ± 20.7	27-142	
Need MV	—	—	68.6 ± 20.8	27-142	0.000*
No MV	—	—	47.2 ± 7.34	27-142	
IPS					
Total	—	—	16.3 ± 3.7	7-26	
Need MV	—	—	17.0 ± 3.4	7-26	0.000*
No MV	—	—	12.9 ± 2.9	7-26	

MV = mechanical ventilation; [a] chi-square test; * *t*-test

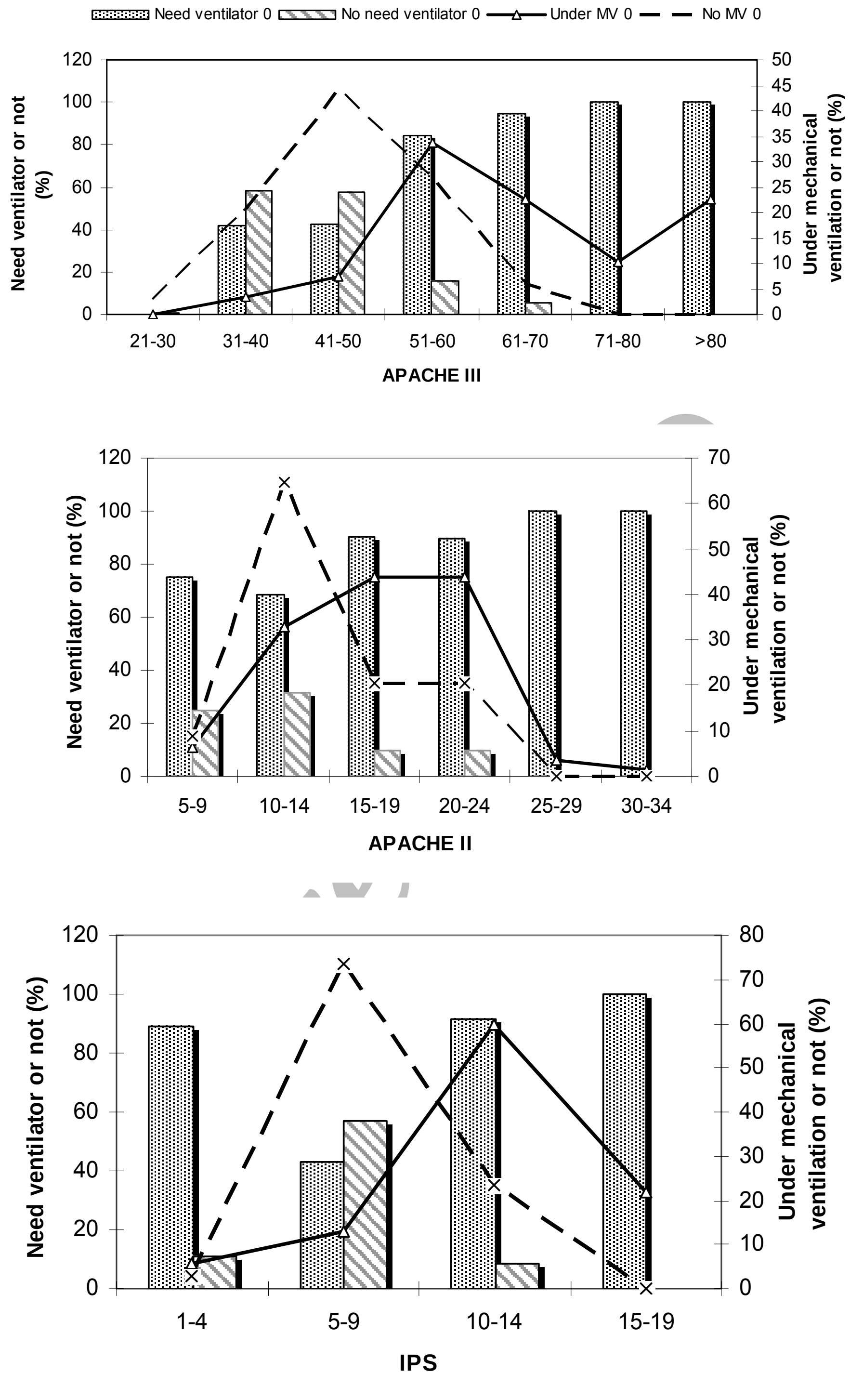

Figure 1. Distribution of the IPS and APACHE II and III scores in the study patients and ventilator need in each category. The higher the scores in APACHE III or II, the higher the ventilator need. The higher the scores in APACHE III, II or IPS, the higher the number of patients under mechanical ventilation. MV = mechanical ventilation.

for prediction of need for mechanical ventilation.

The APACHE III is not much better than the APACHE II and IPS in the prediction of duration of mechanical ventilation because many biases are found in the use of the APACHE system. Firstly, the treatment error is not predictable, especially in surgical patients.[17] Secondly, the data collected on the day of admission may not reflect completely the unforeseen events, which may be major determinants of outcome.[18] And lastly, the co-

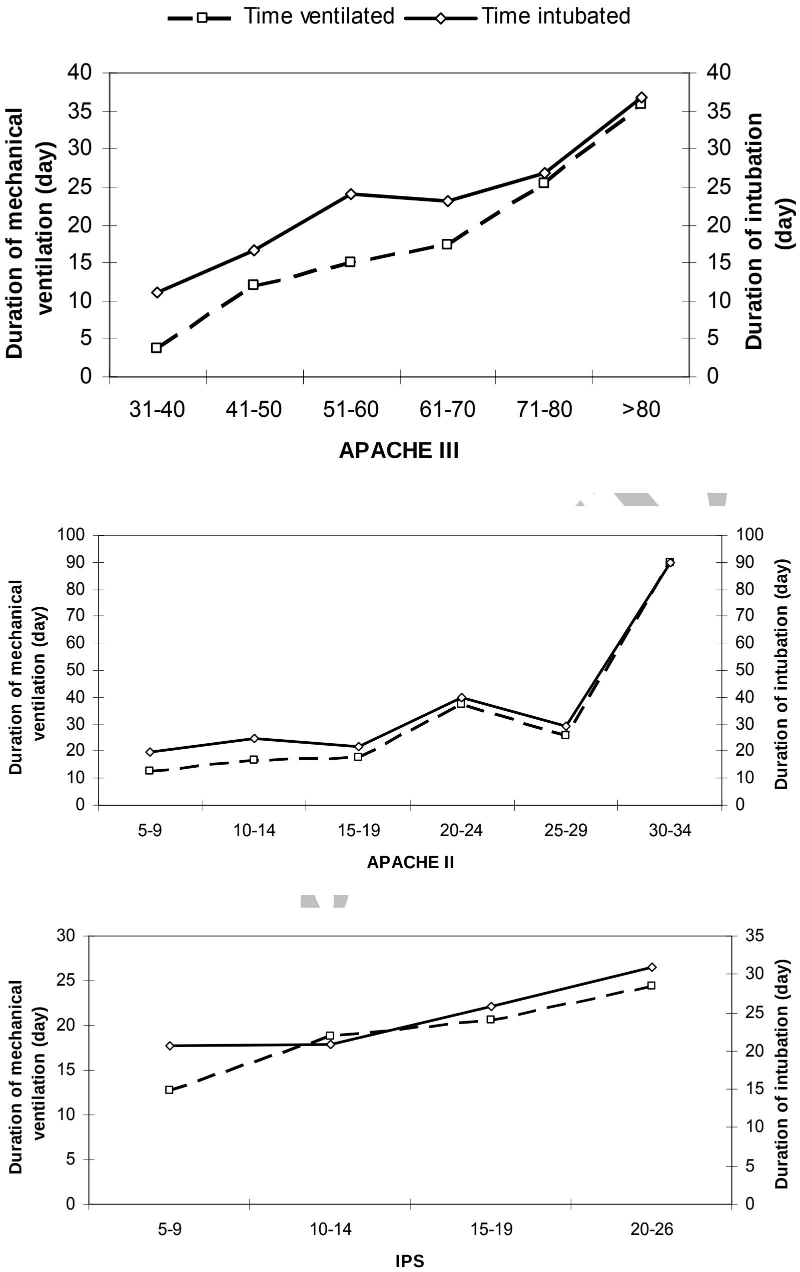

Figure 2. Distribution of the APACHE II and III and IPS scores in the study patients and duration of mechanical ventilation or intubation in each category. The higher the scores in APACHE III and II, and IPS the higher the duration of mechanical ventilation or intubation.

morbidity condition is not taken into account enough in the APACHE system.[11] Although APACHE III yielded a sensitivity of up to 90.4 and 87.7% for prediction of need for mechanical ventilation, this still does not justify its application to individual patients for prediction of outcome in order to change our decision making. However, decision making, in terms of transferring patients from ICU, the reinforcement of medical treatment or surgical intervention, may be changed if we make records of the scoring system for individual patients. The commonly used scoring systems in our study are sufficiently accurate to estimate the probability of need for mechanical ventilation in critically ill patients. However, the sensitivity and calibration of the scoring systems should be improved.[19] The cut-off point is data dependent. The ROC area provides an adequate judgment with no fixed and arbitrary cut-off points in each scoring system for comparison of discrimination.[19]

Table 3. Comparison of the assessment scores in need for mechanical ventilation.

	Cut-off point	Sensitivity (%)	Specificity (%)	Correct prediction outcome (%)	likelihood ratio	ROC area
APACHE III	52	87.7	77.5	86.7	4.0*	0.89 ± 0.03#
APACHE II	12	90.4	32.3	81.1	1.3	0.74 ± 0.04
IPS	12	93.5	42.6	85.6	1.6	0.82 ± 0.04

*APACHE III vs. APACHE II; IPS = *P* < 0.05; # APACHE III vs. APACHE II; IPS = *P* < 0.05

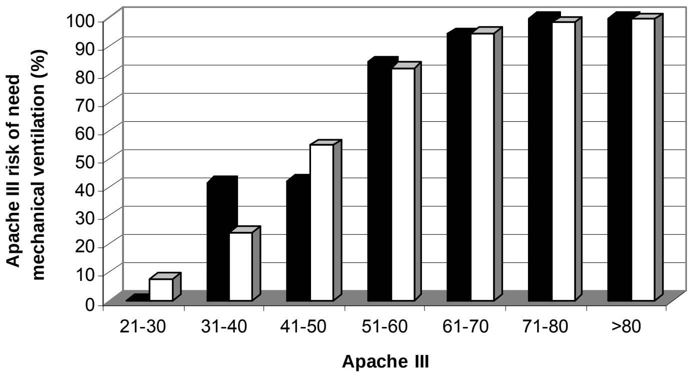

Figure 3. The calibration of APACHE III system on our patients for the observed and predictive risk of need for mechanical ventilation; □ predicted and ■ observed need for mechanical ventilation, Hosmer-Lemeshow statistic = 15.1 with 8 d.f. (*P* = 0.057).

Potential limitation of our study should also be mentioned. Our study was performed in an academic referral hospital; therefore our results may not be applicable to institutions with different patient populations.

In conclusion, for the prediction of need for mechanical ventilation, APACHE III is better than IPS or APACHE II. For prediction of more than 5 days mechanical ventilation, there is no statistical difference between the three scoring systems.

References

1 Tobin MJ. Mechanical ventilation. *N Engl J Med*. 1994; **330:** 1056 – 1061.

2 Smith RA. Principles of mechanical ventilation. In: Ayres SM, Grenvik A, Holbrook PR, Shoemaker WC, eds. *Textbook of Critical Care*. Philadelphia (PA): WB Saunders; 1995: 858 – 867.

3 Chatzicostas C, Roussomoustakaki M, Notas G, Vlachonikolis IG, Samonakis D, Romanos J, et al. A comparison of Child-Pugh, APACHE II and APACHE III scoring systems in predicting hospital mortality of patients with liver cirrhosis. *BMC Gastroenterol*. 2003; **3:** 7.

4 Tillett WS, Francis T Jr. Serological reactions in pneumonia with a nonprotein somatic fraction of pneumococcus. *J Exp Med*. 1930; **52:** 561 – 571.

5 Ebell MH, Preston PS. The effect of the APACHE II score and selected clinical variables on survival following cardiopulmonary resuscitation. *Fam Med*. 1993; **25:** 191 – 196.

6 Rutledge R, Fakhry S, Rutherford E, Muakkassa F, Meyer A. Comparison of APACHE II, trauma score, and injury severity score as predictors of outcome in critically injured trauma patients. *Am J Surg*. 1993; **166:** 224 – 227.

7 Matiæ I, Majeriæ-Kogler V. Comparison of pressure support and T-tube weaning from mechanical ventilation: randomized prospective study. *Croat Med J*. 2004; **45:** 162 – 166.

8 Wagner D, Draper E, Knaus W. APACHE III study design: development of APACHE III. *Crit Care Med*. 1989; **17:** S199 – S203.

9 Chen YC, Chen CY, Hsu HH, Yang CW, Fang JT. APACHE III scoring system in critically ill patients with acute renal failure requiring dialysis. *Dial Transplant*. 2002; **31:** 222 – 233.

10 Ugarte H, Silva E, Mercan D, De Mendonca A, Vincent JL. Procalcitonin as a marker of infection in the intensive care unit. *Crit Care Med*. 1999; **27:** 498 – 504.

11 Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991; **100:** 1619 – 1636.

12 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985; **13:** 818 – 829.

13 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; **143:** 29 – 36.

14 Oh TE, Hutchinson R, Short S, Buckley T, Lin E, Leung D. Verification of the Acute Physiology and Chronic Health Evaluation scoring system in a Hong Kong intensive care unit. *Crit Care Med*. 1993; **21:** 689 – 705.

15 Biggerstaff BJ. Comparing diagnostic tests: a simple graphic using likelihood ratios. *Stat Med*. 2000; **19:** 649 – 463.

16 Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*. 1982; **115:** 92 – 106.

17 Meyer AA, Messick WJ, Young P, Baker CC, Fakhry S, Muakkassa F, et al. Prospective comparison of clinical judgment and APACHE II score in predicting the outcome in critically ill surgical patients. *J Trauma*. 1992; **32:** 747 – 754.

18 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:** 975 – 977.

19 Suter P, Armaganidis A, Beaufils F. Consensus Conference organized by the ESICM and the SRLF. Predicting outcome in ICU patients. *Intens Care Med*. 1994; **20:** 390 – 397.