

External Validation of the Acute Physiology and Chronic Health Evaluation II in Korean Intensive Care Units

Jae Yeol Kim,^{1*} So Yeon Lim,^{2*} Kyeongman Jeon,² Younsuck Koh,³ Chae-Man Lim,³ Shin Ok Koh,⁴
Sungwon Na,⁴ Kyoung Min Lee,⁵ Byung Ho Lee,⁶ Jae-Young Kwon,⁷ Kook Hyun Lee,⁸
Seok-Hwa Yoon,⁹ Jisook Park,¹⁰ and Gee Young Suh²

¹Division of Pulmonary and Critical Care Medicine, Department of Medicine, Chung-Ang University College of Medicine, Seoul;

²Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul;

³Division of Pulmonary and Critical Care Medicine, Department of Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul;

⁴Department of Anesthesiology and Pain Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul; ⁵Anesthesiology and Critical Care Medicine, Konkuk University Hospital, Seoul; ⁶Department of Anesthesiology and Pain Medicine, The Catholic University of Korea, St. Paul's Hospital, Seoul;

⁷Department of Anesthesiology and Pain Medicine, Pusan National University Medical School, Busan; ⁸Department of Anesthesiology and Pain Medicine, Seoul National University Hospital, Seoul; ⁹Department of Anesthesiology and Pain Medicine, Chungnam National University Hospital, Gwangju;

¹⁰School of Media, Seoul Women's University, Seoul, Korea.

Received: December 13, 2011

Revised: March 30, 2012

Accepted: April 8, 2012

Corresponding author: Dr. Gee Young Suh,
Division of Pulmonary and Critical Care
Medicine, Department of Medicine,
Samsung Medical Center, Sungkyunkwan
University School of Medicine,
81 Irwon-ro, Gangnam-gu,
Seoul 135-710, Korea.
Tel: 82-2-3410-3426, Fax: 82-2-3410-6956
E-mail: suhgy@skku.edu

*Jae Yeol Kim and So Yeon Lim contributed
equally to this work.

The authors have no financial conflicts of
interest.

© Copyright:

Yonsei University College of Medicine 2013

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose: This study was designed to validate the usefulness of the Acute Physiology and Chronic Health Evaluation (APACHE) II for predicting hospital mortality of critically ill Korean patients. **Materials and Methods:** We analyzed data on 826 patients who had been admitted to nine intensive care units and were included in the Fever and Antipyretics in Critical Illness Evaluation study cohort. **Results:** Among the patients enrolled, 62% (512/826) were medical and 38% (314/826) were surgical patients. The median APACHE II score was 17 (11 to 23 interquartile range), and the hospital mortality rate was 19.5%. Age, underlying diseases, medical patients, mechanical ventilation, and renal replacement therapy were independently associated with hospital mortality. The calibration of APACHE II was poor ($H=57.54$, $p<0.0001$; $C=55.99$, $p<0.0001$), and the discrimination was modest [area under the receiver operating characteristic (aROC)=0.729]. Calibration was poor for both medical and surgical patients ($H=63.56$, $p<0.0001$; $C=73.83$, $p<0.0001$, and $H=33.92$, $p<0.0001$; $C=33.34$, $p=0.0001$, respectively), while discrimination was poor for medical patients (aROC=0.651) and modest for surgical patients (aROC=0.704). At the predicted risk of 50%, APACHE II had a sensitivity of 36.6% and a specificity of 87.4% for hospital mortality. **Conclusion:** For Koreans, the APACHE II exhibits poor calibration and modest discrimination for hospital mortality. Therefore, a new model is needed to accurately predict mortality in critically ill Korean patients.

Key Words: APACHE II, calibration, discrimination, intensive care units, illness severity

INTRODUCTION

Over the last three decades, severity-of-illness scoring systems, such as the Acute

Physiology and Chronic Health Evaluation (APACHE) Score and the Simplified Acute Physiology Score, have been developed to predict hospital mortality of intensive care unit (ICU) patients.¹⁻⁵ As with other types of prediction scores, these scores must be externally validated when there is a reason to believe that the intended population may constitute a patient mix different from the derivation cohort. In addition, to accurately predict mortality over prolonged periods, such scores should periodically be updated to reflect changes in medical practice and patient mix over time.⁶

Although APACHE II is one of the most widely used prediction models in Korea, the performance of the APACHE II model has not yet been validated in a Korean ICU sample, and has not been updated for Korean patients since its development in 1985.¹ Therefore, the aim of this study was to perform an external validation of the APACHE II by evaluating its performance in a large representative database of Korean ICU patients.

MATERIALS AND METHODS

Patients

This study is an analysis of prospectively collected database for The Fever and Antipyretics in Critical Illness Evaluation (FACE) cohort.⁷ The FACE study was a prospective observational study of the influence of fever and antipyretic treatment on the outcomes of ICU patients in 25 hospitals-15 in Japan and 10 in Korea-that was conducted by the Japan-Korea Intensive Care Study group between September 1 and November 30, 2009. The current study used only data from the Korean centers. Inclusion criteria for the FACE study were all patients who were admitted to the participating ICUs for more than 24 hrs. Exclusion criteria for the FACE study were patients less than 18 years of age and with brain damage. For this study, data from one center that recruited only two patients were excluded to reduce selection bias. The study was approved by the Institutional Review Board, and the requirement for informed consent was waived.

Data collection

The following variables were recorded: 1) general characteristics including age, gender, pre-existing underlying disease, reason for ICU admission, and admission status (medical or surgical patients); 2) severity of illness as assessed

by the APACHE II score; and 3) interventions including need for mechanical ventilation and continuous renal replacement therapy (CRRT). The worst data from the first 24 hours after admission to the ICU was used to calculate the APACHE II score. Hospital mortality was adjusted for the principal diagnostic category which led to ICU admission. The equation used to predict the risk (R) of hospital death was: $\ln(R/1-R) = -3.517 + (\text{APACHE II score} \times 0.146) + (0.603, \text{ if only post-emergency surgery}) + (\text{diagnostic category weight})$.¹

Statistical analysis

For statistical analysis, PASW 17.0 (SPSS Inc., Chicago, IL, USA) and STATA 11.0 (StataCorp LP, TX, USA) were used. Data are shown as medians and interquartile ranges or numbers with percentages. A univariate logistic regression analysis was performed to examine any associations between various risk factors and hospital mortality. Among the variables used in this model, the risk factors with a *p*-value less than 0.25, including age and gender, were selected for multivariate logistic regression analysis. Validation of the scoring system was performed using standard tests to measure calibration and discrimination. Calibration was defined as agreement between individual probabilities and actual outcomes. The Hosmer-Lemeshow goodness-of-fit (GOF) H statistic and GOF C statistic were used to evaluate agreement between observed and predicted mortality rates.⁸ In the Hosmer-Lemeshow GOF test, a *p*-value greater than 0.05 was interpreted as being well calibrated.⁸ The standardized mortality ratio (SMR) was calculated by dividing the number of observed deaths per group by the number of expected deaths per group. To test for statistical significance, we calculated 95% confidence intervals (CI) according to the method described by Hosmer and Lemeshow. Discrimination was defined as the power to distinguish between non-survivors and survivors, and was assessed by calculating the area under the receiver operating characteristic (aROC) curve as described by Hanley and McNeil.^{6,9} The level of discrimination was defined by the level of aROC. An aROC greater than 0.9 was defined as excellent discrimination, $0.8 \leq \text{aROC} < 0.9$ as good discrimination, and $0.7 \leq \text{aROC} < 0.8$ as modest discrimination.^{6,9} A predicted mortality of 50% was used to calculate the sensitivity and specificity of the APACHE II in predicting hospital mortality.¹⁰

Two-tailed $p < 0.05$ was considered statistically significant.

RESULTS

Baseline characteristics and outcomes of patients

Of 1426 patients enrolled in the FACE study, 600 patients were excluded, therefore, data of 826 patients were used in the analysis. The baseline characteristics and outcomes of the patients are shown in Table 1. Overall hospital mortality was 19.5% (161 out of 826 patients). The patients who died were significantly older, and had higher APACHE II scores, and greater numbers of co-morbidities. They had ICU admissions related to either cardiovascular and respiratory problems, and were more likely to be medical patients and to have more interventions such as mechanical ventilation or CRRT (Table 1).

Medical patients were significantly older, and had higher

APACHE II scores, and more co-morbidities. Furthermore, they were more likely to have admissions related to major organ system failures, and had more patients on CRRT (Table 2). Hospital mortality of medical patients was 28.1%, which was significantly higher than 5.4% of surgical patients ($p<0.001$).

Risk factors for hospital mortality of Korean ICU patients

The multivariate logistic regression model was used to determine independent risk factors of hospital mortality by using all variables with a p -value <0.25 in the univariate model. Among the variables used in the model, age (OR=1.035, $p<0.001$, 95% CI=1.019-1.051), liver cirrhosis (OR=5.328, $p<0.001$, CI=2.745-10.342), hematologic malignancy (OR=3.985, $p=0.001$, CI=1.776-8.943), medical patients

Table 1. Baseline Characteristics and Outcomes of Korean ICU Patients

	Total	Survivor	Non-survivor	<i>p</i> value
Number of patients	826	665 (80.5)	161 (19.5)	-
Age, yrs	63 (50, 73)	62 (49, 71)	68 (56, 77)	<0.001
Male	499 (60.4)	393 (59.1)	106 (65.8)	0.127
APACHE II score	17 (11, 23)	15 (10, 21)	23.5 (20, 28)	<0.001
Adjusted PDR	18.5 (7, 38.6)	16.3 (5.5, 31.9)	43.9 (23.2, 66.5)	<0.001
Co-morbidities				
CHF, NYHA IV	34 (4.1)	21 (3.2)	13 (8.1)	0.013
Liver cirrhosis	77 (9.3)	47 (7.1)	30 (18.6)	<0.001
Chronic dialysis	37 (4.5)	23 (3.5)	14 (8.7)	0.009
Hematologic malignancy	41 (5)	21 (3.2)	41 (5.0)	<0.001
Metastatic cancer	65 (7.9)	42 (6.3)	23 (14.3)	0.002
Reason for ICU admission				
Cardiovascular	228 (27.6)	202 (30.4)	26 (16.1)	<0.001
Respiratory	442 (53.5)	339 (51)	103 (64)	0.004
Gastrointestinal	66 (8)	54 (8.1)	12 (7.5)	0.872
Renal or metabolic	68 (8.2)	55 (8.3)	13 (8.1)	1.000
Central nervous system	22 (2.7)	15 (2.3)	7 (4.3)	0.168
Infection-related admission	338 (40.9)	234 (35.2)	104 (64.6)	<0.001
Admission status				
Medical	512 (62)	368 (55.3)	144 (89.4)	
Surgical	314 (38)	297 (44.7)	17 (10.6)	
Readmission	41 (5)	28 (3.4)	13 (1.6)	0.066
Mechanical ventilation	545 (66)	414 (62.3)	131 (81.4)	<0.001
CRRT	126 (15.3)	64 (9.6)	62 (38.5)	<0.001
ICU length of stay (days)	5 (3, 10)	5 (3, 8)	7 (4, 13)	<0.001
Hospital stay (days)	20 (12, 28)	23 (13, 28)	9 (4, 15)	<0.001
ICU mortality	112 (13.6)	-	-	-
28-day mortality	126 (15.3)	-	-	-
Hospital mortality	161 (19.5)	-	-	-

APACHE, Acute Physiology and Chronic Health Evaluation; PDR, predicted death rate; CHF, congestive heart failure; NYHA, New York Heart Association; ICU, intensive care unit; CRRT, continuous renal replacement therapy.

An outline of patient characteristics and primary reasons for ICU admission. Data are expressed by n (%) or median (interquartile range).

Table 2. Baseline Characteristics and Outcomes of Medical and Surgical Patients (n=826)

	Medical patients	Surgical patients	p value
Number of patients	512 (62)	314 (38)	<0.001
Age, yrs	65 (51, 74)	60 (49, 70)	0.003
Male	313 (61.1)	186 (59.2)	0.608
APACHE II score	20 (15, 25)	11 (8, 16)	<0.001
Adjusted PDR	28.8 (14.4, 49.3)	7 (3.1, 18.3)	<0.001
Co-morbidities			
CHF, NYHA IV	29 (5.7)	5 (1.6)	0.004
Liver cirrhosis	55 (10.7)	22 (7.0)	0.084
Chronic dialysis	29 (5.7)	5 (1.6)	0.038
Hematologic malignancy	40 (7.8)	1 (0.3)	<0.001
Metastatic cancer	59 (11.5)	6 (1.9)	<0.001
Reason for ICU admission			
Cardiovascular	75 (14.6)	153 (48.7)	<0.001
Respiratory	319 (62.3)	123 (39.2)	<0.001
Gastrointestinal	45 (8.8)	21 (6.7)	0.294
Renal or metabolic	54 (10.5)	14 (4.5)	0.002
Central nervous system	19 (3.7)	3 (1)	0.024
Infection-related admission	326 (63.7)	12 (3.8)	<0.001
Readmission	28 (5.5)	13 (4.1)	0.416
Mechanical ventilation	331 (64.6)	214 (68.2)	0.326
CRRT	110 (21.5)	16 (5.1)	<0.001
ICU length of stay (days)	7 (4, 13)	4 (3, 5)	<0.001
Hospital stay (days)	23 (10, 28)	19 (13, 28)	0.374
ICU mortality	104 (20.3)	8 (2.5)	<0.001
28-day mortality	115 (22.5)	11 (3.5)	<0.001
Hospital mortality	144 (28.1)	17 (5.4)	<0.001

APACHE, Acute Physiology and Chronic Health Evaluation; PDR, predicted death rate; CHF, congestive heart failure; NYHA, New York Heart Association; ICU, intensive care unit; CRRT, continuous renal replacement therapy.

An outline of patient characteristics and outcomes according to surgical status. Data are expressed by n (%) or median (interquartile range).

(OR=3.446, $p<0.001$, CI=1.768-6.717), mechanical ventilation (OR=2.001, $p=0.007$, CI=1.204-3.325) and CRRT (OR=4.008, $p<0.001$, CI=2.315-6.942) were associated independently with hospital mortality (Table 3).

Performance assessment of APACHE II in Korean ICUs

Calibration and discrimination data are presented in Table 4. Overall, discrepancy between observed and expected mortalities across all of the strata was demonstrated as a Hosmer-Lemeshow GOF H of 57.54 ($p<0.0001$) and a Hosmer-Lemeshow GOF C of 55.99 ($p<0.0001$). The SMR of the APACHE II model was 0.76 (95% CI 0.62-0.85). The overall discriminatory capability of the APACHE II model was modest (aROC=0.729). Using 50% of predicted mortality as the cut-off, the sensitivity and specificity of the APACHE II for hospital mortality were found to be 36.6% and 87.4%, respectively.

The GOF values according to surgical status are listed in

Table 4. Calibrations of the APACHE II were poor among both medical and surgical patients (H=63.56, $p<0.0001$, C=73.83, $p<0.0001$, and H=33.92, $p<0.0001$, C=33.34, $p=0.0001$, respectively). The discrimination of APACHE II of the medical patients was poor (aROC=0.651, 95% CI=0.69-0.97), whereas the discrimination of the surgical patients was modest (aROC=0.704, 95% CI=0.2-0.59).

DISCUSSION

The present study demonstrated that APACHE II had poor calibration and modest discrimination among critically ill Korean patients. The APACHE II also exhibited poor calibration and poor to modest discrimination in groups divided according to surgical status at ICU admission. At the predicted risk of 50%, the APACHE II had a sensitivity and specificity of 36.6% and 87.4% for hospital mortality, respectively.

Developed in 1985 using a database of North American

Table 3. Multivariate Logistic Regression Analysis of the Effects of Various Clinical Parameters on the Odds Ratio for Hospital Mortality

	Odds ratio	95% CI	<i>p</i> value
Age	1.035	1.019, 1.051	<0.001
Male	1.068	0.692, 1.650	0.766
APACHE II	1.010	0.999, 1.021	0.067
Co-morbidities			
CHF, NYHA IV	1.631	0.716, 3.713	0.244
Liver cirrhosis	5.328	2.745, 10.342	<0.001
Chronic dialysis	0.973	0.394, 2.402	0.952
Hematologic malignancy	3.985	1.776, 8.943	0.001
Metastatic cancer	1.753	0.910, 3.376	0.093
Reason for ICU admission			
Cardiovascular	1.576	0.745, 3.333	0.234
Respiratory	1.452	0.762, 2.764	0.257
Central nervous system	1.236	0.352, 4.343	0.741
Infection-related admission	1.312	0.801, 2.150	0.280
Medical patients	3.446	1.768, 6.717	<0.001
Readmission	1.711	0.735, 3.985	0.213
Mechanical ventilation	2.001	1.204, 3.325	0.007
CRRT	4.008	2.315, 6.942	<0.001

APACHE, Acute Physiology and Chronic Health Evaluation; CHF, congestive heart failure; NYHA, New York Heart Association; CRRT, continuous renal replacement therapy; ICU, intensive care unit; CI, confidence interval.

Table 4. Performance Assessment of the APACHE II in Korean ICU Patients

Group	n	APACHE II score	Adjusted APACHE II (%)	GOF H test		GOF C test		SMR	95% CI	aROC
				H	<i>p</i> value	C	<i>p</i> value			
All	826	17 (11, 23)	18.5 (7, 38.6)	57.54	<0.0001	55.99	<0.0001	0.76	0.62-0.85	0.729
Medical patients	512	20 (15, 25)	28.8 (14.4, 49.3)	63.56	<0.0001	73.83	<0.0001	0.85	0.69-0.97	0.651
Surgical patients	314	11 (8, 16)	7 (3.1, 18.2)	33.92	<0.0001	33.34	0.0001	0.40	0.2-0.59	0.704

APACHE, Acute Physiology and Chronic Health Evaluation; GOF, goodness of fitness; SMR, standardized mortality ratio; CI, confidence interval; aROC, area under the receiver operating characteristic; ICU, intensive care unit.

The APACHE II model exhibited poor calibration (H=57.54, $p<0.0001$, C=55.99, $p<0.0001$) and modest discriminative power (aROC=0.729).

ICU patients, the APACHE II uses the most extreme values of 12 physiological variables 24 hours after admission to ICU, chronic health status, age, and the Glasgow coma scale to calculate a score.¹ Hospital mortality is predicted using the APACHE II score, principal diagnostic category which had to ICU admission, and also depending on whether or not the patient required emergency surgery.¹ In the original article, no goodness-of-fit testing was reported for calibration, and the aROC for the APACHE II was 0.863.¹

External validation is essential before routine application of any model to a group of subjects that is different from the group originally used for model development.¹¹ Although the APACHE II has extensively been validated in different regions of the world, it has never been validated in critically ill East Asian patients. It is popular in Korea because of its ease of use and availability, and is often used as a tool to assess severity of illness and to predict mortality of

Korean ICU patients.

To our knowledge, this is the first study which attempted to validate the APACHE II score in a large, multicenter, prospectively collected database in East Asia. In this study, the performance of the APACHE II was poor not only in general ICU patients, but also in patients group divided according to surgical status. The results of this study are consistent with studies conducted in other regions of the world that have attempted to validate the APACHE II at various times. Most of the studies published since 1996 showed that the APACHE II had modest to good discrimination and poor calibration in general ICUs.¹²⁻¹⁴ The only exception was found in a study conducted prospectively at a single center in Italy from 1994 to 1997.¹⁵ The study sample consisted mainly of surgical patients, and patients with ICU length of stays <24 hr were excluded. Furthermore, studies performed with both medical and surgical patients, using

the APACHE II, show poor predictive power for predicting hospital mortality.^{16,17}

There are several reasons for the limited usefulness of the APACHE II score in predicting mortality for critically ill Korean patients today. First, the differences may represent differences between the population from which the original APACHE II sample was derived and current Korean patients. As stated, the APACHE II was developed mainly using patients from North American ICUs who descended from different populations, and have different underlying diseases and different ICU organizational structures from Korean patients. Second, even if the APACHE II had good performance for Koreans in the 1980s when it was developed, it may not be accurate now. Since its development in the 1980s, there have been significant changes in the prevalences of major diseases, diagnostic approaches and therapeutic modalities. For example, therapeutic modalities, such as liver transplantation, CRRT, early goal-directed therapy and low tidal volume ventilation to reduce ventilator-induced lung injury, were not available at that time.¹ Prediction models should periodically be updated to reflect the changes in medical practice and case-mix over time.⁶ Third, the diagnostic category used to adjust for mortality in the APACHE II is vague.¹ More and more patients have several concurrent, complex problems, and selecting only one principal diagnostic category may be very difficult. In addition, some patients (e.g., liver transplant recipients) may be grouped with patients with very different prognoses (e.g., elective abdominal surgery) because these kinds of patients were not seen in the ICU at the time when the APACHE II was developed. Fourth, APACHE II includes a lead-time bias. The physiological variables included in any scoring system are all dynamic and can be influenced by multiple factors, including recent and ongoing resuscitation and therapy.¹⁸ The interpretation of outcome prediction scores needs to take this potential bias into account, especially with recent increased emphasis on the importance of early resuscitation. The APACHE II is vulnerable to the Boyd and Grounds effect,^{1,19} which increases the risk of overestimation of predicted mortality. A recently developed scoring system tried to eliminate this effect by using only data collected during the first hour of ICU stay.

In this study, a commonly used cut-off of 50% was used to evaluate the prediction value of the APACHE II for hospital mortality.¹⁰ At this cut-off point, the sensitivity and specificity were 36.6% and 87.4%, respectively. Specificity of the APACHE II for hospital mortality in this cohort was

lower than that reported by Capuzzo, et al.¹⁵ If we treat the APACHE II as a diagnostic test, the cut-off point of the APACHE II for hospital mortality in our study can be calculated by using the Youden index, giving a predicted mortality of 22%.²⁰ At this cut-off point, sensitivity, specificity, positive predictive value and negative predictive value were 73.9%, 63.8%, 33.1% and 91%, respectively.

In this study, although the APACHE II score was found to be associated with hospital mortality in the univariate model, it only showed borderline statistical significance in the multivariate model ($p=0.067$). This may be due to the fact that components of the APACHE II score include many of the strong outcome predictors of our model, such as age, gender, and comorbidities, although the test for co-linearity did not show significant co-linearity with predicted mortality of the APACHE II for each variable. Another possible reason might be that the APACHE II score is not an accurate predictor of hospital mortality in these days and age in Korean ICU patients, as described above.

One of the limitations of this study is that the data used were drawn from large university hospitals, therefore, the conclusions may not be applicable to patients treated at smaller hospitals in Korea. Also, certain populations that use the ICU (for example, neurosurgery patients or coronary care unit patients), were not represented in this study, thus limiting the applicability of our findings among these patient populations. However, this is the first study that prospectively collected data from nine university-affiliated hospitals and provided a good case-mix of both medical and surgical ICU patients, thus increasing the generalizability of our findings.

In conclusion, the APACHE II prognostic model has poor calibration and modest discriminative power when applied to ICU patients in Korea. Therefore, a new prognostic model that is customized to Korean patients or an updated model is needed.

REFERENCES

1. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13: 818-29.
2. 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;100:1619-36.
3. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American

- multicenter study. *JAMA* 1993;270:2957-63.
4. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3--from evaluation of the patient to evaluation of the intensive care unit. Part 1: objectives, methods and cohort description. *Intensive Care Med* 2005;31:1336-44.
 5. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3--from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005;31:1345-55.
 6. Afessa B, Gajic O, Keegan MT. Severity of illness and organ failure assessment in adult intensive care units. *Crit Care Clin* 2007;23:639-58.
 7. Lee BH, Inui D, Suh GY, Kim JY, Kwon JY, Park J, et al. Association of body temperature and antipyretic treatments with mortality of critically ill patients with and without sepsis: multi-centered prospective observational study. *Crit Care* 2012;16:R33.
 8. Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance based on logistic regression models. *Stat Med* 1995;14:2161-72.
 9. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
 10. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating (statistics for biology and health)*. New York: Springer; 2010.
 11. 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.
 12. Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006;34:1378-88.
 13. Timsit JF, Fosse JP, Troché G, De Lassence A, Alberti C, Garrouste-Orgeas M, et al. Accuracy of a composite score using daily SAPS II and LOD scores for predicting hospital mortality in ICU patients hospitalized for more than 72 h. *Intensive Care Med* 2001;27:1012-21.
 14. Strand K, Flaatten H. Severity scoring in the ICU: a review. *Acta Anaesthesiol Scand* 2008;52:467-78.
 15. Capuzzo M, Valponi V, Sgarbi A, Bortolazzi S, Pavoni V, Gilli G, et al. Validation of severity scoring systems SAPS II and APACHE II in a single-center population. *Intensive Care Med* 2000;26:1779-85.
 16. Fedullo AJ, Swinburne AJ, Wahl GW, Bixby KR. APACHE II score and mortality in respiratory failure due to cardiogenic pulmonary edema. *Crit Care Med* 1988;16:1218-21.
 17. Cerra FB, Negro F, Abrams J. APACHE II score does not predict multiple organ failure or mortality in postoperative surgical patients. *Arch Surg* 1990;125:519-22.
 18. Vincent JL, Bruzzi de Carvalho F. Severity of illness. *Semin Respir Crit Care Med* 2010;31:31-8.
 19. Lim SY, Ham CR, Park SY, Kim S, Park MR, Jeon K, et al. Validation of the Simplified Acute Physiology Score 3 scoring system in a Korean intensive care unit. *Yonsei Med J* 2011;52:59-64.
 20. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000;45:23-41.