Comparison of the APACHE II and APACHE III Scoring Systems in Patients with Respiratory Failure in a Medical Intensive Care Unit

Chien-Wei Hsu, Shue-Ren Wann, Hung-Ting Chiang, Chih- Hsun Lin, Ming-Ho Kung, and Shoa-Lin Lin

Background and purpose: This retrospective study compared the capability of the Acute Physiology and Chronic Health Evaluation (APACHE) II and APACHE III scoring systems to predict outcome and determined the independent predictors of survival in these scoring systems for patients with respiratory failure in a medical intensive care unit (ICU).
Materials and methods: Seven hundred and eight patients with respiratory failure admitted to the medical ICU throughout a 9-year period were studied. Patients with an ICU stay of less than 24 hours, patients under 12 years of age, and burn and surgery patients were excluded. APACHE scores were calculated at 24 hours after admission. Student's *t*-test was used to compare the total APACHE scores of survivor and non-survivor groups. Multivariate logistic regression analysis was used to determine which variables were predictors of mortality. The discriminative power of APACHE scores to predict in-hospital mortality was studied by the area under the receiver operating characteristic curves of the APACHE II and APACHE III systems, respectively.
Results: Both systems showed a significant association between higher scores and higher mortality. The APACHE II system under-predicted the actual hospital mortality rate. The APACHE III systems had a higher discriminative power (area 0.7462) than the APACHE II systems (area 0.6856; $p < 0.05$). The independent predictors of survival as assessed by APACHE II and III systems were respiratory rate, arterial oxygen pressure, oxygen gradient between alveoli and artery, serum creatinine concentration, and the presence of neurologic abnormalities.
Conclusions: The APACHE III systems has greater discriminative power than the APACHE II systems for predicting in-hospital mortality. The variables of oxygenation, mean artery pressure, respiratory rate, serum creatinine concentration, and Glasgow Coma Scale play important roles in predicting survival for patients with respiratory failure.

(*J Formos Med Assoc* 2001;**100**:437–42)

Key words:
Acute Physiology and Chronic Health Evaluation
APACHE
critical care
respiratory failure

Many different scoring systems are used to analyze the severity of disease in patients admitted to the intensive care unit (ICU); including the acute physiology and chronic health evaluation (APACHE), the simplified acute physiology score (SAPS), and the mortality probability model (MPM) [1–7]. These systems have been compared with one another in different populations [1, 8, 9]. The APACHE II systems has been widely used in many ICUs since 1985; it is a simplified modification of the original APACHE systems developed by Knaus et

Intensive Care Unit, Department of Internal Medicine, Veterans General Hospital, Kaohsiung, and National Yang-Ming University, School of Medicine, Taipei.
Received: 15 September 2000. Revised: 20 October 2000. Accepted: 17 April 2001.
Reprint requests and correspondence to: Dr. Shoa-Lin Lin, Intensive Care Unit, Department of Internal Medicine, Kaohsiung Veterans General Hospital, 386 Ta-Chung 1st Road, Kaohsiung City, Taiwan.

al [1]. It is the first systems to use a quantitative evaluation of disease severity in the ICU [10]. In order to more accurately predict hospital mortality risk for critically ill hospitalized adults, Knaus et al developed the APACHE III systems in 1990, based on the APACHE II, by analyzing data from 17,457 patients from a representative sample of 40 American hospitals, determining the predictive variables, and identifying the additional factors that contributed to outcome [3]. Compared with the APACHE II systems, the APACHE III systems employs more variables and is more complex and time-consuming. The value of assessments made with these two systems has been compared in patients with liver cirrhosis [11] and acute head injury [12], in patients undergoing surgery [13], and in all ICU patients [10]. A study performed in an ICU in the UK demonstrated a similar degree of overall goodness-of-fit between the two systems: while the APACHE II showed better calibration, discrimination was better with the APACHE III [14]. Actual hospital mortality was higher than predicted by both models. Overall, the performance of the APACHE III systems was not superior to the APACHE II systems. A multicenter, multinational study comparing the severity of illness in ICU patients showed that the APACHE III systems had a better receiver operating characteristic (ROC) area than the APACHE II systems, and that it performed better than the APACHE II in discrimination and calibration [10].

The aim of this study was to compare the association between mortality rate and APACHE score in the two systems and to compare the independent predictors of survival in the two systems for patients with respiratory failure.

Materials and Methods

Patients with respiratory failure treated in the medical ICU of Kaohsiung Veterans General Hospital from 1991 to 1999 were included in this study. The inclusion criteria were all patients with respiratory failure under ventilator support admitted to the medical ICU who stayed more than 24 hours in the ICU. The exclusion criteria were an ICU stay of less than 24 hours, age less than 12 years, and burn and surgery patients.

APACHE II and III scores were calculated as described by Knaus et al [2, 3]. The worst physiologic values during the first 24 hours of ICU admission were used for analysis. All patients were followed until hospital discharge or death.

Hospital mortality was defined as death occurring before hospital discharge or death within 3 days of discharge with unstable vital signs on ventilator support. Student's *t*-test was used to compare the total APACHE scores between survivor and non-survivor groups. The logistic regression model of the mortality rate developed using APACHE II or III scores was similar to that used in a previous study [2]. The correlation between the observed mortality rate and the predicted mortality rate determined from the equation of the APACHE II model derived by Knaus et al [2] was assessed by linear regression. The area under the ROC curve was evaluated and the differences between two areas under ROC curves were assessed as in the study of Hanley and McNeil [15].

In both systems, the mean score of every variable was calculated in survivor and non-survivor groups. Multivariate logistic regression analysis was used to determine which variables were independent predictors of survival. A p value of less than 0.05 was considered statistically significant.

Results

The study population consisted of 708 patients consecutively admitted to the ICU including 504 males (71.18%), and 204 females (28.82%). A greater percentage of women died (204, 58.82%) than men (504, 48.81%). Ages ranged from 13 to 91 years (mean, 68.12 ± 14.04 years). The causes of respiratory failure were sepsis (35%), airway disease or acute lung injury (23%), complications of renal failure (13%), massive gastrointestinal bleeding (7%), acute neurologic deficit (7%), complications of diabetes mellitus or metabolic disease (5%), drug poisoning (4%), and others (6%). The overall mortality rate was 48.31%. The distributions of APACHE II and III scores are shown in Figure 1. A graphical representation of the best trend of mortality rate is shown in Figures 2 and 3.

The observed and predicted mortality rate (obtained by the APACHE II model) are compared in Figure 4. The APACHE II systems under-predicted the observed mortality rate, but Figure 5 shows a significant linear correlation between these two mortality rates ($p < 0.05$).

The discriminative power of both systems was compared using ROC curves. The area under the ROC curve showed that the APACHE III systems (area 0.7462; standard error 0.0183) provided a better prediction of mortality than the APACHE II systems (area 0.6856; standard error 0.0198; $p < 0.05$) (Fig. 6).

We further analyzed the average score of every variable in the survivor and non-survivor groups. In the APACHE II systems, all mean scores of variables in the non-survivor group were higher than those in the survivor group (Table 1). Multivariate logistic regression analysis revealed a significant difference between

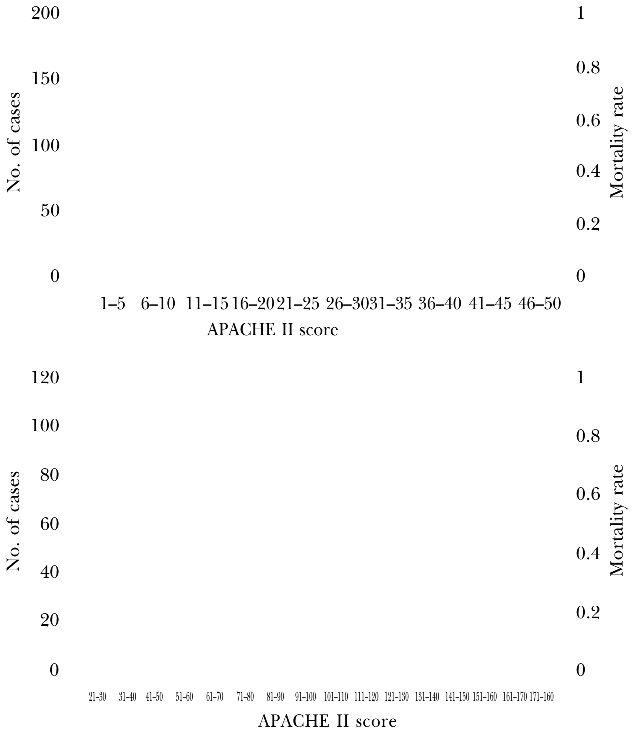

Fig. 1. *Number of patients and mortality rates in each range of APACHE II (upper panel) and III (lower panel) scores. The bar diagram represents the number of patients and the diamond line represents the mortality rate. Higher scores in both systems were correlated with higher mortality rates.*

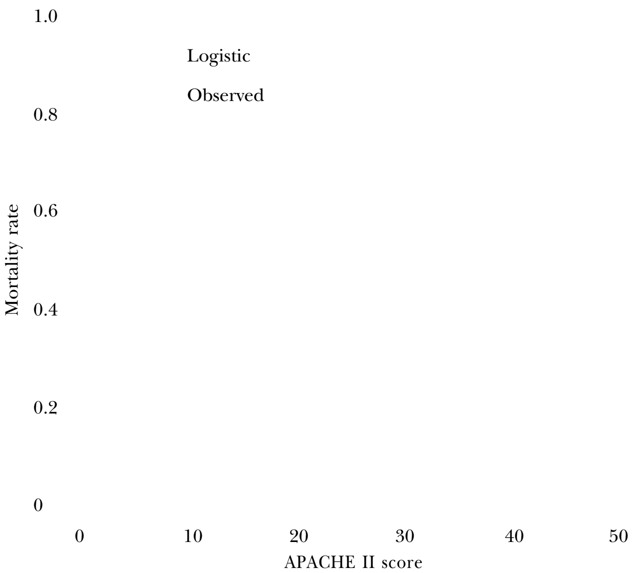

Fig. 2. *The best trend of mortality rate* vs. *APACHE II scores by logistic regression model, Ln(Y/1–Y) = 0.0954*X *– 2.4825, where* X *is the APACHE II score and* Y *is the mortality rate (R^2 = 0.989).*

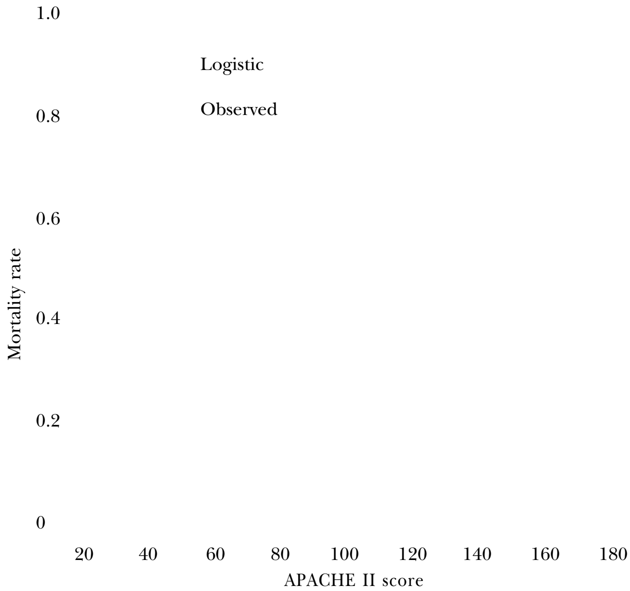

Fig. 3. *The best trend of mortality rate vs. APACHE III scores by logistic regression model, Ln(Y/1–Y) = 0.0326*X *– 3.0038, where* X *is the APACHE III score and* Y *is the mortality rate (R^2 =0.944).*

the survivor and non-survivor groups in body temperature, mean arterial blood pressure (MBP), pulse rate, respiratory rate, arterial oxygen pressure (PaO_2), oxygen gradient between alveoli and artery ($PA\text{-}aO_2$), serum creatinine concentration, and Glasgow Coma Scale scores. For the APACHE III systems, the mean scores of all variables in the non-survivor group were also higher than those in the survivor group, with significant differences in MBP, respiratory rate, PaO_2, $PA\text{-}aO_2$, serum creatinine concentration (for acute

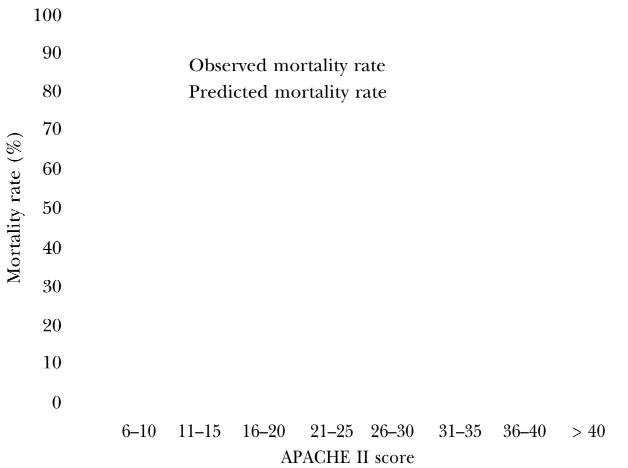

Fig. 4. *Mortality rates predicted by the APACHE II model developed by Knaus et al (solid bars) were compared with the observed mortality rates (open bars) in patients with respiratory failure (*n *= 708).*

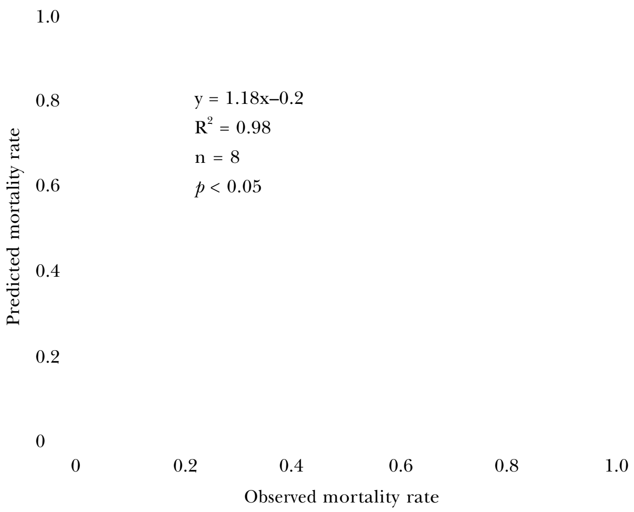

Fig. 5. *A significant correlation was found between the observed mortality rate and predicted mortality rate in the APACHE II scoring system (Y axis, predicted mortality rate; X axis, observed mortality rate).*

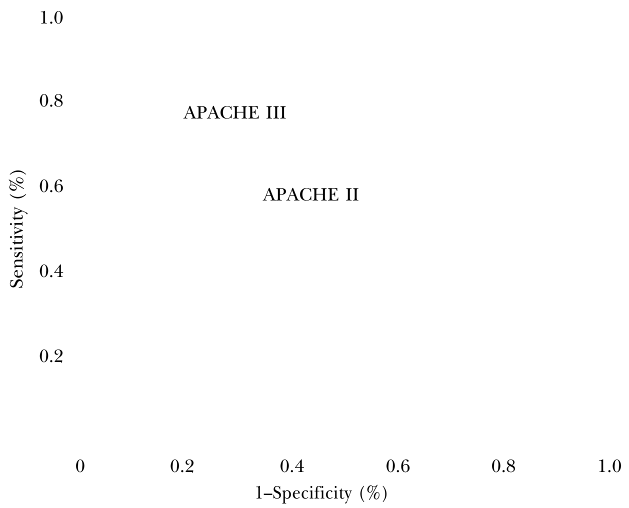

Fig. 6. *Receiver operating characteristic (ROC) curves were drawn at different cut-off values for APACHE II and III. The area under the ROC curve for APACHE III is larger than that for APACHE II.*

renal failure), urine output, glucose, neurologic abnormality, and comorbid condition (Table 2).

Discussion

In this study, the survivor group had significantly lower APACHE II and APACHE III scores than the non-survivor group. Logistic regression analysis revealed that, in both systems, there was a link between higher scores and higher mortality rate.

As shown in Figure 4, analysis of risk prediction using the APACHE II model proposed by Knaus et al for respiratory failure patients showed a pattern of underprediction of actual mortality in this study. The pattern of underprediction may partly reflect that our hospital is a tertiary medical center with only eight beds in the medical ICU to serve the entire 1141-bed hospital.

Table 1. Mean scores of each variable in the APACHE II system in the survivor and non-survivor groups were evaluated by multivariate logistic regression analysis

	All (n = 708)	Survivor (n = 366)	Non-survivor (n = 342)	p* value	Odds ratio
BT	0.60 ± 0.93	0.47 ± 0.83	0.75 ± 1.01	0.04	1.21
MBP	1.86 ± 1.28	1.60 ± 1.22	2.13 ± 1.30	0.02	1.18
Pulse	2.15 ± 0.10	2.00 ± 1.07	2.32 ± 0.89	0.01	1.29
RR	1.72 ± 0.78	1.63 ± 0.76	1.82 ± 0.79	< 0.01	1.42
PaO_2	1.17 ± 1.61	0.82 ± 0.82	1.55 ± 1.74	< 0.01	1.24
$PA\text{-}aO_2$	0.47 ± 1.13	0.43 ± 1.07	0.53 ± 1.20	0.01	1.29
Creatinine	1.78 ± 1.64	1.30 ± 1.02	2.28 ± 1.62	< 0.01	1.36
GCS	6.51 ± 3.58	5.80 ± 3.23	7.28 ± 3.77	< 0.01	1.09
pH	1.22 ± 1.35	1.02 ± 1.20	1.43 ± 1.48	0.36	0.94
Sodium	0.35 ± 0.79	0.33 ± 0.79	0.38 ± 0.78	0.23	0.88
Potassium	0.74 ± 1.03	0.71 ± 1.02	0.78 ± 1.03	0.99	1.00
Hematocrit	0.76 ± 1.04	0.66 ± 0.99	0.86 ± 1.08	0.21	1.12
WBC	0.62 ± 0.91	0.54 ± 0.84	0.72 ± 0.96	0.83	1.02
Age	4.56 ± 1.78	4.51 ± 1.90	4.60 ± 1.66	0.81	1.01
Chronic health point	0.56 ± 1.63	0.44 ± 1.50	0.46 ± 1.66	0.60	1.03
Total	25.07 ± 7.96	22.25 ± 6.98	28.09 ± 7.84	< 0.01	

*Survivor *vs.* non-survivor. BT = body temperature; MBP = mean arterial blood pressure; RR = respiratory rate; PaO_2 = arterial oxygen pressure; $PA\text{-}aO_2$ = gradient between alveolar and arterial oxygen pressure; GCS = Glasgow Coma Scale; WBC = white blood cell count.

Table 2. Mean scores of each variable in the APACHE III system in the survivor and non-survivor groups were evaluated by multivariate logistic regression analysis

	All (n = 708)	Survivor (n = 366)	Non-survivor (n = 342)	p* value	Odds ratio
MBP	9.25 ± 5.62	7.90 ± 4.76	10.71 ± 6.10	0.01	1.05
RR	3.96 ± 3.84	3.41 ± 3.75	4.56 ± 3.85	0.02	1.08
PaO_2	1.46 ± 3.10	1.45 ± 3.08	1.47 ± 3.13	0.01	1.07
$PA\text{-}aO_2$	0.47 ± 1.13	3.20 ± 5.08	6.10 ± 9.10	0.01	1.05
Cr (acute renal failure)	1.91 ± 3.94	0.78 ± 2.72	3.13 ± 4.63	0.02	1.09
Urine output	4.43 ± 4.88	3.17 ± 3.31	5.77 ± 5.84	< 0.01	1.08
Glucose	1.48 ± 2.15	1.11 ± 1.88	1.87 ± 2.35	< 0.01	1.14
Neurologic abnormality	22.04 ± 14.56	19.07 ± 12.57	25.21 ± 15.83	0.01	1.01
Comorbid condition	0.73 ± 2.86	0.25 ± 1.66	1.25 ± 3.67	< 0.01	1.14
Temperature	0.64 ± 1.39	0.46 ± 1.00	0.83 ± 1.69	0.05	1.03
Pulse	7.50 ± 5.61	6.52 ± 5.67	8.54 ± 5.35	0.15	1.12
Hematocrit	2.63 ± 0.99	2.54 ± 1.08	2.72 ± 0.88	0.13	1.16
WBC	0.75 ± 2.17	0.53 ± 1.42	0.99 ± 2.73	0.86	1.01
Cr (chronic renal failure)	2.18 ± 2.94	2.12 ± 2.90	2.24 ± 2.97	0.70	0.98
BUN	7.47 ± 4.86	6.31 ± 4.32	8.71 ± 5.09	0.54	1.05
Sodium	0.72 ± 1.12	0.70 ± 1.11	0.75 ± 1.14	0.15	0.89
Albumin	2.02 ± 3.59	1.45 ± 3.05	2.63 ± 4.00	0.09	1.05
Bilirubin	0.94 ± 2.92	0.60 ± 2.11	1.31 ± 3.56	0.14	1.05
Age	13.28 ± 5.51	13.10 ± 5.80	13.49 ± 5.18	0.55	1.01
Acid-base	4.95 ± 4.12	4.51 ± 3.88	5.41 ± 4.32	0.44	0.98
Total	92.94 ± 31.35	79.16 ± 23.83	107.66 ± 31.75	< 0.01	

*Survivor vs. non-survivor. MBP = mean arterial blood pressure; RR = respiratory rate; PaO_2 = arterial oxygen pressure; $PA\text{-}aO_2$ = gradient between alveolar and arterial oxygen pressure; Cr (acute renal failure) = serum creatinine concentration in patients with acute renal failure; Neurologic = neurologic abnormality; WBC = white blood cell count; Cr (chronic renal failure) = serum creatinine concentration in patients with chronic renal failure; BUN = blood urea nitrogen; Acid-base = acid-base abnormality.

Many critically ill patients were thus aggressively treated before being transferred to the ICU. The extent of physiologic derangement in these patients is likely to have been less marked at the time of ICU admission, which may explain why the predicted mortality rate calculated by the APACHE II model was lower than the actual mortality rate. This phenomenon has been described as lead-time bias [16, 17].

The area under the ROC curve provides an adequate measure to compare discrimination for each scoring systems [18]. Our results demonstrated that the APACHE III systems yields better results than the APACHE II systems ($p < 0.05$). The discriminative power of the APACHE III systems may be increased by the inclusion of more physiologic variables.

Many factors can influence the assessment of outcome in an ICU population, including case mix, severity of illness, quality of care, source of ICU admission (emergency room, hospital floor, another hospital) [19], lead-time bias [16, 17], and interobserver bias [20, 21]; the results may vary greatly among ICUs [22].

The APACHE III systems has more independent predictive variables of survival than the APACHE II systems. The common variables in the APACHE II and III systems are MBP, respiratory rate, PaO_2, $PA\text{-}aO_2$, and serum creatinine concentration (for acute renal failure in APACHE III). Respiratory rate, PaO_2, and $PA\text{-}aO_2$ are directly related to pulmonary condition. Although MBP and serum creatinine concentration are not directly related to pulmonary condition, they are also independent predictors of survival.

The variables used to assess neurologic abnormality in the APACHE III systems are similar to those of the Glasgow Coma Scale, which is one of the physiologic variables in the APACHE II systems. These non-pulmonary parameters were both found to be related to mortality. The variables used to assess comorbid conditions in the APACHE III systems and in the APACHE II systems are similar. However, we found that while comorbid condition was associated with mortality, the chronic health points did not correlate with mortality. The measurement of chronic health points depends on the existence of severe organ insufficiency or immunocompromise evident prior to hospital admission; the score is 5 is given if the patient has one of the following organ insufficiencies: heart, lung, kidney, liver, or immune systems. It may be associated with the prognosis of patients with respiratory failure, but the APACHE II systems lacks explicit definitions for chronic health conditions that contribute to significant variability of data collection [20, 21].

In the APACHE III systems, the scoring method for serum creatinine concentration is divided into two parts: acute renal failure and chronic renal failure. Serum creatinine concentration is an independent predictor of survival for acute renal failure, but not for chronic renal failure. This finding suggests that the role of chronic renal failure is not as important as acute renal failure in predicting the survival of respiratory failure patients.

In conclusion, although the APACHE III systems is more complex and data collection is more time-consuming, it has better discriminative prognostic power than the APACHE II systems. Oxygenation, MBP, respiratory rate, serum creatinine concentration, and Glasgow Coma Scale are the common predictors of survival in the two systems. These variables play important roles in predicting survival for patients with respiratory failure.

ACKNOWLEDGMENT: This study was supported in part by the Kaohsiung Veterans General Hospital, Grant No. VGHKS 88-14 and VGHKS 89-73.

References

1. Knaus WA, Zimmerman JE, Wagner DP, et al: APACHE—acute physiology and chronic health evaluation: a physiologically based classification systems. *Crit Care Med* 1981; 9:591–7.
2. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: a severity of disease classification systems. *Crit Care Med* 1985;13:818–29.
3. Knaus WA, Wagner DP, Draper EA, et al: The APACHE III prognostic systems. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100: 1619–36.
4. Le Gall J, Loirat P, Alperovith A, et al: A simplified acute physiology score for ICU patients. *Crit Care Med* 1984;12: 975–7.
5. Le Gall J, Lemeshow S, Saulnier F: A new simplified acute physiology score (SAPS II) based on a European/ North American multi-center study. *JAMA* 1993;270:2957–63.
6. Lemeshow S, Teres D, Pastides H, et al: A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985;13:519–25.
7. Lemeshow S, Teres D, Klar J, et al: Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993;270:2478–86.
8. Lemeshow S, Teres D, Avrunin JS, et al: A comparison of methods to predict mortality of intensive care unit patients. *Crit Care Med* 1987;15:715–22.
9. Castella X, Gilabert J, Torner F, et al: Mortality prediction models in intensive care: Acute Physiology and Chronic Health Evaluation II and Mortality Prediction Model compared. *Crit Care Med* 1991;19:191–7.
10. Castella X, Artigas A, Bion J, et al: A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. *Crit Care Med* 1995;23:1327–35.
11. Zauner CA, Apsner RC, Kranz A, et al: Outcome prediction for patients with cirrhosis of the liver in a medical ICU: a comparison of the APACHE scores and liver-specific scoring systems. *Intensive Care Med* 1996;22: 559–63.
12. Cho DY, Wang YC: Comparison of the APACHE III, APACHE II and Glasgow Coma Scale in acute head injury for prediction of mortality and functional outcome. *Intensive Care Med* 1997;23:77–84.
13. Barie PS, Hydo LJ, Fischer E: Comparison of APACHE II and III scoring systems for mortality prediction in critical surgical illness. *Arch Surg* 1995;130:77–82.
14. Beck DH, Taylor BL, Millar B, et al: Prediction of outcome from intensive care: a prospective cohort study comparing Acute Physiology and Chronic Health Evaluation II and III prognostic systems in a United Kingdom intensive care unit. *Crit Care Med* 1997;25: 9–15.
15. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
16. Dragsted L, Jorgensen J, Jensen NH, et al: Interhospital comparisons of patient outcome from intensive care: importance of lead-time bias. *Crit Care Med* 1989;17: 418–22.
17. Wagner D, Knaus W, Bergner M: APACHE III study design: analytic plan for evaluation of severity and outcome in intensive care unit patients. Statistical methods. *Crit Care Med* 1989;17:194–8.
18. Suter P, Armaganidis A, Beaufil F, et al: Consensus conference organized by the ESICM and the SRLE. Predicting outcome in ICU patients. *Intensive Care Med* 1994;20:390–7.
19. Borlase BC, Baxter JK, Kenney PR et al: Elective intrahospital admission versus acute interhospital transfers to a surgical intensive care unit: cost and outcome prediction. *J Trauma* 1991;31:915–8.
20. Polderman KH, Thijs LG, Girbes AR: Interobserver variability in the use of APACHE II scores. *Lancet* 1999; 353:380.
21. Chen LM, Martin CM, Morrison TL, et al: Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit Care Med* 1999; 27:1999–2004.
22. Marsh HM, Krishan I, Naessens JM, et al: Assessment of prediction of mortality by using the APACHE II scoring systems in intensive care units. *Mayo Clin Proc* 1990;65: 1627–9.