A Comparison of APACHE II and SAPS II Scoring Systems in Predicting Hospital Mortality in Thai Adult Intensive Care Units

Bodin Khwannimit MD*,
Alan Geater PhD**

** Division of Critical Care, Department of Internal Medicine, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla*
*** Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla*

Objective: *To assess the performance of Acute Physiology and Chronic Health Evaluation II (APACHE II) and Simplified Acute Physiology Score II (SAPS II) in Thai critically ill patients.*
Material and Method: *Prospective observational cohort study conducted between July 1, 2004 and October 31, 2005 in the Intensive Care Unit (ICU) of Songklanagarind Hospital, an 800-beds tertiary referral university teaching hospital.*
Results: *One thousand three hundred sixteen patients were enrolled. There were 310 deaths (23.6%) at hospital discharge. APACHE II and SAPS II predicted hospital mortality 30.5 $\pm$ 28.2 and 30.5 $\pm$ 29.8 respectively. Both models showed excellent discrimination. The discrimination of APACHE II was better than SAPS II (0.911 and 0.888, $p < 0.001$). However, both systems presented a poor calibration. The Hosmer-Lemeshow goodness-of-fit H and C statistics were 66.59 and 66.65 of APACHE II ($p < 0.001$) and 54.01 and 71.44 of SAPS II ($p < 0.001$).*
Conclusion: *APACHE II provided better discrimination than SAPS II, but both models showed poor calibration in over predicting mortality in our ICU patients. Customized or new severity scoring systems should be developed for critically ill patients in Thailand.*

Keywords: *Severity of illness, Intensive care, Mortality prediction, Acute Physiology and Chronic Health Evaluation (APACHE II), Simplified Acute Physiology Score (SAPS II)*

J Med Assoc Thai 2007; 90 (4): 643-52
Full text. e-Journal: *http://www.medassocthai.org/journal*

Severity scoring systems are designed to provide an estimate of the probability of hospital mortality for critically ill patients[1]. Measuring the severity of illness enables hospital administrators to describe ICU populations and can be helpful in clinical decision making and guiding resource allocation[2]. Severity-adjusted outcomes should not be used only to measure individual ICU performance, but can also be used to compare between ICU[1-3]. Severity scoring systems have been around for many years now. The first two introduced were the Acute Physiology and Chronic Health Evaluation score (APACHE) and the Simplified Acute Physiology Score (SAPS). Modified versions of both models have been created since then, notably APACHE II[4] and III[5], and SAPS II[6] and III[7]. The most popular current scales are APACHE II and SAPS II due to their reliability and relative ease of use[1].

Ability of any such systems to accurately predict mortality rate is determined by its discrimination and calibration. A model has good discrimination when it can distinguish accurately between patients who will die or survive. Calibration evaluates the degree of correspondence between the estimated overall probability of death produced by a model and the actual mortality rate. When the observed number of

Correspondence to : Khwannimit B, Division of Critical Care, Department of Internal Medicine, Songklanagarind Hospital, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand. Phone: 074-451452, Fax: 074-429385, E-mail: kbordin@medicine.psu.ac.th

deaths is close to the number expected from the model, the model is considered to be well calibrated[1]. Good discrimination of both APACHE II and SAPS II has been found in several studies conducted in Western countries[8-12]; however, most of these studies reported poor calibration[9-12]. It is not clear to what extent these findings should be extrapolated to ICU patients in different ICUs or different countries[1]. In Asian countries, there have been four published studies in which the data showed good discrimination of APACHE II[13-16], three in which SAPS II exhibited good discrimination[13,14,16] and two which showed poor discrimination in APACHE II[17,18]. All studies reported a lack of fit for both systems[13,16]; however, only a single study reported calibration of these scores in Thailand[14]. Before applying severity scoring systems in a specific country or different type of ICU, their prognostic performance must be validated.

In the present study, the authors evaluated and compared the ability and validity of APACHE II and SAPS II scoring systems in predicting hospital mortality at a mixed medical-surgical ICU of tertiary referral university hospital in southern Thailand.

Material and Method

Patients and Data collection

The present study was performed at Songklanagarind Hospital, an 800-beds tertiary referral university teaching hospital at Prince of Songkla University, Hat Yai, Songkhla. In the hospital, there are two units in the adult ICU: a ten-bed surgical ICU and a nine-bed mixed medical and coronary care unit. The surgical ICU serves all post operative patients. The medical ICU serves all critically ill patients from the Department of Internal Medicine.

Data collection lasted from July 1, 2004 until October 31, 2005. All data were collected concurrently for consecutive ICU admissions. Patients who were excluded from the present study included those who were younger than 15 years of age, suffered burn injuries, had not received attempted cardiac resuscitation, were withdrawn from treatment, died within four hours of admission to ICU or who stayed in the ICU less than 24 hours. If patients had been admitted more than once to the ICU during the study period, only the first admission was included. Approval for the project was obtained from the faculty Ethics Committee.

The following data were collected: basic demographic characteristics including sex, age, and principal diagnostic category leading to ICU admission according to Knaus et al[4]. Type of patient status was defined classically as medicine, scheduled surgery, or unscheduled surgery[6]. The worst physiological values in the 24 hours following ICU admission for APACHE II and SAPS II variable were abstracted from clinical and laboratory records, using the variable definitions reported in the original literature[4,6]. The standard mortality ratio (SMR) was calculated by dividing observed hospital mortality by the predicted hospital mortality. The 95% confidence interval (CI) for the SMRs were calculated by regarding the observed mortality as a Poisson variable, then dividing its 95%CI by the predicted mortality[19]. The presence of organ dysfunction during the first 24 hours in the ICU was assessed by the Sequential organ dysfunction scores (SOFA), as described by Vincent et al[20]. In sedated patients, the Glasgow Coma Score (GCS) was determined either from medical records before sedation or through interviewing the physician who ordered the sedation. However, if a variable could not be measured the GCS was assumed normal. The predicted hospital mortality was calculated using the original formulas of each severity scoring systems[4,6]. ICU, hospital length of stay (LOS) and lead time (the interval from hospital admission to ICU admission) were calculated. Patients were followed up until ICU and hospital discharge in order to registrar their survival status. All data were collected by one well-trained research assistant and then all patients' records were rechecked for accuracy by the authors.

Statistical analysis

Stata 7 software (Stata Corporation, College Station Tx.) were used for statistical analysis. Data are presented as mean $\pm$ SD, when indicated. Student's t test and Wilcoxon's rank sum test were used to compare normally distributed continuous variables and non-parametric data, respectively. Chi-square statistic was used to test for the statistical significance of categorical variables. A p-value less than 0.05 was considered statistically significant.

The ability and accuracy of the models for hospital mortality prediction were determined by examining their discrimination and calibrations. Discrimination was tested by the area under the receiver operating characteristics (ROC) curve, computed by a modification of the Wilcoxon statistic, as described by Hanley and McNeil[21] and 2×2 classification table at decision criterion of 50%. The Hosmer-Lemeshow goodness-of-fit H and C statistics were used to evaluate calibration[22]. Patients were rank-ordered in ten groups according to their probability of death to cal-

culate the H statistic. To calculate the C-statistic, the study population was divided into deciles of predicted risk. The predicted and actual number of survivors and non-survivors were compared statistically using formal goodness-of-fit calculations to determine whether the discrepancies between predicted and actual values were statistically significant ($p < 0.05$).

Results

During the study period, data were collected on 1,316 patients. Overall 236patients (17.9%) died in the ICU and 310 patients (23.6%) died in the hospital. The SMR using both models was 0.77 (95%CI = 0.69-0.86).

The patients' demographic, type of admission and general diagnostic categories of the presented patients are shown in Table 1. The severity of patients' illness as assessed by APACHE II and SAPS II are shown in Table 2.

The area under the ROC curve was 0.911 (95%CI = 0.891-0.930) for APACHE II and 0.888 (95%CI = 0.867-0.909) for SAPS II. When the APACHE II and SAPS II curves are compared, the area under the ROC curve of APACHE II showed to be statistically and significantly, higher than SAPS II system ($p < 0.001$) (Fig. 1).

The results of the 2×2 classification table for APACHE II and SAPS II are shown in Table 3. The false-positive rate was lower for APACHE II than SAPS II (8.45% *vs.* 10.83%).

The Hosmer-Lemeshow goodness-of-fit test H statistic revealed a poor performance for both systems (Table 4). Similar results were obtained, using the Hosmer-Lemeshow goodness-of-fit test C statistic (Table 5). These findings indicated a significant a lack of fit for both models.

Predicted risk of hospital mortality as described by the two models showed a highly significant correlation ($R = 0.907$). It is clear that, although highly correlated, the predicted mortality from the two severity scoring systems are widely dispersed (Fig. 2), mostly in the mid-range of predicted risk of death. The predictions of hospital deaths seem cloudier related in the extremes of risk.

Regarding the bias resulting from case mix, a sub sample of cases was analyzed excluding patients with coronary care and postoperative cardiac surgery

Table 1. Demographic and clinical characteristics of patients in study

	Total (n = 1316) n (%)	Survivors (n = 1006) n (%)	Non-survivors (n = 310) n (%)	p-value
Age (years)	55.6 ± 18.2	55.7 ± 17.9	55.3 ± 19.1	0.797
Male 754 (57.3)	570 (56.7)	184 (59.3)	0.402	
Type of admission				
Medicine	732 (55.6)	496 (49.3)	236 (76.1)	<0.001
Surgery, scheduled	357 (27.2)	337 (33.5)	20 (6.5)	<0.001
Surgery, unscheduled	227 (17.2)	173 (17.2)	54 (17.4)	0.928
Categories of diseases				
Non-operative				
Respiratory disease	86 (6.5)	76 (7.6)	10 (3.2)	0.007
Cardiovascular disease	350 (26.7)	262 (26)	88 (28.4)	0.414
Coronary artery disease	186 (14.2)	172 (17.1)	14 (4.5)	<0.001
Others	164 (12.5)	90 (8.9)	74 (23.8)	<0.001
Sepsis	220 (16.7)	103 (10.2)	117 (37.7)	<0.001
Neurological disease	16 (1.2)	13 (1.3)	3 (1)	0.649
Gastrointestinal disease	30 (2.3)	13 (1.3)	17 (5.5)	<0.001
Other	30 (2.2)	29 (2.9)	1 (0.3)	0.008
Post-operative disease				
Post-CABG*	55 (4.2)	48 (4.8)	7 (2.2)	0.053
Post cardiac surgery	156 (11.9)	148 (14.7)	8 (2.6)	<0.001
Brain and spinal cord	124 (9.4)	110 (10.9)	14 (4.5)	<0.001
Gastrointestinal	94 (7.1)	78 (7.8)	16 (5.2)	0.121
Other	155 (11.8)	126 (12.5)	29 (9.4)	0.001

Table 2. APACHE II and SAPS II scores, predicted risk of hospital mortality and LOS of patients in study

	Total (n = 1316)	Survivors (n = 1006)	Non-survivors (n = 310)	p-value
Age (years)	55.6 ± 18.2	55.7 ± 17.9	55.3 ± 19.1	0.797
APS[#]	14.7 ± 9.1	11.5 ± 6	25.1 ± 9.4	<0.001
APACHE II	18.3 ± 9.6	14.8 ± 6.6	29.5 ± 9.3	<0.001
SAPS II	40.1 ± 20.4	33.0 ± 14.2	63.1 ± 20.5	<0.001
SOFA	5.8 ± 4.2	4.5 ± 3.2	10.1 ± 4.1	<0.001
APACHE II predictionof death (%)	30.5 ± 28.2	19.6 ± 18	66.0 ± 25.8	<0.001
SAPS II prediction of death (%)	30.5 ± 29.8	19.9 ± 20.6	65.0 ± 28.8	<0.001
ICU LOS (day)*	2 (1-5)	2 (1-4)	3 (1-6)	0.023
Hospital LOS (day)*	15 (8-29)	17 (10-31)	8 (2-22)	<0.001
Lead time (day)*	1 (0-5)	1 (0-5)	1 (0-5)	0.275

[#] APS; acute physiology scores of APACHE II
* Median and interquartile range

as in the original SAPS II and APACHE II models. Discrimination and calibrations of both systems after excluding these patients are shown in Table 6. The area under the ROC curve of APACHE II was 0.913 after excluding post coronary artery bypass graft (CABG) patients (as in the original APACHE II model). However, the Hosmer-Lemeshow goodness-of-fit test H and C statistics also showed poor calibration. The SAPS II scoring system showed good discrimination after calculating as in the original SAPS II, but the poor goodness-of-fit persisted.

Discussion

APACHE II and SAPS II are two of the most popular models in common use throughout the world, although later version of each, APACHE III and SAPS III, have also been created. However, there are limitations in the application of these systems arising from

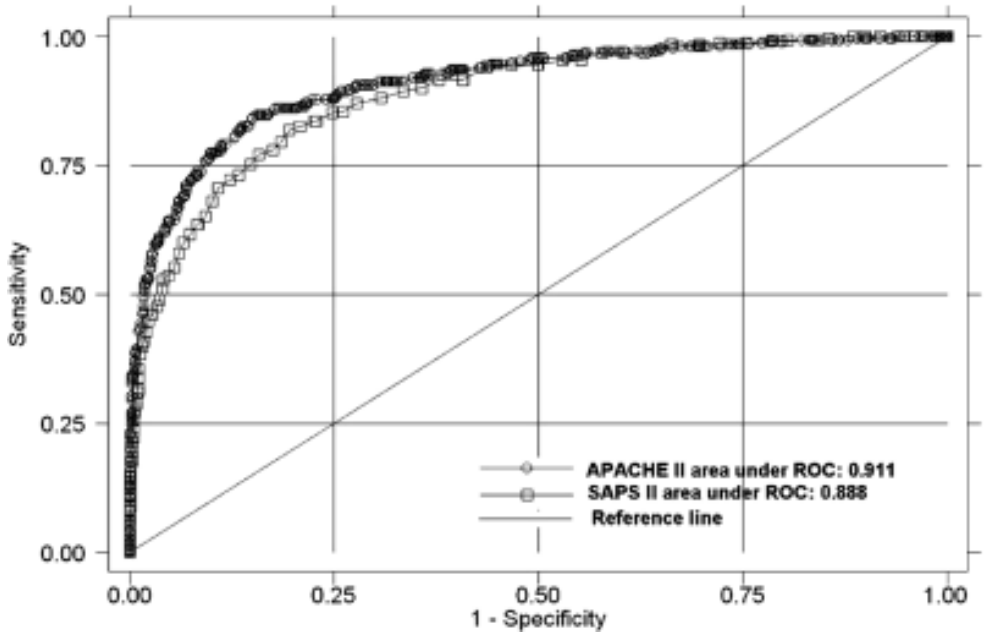


Fig. 1 Receiver operating characteristic curves (ROC) for APACHE II and SAPS II systems

Table 3. Classification table of APACHE II and SAPS II

Decision criterion 50%	APACHE II		SAPS II	
	Predicted to live (n)	Predicted to die (n)	Predicted to live (n)	Predicted to die (n)
Observed survivors	921	85	897	109
Observed non-survivors	81	229	91	219
Sensitivity	73.87 (68.61-78.67)		70.65 (65.23-75.66)	
Specificity	91.55 (89.66-93.20)		89.17 (87.08-91.02)	
Positive predictive value	72.93 (67.65-77.78)		66.77 (61.38-71.85)	
Negative predictive value	91.92 (90.05-93.53)		90.78 (88.81-92.52)	
Overall correct classification	87.38 (85.47-89.13)		84.80 (82.75-86.70)	

* In parenthesis are 95%CI

Table 4. The Hosmer-Lemeshow goodness-of-fit test H statistic for APACHE II and SAPS II

APACHE II predicted death rate	n	Non-survivors		Survivors	
		Observed	Expected	Observed	Expected
0.00-0.10	413	9	23.80	404	389.20
0.11-0.20	268	16	38.32	252	229.68
0.21-0.30	151	17	36.92	134	114.08
0.31-0.40	92	14	32.33	78	59.67
0.41-0.50	78	25	34.81	53	43.19
0.51-0.60	66	30	36.62	36	29.38
0.61-0.70	60	35	39.37	25	20.63
0.71-0.80	60	44	45	16	15.00
0.81-0.90	61	55	51.94	6	9.06
0.91-1.00	67	65	63.07	2	3.93
Total	1316	310	402.19	1006	913.81

Chi-square H = 66.59 *df* =10 p < 0.001

SAPS II predicted death rate	n	Non-survivors		Survivors	
		Observed	Expected	Observed	Expected
0.00-0.10	464	14	22.5	450	441.5
0.11-0.20	238	19	35.13	219	202.87
0.21-0.30	127	18	31.02	109	95.98
0.31-0.40	89	20	30.97	69	58.03
0.41-0.50	70	20	31.51	50	38.49
0.51-0.60	76	33	41.87	43	34.13
0.61-0.70	54	28	35.39	26	18.61
0.71-0.80	53	31	40.29	22	12.71
0.81-0.90	55	43	47.02	12	7.98
0.91-1.00	90	84	86.08	6	3.92
Total	1316	310	401.77	1006	914.23

Chi-square H = 54.01 *df* =10 p < 0.001

Table 5. The Hosmer-Lemeshow goodness-of-fit test C statistic for APACHE II and SAP II

APACHE II predicted death rate	n	Non-survivors		Survivors	
		Observed	Expected	Observed	Expected
0.00-0.05	131	1	3.1	89	86.9
0.06-0.08	132	4	3.4	80	80.6
0.09-0.11	131	4	4.5	89	88.5
0.12-0.16	132	5	4.2	66	66.8
0.17-0.22	132	7	6.7	80	80.3
0.23-0.38	131	12	9.6	71	73.4
0.39-0.48	132	20	18.1	64	65.9
0.49-0.68	131	36	35.5	48	48.5
0.69-0.89	132	55	60.3	33	27.7
0.90-1.00	132	71	69.6	9	10.4
Total	1316	310	402.19	1006	913.81

Chi-square C = 66.65 *df* = 10 $p < 0.001$

SAPS II predicted death rate	n	Non-survivors		Survivors	
		Observed	Expected	Observed	Expected
0.00-0.73	131	2	6.25	129	124.75
0.74-0.82	132	3	9.77	129	122.24
0.83-0.11	131	4	10.94	127	120.06
0.12-0.17	132	4	15.21	128	116.79
0.18-0.25	132	10	22.19	122	109.81
0.26-0.30	131	14	32.64	117	98.36
0.31-0.45	132	17	39.77	115	92.23
0.46-0.68	131	51	59.32	80	71.68
0.69-0.87	132	82	90.15	50	41.85
0.88-1.00	132	123	115.54	9	16.46
Total	1316	310	401.77	1006	914.23

Chi-square C = 71.44 *df* =10 $p < 0.001$

the need to validate them first using independent sample data from different ICUs in different countries such as in Thailand.

In the present study, the authors evaluated the ability and validity of APACHE II and SAPS II systems to accurately predict hospital mortality in a Thai adult mixed-case ICU. Both models showed excellent discrimination, although the authors found that discrimination was better for APACHE II than for SAPS II; the same is true for the percentage of overall correct classification. Good discrimination of both models has been reported in previous studies[8-13,15]. The area under the ROC of both systems in the present study was higher than that in the other reports. Previously reported area under the ROC curve of APACHE II and SAPS II included 0.839 and 0.870 in Greece[10], 0.787 and 0.817 in Portugal[11], 0.83 and 0.79 in Saudi Arabia[13], 0.819 and 0.840 in Tunisia[12] and 0.88 and 0.87 in Hong Kong[16], respectively and 0.88 in the original SAPS II[6]. The ability of the two models to correctly predict group prognosis also was assessed by means of 2×2 decision table. Other reports using this method showed lower correct classification, for instance 85.5% in the original APACHE II[4], and then 83.06% and 82.75% in Greece[10], 78.00% and 76.37% in Portugal[11], 77% and 77% in Saudi Arabia[13] with

Table 6. Discrimination and calibration of APACHE II and SAPS II after excluding coronary and postoperative cardiac surgery cases

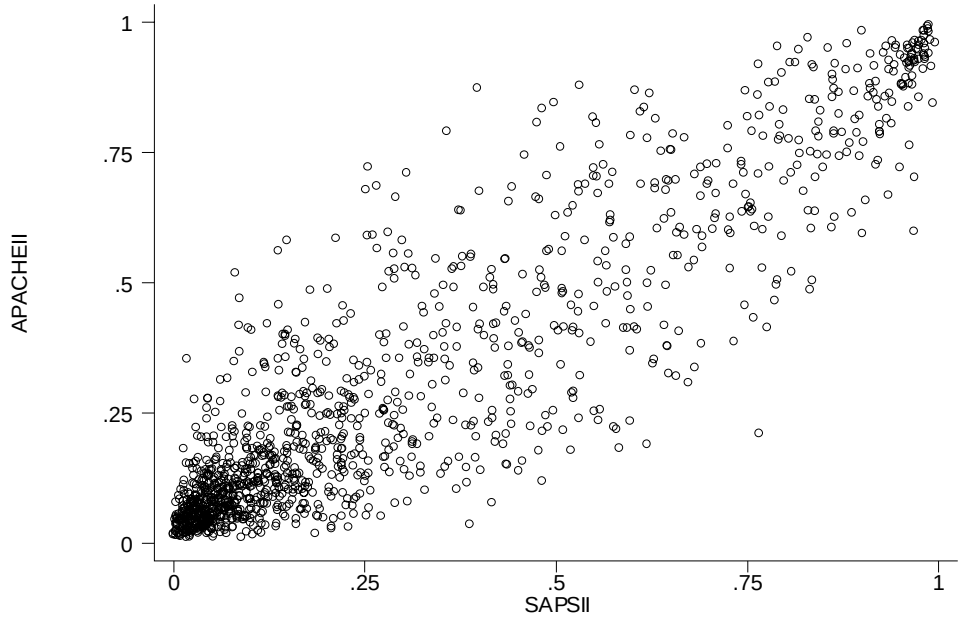
		APACHE II			SAPS II		
	n	AUC*	H chi2	C chi2	AUC*	H chi2	C chi 2
All patients	1316	0.911	66.59**	66.65**	0.888	54.01**	71.44**
Exclude							
- CABG[$]	1261	0.913	66.69**	67.18**	0.891	54.83**	72.35**
- CABG + CAD[$+]	1078	0.915	51.12**	50.86**	0.893	38.23**	55.78**
- CABG + CAD[$+] postoperative cardiac surgery	923	0.904	44.13**	44.5**	0.880	30.77**	47.33**

* AUC; Area Under the Receiver Operating Curve
[$] CABG; Coronary Artery Bypass Graft
[+] CAD; Coronary Artery Disease

the APACHE II and SAPS II models, respectively.

In Thailand, previous reports showed the area under the ROC curve of APACHE II include 0.723[18], 0.788[17] and 0.838[14]. Lertsithichi et al[14] found the area under the ROC curve of SAPS II 0.818 in surgical patients. Limitations of previous reports were relatively small case mix populations and only a single study reported calibration of these scores. Thus, both the APACHE II and SAPS II scoring systems are able to predict group outcome in Thai ICU patients, providing a prognostic estimate that may be helpful in assisting clinical decision making.

However, both models in the present study failed to predict hospital mortality accurately- that is,



Correlation r = p < 0.001

Fig. 2 Correlation of APACHE II and SAPS II predicted hospital mortalities

overall calibration was poor. This lack of overall goodness-of-fit is in similar to the finding of other previous studies[8,10-13,16].

Potential reasons for poor calibration of both scoring systems in the presented population might include the following: 1) differences in data collection and definition; 2) differences in case mix from other studies; 3) lead time bias. The reliability of the data collected is important because poor data can influence the predictions of mortality. Holt et al[23] showed that the main causes of data error in scoring APACHE II are inconsistent choice between highest and lowest value of acute physiologic score and GCS. The variability of GCS determination in sedated patients may affect the predicted death in both models. In the present study, the authors used the pre-sedation GCS in sedated patients as in previous studies[13], an approach which has been shown to be associated with better performance of APACHE II than the approach that is normal GCS for sedated patients[24]. Inclusion criteria of a study can also lead to calibration problems. Coronary care and post-cardiac surgical patients were included in the present study; however, the lack of goodness-of-fit for both models is shown persistently after excluding these populations. The number and type of missing physiological variable may affect the prediction of mortality[25]. In the original APACHE II, physiologic variable were missing in 13% of the cases[4]. Engel et al[26] reported bilirubin, oxygenation, and GCS to be missing in 84%, 34%, and 34% of cases, respectively, in SAPS II database. In the presented records, missing physiological variables were found in only 6% for APACHE II requirements and 6.3% for SAPS II requirement (excluding bilirubin, which was missing in 76.5% of the presented data records).

The potential role of difference in case mix between the presented database and the development database may have had a negative impact of calibration assessment. In general, medical patients have a higher mortality risk than postoperative surgical patients, and in the present study population, medical patients constituted a larger proportion (55.6%) than in the original SAPS II database (48%)[6]. However, when the main diagnostic categories in the presented population are compared with those in the development APACHE II, some interesting differences appear. "Sepsis", diagnostic of "cardiovascular" without one of these principle diagnosis in non-operative patients and "gastrointestinal perforation/obstruction" in postoperative patients which are associated with a high mortality risk account for 16.7%, 3.2% and 3.4% of the presented database, compared with 3.6%, 1.1% and 2.1% in the original APACHE II[4], respectively. On the other hand, the postoperative category of "peripheral vascular surgery", which is associated with a low mortality rate only accounted for 0.8% of our admissions, compared with 9.8% of the development database[4]. These differences could contribute to over-predicted death for the APACHE II model in the presented patients.

Lead time bias is another factor that could affect the accuracy of hospital prediction. Tunnell et al[27] revealed that lead time bias increased the APACHE II and SAPS II scores by 14 and 23 points, respectively, leading to the APACHE II and SAPS II prediction of hospital mortality being increased as much as 42.7% and 33.4%, respectively. In the present study, however, the lead time was no different between survivors and non-survivors, and univariate analysis showed that lead time was not a significant prediction of hospital deaths (odds ratio 1.01, 95%CI = 0.99-1.02 per day, $p = 0.16$). Thus, the authors believed that the influence of lead time bias on calibrations of both models is minimal in the present study.

The present study has some limitations. First, as a single center study there may be bias concerning case mix, quality of ICU care, and ICU policy. Secondly, the relatively small sample size was a relevant limiting factor in performed stratified analysis of calibration of both models. A multicenter study would have the benefit of fewer concerns of case mix and a better sample size.

In conclusion, the present study found two popular severity scoring systems in a Thai ICU to be less accurate than in the original studies. Both models showed poor calibration but excellent discrimination power, although overall the APACHE II performed better than the SAPS II model. A locally customized APACHE II, or a new version of a scoring system such as APACHE III or SAPS III, should also be evaluated in Thai ICUs.

Acknowledgement

The present study was supported through a Faculty of Medicine research grant, Prince of Songkla University.

References

1. Le Gall JR. The use of severity scores in the intensive care unit. Intensive Care Med 2005; 31: 1618-23.
2. Teres D. The value and limits of severity adjusted

mortality for ICU patients. J Crit Care 2004; 19: 257-63.
3. Zimmerman JE, Alzola C, Von Rueden KT. The use of benchmarking to identify top performing critical care units: a preliminary assessment of their policies and practices. J Crit Care 2003; 18: 76-86.
4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13: 818-29.
5. 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.
6. 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.
7. 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.
8. Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. Crit Care Med 1995; 23: 1327-35.
9. Capuzzo M, Valpondi 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.
10. Katsaragakis S, Papadimitropoulos K, Antonakis P, Strergiopoulos S, Konstadoulakis MM, Androulakis G. Comparison of acute physiology and chronic health evaluation II (APACHE II) and simplified acute physiology score II (SAPS II) scoring systems in a single Greek intensive care unit. Crit Care Med 2000; 28: 426-32.
11. Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. Intensive Care Med 1997; 23: 177-86.
12. Nouira S, Belghith M, Elatrous S, Jaafoura M, Ellouzi M, Boujdaria R, et al. Predictive value of severity scoring systems: comparison of four models in Tunisian adult intensive care units. Crit Care Med 1998; 26: 852-9.
13. Arabi Y, Haddad S, Goraj R, Al-Shimemeri A, Al-Malik S. Assessment of performance of four mortality prediction systems in a Saudi Arabian intensive care unit. Crit Care 2002; 6: 166-74.
14. Lertsithichai P, Euanorasetr C. Preliminary evaluation of APACHE II, APACHE III, MPM OII, MPM 24II and SAPS II predictive systems in a surgical ICU. Ramathibodi Med 1997; 20: 32-41.
15. 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: 698-705.
16. Tan IK. APACHE II and SAPS II are poorly calibrated in a Hong Kong intensive care unit. Ann Acad Med Singapore 1998; 27: 318-22.
17. Ratanarat R, Thanakittiwirun M, Vilaichone W, Thongyoo S, Permpikul C. Prediction of mortality by using the standard scoring systems in a medical intensive care unit in Thailand. J Med Assoc Thai 2005; 88: 949-55.
18. Wilairatana P, Noan NS, Chinprasatsak S, Prodeengam K, Kityaporn D, Looareesuwan S. Scoring systems for predicting outcomes of critically ill patients in northeastern Thailand. Southeast Asian J Trop Med Public Health 1995; 26: 66-72.
19. Goldhill DR, Sumner A. Outcome of intensive care patients in a group of British intensive care units. Crit Care Med 1998; 26: 1337-45.
20. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996; 22: 707-10.
21. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making 1984; 4: 137-50.
22. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997; 16: 965-80.
23. Holt AW, Bury LK, Bersten AD, Skowronski GA, Vedig AE. Prospective evaluation of residents and nurses as severity score data collectors. Crit Care Med 1992; 20: 1688-91.
24. Livingston BM, Mackenzie SJ, MacKirdy FN, Howie JC. Should the pre-sedation Glasgow Coma Scale value be used when calculating acute physiology and chronic Health Evaluation scores

for sedated patients? Scottish Intensive Care Society Audit Group. Crit Care Med 2000; 28: 389-94.
25. Afessa B, Keegan MT, Gajic O, Hubmayr RD, Peters SG. The influence of missing components of the Acute Physiology Score of APACHE III on the measurement of ICU performance. Intensive Care Med 2005; 31: 1537-43.
26. Engel JM, Junger A, Bottger S, Benson M, Michel A, Rohrig R, et al. Outcome prediction in a surgical ICU using automatically calculated SAPS II scores. Anaesth Intensive Care 2003; 31: 548-54.
27. Tunnell RD, Millar BW, Smith GB. The effect of lead time bias on severity of illness scoring, mortality prediction and standardised mortality ratio in intensive care - a pilot study. Anaesthesia 1998; 53: 1045-53.

เปรียบเทียบระบบคะแนนความรุนแรง APACHE II และ SAPS II ในการทำนายอัตราตายในโรงพยาบาลของผู้ใหญ่ชาวไทยในหออภิบาลผู้ป่วย

บดินทร์ ขวัญนิมิตร, อลัน กีเตอร์

วัตถุประสงค์*: เพื่อประเมินความสามารถของระบบ APACHE II และ SAPS II ในผู้ป่วยหนักในประเทศไทย*
วัสดุและวิธีการ*: ศึกษาแบบติดตามไปข้างหน้าในช่วง 1 กรกฎาคม พ.ศ. 2547 ถึง 31 ตุลาคม พ.ศ. 2548 ในหออภิบาลผู้ป่วยโรงพยาบาลสงขลานครินทร์ ซึ่งเป็นโรงพยาบาลมหาวิทยาลัยระดับตติยภูมิขนาด 800 เตียง*
ผลการศึกษา*: ผู้ป่วยทั้งหมด 1,316 ราย เสียชีวิตในโรงพยาบาล 310 ราย (ร้อยละ 23.6) ระบบ APACHE II และ SAPS II ทำนายอัตราตายในโรงพยาบาลร้อยละ 30.5 $\pm$ 28.2 และ 30.5 $\pm$ 29.8 ตามลำดับ ระบบคะแนนทั้งสองมีค่าการจำแนกที่ดีมาก โดยระบบคะแนน APACHE II สามารถจำแนกผู้ป่วยรอดและเสียชีวิตได้ดีกว่าระบบ SAPS II (0.911 และ 0.888, $p < 0.001$) อย่างไรก็ตามระบบทั้งสองไม่สามารถทำนายอัตราตายได้เที่ยงตรง Hosmer-Lemeshow goodness-of-fit H และ C statistics เท่ากับ 66.59 และ 66.65 สำหรับระบบ APACHE II ($p < 0.001$) และ 54.01 และ 71.44 สำหรับระบบ SAPS II ($p < 0.001$)*
สรุป*: ระบบ APACHE II จำแนกผู้ป่วยรอดและเสียชีวิตได้ดีกว่าระบบ SAPS II แต่ระบบทั้งสองทำนายอัตราตายไม่เที่ยงตรงโดยทำนายสูงกว่าอัตราตายจริง ควรปรับปรุงหรือหาระบบคะแนนความรุนแรงใหม่ที่เหมาะสมสำหรับผู้ป่วยหนักในประเทศไทย*