Intensive Care Med (2000) 26: 1779–1785
DOI 10.1007/s001340000715

ORIGINAL

M. Capuzzo
V. Valpondi
A. Sgarbi
S. Bortolazzi
V. Pavoni
G. Gilli
G. Candini
G. Gritti
R. Alvisi

Validation of severity scoring systems SAPS II and APACHE II in a single-center population

Received: 20 March 2000
Final revision received: 9 August 2000
Accepted: 12 September 2000
Published online: 19 October 2000

Supported, in part, by a grant from the Italian Ministero dell'Università e della Ricerca Scientifica e Tecnologica.

M. Capuzzo (✉) · V. Valpondi · A. Sgarbi ·
S. Bortolazzi · R. Alvisi
Department of Surgical,
Anesthetic and Radiological Sciences,
Section of Anesthesiology and Intensive
Care, University Hospital of Ferrara,
Corso Giovecca 203, 44100 Ferrara, Italy
E-mail: sar@dns.unife.it
Phone: +39-0532-236306
Fax: +39-0532-247160

V. Pavoni · G. Gritti
Division of Anesthesia and Intensive Care,
University Hospital of Padova, Italy

G. Gilli · G. Candini
Department of Health Physics,
University Hospital of Ferrara, Italy

Abstract *Objective*: To validate two severity scoring systems, the Simplified Acute Physiology Score (SAPS II) and Acute Physiology and Chronic Health Evaluation (APACHE II), in a single-center ICU population.
Design and setting: Prospective data collection in a two four-bed multidisciplinary ICUs of a teaching hospital.
Patients and methods: Data were collected in ICU over 4 years on 1721 consecutively admitted patients (aged 18 years or older, no transferrals, ICU stay at least 24 h) regarding SAPS II, APACHE II, predicted hospital mortality, and survival upon hospital discharge.
Results: At the predicted risk of 0.5, sensitivity was 39.4 % for SAPS II and 31.6 % for APACHE II, specificity 95.6 % and 97.2 %, and correct classification rate 85.6 % and 85.5 %, respectively. The area under the ROC curve was higher than 0.8 for both models. The goodness-of-fit statistic showed no significant difference between observed and predicted hospital mortality ($H = 7.62$ for SAPS II, $H = 3.87$ for APACHE II; and $C = 9.32$ and $C = 5.05$, respectively). Observed hospital mortality of patients with risk of death higher than 60 % was overpredicted by SAPS II and underpredicted by APACHE II. The observed hospital mortality was significantly higher than that predicted by the models in medical patients and in those admitted from the ward.
Conclusions: This study validates both SAPS II and APACHE II scores in an ICU population comprised mainly of surgical patients. The type of ICU admission and the location in the hospital before ICU admission influence the predictive ability of the models.

Key words Intensive care unit · Severity of illness index · Mortality prediction · Hospital mortality · Simplified Acute Physiology Score · Acute Physiology and Chronic Health Evaluation

Introduction

In comparing the effects of treatment and organization on patient outcome physicians require scoring systems with which to measure the severity of patient illness and predict hospital mortality in large intensive care unit (ICU) populations of a general patient mix. Once these measures are demonstrated accurately to predict hospital mortality, they should become powerful tools for comparing the performance of different ICUs [1]. The Simplified Acute Physiology Score (SAPS II) [2] and the Acute Physiology and Chronic Health Evaluation (APACHE II) [3] are severity scoring systems which have been the subject of extensive research. The keystones in analyzing the accuracy of hospital mortality predictions of a model are discrimination and calibration. The former refers to how well the model distinguishes between patients who will die or survive, and

Table 1 Study group by year of ICU admission

	1994	1995	1996	1997	p	Total
No. of admissions	434	500	488	490		1912
No. of readmissions	17 (3.9 %)	27 (5.4 %)	22 (4.5 %)	31 (6.3 %)	0.213	97 (5.1 %)
No excluded	28 (6.5 %)	31 (5.6 %)	18 (3.7 %)	17 (3.5 %)	0.213	94 (4.9 %)
Exclusion due to						
Age (< 18 years)	8	5	1	3		17
From other ICU	5	1	0	3		9
ICU length of stay < 24 h	15	25	17	11		68
No patients included	389	442	448	442	0.238	1721
Type of admission						
Scheduled surgical	208 (53.5 %)	230 (52.0 %)	256 (57.2 %)	244 (55.2 %)		938 (54.5 %)
Urgent surgical	97 (24.9 %)	128 (29.0 %)	114 (25.4 %)	122 (27.6 %)	0.474	461 (26.8 %)
Medical	84 (21.6 %)	84 (19.0 %)	78 (17.4 %)	76 (17.2 %)		322 (18.7 %)
Mean ICU length of stay (days)	6.1 ± 11.9	5.5 ± 8.8	6.1 ± 13.4	5.6 ± 7.6	0.731	5.8 ± 10.7
Mean postICU length of stay (days)	16.0 ± 21.2	14.4 ± 14.0	13.0 ± 18.6	13.7 ± 15.7	0.254	14.2 ± 16.5

p values refer to the variables in the 4 years considered (χ^2 or analysis of variance)

the latter to how closely predictions are correlated with actual outcome across the entire range of risk [4]. Good discrimination has been demonstrated for SAPS II in multicenter studies performed in Italy [5], Portugal [6], Austria [7], as well as in the European group of ICUs EURICUS-I [8]. Also, APACHE II has shown good discriminative ability in Portugal [6] and the United Kingdom [9]. However, most of these studies [5, 6, 7, 8] report calibration as disappointing.

The poor calibration recorded in multicenter studies has been hypothesized as due to one or more of the following factors: statistics (models, sample size), patients (case-mix), and ICU organization (management, lead-time bias, data collection, hospital discharge policy) [10]. The only way in which to study the performance of the model is to isolate it from the other variables. A study carried out in a single-center multidisciplinary ICU should avoid the effect of the ICU-specific variables and allow investigation only of statistics and patients as factors affecting the performance of the model. The aim of the present study was to analyze the predictive accuracy of SAPS II and APACHE II in a population of patients admitted to a single center ICU.

Patients and methods

This study was performed in two four-bed mixed (surgical and medical) ICUs of a 960-bed teaching hospital. In the hospital there are two additional adult ICUs (a ten-bed mixed ICU and a six-bed coronary care unit), no cardiac surgery, no burn units, and no intermediate or step-down beds. The ICUs in which the study was performed serve all thoracic, vascular, and high-risk abdominal surgery patients and about one-half of the medical ward patients of the hospital. The two ICUs have one medical director who makes a daily round in both units and a full-time specialist in intensive care each. A resident physician provides night coverage, with a specialist in anesthesia and intensive care on call. The nurse to patient ratio is 1:2. All ICU admissions were included from 1 January 1994 to 31 December 1997. Patients readmitted during the study period were considered only at the time of their first admission. The patients aged under 18 years and those transferred from other ICUs or staying in the ICU less than 24 h were excluded. During the study period there were 1912 admissions (Table 1). The number of patients included in the study, the type of ICU admission and the ICU and hospital (post-ICU) lengths of stay did not significantly differ over the years considered. The final number of patients included in the study was 1721, 68 % of whom ($n = 1173$) were men. Patients' mean age was 68.3 ± 13.4 years.

The day after ICU admission the worst values on APACHE II and SAPS II variables (worst measurements observed during 24 h following ICU admission) were abstracted from clinical and laboratory records, using the variable definitions reported in the literature for SAPS II [2] and APACHE II [3, 11]. APACHE II, SAPS II, and hospital mortality as predicted by both indices were calculated according to the published techniques [2, 3]. In cases in which data were missing, the score on that variable was considered as normal [3]. Information collected from the forms, SAPS II, APACHE II, computed probability of hospital mortality according to both indices and survival upon hospital discharge of each patient were recorded in a database.

Interobserver reliability was tested in a sample of 81 cases selected from the database, reviewed by "new observers" [12].

Data are presented here as mean ± 1 SD, when indicated. Statistical analysis was carried out using a software package (SPSS version 8.0) and p values less than 0.05 were considered significant. Student's t test and analysis of variance were used for normally distributed continuous variables and the χ^2 statistic for categorical data.

The accuracy of outcome prediction according to SAPS II and APACHE II was assessed using the methods described below. The 2 × 2 decision matrices at the predicted risk of 0.5 was used to compare predicted and observed outcomes, recording true positive (sensitivity), true negative (specificity), and correct classification rate, as percentages [13]. The area under the receiver operating characteristic (ROC) curve [14] was measured to test discrimination. The Lemeshow-Hosmer goodness-of-fit H and C statistics [15] were used to evaluate calibration. The calibration curve was constructed by plotting observed death rate against predicted death rate stratified by 10% risk ranges; linear regression analysis was used and the r^2 value calculated [16].

To evaluate the uniformity of fit in different subgroups we stratified patients by age group, type of ICU admission, and location in the hospital before ICU admission. Discrimination and calibration of the two scores were analyzed in each subgroup. The standardized mortality ratios (SMR) with 95% confidence intervals were calculated [13], and the differences between observed and predicted numbers of hospital deaths were analyzed using the χ^2 test corrected for continuity.

Results

Overall, 18% of patients ($n = 307$) died in the hospital. SAPS II predicted 345.4 hospital deaths (20.1%) and APACHE II 312.5 (18.2%). The SMRs were 0.89 (CI 95% 0.85–0.94) and 0.98 (CI 95% 0.94–1.03), respectively. Considering interobserver reliability, there was no statistically significant difference in probability of hospital mortality calculated using values of SAPS II and APACHE II components reabstracted by different observers, as described in detail elsewhere [12]. At the predicted risk of 0.5, sensitivity was 39.4%, specificity 95.6%, and correct classification rate 85.6% for SAPS II, and 31.6%, 97.2%, and 85.5%, respectively, for APACHE II. The ROC curve is reported in Fig. 1; the area under the ROC curve was similar for SAPS II (0.816) and APACHE II (0.805). The Lemeshow-Hosmer goodness-of-fit *H* and *C* statistics for SAPS II and APACHE II are reported in Tables 2 and 3, respectively. The calibration curve is plotted in Fig. 2. The r^2 value was 0.97 for SAPS II and 0.99 for APACHE II.

The observed hospital mortality rate in patients at risk of death higher than 60% was lower than that predicted by SAPS II and higher than that predicted by APACHE II. When a decision criterion of 60% was applied, out of 58 patients predicted to die by both SAPS II and APACHE II, 49 died and 9 survived. For the 94 patients for whom nonconcordant outcome was predicted by the two scores, the correct classification rate was 45.7% for SAPS II and 54.3% for APACHE II. In patients stratified according to age groups, type of ICU admission and location in the hospital before ICU admission (Table 4), discrimination was generally good in all subgroups (area under the ROC curve > 0.7). Considering the type of ICU admission, the *C* statistic showed a bad fit for both SAPS II and APACHE II in medical cases, and the number of observed deaths was lower that that predicted by SAPS II in scheduled surgical cases. Considering the location in the hospital before ICU admission, both models underestimated mortality in the patients admitted from the ward, and SAPS II overestimated mortality in the patients admitted from operative room or emergency department. The observed hospital mortality was lower than that predicted by SAPS II in the patients admitted from the operative room and emergency department.

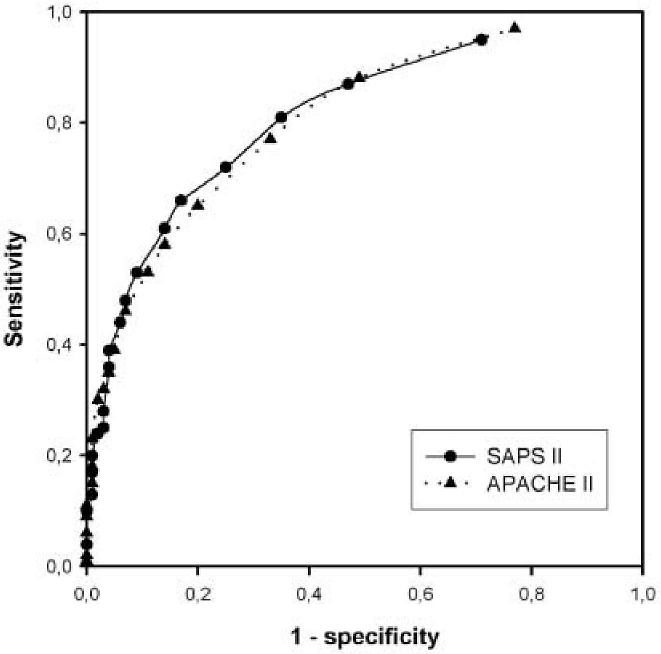

Fig. 1 ROC curve for SAPS II and APACHE II

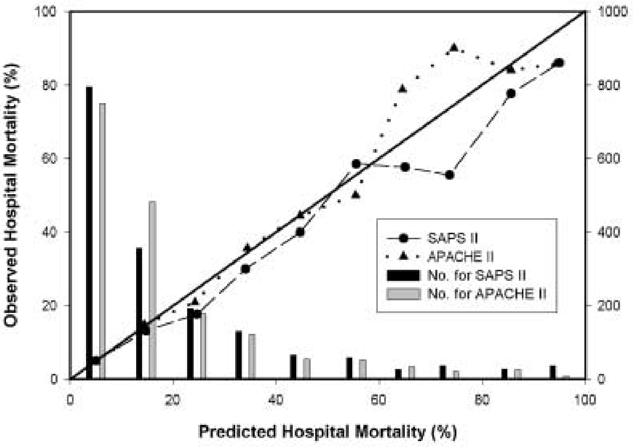

Fig. 2 Calibration curve for SAPS II and APACHE II

Discussion

This study analyzed the predictive accuracy of SAPS II and APACHE II in a group of patients admitted to a single-center ICU over a 4-year period. During this the period the only change at the institutional level was a reduction in the number of hospital beds from 1100 to 960, without changes in ICU admission or discharge policy. The number of ICU medical and nursing staff remained

Table 2 Lemeshow-Hosmer goodness-of-fit *H* and *C* statistics for SAPS II

Risk of death	No. of ICU admissions	Observed survivors	Predicted survivors	Observed deaths	Predicted deaths
H statistic[a]					
0.0–0.1	795	755	755.3	40	39.7
0.1–0.2	356	309	303.7	47	52.3
0.2–0.3	192	158	144.7	34	47.3
0.3–0.4	130	91	85.8	39	44.2
0.4–0.5	65	39	36.0	26	29.0
0.5–0.6	58	24	25.8	34	32.2
0.6–0.7	26	11	9.1	15	16.9
0.7–0.8	36	16	9.5	20	26.5
0.8–0.9	27	6	3.9	21	23.1
0.9–1.0	36	5	1.8	31	34.2
Total	1721	1414	1375.6	307	345.4
C statistic[b]					
0.00–0.02	173	168	170.0	5	3.0
0.02–0.04	172	165	165.9	7	6.1
0.04–0.05	172	164	163.2	8	8.8
0.05–0.07	172	160	159.9	12	12.1
0.07–0.11	172	159	155.1	13	16.9
0.11–0.16	172	148	148.3	24	23.7
0.16–0.22	172	145	138.9	27	33.1
0.22–0.30	172	140	126.0	32	46.0
0.30–0.50	172	108	103.6	64	68.4
0.50–1.00	172	57	44.7	115	127.3
Total	1721	1414	1375.6	307	345.4

[a] $\chi^2 = 7.62$, degrees of freedom 10, $p > 0.5$
[b] $\chi^2 = 9.32$, degrees of freedom = 10, $p > 0.25$

stable. Also, the patient population did not change, as demonstrated by the lack of significant differences in the number of admissions, readmissions, excluded cases, and type of ICU admission (Table 1). Therefore the group of patients studied appears to be homogeneous for the ICUs considered, even if collected over a long period. It has been claimed that the severity systems are not expected to show good calibration over time due to medical progress [17]. Nevertheless, overall good calibration has been reported for APACHE II 10 years after publication [16].

Multicenter studies provide a large sample size within a short period, but the participating ICUs sometimes collect only a few cases each. The mean number of patients given by each unit varies widely between different studies: 14 [5], 52 [6], 113 [8], 193 [7], and 338 [9]. When many individual ICUs collect only a few cases each, the overall sample can accumulate the variability of individual patients and that of individual ICU organization. On the other hand, considering that 75 % of the ICUs involved in the European Prevalence of Infection in Intensive Care study had less than ten beds [18], difficulties in collecting a large sample in a single ICU and in a short time are evident.

The reliability of the data collected is important because it has been suggested that small, consistent differences in scores cause potentially important changes in the predicted mortality [19]. It has been shown that the main causes of data errors in computing APS of APACHE II are the choice between highest or lowest value as worst and the Glasgow Coma Score evaluation [20]. Nevertheless, variations in individual prediction are not reflected in collective predictions, as reported by Chen et al. for APACHE II [21]. Our own review [12] draws similar conclusions and demonstrate the reliability of our data collection.

The mean age of our study group was higher than in other studies [5, 6, 7, 8, 9]. Moreover, the percentage of medical admissions (18.7 %) was lower than that of the original samples used to develop SAPS II (48 %) [2] and APACHE II (47 %) [3]. This feature could have been responsible for a bias. Considering discrimination, the comparison of our results with those of other independent studies at the predicted risk of 0.5 shows the highest specificity for SAPS II (95.6 % vs. 92.4 [8], 87.9 [6], and 84 [22]) and APACHE II (97.2 vs. values ranging from 94.8 [23] to 81 [22]). The same is true for the correct classification rate, which was 85 % for both scores. The area under the ROC curve was higher than 0.8, as reported in studies on SAPS II [5, 7, 8, 24] and APACHE II [9, 24, 25]. Considering calibration, the goodness-of-fit statistics demonstrate that the hospital mortality observed in our patients did not differ significantly from that predicted by both models.

The calibration in our sample appears to be better than reported elsewhere, in studies devoted to SAPS II

Table 3 Lemeshow-Hosmer goodness-of-fit *H* and *C* statistics for APACHE II

Risk of death	No of ICU admissions	Observed survivors	Predicted survivors	Observed deaths	Predicted deaths
H statistic[a]					
0.0–0.1	750	715	708.6	35	41.4
0.1–0.2	482	410	412.5	72	69.5
0.2–0.3	177	140	134.0	37	42.9
0.3–0.4	121	79	79.5	42	41.6
0.4–0.5	54	30	29.9	24	24.1
0.5–0.6	52	26	23.3	26	28.8
0.6–0.7	33	7	11.7	26	21.3
0.7–0.8	20	2	5.0	18	14.9
0.8–0.9	25	4	3.6	21	21.4
0.9–1.0	7	1	0.4	6	6.6
Total	1721	1414	1408.5	307	312.5
C statistic[b]					
0.00–0.03	173	171	168.8	2	4.2
0.03–0.05	172	165	164.9	7	7.1
0.05–0.06	172	162	161.5	10	10.5
0.06–0.09	172	162	158.2	10	13.8
0.09–0.11	172	149	154.3	23	17.7
0.11–0.15	172	151	149.1	21	22.9
0.15–0.19	172	142	142.7	30	29.3
0.19–0.27	172	139	132.8	33	39.2
0.27–0.42	172	113	113.5	59	58.5
0.42–1.00	172	60	62.7	112	109.3
Total	1721	1414	1408.5	307	312.5

[a] $\chi^2 = 3.87$, degrees of freedom = 10, $p > 0.9$
[b] $\chi^2 = 5.05$, degrees of freedom10, $p > 0.9$

Table 4 Uniformity of fit of SAPS II and APACHE II. Stratification by age groups, year of ICU admission, type of ICU admission and location before ICU admission (*OR/ED* operative room/emergency department, *SMR* standardized hospital mortality ratio, *CI* confidence interval)

	n	SAPS II				APACHE II			
		ROC	*H*	*C*	SMR (95% CI)	ROC	*H*	C	SMR (95% CI)
Age groups (years)									
≤55	219	0.783	7.77	17.17	1.00 (0.84–1.24)	0.844	1.94	3.49	0.93 (0.80–1.09)
56–65	320	0.827	4.48	8.35	1.10 (0.96–1.27)	0.819	8.65	10.74	1.15 (1.02–1.32)
66–75	696	0.837	8.45	10.62	0.85 (0.78–0.92)	0.837	5.30	8.66	0.94 (0.87–1.01)
≥76	486	0.791	6.81	7.80	0.82 (0.76–0.89)	0.768	2.43	8.22	0.97 (0.90–1.06)
Type of ICU admission									
Scheduled surgical	938	0.721	15.57	16.32	0.68 (0.63–0.73)***	0.717	6.37	11.97	0.77 (0.73–0.82)
Urgent surgical	461	0.820	8.61	12.15	0.85 (0.79–0.91)	0.778	1.27	6.96	0.95 (0.89–1.02)
Medical	322	0.771	15.74	42.89*	1.17 (1.07–1.30)	0.750	18.05	24.67**	1.22 (1.13–1.34)
Location before ICU									
OR/ED	1434	0.838	17.74	19.55***	0.79 (0.74–0.84)**	0.813	5.23	7.89	0.92 (0.88–0.98)
Ward	287	0.710	30.90*	60.57*	1.22 (1.11–1.35)	0.718	17.66	21.10***	1.13 (1.05–1.24)

*$p < 0.001$, **$p < 0.01$, ***$p < 0.05$

[6, 7, 8, 24] and APACHE II [6, 24, 25, 26] using both *C* and *H* Lemeshow-Hosmer statistics. Unfortunately, the low number of patients in our higher risk groups could have introduced a bias. Zhu et al. [27] showed that the goodness-of-fit statistic increases as hospital outcome diverges increasingly from that observed in the data used from the original model development, although this tendency seems to be diminished when the sample size is small. On the other hand, in very large samples of ICU patients, even if the hospital mortality pattern differs by only one or two percentage points from the original, the models may demonstrate poor fit. According to this simulation study [27], the largest databases on SAPS II [8] and APACHE II [26], considering

10,027 and 8,724 cases, respectively, showed the highest goodness-of-fit values. Moreover, two studies showing poor calibration, and performed on 1733 [7] and 1325 patients [24], reported an observed hospital mortality differing by 9.5 % [7] and 5.1 % [24] from that predicted by SAPS II and by 3.8 % [24] from that predicted by APACHE II.

The results about calibration from the study performed by Moreno et al. [6] in 982 patients appear to be more intriguing because the general sample size seems adequate, and the difference between observed and predicted hospital mortality is low (0.6 % for SAPS II and 1.5 % for APACHE II). Nevertheless, the goodness-of-fit demonstrated a poor calibration [6]. The data analyzed in that study [6] were collected in 19 Portuguese ICUs, and differences among ICUs were quite large, with individual ICU SMR values ranging from 0.69 to 1.72 for SAPS II and from 0.42 to 1.63 for APACHE II. The authors concluded that their results do not allow the use of SAPS II and APACHE II to analyze quality of care, but no one can exclude that the difference in ICUs was responsible for the poor calibration. The studies in which a model (APACHE II) showed a good calibration curve were performed in two Canadian university teaching hospital ICUs of the same town [16], collecting 1724 patients, and in one Hong Kong teaching hospital ICU [28], on 1573 patients. In these studies observed hospital mortality differed by 0.1 % [16] and 2 % [28], respectively, from the predicted one. Unfortunately, the authors applied linear regression analysis to the calibration curve instead of the Lemeshow-Hosmer goodness-of-fit statistics. Their r^2 values were 0.99 [16] and 0.98 [28], which are similar to those found in the present study (0.97 for SAPS II and 0.99 for APACHE II).

Ideally, models should be well calibrated in all subgroups of patients represented in an ICU before they can be safely applied for risk adjustment. When the patients were stratified according to type of ICU admission and location in the hospital before ICU admission, the observed hospital mortality was significantly higher than that predicted by either model in medical admissions and in patients admitted from the ward, as reported by others [28, 29]. Moreover, the calibration was reported to be good in studies in which the proportion of surgical cases was 60 % [28] or 50 % [16] and poor in those studies in which the proportion of surgical cases was 35 % or lower [22, 23, 24]. From a clinical point of view it is not surprising that the medical patients admitted first to the ward and then showing such deterioration as to need intensive care have outcomes which are poorer than those admitted from operative room or emergency department.

Our results are quite different from those obtained by others in the same country [5, 30]. In our study the calibration curve showed a good fit, especially in patients with predicted risk of death lower than 60 %. The lack of fit reported by others was associated with an overall underprediction of mortality, especially in patients with predicted risk of death lower than 50 % [5] or between 30 % and 60 % [30]. One study including 99 ICUs from all parts of Italy and collecting a total sample of 1393 patients during 1 month [5] attributed the lack of fit to the great variability in unmeasured case-mix across ICUs and to the model. In the other, considering 6794 patients from 24 ICUs in northern Italy [30], the lack of fit was suggested to be due to lead-time bias [31] and length of stay in ICU. Nevertheless, the bad fit could be due to the high number of medical admissions in both studies (56.2 % [5] and 43.6 % [30]). Moreover, in one study the most frequent location of patients in the hospital before ICU admission was ward (62.8 % [30]) and in the other 29.3 % of cases were admitted from main hospital and 7.8 % from other ICUs [5].

In conclusion, this study validates both SAPS II and APACHE II scores in an ICU population composed mainly of surgical patients. Generally, discrimination and calibration of the models considered appeared to be good, even if the calibration curve showed hospital mortality rate to be overpredicted by SAPS II and underpredicted by APACHE II in patients with predicted risk of death higher than 60 %. The type of ICU admission and the location in the hospital before ICU admission influenced the predictive ability of the models.

References

1. Consensus Conference Organised by the ESICM and the SRLF (1994) Predicting outcome in ICU patients. Intensive Care Med 20: 390–397
2. Le Gall JR, Lemeshow S, Saulnier F (1993) A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 270: 2957–2963
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) APACHE II: a severity of disease classification system. Crit Care Med 13: 818–829
4. Schuster DP (1992) Predicting outcome after ICU admission. The art and science of assessing risk. Chest 102: 1861–1870
5. Apolone G, Bertolini G, D'Amico R, Iapichino G, Cattaneo A, De Salvo G, Melotti RM (1996) The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: results from GiViTI. Intensive Care Med 22: 1368–1378
6. Moreno R, Morais P (1997) Outcome prediction in Intensive care: results of a prospective, multicentre, Portuguese study. Intensive Care Med 23: 177–186

7. Metnitz PGH, Valentin A, Vesely H, Alberti C, Lang K, Steltzer H, Hiesmayr (1999) Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Intensive Care Med 25: 192–197
8. Moreno R, Reis Miranda D, Fidler V, Van Schilfgaarde R (1998) Evaluation of two outcome prediction models on an independent database. Crit Care Med 26: 50–61
9. Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP (1993) Intensive Care Society's APACHE II study in Britain and Ireland. II. outcome comparison of intensive care units after adjustment for case mix by the American APACHE II method. BMJ 307: 977–981
10. Randolph AG, Guyatt GH, Carlet J, for the Evidence Based Medicine in Critical Care Group (1998) Understanding articles comparing outcomes among intensive care units to rate quality of care Crit Care Med 26: 773–781
11. Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP (1993) Intensive Care Society's APACHE II study in Britain and Ireland. I. variation in case mix of adult admission to general intensive care units and impact on outcome. BMJ 307: 972–977
12. Capuzzo M, Pavoni V, Valpondi V, Paparella L, Sgarbi A, Cingolani E, Alvisi R, Gritti G (1998) Reliability of the general severity scoring systems, APACHE II and SAPS II. Clin Intensive Care 9: 4–10
13. Wassertheil-Smoller S (1990) Biostatistics and epidemiology. Springer, Berlin Heidelberg New York
14. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29–36
15. Lemeshow S, Hosmer DW (1982) A review of goodness of fit statistics for use in the development of logistic regression models Am J Epidemiol 115: 92–106
16. Wong DT, Crofts SL, Gomez M, McGuire GP, Byrick RJ (1995) Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients. Crit Care Med 23: 1177–1183
17. Teres D, Lemeshow S (1999) When to customize a severity model. Intensive Care Med 25: 140-142
18. Vincent JL, Suter P, Bihari D, Bruining H (1997) Organization of intensive care units in Europe: lessons from the EPIC study. Intensive Care Med 23: 1181–1184
19. Goldhill DR, Withington PS (1996) Mortality predicted by APACHE II. Anaesthesia 51: 719–723
20. Holt AW, Bury KL, Bersten AD, Skowronski GA, Vedig AE (1992) Prospective evaluation of residents and nurses as severity score data collectors. Crit Care Med 20: 1688–1691
21. Chen LM, Martin CM, Morrison TL, Sibbald WJ (1999) Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. Crit Care Med 27: 1999–2004
22. Patel PA, Grant BJB (1999) Application of mortality prediction systems to individual intensive care unit. Intensive Care Med 25: 977–982
23. Beck DH, Taylor BL, Millar B, Smith GB (1997) 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 25: 9–15
24. Nouira S, Belghith M, Elatrous S, Jaafoura M, Ellouzi M, Boujdaria R, Gahbiche M, Bouchoucha S, Abroug F (1998) Predictive value of severity scoring systems: comparison of four models in Tunisian adult intensive care units. Crit Care Med 26: 852–859
25. Castella X, Artigas A, Bion J, Aarno K, European/North American Severity Study Group (1995) A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. Crit Care Med 23: 1327–1335
26. Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP (1994) Intensive Care Society's Acute Physiology and Chronic Health Evaluation (APACHE II) study in Britain and Ireland: a prospective, multicenter, cohort study comparing two methods for predicting outcome for intensive care patients. Crit Care Med 22: 1392–1401
27. Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D (1996) Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. Crit Care Med 24: 57–63
28. Oh TE, Hutchinson R, Short S, Buckley T, Lin E, Leung D (1993) Verification of the Acute Physiology and Chronic Health Evaluation scoring system in a Hong Kong intensive care unit. Crit Care Med 21: 698–705
29. Moreno R, Apolone G, Reis Miranda D (1998) Evaluation of the uniformity of fit of general outcome prediction models. Intensive Care Med 24: 40–47
30. Sicignano A, Giudici D, and behalf of ARCHIDIA (2000) Customization of SAPS II for the assessment of severity in Italian ICU patients. Minerva Anesthesiol 66: 139–145
31. Dragsted L, Jorgensen J, Jensen NH, Bonsing E, Jacobsen E, Knaus WA, Qvist J (1989) Interhospital comparisons of patient outcome from intensive care: importance of lead-time bias. Crit Care Med 17: 418–422