Nephrol Dial Transplant (2011) 26: 3894–3901
doi: 10.1093/ndt/gfr201
Advance Access publication 19 April 2011

Performance of the third-generation models of severity scoring systems (APACHE IV, SAPS 3 and MPM-III) in acute kidney injury critically ill patients

Verônica Torres Costa e Silva[1], Isac de Castro[1], Fernando Liaño[2], Alfonso Muriel[3], José R. Rodríguez-Palomares[2] and Luis Yu[1]

[1]Division of Nephrology, University of São Paulo School of Medicine, São Paulo, Brasil, [2]Division of Nephrology, Hospital Universitario Ramón y Cajal, IRYCIS, Madrid, Spain and [3]Clinical Biostatistics, Hospital Universitario Ramón y Cajal, IRYCIS, Madrid, Spain

Correspondence and offprint requests to: Luis Yu; E-mail: luisyu@usp.br

Abstract

Background. Severity scores are useful to guarantee similar disease severity among groups in clinical trials and to enable comparison between different studies. The aim of this study was to assess the performance of the third generation models of severity scoring systems [simplified acute physiology score (SAPS) 3, acute physiology and chronic health evaluation (APACHE) IV and mortality probability model (MPM)-III] in acute kidney injury (AKI) patients in the intensive care unit (ICU).

Methods. Three hundred and sixty-six consecutive AKI critically ill patients were prospectively assessed in six ICUs of an academic tertiary care center. Scores were applied on AKI diagnosis day (DD) and on the day of nephrology consultation (NCD). Discrimination was assessed by area under the receiver operating characteristic curve (AUCROC) and calibration by Hosmer–Lemeshow (HL) goodness-of-fit test.

Results. Hospital mortality rate was 67.8%. SAPS 3 general and Central and South America (CSA) customized equations presented identical good discrimination (AUCROC curve: 0.80 on NCD) and satisfactory HL tests on both analyzed days ($P > 0.100$). CSA SAPS 3 equation predicted mortality more accurately [standardized mortality ratio (SMR) on NCD = 1.00 (95% confidence interval (CI) 0.84–1.34)]. APACHE IV and MPM-III scores presented similar discrimination compared to SAPS 3 on both analyzed days ($P > 0.05$). APACHE IV presented satisfactory HL tests over time ($P > 0.100$) but underestimated mortality [SMR on DD = 1.92 (95% CI 1.61–2.23); SMR on NCD = 1.46 (95% CI 1.48–1.96)]. MPM-III showed unsatisfactory HL test results ($P = 0.027$ on DD; $P = 0.045$ on NCD) and underestimated mortality [SMR on NCD = 2.09 (95% CI 1.48–1.96)].

Conclusions. SAPS 3, especially the geographical customized equation, presented good discrimination and calibration performances, accurately predicting mortality in this group of AKI critically ill patients.

Keywords: acute kidney injury; intensive care unit; prognostic factors; severity scoring systems; third generation models

Introduction

One of the main problems concerning the design of clinical trials in critically ill acute kidney injury (AKI) patients is the lack of validated, well-established scoring systems to stratify the severity of patient disease states and guarantee adequate randomization within a particular study design [1, 2]. It remains unsettled, which are the best models (general or specific scores) for AKI patients and the most appropriate moment for scores application. Also, most studies reported thus far have been limited by small sample sizes and the absence of a uniform AKI definition. Patients in different stages of AKI and diverse severity have been compared, contributing to the discrepancies in the performance of these models. Nevertheless, we have previously demonstrated that simplified acute physiology score (SAPS) II and Stuivenberg Hospital Acute Renal Failure scores presented the best performance in critically ill AKI patients. [3–9].

The RIFLE system [10], now validated in >71 000 patients worldwide, included patients in the early phases of AKI and provided a simple and universal AKI definition, allowing comparisons among studies [11–14]. The RIFLE classification of severity staging also proved to be associated with mortality, becoming the most powerful prognostic AKI stratification system validated so far [14]. However, there are still scant data concerning the assessment of illness severity scores in AKI patients using the RIFLE system.

General intensive care unit (ICU) models usually underestimate the mortality of AKI patients. Most studies have assessed the second generation scores [acute physiology and chronic health evaluation (APACHE) II [15] and APACHE III [16], SAPS II [17], Sepsis-related organ failure assessment [18], Logistic organ dysfunction system [19]], which were developed in the mid 1990s [20]. The recent third generation of ICU scoring systems [SAPS 3 [21], APACHE IV [22] and mortality probability model (MPM) III [23]] are powerful and updated models. Although they were assessed in several groups of ICU patients [24–30], these models have never been evaluated in non-dialysis AKI patients.

The aim of this study was to evaluate the performance of the third generation of severity scoring systems (APACHE IV, SAPS 3 and MPM-III) in a group of critically ill patients with AKI defined according to the RIFLE system criteria.

Patients and methods

Study participants

A prospective observational study was conducted through an active search for AKI cases by daily visits to six ICUs comprising 53 beds, in the Hospital das Clinicas of University of Sao Paulo School of Medicine, Brazil. This is a tertiary academic hospital with 13 ICUs and 128 available beds for critically ill patients; only general clinical and surgical ICUs were chosen. All specific ICUs were excluded, such as cardiac surgery, coronary, bone marrow, solid organ transplantation and pediatric ICUs.

All patients admitted to the selected ICUs were evaluated for renal function in the period between November 2003 and June 2005. AKI was defined as an increase of ≥50% of the baseline serum creatinine (SCr) measurement according to the R (risk) level criteria of the RIFLE system. Baseline SCr was defined as the lowest value obtained during the hospital stay or within 30 days before diagnosis. The exclusion criteria included baseline SCr ≥3.0 mg/dL, previous dialysis, age <18 years, kidney transplantation, an ICU stay shorter than 48 h, urinary tract obstruction and hypovolemia responsive to fluids. Vital signs and hemodynamic and laboratory data were recorded on AKI diagnosis day (DD) and on the day of nephrology consultation (NCD). Severity scores (APACHE IV, SAPS 3 and MPM III) were calculated longitudinally considering the worst value of physiologic variables measured every 24 h. Scoring elements are described in Table 1. A complete description of inclusion and exclusion criteria, data elements, data collection, organ failure and related parameters and management strategies have been previously detailed [9]. Data were collected by an independent single observer, non-member of the ICU or nephrology staff. Nephrology consultation (NC) was solicited by the ICU physician. The primary outcome was in-hospital mortality. The study was approved by the local Ethics Committee and informed consent was not required.

Statistical analysis

All 366 patients were sequentially evaluated. Continuous variables were expressed as mean ± SD or as median with 25th and 75th quartiles as appropriate. Categorical variables were expressed as proportions and analyzed with Pearson's χ^2 test for independent groups. Logistic regression was employed to determine the adjusted odds of in-hospital mortality. Two models were built using variables of each analyzed day. Candidate variables were those with a likelihood ratio significance <0.05 upon bivariate analysis. Multivariable logistic regression models were constructed with backwards variable selection, using a P-value <0.05 for variable retention. The colinearity of the maximal models was evaluated using the criteria proposed by Belsley [31]. Discrimination was assessed using the area under the receiver operating characteristic curve (AUCROC) [32]. Calibration was assessed using the Hosmer–Lemeshow (HL) goodness-of-fit test comparing observed versus expected mortality across deciles of risk [33]. A high P-value (>0.05) indicated a good fit of the model. The standardized mortality ratio (SMR) with respective 95% confidence intervals (CIs) was calculated for each model by dividing the observed by the predicted mortality rate. AUROCs were compared using nonparametric statistics [34]. A two-tailed P-value <0.05 was

Table 1. Scoring elements for APACHE IV, SAPS 3 and MPM-III[a]

APACHE IV	SAPS 3	MPM-III
Age	Age	Age
Heart rate	Heart rate	Heart rate
Mean arterial pressure	Lowest systolic BP	Systolic BP
Mechanical ventilation	Ventilation support/oxygenation	Mechanical ventilation
Glasgow Coma Scale	Glasgow Coma Scale	Coma/stupor (GCS 3–4)
Creatinine and BUN	Creatinine	Chronic renal failure
Urine output	Chronic heart failure	Acute renal failure
Hepatic failure	Cirrhosis	Cirrhosis
Various malignancies, AIDS	Various malignancies, AIDS	Metastatic neoplasm
Emergency surgery	Unplanned/planned admit	Medical/unscheduled surgical
Bilirubin	Bilirubin	
Temperature	Temperature	
Serum pH/PcO2	Lowest pH	
Respiratory rate	Use of vasoactive drugs	CPR before admission
Oxygenation (AaDO2 or PaO2)	Surgical status/anatomic site	Age interaction terms
Hematocrit	Thrombocytopenia	GI bleeding
White cell count	White cell count	Cerebrovascular incident
Sodium, albumin, glucose	Presence of infection	Absence of other risk factors
Admitting diagnosis	Reason for ICU admission	Cardiac dysrhythmias
Pre-ICU location and LOS	Pre-ICU location and LOS	
Comorbidities	Comorbidities	

[a]BP, blood pressure; GCS, Glasgow Coma Scale; AIDS, acquired immunodeficiency syndrome; BUN, blood urea nitrogen; PcO2, capillary oxygen pressure; CPR, cardiopulmonary resuscitation; AaDO2, alveolar-arterial oxygen gradient; GI, gastrointestinal; LOS, length of stay. This table is a simplified version and exact definition may vary. Detailed definitions and notes regarding proper application of each score are depicted in original references.

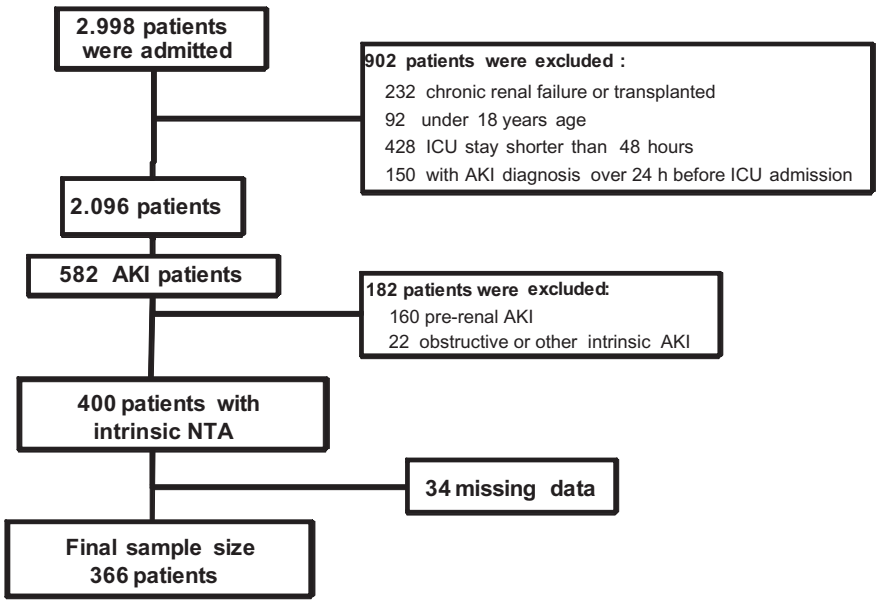

Fig. 1. Study population.

considered significant. Statistical analysis was carried out using SPSS for Windows version 18.0 (Chicago, IL) and SAS (SAS Institute, Cary, NC).

Results

A total of 2998 patients were admitted to the selected ICUs during the study period. The final sample size comprised 366 AKI patients (Figure 1). Mean age was 57.1 ± 18.8 and main AKI-related factors were sepsis (67%) and surgery (22.1%). Patients' origin before ICU admission were emergency room (40.4%), ward (32.8%) and operating room (23.5%). Median hospital length of stay (LOS) before ICU admission was 4.0 (1.0–11) days and AKI diagnosis occurred within the first 2 days after ICU admission in 64% of patients. One hundred and twelve patients (30.6%) required dialysis therapy and overall hospital mortality was 67.8% (Table 2). NC occurred 3 days (1.0–4.0) after DD. Main physiological and laboratory variables on both days are presented in Table 3.

Table 2. Baseline and clinical patients' characteristics[a]

Variables	
Age	57.1 ± 18.8
Men	216 (59%)
Charlson comorbidity index	4 (2–6)
Main comorbidities	
CKD (Stage III or above)	87 (23.7%)
Hypertension	152 (41.5%)
Diabetes mellitus	85 (23.2%)
Heart failure	69 (18.9%)
Solid tumor	85 (23.2%)
AIDS	24 (6.5%)
Chronic liver disease	20 (5.4%)
Type of ICU admission	
Medical	280 (76.5%)
Emergency surgical	54 (14.8%)
Elective surgical	32 (8.7%)
Pre ICU patient origin	
ER	148 (40.4%)
Ward	120 (32.8%)
Operating room	86 (23.5%)
AKI related factors	
Sepsis	244 (66.7%)
Surgery	81 (22.1%)
Bleeding	58 (15.9%)
Cardiogenic shock	50 (13.7%)
Time-related variables	
Pre-ICU LOS (days)	4.0 (1.0–11)
Pre-ICU LOS in ER origin patients (days)	1.0 (0–2.0)
ICU stay before AKI diagnosis (days)	1.0 (1.0–5.0)
Early AKI diagnosis	234 (64%)
ICU stay before RRT start (days)	4.0 (2.0–07)
Outcome variables	
NC	196 (53.5%)
Dialysis need	112 (30.6%)
ICU LOS (days)	13 (8.0–23)
ICU mortality	229 (62.6%)
Hospital LOS (days)	17 (10–36)
Hospital mortality	248 (67.8%)

[a]CKD, chronic kidney disease; AIDS, acquired immunodeficiency syndrome; ER, emergency room; NC, Nephrology consultation; LOS, length of stay. Early AKI (AKI diagnosis within 48 h after ICU admission). Results expressed as mean ± SD, median (25th–75th) and *n* (%).

Table 4 presents the performance of severity scoring systems on DD and on NCD.

SAPS 3 general and Central and South America (CSA) customized equations had a similar discrimination (AUCROC range from 0.73 on DD to 0.80 on NCD) and presented satisfactory HL tests on both analyzed days ($P > 0.10$). SAPS 3 CSA customized equation predicted mortality more accurately [SMR on DD = 1.09 (95% CI 0.84–1.34); SMR on NCD = 1.00 (95% CI 0.84–1.34)] than SAPS 3 general equation (GEq) [SMR on DD = 1.35 (95% CI 1.07–1.63); SMR on NCD = 1.15 (95% CI 0.75–1.55)]. APACHE IV and MPM-III presented similar discrimination as compared to SAPS 3 on both analyzed days ($P > 0.05$). APACHE IV presented satisfactory HL tests over time ($P > 0.10$) but underestimated mortality [SMR on DD = 1.92 (95% CI 1.61–2.23); SMR on NCD = 1.46 (95% CI 1.13–1.79)]. MPM-III showed unsatisfactory HL test results ($P = 0.027$ on DD; $P = 0.045$ on NCD) and fairly underestimated mortality [SMR on DD = 1.89 (95% CI 1.60–2.18); SMR on NCD = 2.09 (95% CI 1.69–2.49)]. The AUCROC of third generation models on NC day are depicted in Figure 2.

Additional prognostic factors were determined by multivariate analysis. Advanced age, lower urine output, longer LOS in the ICU (before AKI diagnosis) and central nervous system (CNS) failure were included in the logistic regression models for mortality on both days of analysis (Table 5). On DD, low albumin and low SCr concentrations, cardiovascular and liver failures were also related to higher mortality. In addition, the AUCROC for death was 0.84. On NCD, the following variables were also related to increased mortality: higher lactate values, respiratory and liver failure. Mortality model discrimination was good, with an AUCROC of 0.88, higher than those obtained on the preceding day ($P < 0.05$). Models presented good sustained calibration over time ($P = 0.50$ and $P = 0.85$, on DD and NCD, respectively).

Discussion

The third generation models of ICU prognostic systems are more complex than their previous counterparts. They have been developed based on larger databases and built using more complex statistical modeling techniques. ICU admission causes were expanded and refined and new important prognostic factors like patient origin, infection site and hospital LOS were included [35, 36].

SAPS 3 score, the last version of the SAPS system published in 2005, was the largest prospective multinational study conducted so far, with enrollment of 19 577 patients in 307 ICUs from 35 countries of the five continents between October and December 2002. The main differences compared with SAPS II are data collection on ±1 h from ICU admission and derivation of seven customized equations for different geographic regions.

In this study, the SAPS 3 GEq and customized CSA equation generated nearly the same prediction despite being mathematically different, resulting in identical observed AUCROC on both analyzed days. The SAPS 3 calibration HL tests (for GEq and CSA equations) were satisfactory on both assessed days and SAPS 3 customized equation for CSA countries presented the best SMR, reflecting the impact of important geographical differences such as patients' living styles and presence of comorbidities and health care systems.

SAPS 3 was the most frequently assessed score among the third generation models with several studies demonstrating good discrimination (AUCROC > 0.80) [24, 26,28–30]. Calibration difficulties were more frequently observed [24, 26, 28]. However, three prospective studies, including the Maccariello *et al.* [8], in which they have assessed a group of 244 AKI dialysis patients, presented similar results: adequate calibration and accurate prediction from CSA-customized equation [29, 30]. In our study, SAPS 3 discrimination was good on NCD [(AUCROC: 0.80 (95% CI 0.73–0.86)], while it was regular on DD [AUCROC: 0.73 (95% CI 0.67–0.79)]. Physiological variables were collected in a 24 h window, which may have influenced model discrimination. Indeed, AKI may not be detected within the short time frame of 1 h. Limiting data collection to the first hour of ICU admission was proposed to avoid the influence of the ICU-delivered care.

Table 3. Patient characteristics on the DD of AKI and NCD[a]

Parameter	Variable	DD ($N = 366$)	NCD ($N = 196$)
Physiological variables	% Mechanical ventilation	73.8	84.7
	Heart rate (per minute)	100 ± 19	98 ± 17
	Systolic BP (mmHg)	120 (108–133)	126 (112–141)
	Mean arterial BP (mmHg)	85 (77–93)	89 (79–98)
	Temperature (°C)	36.5 (36.1–37)	36.5 (36.1–37)
	Urine output (mL/24 h)	1005 (523–1665)	575 (200–1195)
	Respiratory rate	18 (14–23)	19 (15–22)
	Glasgow Coma Scale	8 (6–14)	7 (5–14)
Laboratory variables	Creatinine (mg/dL)	1.9 (1.5–2,6)	3.2 (2.3–4.3)
	Urea (mg/dL)	74 (50–108)	119 (83–165)
	pH	7.32 (7.24–7.39)	7.29 (7.19–7.36)
	Bicarbonate (mEq/L)	17 (14–20)	15 (13–18)
	PO2	90.0 (74.3–111.2)	90.0 (76.8–112.4)
	PO2/FiO2	242 (177–302)	250 (184–339)
	PCO2	32.1 (26.2–39.8)	32.5 (26.6–40)
	Glucose (mg/dL)[b]	135 (108–174)	137 (117–167)
	Sodium (mEq/L)	140 (136–144)	141 (136–146)
	Leukocyte (1000/mm^3)	13,3 (9.0–19.0)	13,3 (8.6–20.9)
	Platelets (1000/mm^3)	168 (101–274)	146 (85–251)
	Hematocrit (%)	29 (26–34)	28 (23–32)
	Total bilirubin (mg/dL)	0.8 (0.4–2.0)	0.9 (0.5–2.5)
	Albumin (g/dL)	2.1 (1.8–2.6)	2.0 (1.7–2.6)

[a]BP, blood pressure.
[b]Capillary glycemia.

Table 4. AUCROCs and HL goodness-of-fit statistics for severity scores on DD of AKI and NCD

Prognostic score	ROC curve		Goodness-of-fit *C*-test		Predicted mortality[a] (mean ± SD)	SMR (95% CI)
	AUC	CI 95%	χ^2	P-value		
DD ($N = 366$)						
APACHE IV	0.74	(0.69–0.79)	6.65	0.574	35.3 ± 21.8	1.92 (1.61–2.23)
SAPS3 (GEq)	0.73	(0.67–0.78)	6.86	0.551	50.1 ± 24.5	1.35 (1.07–1.63)
SAPS3 (CSA)	0.73	(0.67–0.78)	6.33	0.610	61.9 ± 26.0	1.09 (0.83–1.35)
MPM-III	0.73	(0.67–0.78)	17.28	0.027	35.7 ± 23.9	1.89 (1.60–2.18)
NCD ($N = 196$)						
APACHE IV	0.79	(0.74–0.85)	12.86	0.117	46.5 ± 28.0	1.46 (1.13–1.79)
SAPS3 (GEq)	0.80	(0.73–0.86)	10.47	0.163	58.6 ± 23.4	1.15 (0.75–1.55)
SAPS3 (CSA)	0.80	(0.73–0.86)	13.22	0.113	67.1.± 23.8	1.00 (0.61–1.39)
MPM-III	0.81	(0.73–0.88)	15.79	0.045	32.3 ± 23.7	2.09 (1.69–2.49)

[a]The observed mortality was 67.8%.

Nevertheless, in centers with shortage of ICU beds, under treatment during long pre-ICUs stays (our median time was 4.0 days) could actually overestimate mortality in the ICU admission, which could be partially corrected over the following 24 h with appropriate medical care. It is not clear how this change could influence the model's accuracy. Another aspect that may have influenced score's discrimination is the SAPS 3 kidney dysfunction evaluation, which is based exclusively on proposed SCr values (SCr mg/dL stratification: <1.2, 1.2–2.0, 2.1–3.4, ≥ 3.5).

APACHE IV, the last version of APACHE score system, published in 2006, was prospectively developed in 131 988 patients admitted to 104 ICUs in USA, providing predictions of hospital mortality and ICU LOS. APACHE IV has never been evaluated for AKI patients. In our study, this model presented similar discrimination and satisfactory HL tests as compared to SAPS 3. The satisfactory accuracy of APACHE IV is probably explained by the large number of physiological variables, including serum albumin level, a more refined CNS impairment assessment and multiple use of SCr, serum urea level (Sur) and diuresis for kidney function evaluation. Also, the large number of ICU diagnosis admissions (116 detailed options which accounted for 16% of model explanatory power) improved disease identification and calibration. In our study, APACHE IV underestimated mortality, probably due to important differences from the original database: higher mortality rate (13.5 versus 67.8 %), sepsis etiology (5.4 versus 66.7 %) and longer pre-ICU LOS (0.78 versus 4.0 days). Despite commonly shared features, APACHE IV and SAPS 3 have some differences: origin (American versus multinational nature), acute physiologic variable explanatory power (66 versus 25%, respectively), data collection window (24 versus 1 h) and kidney dysfunction parameters. Since these

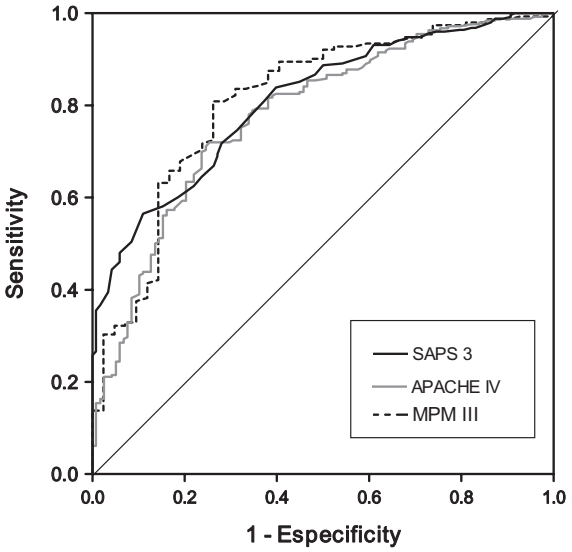

Fig. 2. AUROCs for SAPS 3, APACHE IV and MPM-III on NCD.

Table 5. Predictors of mortality using logistic regression on DD of AKI and NCD[a]

Parameter	Coefficient β	OR	95% CI
DD A1			
Age (per 10 years)	0.365	1.44	1.23–1.67
Creatinine (mg/dL)	−0.230	0.79	0.63–0.99
Urine output (per 100 mL/day)	−0.029	0.97	0.94–0.99
ICU LOS (per 1 day)	0.044	1.04	1.00–1.08
Albumin (g/dL)	−0.491	0.61	0.38–0.96
Cardiovascular failure	1.981	7.24	3.87–13.54
CNS failure	0.794	2.21	1.25–3.91
Liver failure	0.947	2.57	1.47–4.50
NCD A2*			
Age (per 10 years)	0.404	1.49	1.27–1.75
Lactate (mmoL/dL)	0.039	1.04	1.00–1.07
Urine output (per 100 mL/day)	−0.053	0.94	0.92–0.97
ICU LOS (per 1 day)	0.047	1.04	1.01–1.08
CNS failure	1.194	3.29	1.68–6.47
Respiratory failure	1.307	3.69	1.77–7.70
Liver failure	1.088	2.96	1.63–5.39

[a]OR, adjusted odds ratio; A1, area under ROC curve = 0.84 (0.79–0.88), HL $\chi^2 = 0.50$; A2, area under ROC curve = 0.88 (0.83–0.91), HL $\chi^2 = 0.85$.
*$P < 0.05$ versus DD.

two models have never been compared in a large sample of AKI ICU patients, it remains to be defined which score would be more adequate for these patients.

MPM-III presented the worst performance among the third generation models with satisfactory discrimination but inadequate calibration, and the observed mortality was double compared with the predicted mortality. MPM-III model presented satisfactory performance in a retrospective study of 11 300 general ICU patients [25]. However, two recent prospective studies reported mortality underestimation: Soares *et al.* [30] have assessed a group of 717 critically ill cancer patients with an SMR: 3.42 (95% CI 2.63–4.41) and Maccariello *et al.* [8] found a SMR: 2.42 (95% CI 1.95–3.01) assessing dialysis AKI patients.

MPM-III score was developed in a large sample of 124 855 patients admitted to 135 ICUs, mostly in the USA. However, the model has several limitations, including retrospective data collection nature and inclusion of a limited number of variables: age, three physiologic parameters (coma, heart rate and systolic blood pressure), five acute and three chronic diagnosis and a few other parameters (cardiopulmonary resuscitation, mechanical ventilation and admission type). ICU entry diagnosis was not included and AKI was defined only by SCr values (>2.0 mg/dL). In addition, data collection window is also 1 h, leading to the same methodological problems for AKI patients as discussed for SAPS 3.

Logistic models provided further data on outcome prognosis. Low serum albumin levels have been associated with worse prognosis in both general critically ill and AKI patients, probably reflecting the intense inflammatory and hypercatabolic AKI status [37]. However, only APACHE IV includes albumin levels among the assessed variables which should be evaluated in future models. Increased lactate level has been reported as a single prognostic factor in ICU patients, presenting similar discrimination compared with most prognostic scores in the BEST study [20]. Interestingly, no severity scoring system has included lactate level among assessed parameters. Increased ICU LOS before AKI diagnosis has been implicated as a very important prognostic factor in recent studies, usually related to more serious events, such as infection [38]. Association between lower SCr and increased mortality was previously described, which could be explained by factors such as diminished muscle mass, malnutrition and fluid overload [39–41]. This is an important observation since all ICU scores stratify AKI severity according to specific criteria, usually attributing higher severity to increased SCr levels. Currently, there is no scoring system that applies SCr stratification as proposed by the RIFLE or AKIN systems [42]. Furthermore, some models utilize Sur (or blood urea nitrogen), which is more susceptible to several influences.

The strengths of this study rely on the prospective nature, with an active search for new AKI cases in a large sample of critically ill patients. For the first time, the performance of the new third generation scores was assessed in a sample of non-dialysis AKI patients using the RIFLE system criteria. Use of a less severe AKI definition allowed evaluation from early stages of the disease, improving model performance. In addition, only 8.5% of the sample was excluded, minimizing the risk of selection and analysis bias. Additionally, data were collected by one single investigator, minimizing the interobserver effect. Despite the presence of the investigator, she had no contact with ICU physicians on duty and simply collected the data. Thus, there was no influence on the ICU staff decision for NC.

Although almost four hundred patients were included in this study, severity score validation is better assessed in larger samples [43]. The main limitation refers to the modification of SAPS 3 and MPM-III original methodology. Considering this study as the first report with non-dialysis AKI patients, we could not compare our results. Also, general ICU models were developed to be applied on ICU admission day and observed mortality underestimation might have been influenced by the latter application [44].

Furthermore, we could not exclude a possible casemix influence as well as local differences in the delivered care, ICU infrastructure and patient referral [45]. The long pre-ICU LOS was an important factor, which may have influenced models performance. Finally, we had a homogeneous population, which may be difficult to reproduce elsewhere.

Conclusions

In this prospective study, the customized equation of SAPS 3 from CSA countries was the most accurate scoring system among the third generation models for prediction of hospital mortality in AKI critically ill patients. APACHE IV score presented satisfactory performance but underestimated mortality. These results represent an important step forward for the validation of prognostic models, which are essential for the development of clinical trials with AKI critically ill patients.

Acknowledgements. The authors wish to thank Dr Andrew A. Kramer and Dr Marcio Soares for expert analysis and revision of the manuscript.

Conflict of interest statement. None declared.

References

1. Bellomo R. The epidemiology of acute renal failure: 1975 versus 2005. *Curr Opin Crit Care* 2006; 12: 557–560
2. Uchino S. Outcome prediction for patients with acute kidney injury. *Nephron Clin Pract* 2008; 109: c217–c223
3. Douma CE, Redekop WK, van der Meulen JH *et al.* Predicting mortality in intensive care patients with acute renal failure treated with dialysis. *J Am Soc Nephrol* 1997; 8: 111–117
4. Fiaccadori E, Maggiore U, Lombardi M *et al.* Predicting patient outcome from acute renal failure comparing three general severity of illness scoring systems. *Kidney Int* 2000; 58: 283–292
5. Lins RL, Elseviers M, Daelemans R *et al.* Prognostic value of a new scoring system for hospital mortality in acute renal failure. *Clin Nephrol* 2000; 53: 10–17
6. Mehta RL, Pascual MT, Gruta CG *et al.* Refining predictive models in critically ill patients with acute renal failure. *J Am Soc Nephrol* 2002; 13: 1350–1357
7. Chertow GM, Soroko SH, Paganini EP *et al.* Mortality after acute renal failure: models for prognostic stratification and risk adjustment. *Kidney Int* 2006; 70: 1120–1126
8. Maccariello E, Valente C, Nogueira L *et al.* SAPS 3 scores at the start of renal replacement therapy predict mortality in critically ill patients with acute kidney injury. *Kidney Int* 2010; 77: 51–56
9. Costa e Silva VT, de Castro I, Liano F *et al.* Sequential evaluation of prognostic models in the early diagnosis of acute kidney injury in the intensive care unit. *Kidney Int* 2009; 75: 982–986
10. Bellomo R, Ronco C, Kellum JA *et al.* Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group. *Crit Care* 2004; 8: R204–R212
11. Hoste EA, Clermont G, Kersten A *et al.* RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. *Crit Care* 2006; 10: R73
12. Uchino S, Bellomo R, Goldsmith D *et al.* An assessment of the RIFLE criteria for acute renal failure in hospitalized patients. *Crit Care Med* 2006; 34: 1913–1917
13. Ostermann M, Chang RW. Acute kidney injury in the intensive care unit according to RIFLE. *Crit Care Med* 2007; 35: 1837–1843
14. Ricci Z, Cruz D, Ronco C. The RIFLE criteria and mortality in acute kidney injury: a systematic review. *Kidney Int* 2008; 73: 538–546
15. Knaus WA, Draper EA, Wagner DP *et al.* APACHE II:. a severity of disease classification system. *Crit Care Med* 1985; 13: 818–829
16. Knaus WA, Wagner DP, Draper EA *et al.* The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100: 1619–1636
17. 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: 2957–2963
18. de Mendonca A, Vincent JL, Suter PM *et al.* Acute renal failure in the ICU: risk factors and outcome evaluated by the SOFA score. *Intensive Care Med* 2000; 26: 915–921
19. Le Gall JR, Klar J, Lemeshow S *et al.* The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. *JAMA* 1996; 276: 802–810
20. Uchino S, Bellomo R, Morimatsu H *et al.* External validation of severity scoring systems for acute renal failure using a multinational database. *Crit Care Med* 2005; 33: 1961–1967
21. Moreno RP, Metnitz PG, Almeida E *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: 1345–1355
22. Zimmerman JE, Kramer AA, McNair DS *et al.* Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34: 1297–1310
23. Higgins TL, Teres D, Copes WS *et al.* Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). *Crit Care Med* 2007; 35: 827–835
24. Khwannimit B, Bhurayanontachai R. The performance and customization of SAPS 3 admission score in a Thai medical intensive care unit. *Intensive Care Med* 2010; 36: 342–346
25. Kuzniewicz MW, Vasilevskis EE, Lane R *et al.* Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. *Chest* 2008; 133: 1319–1327
26. Ledoux D, Canivet JL, Preiser JC *et al.* SAPS 3 admission score: an external validation in a general intensive care population. *Intensive Care Med* 2008; 34: 1873–1877
27. Metnitz B, Schaden E, Moreno R *et al.* Austrian validation and customization of the SAPS 3 Admission Score. *Intensive Care Med* 2009; 35: 616–622
28. Poole D, Rossi C, Anghileri A *et al.* External validation of the Simplified Acute Physiology Score (SAPS) 3 in a cohort of 28,357 patients from 147 Italian intensive care units. *Intensive Care Med* 2009; 35: 1916–1924
29. Soares M, Salluh JI. Validation of the SAPS 3 admission prognostic model in patients with cancer in need of intensive care. *Intensive Care Med* 2006; 32: 1839–1844
30. Soares M, Silva UV, Teles JM *et al.* Validation of four prognostic scores in patients with cancer admitted to Brazilian intensive care units: results from a prospective multicenter study. *Intensive Care Med* 2010; 36: 1188–1195
31. Belsley DA. *Conditioning Diagnostics: Colinearity and Weak Data in Regression*. New York, NY: Wiley-Interscience, 1991
32. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29–36
33. Hosmer DW, Hosmer T, Le CS *et al.* A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16: 965–980
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44: 837–845
35. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. *Crit Care Med* 2010; 38: 1–7
36. Moreno RP. Outcome prediction in intensive care: why we need to reinvent the wheel. *Curr Opin Crit Care* 2008; 14: 483–484
37. Obialo CI, Okonofua EC, Nzerue MC *et al.* Role of hypoalbuminemia and hypocholesterolemia as copredictors of mortality in acute renal failure. *Kidney Int* 1999; 56: 1058–1063
38. Guerin C, Girard R, Selli JM *et al.* Initial versus delayed acute renal failure in the intensive care unit. A multicenter prospective epidemiological study. Rhone-Alpes Area Study Group on Acute Renal Failure. *Am J Respir Crit Care Med* 2000; 161: 872–879

39. Cerda J, Cerda M, Kilcullen P *et al.* In severe acute kidney injury, a higher serum creatinine is paradoxically associated with better patient survival. *Nephrol Dial Transplant* 2007; 22: 2781–2784
40. Paganini EP, Halstenberg WK, Goormastic M. Risk modeling in acute renal failure requiring dialysis: the introduction of a new model. *Clin Nephrol* 1996; 46: 206–211
41. Paganini EP, Larive B, Kanagasundaram NS. Severity scores and outcomes with acute renal failure in the ICU setting. *Contrib Nephrol* 2001; 181–195
42. Mehta RL, Kellum JA, Shah SV *et al.* Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. *Crit Care* 2007; 11: R31
43. Vergouwe Y, Steyerberg EW, Eijkemans MJ *et al.* Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005; 58: 475–483
44. Vincent JL, Opal SM, Marshall JC. Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. *Crit Care Med* 2010; 38: 283–287
45. Metnitz PG, Lang T, Vesely H *et al.* Ratios of observed to expected mortality are affected by differences in case mix and quality of care. *Intensive Care Med* 2000; 26: 1466–1472

Received for publication: 17.12.10; Accepted in revised form: 22.3.11