Intensive Care Decision Making: Using Prognostic Models for Resource Allocation

Alireza ATASHI[a], Masoumeh SARBAZ[b], Sina MARASHI[a] , Fatemeh HAJIALIASGARI[a] and Saeid ESLAMI[c]

[a] *Assistant Professor, E-health Department, Virtual school, Tehran University of Medical Sciences, Tehran, Iran*

[b] *Ph.D. in Medical Informatics, Assistant Professor, Department of Medical Records and Health Information Technology, School of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran*

[c] *Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*

Abstract. Accurate outcome prediction by the means of available clinical contributing factors will support researchers and administrators in realistic planning, workload determination, resource optimization, and evidence-based quality control process. This study is aimed to evaluate APACHE II and SAPS II prediction models in an Iranian population. A a prospective cross-sectional study was conducted in four tertiary care referral centers located in the top two most populated cities in Iran, from August 2013 to August 2015. The Brier score, Area under the Receiver Operating Characteristics Curve (AUC), and Hosmer-Lemeshow (H-L) goodness-of-fit test were employed to quantify models' performance. A total of 1799 patients (58.5% males and 41.5% females) were included for further score calculation. The overall observed mortality (24.4%) was more than international rates due to APACHE II categories. The Brier score for APACHE II and SAPS II were 0.17 and 0.196, respectively. Both scoring systems were associated with acceptable AUCs (APACHE II = 0.745 and SAPS II = 0.751). However, none of prediction models were fitted to dataset (H-L ρ value < 0.01). With regards to poor performance measures of APACHE II and SAPS II in this study, finding recalibrated version of current prediction models is considered as an obligatory research question before applying it as a clinical prioritization or quality control instrument.

Keywords. Intensive Care Unit, Prediction Models, Performance Measures, Iran.

1. Introduction

Although, as a part of modern medicine, Intensive Care Units (ICUs) have been customized as particular units aiming to provide specific health care services to a particular group of patients who share a common acute disorder, but the continuous time-limited decision making process remains as a significant challenging issue in this area. Regarding vulnerability and rapid fluctuations of vital organs clinical decision making should be accompanied by accurate prioritization (1). It should be also noted that ICUs contribute to a growing proportion of health care expenditures which in turn include internal and external mechanical equipment (2).

Utilizing scoring systems like the Acute Physiology and Chronic Health Evaluation (APACHE) may help to provide a clinical standard for severity prioritization by the means of routine blood investigations. Various contributing factors such as age, duration of acute disorder, special medical consideration e.g. malignancy, immunosuppression or the need for kidney transplant and emergency ICU admission increase the mortality rate in ICU (3).

APACHE II incorporates 14 variables (The most deviating symptoms and laboratory results from normal definitions during the first 24-hour period post-admission), each of which is scored from 0 to 4 and results in an ordinal total score ranging from 0 to 71 in which higher score reflects more severity of acute disorder (4). It was first developed to predict an individual's mortality risk in ICU, however numerous studies evaluated this score as a patient triaging tool (5). This may highlight the fact that APACHE II score may be utilized as a quality control instrument (6).Similar existing scoring systems such as Simplified Acute Physiology Score II (SAPS II) have been rarely evaluated within various countries around the world, but a few studies have confirmed the acceptable predictive power of SAPS II in Europe and North America (7).

Regarding the limited number of patients in previous studies around the country, a multi-center prospective study was conducted in the top two most populated cities in Iran to evaluate the predictive power and to provide performance-related statistics for APACHE II and SAPS II scoring systems.

2. Methods

A prospective cross-sectional study was conducted to collect a pre-specified set of variables in four centers located in Tehran, capital Iran(67% of patients) and Mashhad, northeast Iran(33% of patients) as the top two most populated cities in country, from August 2013 to August 2015. Patients who were admitted due to traumatic surgeries, burn patients, patients underwent cardiac surgery or psychological disorders were excluded with regards to the nature of diagnoses. In addition, any use of psychotropic agents in medication profile or symptoms of dysarthria or paramnesia due to a type of brain disorder were considered as exclusion criteria.

A total of fourteen variables in APACHE II in addition with remaining variables requested by SAPS II were designed as a structured paper form to be filled out for consecutive 1799 adult(>16 yrs.) patients. The highest APACHE II score for each particular patient during the first 24-hour period post-admission was considered as the final score. Regarding predetermined personnel cooperation framework, minimal missing values were included in this study (less than 0.2%) which were excluded. Using online calculators (available at: http://clincalc.com/Error.aspx) APACHE II and SAPS II scores were calculated for each particular patient by two of authors. The Brier score (overall performance), Area under the Receiver Operating Characteristic Curve (AUC) and Hosmer-Lemeshow (H-L) goodness-of-fit test were considered as performance indicators for both models. Analyses were performed using medcalc-13.3.3.0 and R-3.3.1 (Resource Selection package).

3. Results

A total of 1053 (58.5%) males and 746 (41.5%) females were included in this study, 834 patients (46.3%) were post-surgical, N=230 (12.8%) of patients were diabetic and N=859 (47.7%) were post-surgical admitted patients. The overall mortality rate was 24.4% (N=439) and the mean APACHE II score for all patients was 10.8

(±6.129). About 67.4% (N=1213) of patients were associated with APACHE II lower than 15.

Mean APACHE II score for living and dead outcomes were 11.4 and 16.7, respectively (ρ value < 0.01). As expected, mortality rate and APACHE II score were increased similarly. Also, total population was associated with 20 (± 11.43) SAPS II score (Table 1).

Table 1. Comparison of Observed Mortality Rates in ICUs with International Standards regarding APACHE II Score

APACHE II score	Total N (%)	Observed Mortality N (%)	International Standard (%)	ρ value a
APACHE II ≤ 15	1213 (67.4%)	187 (15.4%)	10%	ρ < 0.01 [b]
16 < APACHE II < 19	335 (18.6%)	118 (35.2%)	15%	ρ < 0.01 [b]
20 < APACHE II <30	251 (13.9%)	134 (53.3%)	35%	ρ < 0.01 [b]

[a] Comparison was performed using chi square test.
[b] Observed mortality rate more than international standards.

While APACHE II was associated with better overall performance (Brier score=0.17), SAPS II performed a more acceptable discrimination of alive and dead cases (AUC=0.751). This is while, both scoring systems revealed unsuccessful calibration (H-L ρ value < 0.01) (Table 2 and Figure 1).

Table 2. Performance measures calculated for APACHE II and SAPS II scoring systems.

Scoring System	Overall Performance	Discrimination					Calibration
	Brier Score(min-max) [STD]	Mean	SE	95% CI	Difference	ρ value	H-L Test
APACHE II	0.17(0-0.94) [0.25]	10.745	0.0133	(0.725-0.765)	0.00608	0.4694	Chi2(8) = 98.588, ρ<0.01
SAPS II	0.196(0-0.999) [0.35]	0.751	0.0132	(0.731-0.771)			Chi2(8) = 1608.9, ρ<0.01

AUC: Area Under the ROC Curve, SE: Standard Error, CI: Confidence Interval, H-L: Hosmer-Lemeshow.

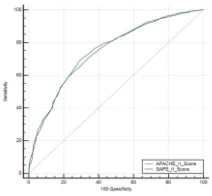

Figure 1. Area under the ROC curve for APACHE II and SAPS II.

4. Discussion

Recruiting patients from four tertiary care referral centers may increase the generalizability of results to a large subset of our target population in Iran. While both of APACHE II and SAPS II scoring systems performed relatively admissible classifications of alive and dead predicted probabilities estimated by APACHE II were closer to observed events (Brier score = 0.17). The H-L goodness-of-fit test revealed poor calibrations for both models (ρ value < 0.01). It is worth noting that observed ICU mortality rates were significantly higher than internationally published standards.

Correct outcome prediction in 75% of cases is similar to accuracy measures reported by Gupta et al. for Indian population (8). Also, an observational study in

Rasoul Akram hospital, Tehran, Iran revealed that observed mortality rate for low-risk patients (APACHE II $\leq$ 15) was comparable to international standards. However, mortality rate for the rest of patients was significantly higher than reported standards which may be due to variability of provided treatments in the center. This may highlight the fact that APACHE II score may be utilized as a quality control instrument (6). Safavi et al. proved that APACHE II was the most accurate prediction tool (sensitivity=90%, specificity=32%, and accuracy=81%) in compare with Infection Probability Control (IPC) and APACHE III to estimate the overall ICU mortality rate (5).

A brief comparison of AUCs may indicate the fact that discriminative ability of APACHE II in Iran is relatively lower than those published in similar studies around the world. The aforementioned issue may be addressed by model recalibration approach which may provide us more accurate outcome predictions in research, practice and policy making (esp. benchmarking) areas. Integration of different clinical prediction models for benchmarking purposes will support researchers and administrators to step forward in severity prioritization, ICU bed allocation scheme, and evidence-based distribution of intensive care capacities.

Prospective data collection approach, minimal missing values, recruiting acceptable number of patients for evaluation purposes, and representativeness of our sample due to geographic situation and annual number of ICU admissions in four included hospitals may be noted as strengths for this study. Regardless of the diagnosis at the time of admission, all patients were included aiming to assess the performance of APACHE II within patients involving with various organ malfunctions. Although, a 2-year sampling duration will adjust the effect of time-related confounders and may guarantee the inclusion of probable seasonal disorders, but time and sample-related limitations remains as an inevitable issue.

With regards to poor performance measures of APACHE II and SAPS II in our included sample, recalibration of current prediction models is considered as an obligatory research question before applying it as a clinical prioritization or quality control instrument.

References

[1] M. Nikravanfard, H. Sheri. *Nursing fundamentals in ICU, CCU and dialyze*. Tehran Nordanesh, Tehran, 2006.

[2] R.D. Truog, D.W. Brock, D.J. Cook et al, Rationing in the intensive care unit, *Critical Care Medicine* **34** (2006), 958-963.

[3] C. Bouch, J. Thompson, *Severity scoring systems in the critically ill patients*, Oxford University Press. 2008.

[4] W.A. Knaus, E.A. Draper, D.P. Wagner et al, APACHE II: a severity of disease classification system, *Critical Care Medicine* **13** (1985), 818-829.

[5] M. Safavi, A. Honarmand, Comparison of infection probability score, APACHE II, and APACHE III scoring systems in predicting need for ventilator and ventilation duration in critically ill patients, *Archives of Iranian medicine* **10** (2007), 354-360.

[6] P. Rahimzadeh, Z. Taghipor, V. Hasani, Estimated mortality rates using the APACHE II scoring system in ICU patients surgery Rasoul Akram Hospital in Tehran, *Iranian Journal of Critical Care Nursing* **11** (2007), 22-8.

[7] P. Prakash, K. Krishna, D. Bhatia, Usefulness of SAPS II Scoring System as an Early Predictor of Outcome in ICU Patients, Journal, *Indian Academy of Clinical Medicine* **7** (2006), 202-205.

[8] R. Gupta, V.K. Arora, Performance evaluation of APACHE II score for an Indian patient with respiratory problems, *The Indian journal of medical research* **119** (2004), 273-282.