

Tanaffos (2010) 9(3), 58-64

©2010 NRITLD, National Research Institute of Tuberculosis and Lung Disease, Iran

## Assessing the Performance of Two Clinical Severity Scoring Systems in the ICU of a Tertiary Respiratory Disease Center

Seyed Mohammad Reza Hashemian<sup>1</sup>, Hamid Reza Jamaati<sup>2,3</sup>, Majid Malekmohammad<sup>2,4</sup>, Elnaz Ehteshami Afshar<sup>2</sup>, Oldoz Alesh<sup>2</sup>, Golnar Radmand<sup>2</sup>, Atefeh Fakharian<sup>2</sup>

<sup>1</sup> Tracheal Disease Research Center, <sup>2</sup>Department of Pulmonary Medicine, <sup>3</sup> Tobacco Prevention and Control Research Center, <sup>4</sup> Lung Transplantation Research Center, NRITLD, Shahid Beheshti University M.C., TEHRAN-IRAN.

### ABSTRACT

**Background:** The aim of this study is to compare the performance of five applied general severity scoring systems and their ability to predict mortality rate for the intensive care unit patients: Simplified Acute Physiology Score II (SAPS II), Mortality Probability Model II at admission (MPM II<sub>0</sub>), at 24 hours (MPM II<sub>24</sub>), at 48 hours (MPM II<sub>48</sub>) and over time (MPM II<sub>overtime</sub>). These scoring systems have been developed in response to an increased emphasis on the evaluation and monitoring of health care services; and also making cost-effective decisions.

**Materials and Methods:** In this historical cohort study, all of the scoring systems were applied to 114 patients and the predicted mortality rate and the Standardized Mortality Ratio (SMR) were calculated for them. Calibration of each model and discriminative powers were evaluated by using Hosmer-Lemeshow goodness of fit test and ROC curve analysis, respectively.

**Results:** The predicted mortalities were not significantly deviated from the main systems (SMR for SAPS II: 0.79, MPM II<sub>0</sub>: 1.10, MPM II<sub>24</sub>: 1.32, MPM II<sub>48</sub>: 1.08 and MPM<sub>overtime</sub>: 1.02). The Hosmer-Lemeshow statistics had the least value for MPM II<sub>48</sub> (C=2.922, p-value=0.939); and the discrimination was best for MPM II<sub>24</sub> (AUC=0.927) followed by SAPS II (AUC=0.903), MPM II<sub>0</sub> (AUC=0.899), MPM II<sub>48</sub> (AUC=0.848) and MPM II<sub>overtime</sub> (AUC=0.861).

**Conclusion:** All five general ICU mortality predictors showed accurate standardized mortality ratio. MPM II<sub>24</sub> had the best discrimination, MPM II<sub>0</sub> had the best SMR before 24 hours and MPM<sub>overtime</sub> had the best SMR after 24 hours. Performance of MPM II and its ease of use make it an efficient model for mortality prediction in our study. (Tanaffos2010; 9(3): 58-64)

**Key words:** Mortality probability model II, Simplified acute physiology score II, Intensive care unit

### INTRODUCTION

In recent decades the emphasis on developing systems to measure the severity of illness in the

intensive care units (ICUs) has increased. Several models have been made for mortality prediction in critically ill patients (1,2). The critical condition of ICU patients that increases the probability of mortality, as well as the expensive services offered (3) in this unit, have attracted more attention to the outcome of ICU patients. In spite of expensive

Correspondence to: Fakharian A

Address: NRITLD, Shaheed Bahonar Ave, Darabad, TEHRAN 19569, P.O:19575/154, IRAN

Email address: afakharian@yahoo.com

Received: 24 February 2010

Accepted: 26 May 2010

services offered in the ICUs, mortality rate is still high. Some believed that services provided in the ICU cannot prevent mortality and may also impose a financial burden on the economy by postponing the time of death. For appraising these claims, a variety of systems have been developed to evaluate the probability of death at the time of ICU admission. By using these indices, in addition to making decisions about the cost effectiveness of these services (4,5) and assess the performance of different ICUs (6), evaluation of the results of new treatments and technologies is also possible.

The main reasons that augmented the importance of these scoring systems are: 1- the scoring systems are used in clinical trials for matching, 2- these systems are used to quantify the severity of illness for the administrative decisions such as resource allocation, 3- the scoring systems assess the ICU performance, and compare the quality of care; and 4- they are used to appraise the prognosis of individual patients (7).

Several systems have been constructed for scoring the severity of illness in specific groups of patients (i.e. children, cancer patients, etc.) (8), and also for general patients. Two of the most common systems that are used for general patients in the ICUs, are: Simplified Acute Physiology Score (SAPS) and Mortality Probability Model (MPM).

In 1990s, SAPS II (9) was developed for mortality prediction from 17 variables. The summation of the scores computed from these variables, could be converted into the probabilities of mortality. The MPM II (10) model, was also made in 1990s. The MPM II system has models that could be used at the time of admission (MPM II<sub>0</sub>); after 24 hours (MPM II<sub>24</sub>); after 48 hours (MPM II<sub>48</sub>); and over the time (MPM II<sub>overtime</sub>) for patients who stay more than 48 hours in the ICU.

Several studies have been performed for evaluating these scoring systems in different

populations. In some studies these systems had good performance (11-14); and in some, they were not suitable for prediction (15,16).

The severity of illness scoring systems has been proven to be suitable for their own populations; but, because their accuracy is very sensitive to patient population changes, they need to be checked for validation and calibration before using in a new population.

Using the severity of illness scoring systems has not been common in the Iranian ICUs. In this study we evaluated the performance of two scoring systems: SAPS II and MPM II (MPM II<sub>0</sub>, MPM II<sub>24</sub>, MPM II<sub>48</sub> and MPM<sub>overtime</sub>) on a small sample of Iranian patients in the ICU of a referral center for respiratory diseases.

## MATERIALS AND METHODS

This historical cohort study was conducted on patients admitted to the intensive care unit of Masih Daneshvari Hospital as a referral center in October and November of 2007. A total of 136 patients were included in the study during two months. Patients' clinical and physiological data were collected using a questionnaire, designed according to the definitions described by the developers of MPM II and SAPS II scoring systems. Data were collected by a physician, from the information provided in patients' records, 24 and 48 hours after ICU admission.

Among 136 patients, 15 were excluded because of age younger than 18, coronary disease or staying less than 8 hours in the ICU; and 7 patients were excluded because of incomplete information. The predicted probability of death in the ICU was calculated by the logistic regression models, suggested in the original articles.

### *Statistical Analysis*

Descriptive statistics are shown as mean±SD for the quantitative and n (%) for the qualitative variables. To compare the quantitative variables

between the two groups, student's t-test and the nonparametric Mann-Whitney test were used. For assessing the relationship between the qualitative variables and the mortality, the Pearson chi-square test and Fisher's exact test were used when necessary.

To compare the performance of the severity of illness scoring systems (SAPS II, MPM II<sub>0</sub>, MPM II<sub>24</sub>, MPM II<sub>48</sub>, and MPM II overtime), the discrimination and calibration of them were assessed.

The calibration (the ability to predict the probability of death) was appraised by using Hosmer-Lemeshow goodness of fit test. Patients were sorted in an increasing order of estimated probability of death; and then were divided into 10 groups of approximately equal number of observations. Then, the number of the expected and observed mortality was compared in these groups by using Pearson statistics (C). The p-value less than 0.05 shows a statistically significant difference. Low value of statistics and high p-values indicate good agreement between the observed and expected number of deaths.

The Standardized Mortality Ratio (SMR) was calculated by dividing the observed mortality by the predicted mortality. The 95% confidence intervals (CIs) for SMRs were calculated by considering the observed mortality as a Poisson variable, then dividing its 95% CI by the predicted mortality.

The discriminative power of the scoring systems was assessed by performing the receiver operating characteristics (ROC) curve analysis and computing the area under the curve (AUC), and the corresponding 95% CI. All statistical analyses were performed using STATA software version 10.

## RESULTS

From 114 patients enrolled in this study, 26(22.8%) died and 78(68.42%) stayed in the ICU for more than 48 hours. There were 64(56.1%) males

and 50(43.9%) females; and the mean age was  $49.01 \pm 18.38$  yrs. Table 1 shows the demographic characteristics of patients. Length of ICU stay for the survivors was more than non-survivors ( $p < 0.0001$ ). Total hours of mechanical ventilation, emergency admission and SAPS II score in non-survivors were significantly more than in survivors.

Table 2 shows the SMR and the comparison between observed and expected mortality in different scoring systems. The SMR of SAPS II score was less than 1 whereas it was more than one for the other systems. The 95% confidence intervals (CI) of SMR for all systems was 1, therefore the observations were not significantly deviated from the expected values of the main systems. For all systems, the predicted probability of death in non-survivors was significantly higher than that of survivors.

In comparison of the Hosmer-Lemeshow goodness of fit statistic, MPM II<sub>24</sub> had the least value of C statistic ( $p = 0.939$ ); and SAPS II showed the worst calibration. None of them showed significant difference between observed and expected values.

Figure 1 demonstrates the calibration plots for the five scoring systems. In this plot, in addition to a fitted line that shows the relation between observed and predicted mortality and an identity line as a reference line, 10 points show 10 groups used for the Hosmer-Lemeshow goodness of fit test. The fitted line shows the deviation from the model. This figure illustrates that the SAPS II model overestimated and the MPM II underestimated the mortality probabilities in this sample.

The AUC was 0.903 (CI 95%: 0.827-0.979) for the SAPS II; 0.899 (CI 95%: 0.830-0.967) for MPM II<sub>0</sub>; 0.927 (CI 95%: 0.866-0.987) for the MPM II<sub>24</sub>; 0.848 (CI 95%: 0.740-0.955) for the MPM II<sub>48</sub> and 0.861 (CI 95%: 0.747-0.975) for the MPM II overtime. This explains that MPM II<sub>24</sub> and SAPS II had better discriminative powers (Table 3).

Table 1. Patients' characteristics

Variables	Total(n=114)	Alive(n=88)	Dead(n=26)	P-value
Age (mean±SD)	49.2±18.4	47.7±17.3	53.4±21.4	0.169
Gender				
Male	64(56.1%)	47(53.4%)	17(65.4%)	0.280
Female	50(43.9%)	41(46.6%)	9(34.6%)	
ICU LOS in days (mean±SD)	9.95±8.7	6.23±7.7	11.09±8.7	<0.0001
Total hours of MV	7.6±10.8	5.1±9.3	16.2±11.1	<0.0001
SAPS II Score	39.14±19.2	33.6±17.05	57.8±13.8	<0.0001
Emergency admission	65(57)	42(47.7%)	23(88.5%)	<0.0001
Previous ICU hospitalization	28(24.6%)	23(26.1%)	5(19.2%)	0.607
Surgery deliberation	86(75.4%)	75(85.2%)	11(42.3%)	<0.0001
CRF	13(11.4%)	6(6.81%)	7(26.9%)	0.071
Cancer	26(22.8%)	21(23.9%)	5(19.2%)	0.792
Infection	68(59.6%)	52(59.1%)	16(61.5%)	0.823

LOS: Length of Stay; MV: Mechanical Ventilation; CRF: Chronic Renal Failure

Table 2. Comparing the scoring systems in mortality prediction.

Scoring System	N	Observed Mortality	Expected Mortality	SMR	CI (95%)	Survivors	Non-Survivors	P-value
SAPS II	114	0.228	0.290	0.79	0.514-1.152	0.19±0.18	0.61±0.24	<0.0001
MPM II <sub>0</sub>	114	0.228	0.208	1.10	0.716-1.606	0.13±0.14	0.48±0.24	<0.0001
MPM II <sub>24</sub>	114	0.228	0.172	1.32	0.628-1.84	0.07±0.1	0.52±0.32	<0.0001
MPM II <sub>48</sub>	78	0.192	0.178	1.08	0.607-1.781	0.12±0.13	0.43±0.33	0.0025
MPM II Over time	78	0.192	0.188	1.02	0.574-1.686	0.12±0.11	0.47±0.34	0.0013

SMR: Standardized Mortality Ratio; CI: Confidence Interval.

Table 3. Comparison of discriminative power and calibration of scoring systems.

Scoring System	AUC <sup>†</sup>	CI <sup>‡</sup> (95%)	C <sub>HL</sub> <sup>*</sup>	P-value
SAPS II	0.903	0.827-0.979	7.659	0.467
MPM II <sub>0</sub>	0.899	0.830-0.967	5.527	0.701
MPM II <sub>24</sub>	0.927	0.866-0.987	5.222	0.734
MPM II <sub>48</sub>	0.848	0.740-0.955	2.922	0.939
MPM II Over time	0.861	0.747-0.975	4.785	0.780

<sup>†</sup> Area under the ROC curve

<sup>‡</sup> Confidence interval

<sup>\*</sup> Hosmer-Lemeshow statistic

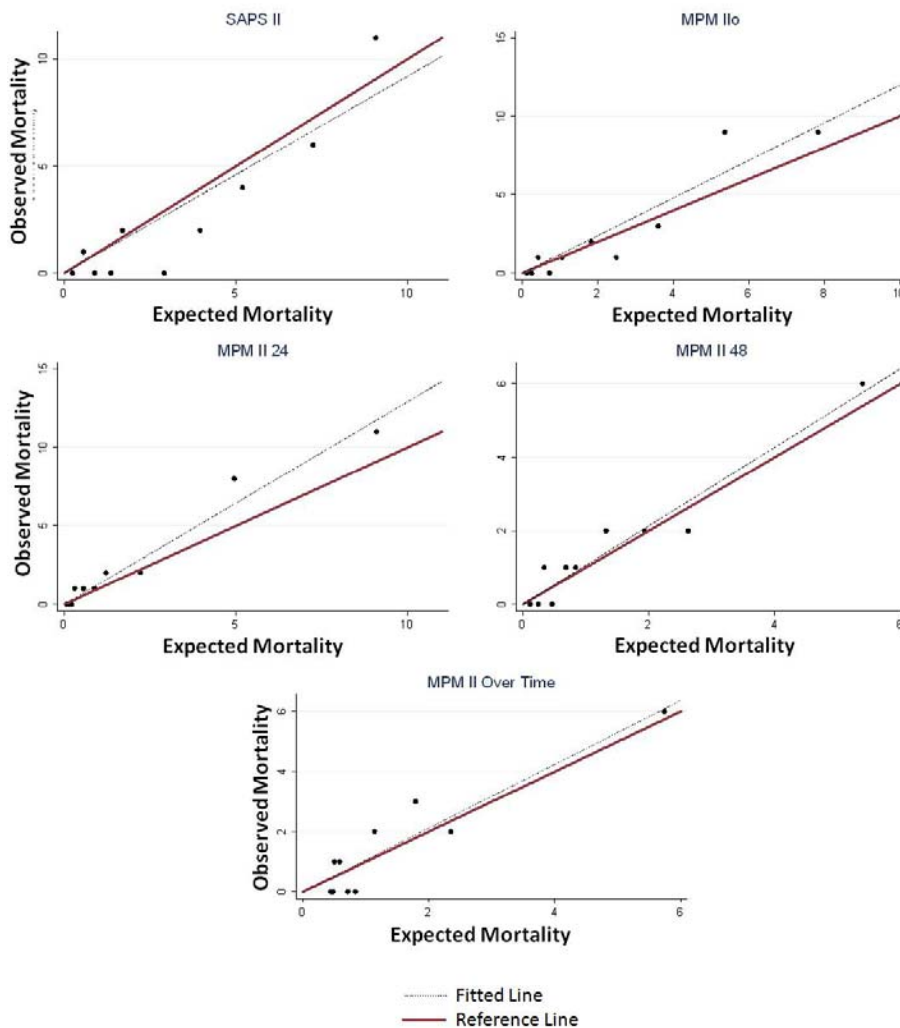


Figure 1. The Calibration plots for different models.

## DISCUSSION

The objective of this study was to assess the performance of the SAPS II and the MPM II models for mortality prediction in the intensive care units. The results demonstrate that the MPM II model had better performance than the SAPS II in prediction of mortality. The discrimination of MPM II24 model in our study (AUC=0.927) was better than that of a similar study conducted in a similar region (AUC=0.84) (12). Although in that study MPM II0 had better discrimination (AUC=0.85) and SAPS II

had the worst (AUC=0.79), in our study MPM II24 was the best and after that SAPS II had a good discriminative power (AUC=0.903).

The SMR for the SAPS II was less than one in our study. However, it was not statistically significant, but shows overestimation of SAPS II for the probability of mortality in our population. Alternatively, the SMRs of MPM II models were more than one (not statistically significant); and demonstrated underestimation for the probability of mortality.

The results also show that the MPM II24 had the best calibration ( $C=5.222$ ) for prediction before 24 hours of admission. For the prediction after 24 hours the best calibration was for the MPM II48 ( $C=2.922$ ).

Moreover, this study showed that although the SMR of MPM II0 was better than MPM II24, the discriminative power and calibration of MPM II24 was better than all other models.

The process of diagnosis and treatment of patients admitted to the ICUs, in most countries is guideline-based. Using these scoring systems for the mortality prediction along with the guideline-based medicine, help to compare ICUs' performances and also in assignment of patients' precedence for using the available facilities, evaluating the result of new interventions, technologies and protocols; and determination of cost-effectiveness in any process; although, we do not have these circumstances in our hospitals.

Furthermore, use of DNR (do not-resuscitate order) expression and the End of Life rules, is not ordinary in our hospitals. As a result, we have a different pattern of death in our ICUs. Therefore, assessing the effects of using this expression on the treatment of these patients could be an important subject for the next studies.

To reach this goal, it is important to choose a suitable index for measuring the probability of mortality and severity of illness for critically ill patients. Several systems have been developed for this purpose. Evaluating the performance of other systems, such as APACHE II and III can be another important issue for further studies.

Limitations of the present survey were the short period of time and having small sample size. A similar study with a larger sample size is recommended to see whether deviations from models are significant or not. If the deviations are significant, these scoring systems will need to be recalibrated before using in the Iranian population. Being a

historical cohort study was another limitation of this study. Hence, there was a risk of having incomplete or missing data. The incomplete data could be from documents of patients with better conditions; and consequently these patients were excluded from our study. Obviously, a prospective study will not have such problems.

## REFERENCES

1. Lewandowski K, Lewandowski M. Scoring systems in the intensive care unit. *Anaesthesist* 2003; 52 (10): 965-87; quiz 988- 9.
2. Ohno-Machado L, Resnic FS, Matheny ME. Prognosis in critical care. *Annu Rev Biomed Eng* 2006; 8: 567-99.
3. Pronovost P, Angus DC. Economics of end-of-life care in the intensive care unit. *Crit Care Med* 2001; 29 (2 Suppl): N46-51.
4. Hsiun Tang ,Chao, Che Ming Yang, Chi Yuang Chuang, Ming Lee Chang, Yu Chwen Huang, and Chin Feng Huang. A comparative study of clinical severity scoring systems in ICUs in Taiwan. *TZU Chi Med J* 2005: 239-45.
5. Glance LG, Osler T, Shinozaki T. Intensive care unit prognostic scoring systems to predict death: a cost-effectiveness analysis. *Crit Care Med* 1998; 26 (11): 1842- 9.
6. Glance LG, Osler TM, Dick A. Rating the quality of intensive care units: is it a function of the intensive care unit scoring system? *Crit Care Med* 2002; 30 (9): 1976- 82.
7. Gregoire, G, Russell JA. Assessment of severity of illness. In *Principle of Critical Care*, eds JB Hall, GA Schmidt, LDH Wood, McGraw Hill , New York, 1998.
8. Soares M, Fontes F, Dantas J, Gadelha D, Cariello P, Nardes F, et al. Performance of six severity-of-illness scores in cancer patients requiring admission to the intensive care unit: a prospective observational study. *Crit Care* 2004; 8 (4): R194- 203.
9. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 12 (11): 975- 7.
10. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on

- an international cohort of intensive care unit patients. *JAMA* 1993; 270 (20): 2478- 86.
11. Lemeshow S, Teres D, Avrunin JS, Pastides H. A comparison of methods to predict mortality of intensive care unit patients. *Crit Care Med* 1987; 15 (8): 715- 22.
  12. 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 (2): 166- 74.
  13. Kim EK, Kwon YD, Hwang JH. Comparing the performance of three severity scoring systems for ICU patients: APACHE III, SAPS II, MPM II. *J Prev Med Public Health* 2005; 38 (3): 276- 82.
  14. Gómez-García S, Ponce-de-León S, Hernández-Martínez MV, Macías AE. Mortality predictive indexes in non-critical inpatients. *Rev Invest Clin* 2007; 59 (3): 192- 6.
  15. Schäfer JH, Maurer A, Jochimsen F, Emde C, Wegscheider K, Arntz HR, et al. Outcome prediction models on admission in a medical intensive care unit: do they predict individual outcome? *Crit Care Med* 1990; 18 (10): 1111- 8.
  16. Del Bufalo C, Morelli A, Bassein L, Fasano L, Quarta CC, Pacilli AM, et al. Severity scores in respiratory intensive care: APACHE II predicted mortality better than SAPS II. *Respir Care* 1995; 40 (10): 1042- 7.