

Performance of APACHE IV in Medical Intensive Care Unit Patients: Comparisons with APACHE II, SAPS 3, and MPM₀ III

Mihye Ko¹, Miyoung Shim¹, Sang-Min Lee^{1,2}, Yujin Kim¹, Soyoung Yoon¹

¹Seoul National University Hospital, Seoul; ²Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea

Background: In this study, we analyze the performance of the Acute Physiology and Chronic Health Evaluation (APACHE) II, APACHE IV, Simplified Acute Physiology Score (SAPS) 3, and Mortality Probability Model (MPM)₀ III in order to determine which system best implements data related to the severity of medical intensive care unit (ICU) patients.

Methods: The present study was a retrospective investigation analyzing the discrimination and calibration of APACHE II, APACHE IV, SAPS 3, and MPM₀ III when used to evaluate medical ICU patients. Data were collected for 788 patients admitted to the ICU from January 1, 2015 to December 31, 2015. All patients were aged 18 years or older with ICU stays of at least 24 hours. The discrimination abilities of the three systems were evaluated using c-statistics, while calibration was evaluated by the Hosmer-Lemeshow test. A severity correction model was created using logistics regression analysis.

Results: For the APACHE IV, SAPS 3, MPM₀ III, and APACHE II systems, the area under the receiver operating characteristic curves was 0.745 for APACHE IV, resulting in the highest discrimination among all four scoring systems. The value was 0.729 for APACHE II, 0.700 for SAP 3, and 0.670 for MPM₀ III. All severity scoring systems showed good calibrations: APACHE II (chi-square, 12.540; P=0.129), APACHE IV (chi-square, 6.959; P=0.541), SAPS 3 (chi-square, 9.290; P=0.318), and MPM₀ III (chi-square, 11.128; P=0.133).

Conclusions: APACHE IV provided the best discrimination and calibration abilities and was useful for quality assessment and predicting mortality in medical ICU patients.

Key Words: APACHE; intensive care unit; Mortality Probability Model; severity of illness index; Simplified Acute Physiology Score

INTRODUCTION

Evaluating the quality of medical care requires evaluating the effectiveness of the treatment provided to a patient. Hence, a valid evaluation must be preceded by assessments of the patient's condition before the treatment is given [1]. In order to evaluate patient condition, intensive care units (ICUs) utilize severity scoring systems [2]. The quality of medical treatments in the ICU can be evaluated objectively by comparing actual mortality with predicted mortality using methods that take into account patient severity.

Severity scoring systems can be divided into two categories, depending on the specific sys-

Original Article

Received: May 10, 2018

Revised: August 24, 2018

Accepted: September 17, 2018

Corresponding author

Miyoung Shim

Seoul National University Hospital,
101 Daehak-ro, Jongno-gu, Seoul
03080, Korea

Tel: +82-2-2072-3425

Fax: +82-2-3676-4108

E-mail: smy219@snuh.org

Copyright © 2018 The Korean Society of Critical Care Medicine

This is an Open Access article distributed under the terms of Creative Attributions Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

tem used to collect data. The first category includes the Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS), and Mortality Probability Model (MPM), all of which quantify the initial condition of the patient within 24 hours of ICU admission. The second category includes organ system failure (OSF), sequential organ failure assessment (SOFA), organ dysfunction and infection system (ODIN), and multiple organ dysfunction score (MODS), all of which measure patient condition repeatedly throughout the admission period.

Among these systems, the severity scoring systems most commonly used in the ICU are APACHE, SAPS, and MPM, which evaluate the initial condition [2]. While the measurement variables used by these systems were initially selected subjectively, all have selected statistically meaningful variables since 1990, thereby improving their performance [3].

Our institution established a severity correction model using APACHE II, which was updated in 1985, and is used to constantly monitor predicted mortality in the ICU via APACHE II scores. However, the perceived need for a new system to measure patient severity has increased over time. In 2014, the performance of APACHE IV and SAPS 3 was compared in a surgical ICU context [4].

In this study, we analyzed APACHE IV, SAPS 3, and MPM₀ III in order to determine which system best implements data about the severity of medical ICU patients. We utilized the selected system to create a severity correction model that can measure predicted mortality rates.

MATERIALS AND METHODS

This research, approved by Institutional Review Board of Seoul National University Hospital (IRB No. H-1605-116-763), was a retrospective investigation that analyzed the discrimination and calibration of APACHE II, APACHE IV, SAPS 3, and MPM₀ III, when used for medical ICU patients.

Patient Population

The subjects of this study were medical ICU patients treated at a university hospital in Seoul from January 1, 2015 to December 31, 2015. Only adults over the age of 18 years were included. We also excluded patients with an ICU stay shorter than 24 hours because the systems evaluated severity using physiological events that appeared within 24 hours of ICU admission.

KEY MESSAGES

- Acute Physiology and Chronic Health Evaluation (APACHE) IV demonstrated the best discrimination and calibration abilities.
- APACHE IV was useful for quality assessment and predicting mortality of medical intensive care unit patients.

Data Collection

The data were collected by two nurses (intraclass correlation coefficient, 0.88), both of whom had served in ICUs for over 5 years and were trained in the use of all survey tools. The data showing highest patient severity were selected from the medical records of intensive care patients. APACHE II, APACHE IV, and SAPS three results reported physiological indices observed within 24 hours of admission, while MPM₀ III reported only those recorded within an hour of admission to the ICU.

Statistical Analysis

The data were analyzed using IBM SPSS ver. 23.0 (IBM Corp., Armonk, NY, USA) and MedCalc. The general characteristics of subjects and mortality rates after severity correction (observed mortality/predicted mortality) were analyzed using descriptive statistics, including percentage, mean, and standard deviation. The discrimination effectiveness of the three systems was evaluated through c-statistics, while calibration was evaluated by the Hosmer-Lemeshow test. A severity correction model was created using logistic regression analysis.

RESULTS

Characteristics of the Study Sample

In this study, we included a total of 788 patients, which includes 792 medical ICU patients admitted in 2015, minus four ICU patients during this period who were under 18. Among the included patients, 636 were in the medical ICU for more than 24 hours, and the other 152 (19%) were excluded due to redirection to a different general ward, redirection to another ICU, or death. A total of 61.4% of the subjects were male. The average age was 63.3 years, while the average length of stay in the ICU was 10.7 days. Among the admitted patients, 96.4% had medical problems, among which the most common was respiratory disease (46.6%). Other general characteristics of the patients are mentioned in Table 1.

During the course of the study, the overall observed mortality rate in the medical ICU was 31.5% (248 patients), while the

Table 1. General characteristics and observed mortality

Variable	Total (n=788)	Survivor (n=540)	Non-survivor (n=248)	P-value
Age (yr)	63.3±15.4	63.3±15.6	63.4±14.9	0.945
Male sex	484 (61.4)	326 (60.4)	158 (63.7)	0.547
ICU stay (day)	10.7±43.2	11.2±50.5	9.7±19.4	0.643
Reason for admission				0.538
Medical admission	760 (96.4)	519 (96.1)	241 (97.2)	
Disease category				0.218
Cardiovascular	103 (13.1)	72 (13.3)	31 (12.5)	
Neurologic	115 (14.6)	84 (15.6)	31 (12.5)	
Metabolic	43 (5.5)	28 (5.2)	15 (6.0)	
Gastrointestinal	86 (10.9)	52 (9.6)	34 (13.7)	
Respiratory	367 (46.6)	250 (46.3)	117 (47.2)	
Infection	17 (2.2)	15 (2.8)	2 (0.8)	
Surgical	50 (6.3)	36 (6.7)	14 (5.7)	
Other	7 (0.9)	3 (0.5)	4 (1.6)	
DNR status				<0.001
DNR	41 (5.2)	18 (3.3)	23 (9.3)	

Values are presented as mean±standard deviation or number (%).
ICU: intensive care unit; DNR: do not resuscitate.

rate for males was 32.6%. The average age of the deceased patients was 63.4 years, and the average length of stay in the ICU was 9.7 days. These results did not differ significantly for deceased or surviving patients. Although mortality rates did not differ between disease types, patients who refused resuscitation demonstrated a much higher mortality rate (56.1%) compared to patients who did not refuse resuscitation (Table 1).

Comparison of the Performances of APACHE IV and Other Severity Scoring Systems (APACHE II, SAPS 3, and MPM₀ III)

The performances of various severity scoring systems were evaluated through discrimination of death prediction and calibration. Discrimination of death prediction refers to the ability of each system to distinguish between the death and survival of a patient, which is illustrated by the receiver operating characteristic (ROC) curve (Figure 1). When the area under the ROC curve (AUC) values were analyzed, APACHE IV (AUC, 0.745) discriminated better than APACHE II (AUC, 0.729), SAPS 3 (AUC, 0.700), or MPM₀ III (AUC, 0.670). Discrimination of APACHE II and APACHE IV were similar (P=0.450). APACHE IV had better discriminatory power than SAPS 3 (P<0.05). The discriminatory performances of MPM₀ III were inferior to those of all the other systems (P<0.01).

The calibration of each system was analyzed using the Hos-

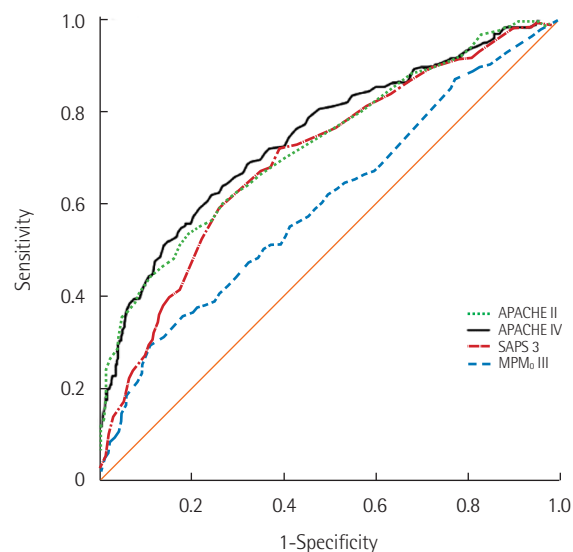


Figure 1. Comparison of the receiver operating characteristic curves for prediction of hospital death by APACHE II, APACHE IV, SAPS 3, and MPM₀ III. APACHE: Acute Physiology and Chronic Health Evaluation; SAPS: Simplified Acute Physiology Score; MPM: Mortality Probability Model.

mer-Lemeshow test, which compares results using each system with the actual simple mortality of the patient. Our findings are displayed in Table 2. All of the severity scoring systems were effective: APACHE II (chi-square, 12.540; P=0.129), APACHE IV (chi-square, 6.959; P=0.541), SAPS 3 (chi-square, 9.290; P=0.318), and MPM₀ III (chi-square, 11.128; P=0.133).

Based on the performances of APACHE IV, SAPS 3, and MPM₀ III, we derived a severity correction model using APACHE IV, which had the highest discrimination and calibration.

$$\text{Logit} \left(\frac{P}{1-P} \right) = -3.347 + 0.029 \times \text{APACHE IV score}$$

According to this equation, the odds of death (the ratio of probability of death compared to probability of surviving) increased 1.03 times when the APACHE VI score increased by 1. The significance probability was P<0.01.

DISCUSSION

The capacity of a severity scoring system to predict death varies between medical and surgical patients [5]. Such disparity presumably arises from the fact that the initial development of conventional systems such as APACHE and SAPS was targeted at large groups of patients with a variety of diseases. When these systems are implemented only for patients with a particular disease, additional factors other than the disease itself can affect predictions of death [6]. Therefore, in order to pre-

Table 2. Hosmer-Lemeshow's H chi-square tests for APACHE II, APACHE IV, SAPS 3, and MPM₀ III

Scoring system	Predicted death rate	Number	Non-survivor		Survivor	
			Observed	Expected	Observed	Expected
APACHE II	0.0≤P<0.1	67	6	5.99	61	61.01
	0.1≤P<0.2	71	15	10.34	56	60.66
	0.2≤P<0.3	65	14	12.08	51	52.92
	0.3≤P<0.4	56	15	12.16	41	43.84
	0.4≤P<0.5	48	11	12.46	37	35.54
	0.5≤P<0.6	62	16	18.47	46	43.53
	0.6≤P<0.7	69	20	24.59	49	44.41
	0.7≤P<0.8	76	28	34.54	48	41.46
	0.8≤P<0.9	65	35	36.21	30	28.79
	0.9≤P<1.0	57	50	43.15	7	13.85
Chi-square 12.540 with 8 DF (P=0.129)						
APACHE IV	0.0≤P<0.1	64	7	5.62	57	58.38
	0.1≤P<0.2	66	13	9.10	53	56.90
	0.2≤P<0.3	64	11	11.56	53	52.44
	0.3≤P<0.4	70	10	15.40	60	54.60
	0.4≤P<0.5	64	18	16.74	46	47.26
	0.5≤P<0.6	66	20	20.22	46	45.78
	0.6≤P<0.7	68	22	25.19	46	42.81
	0.7≤P<0.8	64	28	29.40	36	34.60
	0.8≤P<0.9	65	44	39.54	21	25.46
	0.9≤P<1.0	45	37	37.23	8	7.77
Chi-square 6.959 with 8 DF (P=0.541)						
SAPS 3	0.0≤P<0.1	65	9	8.61	56	56.39
	0.1≤P<0.2	64	11	11.67	53	52.33
	0.2≤P<0.3	62	14	13.57	48	48.43
	0.3≤P<0.4	64	15	15.88	49	48.12
	0.4≤P<0.5	67	11	18.56	56	48.44
	0.5≤P<0.6	74	24	23.71	50	50.29
	0.6≤P<0.7	70	34	26.16	36	43.84
	0.7≤P<0.8	63	31	27.68	32	35.32
	0.8≤P<0.9	64	32	34.35	32	29.65
	0.9≤P<1.0	43	29	29.80	14	13.20
Chi-square 9.290 with 8 DF (P=0.318)						
MPM ₀ III	0.0≤P<0.1	92	22	21.17	70	70.83
	0.1≤P<0.2	34	7	8.61	27	25.39
	0.2≤P<0.3	112	39	30.11	73	81.89
	0.3≤P<0.4	59	13	16.48	46	42.52
	0.4≤P<0.5	64	22	19.38	42	44.62
	0.5≤P<0.6	82	23	27.90	59	54.10
	0.6≤P<0.7	68	18	25.39	50	42.61
	0.7≤P<0.8	62	29	26.66	33	35.34
	0.8≤P<0.9	63	37	34.31	26	28.69
	0.9≤P<1.0	92	22	21.17	70	70.83
Chi-square 11.128 with 7 DF (P=0.133)						

APACHE: Acute Physiology and Chronic Health Evaluation; SAPS: Simplified Acute Physiology Score; MPM: Mortality Probability Model.

dict death rates more accurately, a scoring model that takes into account all of the different characteristics of patients in each ICU has been suggested [7-9].

In this study, we aimed to identify the severity scoring system that best reflects the characteristics of critically ill patients in medical ICUs. Two aspects of such systems were considered: discrimination and calibration. Discrimination in this context is the ability of the system to distinguish whether the subject has deceased or survived, represented by c-statistics. The systems with c-statistics closest to 1 are in descending order as follows: APACHE IV (AUC, 0.745), APACHE II (AUC, 0.729), SAPS 3 (AUC, 0.700), and MPM₀ III (AUC, 0.670). In another study focused on patients treated in our hospital's surgical ICU, we found that the c-statistics of APACHE IV were lower than those of SAPS 3. Nonetheless, the value of APACHE IV was still 0.80, indicating sufficient discrimination capacity [4]. Another previous study also suggested superior discrimination of APACHE and SAPS compared to MPM, which may be due to the larger number of variables considered by APACHE and SAPS, thus allowing these systems to better calculate the intricate relationships between different factors measured in patients [2].

Calibration refers to the closeness between the predicted mortality rate and the simple observed mortality rate, a probabilistic criterion used to evaluate the performance of each system when targeting all patients. In this study, all severity scoring systems satisfied $P > 0.05$ according to the Hosmer-Lemeshow test, therefore signifying excellent performance. Given that P is closer to 1 when the predicted mortality is closer to the simple observed mortality, APACHE IV had the best performance, followed by SAPS 3, MPM₀ III, and APACHE II. In most domestic and international studies, all three systems demonstrated adequate calibration for predicting simple mortality [10]. Some studies, however, suggested poorer calibration, both in surgical patients [4] and internal medicine patients [2]. Groeger et al. [11] suggested that these conflicting results can be explained by differences in the study population.

In general, APACHE IV and SAPS 3 reflected the severity of patients better than MPM₀ III. This result may be due to the number of variables each system implements. If the system requires more variables, it takes a longer time for evaluators to input the values, and the time required by each evaluator varies significantly. If a system requires fewer variables, there are fewer discrepancies between different evaluators, and severity measurements can be completed in a relatively short time. These reasons may explain the benefits of using MPM₀ III. However, patient factors are currently automatically extracted from

electronic medical records. This development eliminates the time-related disadvantages of systems that consider more variables, thereby making the ability to take full account of patient condition the most important criterion for choosing an appropriate severity scoring system. Thus, in this study we suggest a severity correction model ($P < 0.01$) using APACHE IV.

The limitations of this study lie with the data collection system. Whereas many studies evaluating the performance of different severity scoring systems collected data using a prospective cohort system [12-14], we collected data retrospectively in this study. We initially aimed to analyze records from past patients in order to immediately implement our study results in ongoing clinics, but it took a long time to analyze the medical records for all scoring systems. Thankfully, the average severity measured by APACHE II of medical ICU patients was 24.3 in 2014, 23.1 in 2015, and 23.7 in 2016, so there was no significant difference.

ICU patient severity plays the most important role in determining whether the patient will survive or not, and that severity is significantly affected by underlying disease. However, when the underlying disease is the same between patients, the impacts of other major factors on severity decrease, which may lead to erroneous death predictions [11]. The significance of the present study lies in this idea. Instead of implementing conventional severity scoring systems, we established a death prediction model specifically for patients admitted to the ICU for the same reason. Therefore, ICU entry should predict patient death much more accurately. We further recommend that other ICUs in our hospital, including the cardiopulmonary ICU and emergency ICU, should implement a specific severity scoring system that fully takes into account the characteristics of patients unique to each unit.

Besides patient factors, other structural factors play major roles in determining mortality rates, including standardization guidelines for the treatment and an abundance of medical personnel such as doctors and nurses working in the hospital [15-17]. Therefore, the development of a predictive model for death will depend on underlying disease. Such predictive models will serve as capstones for evaluating factors that affect the quality of medical care.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ORCID

Mihye Ko <https://orcid.org/0000-0003-1349-4575>
 Miyoung Shim <https://orcid.org/0000-0003-0349-6729>
 Sang-Min Lee <https://orcid.org/0000-0002-1388-9318>
 Yujin Kim <https://orcid.org/0000-0002-7712-7208>
 Soyoung Yoon <https://orcid.org/0000-0002-9181-1188>

REFERENCES

1. Kwon Y. Health care outcome measurement and risk adjustment. *J Korean Soc Qual Assur Health Care* 2007;13:59-67.
2. Keegan MT, Gajic O, Afessa B. Comparison of APACHE III, APACHE IV, SAPS 3, and MPM0III and influence of resuscitation status on model performance. *Chest* 2012;142:851-8.
3. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36.
4. Lee H, Shon YJ, Kim H, Paik H, Park HP. Validation of the APACHE IV model and its comparison with the APACHE II, SAPS 3, and Korean SAPS 3 models for the prediction of hospital mortality in a Korean surgical intensive care unit. *Korean J Anesthesiol* 2014;67:115-22.
5. Pappachan JV, Millar B, Bennett ED, Smith GB. Comparison of outcome from intensive care admission after adjustment for case mix by the APACHE III prognostic system. *Chest* 1999; 115:802-10.
6. 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;38:276-82.
7. Lee JH, Baek KJ, Han SB, Ahn ST, Shin DW, Kim AJ, et al. Mortality analysis of intensive care units patients using Mortality Probability Models (MPM II). *J Korean Soc Traumatol* 2001; 14:101-7.
8. Carson SS, Bach PB. Predicting mortality in patients suffering from prolonged critical illness: an assessment of four severity-of-illness measures. *Chest* 2001;120:928-33.
9. Schellongowski P, Benesch M, Lang T, Traunmüller F, Zauner C, Laczika K, et al. Comparison of three severity scores for critically ill cancer patients. *Intensive Care Med* 2004;30:430-6.
10. Kang CH, Kim YI, Lee EJ, Park K, Lee JS, Kim Y. The variation in risk adjusted mortality of intensive care units. *Korean J Anesthesiol* 2009;57:698-703.
11. Groeger JS, Lemeshow S, Price K, Nierman DM, White P Jr, Klar J, et al. Multicenter outcome study of cancer patients admitted to the intensive care unit: a probability of mortality model. *J Clin Oncol* 1998;16:761-70.
12. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006;34:1297-310.
13. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3: from evaluation of the patient to evaluation of the intensive care unit. Part 1: objectives, methods and cohort description. *Intensive Care Med* 2005;31:1336-44.
14. 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; 270:2478-86.
15. Carmel S, Rowan K. Variation in intensive care unit outcomes: a search for the evidence on organizational factors. *Curr Opin Crit Care* 2001;7:284-96.
16. Peelen L, de Keizer NF, Peek N, Scheffer GJ, van der Voort PH, de Jonge E. The influence of volume and intensive care unit organization on hospital mortality in patients admitted with severe sepsis: a retrospective multicentre cohort study. *Crit Care* 2007;11:R40.
17. Frick S, Uehlinger DE, Zuercher Zenklusen RM. Medical futility: predicting outcome of intensive care unit patients by nurses and doctors: a prospective comparative study. *Crit Care Med* 2003;31:456-61.