ORIGINAL ARTICLE

# Comparison of the accuracy and correctness of mortality estimates for Intensive Care Unit patients in internal clinics of the Czech Republic using APACHE II, APACHE IV, SAPS 3 and MPMoIII models

**Pavel Sedloň[1,2], Libor Kameník[1,2], Jan Škvařil[1], Martin Malý[1,2], Miloš Táborský[3], Miroslav Zavoral[1,2]**

[1]Department of Internal Medicine, Central Military Hospital, [2]1st Faculty of Medicine, Charles University in Prague; Prague, [3]Department of Internal Medicine I - Cardiology, University Hospital Olomouc; Czech Republic

## ABSTRACT

**Aim** To verify and compare the accuracies of mortality predictions in the Intensive Care Unit (ICU) of the Internal Clinic of Central Military Hospital in Prague, Czech Republic, using model APACHE II and the newer systems of the APACHE IV, SAPS 3 and $MPM_o$III.

**Methods** The data were collected retrospectively between 2011 and 2012, 1000 patients were evaluated. The assessment of the overall accuracy of the mortality predictions was performed using the standardized mortality ratio (SMR), and the calibration was assessed using the Lemeshow-Hosmer "goodness-of-fit" C statistic. Discrimination was evaluated using ROC curves based on calculations of the areas under the curve (AUCs).

**Results** The APACHE II, SAPS 3, and $MPM_o$III systems significantly overestimated the expected mortality, whereas the APACHE IV model led to correct estimations of the overall mortality. The discrimination capabilities of the models assessed according to the constructions of the ROC curves were evaluated as good, only the APACHE II was evaluated as satisfactory. The calibrations of all models were evaluated as unsatisfactory.

**Conclusion** The best mortality estimation for the investigated population sample was provided by the APACHE IV system. The discrimination capabilities of all models for the studied population were satisfactory, but the calibration of all of the systems was unsatisfactory. The conclusions of our study are limited by the relatively small size of the investigated sample and the fact that this study was conducted at only a single site.

**Key words:** intensive care units, Czech Republic, mortality, prognosis, humans, models, calibration, health status indicators, ROC curve, retrospective studies

**Corresponding author:**
Pavel Sedloň
Department of Internal Medicine (Cardiology), 1st Faculty of Medicine, Charles University, Military University Hospital
U Vojenské nemocnice 1200,
Prague 6 Czech Republic
Phone: +420 973 203 043;
Fax: +420 973 203 038;
E-mail: pavel.sedlon@uvn.cz

**Original submission**:
26 April 2015;
**Revised submission:**
20 June 2015;
**Accepted:**
22 June 2016.
doi: 10.17392/860-16

Med Glas (Zenica) 2016; 13(2):82-89

## INTRODUCTION

Prognostic scoring systems have been developed by companies for intensive medicine in an effort to objectively describe and quantify the severity of the conditions of selected groups of critically ill patients (1). Furthermore, an independent evaluation of the severity of a condition allows for the relatively objective assessments of the workloads required by intensive care units and resuscitation care units (2). To a certain extent, such independent evaluations also allow for the comparison of the quality and effectiveness of the care between these facilities (2,3)

There is a large number of prognostic models (3). Some generally focus on covering the entire population of critically ill patients, and some narrowly focus on predicting the prognosis and severity of narrowly defined conditions. The most used general prognostic scoring models include systems such as the Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS) and the Mortality Probability Model (MPM) (4–7). The oldest general complex systems are the APACHE system, which was published in 1981 (8), and the SAPS system from 1983 (9). The APACHE system was revised in 1985 and sub sequently published as the APACHE II system (4). With regard to the rapid development of health care opportunities and improvements in treatment procedures, the prognoses of a number of critical conditions have improved over a relatively short time; therefore, the original prognostic systems have become relatively absolute and require revision. Before using these models, all of these models should be validated for the specific population to which they should be applied (10,11).

The most recent revisions of the mentioned prognostic models were published between 2005 and 2007; e.g., the APACHE IV model in 2006(8), the European SAPS 3 model in 2005 (6,7) and the $MPM_0$III model in 2007 (12). To date no published test has been performed in the Czech Republic. Our work was committed to verifying the applicability of systems, such as the APACHE II and IV, $MPM_o$III and SAPS3, in a set number of ICU patients of an internal clinic in a tertiary care hospital in Prague, Czech Republic.

## PATIENTS AND METHODS

### Patients and study design

The Central Military Hospital in Prague, Czech Republic, is a university hospital of the third type which covers urban referral area including approximately 150.000 inhabitants. The intensive care unit (ICU) of the internal clinic has 9 beds that are fully equipped to provide comprehensive intensive care including artificial ventilation support and continual renal replacement therapy. It admits approximately 800 patients with prevailing internal diagnoses per year to this ICU. The hospital also houses a surgical intensive care unit and a resuscitation care unit (KARIM). It was retrospectively collected data from patients who were consecutively admitted to the internal ICU between 01 January 2011 and 31 August 2012. The data were obtained by revision of medical records and electronic documents stored in an internal hospitalization system database (AMIS). All tests and data were recorded in accordance with standard processes in relation to the underlying disease. An approval from the Central Military Hospital Ethics Committee was obtained before initiation of the study.

In total, data from 1009 patients were duly processed, and the data from 1000 patients were used for our study (i.e., 99.1%); nine patient records were excluded due to incompleteness or the unavailability of significant values. To suppress the variability in the data collection, all values were recorded by a single trained and experienced medical professional and subsequently revised by the authors of the study.

### Methods and statistical analysis

The analyses focused on the following data: the demographic characteristics, APACHE II scores, APACHE IV scores, SAPS3 scores, and the variables of MPM0 III system. Values obtained at the time of admission or within the first 24 hours of hospitalization were used. In accordance with the original methodology of data collection, the values with the greatest deviations from the normal (physiological) values were used. Individual cases were reviewed in cases of missing values. If a missing value was assessed as insignificant, it was taken to be normal. Patients with missing values that were

considered significant were not included in the assessment. Equally to all evaluated models, the mortality of the patients was assessed upon discharge from the hospital. The patients who were transferred to another facility for hospitalization or intensive care were further monitored, and their mortalities were assessed at discharge to home care or when they were moved to facilities that provided long-term nursing and rehabilitative care.

The APACHE II scores and mortality estimates were calculated according to the original equation (4) as were the predictions of mortality based on the APACHE IV, MPMoIII and SAPS3 (6–8,12). The standardized mortality ratio (SMR) was calculated as the ratio of the observed hospitalization mortality and the predicted mortality. The 95% confidence intervals (CI95%) were obtained using the observed mortality as a Poisson variable and then using the 95% CI divided by the predicted mortality. We further verified the general accuracy of the models by determining the Brier score (13).

Because the Brier scores depend on the prevalence of the reference character in the population (mortality in our case), it is not suitable for comparing the accuracies of models in different populations. Therefore, we completed the calculation by determining the "scaled" Brier score ($BS_{scaled}$), which is independent of the starting set. (14) Verification of the system's applicability to studied cohort of patients was conducted by reviewing its calibration and discrimination. Calibration, which expresses the ability of the test to determine the probability of death in accordance with the observed mortality, was calculated using the Hosmer-Lemeshow test (15). Rejection of the null hypothesis, i.e., the assumption of the absence of a difference between the predicted and actual mortality in each group, was performed at a significance level of $p<0.05$.

The abilities of the individual systems to distinguish survivors and non-survivors according to the estimated mortality (discrimination) were assessed using receiver operating characteristic (ROC) curves and 2x2 classification matrices. The ROC curves were constructed as discrimination measurements with distributions per 10% according to the predicted mortality. The obtained curves were compared using the calculated area under the curve (AUC). AUC values > 0.75 were evaluated as satisfactory, AUC values > 0.8 were evaluated as good, and AUC values > 0.9 were evaluated as very good. The classification matrices were processed with the classification criteria of 10%, 30% and 50%, and based on these criteria, the sensitivities and specificities of the individual tests were determined.

## RESULTS

### Demographic characteristics of the patients

The mean age of the patients was 69.9 years, and the representations of both sexes were relatively proportional (56% males versus 44% females). The total mortality at discharge was 10.9% (109 patients). The representation of surgical patients was very low, 11 (1.1%) patients. Total of 638 patients (63.8%) were admitted due to emergency situations. Significantly higher mortality was observed in the group of patients who were moved to the ICU from wards, 47 (23%) patients. The group of the deceased also exhibited longer histories of previous hospitalization. The deceased patients were older (78.5 years vs. 68.8 years; $p <0.001$), and at a lower level of significance a difference in mortality between the sexes (the mortality of the females was 19%, and the mortality of the males was 11%; $p <0.05$) was also found. Factors associated with higher mortality included cardiopulmonary resuscitation before or at the time of admission ant the need for mechanical ventilation ($p<0.001$). In a greater proportion of patients, hospitalization was discontinued with discharge to home care, 702 (70%). Some patients were moved to facilities for long-term follow-up care (n = 189, 19%). A total of 109 (10.9%) patients died during hospitalization (Table 1).

Table 2 shows the actual mortality and the mortalities predicted based on the examinations of the scoring systems. Mortality was predicted correctly only with the APACHE IV system. In all other systems, the SMRs were significantly below 1, which indicated an overestimation of the set estimated mortality. The Brier scores were satisfactory for all models. Values modified using the "scaled" Brier score (13,14) revealed relatively better prediction accuracies for the APACHE IV and SAPS 3 models, a border-

**Table 1. Baseline characteristics of the patients**

| Parameter | Mean ± SD | | | |
|---|---|---|---|---|
| | Total | Survivors | Non-survivors | p |
| **Total** | 1000 | 891 | 109 | |
| **Mean age (years)** | 69.9 ± 16.3 | 68.8 ± 16.7 | 78.5 ± 9,5 | <0.001* |
| **Sex (%)** | | | | |
| **Male** | 560 (56) | 505 (56.7) | 55 (50.5) | <0,05† |
| **Female** | 440 (44) | 386 (43.3) | 54 (49.5) | <0,05† |
| **Type of admission (%)** | | | | |
| **Transfer from ward** | 199 (19.9) | 152 (17) | 47 (43.1) | <0.001 |
| **From other hospitals** | 163 (16.3) | 157 (17.6) | 6 (5.5) | <0.01 |
| **Duration of previous hospitalization** | 1.68 ± 5.34 | 1.42 ± 5.05 | 3.8 ± 6.95 | <0,001* |
| **Need for mechanical ventilation (%)** | 52 (5.2) | 30 (3.3) | 22 (20.2) | <0.001 |
| **Surgical procedure before admission (%)** | 11 (1.1) | 10 (1.1) | 1 | NS |
| **Cardiopulmonary resuscitation (%)** | 39 (3.9) | 28 (3.1) | 11 (10) | <0.01 |
| **Hospital discharge location (%)** | | | | |
| **Home** | 702 (70.2) | 702 (78.7) | 0 | |
| **Other** | 189 (18.9) | 189 (21.2) | 0 | |
| **Death** | 109 (10.9) | 0 | 109 | |
| **Mortality probability, APACHE II (%)** | 16.6 ± 12.8 | 15.1 ± 11.7 | 28.4 ± 15.5 | <0.001* |
| **Mortality probability, APACHE IV (%)** | 9.85 ± 15.1 | 11.5 ± 7.6 | 27.9 ± 25.4 | <0.001* |
| **Mortality probability, SAPS3 (%)** | 16.2 ± 16.4 | 12.9 ± 13.4 | 39.3 ± 22.4 | <0.001* |
| **Mortality probability, MPM0 III (%)** | 13.5 ± 17.1 | 14.1 ± 11.2 | 31.7 ± 26.3 | <0.001* |

*survivors vs non-survivors; †male vs female mortality

line prediction accuracy for the $MPM_0$III model, and a relatively unsatisfactory prediction accuracy for the APACHE II model.

**Table 2. In-hospital mortalities predicted by the systems**

| System | Actual mortality | Predicted mortality | SMR | 95%CI | Brier score | BS scaled |
|---|---|---|---|---|---|---|
| **APACHE II** | 0.109 | 0.166 | 0.66 | 0.63-0.69 | 0.0909 | 6.30% |
| **APACHE IV** | 0.109 | 0.098 | 1.11 | 1.00-1.22 | 0.0806 | 17% |
| **SAPS3** | 0.109 | 0.162 | 0.67 | 0.63-0.71 | 0.0764 | 21.30% |
| **MPM0III** | 0.109 | 0.134 | 0.81 | 0.75-0.87 | 0.0873 | 10% |

SMR, standardized mortality ratio abbreviation in full; BS, Brier score;

## Estimated mortality based on primary diagnosis at admission

The division of the total set into categories according to primary diagnosis at admission (Table 3) resulted in the total accuracies of the mortality predictions slightly decreasing in most of the systems. All systems achieve the best prediction of total mortality for respiratory diseases. The mortality was most overestimated for cardiovascular diseases, particularly in patients with acute myocardial infarction (MI) (Table 4).

## Calibration

For all systems, the $\chi^2$ values were not very satisfactory, and p values <0.05 were observed in all cases. Relatively, the best calibration was observed for the APACHE IV system ($\chi^2$=24.2) with a p value of 0.002, for APACHE II system was$\chi^2$ value slightly higher ($\chi^2$=25.7) with p value of 0.001. The SAPS3 and the $MPM_0$III systems exhibited the worst calibrations in our group, with high $\chi^2$ values (31.1 and 43.2) and p values <0.001.

**Table 3. Standardized mortality ratio (SMR) according to reason for admission**

| Subcategory | No of patients | Died | Standardized mortality ratio | | | |
|---|---|---|---|---|---|---|
| | | | APACHE II | APACHE IV | SAPS 3 | MPMoIII |
| **Cardiovascular** | 610 | 41 | 0.49 (0.46-0.53) | 0.78 (0.69-0.89) | 0.47 (0.44-0.51) | 0.47 (0.43-0.52) |
| **Respiratory** | 112 | 22 | 1.03 (0.91-1.19) | 1.21 (1.00-1.52) | 0.89 (0.76-1.07) | 1.42 (1.13-1.87) |
| **Gastrointestinal and metabolic** | 193 | 12 | 0.51 (0.49-0.59) | 0.91 (0.73-1.2) | 0.47 (0.42-0.55) | 0.88 (0.74-1.09) |
| **Other** | 85 | 33 | 1.84 (1.60-2.17) | 2.23 (1.76-3.03) | 1.27 (1.1-1.52) | 1.75 (1.43-2.24) |

**Table 4. Standardized mortality ratio (SMR) according to presence of myocardial infarction**

| Subcategory | No of patients | Died | Standardized mortality ratio | | | |
|---|---|---|---|---|---|---|
| | | | APACHE II | APACHE IV | SAPS 3 | MPMoIII |
| **Myocardial infarction (MI)(16)** | 200 | 7 | 0.29 (0.27-0.33) | 0.59 (0.49-0.75) | 0.33 (0.30-0.38) | 0.37 (0.31-0.45) |
| **Non-MI patients** | 800 | 102 | 0.72 (0.62-0.83) | 1.17 (1.12-1.24) | 0.72 (0.70-1.24) | 0.88 (0.77-1.03) |

### Discrimination

The worst correctness of the classification was observed with the APACHE II system. At the selected discriminating levels of 30 and 50% all of the other systems reached values of overall classification accuracy of approximately 88 and 89%, respectively (Table 5).

**Table 5. Results of the analysis of the classification matrices**

| System | Died | | Survived | | Sensitivity (95% CI) | Specificity (95% CI) | OCCR |
|---|---|---|---|---|---|---|---|
| | PD | PS | PD | PS | | | |
| **Cutoff 10%** | | | | | | | |
| **APACHE II** | 101 | 8 | 576 | 315 | 92 (86-96) | 35 (32-39) | 42% |
| **APACHE IV** | 72 | 37 | 166 | 725 | 66 (56-74) | 81 (79-84) | 80% |
| **MPMoIII** | 86 | 23 | 293 | 598 | 79 (70-85) | 67 (64-70) | 68% |
| **SAPS3** | 102 | 7 | 386 | 505 | 94 (87-97) | 57 (53-60) | 61% |
| **Cutoff 30%** | | | | | | | |
| **APACHE II** | 48 | 61 | 109 | 782 | 44 (35-54) | 88 (85-90) | 83% |
| **APACHE IV** | 36 | 73 | 43 | 846 | 33 (24-42) | 95 (94-96) | 88% |
| **MPMoIII** | 40 | 69 | 55 | 836 | 37 (28-46) | 94 (92-95) | 88% |
| **SAPS3** | 65 | 44 | 81 | 810 | 60 (50-68) | 90 (88-92) | 88% |
| **Cutoff 50%** | | | | | | | |
| **APACHE II** | 9 | 100 | 13 | 878 | 8 (4-15) | 99 (98-99) | 89% |
| **APACHE IV** | 23 | 86 | 20 | 871 | 21 (15-30) | 98 (97-99) | 89% |
| **MPMoIII** | 26 | 83 | 29 | 862 | 24 (17-33) | 97 (95-98) | 89% |
| **SAPS3** | 37 | 72 | 32 | 859 | 34 (26-43) | 96 (95-97) | 90% |

PD, death predicted; PS, survival predicted; OCCR, overall correct classification rate

ROC curves were constructed for all the measured systems. Good discrimination capabilities were demonstrated by the SAPS 3 (AUC 0.867) and APACHE IV (AUC 0.841) systems, and a slightly worse but still good discrimination capability was also observed for the MPMoIII system (AUC 0.807). As expected, the oldest system, i.e., the APACHE II, exhibited the worst value, although it was still satisfactory (AUC 0.766).

## DISCUSSION

The construction of prognostic scoring systems is always determined by the population sample used for the calibration and subsequent validation of the model. Similarly, the construction of models is always limited by the time of their creation, and the results are always related to the current level of medicine at that time. Finally, the outputs of models are limited by the resources invested in medical care within the given population under examination. All of these limits lead to potential limitations regarding the use of scoring systems in different populations and in different eras (2).

All of the systems under investigation were developed in sample populations with compositions that differed from the set of our workplace. Although the diagnoses and conditions occurring in our group were part of the spectrum of diagnoses of the cohorts of patients used to develop the tested models, their proportions and profiles of the severities of the conditions were different.

In our work, we observed an overestimation of the estimated mortality that applied to all models with the exception of the most recent and most comprehensive model, i.e., the APACHE IV. This overestimation of predicted mortality was particularly strong in the subgroup of patients with acute myocardial infarction (16). This finding correlates well with the decrease in mortality in these patients that has been documented in recent years in connection with the development of acute reperfusion therapy (17). In the Czech Republic, the predominant mode of treatment for patients with acute coronary syndrome is acute PCI, which is expected to lead to significant reductions in mortality in patients who are admitted to the hospital with acute myocardial infarction (18). However, this conclusion may not be universally valid as demonstrated by the study by Nassara Jr. et al. from 2013 that was conducted at the University Hospital in Sao Paulo; in this study, the overall predictions of mortality and the discrimination capabilities of the APACHE IV and SAPS 3 tests for this group of patients were good (19). If we exclude the group of patients with myocardial infarction in our sample from the evaluation, the overall accuracy of the mortality classification according to the SMR significantly improves for all newer systems. Only the APACHE II system also significantly overestimated the overall mortality in this group of patients (SMR 0.72, CI 0.62 to 0.83).

We found a number of studies that compared the prediction accuracies of the APACHE IV, SAPS II and $MPM_0$III systems (20–23). There are also many published papers that have described the functionality of the selected general scoring systems in various intensive care units in different countries (21,24–26). Kuzniewitz

et al. (22) compared the functionalities of the APACHE IV, SAPS II and $MPM_o$III models in a study involving 11,300 patients who were admitted to 35 intensive care units in California. When applied to the US population, the best discrimination capability was observed for the APACHE IV (AUC 0.892) model followed by the SAPS 2 (AUC 0.873) and $MPM_o$III (AUC 0.809) models, and these results agree well with our own. Similarly, Keegan et al. (23) compared the APACHE III, APACHE IV, SAPS 3 and MPMoIII models in 2596 patients from three intensive care units (Rochester, Mayo Medical Centre). Regarding the discrimination capabilities as expressed by ROC curves, these authors found the best results for the APACHE IV (AUC 0.861), satisfactory results for the SAPS (AUC 0.801) and the worst results for the MPMoIII (AUC 0.721). Again, these findings confirm the values in the cases of the SAPS 3and APACHE IV models observed in our work and the quite significant difference for the MPMoIII. The calibrations of the models examined in this study were generally poor with $\chi^2$ values of 21.8 for the MPMoIII ($p<0.05$) and 31 ($p<0.05$) for the APACHE IV, which correspond to the similarly disappointing calibration values for all of the models in our work. Lee et al. (26) compared the APACHE IV with the APACHE II and SAPS 3 models for patients who were admitted to a surgical intensive care unit of a university hospital (South Korea). These authors described good discrimination capabilities of these systems when applied to different populations but distinctly unsatisfactory calibrations for all of the models, and these results are analogous to those from our work

There are many possible reasons for which the calibrations of assessed systems did not reach satisfactory values for the sample under examination. The most frequently mentioned reasons include differences in the evaluated sets or differences in the sizes of cohorts used to develop and validate the models (27). In our work, a major role was also played by the significantly higher proportion of patients with cardiovascular issues, especially the patients with myocardial infarctions, and the practical absence of surgical patients. Poor model calibrations can also be associated with a relatively higher proportions of patients with low overall risks (28). Other causes may also include regional differences in the approach to decisions about terminating care, differences in the diagnoses of certain diseases and regional differences in the manner of providing medical care (15,29).

The results of our work should be evaluated with caution due to the present limitations and restrictions. This work is limited in scope to a single workplace that predominantly provides care to patients with diagnoses of internal medical natures. The number of patients included in the evaluation was not small; however, this number was significantly smaller compared to the large multicentre studies or sets based on which the assessed models were constructed. Therefore, our findings may not be fully transferable and valid for other intensive care units.

In conclusion, our findings revealed that the investigated systems overestimated the expected total mortality with the exception of the APACHE IV, which is the most comprehensive system but is rarely used in our conditions. However, when the patients with diagnosed myocardial infarctions were removed from the evaluation, the total mortality predictions of the $MPM_o$III and SAPS III systems significantly improved. The APACHE II overestimated the mortalities for all groups of patients. In the patients with diagnoses of acute myocardial infarction, the prediction of mortality was dramatically overestimated by all of the investigated models, and these systems should not be used in this patient group without further modifications. Regarding our group, the discrimination capabilities as evaluated with ROC curves were good for all of the newer models and satisfactory for the APACHE II model. The calibrations of all of the systems for our patient population were assessed as unsatisfactory. Nevertheless, the conclusions of our work need to be interpreted with caution due to the limitations and restrictions, particularly the relatively small scale of the studied population with prevailing internal diseases and the fact that the study was conducted at only a single site.

## ACKNOWLEDGMENT

We thank Petra Hajslova who provided the data collection.

## FUNDING

This work is supported by the Intent of Development of the Central Military Hospital, Ministry of Defense, MO 1012.

## TRANSPARENCY DECLARATION

Conflict of interest: None to declare.

## REFERENCES

1. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. Crit Care Med 2011; 39:163–9.
2. 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; 30:1976–82.
3. Strand K, Flaatten H. Severity scoring in the ICU: a review. Acta Anaesthesiol Scand 2008; 52:467–78.
4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13:818–29.
5. Higgins TL, Kramer AA, Nathanson BH, Copes W, Stark M, Teres D. Prospective validation of the intensive care unit admission Mortality Probability Model (MPM0-III). Crit Care Med 2009; 37:1619–23.
6. Metnitz PGH, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR; SAPS 3 Investigators. 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.
7. Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR; SAPS 3 Investigators. 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; 31:1345–55.
8. 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.
9. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R, Villers D. A simplified acute physiology score for ICU patients. Crit Care Med 1984; 12:975–7.
10. Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. A systems update. JAMA 1994; 272:1049–55.
11. Teres D, Lemeshow S. Why severity models should be used with caution. Crit Care Clin 1994; 10:93–110; discussion 111–5.
12. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med 2007; 35:827–35.
13. Brier GW. Verification of forecasts expressed in terms of probability. Mon Wea Rev 1950:1–3.
14. Wu Y-C, Lee W-C. Alternative Performance Measures for Prediction Models. PLoS One 2014; 9:e91249
15. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982; 115:92–106.
16. Thygesen K, Alpert JS, White HD, Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. J Am Coll Cardiol 2007; 50:2173–95.
17. Widimský P1, Budesínský T, Vorác D, Groch L, Zelízko M, Aschermann M, Branny M, St'ásek J, Formánek P; 'PRAGUE' Study Group Investigators. Long distance transport for primary angioplasty vs immediate thrombolysis in acute myocardial infarction. Final results of the randomized national multicentre trial--PRAGUE-2. Eur Heart J 2003; 24:94–104.
18. Widimsky P, Zelizko M, Jansky P, Tousek F, Holm F, Aschermann M, CZECH investigators. The incidence, treatment strategies and outcomes of acute coronary syndromes in the "reperfusion network" of different hospital types in the Czech Republic: results of the Czech evaluation of acute coronary syndromes in hospitalized patients (CZECH) registry. Int J Cardiol 2007; 119:212–9.
19. Nassar Junior AP1, Mocelin AO, Andrade FM, Brauer L, Giannini FP, Nunes AL, Dias CA. SAPS 3, APACHE IV or GRACE: which score to choose for acute coronary syndrome patients in intensive care units? São Paulo Med J Rev Paul Med 2013; 131:173–8.
20. Costa e Silva VT, de Castro I, Liaño F, Muriel A, Rodríguez-Palomares JR, Yu L. Performance of the third-generation models of severity scoring systems (APACHE IV, SAPS 3 and MPM-III) in acute kidney injury critically ill patients. Nephrol Dial Transplant 2011; 26:3894–901.
21. 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; 6:166–74.
22. Kuzniewicz MW, Vasilevskis EE, Lane R, Dean ML, Trivedi NG, Rennie DJ, Clay T, Kotler PL, Dudley RA. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008; 133:1319–27.
23. 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.
24. Katsaragakis S, Papadimitropoulos K, Antonakis P, Strergiopoulos S, Konstadoulakis MM, Androulakis G. Comparison of Acute Physiology and Chronic Health Evaluation II (APACHE II) and Simplified Acute Physiology Score II (SAPS II) scoring systems in a single Greek intensive care unit. Crit Care Med 2000; 28:426–32.

25. Cook DA. Performance of APACHE III models in an Australian ICU. Chest 2000; 118:1732–8.
26. Lee H, Shon Y-J, Kim H, Paik H, Park H-P. 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.
27. Murphy-Filkins R, Teres D, Lemeshow S, Hosmer DW. Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. Crit Care Med 1996; 24:1968–73.
28. Beck DH, Smith GB, Taylor BL. The impact of low-risk intensive care unit admissions on mortality probabilities by SAPS II, APACHE II and APACHE III. Anaesthesia 2002; 57:21–6.
29. Brinkman S1, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, de Keizer NF. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. J Crit Care 2011; 26:105.e11–8.