

# Mortality scoring in ITU

\*Grzegorz Niewiński, Andrzej Kański

2<sup>nd</sup> Department of Anaesthesiology and Intensive Therapy, Medical University of Warsaw

## Abstract

Chronic shortage of ITU beds makes decisions on admission difficult and responsible. The use of computer-based mortality scoring should help in decision-making and for this purpose, a number of different scoring systems have been created; in principle, they should be easy to use, adaptable to all populations of patients and suitable for predicting the risk of mortality during both ITU and hospital stay.

Most of existing scales and scoring systems were included in this review. They are frequently used in ITUs and become a necessary tool to describe ITU populations and to explain differences in mortality. As there are several pitfalls related to the interpretation of the numbers supplied by the systems, they should be used with the knowledge on the severity scoring science. Moreover, the cost and significant workload limit the use of scoring systems; in many cases an extra person has to be employed for collection and analysis of data only.

**Key words:** intensive care, scoring systems

Anestezjol Intens Ter 2012; 44: 47-50

Permanent shortage of ITU beds makes decisions on admission difficult and responsible, not to mention strong emotions that accompany them. Therefore, high hopes are placed in computer-based methods that should aid in decision-making.

Due to rapid development of ITUs in the 60s and 70s of the previous century, new tools to assess the efficacy of intensive therapeutic methods used in severely ill patients were required [1]. At the beginning of the 80s of the 20<sup>th</sup> century, three scales were designed, i.e. Acute Physiology and Chronic Health Evaluation Score (APACHE) [2], Simplified Acute Physiology Score (SAPS) [3] and Mortality Probability Model (MPM) [4]. The second version of APACHE, APACHE II, is of interest, as it is the most commonly used and cited method in the medical literature [5]. A decade later, the new, improved versions of the scales discussed were prepared – APACHE III [6], SAPS II [7] and MPM II [8], based on analysis of thousands of cases of patients hospitalized in many counties; as a result the reliability of these tools was remarkably higher. [9, 10]. In recent years, some newer versions were introduced – APACHE IV, SAPS III and MPM III.

An ideal scale should be characterised by simplicity of use (data obtained during ITU admission should

be enough to fill it), universality (should be suitable to assess all populations of patients), good discrimination and calibration, as well as capability to estimate ITU and hospital death risk [11].

The results obtained using the scales are mainly used to assess the efficacy of treatment and unit organization. Thanks to them the effectiveness and quality of treatment in different ITUs can be compared; moreover, they are useful tools facilitating therapeutic decisions at various stages and levels of therapy.

The use of scales in intensive therapy requires thorough knowledge of data collection and great caution in drawing conclusions.

A universal scoring system enabling the assessment of patients in all ITUs is extremely difficult, if not impossible, to design. ITUs are frequently specialised care units; thus, patients hospitalised at various centres have different characteristics. The entire issue is very complex. On the one hand, the organisation-related factors are decisive, such as the location of ITUs in the health care systems of various countries; on the other hand, the characteristic features of the population are relevant, e.g. behavioural and nutritional habits, addictions, professions, genetic differences. Such variations can be overcome by increasing the sizes of patients' populations in the scoring model.

The scale design should be based on the meticulously planned database, recruiting patients from many ITUs in various countries and subsequently assessed [11]. The objective of this process is to determine the capacity of a given scale to predict the risk of death; therefore, the process should be carried out based on another group of patients (assessment group) than the one used for data collection. The newest versions of scales, published in the recent five years and based on the analysis of thousands of patients, have not been thoroughly assessed. Therefore, there are some doubts whether their designs are correct and the results they provide are reliable [12].

The long-term collection of the data needed to structure and assess a scale model may also generate errors associated with the changes in the population examined, the organisation of ITUs and hospitals as well as treatment methods. Unfortunately, the size of the population considered has not been defined [11].

The scoring system model should be created based on the real population of ITU patients; hence we cannot expect that the population examined during designing the model will be representative for a larger group of patients. This may not be so important in Poland, where ITUs mainly differ in terms of the available equipment rather than the types of patients, yet it does matter in the United States and many European countries, where ITUs have different profiles, e.g. general, trauma, cardiac or neurosurgical. Moreover, since the scales based on diagnostic variables do not always provide better assessment than the scales based on physiological variables, their most recent versions include larger groups of patients and use more advanced statistical models [13].

Another relevant issue is the lack of prognostic nature of scales. In many cases, the risk of mortality of low-death risk patients is underestimated whereas the mortality in the highest death risk patients is overestimated. For these reasons, the actions were undertaken to change the formula calculating the risk of death or to restructure the relevance of the factors affecting the formula. In some cases, the recalibrating actions taken improved the accuracy of prognosis [14, 15], yet the pivotal model-, physician- or patient-related problems were not eliminated. The problems related to a physician or a person entering the data resulted from the differences in interpretation of the same definitions, patients' conditions, etc [16, 17]. The patient-related problems regarded the changes in characteristics of the examined population [18]: age distribution, disease severity, new methods of therapy. The model-related problems resulted from the lack of diagnostically important information [6], e.g. data concerning infections (their aetiology and location) [19, 20]; moreover, they were associated with the errors during entering and statistical analysis of data [1, 21].

The scales have to be periodically remodelled due to ongoing advances in medicine. Their reliability decreases with time because of increasingly better therapeutic and organisational methods as well as availability of novel medical devices [22]. Furthermore, the profile of patients undergoes changes. As a result, the statistical relevance

of the analysed factors alters and new parameters are introduced to mathematical models. Therefore, the scales have to be continuously widened and become increasingly complex (e.g. APACHE IV, in which about 37 min are needed to fill its electronic form) [23]. It is depressing that even such a monstrously widened scale does not consider many important factors affecting the treatment outcomes, e.g. wishes of patients and their families, malnutrition, past bedridden periods, lifestyle (obesity, alcoholism, nicotine), lost will to live, genetic differences among the populations, equipment and organisational structure of a particular ITU or hospital and skills of the therapeutic personnel [24, 25].

The major potential danger while using the scoring scales is their incorrect interpretation. The scales should not be used to assess the risk of death of a particular patient. The calculated probability of death should regard the entire population of patients, as the scale will never show who is going to survive or die in a particular group [24]. It is emphasized that prognosis for a given patient is based on clinical and not statistical assessment [26]. The prognostic scales used in intensive therapy should be thus considered the tools helping the clinicians [11, 24] as they enable the objective assessment of the patient's condition (severity of a disease) and reflect only the probability of death in a similar group of patients. To calculate the probability of death based on APACHE IV, it is necessary to establish the diagnosis on admission to ITU. The diagnosis is subjectively chosen out of four hundred and twenty-six diagnoses collected in 20 groups. The choice may be extremely difficult and troublesome in cases of multiple organ failure. The optimal choice requires experience and knowledge of all details of scales. It should be remembered that the diagnosis decides about the coefficient, based on which the risk of death is determined.

Other causes of incorrect prognoses of death involve imprecise definitions or non-uniform interpretation of definitions by researchers. Some studies demonstrated possible different interpretations of data introduced to APACHE II, which leads to 15% differences in the final scoring [27]. In this respect, the MPM II model is much less sensitive, as even marked differences in the data introduced do not reduce high consistency of predicting the risk of death [28].

Paradoxically, the scales predicting the risk of death in patients admitted to ITUs based on "physiological vital parameters" are likely to lead to serious bias. The values of these parameters can change due to resuscitations preceding the ITU hospitalization. As a result, the scoring of risk of death is lower, although the patient's condition on admission is still critical. This phenomenon is called "lead-time bias". The study carried out in the group of 76 patients revealed that the assessment of death risk based on the data preceding the ITU admission improved its accuracy. Moreover, it was demonstrated that incorrect assessment of death risk is particularly dependent on the pre-admission correction of the following parameters: heart rate, arterial blood pressure, breathing rate, oxygenation, pH and blood glucose concentration [29].

A similar source of bias distorting the final prognosis of death is the Boud and Ground effect [30]. This results from the fact that the data about the general patient condition are gathered during the first hours of ITU stay, when the physiological parameters may reach their extreme values due to numerous procedures performed. Furthermore, it is stressed that the systems monitoring vital parameters tend to record the extreme values, which is also likely to lead to the Boud and Ground effect [31]. Consequently, the risk of death is artificially overestimated.

In cases of inter-hospital transport of patients, the hospital mortality as a target point should be accepted with great caution, as it is difficult to follow what was going on with the patient. The longer the ITU stay, the less predictive the scales concerning the risk of death based on ITU admission data are [32]. Therefore, the risk of death for such patients should be assessed using the scales based on parameters calculated in the successive days of hospitalisation [33].

The main cause of death of ITU patients is multiple organ failure [34]. The majority of scales do not consider progressive multiple organ failure developing since the first day of hospitalisation. The only exception is the MPM scale involving the second and/or third day of ITU stay.

The interpretation of the risk of death based on general scales is difficult and unsuccessful, therefore the multiple organ failure scales were prepared, i.e. the Sepsis-related Organ Failure Score (SOFA), Multiple Organ Dysfunction Score (MODS), Logistic Organ Dysfunction System (LODS). Everyday assessment of multiple organ failure provides many interesting details, which may be of a prognostic value [35]. However, the prognostic capacities of these three scales were not found to be markedly clinically different [36]. Moreover, the usefulness of LODS to estimate the risk of death of ITU patients was not confirmed [37]; the scales assessing multiple organ failure were not designed to determine the severity of the patient's condition. Their everyday use is grounded only when they reflect the changes in the insufficiency of a particular organ [38].

It should be remembered that the scales might overestimate the risk of death prognosis in some clinical conditions. For instance, in cases of diabetic ketoacidosis or prolonged action of general anaesthetics, the score on ITU admission is high, even though the conditions are potentially reversible and their effect on the mortality is low. This shortcoming was corrected in the new versions of scoring scales [11]. Moreover, the prognosis concerning mortality should consider the fact that the function determining the scale values is not linear, thus claiming that the risk of death at the score of 20 is doubled compared to the score of 10 is a misconception [11]. Furthermore, each scale has its time window, during which the patient admitted to ITU is assessed. The assessment carried out beyond the time window, e.g. on the second day of ITU stay, is likely to lead to misguided conclusions [25].

The common use of prognostic scales is profoundly limited by high costs of data collection.

Irrespective of the complexity of scoring forms, their use facilitates the clinical decision-making. Moreover, the scoring assessment of the severity of patients' conditions enables the control of the quality of ITU treatment and its effective management.

## REFERENCES

1. *Moreno R, Miranda DR, Matos R, Fevereiro T*: Mortality after discharge from intensive care: the impact of organ system failure and nursing workload use at discharge. *Inten Care Med* 2001; 27: 999-1004.
2. *Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE*: APACHE – acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981; 9: 591-597.
3. *Le Gall J-R, Loirat P, Alperovitch A*: Simplified acute physiological score for intensive care patients. *Lancet* 1983; 8352: 741.
4. *Lemeshow S, Teres D, Avrunin JS, Gage RW*: Refining intensive care unit outcome prediction by using changing probabilities of mortality. *Crit Care Med* 1988; 16: 470-477.
5. *Knaus WA, Draper EA, Wagner DP, Zimmerman JE*: APACHE II: a severity of disease classification system. *Crit Care Med* 1985; 13: 818-829.
6. *Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T*: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100: 1619-1636.
7. *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.
8. *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: 2478-2486.
9. *Castella X, Artigas A, Bion J, Kari A*: The European/North American Severity Study Group. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. *Crit Care Med* 1995; 23: 1327-1335.
10. *Bertolini G, D'Amico R, Apolone G, Cattaneo A, Ravizza A, Iapichino G, Brazzi L, Melotti RM*: Predicting outcome in the intensive care unit using scoring systems: is new better? A comparison of SAPS and SAPS II in a cohort of 1.393 patients. *Med Care* 1998; 36: 1371-1382.
11. *Bouch DCh, Thompson JP*: Severity scoring systems in the critically ill. *Continuing Education in Anaesthesia, Critical Care & Pain* 2008; 5: 181-185.
12. *Mourouga P, Goldfrad C, Rowan KM*: Does it fit? Is it good? Assessment of scoring systems. *Curr Opin Crit Care* 2000; 6: 176-180.
13. *Strand K, Flaatten H*: Severity Scoring in the ICU: a review. *Acta Anaesthesiol Scand* 2008; 52: 467-478.
14. *Moreno R, Reis Miranda D, Fidler V, Van Schilfgaarde R*: Evaluation of two outcome predictors on an independent database. *Crit Care Med* 1998; 26: 50-61.
15. *Metnitz PG, Valentin A, Vesely H, Alberti C, Lang T, Lenz K, Steltzer H, Hiesmayr M*: Prognostic performance and

- customization of the SAPS II: results of a multicenter Austrian study. *Inten Care Med* 1999; 25: 192-197.
16. Rowan K: The reliability of case mix measurements in intensive care. *Curr Opin Crit Care* 1996 2: 209-213.
  17. Fery-Lemmonier E, Landais P, Kleinknecht D, Brivet F: Evaluation of severity scoring systems in the ICUs: translation, conversion and definitions ambiguities as a source of inter-observer variability in APACHE II, SAPS and OSF. *Inten Care Med* 1995; 21: 356-360.
  18. Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin J, Teres D: Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med* 1996; 24: 57-63.
  19. Alberti C, Brun-Buisson C, Burchardi H, Martin C, Goodman S, Artigas A, Sicignano A, Palazzo M, Moreno R, Boulme R, Lepage E, Le Gall R: Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. *Intensive Care Med* 2002; 28: 108-121.
  20. Azoulay E, Alberti C, Legendre I, Buisson CB, Le Gall JR: Post-ICU mortality in critically ill infected patients: an international study. *Intensive Care Med* 2005; 31: 56-63.
  21. Pollack MM, Alexander SR, Clarke N Ruttimann UE, Tesselaar HM, Bachulis AC: Improved outcomes from tertiary center pediatric intensive care: a statewide comparison of tertiary and nontertiary care facilities. *Crit Care Med* 1990; 19:150-159.
  22. Luaces O, Taboada F, Albaiceta GM, Domínguez LA, Enríquez P, Bahamonde A: GRECIA Group: Predicting the probability of survival in intensive care unit patients from a small number of variables and training examples. *Artif Intell Med* 2009; 45: 63-76.
  23. 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: 1319-1327.
  24. Le Gall JR: Modeling the severity of illness of ICU patients. A systems update. *JAMA* 1994; 272: 1049-1055.
  25. Le Gall JR: The use of severity scoring systems in the intensive care unit. *Inten Care Med* 2005; 31: 1618-1623.
  26. Flaatten H: Prognostic scoring systems in the ICU. *Acta Anaesthesiol Scand* 2006; 50: 1175-1176.
  27. Polderman KH, Christiaans HM, Wester JP, Spijkstra JJ, Girbes AR: Intraobserver variability in APACHE II scoring. *Inten Care Med* 2001; 27: 1550-1552.
  28. Rue M, Valero C, Quintana S, Artigas A, Alvarez M: Interobserver variability of the measurement of the mortality probability models (MPM II) in the assessment of severity of illness. *Inten Care Med* 2000; 26: 286-291.
  29. Tunnell RD, Millar BW, Smith GB: The effect of lead time bias on severity of illness scoring, mortality prediction and standardized mortality ratio in intensive care – a pilot study. *Anaesthesia* 1998; 53: 1045-1053.
  30. Boyd O, Grounds M: Can standardized mortality ratio be used to compare quality of intensive care unit performance? *Crit Care Med* 1994; 22: 1706-1709.
  31. Bosman RJ, Oudemane van Straaten HM, Zandstra DF: The use of intensive care information systems alters outcome prediction. *Inten Care Med* 1998; 24: 953-958.
  32. Sicignano A, Carozzi C, Giudici D, Merli G, Arlati S, Pulici M: The influence of length of stay in ICU on power of discrimination of a multipurpose severity score (SAPS). *Inten Care Med* 1996; 22: 1048-1051.
  33. Wagner DP, Knaus WA, Harrell FE, Zimmerman JE, Watts C: Daily prognostic estimates for critically ill adults in intensive care units: results from a prospective, multicenter, inception cohort analysis. *Crit Care Med* 1994; 22: 1359-1372.
  34. Tran DD, Groeneveld ABJ, van der Meulen J, Nauta JJ, Strack van Schijndel RJ, Thijs LG: Age, chronic disease, sepsis, organ system failure, and mortality in a medical intensive care unit. *Crit Care Med* 1990; 18: 474-479.
  35. Ulvik A, Wentzel-Larsen T, Flaatten H: Trauma patients in the intensive care unit: short- and long-term survival and predictors of 30-day mortality. *Acta Anaesthesiol Scand* 2007; 51: 171-177.
  36. Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K: Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006; 34: 1378-1388.
  37. Metnitz PG, Lang T, Valentin A, Hiesmayr M, Metnitz PG: Evaluation of the logistic organ dysfunction system for the assessment of organ dysfunction and mortality in critically ill patients. *Inten Care Med* 2001; 27: 992-998.
  38. Laurila J, Laurila PA, Saarnio J, Koivukangas V, Syrjälä H, Ala-Kokko TI: Organ system dysfunction following open cholecystectomy for acute acalculous cholecystitis in critically ill patients. *Acta Anaesthesiol Scand* 2006; 5: 173-179.

received: 01.09.2011  
accepted: 27.12.2011

address:

\*Grzegorz Niewiński  
II Klinika Anestezjologii i Intensywnej Terapii  
Warszawski Uniwersytet Medyczny  
ul. Banacha 1a, 02-097 Warszawa  
tel.: 22 599 20 02