CRITICAL CARE

Clinical review: Scoring systems in the critically ill

Jean-Louis Vincent*[1] and Rui Moreno[2]

Abstract

General illness severity scores are widely used in the ICU to predict outcome, characterize disease severity and degree of organ dysfunction, and assess resource use. In this article we review the most commonly used scoring systems in each of these three groups. We examine the history of the development of the initial major systems in each group, discuss the construction of subsequent versions, and, when available, provide recent comparative data regarding their performance. Importantly, the different types of scores should be seen as complementary, rather than competitive and mutually exclusive. It is possible that their combined use could provide a more accurate indication of disease severity and prognosis. All these scoring systems will need to be updated with time as ICU populations change and new diagnostic, therapeutic and prognostic techniques become available.

Introduction

Scoring systems used in critically ill patients can be broadly divided into those that are specific for an organ or disease (for example, the Glasgow Coma Scale (GCS)) and those that are generic for all ICU patients. In this article, we focus on the generic scores, which can broadly be divided into scores that assess disease severity on admission and use it to predict outcome (for example, Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS), Mortality Probability Model (MPM)), scores that assess the presence and severity of organ dysfunction (for example, Multiple Organ Dysfunction Score (MODS), Sequential Organ Failure Assessment (SOFA)), and scores that assess nursing workload use (for example, Therapeutic Intervention Scoring System (TISS), Nine Equivalents of Nursing Manpower Use Score (NEMS)).

*Correspondence: jlvincen@ulb.ac.be
[1]Department of Intensive Care, Erasme University Hospital, Route de Lennik 808, 1070 Brussels, Belgium
Full list of author information is available at the end of the article

BioMed Central

The objective of this review is to give the intensivist without any particular knowledge or expertise in this area an overview of the current status of these instruments and their possible applications. For a more detailed explanation of the development, application and limitations of these models, the reader is referred to a recent review [1].

Outcome prediction scores

The original outcome prediction scores were developed more than 25 years ago to provide an indication of the risk of death of groups of ICU patients; they were not designed for individual prognostication. Patient demographics, disease prevalence, and intensive care practice have changed considerably since [2], and statistical and computational techniques have also progressed. As a result, all three of the major scores in this category have been recently updated to ensure their continued accuracy in today's ICU (Table 1).

Acute Physiology and Chronic Health Evaluation

The original APACHE score was developed in 1981 to classify groups of patients according to severity of illness and was divided into two sections: a physiology score to assess the degree of acute illness; and a preadmission evaluation to determine the chronic health status of the patient [3]. In 1985, the original model was revised and simplified to create APACHE II [4], now the world's most widely used severity of illness score. In APACHE II, there are just 12 physiological variables, compared to 34 in the original score. The effects of age and chronic health status are incorporated directly into the model, weighted according to their relative impact, to give a single score with a maximum of 71. The worst value recorded during the first 24 hours of a patient's admission to the ICU is used for each physiological variable. The principal diagnosis leading to ICU admission is added as a category weight so that the predicted mortality is computed based on the patient's APACHE II score and their principal diagnosis at admission. The reason for ICU admission is, therefore, an important variable in predicting mortality, even when previous health status and the degree of acute physiological dysfunction are similar.

APACHE III was developed in 1991 [5] and was validated and further updated in 1998 [6]. Equations for

Table 1. Comparison of general outcome prediction models

Characteristics	APACHE [3]	SAPS [10]	APACHE II [4]	MPM[a] [14]	APACHE III [5]	SAPS II [11]	MPM II[b] [15]	SAPS 3 [12]	APACHE IV [8]	MPM III [17]
Year	1981	1984	1985	1985	1991	1993	1993	2005	2006	2007
Countries	1	1	1	1	1	12	12	35	1	1
ICUs	2	8	13	1	40	137	140	303	104	135
Patients	705	679	5,815	2,783	17,440	12,997	19,124	16,784	110,558	124,855
Selection of variables and their weights	Panel of experts	Panel of experts	Panel of experts	Multiple logistic regression	Multiple logistic regression	Multiple logistic regression	Multiple logistic regression	Multiple logistic regression	Multiple logistic regression	Multiple logistic regression
Variables										
Age	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin	No	No	No	No	Yes	No	No	Yes	Yes	No
Surgical status	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Chronic health status	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Physiology	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Acute diagnosis	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Number of variables	34	14	17	11	26	17	15[c]	20	142	16[d]
Score	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Mortality prediction	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[a]These models are based on previous versions, developed by the same authors. [b]The numbers presented are those for the admission component of the model (MPM$_0$ II). [c]MPM$_{24}$ II uses only 13 variables. [d]Plus 7 interaction terms. APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score; MPM, Mortality Probability Model. Adapted from [64] with permission.

predicting risk-adjusted ICU length of stay were also developed using the APACHE III model [7]. Most recently, APACHE IV was developed using a database of over 100,000 patients admitted to 104 ICUs in 45 hospitals in the USA in 2002/2003, and remodeling APACHE III with the same physiological variables and weights but different predictor variables and refined statistical methods [8]. APACHE IV again provides ICU length of stay prediction equations, which can provide benchmarks for the assessment and comparison of ICU efficiency and resource use [9].

Simplified Acute Physiology Score
SAPS, developed and validated in France in 1984, used 13 weighted physiological variables and age to predict risk of death in ICU patients [10]. Like the APACHE scores, SAPS was calculated from the worst values obtained during the first 24 hours of ICU admission. In 1993, Le Gall and colleagues [11] used logistic regression analysis to develop SAPS II, which includes 17 variables: 12 physiological variables, age, type of admission, and 3 variables related to underlying disease. The SAPS II score was validated using data from consecutive admissions to 137 ICUs in 12 countries [11].

In 2005, a completely new SAPS model, the SAPS 3, was created. Complex statistical techniques were used to select and weight variables using a database of 16,784 patients from 303 ICUs in 35 countries [12]. The SAPS 3 score includes 20 variables divided into three subscores related to patient characteristics prior to admission, the circumstance of the admission, and the degree of physiological derangement within 1 hour (in contrast to the 24-hour time window in the SAPS II model) before or after ICU admission. The total score can range from 0 to 217. Unlike the other scores, SAPS 3 includes customized equations for prediction of hospital mortality in seven geographical regions: Australasia; Central, South America; Central, Western Europe; Eastern Europe; North Europe; Southern Europe, Mediterranean; and North America. It should be noted that the sample size for development of some of these equations was relatively small, which may compromise their prognostic accuracy. The SAPS 3 score has been shown to exhibit good discrimination, calibration, and goodness of fit [12]. SAPS 3 has also been used to examine variability in resource use between ICUs using the standardized resource use parameter based on the length of stay in the ICU adjusted for severity of acute illness [13].

Mortality Probability Model
The first MPM, developed from data from patients in one ICU, consisted of an admission model using seven admission variables, and a 24-hour model using seven 24-hour variables [14]. A revised MPM, MPM II, was

developed in 1993 using logistic regression techniques on a large database of 12,610 ICU patients from 12 countries [15]. MPM II also consists of two scores: MPM_0, the admission model, which contains 15 variables; and MPM_{24} the 24-hour model, which contains 5 of the admission variables and 8 additional variables and is designed for patients who stay in the ICU for more than 24 hours. Unlike the APACHE and SAPS systems where variables are weighted, in MPM II each variable (except age, which is entered as the actual age in years), is designated as present or absent and given a score of 1 or 0 accordingly. A logistic regression equation is then used to provide a probability of hospital mortality. The authors also developed a Weighted Hospital Days scale (WHD-94) by subjectively assigning weights to days in the ICU and to hospital days after ICU discharge from the first ICU stay, and an equation to predict an ICU's mean WHD-94, thus providing an index of resource utilization [16].

MPM_0 has recently been updated using a database of 124,885 patients from 135 ICUs in 98 hospitals (all in North America except for one in Brazil) collected in 2001 to 2004 [17]. MPM_0-III uses 16 variables, including 3 physiological parameters, obtained within 1 hour of ICU admission to estimate mortality probability at hospital discharge; the MPM_0 characterization is, therefore, based on patient condition largely before ICU care begins. The WHD-94 predictive equation has also been updated [18].

Discussion

Several studies have compared the different outcome prediction scoring systems. For example, in a study of 10,393 patients from Scottish ICUs, Livingston and colleagues [19] compared the APACHE II and III, an APACHE II using United Kingdom-derived coefficients (UK APACHE II), SAPS II, and MPM_0 and MPM_{24}. These authors reported that all models showed good discrimination, although observed mortality was significantly different from that predicted by all models. SAPS II had the best performance overall, but APACHE II had better calibration. In a retrospective study of 11,300 patients from 35 hospitals in California, Kuzniewicz and colleagues [20] recently used logistic regression to re-estimate the coefficients for the APACHE IV, MPM_0-III and SAPS II scores and applied the new equations to assess risk-adjusted mortality rates. These authors noted that discrimination and calibration were adequate for all models, with discrimination of APACHE IV slightly better than that of the other two scores (area under the receiver operating characteristic curve 0.892 for APACHE IV, 0.873 for SAPS II, and 0.809 for MPM_0 III, $P < 0.001$).

In addition to using a more geographically heterogeneous database for development, the SAPS 3 model attempted to address any geographic variation by providing separate customized equations of different geographical regions. Nevertheless, local customization may still help improve the calibration of these scores in individual countries or regions as demonstrated for the APACHE III in Cleveland, Ohio [21], or more recently for the SAPS 3 score in Austria [22]. In a retrospective analysis of prospectively collected data from a surgical ICU, Sakr and colleagues [23] reported that the discriminative ability of SAPS 3 was similar to that of APACHE II and SAPS II (area under the receiver operating characteristic curve 0.80 for APACHE II, 0.83 for SAPS II, and 0.84 for SAPS 3). All three scores had poor calibration, which improved after customization to the local population. In the UK, investigators have developed a new scoring system specifically for use in UK ICU patients [24]. This score uses elements of the APACHE, SAPS, and MPM systems and was developed using the large Intensive Care National Audit and Research Centre (ICNARC) database and calibrated for adult critically ill patients admitted to ICUs in the UK. It performed better than SAPS II, APACHE II and III, and MPM II [24], but has not been compared to the latest versions of these scores.

When using these instruments, in addition to the issues related to local customization and regular updates discussed above, a few important limitations should be kept in mind. First, all general outcome prediction models can only at their best predict the behavior of a group of patients that exactly matches the patients in the development population. For example, the APACHE and MPM scores were largely based on North American populations and the SAPS score on European patients, while SAPS 3 developers used a database that included a geographically more heterogeneous group of patients [12]. In addition, in most of the scores, specific populations were excluded from the original databases (for example, patients with burns, patients aged less than 16 or 18 years, patients with a very short length of ICU stay, and so on).

Second, the accuracy of any scoring system is highly dependent on the quality of the input. To be used correctly, the definitions, time of data collection, rules for missing data, and so on must exactly match those applied when building the model. The reported reliability of the systems (intra- and inter-observer) must also be taken into account.

Third, there is an inherent bias in many of the derived equations used to predict mortality in that they are created from a limited population of patients from ICUs that are specifically interested in measuring (and improving) ICU performance.

Fourth, the outcome used in all these instruments is the vital status at hospital discharge; consequently, the use of other outcome measures (such as the vital status at

Table 2. Comparison of three organ dysfunction scores

Characteristics	LODS [29]	MODS [30]	SOFA [31]
Year of publication	1996	1995	1996
Selection of variables and their weights	Multiple logistic regression	Literature review and logistic regression	Panel of experts
Variables used to assess organ dysfunction			
Neurologic	Glasgow Coma Scale	Glasgow Coma Scale	Glasgow Coma Scale
Cardiovascular	Heart rate, systolic blood pressure	Pressure-adjusted heart rate	Mean arterial blood pressure, vasopressor use
Renal	Serum urea or urea nitrogen, creatinine, urine output	Serum creatinine	Serum creatinine, urine output
Respiratory	PaO_2/FiO_2 ratio, mechanical ventilation	PaO_2/FiO_2 ratio	PaO_2/FiO_2 ratio, mechanical ventilation
Hematologic	White blood cell count, platelet count	Platelet count	Platelet count
Hepatic	Serum bilirubin, prothrombin time	Serum bilirubin	Serum bilirubin

LODS, Logistic Organ Dysfunction Score; MODS, Multiple Organ Dysfunction Score; SOFA, Sequential Organ Dysfunction Score.

ICU discharge) will compromise the accuracy of the predictive equations. Nevertheless, some models have additional equations to assess use of resources, usually measured as risk-adjusted, weighted, ICU- or hospital days [9,13,18].

Fifth, the statistical methodology used to assess calibration of a predictive model, most commonly the Hosmer-Lemeshow statistic, may be influenced by various factors, including the number of covariates being assessed, the manner in which observations with equal probabilities of outcome are sorted, and the sample size (both small and large) [25]. Interpretation of the accuracy of predictive models should, therefore, include some knowledge of the statistical tests used. Different statistical techniques may be required for the larger models increasingly used to develop predictive models, such as the use of calibration graphs and, more recently, the Cox test of calibration and related statistics [26].

Sixth, despite the fact that predictive models have been developed in large populations, in almost all cases when they are applied to new populations calibration deteriorates, although discrimination hardly changes. Two recent examples of this effect were given in validation studies of SAPS 3 in Austria and in Italy [22,27].

Seventh, the use of automatic patient data management systems can, by changing the sampling rate for the physiological variables, change the accuracy of the model. Bosman and colleagues [28] reported that predicted mortality was greater with data management charting than with manual charting for APACHE II, SAPS II, and MPM II.

Organ dysfunction scores

Organ failure scores are primarily designed to describe the degree of organ dysfunction rather than to predict survival. The severity of organ dysfunction varies widely among individuals and within an individual over time and organ failure scores must be able to take both time and severity into account. Many organ dysfunction scores have been developed over the past few decades, but we will limit our discussion to three of the scores most commonly used in general ICU patients: the Logistic Organ Dysfunction System (LODS) [29], MODS [30], and SOFA [31] (Table 2).

Logistic Organ Dysfunction Score

The LODS was developed using a database of 13,152 admissions to 137 ICUs in 12 countries [29]. Using multiple logistic regression, 12 variables were selected to represent the function of six organ systems (neurologic, cardiovascular, renal, pulmonary, hematologic, hepatic). The worst value for each variable in the first 24 hours of admission is recorded, and for each system, a score of 0 (no dysfunction) to 5 (maximum dysfunction) is awarded. Unlike the MODS and SOFA scores, LODS is a weighted system: for the respiratory and coagulation systems, the maximum score allowed is 3, and for the liver the maximum score is 1. LODS values, therefore, can range from 0 to 22.

The LODS lies somewhere between a mortality prediction score and an organ failure score as it combines a global score summarizing the total degree of organ dysfunction across the organ systems, and a logistic regression equation that can be used to convert the score into a probability of mortality. Within organ systems, greater severity of organ dysfunction was consistently associated with higher mortality [32], and a LODS of 22 was associated with a mortality of 99.7% [29]. The LODS was not initially validated for repeated use during the ICU stay, but in a study of 1,685 patients in French ICUs, the

LODS was accurate in characterizing the progression of organ dysfunction during the first week of ICU stay [33].

Multiple Organ Dysfunction Score

The development of the MODS was based on a literature review of 30 publications that had characterized organ dysfunction [30,34]. Seven organ systems were then selected for further consideration (respiratory, cardiovascular, renal, hepatic, hematological, central nervous system, gastrointestinal), and variables for each organ system were chosen according to a set of 'ideal descriptor' criteria (Table 3). No accurate descriptor of gastrointestinal function could be identified, so this system was not included in the final model. For the cardiovascular system, Marshall and colleagues [30] created a composite variable, the pressure-adjusted heart rate (heart rate × central venous pressure/mean arterial pressure); in patients without a central line, this variable is assumed to be normal. For each of the six organs, the first parameters of the day are used to calculate the score and a score of 0 (normal) to 4 (most dysfunction) is awarded, giving a total maximum score of 24. The score was developed in 336 patients admitted to one surgical ICU and validated in 356 patients admitted to the same ICU [30]. Although not designed to predict ICU mortality, increasing MODS values do correlate with ICU outcome [30]. ICU mortality also increases with increasing numbers of failing organ systems [30,35]. The delta MODS, defined as the difference between the MODS at admission and the maximum score, may be more predictive of outcome than individual scores [30].

Sequential Organ Failure Assessment

The SOFA was developed in 1994 during a consensus conference [31]. Six organ systems (respiratory, cardiovascular, renal, hepatic, central nervous, coagulation) were selected based on a review of the literature, and the function of each is scored from 0 (normal function) to 4 (most abnormal), giving a possible score of 0 to 24. Unlike the MODS score in which the first value of each day is used, for the SOFA score, the worst value on each day is recorded. Another key difference is in the cardiovascular component; instead of the composite variable, the SOFA score uses a treatment-related variable (dose of vasopressor agents). This is not ideal, as treatment protocols vary among institutions, among patients and over time, but it is difficult to avoid, especially for the cardiovascular system.

The SOFA was initially validated in a mixed, medical-surgical ICU population [31,36] and has since been validated and applied in various patient groups [37-39]. In a prospective analysis of 1,449 patients, a maximum total SOFA score greater than 15 correlated with a mortality rate of 90% [40]. Changes in SOFA score over

Table 3. 'Ideal' descriptors of organ dysfunction in ICU patients

Simple and inexpensive
Routinely available in all ICUs
Reliable (intra and inter-observer)
Objective (that is, observer independent)
Specific to the function of the organ in question
Therapy independent
Sequential (available at ICU admission or shortly thereafter and then at fixed periods of time)
Not affected by transient, reversible abnormalities associated with therapeutic or practical interventions
Reflect acute dysfunction of the organ in question but not chronic dysfunction
Reproducible in large, heterogeneous groups of ICU patients
Reproducible in several types of ICUs from different regions of the globe
Abnormal in one direction only
Using continuous rather than dichotomous variables

Modified from [34].

time are also useful in predicting outcome. In a prospective study of 352 ICU patients, an increase in SOFA score during the first 48 hours in the ICU, independent of the initial score, predicted a mortality rate of at least 50%, while a decrease was associated with an ICU mortality rate of just 27% [41]. In a prospective observational study of 1,340 patients with multiple organ dysfunction syndrome, Cabrè and colleagues [42] reported 100% mortality for patients with age over 60 years, a total maximum SOFA greater than 13 on any of the first 5 days of ICU admission, minimum SOFA greater than 10 at all times, and a positive or unchanged SOFA trend over the first 5 days of ICU admission.

Discussion

Several studies have directly compared the various organ dysfunction scoring systems. Pettilä and colleagues [43] reported comparable discriminative power of APACHE III, LODS, SOFA, and MODS to predict hospital mortality in a single center study. Peres Bota and colleagues [44] reported no significant differences between MODS and SOFA for mortality prediction in 949 general ICU patients. However, when using the cardiovascular component, outcome prediction was better for the SOFA score at all time intervals compared to the MODS, a finding confirmed by other studies [45]. In a multicenter study, Timsit and colleagues [33] reported good accuracy and internal consistency for both the SOFA and LODS. However, in a Canadian study of 1,436 ICU patients [45], SOFA and MODS had only a modest ability to discriminate between survivors and non-survivors. More recently, SOFA was reported to have superior discriminative ability for

hospital mortality and unfavorable neurologic outcome compared to MODS in patients with brain injury [46].

Severity assessment based on nursing workload use

The Therapeutic Intervention Scoring System (TISS)

TISS was originally developed in 1974 to assess severity of illness and compare patient care based on the measurement of nursing workload [47]. The original score included 57 therapeutic activities with points assigned for each activity conducted during a 24-hour period; higher values were given for more specialized or time-consuming activities. In 1983, the score was updated and expanded to include 76 items [48]. However, TISS-76 was criticized for being too time-consuming and cumbersome, and in 1996, a simplified version was devised using advanced statistical analysis [49]. TISS-28 includes just 28 items, divided into 7 groups: basic activities, ventilatory support, cardiovascular support, renal support, neurological support, metabolic support, and specific interventions. The scoring is weighted to give a total score of 78. TISS-28 was validated in 22 Dutch ICUs [49] and in 19 ICUs in Portugal [50]. According to this system, each nurse can provide care for 46.35 TISS-28 points per shift, with each TISS-28 point requiring 10.6 minutes of each nurse's shift. This information can be useful for planning the allocation of nursing manpower, to evaluate the efficacy in the use of nursing workload use and to objectively classify ICUs based on the amount (and not the complexity) of care provided [51].

Nine Equivalents of Nursing Manpower Use Score

NEMS was derived from the TISS-28 with the aim of creating a simpler system that would be more widely used [52]. Nursing activities are separated into nine categories: basic monitoring, intravenous medication, mechanical ventilatory support, supplementary ventilatory care, single vasoactive medication, multiple vasoactive medication, dialysis techniques, specific interventions in the ICU, specific interventions outside the ICU. Each of these is awarded weighted points, giving a maximum score of 56. NEMS has been validated in large cohorts of ICU patients and is easy to use with almost no interrater variability [53]. Again, this system can be used to evaluate the efficacy of nursing workload use at the ICU level so as to objectively classify ICUs based on the amount (and not only on the complexity) of care provided [51].

Nursing Activities Score

Based on the TISS-28, the Nursing Activities Score (NAS) includes several additional nursing activities not necessarily related to the severity of illness of the patients [54]. The list of items was developed by consensus. The average time consumption of the activities was determined from a 1-week observational cross-sectional study and the results compared with those of the TISS-28 items in a cohort of 99 ICUs in 15 countries. At the end of this process, a total of five new items and 14 sub-items describing nursing activities in the ICU (for example, monitoring, care of relatives, administrative tasks) were added to the TISS-28 list. The new activities accounted for 60% of the average nursing time; and in the development study, NAS activities accounted for 81% of the nursing time (versus 43% in TISS-28) [54].

Discussion

These scores have been used mainly to assess nurse staffing in the ICU, although higher scores are associated with worse outcomes [55,56]. All the scores are limited by the items included, and can be prone to subjective interpretation and influenced by patient case-mix, local admission and discharge policies, and local management protocols. Use of these scores to compare units may, therefore, be difficult; however, within a unit they can provide a valuable indication of changing workload needs. These scores may also be used to estimate overall costs for groups of ICU patients, although they are less reliable on an individual patient basis [57]. Instruments, such as the Work Utilization Ratio, which evaluates the total number of points actually scored divided by the total possible points, have been proposed to evaluate the effectiveness of the use of nursing workload resources [51]. A recent position statement by the European Federation of Critical Care Nursing Associations recommends that all units use such a system on a regular basis to monitor the efficiency of the use of nursing manpower [58].

Other uses of scoring systems

In addition to their use in outcome prediction, organ function assessment, and nursing workload evaluation discussed above, scoring systems have several other potential uses, including use in clinical trials for case-mix comparisons and use in the assessment and comparison of ICU quality and performance.

Clinical trials

Scoring systems are increasingly being incorporated into clinical trial design. Outcome prediction scores, such as APACHE and SAPS, have been used for some time to compare patient populations in clinical trials and even for the identification of eligible patients for inclusion. The analysis of results from one recent randomized controlled study [59], which showed improved outcomes in patients with higher APACHE II scores, led to the drug under investigation, drotrecogin alfa (activated), being licensed in the United States for use only in patients with severe sepsis who are at a high risk of death, that is, those with

an APACHE II score above 25. However, this is a controversial approach and these scores were not designed for this purpose [60].

The realization that mortality alone is inadequate as an outcome measure for interventional studies in ICU patients has led many trials, especially in sepsis, to include an organ dysfunction score as part of ongoing patient assessment so that effects on morbidity can also be evaluated. Increased economic pressure has also led to greater concerns about cost-effectiveness of new and established interventions and nursing workload scores are also being incorporated into clinical trial design, particularly for interventions likely to impact on nursing workload.

Assessment of ICU performance

Costs of care for an ICU patient have been estimated as being three times the costs of care for a general ward patient [61]. Monitoring ICU performance is, therefore, increasingly important in the fight to control hospital expenses. While crude mortality data may offer some global guidance as to ICU performance, adjusting mortality rates according to disease severity, by using outcome prediction scores to calculate the standardized mortality ratio, can help improve quality assessment. Such severity-adjusted indicators can be used to assess performance of a single ICU over time or to compare several or more units. However, this approach has several limitations, including potential effects of pre-ICU admission factors, implications of different ICU discharge policies [62], and effects of different patient case-mix and hence disease severity between units or in the same unit at different times [63]. Nevertheless, there are large variations in risk-adjusted mortality rates among hospitals [20] and repeated quality assessment may help determine the reasons underlying these differences and enable programs to be developed to improve performance.

Conclusions

General illness severity scores are widely used in the ICU to assess resource use, predict outcome, and characterize disease severity and degree of organ dysfunction. All the scores were developed to be used in mixed groups of ICU patients and their accuracy in subgroups of patients can be questioned; disease-specific scoring systems are increasingly being developed. As ICU populations change and new diagnostic, therapeutic and prognostic techniques become available, all the scoring systems will need to be updated. Importantly, the different scoring systems have different purposes and measure different parameters; we believe they should be seen as complementing each other, rather than competing with one another. For example, outcome prediction models cannot

be used to assess the severity of individual organ dysfunctions or to monitor patient progress over time. Although organ dysfunction scores correlate with outcomes, this is not what they were developed for and outcome prediction should be left to scores such as the APACHE and SAPS systems. The workload scores complete the picture by offering information on how the patient's disease will impact on staffing requirement and resource use. We envisage that, increasingly, all patients will be initially evaluated using a general outcome prediction model computed on admission or within the first 24 hours, then by repeated organ failure (for example, SOFA) and nursing workload (for example, TISS-28) scores during their ICU stay. When used together, these three approaches could provide a more accurate indication of disease severity and prognosis, which could be of help both to the clinician in charge of the patient and to the manager involved in resource allocation and performance assessment.

Abbreviations

APACHE = Acute Physiology and Chronic Health Evaluation; LODS = Logistic Organ Dysfunction Score; MODS = Multiple Organ Dysfunction Score; MPM = Mortality Probability Model; NAS = Nursing Activities Score; NEMS = Nine Equivalents of Nursing Manpower Use Score; SAPS = Simplified Acute Physiology Score; SOFA = Sequential Organ Failure Assessment; TISS = Therapeutic Intervention Scoring System; WHD-94 = Weighted Hospital Days scale.

Author details

[1]Department of Intensive Care, Erasme University Hospital, Route de Lennik 808, 1070 Brussels, Belgium. [2]Department of Intensive Care, Hospital de St Antonio dos Capuchos, Centro Hospitalar de Lisboa Central, EPE, 1169-050 Lisbon, Portugal.

Competing interests

The authors declare that they have no competing interests.

References

1. Moreno RP: **Outcome prediction in intensive care: why we need to reinvent the wheel.** *Curr Opin Crit Care* 2008, **14**:483-484.
2. Moreno R, Jordan B, Metnitz P: **The changing prognostic determinants in the critically ill patient.** In *2007 Yearbook of Intensive care and Emergency Medicine.* Edited by Vincent JL. Heidelberg: Springer; 2007:899-907.
3. 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.
4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: **APACHE II: A severity of disease classification system.** *Crit Care Med* 1985, **13**:818-829.
5. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, Harrell FE: **The APACHE III prognostic system: Risk prediction of hospital mortality for critically ill hospitalized adults.** *Chest* 1991, **100**:1619-1636.
6. Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C, Knaus WA: **Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database.** *Crit Care Med* 1998, **26**:1317-1326.
7. Knaus WA, Wagner DP, Zimmerman JE, Draper EA: **Variations in mortality and length of stay in intensive care units.** *Ann Intern Med* 1993, **118**:753-761.
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**:1297-1310.
9. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL: **Intensive care unit length of stay: Benchmarking based on Acute Physiology and**

Chronic Health Evaluation (APACHE) IV. *Crit Care Med* 2006, **34**:2517-2529.

10. Le Gall J-R, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R: A simplified acute physiology score for ICU patients. *Crit Care Med* 1984, **12**:975-977.

11. Le Gall J-R, 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.

12. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR: 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**:1345-1355.

13. Rothen HU, Stricker K, Einfalt J, Bauer P, Metnitz PG, Moreno RP, Takala J: Variability in outcome and resource use in intensive care units. *Intensive Care Med* 2007, **33**:1329-1336.

14. Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS: A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985, **13**:519-525.

15. 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.

16. Rapoport J, Teres D, Lemeshow S, Gehlbach S: A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study. *Crit Care Med* 1994, **22**:1385-1391.

17. 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**:827-835.

18. Nathanson BH, Higgins TL, Teres D, Copes WS, Kramer A, Stark M: A revised method to assess intensive care unit clinical performance and resource utilization. *Crit Care Med* 2007, **35**:1853-1862.

19. Livingston BM, MacKirdy FN, Howie JC, Jones R, Norrie JD: Assessment of the performance of five intensive care scoring models within a large Scottish database. *Crit Care Med* 2000, **28**:1820-1827.

20. 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.

21. Sirio CA, Shepardson LB, Rotondi AJ, Cooper GS, Angus DC, Harper DL, Rosenthal GE: Community-wide assessment of intensive care outcomes using a physiologically based prognostic measure: implications for critical care delivery from Cleveland Health Quality Choice. *Chest* 1999, **115**:793-801.

22. Metnitz B, Schaden E, Moreno R, Le Gall JR, Bauer P, Metnitz PG: Austrian validation and customization of the SAPS 3 Admission Score. *Intensive Care Med* 2009, **35**:616-622.

23. Sakr Y, Krauss C, Amaral AC, Rea-Neto A, Specht M, Reinhart K, Marx G: Comparison of the performance of SAPS II, SAPS 3, APACHE II, and their customized prognostic models in a surgical intensive care unit. *Br J Anaesth* 2008, **101**:798-803.

24. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K: A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med* 2007, **35**:1091-1098.

25. Kramer AA, Zimmerman JE: Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med* 2007, **35**:2052-2056.

26. Miller ME, Hui SL, Tierney WM: Validation techniques for logistic regression models. *Stat Med* 1991, **10**:1213-1226.

27. Poole D, Rossi C, Anghileri A, Giardino M, Latronico N, Radrizzani D, Langer M, Bertolini G: External validation of the Simplified Acute Physiology Score (SAPS) 3 in a cohort of 28,357 patients from 147 Italian intensive care units. *Intensive Care Med* 2009, **35**:1916-1924.

28. Bosman RJ, Oudemane van Straaten HM, Zandstra DF: The use of intensive care information systems alters outcome prediction. *Intensive Care Med* 1998, **24**:953-958.

29. Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, Teres D, ICU Scoring Group.: The logistic organ dysfunction system: A new way to assess organ dysfunction in the intensive care unit. *JAMA* 1996, **276**:802-810.

30. Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ: Multiple organ dysfunction score: A reliable descriptor of a complex clinical outcome. *Crit Care Med* 1995, **23**:1638-1652.

31. Vincent JL, Moreno R, Takala J, Willatts S, de Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996, **22**:707-710.

32. Metnitz PG, Lang T, Valentin A, Steltzer H, Krenn CG, Le Gall JR: Evaluation of the logistic organ dysfunction system for the assessment of organ dysfunction and mortality in critically ill patients. *Intensive Care Med* 2001, **27**:992-998.

33. Timsit JF, Fosse JP, Troche G, De Lassence A, Alberti C, Garrouste-Org, Bornstain C, Adrie C, Cheval C, Chevret S: Calibration and discrimination by daily Logistic Organ Dysfunction scoring comparatively with daily Sequential Organ Failure Assessment scoring for predicting hospital mortality in critically ill patients. *Crit Care Med* 2002, **30**:2003-2013.

34. Marshall JC: Multiple organ dysfunction syndrome. In *Clincial Trials for the Treatment of Sepsis*. Edited by Sibbald WJ, Vincent JL. Heidelberg: Springer-Verlag; 1995:122-138.

35. Cook R, Cook D, Tilley J, Lee K, Marshall J: Multiple organ dysfunction: baseline and serial component scores. *Crit Care Med* 2001, **29**:2046-2050.

36. Moreno R, Vincent JL, Matos A, de Mendonça A, Cantraine F, Thijs J, Takala J, Sprung C, Antonelli M, Bruining H, Willatts S: The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. *Intensive Care Med* 1999, **25**:686-696.

37. Ceriani R, Mazzoni M, Bortone F, Gandini S, Solinas C, Susini G, Parodi O: Application of the sequential organ failure assessment score to cardiac surgical patients. *Chest* 2003, **123**:1229-1239.

38. Lorente JA, Vallejo A, Galeiras R, Tomicic V, Zamora J, Cerda E, De La Cal MA, Esteban A: Organ dysfunction as estmated by the SOFA score is related to outcome in critically ill burned patients. *Shock* 2009, **31**:125-131.

39. Vosylius S, Sipylaite J, Ivaskevicius J: Sequential organ failure assessment score as the determinant of outcome for patients with severe sepsis. *Croat Med J* 2004, **45**:715-720.

40. Vincent JL, de Mendonça A, Cantraine F, Moreno R, Takala J, Suter P, Sprung C, Colardyn FC, Blecher S: Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a multicentric, prospective study. *Crit Care Med* 1998, **26**:1793-1800.

41. Lopes Ferreira F, Peres Bota D, Bross A, Melot C, Vincent JL: Serial evaluation of the SOFA score to predict outcome. *JAMA* 2001, **286**:1754-1758.

42. Cabrè L, Mancebo J, Solsona JF, Saura P, Gich I, Blanch L, Carrasco G, Martin MC: Multicenter study of the multiple organ dysfunction syndrome in intensive care units: the usefulness of Sequential Organ Failure Assessment scores in decision making. *Intensive Care Med* 2005, **31**:927-933.

43. Pettilä V, Pettila M, Sarna S, Voutilainen P, Takkunen O: Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. *Crit Care Med* 2002, **30**:1705-1711.

44. Peres BD, Melot C, Lopes FF, Nguyen B, V, Vincent JL: The Multiple Organ Dysfunction Score (MODS) versus the Sequential Organ Failure Assessment (SOFA) score in outcome prediction. *Intensive Care Med* 2002, **28**:1619-1624.

45. Zygun DA, Laupland KB, Fick GH, Sandham JD, Doig CJ: Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients. *Can J Anaesth* 2005, **52**:302-308.

46. Zygun D, Berthiaume L, Laupland K, Kortbeek J, Doig C: SOFA is superior to MOD score for the determination of non-neurologic organ dysfunction in patients with severe traumatic brain injury: a cohort study. *Crit Care* 2006, **10**:R115.

47. Cullen DJ, Civetta JM, Briggs BA, Ferrara LC: Therapeutic intervention scoring system: a method for quantitative comparison of patient care. *Crit Care Med* 1974, **2**:57-60.

48. Keene AR, Cullen DJ: Therapeutic intervention scoring system: Update 1983. *Crit Care Med* 1983, **11**:1-3.

49. Miranda DR, de Rijk A, Schaufeli W: Simplified Therapeutic Intervention Scoring System: the TISS-28 items - results from a multicenter study. *Crit Care Med* 1996, **24**:64-73.

50. Moreno R, Morais P: Validation of the simplified therapeutic intervention scoring system on an independent database. *Intensive Care Med* 1997, **23**:640-644.

51. Moreno R, Reis MD: Nursing staff in intensive care in Europe: the mismatch between planning and practice. *Chest* 1998, **113**:752-758.

52. Reis MD, Moreno R, Iapichino G: Nine equivalents of nursing manpower use score (NEMS). *Intensive Care Med* 1997, **23**:760-765.

53. Rothen HU, Kung V, Ryser DH, Zurcher R, Regli B: Validation of "nine

equivalents of nursing manpower use score" on an independent data sample. *Intensive Care Med* 1999, **25**:606-611.

54. Miranda DR, Nap R, de Rijk A, Schaufeli W, Iapichino G: **Nursing activities score.** *Crit Care Med* 2003, **31**:374-382.

55. Padilha KG, Sousa RM, Kimura M, Miyadahira AM, da Cruz DA, Vattimo MF, Fusco SR, de Campos ME, Mendes EM, Mayor ER: **Nursing workload in intensive care units: a study using the Therapeutic Intervention Scoring System-28 (TISS-28).** *Intensive Crit Care Nurs* 2007, **23**:162-169.

56. Padilha KG, de Sousa RM, Queijo AF, Mendes AM, Reis MD: **Nursing Activities Score in the intensive care unit: analysis of the related factors.** *Intensive Crit Care Nurs* 2008, **24**:197-204.

57. Dickie H, Vedio A, Dundas R, Treacher DF, Leach RM: **Relationship between TISS and ICU cost.** *Intensive Care Med* 1998, **24**:1009-1017.

58. European Federation of Critical Care Nursing Associations: **Position Statement on Workforce Requirements Within European Critical Care Nursing.** 2007 [http://www.efccna.org/downloads/Position%20Statement%20Workforce%20EfCCNa%202007.pdf]

59. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, Steingrub JS, Garber GE, Helterbrand JD, Ely EW, Fisher CJ Jr: **Efficacy and safety of recombinant human activated protein C for severe sepsis.** *N Engl J Med* 2001, **344**:699-709.

60. Vincent JL, Opal SM, Marshall JC: **Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions.** *Crit Care Med* 2010, **38**:283-287.

61. Cooper LM, Linde-Zwirble WT: **Medicare intensive care unit use: analysis of incidence, cost, and payment.** *Crit Care Med* 2004, **32**:2247-2253.

62. Kahn JM, Kramer AA, Rubenfeld GD: **Transferring critically ill patients out of hospital improves the standardized mortality ratio: a simulation study.** *Chest* 2007, **131**:68-75.

63. Glance LG, Osler T, Shinozaki T: **Effect of varying the case mix on the standardized mortality ratio and W statistic: A simulation study.** *Chest* 2000, **117**:1112-1117.

64. Moreno R, Metnitz P: **Tools for the evaluation of patients and intensive care units.** In *Principles of Diagnosis and Management in the Adult*. Edited by Parrillo JE, Dellinger RP. Philadelphia: Mosby, Elsevier; 2008:1547-1565.