Predictive scoring systems in the intensive care unit

Author
Mark A Kelley, MD, MACP

Section Editor
Scott Manaker, MD, PhD

Deputy Editor
Geraldine Finlay, MD

Contributor disclosures

All topics are updated as new evidence becomes available and our peer review process is complete.
Literature review current through: Jun 2016. | **This topic last updated:** Mar 17, 2016.

INTRODUCTION — Predictive scoring systems are measures of disease severity that are used to predict outcomes, typically mortality, of patients in the intensive care unit (ICU). Such measurements are helpful for standardizing research and comparing the quality of patient care across ICUs. The common validated predictive scoring systems and their uses in the ICU are described in this topic.

COMPONENTS — Scoring systems are typically developed using prospectively collected data from a large number of patients from several intensive care units (ICUs):

- Data include previous and current clinical health information (eg, comorbidities, admission diagnosis) as well as physiologic and laboratory data (eg, mean arterial pressure, partial pressure of oxygen). (See 'Data collection' below.)
- The data are used to determine a numerical severity of illness score. (See 'Calculations' below.)
- The score, in turn, determines outcomes at hospital discharge including mortality, and sometimes length of stay. (See 'Outcome measures' below.)
- All instruments require validation prior to use. They can only be applied to patients admitted to the ICU since the score can only reliably predict outcomes for populations that were included in the derivation data sets. (See 'Validation and customization' below and 'No generalizability' below.)
- All instruments should be periodically updated as they deteriorate over time due to worsening calibration which, in turn, leads to poor discrimination. (See 'Key characteristics' below.)

Data collection — The individual components and the point in time for data collection vary among the predictive scoring systems, which can lead to differences in efficacy (see 'Predictive scoring systems' below):

- **Variables measured** – The Acute Physiologic and Chronic Health Evaluation (APACHE) scoring system requires collection of a large number of physiological and general health data, while the other instruments use concise and easily measured physiological categories to facilitate data recording [1].
- **Timing of the measurements** – APACHE uses the worst physiologic values measured within 24 hours of admission to the ICU. The updated versions of the Simplified Acute Physiologic Score (SAPS) and the Mortality Prediction Model (MPM_0) use data collected within one hour of ICU admission, while the Sequential (sepsis-related) Organ Failure Assessment (SOFA) instrument uses data collected 24 hours after admission and every 48 hours thereafter. (See 'Acute Physiologic and Chronic Health Evaluation (APACHE)' below and 'Simplified Acute Physiologic Score (SAPS)' below and 'Mortality Prediction Model (MPM0)' below and 'Sequential (sepsis-related) Organ Failure Assessment (SOFA)' below.)

Calculations — All predictive instruments provide a severity score for each patient, which is calculated differently for each instrument:

- For APACHE, the predicted mortality is based upon the sum of categorical variables entered into a computer-generated algorithm. (See 'Acute Physiologic and Chronic Health Evaluation (APACHE)' below.)
- For SAPS and MPM_0, the severity score, which is based upon the sum of the variables, is entered into an equation that calculates a predicted mortality. (See 'Simplified Acute Physiologic Score (SAPS)' below and 'Mortality Prediction Model (MPM0)' below.)

- For SOFA, sequential severity scores plot the trajectory of the clinical course to provide a semiquantitative assessment of mortality, based upon multi-organ failure. (See 'Sequential (sepsis-related) Organ Failure Assessment (SOFA)' below.)

Outcome measures — A predictive scoring system should measure an important outcome(s), typically mortality. All scoring systems should be validated for a specific outcome in the population of interest before use. If they do not accurately predict the outcome of interest, they can be customized to improve performance. (See 'Validation and customization' below.)

Mortality — All ICU scoring systems predict the likelihood of hospital mortality for patients admitted to the critical care unit.

Length of stay — Some investigators have also applied scoring systems to predict length of stay (eg, Acute Physiologic and Chronic Health Evaluation-IV), although not all scores have this ability.

Other — While many scoring systems use a calculated score to quantify disease severity, the SOFA score is also used to facilitate the identification of patients at risk of dying from sepsis. (See 'Sequential (sepsis-related) Organ Failure Assessment (SOFA)' below and "Sepsis syndromes in adults: Epidemiology, definitions, clinical presentation, diagnosis, and prognosis", section on 'Definitions'.)

Validation and customization — Predictive scoring instruments perform best in the population in which the score was originally derived. Thus, many experts propose external validation and customization on an institutional, regional, or national level. As an example, SAPS 3 has several customized equations for seven different geographic regions worldwide [2,3]. Several investigators have performed external validation of the major scoring systems, the details of which are discussed separately. (See 'Predictive scoring systems' below.)

Key characteristics — Accurate discrimination and calibration are two key characteristics that should be met by all predictive scoring systems. Traditionally, calibration can deteriorate over time largely due to the effects of case-mix and altered patient interventions, which often results in an overestimation of mortality, and consequently poor discrimination. (See 'Poor performance over time' below.)

Discrimination — Discrimination describes the accuracy of a given prediction (ie, its ability to discriminate between survivors and non survivors). As an example, if a scoring instrument predicts a mortality of 90 percent, discrimination is perfect if the observed mortality is 90 percent.

Calibration — Calibration describes how the instrument performs over a wide range of predicted mortalities (ie, the agreement between observed and expected numbers of survivors and non survivors across all probabilities of death). In the example above, a predictive instrument would be highly calibrated if it were accurate at mortalities of 90, 50, and 20 percent. Calibration is sensitive to alterations in case-mix and patient care/interventions.

PREDICTIVE SCORING SYSTEMS — Predictive scoring systems are typically used to predict mortality in general intensive care unit (ICU) patients (ie, mixed patient populations). They are occasionally customized to predict outcomes in specific patient populations (eg, liver failure). However, none of these can be used to predict outcomes in a specific individual patient. (See 'General intensive care unit patients' below and 'Specific ICU populations' below and 'Uses' below and 'No generalizability' below.)

General intensive care unit patients — The three major predictive scoring systems used to predict mortality in general ICU patients are the Acute Physiologic and Chronic Health Evaluation (APACHE) scoring system, the Simplified Acute Physiologic Score (SAPS), and the Mortality Prediction Model (MPM_0). APACHE-IV, SAPS 3, and MPM_0-III are third or fourth generation versions that have used data from a larger number of patients and more sophisticated analyses, compared to first and second generation scores.

Acute Physiologic and Chronic Health Evaluation (APACHE) — The APACHE scoring system is widely used in the United States, of which there are four versions (APACHE I through IV) [4-11]. APACHE instruments require the input of a large number of clinical variables, from which a severity score is derived. The resulting severity score is entered into computer-generated logistical regression equation, which predicts hospital mortality and in some cases length of stay. Th required variables differ among the versions, but generally include factors such as age, diagnosis, prior treatment location and numerous acute physiologic and chronic health variables.

APACHE-IV, the latest version, uses 129 variables derived from the worst values from the initial 24 hours of intensive care unit (ICU) admission [10]. Data were generated from over 110,000 patients from 104 ICUs across the United States. Compared to APACHE II and III, APACHE-IV uses a larger set of variables, a new logistical regression equation, and new statistical modeling. The value of this version was illustrated by an observational study of 110,588 consecutive ICU admissions that found that APACHE IV predicted mortality more accurately than APACHE III [10]. A similar study found that APACHE IV predicted ICU length of stay, which is also an advantage compared to APACHE II and III as well as SAPS 3 and MPM0-III [11]. External validation studies also suggest superior discrimination with APACHE-IV when compared with other scoring systems [1,12]. (See 'Discrimination' above.)

- **Advantages** – The major advantages of APACHE-IV is that, similar to other APACHE systems, it is accurate in discriminating between survivors and non survivors. In addition, it can predict length of ICU stay. (See 'Discrimination' above and 'Other' below.)

- **Disadvantages** – When compared to their discriminatory capacity, APACHE systems including APACHE-IV, have less impressive calibration (ie, deteriorating performance over time due to changes in case-mix and new therapies), such that periodic updating is necessary. However, when compared with other predictive scoring systems, it may be less affected by case-mix. (See 'Calibration' above.).

  In addition, the burden of data entry for APACHE instruments is significant since, compared with other instruments, the number of input variables is high, particularly when the collection of data is manual [1]. This burden may lessen as electronic records systems can automatically extract the relevant data.

  Another cited disadvantage is that the data are derived only from ICUs within the United States and therefore, may not be accurate for patients in other countries.

Although APACHE-IV is the most up-to-date version, some centers still use older versions including APACHE II (calculator 1).

Simplified Acute Physiologic Score (SAPS) — The Simplified Acute Physiologic Score instruments (SAPS 1, SAPS 2 (table 1), SAPS 3 (table 2)) calculate a severity score using the worst values measured within the first hour of admission to the ICU [2,3,13-19]. Several of the variables are dichotomous, meaning that they are assigned a specific value depending upon whether or not they are either present or absent. The others are continuous (eg, age range) and points are assigned to ranges of values. All scores are entered into a mathematical formula, which predicts hospital mortality. The number of variables used is significantly smaller than the APACHE scores (eg, SAPS 3 uses 20 variables).

The latest version, SAPS 3, was designed and validated using data from almost 20,000 patients admitted to over 300 ICUs in 35 countries [2,3,17,20]. External validation in other ICU populations reported that SAPS 3 had good discrimination, but poorer calibration, especially when compared with APACHE-IV and MPM_0-III [18-22]. Additional external validation in an Italian ICU suggested that SAPS 3 over predicted mortality [22]. Nonetheless, it may be less affected by interventions than some of the other predictive scores. (See 'Discrimination' above and 'Calibration' above.)

- **Advantages** – The main advantages of SAPS instruments are that data extraction is easier when compared with APACHE, such that they are consequently less labor-intensive. In fact, an automated version that can extract data from the electronic record (also known as eSAPS 3) has also been reported [20]. In addition, it may have greater potential for international use, since data are derived from more than one country.

- **Disadvantages** – Although SAPS 3 does not predict the length of stay (unlike APACHE IV), it has been used as a tool in studies comparing resource use between ICUs [23].

Mortality Prediction Model (MPM0) — There are three versions of the Mortality Prediction Model (MPM_0-I, MPM_0-II, MPM_0-III) [24-27]. A severity score is calculated from variables, as assessed at the time of ICU admission (hence the term "0"). Except for age, all of the variables are dichotomous (ie, either present or absent). As an example, a systolic blood pressure ≤90 mmHg is worth one point, while all other systolic blood pressure values are assigned zero points. The final score is entered into a mathematical formula whose solution provides the predicted mortality.

While we prefer the use of MPM_0-III, MPM_0-II is more widely used (table 3) and a score can be calculated at this site: http://intensivecarenetwork.com/Calculators/Files/Mpm.html. The MPM_0-II severity score can be re-measured serially (eg 24 and 72 hours) allowing it be compared to the SAPS, APACHE, and Sequential (sepsis-related) Organ Failure Assessment (SOFA), all of which have incorporated disease progression in to their predictions (calculator 2) (table 4). The

latest version, MPM_0-III was derived from over 124,000 patients in 135 ICUs, most of which were located in the United States [27]. The MPM_0-III, has good discrimination and calibration, and has been externally validated in additional ICU populations [1,26,27]. There is some evidence that MPM_0-III provides more accurate prediction of ICU mortality, than MPM_0-II [27].

- **Advantages** – MPM systems have the lowest extraction burden among the three major predictive scoring systems (APACHE, SAPS, MPM) since they do not use laboratory data but rather clinical and physiologic data only.
- **Disadvantages** – MPM doesn't typically predict length of stay but older versions have been customized by some investigators for this purpose [24]. It may be more susceptible to poor performance due to the effects of case-mix over time than the other predictive instruments (APACHE, SAPS). It may be less accurate in ICUs outside the United States.

Sequential (sepsis-related) Organ Failure Assessment (SOFA) — The SOFA score was initially designed to sequentially assess the severity of organ dysfunction in patients who were critically ill from sepsis. The original SOFA instrument was derived from a cohort of 1449 patients admitted to 40 ICUs in 16 countries [28]. Since multiple organ dysfunction is common in critically ill patients, it has since been used to predict mortality in those with organ failure from other causes including those with acute liver failure from acetaminophen overdose, chronic liver failure (CLIF-SOFA) and, cancer, as well as in patients who have undergone cardiac surgery or hematopoietic stem cell transplant [29-34].

SOFA uses simple measurements of major organ function to calculate a severity score, which is available at the following site http://clincalc.com/IcuMortality/SOFA.aspx. The scores are calculated 24 hours after admission to the ICU and every 48 hours thereafter (thus, the term "Sequential" Organ Failure Assessment). The mean and the highest scores are most predictive of mortality. In addition, scores that increase by about 30 percent are associated with a mortality of at least 50 percent [35].

The SOFA severity score is based upon the following measurements of organ function (http://clincalc.com/IcuMortality/SOFA.aspx):

- Respiratory system – the ratio of arterial oxygen tension to fraction of inspired oxygen (PaO_2/FiO_2)
- Cardiovascular system – the amount of vasoactive medication necessary to prevent hypotension
- Hepatic system – the bilirubin level
- Coagulation system – the platelet concentration
- Neurologic system – the Glasgow coma score
- Renal system – the serum creatinine or urine output

The SOFA score has been endorsed by the Society of Critical Care Medicine (SCCM) and the European Society of Intensive Care Medicine (ESICM) as a tool to facilitate the identification of patients at risk of dying from sepsis [36-38]:

- **Sepsis** – Sepsis is now defined as life-threatening organ dysfunction caused by a dysregulated host response to infection. As an organ dysfunction score, SOFA can be used to identify those whose organ dysfunction is "life-threatening" such that an increase in the SOFA score ≥2 is associated with a mortality of ≥10 percent.
- **Septic shock** – Patients with a SOFA score ≥2 who also have a vasopressor requirement and an elevated lactate >2 mmol/L (>18 mg/dL) despite adequate fluid resuscitation have a predicted mortality of 40 percent.

The validity of this score was derived from millions of ICU electronic health record encounters both inside and outside the United States. Among critically ill patients with suspected sepsis, the predictive validity of the SOFA score for in-hospital mortality was superior to that for the systemic inflammatory response criteria (SIRS; area under the receiver operating characteristic curve 0.74 versus 0.64).

Importantly, despite the SCCM/ESICM endorsement of the SOFA score, many experts caution clinicians regarding the use of SOFA. The SOFA score does not diagnose sepsis, identify those whose organ dysfunction is truly due to infectior or determine individual treatment strategies or individual outcome. Rather, the SOFA score helps identify patients, as a group, who potentially have a high risk of death from infection. The definition and management of sepsis is discussed in detail separately. (See "Sepsis syndromes in adults: Epidemiology, definitions, clinical presentation, diagnosis, and

prognosis" and "Evaluation and management of suspected sepsis and septic shock in adults".)

The quick SOFA (qSOFA) score has also been proposed by the SCCM/ESICM [36-38] as a tool to help identify patients with early sepsis outside of the ICU. Patients are assigned one point each for the following clinical features which can be easily measured at the bedside: respiratory rate ≥22/min, altered mentation, and systolic blood pressure ≤100 mmHg. Patients with two or more of these features were reported to have a poor outcomes from sepsis. The qSOFA requires additional prospective validation before it can be routinely used to identify those at risk of death from sepsis outside the ICU.

Specific ICU populations — Predictive scoring systems are occasionally customized (institutional, regional, national) or generated to predict mortality in a specific population of critically ill patients (eg, cardiac surgery patients). Other than the SOFA score, their use is not widespread.

Sepsis — The SOFA score, which is frequently used in patients with organ failure from sepsis, is discussed above. (See 'Sequential (sepsis-related) Organ Failure Assessment (SOFA)' above.)

Other — APACE, MPM_0, SAPS 3, and SOFA have been evaluated and found to have mixed performance in the following specific subgroups of critically ill patients [29-33,39-46]:

- Patients with cancer and solid organ transplant
- Patients with acute kidney injury and on renal replacement therapy
- Patients with respiratory failure receiving extracorporeal membrane oxygenation (ECMO) therapy
- Patients with acute coronary syndrome, post cardiac surgery, and cardiac arrest
- Patients with acute and chronic liver disease

In addition, specialized scoring systems created for specific populations have also been reported including the Global Registry of Acute Coronary Events [47], the European System for Cardiac Operative Risk Evaluation (EUROSCORE2) [48], and the Cardiac Surgery Score (CASUS) [48].

Common scoring systems are available at this site: http://intensivecarenetwork.com/124-icu-calculators/

CHOOSING A PREDICTIVE SCORING SYSTEM — No single instrument has convincing or proven superiority to another in its ability to predict death. When choosing a predictive scoring system, it is important to use a score that was developed and validated recently and one that accurately predicts the outcomes in the population of interest (institutionally, regionally, or nationally) (see 'Comparative efficacy' below). Other factors that should be taken in to consideration include feasibility, ease of use, and cost (see 'Ease of use' below and 'Availability' below). Importantly, once chosen, the instrument should be updated periodically to reflect contemporary practice and patient demographics, thereby avoiding deterioration in performance over time (ie, worsening discrimination and calibration). (See 'Key characteristics' above.)

Typically, most intensive care units (ICUs) and researchers use general predictive scoring systems, including Acute physiologic and Chronic Health Evaluation (APACHE), Simplified Acute Physiologic Score (SAPS), Mortality Prediction Model (MPM_0), and Sequential (sepsis-related) Organ Failure Assessment (SOFA). However, clinicians should be aware of what the limitations are in a specific population of interest. As an example, SAPS 3 may underestimate mortality in patients with cancer and solid organ transplant, and SOFA may be more useful in a population with sepsis [39,49-51]. (See 'General intensive care unit patients' above and 'Limitations' below.).

Alternatively, general instruments can be customized to a specific population (institutional, regional, national) or specialized scoring systems can be used for the same purpose. However, none of the specialized instruments are widely or routinely used and we believe that they should only be used when necessary (eg, when a specific population comprises a large percentage of the total ICU population or when a quality improvement program is in place for a specific ICU population). (See 'Other' above and 'Specific ICU populations' above.)

In general, APACHE-IV tends to be more accurate for predicting mortality than the other instruments, but is more burdensome, dependent upon proprietary software, and is more costly. In addition, APACHE-IV can provide algorithms to predict length of stay and is less affected by the case-mix but may not be applied to ICUs outside the United States. In contrast, the other systems are easier to use, cheaper, can be applied internationally (SAPS) but do not provide reliable prediction of LOS, and are more susceptible to case-mix effects. The advantages and disadvantages of three of the most commonly used scoring systems, APACHE, MPM_0, and SAPS are discussed separately. (See 'Acute Physiologic and

Chronic Health Evaluation (APACHE)' above and 'Simplified Acute Physiologic Score (SAPS)' above and 'Mortality Prediction Model (MPM0)' above and 'Sequential (sepsis-related) Organ Failure Assessment (SOFA)' above.)

Comparative efficacy — Large randomized trials demonstrating superiority of one predictive system over another are lacking [1,19,52-54].

- One retrospective study of 11,300 ICU patients from 35 hospitals compared the MPM_0-III, SAPS 2, and APACHE-IV instruments [1]. APACHE-IV offered the best predictive accuracy. However, MPM_0-III proved to be an effective alternative when cost and the complexity of data collection were considered [1].
- A systemic review of the SOFA, SAPS 2, APACHE-II, and APACHE-III scoring systems found that the APACHE systems were slightly superior to the SAPS 2 and SOFA systems in predicting ICU mortality [52]. The accuracy of both the SAPS 2 and APACHE instruments improved when combined with the assessment of SOFA scores.
- In another study of nearly 6000 patients from three Brazilian ICUs, the three models studied, APACHE-IV, SAPS 3, and MPM_0-III, showed good discrimination for all models but poor calibration [19]. APACHE IV showed better discrimination than SAPS 3 and MPM_0-III.
- In another study, APACHE-IV had superior discrimination and calibration compared to MPM_0-III [55].

Ease of use — Developing predictive scoring systems involves extracting data that are either manually or electronically entered in to a database, which can be time-consuming. Thus, in choosing an instrument, many experts prefer one that is not labor-intensive and is easy to use.

While some of these scoring systems are publically available for hospitals and clinicians to use, they sometimes require further customization on an institutional, regional, or national basis, further adding to the associated labor and cost.

APACHE systems are more burdensome as they collect more data and the software is specialized, while SAPS 3, MPM_0-III, and SOFA are easier to use since they extract less data. The development of electronic records may help reduce this burden.

Availability — In the past, most of these scores and predictive calculations were not readily accessible to clinicians at the bedside. However, now many of them are publicly available (eg, at this site http://intensivecarenetwork.com/124-icu-calculators/).

USES — Although predictive scores are of little assistance to the management of individual patients, they can be used by researchers in clinical trials to ensure similar baseline risks between comparative groups, and by institutions and healthcare administrative officials to examine intensive care unit (ICU) performance [56].

Research — Severity scores are most commonly used by researchers to facilitate the evaluation of various interventions by ensuring that patients with similar baseline risk are being compared [57-61]. This is particularly common among trials of potential therapies for sepsis [58,59] or acute respiratory distress syndrome [60,61].

Quality care bench marks — Severity scores facilitate the evaluation of the quality of care by ensuring that patients with similar baseline risk of death are being compared. As examples, studies that compared open with closed ICUs [62-67], the outcomes of different ICUs within a hospital [68], and the outcomes of ICUs in different hospitals [68-70] have used predictive scoring systems (usually Acute Physiologic and Chronic Health Evaluation [APACHE]) for this reason.

Similarly, severity scores have been used to manage hospital resources, assigning patients with lower severity scores to less expensive settings [71].

LIMITATIONS — The intensive care unit (ICU) is the ideal setting for predictive scoring systems because the population is well-defined, patient care is well-circumscribed, and the severity of illness in the ICU is the major determinant of hospital mortality. Despite this, predictive scoring systems have important limitations [72,73].

No generalizability — These instruments were developed from, and validated in, patients admitted to ICUs across many institutions. However, the scoring systems have not been validated in other hospitalized patients and do not apply to any other population, other than those admitted to a critical care setting (eg, patients with stem cell transplant, patients in high dependency units).

Similarly, these scoring systems do not accurately assess outcomes associated with specific diseases (obstetric patient

[74-76]. This limitation may also apply to specialized ICUs with high populations of patients with a specific disease (eg, patients with hematopoietic stem cell transplant) or specific institutions or regions with a high percentage of patients with a specific disease (eg, substance abuse, transplant); in such cases some investigators and institutions have used their own validation subsets to adjust for this flaw [77]. (See 'Validation and customization' above and 'Specific ICU populations' above.)

Poor performance over time — Performances of instruments deteriorate over time. This is known as worsening of discrimination and/or calibration. Failure to update predictive scoring systems can, in particular, lead to gradual loss of calibration. This results in overestimating mortality for any given severity score [19]. Thus, periodically updating a predictive system to reflect contemporary practice and patient demographics is critical for maintaining accuracy. (See 'Key characteristics' above.)

Lead-time bias — Patients who are transferred from other hospitals and ICUs have a mortality that is higher than predicted by the APACHE-II system, a phenomenon known as lead-time bias [5]. To address this flaw, treatment location was added as a variable to subsequent APACHE scores. It is uncertain how lead-time bias affects the other predictive scoring systems.

SUMMARY AND RECOMMENDATIONS

- Predictive scoring systems are measures of disease severity that are used to predict outcomes, typically mortality, of patient populations in the intensive care unit (ICU). They are not useful to predict outcomes in a single individual. (See 'Introduction' above.)
- A numerical severity of illness score is typically developed using prospectively collected data from a large number of patients from several ICUs. The score, in turn, determines outcomes at hospital discharge including mortality, and sometimes length of stay. All instruments require validation and can only be applied to ICU patients. (See 'Components' above.)
- The four major ICU predictive scoring systems are the Acute Physiologic and Chronic Health Evaluation (APACHE) scoring system, Simplified Acute Physiologic Score (SAPS), Mortality Prediction Model (MPM_0), and Sepsis-related (or Sequential) Organ Failure Assessment (SOFA). All have been validated and determined to be reliable for patients in the ICU. In addition, the SOFA score has been used as a tool to facilitate the identification of patient populations at risk of dying from sepsis. These instruments can also be customized for specific populations, although this is not widely done. (See 'Predictive scoring systems' above.)
- No single instrument has convincing or proven superiority to another in its ability to predict mortality, although APACHE systems tend to be more accurate than others. When choosing a predictive scoring system, factors that should be taken in to consideration include performance in the population of interest, feasibility, ease of use, and availability. Importantly, once chosen, the instrument should be updated periodically to reflect contemporary practice and patient demographics, thereby avoiding deteriorating performance over time. (See 'Choosing a predictive scoring system' above.)
- Although predictive scores are of little assistance to the management of individual patients, they can be used by researchers in clinical trials to ensure similar baseline risks between comparative groups, and by institutions and healthcare administrative officials to examine ICU performance. Predictive scoring systems have important limitations including poor generalizability, deterioration with time, and possibly lead-time bias. (See 'Uses' above and 'Limitations' above.)

REFERENCES

1. Kuzniewicz MW, Vasilevskis EE, Lane R, et al. Variation in ICU risk-adjusted mortality: impact of methods of assessment and potential confounders. Chest 2008; 133:1319.
2. Moreno RP, Metnitz PG, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive cai unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med 2005; 31:1345.

3. Metnitz PG, Moreno RP, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005; 31:1336.
4. Ho KM, Dobb GJ, Knuiman M, et al. A comparison of admission and worst 24-hour Acute Physiology and Chronic Health Evaluation II scores in predicting hospital mortality: a retrospective cohort study. Crit Care 2006; 10:R4.
5. Escarce JJ, Kelley MA. Admission source to the medical intensive care unit predicts hospital death independent of APACHE II score. JAMA 1990; 264:2389.
6. Capuzzo M, Valpondi V, Sgarbi A, et al. Validation of severity scoring systems SAPS II and APACHE II in a single-center population. Intensive Care Med 2000; 26:1779.
7. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991; 100:1619.
8. Wagner DP, Knaus WA, Harrell FE, et al. 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.
9. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. JAMA 1995; 274:1591.
10. 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.
11. Zimmerman JE, Kramer AA, McNair DS, et al. Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. Crit Care Med 2006; 34:2517.
12. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, et al. 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.
13. 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.
14. Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. Crit Care Med 1995; 23:1327.
15. Auriant I, Vinatier I, Thaler F, et al. Simplified acute physiology score II for measuring severity of illness in intermediate care units. Crit Care Med 1998; 26:1368.
16. Metnitz PG, Valentin A, Vesely H, et al. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Simplified Acute Physiology Score. Intensive Care Med 1999; 25:192.
17. Ledoux D, Canivet JL, Preiser JC, et al. SAPS 3 admission score: an external validation in a general intensive care population. Intensive Care Med 2008; 34:1873.
18. Poole D, Rossi C, Anghileri A, et al. 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.
19. Nassar AP Jr, Mocelin AO, Nunes AL, et al. Caution when using prognostic models: a prospective comparison of 3 recent prognostic models. J Crit Care 2012; 27:423.e1.
20. Liu V, Turk BJ, Ragins AI, et al. An electronic Simplified Acute Physiology Score-based risk adjustment score for critical illness in an integrated healthcare system. Crit Care Med 2013; 41:41.
21. Metnitz B, Schaden E, Moreno R, et al. Austrian validation and customization of the SAPS 3 Admission Score. Intensive Care Med 2009; 35:616.
22. Poole D, Rossi C, Latronico N, et al. Comparison between SAPS II and SAPS 3 in predicting hospital mortality in a cohort of 103 Italian ICUs. Is new always better? Intensive Care Med 2012; 38:1280.
23. Rothen HU, Stricker K, Einfalt J, et al. Variability in outcome and resource use in intensive care units. Intensive Care Med 2007; 33:1329.
24. Lemeshow S, Teres D, Klar J, et al. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. JAMA 1993; 270:2478.
25. Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. A systems update. JAMA 1994; 272:1049.
26. Higgins TL, Kramer AA, Nathanson BH, et al. Prospective validation of the intensive care unit admission Mortality Probability Model (MPM0-III). Crit Care Med 2009; 37:1619.
27. Higgins TL, Teres D, Copes WS, et al. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med 2007; 35:827.
28. Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-

related problems" of the European Society of Intensive Care Medicine. Crit Care Med 1998; 26:1793.

29. ffrench-O'Carroll R, Frohlich S, Murphy N, Conlon N. Predictors of outcome in decompensated liver disease: validation of the SOFA-L score. Ir Med J 2015; 108:114.
30. Pan HC, Jenq CC, Tsai MH, et al. Scoring systems for 6-month mortality in critically ill cirrhotic patients: a prospective analysis of chronic liver failure - sequential organ failure assessment score (CLIF-SOFA). Aliment Pharmacol Ther 2014; 40:1056.
31. Craig DG, Reid TW, Wright EC, et al. The sequential organ failure assessment (SOFA) score is prognostically superior to the model for end-stage liver disease (MELD) and MELD variants following paracetamol (acetaminophen) overdose. Aliment Pharmacol Ther 2012; 35:705.
32. Badreldin AM, Doerr F, Ismail MM, et al. Comparison between Sequential Organ Failure Assessment score (SOFA) and Cardiac Surgery Score (CASUS) for mortality prediction after cardiac surgery. Thorac Cardiovasc Surg 2012; 60:35.
33. Namendys-Silva SA, Texcocano-Becerra J, Herrera-Gómez A. Application of the Sequential Organ Failure Assessment (SOFA) score to patients with cancer admitted to the intensive care unit. Am J Hosp Palliat Care 2009; 26:341.
34. Gilli K, Remberger M, Hjelmqvist H, et al. Sequential Organ Failure Assessment predicts the outcome of SCT recipients admitted to intensive care unit. Bone Marrow Transplant 2010; 45:682.
35. Ferreira FL, Bota DP, Bross A, et al. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001; 286:1754.
36. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016; 315:801.
37. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016; 315:762.
38. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016; 315:775.
39. Costa e Silva VT, de Castro I, Liaño F, et al. 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.
40. Soares M, Salluh JI. Validation of the SAPS 3 admission prognostic model in patients with cancer in need of intensive care. Intensive Care Med 2006; 32:1839.
41. Oliveira VM, Brauner JS, Rodrigues Filho E, et al. Is SAPS 3 better than APACHE II at predicting mortality in critically ill transplant patients? Clinics (Sao Paulo) 2013; 68:153.
42. Tsai CW, Lin YF, Wu VC, et al. SAPS 3 at dialysis commencement is predictive of hospital mortality in patients supported by extracorporeal membrane oxygenation and acute dialysis. Eur J Cardiothorac Surg 2008; 34:1158.
43. Doerr F, Badreldin AM, Can F, et al. SAPS 3 is not superior to SAPS 2 in cardiac surgery patients. Scand Cardiovasc J 2014; 48:111.
44. Khwannimit B, Bhurayanontachai R. A comparison of the performance of Simplified Acute Physiology Score 3 with old standard severity scores and customized scores in a mixed medical-coronary care unit. Minerva Anestesiol 2011; 77:305.
45. Salciccioli JD, Cristia C, Chase M, et al. Performance of SAPS II and SAPS III scores in post-cardiac arrest. Minerva Anestesiol 2012; 78:1341.
46. Muller G, Flecher E, Lebreton G, et al. The ENCOURAGE mortality risk score and analysis of long-term outcomes after VA-ECMO for acute myocardial infarction with cardiogenic shock. Intensive Care Med 2016; 42:370.
47. Nassar Junior AP, Mocelin AO, Andrade FM, et al. SAPS 3, APACHE IV or GRACE: which score to choose for acute coronary syndrome patients in intensive care units? Sao Paulo Med J 2013; 131:173.
48. Tsaousi GG, Pitsis AA, Ioannidis GD, et al. Implementation of EuroSCORE II as an adjunct to APACHE II model and SOFA score, for refining the prognostic accuracy in cardiac surgical patients. J Cardiovasc Surg (Torino) 2015; 56:919.
49. Soares M, Silva UV, Teles JM, et al. Validation of four prognostic scores in patients with cancer admitted to Brazilian intensive care units: results from a prospective multicenter study. Intensive Care Med 2010; 36:1188.
50. Wu VC, Tsai HB, Yeh YC, et al. Patients supported by extracorporeal membrane oxygenation and acute dialysis: acute physiology and chronic health evaluation score in predicting hospital mortality. Artif Organs 2010; 34:828.
51. Maccariello E, Valente C, Nogueira L, et al. SAPS 3 scores at the start of renal replacement therapy predict mortality in critically ill patients with acute kidney injury. Kidney Int 2010; 77:51.

52. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. Crit Care 2008; 12:R161.
53. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. Crit Care Med 2011; 39:163.
54. Juneja D, Singh O, Nasa P, Dang R. Comparison of newer scoring systems with the conventional scoring systems in general intensive care population. Minerva Anestesiol 2012; 78:194.
55. Kramer AA, Higgins TL, Zimmerman JE. Comparison of the Mortality Probability Admission Model III, National Quality Forum, and Acute Physiology and Chronic Health Evaluation IV hospital mortality models: implications for national benchmarking*. Crit Care Med 2014; 42:544.
56. Kollef MH, Schuster DP. Predicting intensive care unit outcome with scoring systems. Underlying concepts and principles. Crit Care Clin 1994; 10:1.
57. 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.
58. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001; 345:1368.
59. Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 2001; 344:699.
60. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. N Engl J Med 2000; 342:1301.
61. Anzueto A, Baughman RP, Guntupalli KK, et al. Aerosolized surfactant in adults with sepsis-induced acute respiratory distress syndrome. Exosurf Acute Respiratory Distress Syndrome Sepsis Study Group. N Engl J Med 1996; 334:1417.
62. Pronovost PJ, Angus DC, Dorman T, et al. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. JAMA 2002; 288:2151.
63. Multz AS, Chalfin DB, Samson IM, et al. A "closed" medical intensive care unit (MICU) improves resource utilization when compared with an "open" MICU. Am J Respir Crit Care Med 1998; 157:1468.
64. Dimick JB, Pronovost PJ, Heitmiller RF, Lipsett PA. Intensive care unit physician staffing is associated with decreased length of stay, hospital cost, and complications after esophageal resection. Crit Care Med 2001; 29:753.
65. Li TC, Phillips MC, Shaw L, et al. On-site physician staffing in a community hospital intensive care unit. Impact on test and procedure use and on patient outcome. JAMA 1984; 252:2023.
66. Brown JJ, Sullivan G. Effect on ICU mortality of a full-time critical care specialist. Chest 1989; 96:127.
67. Carson SS, Stocking C, Podsadecki T, et al. Effects of organizational change in the medical intensive care unit of a teaching hospital: a comparison of 'open' and 'closed' formats. JAMA 1996; 276:322.
68. Afessa B, Keegan MT, Hubmayr RD, et al. Evaluating the performance of an institution using an intensive care unit benchmark. Mayo Clin Proc 2005; 80:174.
69. Zimmerman JE, Alzola C, Von Rueden KT. The use of benchmarking to identify top performing critical care units: a preliminary assessment of their policies and practices. J Crit Care 2003; 18:76.
70. Zimmerman JE, Shortell SM, Knaus WA, et al. Value and cost of teaching hospitals: a prospective, multicenter, inception cohort study. Crit Care Med 1993; 21:1432.
71. Zimmerman JE, Wagner DP, Knaus WA, et al. The use of risk predictions to identify candidates for intermediate care units. Implications for intensive care utilization and cost. Chest 1995; 108:490.
72. Katsaragakis S, Papadimitropoulos K, Antonakis P, et al. 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.
73. Patel PA, Grant BJ. Application of mortality prediction systems to individual intensive care units. Intensive Care Med 1999; 25:977.
74. Barie PS, Hydo LJ, Fischer E. Comparison of APACHE II and III scoring systems for mortality prediction in critical surgical illness. Arch Surg 1995; 130:77.
75. Brown MC, Crede WB. Predictive ability of acute physiology and chronic health evaluation II scoring applied to human immunodeficiency virus-positive patients. Crit Care Med 1995; 23:848.
76. Lewinsohn G, Herman A, Leonov Y, Klinowski E. Critically ill obstetrical patients: outcome and predictability. Crit Care Med 1994; 22:1412.
77. Sakr Y, Krauss C, Amaral AC, et al. 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.

Topic 1655 Version 12.0