Severity of illness scoring systems in the intensive care unit

Mark T. Keegan, MB, MRCPI; Ognjen Gajic, MD; Bekele Afessa, MD

Objective: **Adult intensive care unit prognostic models have been used for predicting patient outcome for three decades. The goal of this review is to describe the different versions of the main adult intensive care unit prognostic models and discuss their potential roles.**

Data Source: **PubMed search and review of the relevant medical literature.**

Summary: **The main prognostic models for assessing the overall severity of illness in critically ill adults are Acute Physiology and Chronic Health Evaluation, Simplified Acute Physiology Score, and Mortality Probability Model. Simplified Acute Physiology Score and Mortality Probability Model have been updated to their third versions and Acute Physiology and Chronic Health Evaluation to its fourth version. The development of prognostic models is usually followed by internal and external validation and performance assessment. Performance is assessed by area under the receiver operating characteristic curve for discrimination and Hosmer-Lemeshow statistic for calibration. The areas under the receiver operating characteristic curve of Simplified Acute Physiology Score 3, Acute Physiology and Chronic Health Evaluation IV, and Mortality Probability Model$_0$ III were 0.85, 0.88, and 0.82, respectively, and all these three fourth-generation models had good calibration. The models have been extensively used for case-mix adjustment in clinical research and epidemiology, but their role in benchmarking, performance improvement, resource use, and clinical decision support has been less well studied.**

Conclusions: **The fourth-generation Acute Physiology and Chronic Health Evaluation, Simplified Acute Physiology Score 3, Acute Physiology and Chronic Health Evaluation IV, and Mortality Probability Model0 III adult prognostic models, perform well in predicting mortality. Future studies are needed to determine their roles for benchmarking, performance improvement, resource use, and clinical decision support. (Crit Care Med 2011; 39:163–169)**

KEY WORDS: APACHE; benchmarking; critical care; intensive care unit; mortality; outcome assessment; statistical models

The severity of illness assessment scoring systems may be disease- and organ-specific, or global. The first intensive care unit (ICU) prognostic model used to assess patients' overall disease severity was the Therapeutic Intervention Scoring System (1). The Therapeutic Intervention Scoring System was first described as a severity index based on treatment intensity. Its performance (discrimination and calibration) in predicting mortality was not well described. It is a direct measure of treatment intensity, not necessarily disease severity. As a result, its current application is limited to the assessment of workload and resource allocation in the ICU (2, 3). During the last three decades, several physiological-based ICU prognostic models have emerged. The main prognostic models for assessing the overall severity of illness in critically ill adults are Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS), and Mortality Probability Model (MPM) (4–13). Pediatric Risk of Mortality and Pediatric Index of Mortality scores are used in critically ill pediatric patients (14–16). There are also scores primarily designed to describe the degree of organ dysfunction, not survival, in the critically ill (17–20). This review focuses on the global prognostic models used to assess the severity of illness in critically ill adults.

Model Development, Validation, Performance, and Customization. In a recent publication, our group reviewed the basics of model development (4). The commonly used predictor variables include age, comorbidities, physiological abnormalities, acute diagnoses, and lead-time bias. Lead-time bias refers to the inaccuracy in risk prediction that occurs when treatment and measurement occur at different times (21, 22). Lead-time bias has most effect in medical patients and emergency admissions (22). The main outcome measure is usually short-term mortality. The APACHE III and IV models have also included length of ICU and hospital stay and duration of mechanical ventilation (23–26). The relationships between the predictor and outcome variables of the development model need independent validation (27, 28). A mortality prognostic model must differentiate between survivors and nonsurvivors and be well calibrated and reliable (29). It also has to be periodically updated to reflect the change in medical practice and case mix over time (4, 30). The performance of the ICU prognostic models is assessed by the area under the receiver operating characteristic curve for discrimination and the Hosmer-Lemeshow statistic for calibration (4, 28, 31–33). The area under the receiver operating characteristic curve is the measure of how well a model differentiates between groups (4). Area under the receiver operating characteristic curves of 1, 0.90–0.99, 0.80–0.89, 0.70–0.79, 0.60–0.69, and <0.60 are considered to be perfect, excellent, very good, good, moderate, and poor, respectively (4). Calibration refers to the correlation between the predicted and actual outcome for the entire range of risk (34). The calibration is considered good if the Hosmer-Lemeshow statistic p value is $>.05$.

From the Department of Anesthesiology (MTK), Mayo Clinic, Rochester, MN; the Department of Internal Medicine (OG), Division of Critical Care, Mayo Clinic, Rochester, MN; and the Department of Internal Medicine (BA), Division of Pulmonary and Critical Care Medicine, Rochester, MN.

The authors have not disclosed any potential conflicts of interest.

For information regarding this article, E-mail: afessa.bekele@mayo.edu

DOI: 10.1097/CCM.0b013e3181f96f81

Because of changes in case-mix and clinical practice, the performances of prognostic models deteriorate over time (35). To counterbalance the deterioration, models are often customized (36, 37) by adding new predictor variables at times (38, 39). The customized SAPS3 equations are already undergoing further customization for region-specific benchmarking purposes (40, 41). Recent data suggest that MPM_0 III may not be significantly affected by case-mix (42). There are limited data to recommend how often the performance of such models should be assessed. We believe model performance should be assessed periodically and model upgrading may be needed every 4 yrs.

Adult ICU Prognostic Models. The main adult ICU prognostic models include APACHE, SAPS, and MPM (4–13, 43, 44). Recent reviews have addressed the main components of these models (4, 45). In addition to predicting mortality, APACHE III and IV provide predictions for ICU and hospital length of stay, duration of mechanical ventilation, risk of needing an active treatment during the ICU stay, and potential transfer from the ICU (4, 23, 24, 26). A recent multicentered study from California developed ICU length of stay prediction models based on SAPS II, MPM_0 III, and recalibrated APACHE IV length of stay models (46). The study showed that APACHE IV and MPM_0 III were more accurate than SAPS II for predicting ICU length of stay and APACHE IV was the most accurate and best calibrated model.

The history of the current adult ICU prognostic models goes back to development of the original APACHE three decades ago (7). Its second generation, APACHE II, is an ICU prognostic scoring system that is the most widely used in the medical literature (5). APACHE III was narrowly disseminated because of its proprietary nature. SAPS I was developed on data from eight ICUs in France (9) and SAPS II from 137 ICUs in 12 countries (8). The MPM I model was created from a small number of easily available variables from a single medical center (10). Fifteen variables were used in the MPM_0 II (43).

Fourth-Generation Adult ICU Prognostic Models. Studies evaluating the performance of the older generation adult prognostic models showed performance degradation over time manifested by worsening discrimination and calibration of the model (47). This led to the development of fourth-generation adult ICU prognostic models. There are differences among these models. Data for SAPS3 were collected as part of a research project. The data for APACHE IV and MPM III were obtained from ICUs that had bought the APACHE or Project Impact Critical Care systems (both owned by Cerner Corporation, Kansas City, MO). Because institutions that participated in the development of either of these models were not randomly selected, the findings may not apply to other ICUs. Because several medical centers continue to participate in Cerner-owned APACHE and Project Impact activities, upgrades and newer versions are likely to be developed when APACHE IV and MPM III show performance degradation (48).

The calibration of the initial fourth-generation models was good as well as the discrimination. The areas under the receiver operating characteristic curve of SAPS3, APACHE IV, and MPM_0 III were 0.85, 0.88, and 0.82, respectively (11–13, 49).

APACHE IV and MPM_0 III were validated in a multicentered study of 11,300 ICU patients from California (50). APACHE IV had better discrimination and longer data extraction time than MPM_0 III (50). MPM_0 III was recently validated on 55,459 patients from 103 ICUs, 25 of which did not participate in the original development (48). The fact that all three fourth-generation models are free from charge may help their use for research, healthcare delivery, and performance measure. APACHE IV is the most complex and may require software support. MPM_0 III is the least complex.

A recent focus of SAPS3 researchers has been highlighting the need for institutional or regional customization (40, 41, 51–56). All patients included in the development of APACHE IV and MPM III were from the United States. In contrast, patients from five continents were included in the development of SAPS3. With its customized models, SAPS3 appears to be a good candidate for an international benchmark. However, the number of patients included from some of the countries is very small and the results may not be generalizable. There are external SAPS3 validation data from different countries. However, most of these data are limited to small sample size and a narrow patient case-mix (11–13, 49, 51, 52, 54–59). A recent SAPS3 external validation study of 28,357 patients from 147 Italian ICUs showed good discrimination but poor calibration (53). Similar findings were noted in an Austrian multicentered study (41).

Because MPM_0 III and SAPS3 are based on data obtained within 1 hr of ICU admission, they can be used to assess severity of illness before ICU interventions take place. The prognostic models assume missing data as normal, which may adversely affect the performance of the severity scores (60). Because of the multiplicity of data to be collected, missing data may have the highest impact on the performance of APACHE IV (Table 1). The performances of prognostic models in predicting outcome are likely to be compromised by the lack of uniformity in data acquisition (61). Several ICUs use computer interfaces with their laboratory and bedside monitor systems to extract data. Others still enter data manually. SAPS3 was calibrated for manual data acquisition.

All patients included in the development of APACHE IV and MPM III were from the United States. In contrast, patients from five continents were included in the development of SAPS3. However, the number of patients included from some of the countries was very small and the results may not be generalizable.

Model Use. Knowledge about the probability of clinical outcome may provide help to healthcare policymakers, hospital administrators, clinicians, patients, and their families in selecting treatment options. Rising costs of health care and concerns about quality of care have led to efforts aimed at determining outcome associated with medical services. Quality of care can be measured by comparing observed and predicted outcomes. Discordance between the predicted and observed outcomes is considered to indicate better or worse than average quality of care. The ICU adult prognostic models are attractive to measure predicted outcomes in the critically ill. However, the predictor variables should be resistant to manipulation and subjectivity and the models should be reliable and valid before they are applied to assess quality of care (28). There are several factors unrelated to quality of care such as patients' preferences for life support and response to disease, the surrounding environment, and effect of treatment that influence outcome and may have not been included in the prognostic models (47). Most of the prognostic models do not include patients' preferences for life support as a predictor variable. However, the MPM investigators have persistently shown that

Table 1. Variables included in the fourth-generation prognostic models

Predictor Variables	SAPS 3 (11, 12)	APACHE IV (13)	MPM_0 III (49)
Age	Yes	Yes	Yes
Length of hospital stay before ICU admission	Yes	Yes	No
ICU admission source	3	8	No
Type of ICU admission	Yes	Yes	Yes
Chronic comorbidities	6	7	3
CPR before ICU admission	No	No	Yes
Resuscitation status	No	No	Yes
Surgical status at ICU admission	Yes	Yes	No
Anatomic site of surgery	5	No	No
Reasons for ICU admission/acute diagnosis	10	116	5
Acute infection at ICU admission	Yes	No	No
Mechanical ventilation	Yes	Yes	Yes
Vasoactive drug therapy before ICU admission	Yes	No	No
Clinical physiological variables	4	6	3
Laboratory physiological variables	6	10	0

SAPS, Simplified Acute Physiology Score; APACHE, Acute Physiology and Chronic Health Evaluation; MPM, Mortality Probability Model; ICU, intensive care unit; CPR, cardiopulmonary resuscitation.

Modified with permission from Afessa et al (4).

Table 2. Potential uses of adult intensive care unit prognostic models

National level
Benchmarking
Institution level
Internal use of quality improvement efforts such as comparing performance of a hospital against the national average
Institutional self-monitoring for competitive or contractual reasons
Monitoring by regulatory agencies and payors
Adjustment of outcomes in clinical trials
Helping patients select among different hospitals
Physician level
Quality improvement for individual physician
Institution's use of outcome information on individual physicians
Help patients select physician
Use of persistent poor performance for sanctions or license withdrawal
Patient level
Help patients in the decisionmaking process
Resource allocation

the Do-Not-Resuscitate status of a patient influences outcome (10, 44, 49). We have indicated previously that lead-time bias influences patient outcome (21, 22). However, most of the prognostic models do not include any measure of lead-time bias. The specific ICU admission reason is one of the main predictors of outcome. APACHE IV includes 116 ICU admission reasons compared with ten and five for SAPS3 and MPM III, respectively (11–13, 49). Despite their limitations, the predictive models have potential uses at the national, hospital, physician, and patient levels (Table 2) (28).

Benchmarking. Clinicians and investigators need to know why some ICUs save more lives than others (62). Transparency and severity-adjusted data analysis linked to the process of care are likely to lead to this path. The *US News and World Report* has been ranking thousands of hospitals in the United States for several years (63). Thompson Reuters has started reporting the 100 top US hospitals annually (64). Several Internet web sites claim to provide physician ratings. An organization had published ranking of ICUs based on their performances (65). Most of these rankings are based on administrative data. The Centers for Medicare and Medicaid Services is planning to transition from fee-for-service payment systems to value-based purchasing (66, 67). Compared with the severity models derived from administrative data, the ICU adult prognostic models are better tools for risk-adjusted quality assessment. Administrative data do not distinguish between medical conditions present at hospital admission and complications that occur after admission (68, 69). Including complications as pre-existing conditions weakens the performance of a model in risk stratification (68). Compared with models based on administrative data, the ICU severity score models are based on predictor variables available at ICU admission and their performances have been well described. Although not yet implemented, the Joint Commission has a plan requiring hospitals to publicly report their risk-adjusted mortality and length of stay as part of the ICU quality core measures (70). Some states have already started participating in ICU prognostic model severity-adjusted benchmarking (46, 50). Standardized mortality ratio is widely used to evaluate performance. The standardized mortality ratio should be reported with its 95% confidence intervals (71). Similar measures can be established for other outcomes such as length of stay (72). Benchmarking helps to identify variations in clinical outcome and changes in practice patterns over time (73). The appropriate application of benchmarking at the national and community levels may provide reliable information to regulatory agencies, payers, healthcare providers, and patients. However, it requires buy-in by governments, payers, hospitals, healthcare providers, and the public. Over two decades ago, hospitals, physicians, and employers from the Cleveland metropolitan area formed the Cleveland Health Quality Choice to implement a standardized measurement system to evaluate patient outcome in the ICU (73). They used the APACHE III prognostic system for severity adjustment. Although the overall study was compromised by the use of a nonrecalibrated model and changing hospital discharge practices, it highlighted the feasibility of community-based benchmarking. Hospital discharge practices influence hospital mortality. In a multicentered study from California, Vasilevkis et al described the association between acute care hospital transfers and early postdischarge mortality and recommended using the standardized mortality ratio based on 30-day, instead of hospital, mortality (74). Benchmarking provides opportunities to improve performance based on findings from good and bad performers (72, 75–77). Internal benchmarking can also be used to highlight weaknesses and strengths within the same institution (78).

Although hospital mortality rates is usually used to judge hospital performance, it has several weaknesses, including the limitation of the risk adjustment to factors, that are identifiable and measurable (79). In addition to mortality, there are other important outcome measures that can be used for benchmarking. The APACHE prognostic system has models for predicting ICU length of stay (23, 26) and duration of mechanical ventilation (24). The MPM researchers introduced and subsequently revised a two-dimensional graphic tool (Rapoport-Teres graph) for benchmarking performance and resource use (80, 81). The Rapoport-Teres graph is constructed by plotting the normalized differences between actual

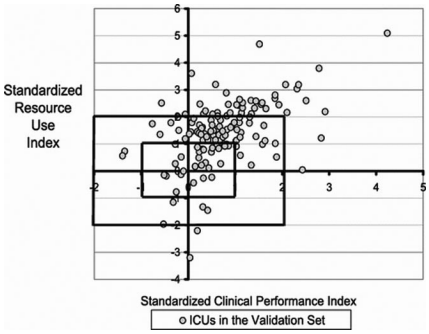

Figure 1. The revised Rapoport-Teres graph for intensive care units in the Project IMPACT validation set. Rectangles mark 1 and 2 standard deviations from the origin. Reprinted with permission from Nathanson et al (81).

and predicted survival rates (standardized clinical performance index) on the x-axis and normalized differences expected and actual weighted hospital days values (standardized resource use index) on the y-axis (80, 81) (Fig. 1). A recent publication reported the standardized hospital length of stay, based on the adult ICU prognostic models, of several hospitals from California (46). The APACHE III database also provides accessories to track low-risk monitor admissions and readmissions (25, 72, 78). Previous ICU benchmark studies based on severity-adjusted outcome have identified policies and practices associated with ICUs that perform well and with good patient outcome: the existence of an alternative to ICU care, a mechanism for improving selection for ICU care, a mechanism to facilitate patient throughput, a mechanism to facilitate ICU discharge planning, reducing excess capacity, matching staffing to workload, process-related guidelines or protocols, care guidelines or protocols for high-volume diagnoses and care processes, performance monitoring and review, and empowering the medical director of the ICU to play an active role (72, 75). Implementation of such policies and practices is likely to improve patient outcome.

Performance Improvement. A well-performing prognostic model helps to make meaningful comparisons of a hospital's current performance with its past. This will allow hospitals to identify their weaknesses and initiate interventions aimed at quality improvement and allow patients and third party payers to choose healthcare providers based on performance. However, changes in case-mix, practice patterns, and other secular trends may also influence differences in outcomes. The ICU prognostic models may facilitate the accreditation process by external organizations. The ICU severity models may also serve as tools for evaluation of the impact of new therapies as well as organizational and process of care changes (72, 75, 78).

The APACHE Critical Care Series and Project Impact have advanced the prognostic models by adding accessories to track readmission, sentinel event, reimbursement, and resource consumption (82). They regularly provide standardized and customized reports of outcome. Based on data from the APACHE III database, Zimmerman et al (72) have highlighted the policies and practices of ICUs with low mortality rate and efficient resource use). They have described the structural characteristics and process of care in ICUs with good performance.

Resource Use. Accurate estimation of severity of illness has the potential to help in the appropriate allocation of scarce ICU resources. With the scarcity of ICU beds in many hospitals, avoiding unnecessary ICU admission and transferring patients who do not need ICU are important. MPM_0 III and SAPS3 have the potential to be used as decision support for ICU admission triage because most of their predictor variables are available at admission (11, 12, 49). The Critical Care Series of the APACHE III clinical support system provides estimates for the risk of requiring specific critical care interventions and potential transfer from the ICU, including providing care in an intermediate unit with reduced cost (25, 82, 83). Using APACHE III data, Seneff et al (24) reported an accurate prediction of the average duration of mechanical ventilation for groups of ICU patients. The MPM researchers developed and updated models for predicting patients' weighted hospital days (80, 81). Such predictions may be useful for resource allocation.

Clinical Decision Support. Probabilities of hospital mortality provide meaningful information to physicians when discussing patient prognosis. However, probabilities should not be used for making treatment decisions at the individual patient level (84). Patient and caregiver preferences as well as their spiritual and cultural beliefs have to be taken into account during the decisionmaking process by patients and family members. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments showed that survival estimates combining an objective prognosis with a physician's clinical estimate had better ability to identify patients with high probabilities of survival or death (85). This can be attributed to the fact that physician estimates of low ICU survival may lead to subsequent life support limitation (86). Currently, most patients and their families rely on prognostic information given to them by the physicians to make decisions. However, because of the biases of subjective estimates, physicians' ability to correctly predict mortality is highly variable (87, 88). Assessment of futility is another important potential application for the use of severity of illness systems. Trends in the severity of illness provide important prognostic information (89). In patients with high risk of death at ICU admission, lack of improvement in severity score indicates poor prognosis (89–91). Awaiting studies addressing their role in improving the clinicians' estimates, the probabilities derived from the prognostic models should be used as "the drunken man uses the lamppost, for support rather than illumination" in making a clinical decision (92). Prognostic models will need to be subjected to the same scrutiny as drugs before they are used in decisions that impact on healthcare delivery and individual patient care.

Although there is scarcity of data supporting the use of severity scores for individual patient care, APACHE II-based administration of activated protein C for severe sepsis/septic shock has become an accepted practice. The initial randomized clinical trial showed that activated protein C reduces the mortality of patients with severe sepsis/septic shock (93). However, *post hoc* analysis suggested this benefit may not extend to patients with lower severity scores. A subsequent clinical trial showed no mortality benefit in patients with APACHE score <25 (94).

Limitations. There are several limitations inherent in the ICU prognostic models (95). Errors in data collection and entry and flaws in model development and validation weaken the performance of prognostic models. All adult ICU severity score models, including the three fourth-generation ones, were developed in nonrandomly selected ICUs, compromising the generalizability of their findings (11–13, 26, 49). Application of prognostic models requires unambiguous definitions of predictor and outcome variables and reproducible measurements easily available in clinical practice (96). Predictor variables may not be easily measured and certain laboratory values may not be rou-

tinely obtained. Lack of standardization in obtaining predictor values leads to missing data, compromising the performance of a model. A few years usually pass between performing a study and publishing its result. By the time the prognostic model studies are published, their prognostic accuracy may have degraded (96). Several factors, including lead-time bias, pre-ICU location, acute diagnosis, physiological reserve, and patients' preferences for life support, influence mortality. Most of these prognostically important variables are not included in some of the latest prognostic models. Although the ICU models perform reasonably well in the general ICU population, they are far from perfect in identifying which individual patient will live or die. Most importantly, long-term survival and quality of life are not forecast by the prediction models.

Future Directions. Prognostic research has received limited attention compared with etiologic, diagnostic, and therapeutic research (97). Data addressing the impact of adult ICU prognostic models on healthcare providers' behavior and patient outcome are scarce. Currently, existing hospital and healthcare provider ranking systems are based on administrative databases and are greatly influenced by the public relation policies of the individual hospital or healthcare provider. The development and application of robust prognostic models are prerequisites for meaningful ranking. The level of clinical detail, ICU-specific diagnoses, and variables make the current adult prognostic models attractive for use in epidemiologic and critical care outcomes research (98).

Future studies are needed to determine the role of the ICU severity scores in clinical practice. The APACHE II, III, and IV models provide risk stratification based on the worst value of the firs 24 hrs (5–7, 13). MPM_0 III and SAPS3 provide risk stratification based on data available within 1 hr of ICU admission (11, 12, 49). If these models can be modified to include values available before ICU admission, they can be incorporated in the ICU admission criteria. Risk stratification based on severity scores can identify patients who are at low risk of mortality and in whom ICU-level life-sustaining interventions are unlikely to be required (25). The identification of such patients can lead to their treatment in the non-ICU setting at reduced cost. The APACHE III and IV prognostic models provide daily risk stratification (6, 13). The potential use of the trend in risk stratification inpatient disposition from the ICU needs further studies (90, 91). Internal and external ICU benchmark studies have shown that differences in policies and practices may partly explain the variations in performance among ICUs (72, 75, 76, 78). The impact of implementing the best policies and practices requires future studies. There is scarcity of data advocating the use of ICU severity scores in selecting treatment for individual patients. The lack of benefit of activated protein C in patients with severe sepsis and APACHE II score <25 highlights the potential role of risk stratification based on severity scores in selecting treatment (94). Before their full potential is realized, future prognostic models will have to include not only the improvement in the assessment of baseline severity of illness, but also the meaningful patient outcome beyond hospital mortality.

REFERENCES

1. Cullen DJ, Civetta JM, Briggs BA, et al: Therapeutic intervention scoring system: A method for quantitative comparison of patient care. *Crit Care Med* 1974; 2:57–60
2. Beck DH, McQuillan P, Smith GB: Waiting for the break of dawn? The effects of discharge time, discharge TISS scores and discharge facility on hospital mortality after intensive care. *Intensive Care Med* 2002; 28: 1287–1293
3. Padilha KG, de Sousa RM, Queijo AF, et al: Nursing Activities Score in the intensive care unit: Analysis of the related factors. *Intensive Crit Care Nurs* 2008; 24:197–204
4. Afessa B, Gajic O, Keegan MT: Severity of illness and organ failure assessment in adult intensive care units. *Crit Care Clin* 2007; 23:639–658
5. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: A severity of disease classification system. *Crit Care Med* 1985; 13: 818–829
6. 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: 1619–1636
7. Knaus WA, Zimmerman JE, Wagner DP, et al: APACHE—Acute Physiology and Chronic Health Evaluation: A physiologically based classification system. *Crit Care Med* 1981; 9:591–597
8. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957–2963
9. Le Gall JR, Loirat P, Alperovitch A, et al: A Simplified Acute Physiology Score for ICU patients. *Crit Care Med* 1984; 12:975–977
10. Lemeshow S, Teres D, Pastides H, et al: A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985; 13:519–525
11. 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; 31:1336–1344
12. Moreno RP, Metnitz PG, Almeida E, et al: 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. Zimmerman JE, Kramer AA, McNair DS, et al: Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34:1297–1310
14. Pollack MM, Ruttimann UE, Getson PR: Pediatric Risk of Mortality (PRISM) score. *Crit Care Med* 1988; 16:1110–1116
15. Pollack MM, Patel KM, Ruttimann UE: PRISM III: An updated Pediatric Risk of Mortality score. *Crit Care Med* 1996; 24:743–752
16. Shann F, Pearson G, Slater A, et al: Paediatric Index of Mortality (PIM): A mortality prediction model for children in intensive care. *Intensive Care Med* 1997; 23:201–207
17. Le Gall JR, Klar J, Lemeshow S, et al: The Logistic Organ Dysfunction system: A new way to assess organ dysfunction in the intensive care unit ICU Scoring Group. *JAMA* 1996; 276:802–810
18. Marshall JC, Cook DJ, Christou NV, et al: Multiple organ dysfunction score: A reliable descriptor of a complex clinical outcome. *Crit Care Med* 1995; 23:1638–1652
19. Vincent JL, de Mendonca 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; 26:1793–1800
20. Vincent JL, Moreno R, Takala J, et al: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; 22:707–710
21. Dragsted L, Jorgensen J, Jensen NH, et al: Interhospital comparisons of patient outcome from intensive care: Importance of lead-time bias. *Crit Care Med* 1989; 17: 418–422
22. Tunnell RD, Millar BW, Smith GB: The effect of lead time bias on severity of illness scoring, mortality prediction and standardised mortality ratio in intensive care —A pilot study. *Anaesthesia* 1998; 53:1045–1053
23. Knaus WA, Wagner DP, Zimmerman JE, et al: Variations in mortality and length of stay

in intensive care units. *Ann Intern Med* 1993; 118:753–761
24. Seneff MG, Zimmerman JE, Knaus WA, et al: Predicting the duration of mechanical ventilation: The importance of disease and patient characteristics. *Chest* 1996; 110:469–479
25. 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; 108:490–499
26. 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; 34:1517–1529
27. Altman DG, Vergouwe Y, Royston P, et al: Prognosis and prognostic research: Validating a prognostic model. *BMJ* 2009; 338:b605
28. Hadron DC, Keeler EB, ROgers WH, Brook RH: Assessing the performance of mortality models. Available at: http://www.rand.org. Accessed May 20, 2010
29. Watts CM, Knaus WA: The case for using objective scoring systems to predict intensive care unit outcome. *Crit Care Clin* 1994; 10: 73–89
30. Higgins TL, Teres D, Nathanson B: Outcome prediction in critical care: The Mortality Probability Models. *Curr Opin Crit Care* 2008; 14:498–505
31. Hanley JA, McNeil BJ: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148:839–843
32. Hosmer DW, Lemeshow S: Assessing the fit of the model. *In*: Applied Logistic Regresson. Hosmer DW, Lemeshow S (Eds). Second Edition. New York, Q Wiley-Interscience Publication, 2000, pp 143–202
33. Royston P, Moons KG, Altman DG, et al: Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009; 338:604
34. Lemeshow S, Hosmer DW Jr: A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 115:92–106
35. Murphy-Filkins R, Teres D, Lemeshow S, et al: Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: How to distinguish a general from a specialty intensive care unit. *Crit Care Med* 1996; 24:1968–1973
36. Beck DH, Smith GB, Pappachan JV: The effects of two methods for customising the original SAPS II model for intensive care patients from South England. *Anaesthesia* 2002; 57:785–793
37. Moreno R, Apolone G: Impact of different customization strategies in the performance of a general severity score. *Crit Care Med* 1997; 25:2001–2008
38. Le Gall JR, Neumann A, Hemery F, et al: Mortality prediction using SAPS II: An update for French intensive care units. *Crit Care* 2005; 9:R645–R652
39. Rivera-Fernandez R, Vazquez-Mata G, Bravo M, et al: The APACHE III prognostic system: Customized mortality predictions for Spanish ICU patients. *Intensive Care Med* 1998; 24:574–581
40. Khwannimit B, Bhurayanontachai R: The performance and customization of SAPS 3 admission score in a Thai medical intensive care unit. *Intensive Care Med* 2010; 36: 342–346
41. Metnitz B, Schaden E, Moreno R, et al: Austrian validation and customization of the SAPS 3 admission score. *Intensive Care Med* 2009; 35:616–622
42. Nathanson BH, Higgins TL, Kramer AA, et al: Subgroup mortality probability models: Are they necessary for specialized intensive care units? *Crit Care Med* 2009; 37:2375–2386
43. Lemeshow S, Klar J, Teres D, et al: Mortality probability models for patients in the intensive care unit for 48 or 72 hours: A prospective, multicenter study. *Crit Care Med* 1994; 22:1351–1358
44. 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; 270:2478–2486
45. Vincent JL, Moreno R: Clinical review: Scoring systems in the critically ill. *Crit Care* 2010; 14:207
46. Vasilevskis EE, Kuzniewicz MW, Cason BA, et al: Mortality Probability Model III and Simplified Acute Physiology Score II: Assessing their value in predicting length of stay and comparison to APACHE IV. *Chest* 2009; 136:89–101
47. Ohno-Machado L, Resnic FS, Matheny ME: Prognosis in critical care. *Annu Rev Biomed Eng* 2006; 8:567–599
48. 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: 1619–1623
49. 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; 35:827–835
50. 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: 1319–1327
51. 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; 34:1873–1877
52. Mbongo CL, Monedero P, Guillen-Grima F, et al: Performance of SAPS3, compared with APACHE II and SOFA, to predict hospital mortality in a general ICU in southern Europe. *Eur J Anaesthesiol* 2009; 26:940–945
53. 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; 35:1916–1924
54. 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; 101:798–803
55. Silva Junior JM, Malbouisson LM, Nuevo HL, et al: Applicability of the Simplified Acute Physiology Score (SAPS 3) in Brazilian hospitals. *Rev Bras Anestesiol* 2010; 60:20–31
56. 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; 32:1839–1844
57. 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; 77:51–56
58. Strand K, Soreide E, Aardal S, et al: A comparison of SAPS II and SAPS 3 in a Norwegian intensive care unit population. *Acta Anaesthesiol Scand* 2009; 53:595–600
59. Capuzzo M, Scaramuzza A, Vaccarini B, et al: Validation of SAPS 3 admission score and comparison with SAPS II. *Acta Anaesthesiol Scand* 2009; 53:589–594
60. Afessa B, Keegan MT, Gajic O, et al: The influence of missing components of the Acute Physiology Score of APACHE III on the measurement of ICU performance. *Intensive Care Med* 2005; 31:1537–1543
61. 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
62. Suter PM: Some ICUs save more lives than others: We need to know why! *Intensive Care Med* 2005; 31:1301–1302
63. America's Best Hospitals. Available at: http://health.usnews.com/best-hospitals. Accessed May 19, 2010
64. 100 Top Hospitals: Healthsystem Quality/Efficiency Benchmarks Study. Available at: http://www.100tophospitals.com/assets/100TopHealthSystem. Accessed May 19, 2010
65. 100 Top Hospitals™: ICU Benchmarks for Success—2000. Available at: http://www.100tophospitals.com. Accessed February 3, 2007
66. US Department of Health and Human Services Development of a plan to transition to a Medicare value-based purchasing program for physician and other professional services. Available at: http://www.cms.gov. Accessed May 19, 2010
67. US Department of Health and Human Services Plan to implement a Medicare value-based purchasing program Available at: http://www. cms.gov. Accessed May 19, 2010
68. Glance LG, Dick AW, Osler TM, et al: Accuracy of hospital report cards based on administrative data. *Health Serv Res* 2006; 41: 1413–1437
69. Glance LG, Dick AW, Osler TM, et al: Does date stamping ICD-9-CM codes increase the value of clinical information in administrative data? *Health Serv Res* 2006; 41:231–251
70. National Hospital Quality Measures—ICU. Available at: http://www.jointcommission.org. Accessed May 19, 2010

71. Hosmer DW, Lemeshow S: Confidence interval estimates of an index of quality performance based on logistic regression models. *Stat Med* 1995; 14:2161–2172
72. 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; 18:76–86
73. Sirio CA, Shepardson LB, Rotondi AJ, et al: 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
74. Vasilevskis EE, Kuzniewicz MW, Dean ML, et al: Relationship between discharge practices and intensive care unit in-hospital mortality performance: Evidence of a discharge bias. *Med Care* 2009; 47:803–812
75. DePorter J: UHC operations improvement: Adult ICU benchmarking project summary University HealthSystem Consortium. *Best Pract Benchmarking Healthc* 1997; 2:147–153
76. Glance LG, Osler TM, Dick AW: Identifying quality outliers in a large, multiple-institution database by using customized versions of the Simplified Acute Physiology Score II and the Mortality Probability Model II0. *Crit Care Med* 2002; 30:1995–2002
77. Glance LG, Osler TM, Dick A: Rating the quality of intensive care units: Is it a function of the intensive care unit scoring system? *Crit Care Med* 2002; 30:1976–1982
78. Afessa B, Keegan MT, Hubmayr RD, et al: Evaluating the performance of an institution using an intensive care unit benchmark. *Mayo Clin Proc* 2005; 80:174–180
79. Lilford R, Mohammed MA, Spiegelhalter D, et al: Use and misuse of process and outcome data in managing performance of acute medical care: Avoiding institutional stigma. *Lancet* 2004; 363:1147–1154
80. Rapoport J, Teres D, Lemeshow S, et al: 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
81. Nathanson BH, Higgins TL, Teres D, et al: A revised method to assess intensive care unit clinical performance and resource utilization. *Crit Care Med* 2007; 35:1853–1862
82. Sakallaris BR, Jastremski CA, Von Rueden KT: Clinical decision support systems for outcome measurement and management. *AACN Clin Issues* 2000; 11:351–362
83. Zimmerman JE, Wagner DP, Draper EA, et al: Improving intensive care unit discharge decisions: Supplementing physician judgment with predictions of next day risk for life support. *Crit Care Med* 1994; 22:1373–1384
84. Teres D, Lemeshow S: Why severity models should be used with caution. *Crit Care Clin* 1994; 10:93–110
85. Knaus WA, Harrell FE Jr, Lynn J, et al: The SUPPORT prognostic model: Objective estimates of survival for seriously ill hospitalized adults: Study to understand prognoses and preferences for outcomes and risks of treatments. *Ann Intern Med* 1995; 122:191–203
86. Rocker G, Cook D, Sjokvist P, et al: Clinician predictions of intensive care unit mortality. *Crit Care Med* 2004; 32:1149–1154
87. Poses RM, McClish DK, Bekes C, et al: Ego bias, reverse ego bias, and physicians' prognostic. *Crit Care Med* 1991; 19:1533–1539
88. Sinuff T, Adhikari NK, Cook DJ, et al: Mortality predictions in the intensive care unit: Comparing physicians with scoring systems. *Crit Care Med* 2006; 34:878–885
89. Chang RW, Jacobs S, Lee B: Predicting outcome among intensive care unit patients using computerised trend analysis of daily APACHE II scores corrected for organ system failure. *Intensive Care Med* 1988; 14: 558–566
90. Afessa B, Keegan MT, Mohammad Z, et al: Identifying potentially ineffective care in the sickest critically ill patients on the third ICU day. *Chest* 2004; 126:1905–1909
91. Fleegler BM, Jackson DK, Turnbull J, et al: Identifying potentially ineffective care in a community hospital. *Crit Care Med* 2002; 30:1803–1807
92. TPN and APACHE. *Lancet* 1986; 1:1478
93. 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; 344:699–709
94. Abraham E, Laterre PF, Garg R, et al: Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med* 2005; 353:1332–1341
95. Cowen JS, Kelley MA: Errors and bias in using predictive scoring systems. *Crit Care Clin* 1994; 10:53–72
96. Moons KG, Altman DG, Vergouwe Y, et al: Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ* 2009; 338:606
97. Moons KG, Royston P, Vergouwe Y, et al: Prognosis and prognostic research: What, why, and how? *BMJ* 2009; 338:b375
98. Rubenfeld GD, Angus DC, Pinsky MR, et al: Outcomes research in critical care: Results of the American Thoracic Society Critical Care Assembly Workshop on Outcomes Research: The members of the Outcomes Research Workshop. *Am J Respir Crit Care Med* 1999; 160:358–367