Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions*

Jean-Louis Vincent, MD, PhD, FCCM; Steven M. Opal, MD; John C. Marshall, MD, FRCSC

Objective: **Severity scores such as Acute Physiology and Chronic Health Evaluation II have been advocated as entry criteria for clinical trials and in clinical decision-making. We present ten reasons why we believe this approach is not appropriate and may even be detrimental.**

Data Sources: **Available relevant literature from authors' personal databases and personal knowledge of past and future clinical trial development.**

Data Synthesis: **Severity scores were not designed for use in individual patients or for therapeutic decision-making for specific interventions. Difficulties with the time window needed to calculate these scores and the need to administer therapies early further limit their use in this context. The complex nature of the scores makes it difficult to use them at the bedside and there is considerable interobserver variability in score calculation. Inclusion of chronic health and age points in severity scores may prevent younger, previously healthy patients, with similar acute physiological dysfunction and therefore total lower severity scores, from trial inclusion or from receiving therapies that may be beneficial.**

Conclusions: **We believe severity of illness scores are poor surrogates for risk stratification and should not be used as a criterion for patient enrollment into clinical trials or as the basis for individual treatment decisions. (Crit Care Med 2009; 38: 000–000)**

KEY WORDS: APACHE II; risk prediction; clinical trial; sepsis studies; organ failure scores; outcomes; critical illness

Whether or not we should use severity scores as an entry criterion for clinical trials or in deciding whether to start a specific therapeutic intervention is a timely question. The rationale for using severity scores in this way is the seemingly logical proposition that the intervention should be limited to the sickest group of patients, who presumably would be the most likely to benefit. This approach has been applied to drugs already available and is currently being used in ongoing trials of new therapeutic agents. One key example of how severity scores have been used in this context is shown in the package insert for drotrecogin alfa (activated), which indicates that the drug should be restricted to the sickest patients, that is, those with an Acute Physiology and Chronic Health Evaluation (APACHE) II score >25. This recommendation followed a subgroup analysis of the data from the initial randomized, controlled clinical trial (PROWESS) (1), which showed that patients with an APACHE II score >25 were more likely to benefit from the drug. Future clinical trials of new therapeutic interventions may limit enrollment to patients with a severity score above a threshold value to maximize the chances of success.

In an analysis of clinical trials of anti-inflammatory agents in patients with sepsis, Eichacker et al reported that treatment efficacy was dependent on the risk of death (2). These authors propose that the pathophysiological events found in the sickest patients closely reflect the pathology found in preclinical animal models of sepsis, in which placebo mortality rates approaching 100% are the norm. However, the assumption that there is a linear relationship between severity and potential to respond to therapy may be too simplistic. Indeed, some interventions may have greater benefit in patients with moderate disease severity. As examples, antibiotics may be more effective in less severely ill patients (3) and in the Transfusion Requirements in Critical Care trial, the benefits of the restrictive transfusion strategy were more marked in the group with low APACHE II scores (4). One could, therefore, consider the relationship between disease severity and potential to respond to an intervention to be bell-shaped rather than linear (Fig. 1*B*). At both extremes of disease severity, the chances of detecting a survival benefit from any new therapy are limited with patients highly likely to survive (low severity) or die (high severity) regardless of the therapeutic intervention. In this case, study enrollment should be limited to a midrange of severity scores to focus on patients who are sick enough to benefit but not so sick that they are about to die. An approach based on this suggested bell-shaped relationship has been proposed in an ongoing sepsis trial and has already been used in some studies. As one example, Gutierrez et al (5), in a study of gastric tonometry, selected patients with an APACHE II score between 15 and 25; they showed that prevention of a decrease in intramucosal pH was associated with reduced mortality rates.

In view of this, we believe that a cutoff severity score threshold should not be used as a criterion for patient inclusion in

*See also p. xxx.

From the Department of Intensive Care (J-LV), Erasme University Hospital, Free University of Brussels, Belgium; Infectious Disease Division (SMO), Memorial Hospital of Rhode Island, The Warren Alpert Medical School of Brown University, Providence, RI; and Department of Surgery and the Interdepartmental Division of Critical Care Medicine (JCM), University of Toronto, St. Michael's Hospital, Toronto, Ontario, Canada.

J-LV has consulted for Eli Lilly, GSK, Artisan, AstraZeneca, Eisai, Wyeth, and NovoNordisk. J-LV has received honoraria/speaking fees from GSK, Eli Lilly, and Roche. J-LV has received grant support from Esai, Artisan, and Eli Lilly. The remaining authors have not disclosed any potential conflicts of interest.

For information regarding this article, E-mail: jlvincen@ulb.ac.be

DOI: 10.1097/CCM.0b013e3181b785a2

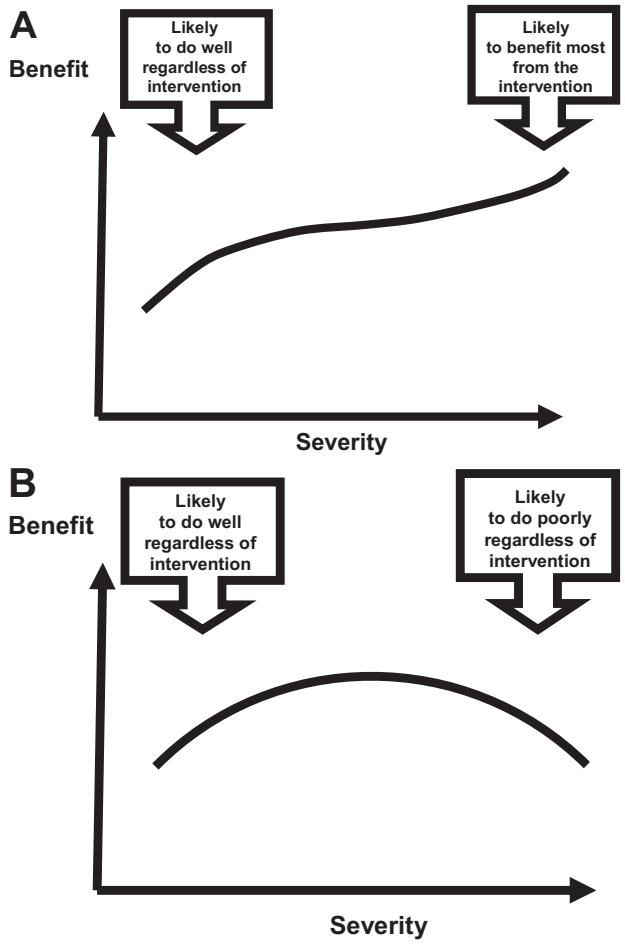

Figure 1. Schematic representation of two models of the potential benefit of a therapeutic intervention versus the severity of the disease state as assessed by a severity score. *A*, A direct correlation between sepsis severity and the potential to respond to therapy. This type of treatment to disease severity interaction corresponds with the findings in the PROWESS study (1), which reported greatest benefit in patients with greatest disease severity. *B*, The chances of deriving a survival benefit from a new treatment intervention in patients with a very low, or very high, disease severity are low in contrast to those patients with moderate disease severity in which the likelihood of treatment benefit is greatest. This model corresponds with the findings from the Transfusion Requirements in Critical Care trial (4), in which benefit was seen in patients who were moderately sick.

clinical trials or to direct therapy and will put forward ten arguments against this approach. Importantly, although we focus on the APACHE II score (6), because it is the most frequently used and is the favored system for assessing mortality risk by the Food and Drug Administration in the United States, these arguments could apply equally to the Simplified Acute Physiology Score (7) or even the Mortality Probability Model (8); more modern scores such as the APACHE 3 or 4 or the Simplified Acute Physiology Score 3 will be confronted with the same problems.

Arguments Against Use of Scoring Systems for Trial Inclusion or to Determine Therapy. Severity scores were not designed for this purpose. APACHE, Simplified Acute Physiology Score, and Mortality Probability Model scores are composite scores developed to assess the risk of hospital death and calibrated to maximize predictive capacity across a spectrum of illnesses rather than to detect differential clinical response of a specific intervention for a particular disease. Despite well-recognized difficulties and potential pitfalls in measurement, severity scores have proved to be remarkably reliable predictors of short-term mortality for cohorts of patients admitted with critical illness. However, selecting patients for a given therapy based on severity of illness cutoff points may exclude a wider group of patients who might benefit. We believe that the decision to treat such patients might be better based on knowledge of potentially reversible pathophysiological processes and measures of organ dysfunction, which on some occasions might be poorly represented by acuity scores. We are interested not only in the risk of death, but even more importantly in the probability that the therapy can favorably alter that risk. Although it is intuitively apparent that the greater the mortality risk, the greater the potential signal for benefit if an intervention is efficacious in that disease, the corollary that mortality risk predicts efficacy, does not follow. There may be better options to predict response to therapy based on the actual pathophysiology of the disease state rather than on a simple severity of illness score. Furthermore, these predictors of response may be specific to the therapy planned. This principle is embedded in contemporary multimodal treatment of cancer in which surgical intervention is generally limited to those patients without evidence of distant spread, adjuvant chemotherapy is given to those at higher risk for distant recurrence (predicted, for example, by the presence of vascular invasion or regional nodal metastases), and salvage therapy is provided only to those with advanced disease.

Generic severity scores are not exclusively measures of the severity of physiological dysfunction but are risk predictors that combine measures of potentially modifiable severity (the acute physiology score) and nonmodifiable patient risk factors of age and comorbidities. The value can be heavily influenced by the latter (up to 11 points can be obtained for age and comorbidities in APACHE II, for example), but this does not necessarily reflect the potential to benefit from the therapeutic intervention. Because age is an important part of the score, when other components of the score are the same, older patients will have higher scores than younger patients. Therefore, older patients are intrinsically more likely than younger patients to receive interventions that are reserved for use only in patients above a certain cutoff score. Furthermore, younger patients are less likely to have chronic health points related to comorbid illnesses. Although chronic health points clearly contribute to the global assessment of intensive care unit mortality risk, it is less clear whether such patients are optimal candidates for therapeutic interventions with experimental agents.

For example, consider two intensive care unit patients. The first is an 82-yr-

old, nonambulatory woman residing in a skilled nursing facility with multiple comorbid illnesses and a minor infection who now develops acute kidney injury. She is admitted to an intensive care unit with a calculated APACHE II score of 28 and a predicted risk of death of approximately 42%. The second patient is a previously healthy 40 yr old who presents with invasive pneumococcal disease complicated by hypotension, acute kidney injury, thrombocytopenia, and lactic acidosis. The risk of death is similar (approximately 42%), but the APACHE II is 21 (normally it is 0 = no age points, no chronic health points, no Glasgow Coma Scale points).

The first patient is a typical patient we see in geriatric medicine with multiple medical illnesses, medications, and complications of disease, age, and therapies. The second patient has a disease process that presents in a similar way to animal models of severe infection and sepsis. Like in animal models, in this case, the patient was previously healthy and became septic as a result of massive pathogen exposure. Both patients might benefit equally well from antimicrobial therapy. The second patient might benefit substantially from an adjuvant anticytokine agent, whereas the first patient might not benefit at all or may be made worse by the same anticytokine agent. Some time after the 28-day study period has concluded, the first patient (if she survives) will eventually go back to the nursing home in a more fragile state with the same pre-existing morbidities and with further loss of residual physiological reserve. The second patient (assuming he survives) may need a long convalescent period but will likely recover eventually, go back to work, and return to society with a good long-term prognosis.

Do these two patients have the same likelihood of response to the experimental therapy? Let us assume that the experimental agent actually works to prevent ongoing inflammation-induced remote organ injury. The first patient has fixed organ disease that will not recover. The second patient might have entirely reversible physiological injury resulting from systemic inflammation and might respond dramatically to the test agent. What if the entry criterion is solely based on an arbitrary cutoff point such as an APACHE II score of 24? The first patient is eligible, but the second patient, who is more likely to respond to the therapy, is not. The study is populated with patients like the first and excludes patients like the second.

Overall mortality risk is likely to be correlated to some degree with responsiveness to an experimental therapy, but a linear observed-to-expected response relationship across a broad range of APACHE II scores is not likely. The acute physiology score component of the APACHE II score might prove to be a better predictor of individual responsiveness with specific therapies than the entire score. This could readily be tested in future clinical trials.

The relationship between severity scores and outcome is not straightforward. Although innovations and advances in the process of supportive care in critical illness have improved outcomes over time, inappropriate applications of severity scores further compromise their predictability. Although a rise in score is often associated with increased hospital mortality, severity scores and their associated calculated risk of death are not synonymous (9). This issue is illustrated by two recent studies with drotrecogin alfa (activated): the PROWESS (1) and the ADDRESS (10) trials show that the same APACHE II scores can result in quite different mortality rates. For patients with an APACHE II score between 25 and 29, the mortality rates were 35.8% in the PROWESS study but only 22.0% in the ADDRESS trial; for patients with APACHE II scores >29, mortality rates were 49% in the PROWESS but only 32.5% in the ADDRESS study. This was consistent with the spirit of the ADDRESS trial; if an investigator thought that the patient was at low risk for death despite a high APACHE II score or multiorgan failure, the protocol permitted enrollment of the patient. Furthermore, because management has improved since the APACHE II score was developed, it may now overestimate expected mortality. Potentially, therefore, the risk of mortality associated with an APACHE II score of >25 may now be less than it was and patients may therefore be allocated to treatments that should in fact now be reserved for patients with a higher risk of death. Recent revisions of the APACHE and Simplified Acute Physiology Score scores have tried to address this issue (11, 12), but frequent calibrations would need to be performed to limit this effect.

The APACHE II score is observer-dependent with several studies showing that different physicians calculate different APACHE II values from the same data (13–15); even the same physician can achieve different scores when reviewing the same data at different times (16). The scores may even be manipulated to fit a required APACHE II cutoff point (15). One of the components of the APACHE II score, the Glasgow Coma Scale score, is also notoriously subjective (17, 18), particularly in patients receiving sedation (19).

Most severity scores are designed to be calculated during the first 24 hrs in the intensive care unit. Many physiological and biochemical abnormalities used in calculating the acute physiology score of APACHE II can be corrected or controlled by expert supportive care measures in the days after intensive care unit admission. How valid an indicator of mortality is the APACHE II score calculated at day 7 in the intensive care unit versus day 1? Age and chronic health points are unchanged, but the acute physiology score component may be markedly different. The use of the APACHE II score when the patient is already in the intensive care unit has never been properly validated, although some studies have suggested using the changes in (delta) APACHE II (20, 21). In the PROWESS/ENHANCE studies, the severity score used was calculated over the 24 hrs before study entry and not within 24 hrs of admission; this is an incorrect application of the APACHE II score.

A related inaccuracy, particularly in scores that include diagnostic category, is that the correct diagnosis may not be immediately obvious. When such scores are used correctly, the diagnostic category should be selected according to the diagnosis made on admission and not the diagnosis, which becomes apparent after several days on the intensive care unit with the benefit of laboratory or imaging results. Again, in the PROWESS/ENHANCE studies, the incorrect timing of the APACHE II score may have affected its diagnostic category component.

The 24-hr window, chosen to collect the worst values over a reasonable time period, can result in artifactual increases or decreases in the score. Increasingly, we are encouraged to start therapies early, even before intensive care unit admission when the patient is still in the emergency room. Acutely ill septic patients who undergo effective early resuscitation such as suggested in the randomized trial conducted by Rivers and colleagues (22) will have lower acute physiology score values than patients who do not undergo early resuscitation as a result of the effective therapy, although

Table 1. Differences in disease severity between patients enrolled in the PROWESS (1) and ENHANCE (23) studies

	PROWESS (treatment)	ENHANCE (treatment)
No. of patients	850	2378
APACHE II score	24.6	22.0
Shock	70%	76%
Vasopressors	61%	74%
Mechanical ventilation	73%	82%
SOFA respiratory 2–4	88.9%	89.6%
SOFA cardiovascular 2–4	68.0%	78.9%
SOFA renal 2–4	34.3%	40.7%
SOFA hepatic 2–4	17.6%	21.2%
SOFA hematology 2–4	22.2%	26.8%
Mean no. of organ failures	2.4	2.7
No. of organ failures		
≥2	75%	84%
≥3	43%	55%

APACHE, Acute Physiology and Chronic Health Evaluation; SOFA, sequential organ failure assessment.

the two groups may in fact have had comparable illness severity on entry into the emergency department. Other scoring systems such as Mortality Probability $Model_0$, which are measured on admission, do not have this bias but are seldom used in the context of clinical trials.

The use of a severity score cutoff as an entry criterion or indication for therapy penalizes for good early care. Values will be influenced by treatments initiated during the 24-hr period and may, therefore, reflect bad (or good) management rather than actual disease severity. Hence, centers with less optimal early care may be more likely to either apply the new intervention (if already available) or to enroll patients in the clinical trial. In the latter case, this may result in more patients being included by centers with poorer performances putting the robustness of the treatment effect at risk.

The calculation of the score typically requires the collection of the worst data obtained during the specified time interval (as indicated previously, optimally 24 hrs). Thus, a longer collection time or more frequent data sampling is more likely to result in a higher score. It may, therefore, be tempting to delay study enrollment and repeat the measurement to increase the score above the given threshold. As an example, one could repeat blood gas measurements to obtain an oxygenation index that would increase the score. Furthermore, this may encourage initiation of treatment to be postponed, although benefit is usually greater when therapy is started early.

Severity scores are influenced by the lead time bias. In particular, the ENHANCE trial (23) has shown that the APACHE II score decreases with time despite similar disease severity. An important difference between ENHANCE and PROWESS (1) was the time delay between the onset of severe sepsis and the administration of the drug; it was longer in ENHANCE than in PROWESS (26.1 vs. 17.5 hrs) as a result of a difference in protocol. However, the criteria used to define organ failure were identical. The APACHE II score was slightly lower in the ENHANCE than in the PROWESS study (22.0 versus 24.6), yet the patients seemed sicker in terms of the higher number of patients with circulatory shock (76% vs. 70%), patients treated by mechanical ventilation (82% vs. 73%), or any organ-related Sequential Organ Failure Assessment score (Table 1).

Furthermore, data were analyzed according to the time delay between the onset of organ failure and the intervention ($<$ or $>$24 hrs); one can anticipate that patients treated later were more severely ill; indeed, one would be less likely to consider enrollment of patients who have already improved substantially than patients who are still critically ill. This observation was supported by the fact that patients treated later were more likely to need vasopressor support, were more commonly treated with mechanical ventilation, and were more likely to have multiple organ failure with a higher number of failing organs and a higher total Sequential Organ Failure Assessment score (10.1 vs. 9.3, $p < 0.001$), but a somewhat lower APACHE II score (21.6 vs. 22.5, $p = 0.01$; Table 2). The most likely explanation is that many elements included in the APACHE II score are corrected with time such as hypotension (masked by vasopressor therapy), metabolic acidosis, and sodium abnormalities. The APACHE II score was developed and calibrated for use within the first 24 hrs of intensive care unit admission. It should still be used in this fashion regardless of the time of enrollment of intensive care unit patients into a clinical trial.

Severity scores were developed for assessment of the mortality risk of populations of patients and not for decisions concerning individuals within those populations. They are particularly useful to compare populations of patients with similar sets of risk factors and still have a place in clinical trials that include several groups of patients to ensure that disease severity is similar among groups and to enable comparability between different studies. It is less clear whether they can be effectively applied to make clinical judgments about treatment options regarding individual patients. Although this does not necessarily invalidate the use of prediction models to assess the risk of individual patients, the confidence interval for predictions for an individual patient will be considerably wider than the confidence interval for the average outcome for a group of patients with identical risk factors. Hence, application of severity scoring systems to individual patients results in frequent misclassification (15), and changing patient mixes can affect the performance of severity scores (24).

Do We Have Alternatives? It is not our intention to criticize severity scores *per se*; they are valuable tools when used for the purpose for which they were developed. Variation of observed mortality rates from predicted outcomes between treatment and placebo groups in a clinical trial, using a correctly calculated APACHE II score from the first 24 hrs in the intensive care unit, can be highly informative. Our aim is to challenge their use, out of their original context, for clinical trial enrollment and therapeutic decision-making. We do not suggest that scores should be replaced by a clinician's opinion or "gestalt." There could be better measures to assess responsiveness to new therapies based on sets of physiologically relevant biomarkers, transcript profiles, or other systems' biology methodologies. Furthermore, it is intuitively plausible that the optimal approach to risk stratification will vary with the therapy under consideration. These questions will need to be evaluated in future clinical trials.

We believe that, at present, clinical judgments on individual patient decisions rest primarily on the physician's reasoning at the bedside, taking into account all the available evidence, especially if specific diagnostic criteria can be identified. When decision tools and technologies have been shown to improve those clinician decisions, they should rapidly be brought to the bedside to improve patient care. Until such time as new methodologies and scoring systems have been clinically validated to improve

Table 2. Differences inpatient characteristics according to length of time after development of first organ dysfunction and before treatment in patients enrolled in the ENHANCE study (23)

	0–24 Hrs	>24 Hrs	*p*
Male, %	55.9	60.2	0.03
Age, years	57.8	60.4	0.004
Surgical, %	36.6	44.9	<0.001
Vasopressors, %	71.2	75.9	0.009
Mechanical ventilation, %	75.4	87.8	<0.001
No. of failing organs	2.6	2.8	<0.001
MOF, %	79.4	89.1	<0.001
SOFA total	9.3	10.1	<0.001
APACHE II	22.5	21.6	0.01

MOF, multiple organ failure; SOFA, sequential organ failure assessment; APACHE, Acute Physiology and Chronic Health Evaluation.

individual patient care ("personalized medicine"), such measures should be considered research methods and not standards of care. In clinical trial enrollment, more objective criteria are needed. For this purpose, organ failure scores may be better than severity scores, although this is not proven and needs to be explored further. The use of organ failure scores can avoid a number of problems related to chronic health points and the 24-hr period, but some of the same limitations would still apply. An alternative would be the use of one or more biomarkers to establish that a process is present and to identify differential potential to respond to treatment. The concept of PIRO (25) is also grounded in the need for better risk stratification.

In clinical practice, cutoff thresholds of scores are never used to decide on a therapeutic intervention, with the possible exception of a Glasgow Coma Scale score <8 to decide whether intracranial pressure monitoring should be started in a patient with severe head injury, but the Glasgow Coma Scale is still relatively simple (calculated mentally), and not the only factor taken into consideration (other elements like the computed tomography scan are also considered). Severity scores cannot be calculated mentally; they ideally need computer support, which although widely available, has not been used clinically. Ultimately, severity scores are surrogates for the true risk factors we are trying to identify. Their validity as surrogate measures for individual risk prediction and as entry criteria for interventional trials can be rigorously tested. The regulatory agencies argue against surrogate outcomes in trials and need, therefore, to recognize that severity scores have the same limitations. More specific entry criteria are needed and will become the standard of care as systems biology becomes an integral part of clinical trials.

REFERENCES

1. 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
2. Eichacker PQ, Parent C, Kalil A, et al: Risk and the efficacy of antiinflammatory agents: Retrospective and confirmatory studies of sepsis. *Am J Respir Crit Care Med* 2002; 166:1197–1205
3. Clec'h C, Timsit JF, De Lassence A, et al: Efficacy of adequate early antibiotic therapy in ventilator-associated pneumonia: Influence of disease severity. *Intensive Care Med* 2004; 30:1327–1333
4. Hebert PC, Wells G, Blajchman MA, et al: A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *N Engl J Med* 1999; 340:409–417
5. Gutierrez G, Palizas F, Doglio G, et al: Gastric intramucosal pH as a therapeutic index of tissue oxygenation in critically ill patients. *Lancet* 1992; 339:195–199
6. Knaus WA, Draper EA, Wagner DP, et al: APACHE II: A severity of disease classification system. *Crit Care Med* 1985; 13: 818–829
7. Le Gall J-R, Lemeshow S, Saulnier F: A new Simplified Acute Physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270: 2957–2963
8. 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
9. Civetta JM: 'New and improved' scoring systems. *Crit Care Med* 1990; 18:1487–1490
10. 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
11. 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
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. Chen LM, Martin CM, Morrison TL, et al: Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit Care Med* 1999; 27: 1999–2004
14. Polderman KH, Jorna EM, Girbes AR: Interobserver variability in APACHE II scoring: Effect of strict guidelines and training. *Intensive Care Med* 2001; 27:1365–1369
15. Booth FV, Short M, Shorr AF, et al: Application of a population-based severity scoring system to individual patients results in frequent misclassification. *Crit Care* 2005; 9:R522–R529
16. Polderman KH, Christiaans HM, Wester JP, et al: Intra-observer variability in APACHE II scoring. *Intensive Care Med* 2001; 27: 1550–1552
17. Rowley G, Fielding K: Reliability and accuracy of the Glasgow Coma Scale with experienced and inexperienced users. *Lancet* 1991; 337:535–538
18. Gill MR, Reiley DG, Green SM: Interrater reliability of Glasgow Coma Scale scores in the emergency department. *Ann Emerg Med* 2004; 43:215–223
19. Livingston BM, Mackenzie SJ, MacKirdy FN, et al: Should the pre-sedation Glasgow Coma Scale value be used when calculating Acute Physiology and Chronic Health Evaluation scores for sedated patients? *Crit Care Med* 2000; 28:389–394
20. Chang RWS, Jacobs S, Lee B: Predicting outcome among intensive care unit patients using computerized trend analysis of daily APACHE II scores corrected for organ system failure. *Intensive Care Med* 1988; 14: 558–566
21. Yzerman EPF, Boelens HAM, Tjhie JHT, et al: Delta APACHE II for predicting course and outcome of *Staphylococcus aureus* bacteremia and its relation to host defense. *J Infect Dis* 1996; 173:914–919
22. 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; 345:1368–1377
23. Vincent JL, Bernard GR, Beale R, et al: Drotrecogin alfa (activated) treatment in severe sepsis from the global open-label trial ENHANCE. *Crit Care Med* 2005; 33: 2266–2277
24. 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
25. Levy MM, Fink MP, Marshall JC, et al: 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003; 31:1250–1256