# Evidence should not be viewed in isolation

Rui Moreno, MD, PhD; Andrew Rhodes, FRCP, FRCA

**After decades of negative multicenter, randomized, placebo-controlled, double-blind clinical trials in patients with critical illness, the last decade witnessed the publication of a series of studies demonstrating an improved outcome for a variety of interventions. These positive studies and interventions have subsequently been brought together into "bundles" of care. This rush to change clinicians' practice has often been with little regard for the context in which the evidence was generated and the problems created by the difference between efficacy and effectiveness. This methodology created a bypass mechanism through which implementation of a package of interventions could be fast-tracked to a much larger group of critically ill patients that differ significantly from those in whom the evidence was generated, without being validated as a bundle. The risk of this process is that the bundle tools will be used for issues such as quality control, performance evaluation, and legal and regulatory purposes without the necessary background of robust validation studies. This has resulted in a "two-weight, two-measures" situation that needs to be resolved to avoid the risk of taking interventions completely out of the context in which the positive results were generated and putting patients and professionals at risk of being harmed by the package of interventions. (Crit Care Med 2010; 38[Suppl.]:S528–S533)**

**Key Words: infection; sepsis; septic shock; clinical trials; bundles of care; methodology; outcomes; intensive care**

> It's more important to know what sort of person this disease has, than what sort of disease this person has.
>
> —William Osler (1849–1919)

The care of critically ill patients is complex. Although many of those patients present with similar syndromic patterns, the pathophysiological processes underpinning these phenotypes are highly complex, in particular in the myriad of ways that these interact with the genotypic makeup of the patients and their often multiple premorbid disease processes. Even when taking all the above into account, the multiple variations in treatment modalities and organ support techniques used on each separate patient makes them highly individual in their own right. This makes our attempts to get homogeneous groupings for comparison extremely difficult. It should therefore not be a surprise to learn that the history of intensive care medicine is one of failed or negative randomized controlled studies.

From the Unidade de Cuidados Intensivos Polivalente (RM), Hospital de St. António dos Capuchos, Centro Hospitalar de Lisboa Central, Lisboa, Portugal; and the Department of Intensive Care Medicine (AR), St George's Healthcare NHS Trust, London, UK.

The authors have not disclosed any potential conflicts of interest.

For information regarding this article, E-mail: r.moreno@mail.telepac.pt

Copyright © 2010 by the Society of Critical Care Medicine and Lippincott Williams & Wilkins

**DOI: 10.1097/CCM.0b013e3181f1cd02**

### *History of trial design in intensive care medicine*

In the 1980s and 1990s, a long series of studies were published with one factor in common: they were all negative. This was especially the case for randomized controlled trials (RCTs) of severe infection, sepsis, and septic shock. The time, efforts, and resources expended on these studies were hugely relevant to our small and relatively impoverished specialty, so this series of "failed" attempts at identifying evidence-based approaches generated both significant controversy and discussion. Fortunately, we learned a lot from these discussions, and the design and process control of RCTs has changed significantly in subsequent decades. Indeed, the debate challenged nearly all of our previously held assumptions underlying the appropriateness and control of phase III, placebo-controlled, double-blind RCTs, especially those using 28-day hospital mortality as a primary end point (1).

The situation changed dramatically in the early 2000s with the publication on March 8, 2001, of the Recombinant Human Activated Protein C Worldwide Evaluation in Severe Sepsis (PROWESS) Study demonstrating a significant reduction in mortality in patients with severe sepsis treated with drotrecogin alfa activated compared with placebo (2). In less than a decade, many new interventions, or new ways of applying old interventions, have been proposed to decrease the mortality and/or morbidity of sepsis. The interventions proposed, based on studies with very different levels of evidence and different methodologic characteristics, include the use of early goal-directed resuscitation of the septic patient during the first 6 hrs after recognition in the emergency department (3), the administration of broad-spectrum antibiotic therapy as soon as possible after diagnosis of infection-related hypotension (if possible within 1 hr of the diagnosis) (4), the use of low-dose steroid therapy given to patients with more severe forms of the syndrome (for example, patients who remain hypotensive despite adequate fluid resuscitation and vasopressors) (5), and the rigid control of glycemia in postoperative patients (6). Many more interventions have been evaluated and published, ranging from ventilatory strategies to renal replacement therapy with different levels of evidence and impact on clinician behavior.

### *The Surviving Sepsis Campaign*

In 2004, the Surviving Sepsis Campaign group published a set of recommendations based on the group analysis of the available literature (7). This effort was then further updated and expanded



Copyright (c) Society of Critical Care Medicine and Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

with new methods to access and weigh the evidence in 2008 (8). Although these recommendations were grounded in evidence-based processes, the publication has proved subsequently to be highly controversial and has further engendered debate that questions not only our ability to perform robust studies, but also our methodologies of appraising and grading evidence to generate rational evidence-based practice guidelines (9). Leaving the question of the origins of, and motivations for funding the most contentious part of the whole Surviving Sepsis Campaign was the translation of the evidence-based recommendations into a "bundle" methodology with the attempt of changing clinicians' behaviors and practices (10). The motivation behind this process change was fundamentally correct; however, the perception of it being used as a marketing tool by industry and the inconsistent levels of evidence that underpinned the individual components of the bundles led to much criticism in academic journals and in the lay press. The publication of these "sepsis" bundles and subsequent attempts to force clinicians to adopt each individual component of the bundle generated the greatest controversy we have seen in our specialty (9). This has been far more heated than previous 'hot topics' such as the importance of tissue oxygen debt as a determinant of postoperative organ failure (11) or the uproar surrounding the linkage of the pulmonary artery catheter to excess mortality (12). The temperature was raised to such an extent that major bodies felt impelled to speak out with pleas for moderation, e,g., the European Society of Intensive Care Medicine (13), the Society of Critical Care Medicine (14), and the Australian and New Zealand Intensive Care Society (15). Central to these debates was the concern held by many about our limited knowledge and expertise in translating data extracted from RCTs into clinical recommendations that could be trusted and relied on.

### *Risk assessment in trial design*

A major issue that trial designers have had to face is how to direct and test a therapy in the most appropriate risk group of patients. Vincent (16) describes a bell-shaped curve whereby a tested treatment is only likely to benefit patients within the middle part of the risk–benefit ratio. If the risk is too high, the patient will probably die anyway; too low and it is unlikely that the treatment effect can be proved in a conventionally sized trial. Our ability to predict the correct risk grouping of patients within a trial setting is thus vitally important and sadly our skills, although improving, are still lacking.

Traditionally researchers have used general severity scores and outcome prediction models such as the Acute Physiology and Chronic Health Evaluation (APACHE) model (17) or the Simplified Acute Physiology Score (18) for risk assessment in a trial setting. We believe that this is fundamentally flawed. APACHE II was originally designed to assess the risk of a patient dying within the hospital inpatient episode. It was not designed to assess survival at a discrete time point such as 28 days, which is what many studies have as their primary end point. This difference may only be small but fundamentally affects the logistically derived variables within the equation. In other words, the risk assessment is likely to be imprecise. This has been clearly demonstrated by us when comparing predictions based on logistic regression models with those based on multilevel Cox models, the former aiming to forecast a certain outcome at a given point in time, while the latter looked for a precise duration of follow-up (19). On top of this, the APACHE II score (as with similar models) was derived after the collection of the worst data within the 24-hr period after intensive care admission. Again, this is not how it is often used in clinical trial settings. The summation of these issues results in a risk assessment that allows some patients with a risk too high to be enrolled and some who would potentially benefit to be excluded. This fact has been well explored by the developers of these models, but the lessons have tended to be forgotten over time (20–23). A good example of where the use of the APACHE II score for this purpose can cause problems is in the interpretation of the PROWESS and ADDRESS studies. The PROWESS study demonstrated that survival benefit increased as the APACHE II quartiles increased. This led the Food and Drug Administration to license activated protein C in the United States for patients with severe sepsis or septic shock and an APACHE II score >25. To better understand the lower-risk group of patients, the Administration of Drotrecogin Alfa (Activated) in Early Stage Severe Sepsis study was initiated. This study was a placebo-controlled RCT for patients with a lower risk of death. This study was stopped early on grounds of futility. Of note, however, was the fact that some patients deemed to be at a low risk of death were enrolled although they had a high APACHE II score. The subgroup of patients in this study with an APACHE II score >25 seemed to derive no benefit from the drug. It is difficult to disentangle exactly what this means; did the initial study overestimate the treatment effect or was the risk assessment tool giving false information? It is clear that clinicians are able to identify factors associated with improved survival that were not included in the APACHE II score and were of importance to the patients' outcome.

### *New models for risk assessment*

In recent years, many groups have studied these risk assessment tools to better refine the equations so that they can be used to provide a better risk assessment both for hospital and 28-day survival. In a general population of critically ill patients, it was possible to identify from the Simplified Acute Physiology Score 3 model a number of factors that were already present on hospital admission (in other words, predisposition) and that were responsible for 46% of the explanatory power of the model (24). This compares with the 45% found in the model specifically for patients with severe infection and sepsis (25). Although, for a number of methodologic reasons, these two models should not be directly compared (26), a striking observation does become apparent, namely that acute physiology seems to contribute far less to outcome than it is often given credit for (27.4% in the general model and 35.3% in the sepsis model). This clearly differs from the 73% contribution attributed to these variables in the old APACHE III model (17). This is of direct relevance to many ongoing trials with inclusion and exclusion criteria based mainly on the presence, severity, and duration of organ failures rather than on either the predisposition of the patient or on the characteristics of the infection such as the agent, site, or extension (27). It is clear that in future years we will have to give far greater credence to the predisposition of patients to critical illness and the characteristics of particular infections rather than the ultimate response in terms of organ dysfunction and failure. This has recently been validated by several groups who have devel-



Copyright (c) Society of Critical Care Medicine and Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

oped models based around the Predisposition, Insult/infection, Response, and Organ dysfunction (PIRO) concept of risk assessment (25, 28).

This lack of strict baseline evaluation of the patient characteristics to which the intervention (or the bundle of interventions) will be applied probably explains why apparently similar studies achieve different results. An example of this is the discrepancy between the results of the Corticosteroid Therapy of Septic Shock study, comparing the use of hydrocortisone with placebo in patients with septic shock (29), with a negative result of the intervention on all cause, 28-day mortality when compared with almost the same design in a much more severe cohort of patients (5). Despite many arguments that can be raised resulting from different amounts of hydrocortisone or the concomitant use of fludrocortisone, it is probably the baseline risk of death of the patients in both studies that explains the different results, as shown by the death rate in the placebo arms of both studies: 61% in Annane's study and 31% in Sprung's study. The same is true for other recent studies, a classic example being the discordance in outcomes of a policy designed to rigidly control blood glucose levels when tested by Van den Berghe (6) and a German network (30). Consequently, new trial designs are needed to minimize some of these problems (31–33), but it is possibly more important to better understand the prognostic determinants of the patient with sepsis, severe sepsis, and septic shock, and more important than that, that these determinants are changing.

### The changing face of intensive care medicine

We must realize that our field of science faces some unique challenges. Contrary to the impact of natural diseases on physiology, which usually evolves slowly over decades or centuries, our models become old and outdated very quickly. Changes in the baseline characteristics of the populations, driven more often by changes in lifestyle than by true (genetic) changes, changes in the way we organize and deliver health care, and in the way we prevent, diagnose, and treat major diseases all have a major impact on the accuracy of our instruments. One of the consequences of these changes is a continuous pressure on the representativeness of our patient databases and in the way we model the outcome of our patients based on a set of predictive variables. Life is made of change, and in regard to intensive care medicine, most of the change has been positive.

The case-mix of a modern day ICU is very different to that of 20 years ago. The mean age of the admitted patients has increased, as have the number and severity of their comorbidities. The diagnoses have changed, as has the complexity of interventions being offered. The emphasis has shifted towards early detection and correction of physiological derangements as the degree of physiological reserve of our patients is lower today than it used to be, and toward a more multidisciplinary approach with a stronger focus upon effectiveness and safety of our practices. The final result of this complex and poorly understood interaction between patient characteristics and medical interventions has been a globally positive impact on outcomes. Examples include sepsis and the acute respiratory distress syndrome where risk-adjusted mortality is lower today than ever before (34). For developers of outcome prediction models, these changes in patient characteristics, healthcare delivery models, and outcomes have required them to regularly reinvent the wheel–first, to assure that the databases in which the models are developed reflect the underlying characteristics of patients and healthcare delivery systems and, second, to assure that the relationship between patient-related characteristics and outcome is taken into account.

All new severity scores are more complex than their old counterparts, having been developed in larger databases, and built using more complex modeling techniques. They are, on the other hand, more bound to the specificities of their development databases, a fact that can affect their use outside their development population. The Simplified Acute Physiology Score 3 admission model was built based on a database of 19,577 patients consecutively admitted to 307 ICUs all over the world from October 14 to December 15, 2002 (24, 35); the APACHE IV model was based on a database of 110,558 consecutive admissions during 2002 and 2003 to 104 ICUs in 45 United States hospitals participating in the APACHE III database (36); the updated Mortality Probability Admission Model (MPM0-III) resulted from a retrospective analysis of data from 124,855 patients admitted to 135 ICUs in 98 North American hospitals participating in the Project IMPACT (Cerner, Bel Air, MD), a USA database of data from ICU patients registered between 2001 and 2004 (37); and the Intensive Care National Audit & Research Centre (ICNARC) model is based on prospective data from 216,626 critical care admissions from 163 adult, general critical care units in England, Wales, and Northern Ireland, December 1995 to August 2003 (38). For all these developers, compromises had to be made regarding the heterogeneity of the databases (and better generalizability of the results) vs. homogeneity and a more limited range of application, as recently discussed by Hutchings et al (39).

It is too soon to be sure whether this third generation of outcome prediction models will perform better in the long run than their old counterparts, as has been shown for the second generation of outcome prediction models over their predecessors (40), or to know the degree of confidence in using these instruments in populations that differ from the development population, a crucial issue when choosing which method to use in a particular ICU. Our knowledge and level of expertise in medical statistics is still too crude to allow us to control completely for differences in patient characteristics and healthcare delivery models that are present when we compare demographic data from patients admitted to an ICU in Boston with those in one from Sydney, or from an ICU in Hong Kong with one from Lisbon. By so doing, we can be sure that outcome assessment models will provide us with an important framework for assessing the severity of illness of our patients, and that this may eventually help the individual ICU to spot and to monitor the consequences of such activity, which in the past was known as the continuous need to reinvent the wheel (41).

### The challenge of converting data into recommendations

There are many parallels between the development of the scoring systems as described previously and the assessment of evidence generated from RCTs. In particular, it is vitally important to understand the concepts of internal and external validity within the trial setting. The internal validity describes the accuracy of the conclusions regarding an intervention's effect on a given group of patients under that study's specific circumstances. It is highly dependent on how



Copyright (c) Society of Critical Care Medicine and Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

controlling how the study was conducted, how biases were prevented, and other factors that may have influenced the outcome. The external validity describes how generalizable the results of the study are to the wider population. Many studies strive to homogenize their recruitment by having stringent inclusion and exclusion criteria. This, by definition, will reduce the external validity of the conclusions. Unless these factors are taken into account when weighing the evidence in the round, then extrapolating these conclusions can lead to poor clinical practice recommendations and potential harm.

Over the last few years, despite significant improvements in our methods for evaluating and weighing evidence generated from clinical research such as the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (42) or more recently the GRADE grid system (43), no significant research results has been published about how to analyze the incorporation of different interventions into a single bundle of care and, more importantly, in defining as to which patients this single bundle of care is potentially beneficial. In the natural enthusiasm and academic bias of rushing from the results of a RCT (usually with a very select population that does not compare with the baseline characteristics of the population to which the bundle will be applied), and in the wish to increase the target population, researchers can be tempted to forget the ultimate aim of any intensive care clinician: to decrease the morbidity, the mortality, and the consequences of critical illness to the patients and their relatives and to achieve that aim in a socially responsible and equitable fashion.

## The bundle of care

Central to this whole discussion is the concept of bundles of care. Originally proposed by the Institute of Healthcare Improvement, a "not-for-profit organization driving the improvement of health by advancing the quality and value of healthcare," bundles have been defined as "a group of interventions related to a disease process that, when executed together, result in better outcomes than when implemented individually. The individual bundle elements are built on upon [sic] evidence-based practices. The science behind the elements of a bundle is so well-established that their implementation should be considered a generally accepted practice" (44). Although the described controversy surrounds the sepsis bundles, similar controversies have been generated by other bundles of care. These include the ventilator bundle or the ventilator-associated pneumonia bundle, a series of measures to prevent ventilator-associated pneumonia (45), and the catheter bundles, a series of measures to prevent catheter-related bloodstream infections (46). As pointed out recently by Marwick and Davey (47), distinguishing the proposed use of any bundles for quality improvement, research, or judgment can be extremely difficult. Even more controversial is the slow but constant movement to change the aim of bundles from methods for changing clinical practice to methods for measuring clinical or organizational performance, especially when data demonstrate clearly that not all elements of a bundle are equally important; in the future, some may even be shown to be antagonist (48, 49). This raises the question as to whether similar standards that have been used to assess evidence have been applied to bundle generation. Here the answer is clearly no.

It is imperative that when assessing evidence for incorporation into a bundle, the grounding and rationale of that evidence must be robust. To do this we need to have reliable, open, and transparent instruments that can weigh the evidence, taking into account both the internal and external validity of the underpinning studies. If the rationale for using any specific variable is not strong, then the bundle methodology may force its use in inappropriate patients with resulting harm (50, 51). This key concept is the difference between the efficacy of an intervention and its effectiveness in clinical practice. It is also a difference between studying a single intervention (the classic field of RCT) or a complex package of interventions (the last step in clinical research before the results can be incorporated in standard care). Only after all these effects have been taken into account can a tentative bundle of care be drafted. We strongly believe that any bundle of care, once drafted, should then enter a validation phase in an independent cohort of patients before being used to mandate any change in practice. This is especially important before administrators or funders of healthcare take control of these tools to drive quality control, legal or benchmarking issues, or for reimbursement purposes. This validation should test the effect of the different elements of the bundle in patients with different baseline characteristics and be evaluated with the same rigor that we demand for other fields of science.

## Summary

Clinical trial design has improved markedly in recent years, although we believe that there are still major improvements that must be made. One area that has often been neglected in our rush for positive results has been the skill, or science, of assessing evidence and bringing heterogeneous studies together and making clinical practice recommendations. All too often we rush into incorporating the latest result into our therapeutic armamentarium, often with subsequent regret. This attitude creates a paradox: we test individual interventions; we add these interventions together and apply them to patients; and we forget the fact that the patient is treated with packages of interventions, never previously evaluated together.

Also, we test different interventions independently from the place and timing of their original application, and usually without a proper interrogation of baseline patient characteristics that can act as a confounder or (worse) as an effect modifier, and we expect that the placebo group remains static, which means that medical practice should remain static during the trial and in the years that follow when the results of the trial are incorporated into the definitions of "best practice."

Consequently, we must return to real life: take interventions and integrate them at a local level in a local package; using cumulative deviation between groups to decide if and how much it affects mortality or other outcome measures; and to continually model outcomes in the control and intervention groups, testing whether the new intervention increases the explanatory power of the model (i.e. it has an effect on the outcome of interest, either positive or negative).

Now is the time to address and resolve the issue. We should never evaluate evidence in isolation but always within the context of the patient's individual characteristics before assuming that the whole is better than the sum of the individual parts, that compliance is assured, and that the effectiveness of the bundle is greater than its potential side effects. Reading the literature, it seems to be based more on reli-



Copyright (c) Society of Critical Care Medicine and Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

gious belief than on scientific grounds. Now, it is the time to change.

> Science never proves a hypothesis. It offers various levels of disproof.
>
> —Karl Popper

## REFERENCES


1. Sibbald WJ, Vincent J-L: Round table conference on clinical trials for the treatment of sepsis Brussels, March 12–14, 1994. *Intensive Care Med* 1995; 21:184–189
2. Bernard GR, Vincent J-L, Laterre P-F, et al: Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 2001; 344:699–709
3. 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
4. Kumar A, Roberts D, Wood KE, et al: Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34:1589–1596
5. Annane D, Sébille V, Charpentier C, et al: Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA* 2002; 288: 862–871
6. Van den Berghe G, Wouters P, Weekers F, et al: Intensive insulin therapy in critically ill patients. *N Engl J Med* 2001; 345:1359–1367
7. Dellinger RP, Carlet JM, Masur H, et al: Surviving Sepsis Campaign guidelines for the management of severe sepsis and septic shock. *Intensive Care Med* 2004; 30:536–555
8. Dellinger RP, Levy MM, Carlet JM, et al: Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Med* 2008; 34:17–60
9. Eichacker PQ, Natanson C, Danner RL: Surviving sepsis—Practice guidelines, marketing campaigns, and Eli Lilly. *N Engl J Med* 2006; 355:1640–1642
10. Levy MM, Pronovost PJ, Dellinger RP, et al: Sepsis change bundles: Converting guidelines into meaningful change in behavior and clinical outcome. *Crit Care Med* 2004; 320(Suppl):S595–S597
11. Shoemaker WC, Appel PL, Kram HB: Tissue oxygen debt as a determinant of lethal and nonlethal postoperative organ failure. *Crit Care Med* 1988; 16:1117–1120
12. Connors AF, Speroff T, Dawson NV, et al: The effectiveness of right heart catheterization in the initial care of critically ill patients. *JAMA* 1996; 276:889–897
13. Marco Ranieri V, Moreno RP, Rhodes A: The European Society of Intensive Care Medicine (ESICM) and the Surviving Sepsis Campaign (SSC). *Intensive Care Med* 2007; 33:423–425
14. Barie PS: An opinion too far—The campaign against the Surviving Sepsis Campaign. *Surg Infect* 2007; 7:485–488
15. Hicks P, Cooper DJ, Webb S, et al: The Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008. An assessment by the Australian and New Zealand Intensive Care Society. *Anaesth Intensive Care* 2008; 36: 149–151
16. Vincent J-L, Opal SM, Marshall JC: Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions. *Crit Care Med* 2010; 38: 283–287
17. 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
18. Le Gall J-R, Loirat P, Alperovitch A: Simplified acute physiological score for intensive care patients. *Lancet* 1983; ii:741
19. Moreno RP, Metnitz PGH, Metnitz B, et al: Modeling in-hospital patient survival during the first 28 days after intensive care unit admission: A prognostic model for clinical trials in general critically ill patients. *J Crit Care* 2008; 23:339–348
20. Knaus WA, Harrell FE, Fisher CJ, et al: The clinical evaluation of new drugs for sepsis: A prospective study design based on survival analysis. *JAMA* 1993; 270:1233–1241
21. Knaus WA: Principles of severity stratification and outcome prediction in sepsis and shock. *Intensive Care Med* 1994; 20:S115
22. Knaus WA, Wagner DP, Harrell FE, et al: What determines prognosis in sepsis? Evidence for a comprehensive individual patient risk assessment approach to the design and analysis of clinical trials. *In*: Sepsis: Current Perspectives in Pathophysiology and Therapy. Reinhart K, Eyrich K, Sprung C (Eds). Berlin, Heidelberg, Springer-Verlag, 1994, pp 23–37
23. Knaus WA, Harrell FE, LaBrecque JF, et al: Use of predicted risk of mortality to evaluate the efficacy of anticytokine therapy in sepsis. *Crit Care Med* 1996; 24:46–56
24. 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
25. Moreno RP, Metnitz B, Adler L, et al: Sepsis mortality prediction based on predisposition, infection and response. *Intensive Care Med* 2008; 34:496–504
26. Moreno R, Jordan B, Metnitz P: The changing prognostic determinants in the critically ill patient. *In*: Yearbook of Intensive Care and Emergency Medicine. Vincent JL (Ed). Berlin, Springer-Verlag, 2007, pp 899–907
27. Finfer S, Marco Ranieri V, Thompson BT, et al: Design, conduct, analysis and reporting of a multi-national placebo-controlled trial of activated protein C for persistent septic shock. *Intensive Care Med* 2008; 34: 1935–1947
28. Rubulotta F, Marshall JC, Ramsay G, et al: Predisposition, insult/infection, response, and organ dysfunction: A new model for staging severe sepsis. *Crit Care Med* 2009; 37: 1329–1335
29. Sprung CL, Annane D, Keh D, et al: Hydrocortisone therapy for patients with septic shock. *N Engl J Med* 2008; 358:111–124
30. Brunkhorst FM, Engel C, Bloos F, et al: Intensive insulin therapy and pentastarch resuscitation in severe sepsis. *N Engl J Med* 2008; 358:125–139
31. Minneci PC, Eichacker PQ, Danner RL, et al: The importance of usual care control groups for safety monitoring and validity during critical care research. *Intensive Care Med* 2008; 34:942–947
32. Natanson C, Esposito CJ, Banks SM: The siren's songs of confirmatory sepsis trials: selection bias and sampling error. *Crit Care Med* 1998; 26:1927–1931
33. Angus DC, Mira J-P, Vincent J-L: Improving clinical trials in the critically ill. *Crit Care Med* 2010; 38:527–532
34. Hutchings A, Durand MA, Grieve R, et al: Evaluation of modernisation of adult critical care services in England: Time series and cost effectiveness analysis. *BMJ* 2009; 339: b4353
35. 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
36. 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
37. 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
38. Harrison DA, Parry GJ, Carpenter JR, et al: A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med* 2007; 35:1091–1098
39. Hutchings A, Durand MA, Grieve R, 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
40. Bertolini G, D'Amico R, Apolone G, et al: Predicting outcome in the intensive care unit using scoring systems: Is new better? A comparison of SAPS and SAPS II in a cohort of 1,393 patients. *Med Care* 1998; 36: 1371–1382
41. Moreno RP: Outcome prediction in intensive care: Why we need to reinvent the wheel. *Curr Opin Crit Care* 2008; 14:483–484
42. GRADE Working Group: Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328:1–8
43. Jaeschke R, Guyatt GH, Dellinger P, et al: Use




Copyright (c) Society of Critical Care Medicine and Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. *BMJ* 2008; 337:a744
44. Singer M: The Surviving Sepsis guidelines: Evidence-based or evidence-biased? *Crit Care Resusc* 2006; 8:244–245
45. Wip C, Napolitano L: Bundles to prevent ventilator-associated pneumonia: How valuable are they? *Curr Opin Infect Dis* 2009; 22: 159–166
46. Pronovost P, Needham D, Berenholtz S, et al: An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006; 355:2725–2732
47. Marwick C, Davey P: Care bundles: The holy grail of infectious risk management in hospital? *Curr Opin Infect Dis* 2009; 22:364–369
48. Gao F, Melody T, Daniels DF, et al: The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: A prospective observational study. *Crit Care* 2005; 9:R764–R770
49. Ferrer R, Artigas A, Levy MM, et al: Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *JAMA* 2008; 299:2294 –2303
50. Bertolini G, Rossi C, Anghileri A, et al: Use of Drotrecogin alfa (activated) in Italian intensive care units: the results of a nationwide survey. *Intensive Care Med* 2007; 33: 426–434
51. Kanji S, Perreault MM, Chant C, et al: Evaluating the use of Drotrecogin alfa (activated) in adult severe sepsis: A Canadian multicenter observational study. *Intensive Care Med* 2007; 33:517–523



Copyright (c) Society of Critical Care Medicine and Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.