

Outcome measures for clinical research in sepsis: A report of the 2nd Cambridge Colloquium of the International Sepsis Forum

John C. Marshall, MD; Jean-Louis Vincent, MD, PhD; Gordon Guyatt, MD, MSc; Derek C. Angus, MD, MPH; Edward Abraham, MD; Gordon Bernard, MD; Claire Bombardier, MD; Thierry Calandra, MD, PhD; Henrik Stig Jørgensen, MD, DMSci; Richard Sylvester, ScD; Maarten Boers, MSc, MD, PhD

Background and Objectives: Sepsis is the leading cause of morbidity and mortality for patients admitted to an intensive care unit. The evaluation of new therapies has been hampered by the underdevelopment of outcome measures used to detect biological activity and patient-centered benefit in a complex and highly heterogeneous patient population. We sought to evaluate existing approaches and to draw on insights from other disciplines to propose a comprehensive approach to outcome evaluation in sepsis clinical trials.

Methods: An expert colloquium organized by the International Sepsis Forum brought together sepsis researchers, clinical epidemiologists, and experts in the development and implementation of outcome measures in rheumatology, neurology, and oncology.

Results: The translation of an evolving understanding of the biology of sepsis into effective new therapies for critically ill patients requires a reevaluation of the end points used to determine response to intervention. These represent a continuum that measures biological activity against the target at one end and sustained improvement in survival or quality of life at the other. Early phase research should determine whether an intervention works *in vivo*, using measures that are responsive and informative to provide proof of principle, to

aid in selecting optimal patient populations for study, and to gain insights into optimal dose and duration of therapy. After *in vivo* biology has been demonstrated and the possibility of efficacy inferred by plausible improvements in surrogate physiologic measures, definitive studies should seek robust evidence of benefit using end points that measure important, patient-centered benefit, including intermediate and longer term survival and health-related quality of life. Nonmortal measures of benefit assume particular importance for populations, such as children, whose mortality risk is low, or who have significant rates of comorbidities that independently limit survival. Composite measures that integrate morbidity and mortality effects may provide the most meaningful information about therapeutic efficacy.

Conclusions: The development of explicit, hypothesis-driven, and iterative approaches to outcome measure development, patterned on approaches used in the fields of rheumatology and oncology, may improve the conduct of clinical studies in the critically ill. (Crit Care Med 2005; 33:1708–1716)

KEY WORDS: sepsis; morbidity; mortality; intensive care unit; outcome measures; outcome evaluation

LEARNING OBJECTIVES

On completion of this article, the reader should be able to:

1. Describe direct measures of outcomes in sepsis.
2. Describe surrogate measures of outcomes in sepsis.
3. Use this information in a clinical setting.

Dr. Marshall has disclosed that he was formerly a consultant for Edwards and Wyeth-Ayerst and is currently a consultant for BRAHMS Diagnostics, and GlaxoSmithKline. Dr. Bernard has disclosed that he has been the recipient of grant/research funds from Eli Lilly, Novo Nordisk, and Takeda Pharmaceuticals. Dr. Bombardier has disclosed that she has been the recipient of grant/research funds from Abbott and Amgen, is a consultant/advisor for and is on the speakers bureau of Merck, Pfizer, Schering-Plough Corp., and Wyeth. Dr. Calandra has disclosed that he is the recipient of direct grant/research funding from Baxter, Wako, and Merck and was the recipient of direct grant/research funding from Pfizer, Natimmune, Bristol-Myers Squibb in the past; a consultant/advisor for Baxter, Pfizer, Merck, GlaxoSmithKline, CAT, Roche; and on the speakers bureau of Pfizer and Merck and was formerly on the speakers bureau of GlaxoSmithKline. The remaining authors have disclosed that they have no financial relationships or interest in any commercial companies pertaining to this educational activity.

Wolters Kluwer Health has identified and resolved all faculty conflicts of interests regarding this educational activity.

Visit the *Critical Care Medicine* Web site (www.ccmjournal.org) for information on obtaining continuing medical education credit.

Professor, Department of Surgery, Interdepartmental Division of Critical Care Medicine, University of Toronto, Attending Surgeon, Intensivist, St. Michael's Hospital, Toronto, Canada (JCM); Head, Department of Intensive Care, Erasme University Hospital, Brussels, Belgium (JLV); Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada (GG); Professor and Vice-Chair, Department of Critical Care Medicine, Director, The CRISMA Laboratory, Univer-

sity of Pittsburgh School of Medicine, Pittsburgh, PA (DCA); Roger Sherman Mitchell Professor, Pulmonary and Critical Care Medicine, University of Colorado Health Sciences Center, Denver, CO (EA); Melinda Owen Bass Professor of Medicine, Director, Division of Allergy, Pulmonary and Critical Care, Vanderbilt University School of Medicine, Nashville, TN (GB); Professor of Medicine, Director, Division of Rheumatology, University of Toronto, Senior Scientist, Institute for Work and Health, Director,

Clinical Decision-Making and Health Care Division, Toronto General Hospital Research Institute, University Health Network, Rheumatologist, Mount Sinai Hospital, Canada Research Chair in Knowledge Transfer for Musculoskeletal Care, Toronto, Canada (CB); Associate Professor, Department of Infectious Diseases, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland (TC); Head of Stroke Unit, Department of Neurology, Hillerød Hospital, Hillerød, Denmark (HSJ); Assistant Director, Bio-

Sepsis is the leading cause of morbidity and mortality for patients admitted to an intensive care unit (ICU) (1, 2). Defined as the systemic host response to invasive infection, sepsis is considered to be severe when it results in organ dysfunction; when that dysfunction jeopardizes oxygen delivery to the tissues, septic shock is said to be present (3). The syndrome has proven challenging to characterize (4), and beyond measures directed toward the eradication of its infectious causes—antibiotics and surgical source control—treatment is essentially supportive. Recognition that sepsis arises through the activation of innate immunity has raised the prospect of targeting specific host-derived mediators of this response (5, 6). However, although billions of dollars have been spent in the evaluation of dozens of promising strategies, only one new therapeutic agent has been licensed for the treatment of sepsis (7).

The study of sepsis presents unique challenges. Sepsis is a syndrome, not a disease, and lacks a reliable measure of disease activity, comparable, for example, to viral load in AIDS, tumor burden in oncology, or left ventricular function in cardiology. The inflammatory response that mediates the syndrome is both adaptive and injurious, containing infection, but at the cost of bystander injury. Finally, sepsis is often a complication of other diseases occurring in patients with life-limiting comorbidities, and an intervention that increases survival may not improve quality of life but merely prolong suffering.

Clinical trials of novel sepsis therapies typically recruit heterogeneous populations of patients, using the nonspecific physiologic criteria of sepsis syndrome (8) or the systemic inflammatory response syndrome (3); therapeutic efficacy is measured as a reduction in 28-day all-cause mortality. The relative risk reduction seen with a diverse group of interventions in this heterogeneous

population has never exceeded 20% and more typically averages 5–15% (5). Most phase III trials have been underpowered to detect such a small effect. Equally important, therapeutic efficacy varies across patient subgroups (9), yet little has been done to understand this variability in order to improve trial design and to facilitate the use of new therapies.

These considerations prompted the International Sepsis Forum to convene a colloquium on outcome measures for sepsis research, bringing together experienced clinical researchers in sepsis and experts in outcome measures in other disciplines. We review the unique challenges of sepsis research, provide an overview of the basic principles of outcome measurement in clinical trials and the insights that have arisen through work in other disciplines, and develop an integrated model for measuring the impact of intervention on outcomes relevant to sepsis that span a spectrum from biological proof of principle to documentation of robust and sustained clinical benefit.

SEPSIS: NATURAL HISTORY AND BURDEN OF ILLNESS

Sepsis is common, and its incidence in the developed world is increasing (10), affecting up to 30–50% of ICU patients (11–14). Severe sepsis rivals myocardial infarction as a modifiable cause of mortality in the United States and generates costs to the health care system that approach \$17 billion annually (2).

The increased mortality risk of sepsis starts early and persists for years. Pooled data from the placebo arms of five large multiple-center trials show that mortality rate increases from 4% at 24 hrs, to 9% by 3 days, and to 16% by the end of the first week. By 28 days, the landmark time point favored for pivotal trials for drug registration, mortality rate is approximately 30% (Fig. 1). There is considerable international variation in mortality rates, with higher rates in Greece, Portugal, and Italy than in Scandinavia or Switzerland (15). The elevated risk of death is prolonged: Quartin et al. (16) found that the mean life span of sepsis survivors was reduced by 4 yrs; a similar reduction has been reported for patients hospitalized with community-acquired pneumonia (17). Premorbid health status, acute organ dysfunction, and development of the acute respiratory distress syndrome are all independent predictors of long-term survival (18).

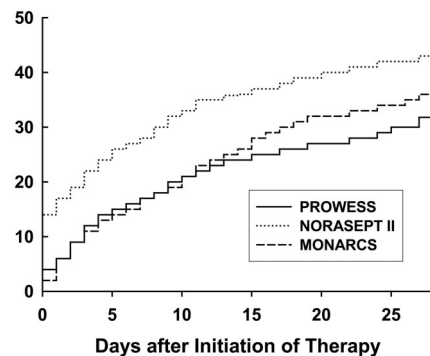


Figure 1. Kaplan-Meier curves of placebo mortality rates in three recent large, multiple-center trials of novel therapies for sepsis (7, 46, 73). Differences in early mortality rate reflect differences in study entry criteria; however, maximal mortality rates occur over the first 2 wks following diagnosis, but those rates continue to increase to 28 days and, although not measured in these studies, even beyond 28 days.

The early morbidity of sepsis is reflected in deranged organ function and the need for ICU supportive care; however, this morbidity is not experienced primarily by the patient but rather by the patient's family and loved ones. After the acute illness resolves, the patient may still require a lengthy hospital stay and subsequent rehabilitation, with the attendant physical, emotional, and financial burdens; long-term morbidity is reflected in reduced health-related quality of life (19–21). Reduced physical function and muscular weakness persist for ≥ 1 yr following the acute illness (22, 23), and survivors experience difficulties with work and normal activities of daily living (18).

OUTCOME MEASURES: A TEMPLATE FOR CLINICAL RESEARCH

Outcomes in clinical trials are events or states that are potentially modified by an intervention. Outcome measures are chosen to capture treatment effects that are important to the patient and/or informative about the disease process and the consequences of intervention in the population studied. Clinical outcomes may be of direct importance to the patient (death, time to death, morbidity, quality of life), or they may be surrogate measures, important because they are causally linked to the outcome of interest, capture the impact of the intervention on that outcome, are believed to be directly related to patient-important outcomes, and are easier or more efficient to mea-

statistics, EORTC Data Center, Brussels, Belgium (RS); and Professor, Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands (MB).

Address requests for reprints to: John C. Marshall, MD, Room 4-007, Bond Wing, St. Michael's Hospital, 30 Bond Street, Toronto, Ontario, Canada M5B 1W8. E-mail: jc.marshall@utoronto.ca

Copyright © 2005 by the Society of Critical Care Medicine and Lippincott Williams & Wilkins

DOI: 10.1097/01.CCM.0000174478.70338.03

Table 1. Outcome measures in sepsis clinical research

Domain of Interest	Examples
Biological activity of the intervention	
On therapeutic target	<ul style="list-style-type: none"> ● Alteration in target (e.g., reduced TNF) ● Change in downstream marker (e.g., IL-6)
On physiologic consequences	<ul style="list-style-type: none"> ● Extent, or time to resolution, of study entry criteria ● Alterations in individual physiologic variables (e.g., PO_2/FIO_2 ratio, MAP) ● Aggregate physiologic scores (e.g., MODS, APACHE)
Clinical alterations	
On short-term morbidity rate	<ul style="list-style-type: none"> ● Alterations in individual measures of ICU intervention (e.g., duration of ventilation, ICU stay, intervention-free days) ● Aggregate intervention scores (e.g., SOFA) ● Potential adverse consequences of therapy (e.g., rates of superinfection, recurrence, resistance, opportunistic infection) ● Toxicity directly related to mode of action (e.g., hemorrhage)
Patient-centered benefit	
Short-term mortality rate	● 14-, 28-, or 30-day mortality rate
Long-term mortality rate	● 90-day, 6-month, 1-yr, 3-yr
Patient-centered morbidity rate	<ul style="list-style-type: none"> ● Health-related quality of life ● Time course of return to premorbid function
Costs	● Cost-effectiveness

TNF, tumor necrosis factor; IL, interleukin; MAP, mean arterial pressure; MODS, Multiple Organ Dysfunction Syndrome; APACHE, Acute Physiology and Chronic Health Evaluation; ICU, intensive care unit; SOFA, Sequential Organ Failure Assessment.

sure (24, 25). However, improvement in a surrogate measurement does not necessarily translate into clinical benefit (26). Although growth hormone induces positive nitrogen balance, its use increases mortality rate (27). Similarly, high tidal volume mechanical ventilation normalizes arterial blood gases but increases mortality rate (28), as does blood transfusion to correct anemia (29) or inhibition of nitric oxide to increase blood pressure (30).

Patient-centered benefit implies improved survival and/or an improved quality of life. Since an efficacious intervention does not ultimately prevent but only delays death, mortality end points actually reflect differential times to death; the relevant time frame depends on the disorder being treated. Survival to 1 hr may be a relevant measure of efficacy for a therapy for cardiac arrest (31), whereas the survival benefit of therapies for cancer or diabetes is typically measured in years or decades. The optimal timeframe for the ascertainment of mortality should be that during which the intervention can plausibly be expected to affect survival and should encompass the potential impact of the intervention—both positive and negative.

In sepsis research, 28-day all-cause mortality has been used as the primary measure of clinical effect. But survival to

day 28, only to die in the ICU of the complications of a nonresolving illness, is not truly therapeutic success; conversely, since sepsis is frequently a complication of other life-limiting diseases, death despite resolution of the septic episode is not necessarily therapeutic failure. Indeed short-term survival can be considered a surrogate measure, for it is only desirable if it predicts long-term survival with an acceptable quality of life. On the other hand, long-term survival is confounded by deaths from other causes.

As the duration of survival increases, so does the importance of the quality of that survival. Nonmortal, patient-centered outcomes are typically specific to the disease studied and to the anticipated activity of the intervention studied—for example, rates of reinfarction or stroke in cardiology trials (32) or symptom scales in studies of inflammatory bowel disease (33) or chronic lung disease (34). Longer term health-related quality of life can be measured using well-validated measurement tools such as the short form-36 (35). The morbidity of sepsis is surprisingly poorly characterized. Certain causes of sepsis result in characteristic morbidity, for example, digit or limb loss in meningococemia or fistulas and abdominal wall defects in peritonitis. The early morbidity of sepsis primarily reflects the interventions, such as ventila-

tion, dialysis, or vasoactive therapy, used to support failing organ function, whereas late morbidity—neuromuscular weakness, cognitive dysfunction, and neuropsychiatric sequelae—is a consequence of prolonged ICU care.

Insights From Other Disciplines

The need for outcome measures that encompass the full burden of illness, and that are responsive to change associated with therapy or disease progression, is not unique to sepsis.

Cancer is a potentially lethal disease whose mortality risk evolves over many years. Phase I clinical trials of novel therapies determine the maximum tolerated dose and toxicity using standardized criteria. Phase II trials evaluate antitumor activity using the objective response rate as the primary end point, with duration of response and acute toxicity as secondary outcomes. The Response Evaluation Criteria in Solid Tumors (RECIST) guidelines (36), developed jointly by European and North American organizations, provide a measure of overall response to treatment. Outcome measures for phase III trials are chosen to detect benefit that is important to the patient and include duration of survival, time to progression, time to treatment failure, toxicity, quality of life, and symptom control; cost-effectiveness may also be evaluated. End points in oncology trials vary with the disease stage and the treatment being tested. For early-stage cancers, the primary focus is cure; in more advanced stages, it may be tumor shrinkage, and in end-stage disease, palliation. The FDA accepts end points other than survival—for example, tumor response, disease-free survival, time to tumor progression, and pain relief—for the approval of new oncology drugs (37).

Patients with stroke experience both acute mortality and long-term disability. Recognizing that therapy can affect multiple outcome domains, the criteria of the World Health Organization consider illness from the perspectives of pathology, impairment, activity level, participation, and quality of life. However, optimal cut-offs are poorly defined, and arbitrary decisions regarding the minimal acceptable change have important implications for drug registration. For example, recombinant tissue plasminogen activator was approved by the FDA for use in stroke patients on the basis of a clinical trial showing improved outcome at 3 months

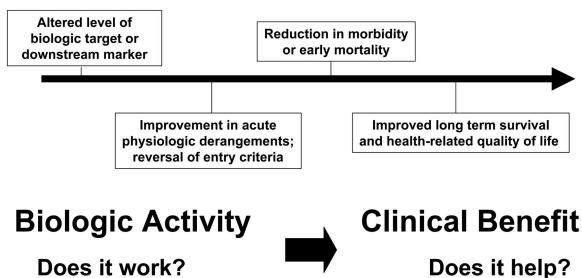


Figure 2. End points in clinical trials reflect a spectrum of questions, ranging from whether an intervention is biologically active through to whether it produces long-term clinical benefit. In sepsis research, where biological activity may not translate into clinical benefit, it is important that a systematic evaluation of the various domains of efficacy be undertaken.

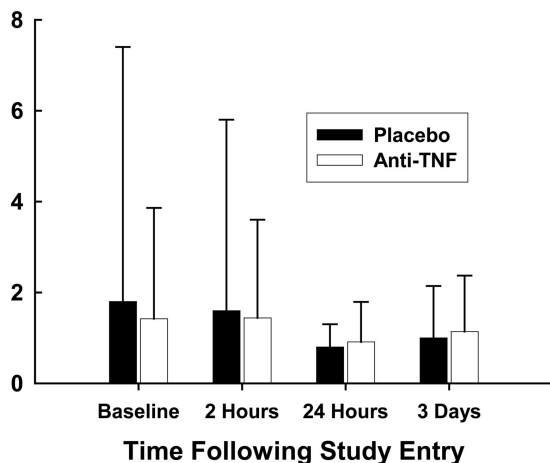


Figure 3. In a randomized, double-blind, multiple-center, placebo-controlled trial of the efficacy of an antibody to tumor necrosis factor (*TNF*) in septic shock (46), patients receiving the antibody (*open bars*) showed no evidence of *in vitro* neutralization of *TNF* bioactivity (mean \pm SD).

(38). Tissue plasminogen activator failed to meet two different primary end points in a second European trial (39), and the drug was not licensed in Europe. Had the American outcome measures been applied to the European population, the trial would have been considered positive (40).

Finally, since rheumatic diseases rarely kill, outcome measures focus exclusively on morbidity. The World Health Organization and the International League Against Rheumatism in 1993 ratified a core set of seven outcome measures for rheumatoid arthritis: three health care provider measures, three patient-centered measures, and one laboratory measure, a development facilitated by the OMERACT initiative (discussed subsequently) (41). Measures were selected based on criteria for validity, reliability, responsiveness, and feasibility. From these measures responder criteria were developed (42, 43) that are now

widely used and accepted by regulatory agencies (44).

OUTCOME MEASURES FOR SEPSIS RESEARCH: A CONCEPTUAL FRAMEWORK

Industry-funded trials in sepsis are undertaken primarily to achieve regulatory approval for novel therapeutic agents. They are inherently inefficient in addressing the questions that may move a novel strategy forward in development or that define how an effective therapy is best used in practice. Multiple domains of interest are relevant to the development and evaluation of new therapies for sepsis (Table 1). They represent a continuum that addresses the questions, “Does it work?”—an efficacy analysis—and “Does it help?”—an effectiveness analysis (Fig. 2).

Clinical trials proceed from the generic hypothesis that an intervention

having a particular biological activity can modify a pathologic state in a patient with a disease and so improve outcomes that are important to the patient. For example, since myocardial infarction results from occlusion of a narrowed coronary artery, myocardial ischemia can be treated by measures that increase coronary blood flow (angioplasty, thrombolysis, or bypass grafting) or that reduce the oxygen demands of the myocardium (β -blockers), and the risk of subsequent myocardial infarction can be reduced by agents that inhibit intravascular thrombosis (aspirin or anticoagulants) (45). Each strategy targets a discrete aspect of a well-defined pathologic process and will only benefit patients in whom that pathologic process is present. Not all patients will benefit: Ischemic damage may be irreversible despite restoration of blood flow, the antithrombotic effect may be short-lived and nonsustained, or the adverse bleeding sequelae associated with therapy may cause greater morbidity than that resulting from the original disease. However, it is implausible that a therapy that lacks *in vivo* biological efficacy will provide robust long-term clinical benefit. For example, a treatment for arthritis that does not reduce markers of inflammation is unlikely to improve joint symptomatology, a treatment for stroke that does not reduce infarct size is unlikely to improve functional outcome, and a treatment for cancer that does not inhibit tumor growth is unlikely to improve progression-free survival.

This therapeutic model is sufficiently ingrained in clinical thinking that its application is almost reflexive. In a complex disease such as sepsis, a particular intervention may help some patients, harm others, and have no effect in the remainder, and the “Does it work?” analysis becomes a prerequisite to identify an appropriate population to test the “Does it help?” question.

“Does It Work?”

Measures of Biological Activity. It is intuitively obvious that a new therapy must have biological activity in the population of interest—an anti-pseudomonal antibiotic should kill the organism in patients with *Pseudomonas* infection, or an inhibitor of nuclear factor- κ B should reduce the transcription of nuclear factor- κ B-dependent genes *in vivo*. Biological activity *in vivo* can be established by demonstration of a change in the immediate

Table 2. Modeling outcome in critical illness: Using the construct of organ dysfunction

Objective	Approach	Uses
To quantify the baseline severity of organ dysfunction	Calculate organ dysfunction score on day of admission (admission score)	To evaluate comparability of populations at baseline and assess extent of preexisting derangement
To quantify severity of organ dysfunction at a point in time	Calculate score on a particular ICU day (daily score)	To determine the intensity of resource utilization or the evolution or resolution of organ dysfunction at a discrete point in time
To measure aggregate severity of organ dysfunction over ICU stay	Sum the individual worst scores for each organ system over a defined time interval (aggregate score) Sum daily scores over a defined period of time or measure area under curve	To determine severity of physiologic derangement over a defined time interval (e.g., ICU stay)
To quantify new organ dysfunction arising following ICU admission	Calculate difference between aggregate and admission scores (δ score)	To measure organ dysfunction attributable to events occurring following ICU admission
To provide a combined measure of morbidity and mortality rate	Adjust aggregate score so that all patients dying receiving maximal number of points (mortality-adjusted score)	To create a single measure that integrates impact of morbidity in survivors and mortality for nonsurvivors

ICU, intensive care unit. Adapted from Ref. 56.

target of therapy or in a downstream marker whose expression is directly dependent on the target of therapy. These basic proof of concept studies has been strikingly absent from many sepsis trials. In the NORASEPT II study of an antibody to tumor necrosis factor (TNF) (46), for example, only 40% of patients had detectable levels of circulating TNF, and although treated patients had detectable levels of antibody, TNF activity as measured by bioassay was not altered (Fig. 3). Thus, the absence of a significant impact on mortality rate may simply reflect recruitment of patients who would not be expected to benefit from therapy and a biologically inactive intervention.

Documentation of the biological consequences of intervention both establishes proof of principle and points to potential harm that might be minimized through careful articulation of exclusion criteria. For example, granulocyte colony-stimulating factor induces neutrophilia and so might accelerate bacterial clearance at the cost of neutrophil-mediated tissue injury (47). Evidence of the former could be obtained by quantitative cultures of airway secretions and of the latter through measurement of the PO_2/FIO_2 ratio. A therapy that neutralizes TNF may lower circulating levels of interleukin-6 but at the cost of impaired microbial clearance or increased susceptibility to subsequent infection, (48).

Measures of Physiologic Effect and Organ Dysfunction. If sepsis syndrome or systemic inflammatory response syndrome criteria delineate a discrete pro-

cess that therapy can modulate, then a therapy that works should reverse these physiologic criteria more rapidly or more completely. This hypothesis has been explicitly tested in a sepsis trial only once, in a study of ibuprofen in sepsis (49).

The acute physiologic sequelae of sepsis produce the multiple organ dysfunction syndrome; prevention of new organ dysfunction, or the more rapid resolution of existing dysfunction, provides a measure of immediate physiologic efficacy. The severity of organ dysfunction can be measured using an organ dysfunction scale (50–54). Each of the commonly used scales incorporates the same six organ systems—lung, cardiovascular, renal, hepatic, coagulation, central nervous systems—and many of the same variables. Their differences lie in the measurement of cardiovascular dysfunction, the use of physiologic derangements only as in the multiple organ dysfunction score (50) or both physiologic and therapeutic variables as in the sequential organ failure assessment (51), the method of development (systematic literature review for multiple organ dysfunction syndrome, consensus process for sequential organ failure assessment), and the time of data acquisition—whether the worst value on a particular day or a representative value. Their performance is similar, with minor differences arising from the measurement of cardiovascular and renal dysfunction (55).

An organ dysfunction scale can be used to stratify patients at study entry and to model outcome in a variety of ways

(56, 57) (Table 2). In particular, it can be used to define the specific domains of physiologic function that are most affected by therapy. In the PROWESS trial of activated protein C, for example, the greatest physiologic effects were seen in the respiratory, cardiovascular, and hematologic domains of the sequential organ failure assessment score (58), whereas in the MONARCS trial of an antibody to tumor necrosis factor, therapy had the greatest impact on the neurologic and respiratory components of the multiple organ dysfunction score (59), and a study of pentoxifylline in severe sepsis showed that treatment improved the respiratory and cardiovascular components (60). In general, organ dysfunction scales have not proven to be more sensitive than mortality measures in detecting the aggregate effects of intervention (7, 28, 29, 61); an exception has been the MONARCS trial, where an early and strongly significant improvement in sequential organ failure assessment and multiple organ dysfunction scores was seen in treated patients, although the final mortality benefit was of borderline statistical significance (62).

Does It Help?

Mortality. Mortality is an intuitively attractive outcome measure for a disease process that claims the lives of a third of its victims over the first month. However, the use of mortality as the sole, or even primary, end point in sepsis research has

Table 3. Conclusions and recommendations

1. Patients with sepsis comprise a highly heterogeneous population that lacks a single distinctive clinical phenotype or an objective histologic or biochemical marker. Clinical trials in this population evaluate interventions that target disparate biochemical mediators, physiologic processes, or clinical complications. It is therefore important that the selection of outcome measures reflect the specific pathophysiologic abnormality being treated, emphasize effects that can plausibly be anticipated to occur, and be maximally informative and important.
2. Early-phase clinical trials in sepsis should emphasize outcome measures that are informative—that detect biologic activity in the population of interest and identify specific populations in whom biologic activity is maximized.
3. Phase III clinical trials should emphasize outcomes that are important—primarily to patients but also to families and caregivers. In addition to conventional mortality outcomes, nonmortal outcomes, including health-related quality of life, should be used as secondary endpoints and as primary outcomes when the baseline risk of death is low.
4. Although 28-day all-cause mortality is a useful measure of clinical benefit, it has important limitations: It may fail to capture longer term mortality effects, ignores the substantial patient-centered morbidity of sepsis, and is insensitive to nonmortal benefits in populations at lower risk of death, including children.
5. Because the increased mortality risk of sepsis is prolonged, and treatments that improve short-term survival may do so at the cost of increased patient-centered morbidity and longer term mortality, composite measures that integrate morbidity and mortality effects may provide the most meaningful information about therapeutic efficacy.
6. The development and validation of outcome measures for sepsis research present unique and particularly complicated challenges. Future research would benefit from the establishment of a dedicated, investigator-driven collaboration analogous to the Outcome Measures in Rheumatology (OMERACT) collaboration, with the active involvement of patients or their representatives, regulatory agencies, and organizations charged with the articulation of global health priorities.

important limitations (63), particularly when mortality rates are low.

First, mortality is time-dependent and confounded by unrelated intercurrent events. Measured early, mortality differences may fail to include late disease-related deaths; later mortality estimates are increasingly contaminated by deaths unrelated to the disease that the therapy targets. Second, although mortality is definitive, it is not informative and so is of little use to define optimal populations for study or determine the optimal dose or duration of treatment. Finally, for patients with significant life-altering comorbidities, mortality may not be the most important patient-centered outcome. Patients and families frequently express a wish for aggressive ICU treatment if a return to an independent existence is possible but opt for comfort measures if the outcome is likely to be a further diminution in independent quality of life (64, 65).

Clinically Important or Patient-Centered Measures of ICU Morbidity

Nonmortal measures of ICU outcome are particularly relevant when mortality rates are low. The length of ICU stay provides an aggregate measure of disease burden for survivors of sepsis but is confounded by differences in criteria for ICU discharge between institutions and the variable availability of step-down beds within institutions. Pediatric meningococemia results in a low mortality rate but significant tissue loss, and so functional outcome is a relevant measure of

therapeutic efficacy (66). Beyond this, however, the patient-centered benefits of reduced ICU morbidity are poorly defined and outcome measures correspondingly underdeveloped (67).

Composite Measures

Failure-Free Days. A treatment that reduces mortality rate may also reduce morbidity rate, but it may paradoxically increase net morbidity, because of the salvage of patients who would otherwise have died. Evaluation of the net effects of therapy would ideally integrate morbidity and mortality rates.

End points other than mortality are potentially confounded, because data are no longer available for patients who have died and because the deaths do not occur randomly but are disproportionately more common in patients with comorbid events of interest. The *morbidity-free day concept* avoids this confounding by measuring the number of days over a specified time interval that a patient is both alive and free of the morbid event. They are counted as 0 if patient dies before day 28 or experiences the morbidity of interest (e.g., ventilation) for >28 days, and 28—the number of days the patient is free from that morbidity or intervention if the patient survives beyond 28 days. The concept can be applied to many morbid outcomes, for example, ventilator-free, dialysis-free, or ICU-free days. The time frame can be modified to reflect the anticipated interval for resolution of the relevant morbidity. In a trial of ventilation with lower tidal volumes, the number of ventilator-free days, and other organ fail-

ure-free days, was increased in patients treated using the low tidal volume strategy (28). Inhaled nitric oxide was approved for use in neonates based solely on increased extracorporeal membrane oxygenation-free days (68). The free-day outcome requires complex statistical modeling but may detect clinically important effects using smaller sample sizes (69).

Morbidity and mortality can be integrated using an organ dysfunction scale by calculating a mortality-adjusted score: Measured scores are recorded for surviving patients, whereas nonsurvivors receive the maximum possible score plus one. This approach has been used to show the early benefits of anti-TNF therapy (62) and of a conservative transfusion strategy in critically ill patients (29).

Finally, more than most other illnesses, sepsis is associated with a need for intensive medical support, convalescence, and rehabilitation over a variable period of time. The aggregate morbidity of sepsis might be quantified simply by tracking the temporal course of the patient's progress through the hospital and convalescent care. Using a scale anchored at one end in death and at the other in a full return to premorbid health status in a time-dependent analysis, a therapy would be demonstrated to be effective if it delayed death and/or accelerated recovery.

Integrating Mortality With Quality of Life Impairment

Multiple-attribute utility indexes integrate death and quality of life impairment and work well across diseases, conditions,

and health states. These measures exist in a variety of disciplines and are anchored in death and full health, scoring optimal health statuses according to societal values and preferences (70). The development of comprehensive measures specific to ICU patients would necessitate such modifications as using family members to provide responses on the patient's behalf but would ground ICU outcome measurement within a well-established theoretical and practical model of health status measurement that facilitates economic analysis.

OUTCOME MEASURES— DEVELOPMENT, VALIDATION, AND IMPLEMENTATION

Sepsis research has been conducted largely through large, multiple-center, industry-funded trials, undertaken with the objective of bringing a specific therapeutic strategy to market and using entry criteria and outcome measures developed by *ad hoc* consensus processes. An alternate approach to outcome measure development is the Outcome Measures in Rheumatology (OMERACT) collaboration (www.omeract.org), an informal international network of investigators in rheumatology whose aim is to define areas of agreement and disagreement and to prioritize the research agenda for outcomes in rheumatologic diseases (71). OMERACT operates under the aegis of the International League for Rheumatology and the World Health Organization, with the active involvement of regulatory agencies and industry. OMERACT initiatives focus on the selection of outcome measures for studies in specific rheumatologic diseases, using the OMERACT filter that considers candidate measure from three perspectives: truth, discrimination, and feasibility (72). The process is iterative, with the assumption that future data may refine or modify the recommendations.

CONCLUSIONS AND RECOMMENDATIONS

The challenge inherent in evaluating novel treatments for sepsis is that we do not know *a priori* which, if any, subgroups of patients might benefit from therapy or how that benefit might be manifest early in the course of disease so that therapy can be titrated or modified. Recognizing this, the colloquium proposed six explicit recommendations (Table 3).

Preclinical data from animal models of sepsis do not provide sufficient proof of principle to justify large clinical trials whose primary objective is demonstration of patient-centered benefit. This conventional approach to sepsis research has almost certainly resulted in the abandonment of potentially useful therapeutic approaches and has rendered the uptake of new therapies more difficult. An intermediate, translational research program may increase the probability that truly efficacious therapies will be successfully shepherded through to clinical practice, whereas ineffective or even harmful approaches will be abandoned at an earlier stage.

The primary objectives of early-stage clinical research in sepsis should be to define an appropriate at-risk population, confirm biological activity *in vivo*, optimize the dose and duration of therapy, and demonstrate plausibility reflected in attenuation of the biochemical and physiologic consequences of the disease and reversal of the study entry criteria. Such studies require the use of outcome measures selected for their biological relevance and sensitivity to the anticipated consequences of intervention—both beneficial and detrimental. Measures of patient-centered benefit are important secondary outcomes in such studies, whose primary focus is documenting activity and identifying a population in whom that activity is maximal: The principle objective is to answer the question, “Does it work?”

Patient-centered outcome assumes primary importance in subsequent studies undertaken to determine whether an intervention that has biological activity *in vivo* will yield net clinical benefit and so answer the question, “Does it help?” Sepsis is a lethal disease, and prevention of its significant mortality rate will always remain a primary objective of the treating clinician. The excess mortality rate attributable to sepsis evolves over prolonged period, and so pivotal trials should assess mortality beyond 28 days, measuring it at least until 90 days and preferably longer.

However, survival is not the only or even the most important objective of therapy. Patient-centered morbidity is difficult to quantify, because the patient is frequently unaware of his or her condition. A diminishing dependence on supportive technology and a return to an independent existence are intuitively attractive objectives, and the rate at which

The development of explicit, hypothesis-driven, and iterative approaches to outcome measure development, patterned on approaches used in the fields of rheumatology and oncology, may improve the conduct of clinical studies in the critically ill.

this independence is achieved might be readily measured. Patients are rarely conscious participants in the early management of sepsis, but as the victims of the longer term effects of the disease, they are the best arbiters of what constitutes therapeutic benefit. Future initiatives in outcome measure development must make a conscious effort to incorporate their perspectives.

Initiatives such as OMERACT underline the value of developing explicit and iterative mechanisms to identify the need for new study end points and to develop and validate these for use in clinical trials. The OMERACT model is clearly applicable to critical care. The active involvement of clinical researchers, industry, regulatory agents, and patients will ensure that measures developed are optimally relevant and acceptable.

ACKNOWLEDGMENTS

This article is based on the 2nd annual Cambridge Colloquium organized by the International Sepsis Forum held in Cambridge, UK, June 29 to July 1, 2003. The participants were (in alphabetical order) Edward Abraham, Denver, CO; Derek Angus, Pittsburgh, PA; Gordon Bernard, Nashville, TN; Maarten Boers, Amsterdam, The Netherlands; Claire Bombardier, Toronto, ON, Canada; Thierry Calandra, Lausanne, Switzerland; Jonathan Cohen, Brighton, UK; R. Phillip Dellinger, Camden, NJ; Jean-Francois Dhainaut, Paris, France; Simon Finfer, Victoria, Australia; Luciano Gattinoni,

Milan, Italy; Gordon Guyatt, Toronto, ON, Canada; Margaret Herridge, Toronto, ON, Canada; Clare Hibbert, Sheffield, UK; Henrick Stig Jørgensen, Copenhagen, Denmark; John Marshall, Toronto, ON, Canada; Steven Opal, Providence, RI; Adrienne Randolph, Boston, MA; Charles Sprung, Jerusalem, Israel; Richard Sylvester, Brussels, Belgium; Jean-Louis Vincent, Brussels, Belgium; John Ware, Lincoln, NE.

Sponsored by the International Sepsis Forum; we thank Dr. Karen Pickett for her assistance in preparing the report.

REFERENCES

1. Wheeler AP, Bernard GR: Current concepts: Treating patients with severe sepsis. *N Engl J Med* 1999; 340:207-214
2. Angus DC, Linde-Zwirble WT, Lidicker J, et al: Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001; 29:1303-1310
3. Bone RC, Balk RA, Cerra FB, et al: ACCP/SCCM consensus conference. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* 1992; 101:1644-1655
4. Levy MM, Fink M, Marshall JC, et al: 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Crit Care Med* 2003; 34:1250-1256
5. Marshall JC: Such stuff as dreams are made on: Mediator-targeted therapy in sepsis. *Nature Rev Drug Disc* 2003; 2:391-405
6. Vincent J-L, Sun Q, Dubois M-J: Clinical trials of immunomodulatory therapies in severe sepsis and septic shock. *Clin Infect Dis* 2002; 34:1084-1093
7. 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
8. Bone RC, Fisher CJ, Clemmer TP, et al: Sepsis syndrome: a valid clinical entity. *Crit Care Med* 1989; 17:389-393
9. Macias WL, Vallet B, Bernard GR, et al: Sources of variability on the estimate of treatment effect in the PROWESS trial: Implications for the design and conduct of future studies in severe sepsis. *Crit Care Med* 2004; 32:2385-2391
10. Martin GS, Mannino DM, Eaton S, et al: The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 2003; 348:1546-1554
11. Sands KE, Bates DW, Lanken PN, et al: Epidemiology of sepsis syndrome in 8 academic medical centers. Academic Medical Center Consortium Sepsis Project Working Group. *JAMA* 1997; 278:234-240
12. Vincent JL, Bihari D, Suter PM, et al: The prevalence of nosocomial infection in intensive care units in Europe—The results of the EPIC study. *JAMA* 1995; 274:639-644
13. Alberti C, Brun-Buisson C, Burchardi H, et al: Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. *Intensive Care Med* 2002; 28:108-121
14. Vincent JL, Sakr YL, Carlet J, et al: Sepsis occurrence in acutely ill patients (SOAP): Results of a large European multicenter study. *Am J Respir Crit Care Med* 2003; 167:A837
15. Vincent JL: Nosocomial infections in adult intensive-care units. *Lancet* 2003; 361:2068-2077
16. Quartin AA, Schein RM, Kett DH, et al: Magnitude and duration of the effect of sepsis on survival. Department of Veterans Affairs Systemic Sepsis Cooperative Studies Group. *JAMA* 1997; 277:1058-1063
17. Kaplan V, Clermont G, Griffin MF, et al: Pneumonia: Still the old man's friend? *Arch Intern Med* 2003; 163:317-323
18. Perl TM, Dvorak L, Hwang T, et al: Long-term survival and function after suspected gram-negative sepsis. *JAMA* 1995; 274:338-345
19. Davidson TA, Caldwell ES, Curtis JR, et al: Reduced quality of life in survivors of acute respiratory distress syndrome compared with critically ill control patients. *JAMA* 1999; 281:354-360
20. Kaarlola A, Pettila V, Kekki P: Quality of life six years after intensive care. *Intensive Care Med* 2003; 29:1294-1299
21. Garcia Lizana F, Peres Bota D, De Cubber M, et al: Long-term outcome in ICU patients: What about quality of life? *Intensive Care Med* 2004; 29:1286-1293
22. Herridge MS, Cheung AM, Tansey CM, et al: One-year outcomes in survivors of the acute respiratory distress syndrome. *N Engl J Med* 2003; 348:683-693
23. De Jonghe B, Sharshar T, Lefaucheur JP, et al: Paresis acquired in the intensive care unit: a prospective multicenter study. *JAMA* 2002; 288:2859-2867
24. Bucher H, Guyatt G, Cook D, et al: Surrogate outcomes. In: *The Users' Guide to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Guyatt G, Rennie D (Eds). AMA Publications, 2002
25. Prentice RL: Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat Med* 1989; 8:431-440
26. Fleming TR, DeMets DL: Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med* 1996; 125:605-613
27. Takala J, Ruokonen E, Webster NR, et al: Increased mortality associated with growth hormone treatment in critically ill adults. *N Engl J Med* 1999; 341:785-792
28. Brower RG, Matthay MA, Morris A, et al: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000; 342:1301-1308
29. Hebert PC, Wells G, Blajchman MA, et al: A multicentre randomized controlled clinical trial of transfusion requirements in critical care. *N Engl J Med* 1999; 340:409-417
30. Lopez A, Lorente JA, Steingrub J, et al: Multiple-center, randomized, placebo-controlled, double-blind study of the nitric oxide synthase inhibitor 546C88: Effect on survival in patients with septic shock. *Crit Care Med* 2004; 32:21-30
31. Wolcke BB, Mauer DK, Schoefmann MF, et al: Comparison of standard cardiopulmonary resuscitation versus the combination of active compression-decompression cardiopulmonary resuscitation and an inspiratory impedance threshold device for out of hospital cardiac arrest. *Circulation* 2003; 108:2201-2205
32. Cannon CP, Braunwald E, McCabe CH, et al: Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med* 2004; 350:1495-1504
33. Best WR, Becktel JM, Singleton JW, et al: Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology* 1976; 70:439-444
34. Guyatt GH, King DR, Feeny DH, et al: Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation. *J Clin Epidemiol* 1999; 52:187-192
35. Ware JJ, Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30:473-483
36. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000; 92:205-216
37. Johnson JR, Williams G, Pazdur R: End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003; 21:1404-1411
38. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N Engl J Med* 1995; 333:1581-1587
39. Hacke W, Kaste M, Fieschi C, et al: Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Study (ECASS). *JAMA* 1995; 274:1017-1025
40. Hacke W, Bluhmki E, Steiner T, et al: Dichotomized efficacy end points and global end-point analysis applied to the ECASS intention-to-treat data set: Post hoc analysis of ECASS I. *Stroke* 1998; 29:2073-2075
41. Boers M, Tugwell P, Felson DT, et al: World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying anti-rheumatic drugs in rheumatoid arthritis

- clinical trials. *J Rheumatol* 1994; 41(Suppl): 86–89
42. Felson DT, Anderson JJ, Boers M, et al: The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993; 36:729–740
 43. van Gestel AM, Prevoo ML, van't Hof MA, et al: Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum* 1996; 39:34–40
 44. van Gestel AM, Anderson JJ, van Riel PL, et al: ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. American College of Rheumatology European League of Associations for Rheumatology. *J Rheumatol* 1999; 26: 705–711
 45. Boersma E, Mercado N, Poldermans D, et al: Acute myocardial infarction. *Lancet* 2003; 361:847–858
 46. Abraham E, Anzueto A, Gutierrez G, et al: Double-blind randomised controlled trial of monoclonal antibody to human tumor necrosis factor in treatment of septic shock. *Lancet* 1998; 351:929–933
 47. Azoulay E, Attalah H, Harf A, et al: Granulocyte colony-stimulating factor or neutrophil-induced pulmonary toxicity: Myth or reality? Systematic review of clinical case reports and experimental data. *Chest* 2001; 120: 1695–1701
 48. Bresnihan B, Cunnane G: Infection complications associated with the use of biologic agents. *Rheum Dis Clin North Am* 2003; 29: 185–202
 49. Bernard GR, Wheeler AP, Russell JA, et al: The effects of ibuprofen on the physiology and survival of patients with sepsis. *N Engl J Med* 1997; 336:912–918
 50. 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
 51. Vincent JL, Moreno R, Takala J, et al: The sepsis-related organ failure assessment (SOFA) score to describe organ dysfunction/failure. *Intensive Care Med* 1996; 22:707–710
 52. Bernard G: The Brussels score. *Sepsis* 1997; 1:43–44
 53. Le Gall J-R, Klar J, Lemeshow S, et al: The logistic organ dysfunction system. *JAMA* 1996; 276:802–810
 54. Goris RJA, te Boekhorst TPA, Nuytinck JKS, et al: Multiple organ failure. Generalized autodestructive inflammation? *Arch Surg* 1985; 120:1109–1115
 55. Peres BD, Melot C, Lopes FF, et al: The Multiple Organ Dysfunction Score (MODS) versus the Sequential Organ Failure Assessment (SOFA) score in outcome prediction. *Intensive Care Med* 2002; 28:1619–1624
 56. Marshall JC: Charting the course of critical illness: Prognostication and outcome description in the intensive care unit. *Crit Care Med* 1999; 27:676–678
 57. Ferreira FL, Bota DP, Bross A, et al: Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; 286:1754–1758
 58. Vincent J-L, Angus DC, Artigas A, et al: Effects of drotrecogin alfa (activated) on organ dysfunction in the PROWESS trial. *Crit Care Med* 2003; 31:834–840
 59. Marshall JC, Panacek EA, Teoh L, et al: Modelling organ dysfunction as a risk factor, outcome, and measure of biologic effect in sepsis: Results of the MONARCS trial. *Crit Care Med* 2001; 28:A46
 60. Staubach KH, Schröder J, Stüber F, et al: Effect of pentoxifylline in severe sepsis. Results of a randomized, double-blind, placebo-controlled study. *Arch Surg* 1998; 133: 94–100
 61. 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
 62. Panacek EA, Marshall JC, Albertson TE, et al: Efficacy and safety of the monoclonal anti-TNF antibody F(ab')₂ fragment afelimomab in patients with severe sepsis stratified by IL-6 level. *Crit Care Med* 2004; 32:2173–2182
 63. Petros AJ, Marshall JC, van Saene HKF: Is mortality an appropriate endpoint for clinical trials in critical illness? *Lancet* 1995; 345: 369–371
 64. Thompson BT, Cox PN, Antonelli M, et al: Challenges in end-of-life care in the ICU: statement of the 5th International Consensus Conference in Critical Care: Brussels, Belgium, April 2003: Executive summary. *Crit Care Med* 2004; 32:1781–1784
 65. Boyce CL, Loewenstein G, Arnold RM, et al: Quality of death: assessing the importance placed on end-of-life treatment in the intensive-care unit. *Med Care* 2004; 42:423–431
 66. Levin M, Quint PA, Goldstein B, et al: Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal sepsis: A randomised trial. *Lancet* 2000; 356: 961–967
 67. Black NA, Jenkinson C, Hayes JA, et al: Review of outcome measures used in adult critical care. *Crit Care Med* 2005; 29:2119–2124
 68. Inhaled nitric oxide in full-term and nearly full-term infants with hypoxic respiratory failure. The Neonatal Inhaled Nitric Oxide Study Group. *N Engl J Med* 1997; 336: 597–604
 69. Schoenfeld DA, Bernard GR: Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. *Crit Care Med* 2002; 30:1772–1777
 70. Patrick DL, Erickson P: Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation. New York, Oxford University Press, 1993
 71. Molenaar E, van der Heijde D, Boers M: Update on outcome assessment in rheumatic disorders. *Curr Opin Rheumatol* 2000; 12: 91–98
 72. Boers M, Brooks P, Strand V, et al: The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998; 25: 198–199
 73. Panacek EA, Marshall JC, Fischkoff S, et al: MONARCS Study Group. Neutralization of TNF by a monoclonal antibody improves survival and reduces organ dysfunction in human sepsis: Results of the MONARCS trial. *Chest* 2000; 118:88S