Sepsis 1997;1:21–25

Outcome Measures: Methodologic Principles

Gordon H. Guyatt,[1] *Maureen O. Meade*[2]

[1]*Departments of Clinical Epidemiology and Biostatistics and Medicine McMaster University, Hamilton, Canada,*
[2]*Department of Medicine, University of Toronto, Toronto, Canada*

Introduction

Intensivists focus their efforts on improving their patients' physiological status in the hopes of increasing patients' chances of surviving a critical illness, and speeding their recovery. Most of the time, the interventions we use have not been rigorously tested, but their physiological rationale is strong, and patients appear to improve with treatment. Tailoring therapy on the basis of the physiologic effects of interventions will always be a crucial part of ICU practice.

This approach, however, has its limitations. These limitations become particularly evident in the investigation of new drugs or other interventions for critically ill patients. We required randomized trials to inform us that we had probably been harming, and were certainly not benefiting, our critically ill septic patients when we administered intravenous steroids [1]. Highly-touted strategies for weaning patients from ventilators have failed to fulfil their promise [2,3]. While it is still likely that subgroups of patients benefit from having right heart catheters introduced, it is even more likely that the net effect in many patients has been harm [4,5,6,7,8].

Our intuitive reliance on physiological rationale, and the limitations of this approach, raise questions about how we should test our interventions. In this paper, we will deal with issues investigators should consider in choosing outcome measures for clinical trials. We believe that the same principles are applicable to all areas of clinical investigation. Thus, while we try to use as many examples as possible from the intensive care unit setting, and from sepsis research in particular, our examples will come from disparate areas.

Requirements for Suitable Outcome Measures for Clinical Trials

In this section we shall outline the properties that could be found in the ideal outcome measure for a randomized trial designed to test an intervention for possible use in clinical practice (Table 1).

1) Direct, Close Biological Link to the Intervention

Biological interventions are intended, foremost, to produce biological effects. Sometimes, investigators stumble upon clinical effects without understanding the underlying mechanisms. The clinical use of drugs like aspirin and digoxin preceded, by well over 100 years, the discovery of the underlying biology of their mechanisms of action; for opiate narcotics, the hiatus was thousands of years. Fortuitous discoveries of beneficial clinical effects are unusual in modern clinical investigation, and most of the time investigators develop interventions on the basis of an understanding of their biological impact, and a hypothesis about the downstream consequences of this impact.

The closer the link between the biological mechanism and our target outcome, the more likely we are to demonstrate a treatment effect. By this criterion, the most suitable outcome for a drug intended to inhibit the mediators of inflammation is the in vivo inhibition of these mediators. As one begins to measure secondary or tertiary effects that flow from the underlying mechanism, random error or competing influences begin to intrude, obscurring links between intervention and outcome.

Table 1. *Characteristics of an ideal outcome measure*

The ideal outcome measure for a clinical trial of a new therapy would be:

1. Directly and closely linked biologically to the experimental intervention.
2. Important to the patient, and secondarily to health care providers and policy makers.
3. Responsive to the experimental intervention.

2) Importance

While some studies may be interested in biological effects for their own sake, in this discussion we address the assessment of treatments for clinical use. One cannot make a strong argument for use of an intervention

Address for correspondence: Dr. Gordon H. Guyatt, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont L8N 3Z5, Canada.

E-mail: guyatt@fhs.mcmaster.ca

unless it has a beneficial effect on an endpoint that someone values. The viewpoint for the value or importance may be the clinician, the healthcare decision maker, or the patients. Historically, clinical investigators have been ready to adopt their own, or other clinicians', judgement of what is important. Given their training, which emphasized physiology and mechanistic thinking, they tended to be satisfied with biological endpoints, implicitly considering them important in and of themselves.

We would argue that however compelling clinicians may find the felicitous biological effects of a treatment, unless it alters a variable that patients consider important, we cannot advise its clinical use. Our primary goal is, after all, to improve patients' health status. We are, for instance, ultimately interested not in improving the spirometry of ambulatory patients with chronic airflow limitation but in decreasing their dyspnea and improving their capacity to function and enjoy life. As health care resources become increasingly constrained, we cannot ethically justify the allocation of scarce resources to providing an intervention that doesn't benefit patients in a way they think is important.

Health policy-makers may have perspectives that differ from patients. They are likely, for instance, to be more preoccupied with costs, and less concerned with alleviating symptoms than improving more easily measured aspects of morbidity and mortality. Nevertheless, within a public health care system, decision-makers should be acting on behalf of patients, and of the general public. Thus, they cannot dismiss outcomes that patients would consider important. Once again, we arrive at the primacy of the patient in determining what outcomes are important.

How can one judge whether an outcome is important to patients? Imagine suggesting to patients that the outcome in question, and no other, would change with treatment. If patients would be willing to undergo whatever risks and inconvenience are associated with the treatment, and willing to pay (either themselves, or as members of society) the associated costs, then the outcome is important to them.

In general, for physiological outcomes, patients would respond in the negative to this question. Families of critically ill patients may become focused on physiological endpoints because they understand that improvement in cardiac or respiratory function is likely to bode well for their loved one's survival. If they believed that the physiological measures were not tied to the likelihood of survival, they would quickly lose interest.

If there is any doubt about how they would respond, investigators can put the question directly to patients. This is not a question for the ICU bedside, or for families in the "quiet room". It may be a question, however, for the research laboratory. Investigators are increasingly addressing the development of methods for ascertaining patients' values [9].

3) Responsiveness to the Intervention

There is no point in measuring an outcome that will not be affected by the intervention. Many interventions are designed to improve symptoms or lower costs, and we would have no expectation that they would increase longevity. Within the ICU, we would have no expectation that, for instance, a different weaning protocol for ventilated patients would change mortality. This being the case, there would be no point in measuring mortality as an outcome in a clinical trial comparing weaning strategies.

The Competing Nature of our Requirements of Outcome Measures

Unfortunately, our three criteria for an ideal outcome are not complementary. The first and third criteria tend to go together, but important outcomes are unlikely to be highly responsive to our interventions. In fact, the first and second criteria will usually be mutually exclusive.

How can we deal with this dilemma? First, we must prioritize our criteria. We cannot recommend clinicians use a treatment unless we have convincing evidence that patients benefit on a domain—generally either quantity or quality of life—that they consider important. Therefore, if our investigations are intended to direct clinical practice, the criterion of importance is paramount.

If we strive to know that treatments lead to important benefits for patients, we are left with the problem that the important outcomes may be very distant in the causal chain from the biological action of the intervention. They are thus likely to be unresponsive to the intervention.

There are two possible strategies to deal with this problem. One is to choose the important outcome which is most closely linked to the intervention, and therefore most likely to be responsive. A second is to choose what is called a surrogate outcome: an outcome that is not itself important, but is very closely linked to an outcome that is important. If the link between the surrogate and the important outcome is sufficiently strong, one might infer that a treatment that improves the surrogate will also improve the important outcome.

Strategy 1 Use of Outcome Measure Closely Linked to the Intervention

The first of these strategies is limited by the dearth of clearly important outcomes of ICU interventions. The most obvious is prolonging life. Even prolongation of life will not always be important. If the quality of the life that is prolonged is sufficiently poor (in the extreme, a persistent vegetative state) the outcome loses its importance. While in most situations mortality will retain its importance, is sufficiently removed from direct biological effects, then we require large and some-

times prohibitive sample sizes to demonstrate that an interventions prolongs life.

Another category of outcomes that is important to patients is quality of life. Establishing that quality of life is improved by an ICU intervention requires long-term followup and measurement of post-hospitalization status. Conducting such followup entails expense and logistical challenges.

A final category of important outcomes is costs. Patients value reduction in costs, either charges to themselves as individuals, or costs to the society of which they are a part. In the view of the health economist, reduction in costs represents decreased resource expenditure that can now be devoted to alternative uses. Reductions in health care services as a function of cost constraints have made the public highly aware of this issue. If investigators focus on costs as primary outcomes, however, it is incumbent on them to demonstrate that there are no adverse effects on quantity or quality of life.

Given the paucity of important outcomes, and the difficulties in demonstrating treatment-related changes in these outcomes, investigators reasonably consider the second solution, the use of a surrogate endpoint. We will now address the issue of surrogate endpoints in some detail.

Strategy 2: Use of Surrogate Outcomes

We feel comfortable using a surrogate endpoint when we are confident that modification of that surrogate will lead to changes in clinically important outcomes. For instance, investigators and regulatory agencies will license new medications on the basis of their ability to reduce blood pressure. The assumption is that the reduction of blood pressure will result in a decrease in clinically important adverse outcomes such as stroke, myocardial infarction, and death.

There are two requirements for a surrogate outcome to be valid. First, we must demonstrate an association between the surrogate and the important endpoint. This is relatively straightforward. Second, we must be convinced of the etiological relationship between the surrogate and the important outcome in order to justify the inference that changes in the former will result in changes in the latter. This second requirement presents a much greater challenge. We will illustrate some of the difficulties by way of example.

1) Association, not Cause

Surrogate outcomes may be associated, but not causally associated, with important outcomes. For example, there is an association internationally between per capita television sets and coronary artery disease. No one would infer from this association that we could reduce coronary artery disease by eliminating television sets. The reason we wouldn't make this foolish mistake is that there is no plausible biological link between the number of television sets and the incidence of coronary artery disease.

When there is a plausible biological link, however, we are at risk of making incorrect inferences. Consider the association between supraphysiologic tissue oxygen delivery and reduced mortality in critically ill patients. Investigators who first observed this association reasonably suggested that increasing tissue oxygen delivery to supraphysiologic levels might improve outcome [10]. Had they been correct, this would have ultimately allowed clinical investigators to examine the effectiveness of interventions on the surrogate outcome, tissue oxygen delivery, confident that treatments that improved this surrogate endpoint would lower mortality.

Subsequent studies have raised considerable doubts about the initial hypothesis [11,12]. If supraphysiologic tissue oxygen delivery doesn't, in fact, result in decreased mortality, what is the explanation for the association? One plausible explanation is that patients destined to do well are more likely to achieve supraphysiologic tissue oxygen delivery than those destined to do badly. If this is the case then the apparent etiologic agent becomes nothing more than a marker for patients' propensity to do well or badly. In other words, the association is not a causal association.

Another example from the ICU literature addresses our assumption of the causal link between ICU morbidities and mortality. Intensivists generally believe that ventilator-associated pneumonia is an important source of ICU mortality. If their belief in the causal link is correct, reduction in pneumonia will lead to reduction in mortality, and pneumonia becomes a valid substitute endpoint for mortality. However, it is possible that in at least some circumstances the belief is mistaken, and reduction in ventilator-associated pneumonia will not be associated with increased likelihood of survival. Were this the case, we would have limited interest in interventions that reduced pneumonia.

Randomized trials of antimicrobial prophylaxis in the ICU have demonstrated substantial reductions in pneumonia, but borderline effects on mortality [13,14]. This suggests that patients die *with*, rather than *of*, pneumonia. In other words, the intervention may reduce pneumonia in those destined to die anyway, without preventing any of those deaths. Again, the anticipated causal link would have failed.

2) Different Interventions, Different Mechanisms

We have already alluded to the readiness with which we have accepted lowering blood pressure as a surrogate for intrinsically important endpoints such as avoiding strokes and increasing longevity. In patients with essential hypertension only two classes of drugs, beta blockers and diuretics, have demonstrated ability to reduce important outcomes. We have assumed that

if other classes of drugs could lower blood pressure, they would also reduce strokes and deaths.

Recently, the possibility that use of short-acting calcium antagonists to treat hypertension could actually increase adverse outcomes has shaken this assumption [15,16]. It may be important not only whether you lower blood pressure, but the physiologic mechanism by which the reduction is achieved. If this is the case, a surrogate outcome may be adequate for one agent or class of agents, but not for another.

The treatment of heart failure provides a second example of this phenomenon. A variety of drugs decrease afterload on the heart, and as a result increase cardiac output. Only angiotensin-converting enzymes have, however, a marked positive effect in reduction of mortality [17,18]. Again, the physiologic mechanism whereby the surrogate outcome is achieved determines its link to the important endpoint.

3) Spurious Biological Link

For decades cardiologists were ready to use drugs that suppressed non-lethal ventricular arrhythmias under the reasonable assumption that these agents would also suppress lethal arrhythmias. Non-lethal arrhythmias were thus used as a surrogate for fatal events. Unfortunately, when investigators conducted randomized trials looking at the effect of the antiarrhythmic drugs ecainide, flecainide and moricizine—which were extremely effective in reducing non-lethal arrhythmias—they found an increase in mortality [19]. Clearly, the biologic link between the intervention and the surrogate endpoint differed in an unexpected fashion from the link with the important outcome.

4) Unanticipated Adverse Effects

Sometimes (in contrast to the previous examples) the link between the surrogate and the important endpoint may be causal, and hold for the intervention under study, and yet the expected result may not be achieved because of unexpected toxicity. The repeated demonstration that blood lipid levels were associated with coronary artery disease risk led to the hypothesis that lowering lipid levels could decrease myocardial infarction and resulting deaths. If we accepted this hypothesis and the substitute endpoint, interventions would need only show a reduction in lipid levels to be deemed effective. Subsequent randomized trials showed that fibrates reduce both lipid levels and coronary mortality. However, because of an increase in mortality from other causes, the overall effect of the intervention on total mortality was harm [20].

Implications for Outcomes in Sepsis Trials

We have argued that outcomes important to patients in the intensive care unit are limited. Long-term health-related quality of life is important, and may be influenced by morbidity associated with an episode of sepsis. Costs and feasibility considerations make this an impractical measure for most sepsis trials. Mortality is important, but plausible reductions in mortality will require large sample sizes. Reductions in cost are important, but are likely to be achieved only through reductions in the duration of mechanical ventilation, hospitalization, or ICU stay. Reduction in stay may be influenced by many factors, including the availability of beds outside of the ICU. Furthermore, the decision to discharge is open to bias in unblinded studies. Finally, cost reduction studies must reassure us that the intervention is not accompanied by unanticipated increases in mortality.

These limitations of important outcomes lead us to consideration of substitute measures. To the extent that we can reduce the severity of physiologic abnormalities such as those associated with multi-organ failure as measured by scoring systems for organ dysfunction we might expect a reduction in important outcomes, including mortality. This expectation may, however, be compromised by each of the mechanisms suggested above. An intervention may mask or slow the development of physiologic abberations without influencing the ultimate outcome. This may be true for some interventions, but not others. One way to reduce apparent physiological abnormalities in an intervention relative to a control group would be to cause the earlier demise of the sickest patients in the intervention group. Finally, interventions may indeed reduce physiological abnormalities and the associated mortality in some patients while increasing mortality in ways unassociated with pre-morbid physiological derangement (such as sudden, unanticipated lethal ventricular arrhythmias or pulmonary embolism) in others.

On the other hand, these substitute outcomes may retain a strong causal link with mortality, and interventions that reduce the substitute outcome may invariably be associated with important clinical benefit. The only way of knowing whether we can rely on a surrogate outcome is to conduct a series of randomized trials with effective interventions and to demonstrate a consistent association between (i) a reduction in the proposed biological abnormalities, (ii) surrogate endpoints and (iii) survival in each of these trials. Even under these circumstances, it is possible that for the next intervention the underlying mechanism will differ in such a way that the causal link demonstrated in the other trials will fail.

Even if we accept that physiologic endpoints are likely to have a limited role in definitively demonstrating treatment benefit, they may play an important role in the orderly investigation of new treatments. We may be in error if we respond to laboratory or animal studies suggesting a drug decreases mediator release by immediately launching a large trial examining the effect of that drug on mortality. It may be more efficient to conduct a smaller study examining the drug effect on physiological variables. If the results are posi-

tive, this would not warrant clinical application, but rather would justify organization of the larger study looking at effects on mortality.

We conclude that while substitute outcomes may provide strong supportive evidence of the efficacy of interventions in sepsis, they are unlikely to obviate the need for demonstration of unequivocal important benefit in at least one randomized trial. A consistent strong association between improvement with an intervention in both a substitute and intrinsically important endpoint will strengthen the validity of any surrogate outcomes we consider.

References

1. Cronin L, Cook DJ, Carlet J, et al. Corticosteroid treatment for sepsis: A critical appraisal and meta-analysis of the literature. *Crit Care Med* 1995;23:1430–1439.
2. Brochard L, Rauss A, Benito S, et al. Comparison of three methods of gradual withdrawal from mechanical ventilation. *Amer J Respir Crit Care Med* 1994;150:896–903.
3. Esteban A, Frutos F, Tobin MJ, et al. A comparison of four methods of weaning patients from mechanical ventilation. Spanish Lung Collaborative Group. *N Engl J Med* 1995;332: 345–350.
4. Connors AF, Speroff T, Dawson NV, et al., for the SUPPORT Investigators. The effectiveness of right heart catheterization in the initial care of critically ill patients. *JAMA* 1996;276:889–897.
5. Wu AW, Rubin HR, Rosen MJ. Are elderly people less responsive to intensive care? *J Am Geriatr Soc* 1990;38: 621–627.
6. Gore JM, Goldberg RJ, Spodick DH, Alpert JS, Dalen JE. A community-wide assessment of the use of pulmonary artery catheters in patients with acute myocardial infarction. *Chest* 1987;92:721–727.
7. Zion MM, Balkin J, Rosenmann D, et al., for the SPRINT Study Group. Use of pulmonary artery catheters in patients with acute myocardial infarction. Analysis of experience in 5,841 patients in the SPRINT registry. *Chest* 1990;98: 1331–1335.
8. The Ontario Intensive Care Group. Randomized control trial of right heart catheterization in the critically ill. *J Intens Care Med* 1991;144:1133–1149.
9. Llewellyn-Thomas H. Patients' health-care decision making: A framework for descriptive and experimental investigations. *Med Dec Making* 1995;15:101–106.
10. Shoemaker WC, Montgomery ES, Kaplan E, Elwyn DH. Physiologic patterns in surviving and nonsurviving shock patients: use of sequential cardiorespiratory variables in defining criteria for therapeutic goals and early warning of death. *Arch Surg* 1973;106:630–636.
11. Heyland DK, Cook DJ, King D, Kernerman P, Brun-Buisson C. Maximizing oxygen delivery in critically ill patients: a methodologic appraisal of the evidence. *Crit Care Med* 1996;24:517–524.
12. Chittock DR. Russell JA. Oxygen delivery and consumption during sepsis. *Clinics in Chest Med* 1996;17:263–78.
13. Heyland DK. Cook DJ. Jaeschke R. Griffith L. Lee HN. Guyatt GH. Selective decontamination of the digestive tract. An overview. *Chest* 1994;105:1221–9.
14. Sun X. Wagner DP. Knaus WA. Does selective decontamination of the digestive tract reduce mortality for severely ill patients?. *Crit Care Med* 1996;24:753–5.
15. Psaty BM, Heckbert SR, Koepsell TD, et al. The risk of myocardial infarction associated with antihypertensive drug therapies. *JAMA* 1995;274:620.
16. Furberg CD, Psaty BM, Meyer JV. Nifedipine: dose-related increase in mortality in patients with coronary heart disease. *Circulation* 1995;92:1326.
17. The SOLVD Investigators. Effects of the angiotensin converting inhibitor enalapril on survival in patients with reduced left ventricular ejection fraction and congestive heart failure. *N Engl J Med* 1991;325:293.
18. The CONSENSUS Trial Group. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study. *N Engl J Med* 1987;316:1429.
19. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: Effect of ecainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;321:406.
20. Muldoon MF. Manuck SB. Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *BMJ* 1990;301:309–14.