The Design of Randomized Clinical Trials in Critically Ill Patients*

Paul C. Hébert, MD, MHS(Epid); Deborah J. Cook, MD, MSc(Epid), FCCP; George Wells, PhD; and John Marshall, MD

There are a number of difficulties in the conduct of randomized trials in the critically ill. These include difficulties in the definition of diseases and syndromes, a heterogenous population of patients undergoing a variety of therapeutic interventions, and outcomes that may not be able to discriminate between beneficial and risky therapies. Following a brief description of different randomized clinical trials (RCTs) and design philosophies, we outline the effects of different design choices in the complex critical care environment. Once the study topic has been determined to be relevant and important, then the potential investigator must establish whether efficacy or effectiveness will be the focus of the RCT. If an effectiveness design philosophy is chosen, then broad representation of study sites, liberal eligibility criteria, easily implemented intervention study protocols, and patient-centered outcomes should be chosen. The potential investigator wishing to establish efficacy will conduct the study in the centers of excellence and adopt stringent eligibility criteria, rigorous study protocols, and opt for outcomes that will be sensitive to change. In conclusion, we describe some of the major challenges and possible solutions to help a potential investigator through the myriad of difficulties in initiating an RCT in a complex environment.

(CHEST 2002; 121:1290–1300)

Key words: critical care; methodology; randomized trials; study protocols

Abbreviations: MODS = multiple-organ dysfunction syndrome; RCT = randomized clinical trial; TRICC = Transfusion Requirements in Critical Care

Randomized clinical trials (RCTs) have evolved to become the “gold standard” clinical research design used to distinguish the risks and benefits of therapeutic interventions.1–4 In 1948, for the first time, a controlled clinical trial made use of random allocation, a control group, and blinding. Additional principles guiding the design of RCTs were first elaborated by Sir Austin Bradford Hill in the 1960s.5–7

Many important questions regarding the management of critically ill patients have not been subjected to well-designed and executed RCTs. Consequently, clinicians frequently base their therapeutic decisions on suboptimal levels of clinical evidence including observational studies, poorly controlled clinical trials, or laboratory studies.3,8 The complex nature of critical illness, and a host of methodologic challenges have hampered the development and execution of clinical trials in this discipline. In this article, we will outline some of the methodologic issues central to the development and conduct of critical care RCTs.

What Is Unique About Critical Care RCTs?

A unique aspect of critical care research is that patients’ eligibility is primarily defined by location of care in the ICU rather than by the presence of a specific disease. Additionally, many clinical entities in the ICU are often nonspecific constellations of physiologic and biological abnormalities forming syn-
dromes, rather than well-defined disease entities such as breast or lung cancer. Finally, the pathologic processes affecting critically ill patients, resulting in homeostatic disturbances severe enough to result in ICU admission, are often in their late stages at the time of ICU admission.

In critical care, as in other disciplines such as surgery, study interventions are administered in conjunction with other treatment modalities by skilled multidisciplinary teams. The large number of therapeutic interventions required in the care of the critically ill also creates special challenges when performing clinical trials in this field. This is because of the significant number of therapeutic choices faced by the clinical investigator.

The selection of outcomes for RCTs in the critical care setting also poses unique challenges. Until recently, the choice of mortality as an RCT outcome was widely advocated by critical care researchers, the pharmaceutical industry, and government agencies such as the US Food and Drug Administration. A mortality rate, ascertained at 28 days or 30 days, is still considered the “gold standard” for the evaluation of ICU therapeutic interventions applying for licensure. However, in the past several years, a large number of clinical trials with negative outcomes have led investigators to suggest that the choice of mortality as an outcome may have significant limitations. As a primary outcome in an RCT, mortality might be too insensitive to detect the benefits of interventions when small but clinically important differences truly exist. These unique aspects of patients, interventions, and outcomes in critical care must first be considered in light of the research questions being asked and the design philosophy chosen to address these questions.

**Overall Design Approaches**

The ideal RCT establishes whether therapeutic interventions work, and determines the overall benefits and risks of each alternative in predefined patient populations. This is accomplished by minimizing the influence of chance, bias, and confounding through appropriate methodology. In addition, the ideal RCT should attempt to fulfill its objectives with the fewest patients possible (often termed **statistical efficiency**). Unfortunately, these objectives are frequently in direct conflict rather than complementary. More importantly, economic considerations often limit our ability to fulfill all of these objectives. For instance, by maximizing the efficiency of a study, investigators might sacrifice their ability to draw conclusions in clinically important subgroups because of the much smaller sample size.

The most important consequence of these conflicting objectives is that choices made in the design of RCTs must focus on whether an intervention works or whether it results in more good than harm for patients. Trials that attempt to determine therapeutic “efficacy” address the question of “Will the therapy work under optimal conditions?” while trials attempting to determine therapeutic “effectiveness” address the question of “Will the therapy do more good than harm under usual practice conditions in all patients who are offered the intervention?” Clearly, both questions will yield useful information for health practitioners. Efficacy is often established first, and then the intervention is evaluated for its effectiveness. In pivotal RCTs used in the final phase of obtaining regulatory approval (phase III trials), pharmaceutical companies primarily wish to demonstrate that their product has proven efficacy. Rarely are attempts made to demonstrate therapeutic effectiveness in larger RCTs.

The design characteristics of efficacy and effectiveness trials tend to differ considerably (Table 1). As a consequence of design choices, inferences, and threats to the validity of effectiveness and efficacy trials are different. Therefore, one of the first steps in planning an RCT is to determine which of these two design approaches will best reflect the primary study question. Efficacy trials often opt for restricted eligibility, rigorous treatment protocols, and disease-specific outcomes responsive to the potential benefits of the experimental intervention. By using this approach, efficacy studies attempt to maximize internal validity, defined as the extent to which the experimental findings represent the true effect in study participants. Effectiveness trials would enroll most patients, introduce the interventions into the community at large with few controls, and monitor easily measured outcomes that are considered important to patients. As a consequence, the effectiveness approach will attempt to maximize external validity defined as the extent to which the experimental findings in the study represent the true effect in the target population. Hence, there are often trade-offs between the two forms of validity and their design approaches, as efficacy studies maximize internal validity at the expense of external validity while effectiveness studies optimally assess external validity (Fig 1).

Many of the features of efficacy trials are exemplified in a study by Morris and colleagues. These investigators studied two modalities of ventilatory support (extracorporeal carbon dioxide removal vs standardized optimal mechanical ventilation) in patients with ARDS. Interventions were controlled via computerized algorithms, and developed and implemented by a specialized critical care team in selected patients with ARDS. The comparable decrease in
mortality observed following both interventions may be explained by overall improvements in patient care as a result of the extensive deliberations necessary for the development and implementation of the detailed ventilation algorithms. Reproducing similar outcomes in other centers without these careful developmental steps would be very challenging.

The level of control in all aspects of study design exercised by Morris and colleagues may be contrasted to the approach adopted in a study comparing restrictive to liberal transfusion strategies in the critically ill. In the Transfusion Requirements in Critical Care (TRICC) trial, a large number of clinical centers enrolled patients using broad eligibility criteria, followed simple treatment strategies for the administration of packed RBCs, and ascer-

Table 1—Comparison of Study Characteristics Using Either an Efficacy or an Effectiveness Approach When Designing a Study

<table>
<thead>
<tr>
<th>Study Characteristics</th>
<th>Efficacy Trial</th>
<th>Effectiveness Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research question</td>
<td>Will the intervention work under ideal conditions?</td>
<td>Will the intervention result in more good than harm under usual practice conditions?</td>
</tr>
<tr>
<td>Setting</td>
<td>Restricted to specialized centers</td>
<td>Open to all institutions</td>
</tr>
<tr>
<td>Patient selection</td>
<td>Selected, well-defined patients</td>
<td>A wide range of patients identified using broad eligibility criteria</td>
</tr>
<tr>
<td>Study design</td>
<td>Smaller RCT using parallel group or factorial or other approaches (crossover design)</td>
<td>Large multicenter RCT using parallel groups, or factorial cluster</td>
</tr>
<tr>
<td>Baseline assessment</td>
<td>Elaborate and detailed</td>
<td>Simple and clinician friendly</td>
</tr>
<tr>
<td>Study intervention</td>
<td>Tightly protocolized</td>
<td>Implemented in usual clinical practice</td>
</tr>
<tr>
<td>Co-interventions</td>
<td>Tightly controlled protocols for many aspects of care</td>
<td>All therapy based on local practice/experience/minimal control</td>
</tr>
<tr>
<td>Compliance</td>
<td>Compliance essential</td>
<td>Noncompliance expected and considered in sample size/analysis</td>
</tr>
<tr>
<td>End points</td>
<td>Disease-related</td>
<td>Patient-related such as all-cause mortality or quality of life</td>
</tr>
<tr>
<td>Analysis</td>
<td>By treatment received</td>
<td>Intention to treat</td>
</tr>
<tr>
<td>Sample size</td>
<td>Generally &lt; 1,000 and often &lt; 100 patients</td>
<td>All patients included</td>
</tr>
<tr>
<td>Data management</td>
<td>Elaborate</td>
<td>Minimal and simple</td>
</tr>
<tr>
<td>Data collection</td>
<td>Detailed and rigorous</td>
<td>Minimal</td>
</tr>
<tr>
<td>Data monitoring*</td>
<td>Significant interventions and support from research staff</td>
<td>Minimal support and interventions from research team</td>
</tr>
</tbody>
</table>

*Data monitoring refers to the review of source documents and adjudication/verification of outcomes.

Figure 1. Implication of design approaches on internal and external validity.
tained mortality rates and rates of organ failure. This approach would be considered more of a hybrid or combined approach especially when compared to the prototypical example of effectiveness trials, the very large International Study of Infarct Survival trials in acute myocardial infarction. In critical care, there are few examples of such large trials. Indeed, the largest trials have enrolled only a few thousand patients. RCTs in sepsis syndrome and septic shock have been successfully conducted using a hybrid approach rather than a true large, simple trial design. Most of the studies in this field collected significant amounts of data, implemented reasonably detailed but flexible treatment protocols, enrolled heterogeneous patient populations, and evaluated 28-day mortality rates. Therefore, many of the design characteristics adopted in sepsis RCTs could be considered a compromise between efficacy and effectiveness trial approaches. This approach was successfully used in the recently published activated protein C study published in the New England Journal of Medicine.

At this juncture, we suggest that a compromise between the two extreme design approaches would be desirable for most multicenter trials. Although providing important information in cardiovascular and cancer care, large simple trials (effectiveness trials) have not been used successfully in the intensive-care setting.

RCT Design Alternative

Once investigators have chosen whether an efficacy, effectiveness, or a hybrid approach will best answer the research question, there are several design options that may be considered (Table 2). A two-group, parallel design is the most common of RCT design choices. In this design, often the simplest to plan, implement, analyze, and interpret, patients are randomly allocated to one of two therapeutic interventions and followed forward in time. Parallel group designs may also be used to independently compare three or more treatments. This is, by far, the most frequently adopted choice of RCT design.

The use of factorial designs may also be considered when a number of therapies are being evaluated in combinations. For instance, in a 2-by-2 factorial, two interventions are tested both alone and in combination, and compared to a control group (usually a placebo). This means that investigators can efficiently test two interventions with only marginal increases in sample size. In addition, the benefits of treatment combinations can be evaluated in a controlled manner. This design is most useful when interactions are either very strong or nonexistent. Thus, before embarking on a large, more complex factorial study, investigators should expect either strong additive or synergistic effects from combined therapy or none at all. Prospective investigators should realize that detecting interactions is also more difficult and requires much larger sample size as compared to comparison of either therapy with a placebo. Factorial designs have been used very successfully to evaluate thrombolytic therapy in combination with an antiplatelet agent (acetylsalicylic acid) in acute myocardial infarct and unstable angina. Factorial designs have rarely been employed in the field of critical care, despite the number of potential research

<table>
<thead>
<tr>
<th>Type of RCT</th>
<th>Description</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-group parallel</td>
<td>Patients randomized to one of two groups</td>
<td>Simplest approach widely used and accepted</td>
<td>Limited to simple comparisons</td>
</tr>
<tr>
<td>Factorial</td>
<td>Patients randomized to one of four groups; therapy A, or control</td>
<td>Combinations of therapy may be compared</td>
<td>Larger sample size required</td>
</tr>
<tr>
<td>(2 × 2)</td>
<td></td>
<td>Ongoing evaluations of therapy</td>
<td>More complex design; therapy B, therapy A plus B</td>
</tr>
<tr>
<td>Sequential study</td>
<td>Pairs of outcomes continuously compared in patients randomized to one of two therapies</td>
<td></td>
<td>Limited uses (effectiveness only)</td>
</tr>
<tr>
<td>Two-period crossover</td>
<td>Patients allocated to one of two therapies then receive required other therapy in second treatment period</td>
<td>Smaller sample</td>
<td>Not a well-accepted approach</td>
</tr>
<tr>
<td>N-of-1 study</td>
<td>Single patient sequentially and repeatedly receives a therapy and a placebo</td>
<td>Optimal method of determining if a therapy is beneficial for a given patient</td>
<td>Sample size unpredictable</td>
</tr>
<tr>
<td>Cluster design</td>
<td>Groups of patients are randomized</td>
<td>Ideal for program or guidelines evaluation</td>
<td>Concealment of randomization a concern</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Major concern with carryover effect from first to second period</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Results not generalizable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Difficult in unstable patients</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Very labor intensive</td>
</tr>
</tbody>
</table>
questions for which this approach may be optimal, such as nutritional studies and ventilation studies where two or more nutrients or ventilation strategies may be compared alone or in combination.

Factorial designs imply concurrent comparisons between at least two therapies. It is also possible to implement a design that compares interventions sequentially. For example, one might compare two therapies in the early treatment of a disease followed by the evaluation of a second intervention(s) in the late phase of care several days later. One such example is the approach adopted by the National Institutes of Health ARDS trials network, where the RCT evaluated two ventilatory strategies (12 mL/kg vs 6 mL/kg of tidal volume) in conjunction with ketoconazole, 400 mg/d, vs placebo. The optimal use of this design requires that the outcome from the initial portion of the trial be ascertained prior to initiation of the second study.

Both the simple parallel-group design and a factorial design are generally implemented with the understanding that the sample size is fixed according to pre-established assumptions prior to the commencement of enrollment. There are other experimental designs that are more responsive to patient outcomes as the study progresses. Sequential designs set boundaries for significance levels that consider the increasing number of comparisons and sample size throughout the study. True sequential studies randomly allocate patients to receive one of two therapies. Pairs of patients are then sequentially compared. The study is terminated as soon as one of the significance boundaries is crossed. This design was successfully used by Meduri and colleagues to establish the benefits of IV methylprednisolone in the treatment of late-phase ARDS. The authors demonstrated that high-dose methylprednisolone was associated with improvement in lung injury, multiple-organ dysfunction syndrome (MODS) score, and mortality in 24 patients. This study question had all of the necessary attributes for this design. The population was well defined and homogenous; more importantly, the study end points were easily ascertained within a very short time following randomization. In critical care, the sequential design may be limited to select patients in whom a dichotomous outcome is promptly available for analysis, for example, progression of disease or intubation status (yes or no). Therefore, this approach may be considered when performing efficacy evaluations. One of the major concerns with the design may be its inability to conceal the randomization process and the uncertainty of not knowing the exact sample size in advance. From this methodology, several biostatisticians have developed methods of performing interim analyses in large clinical trials referred to as group sequential methods.

Another RCT design option particularly amenable to an efficacy evaluation is a two-period crossover study in which patients are used as their own controls. In a two-period crossover trial, patients are randomized to one of two therapies for a fixed period of time and then proceed to receive the other therapy in a second comparable interval. Significant gains in efficiency are made by minimizing “between-subject” variability in this manner. Cooper et al determined the hemodynamic consequences of sodium bicarbonate by randomly allocating critically ill patients with lactic acidosis to receive either 2 nmol/kg of sodium bicarbonate or an equimolar amount of sodium chloride, followed by the other therapy. The authors determined that both sodium bicarbonate and sodium chloride equally increased left ventricular filling pressures and cardiac output without significantly changing arterial BP. One of the fundamental assumptions underlying this design is a “carryover effect”: a treatment effect from the first period does not persist through the second period introduced. As a second example, Wright and colleagues examined whether bronchodilators decreased airflow resistance in patients with ARDS. The authors demonstrated that airways resistance, a short-term physiologic end point, was decreased in patients with ARDS. This was the optimal design choice given the reversibility of the outcome and the intervention. In our example, the administration of sodium bicarbonate in the first period may have altered acid-base status or calcium homeostasis in the second treatment period, potentially resulting in a bias toward the null hypothesis. Crossover studies are therefore best suited to relatively stable conditions (stability required during the study), interventions with rapid onset of action and a very short half-life (biological effect must disappear prior to second treatment period), and rapidly modifiable end points such as hemodynamic and respiratory measures.

All designs discussed so far have described the evaluation of interventions for individual patients. However, it is sometimes necessary to evaluate therapies, protocols, guidelines, or treatment programs for groups of individuals. Using this design, groups such as ICU and physician practices, often referred to as clusters, are randomized to receive alternative interventions. A cluster design may be the most appropriate for evaluating interventions such as antibiotic protocols, weaning guidelines, or early discharge programs. One of the major concerns is the possibility of large variations between clusters that may make it difficult to detect actual differences between therapies. Partial clustering in the allocation of patients was used in a study of hyperbaric oxygen therapy for acute carbon monoxide poisoning. With access to a single chamber, the investigators were faced with only enrolling one
patient at a time or allocating all patients who were poisoned in the same incident to be treated at the same time in one cluster. In this well-conducted RCT, the authors not only did not find any benefit to hyperbaric oxygen but may have detected the possibility of harm.

THE PATIENT POPULATION IN CRITICAL CARE RCTS

One of the difficulties faced by investigators is that potential study participants must usually be admitted to an ICU in order to be considered critically ill. Eligibility or selection of patients based on a location creates difficulties in precisely defining the entry point into the clinical trial. For example, let us assume that an investigator states that patients will remain eligible only for the first 24 h following ICU admission. By adopting a broad definition for the term ICU, the clock may start ticking from the time of admission to the recovery room, emergency department, the ICU itself, a high-dependency unit, or the ICU of another institution. If a time restriction helps define the study population, then the investigator must clarify whether each of these venues of care is suitable given the target population and study question. If possible, time restrictions are more likely to be meaningful if based on a specific point in the patient’s disease rather than a location alone. Further, the use of time restrictions is optimal when there are easily defined milestones in the disease process. For instance, several studies evaluating immunotherapy in septic shock make use of a time restriction initiated by the development of hypotension.

Second, selection is often based on disease definitions, comprised of a constellation of physiologic and other biological abnormalities rather than well-defined disease entities. One of the major challenges facing critical care investigators in the past several years has been to develop definitions for disease processes such as sepsis, septic shock, and ARDS. Unfortunately, few diseases in critical care are simple and clearly defined by a pathologic process such as myocardial infarction or a diagnosis of cancer. Consequently, clinical syndromes are often characterized by using alterations in physiologic, immunologic, and biochemical parameters. Recently, expert opinions from consensus conferences have helped in the formulation of these definitions. Prior to undertaking any clinical trial, it is not only important to understand the pathophysiology of the disease or syndrome, but investigators should also have a sound appreciation of its epidemiology. A detailed understanding of disease incidence and risk factors, as well as some of the limitations in using the definition of the clinical syndrome to select patients, is essential in the planning of an RCT. In assessing proposed or established syndrome definitions, investigators should question their validity and reproducibility (or reliability) as well as whether definitions are sufficiently well established and user friendly to warrant their use in a clinical trial.

In addition to concerns related to location and disease definitions, the choice of either an efficacy or effectiveness approach will have a substantial impact on the selection of the study population. Specifically, in choosing an efficacy approach, investigators usually perform the study in a well-defined patient population where the intervention has the highest probability of demonstrating an effect. This may be done by narrowly defining the patient population through the use of restrictive eligibility criteria and disease definitions as well as selecting specialized centers with clinical expertise in the field. Choosing a narrowly defined study population will decrease overall variability attributed to patient selection but may potentially hamper patient recruitment and jeopardize the generalizability of the study results. Despite these concerns, this approach has been successfully used by the National Institutes of Health ARDS Network.

When defining the eligibility criteria for an effectiveness trial, investigators should consider utilizing more liberal criteria in a wide range of clinical settings. Thus, as the study is being designed, medical or surgical critically ill patients with a broad range of primary diagnoses (or underlying conditions) from a range of tertiary-care centers might be considered for enrollment in the study. Liberal selection of study centers and more permissive eligibility have been adopted in many studies performed by the Canadian Critical Care Trials Group, including the TRICC trial, and the study by Cook et al comparing sucralfate and ranitidine.

On the spectrum between highly selected patients (efficacy) and all ICU patients (effectiveness), we suggest that critical care investigators should consider a number of factors in making the decision. In practice, considerations such as the spectrum of biological activity of the intervention (wide or narrow), funding and resource constraints, the prevalence of the specific condition or disease process, the frequency of the primary outcome, as well as the scientific or clinical interest of the investigative team, will impact on choices made in the selection criteria for potential study participants and study sites. Targeting high-risk patients and possibly centers where the condition of interest is more prevalent may be used as a strategy to maximize the use of study resources.

STUDY INTERVENTIONS

The complexity and multiplicity of interventions used in the care of critically ill patients present
unique challenges for the clinical investigator planning an RCT. In general, as the number of potential interventions increase, available options for the pursuit of optimal patient care exponentially rises. A major consequence of complex care is increased biological variation and practice variation that ultimately increases experimental noise, and thereby makes it more difficult to detect therapeutic benefits when they are truly present.

To cope with concerns regarding the complexity of care, investigators must consider the degree of control or constraints imposed on experimental and nonexperimental interventions that will be adopted in an RCT. Experimental constraints on study interventions can be implemented by instituting rigorous treatment protocols or by the selection of study centers. Thus, a number of choices face the clinical investigator. For example, should the administration of antibiotics be standardized in a septic shock trial? Should the ventilatory management be tightly controlled in an RCT of a weaning intervention?

As outlined in the previous section, the answers to these questions will partially depend on whether investigators wish to evaluate therapeutic efficacy or effectiveness. Elaborate study protocols detailing the use of experimental and nonexperimental therapies characterize efficacy evaluations. It is expected that the development and implementation of elaborate treatment protocols will decrease overall variability attributed to the confounding influence of co-interventions. Decreased variation in the study may enable the detection of smaller clinically important treatment differences, if truly present. However, the development of treatment protocols themselves may improve patient care by decreasing unnecessary practice variation, by increasing the general knowledge of practitioners, or by adopting evidence-based practices in participating centers. Just as critical paths are not easily implemented at a site that did not participate in their development, elaborate study protocols may not be easily adopted in a wide variety of practice settings. Also, elaborate treatment protocols may jeopardize accrual of study participants and physician compliance.

An alternative approach would be to let therapeutic decisions (other than the experimental therapy) devolve onto the attending physician. Allowing the attending physician complete autonomy will maximize the generalizability (increased external validity) of study results but potentially increase the effect of confounding from co-interventions (decreased internal validity). The number and intensity of co-interventions invariably magnify underlying random error in an RCT, potentially leading investigators to falsely conclude that there are no benefits to a promising new therapy. To cope with increased variation, investigators must plan to substantially increase the sample size because of diminished benefits of the experimental therapy. Selecting specific ICUs to participate in an RCT may also be a worthwhile method of ensuring compliance with the protocols for the study intervention and co-intervention.

For all interventions, the blinding of the care team to the study interventions should be seriously considered because this study maneuver has been shown to minimize co-interventions and biases in ascertaining outcomes. Although blinding maneuvers are vitally important, feasibility and patient safety sometimes do not permit blinding of the study intervention. This is more problematic in nonpharmaceutical interventions. However, a number of examples exist in which study investigators have successfully implemented complex blinding maneuvers without jeopardizing either safety or feasibility. Pilot studies are recommended to determine whether blinding can be maintained safely and successfully. If, during a pilot study, caregivers are able to discern which treatment is being administered, then the blinding process should be reconsidered, improved, or possibly abandoned. When double blinding is not possible, for example in the evaluation of surgical techniques and new devices, other safeguards to minimize bias should include the selection of objective outcomes as well as independent and blinded outcome assessments if more subjective outcomes are chosen. In order to minimize differences in therapy due to inability to blind, regimented treatment protocols should also be considered. In addition, the influence of co-interventions can be tested post hoc using multivariate statistical techniques.

When treatment protocols are complex or controversial, compliance may also be a concern. We suggest investigators consider some or all of the following strategies to improve adherence to study protocols. There are several ways to increase compliance with study protocols: (1) by making study protocols simple and easy to implement; (2) by developing the protocols with as many stakeholders as possible; (3) by extensive dissemination of the study protocol and its rationale in participating study centers; (4) by obtaining formal agreements to respect the protocol from all ICU physicians and other potential collaborators; (5) by implementing a mechanism to minimize crossovers (such as consulting the site investigator and the study chair prior to crossing over), and by developing objective crossover criteria when this is a concern.

**Outcome Measures**

In most clinical trials, a number of potential outcomes, both fatal and nonfatal, are considered by
the clinical investigative team. An outcome is defined as a measurement (ie, arterial BP) or an event (ie, death) potentially modified following the implementation of an intervention. If all are given equal consideration, concerns arise about multiple comparisons and interpretation of a study with heterogeneous findings. Thus, it is important to choose a primary outcome that will determine the therapeutic success or failure of an intervention, as well as secondary outcomes that will provide supportive evidence in secondary analyses. As a corollary, a predefined hierarchy implies that the investigators believe that clinically or statistically important differences in secondary outcomes, in the absence of important changes in the primary outcome, will not be interpreted as strong evidence of therapeutic benefit. The primary outcome is also essential in determining the sample size requirements in a clinical trial. Thus, once a decision has been made to determine either therapeutic efficacy or effectiveness (or possibly a hybrid approach), the second task facing investigators is ranking outcomes as primary and secondary (Table 3).

The choice of study outcome is one of the most important design considerations to be made by our investigation. There are, however, a number of factors that should be considered prior to selection of an outcome. The primary outcomes should be considered clinically important and easily ascertained. By fulfilling these two criteria, the investigator will have a much greater chance of influencing clinical practice once a study has been completed and published. All-cause mortality, arguably the most important outcome following critical illness, is generally considered easy to ascertain. Even mortality rates may be influenced by factors such as the period of ascertainment (ie, 28-day vs 60-day mortality). Investigators must be able to determine if an event has occurred or not in each study participant. For example, it may be difficult to determine if a patient has had a myocardial infarction while in ICU because ECGs and serial myocardial enzyme values may not be diagnostic of a cardiac event.

Outcomes should also measure what they are supposed to measure (validity), they should be precise, and they should be reproducible. There is little doubt that all-cause mortality would meet all of these criteria, but cause-specific mortality and a quality-of-life scale may or may not be a valid and reliable assessment of a patient’s health status following an episode of septic shock. Finally, an outcome must be able to detect a clinically important true positive or negative change in the patient’s condition following a therapy. In critically ill patients, the ability to discriminate or detect the potential benefits of therapy may be less than optimal using mortality rates ascertained at 30 days as the primary outcome.11 Because few sepsis studies have shown any significant impact on 30-day mortality, many investigators11,47 have suggested that other outcomes should be considered in RCTs evaluating therapeutic effectiveness.

However, the ability to discriminate between beneficial and risky therapies may be modified by specific design choices including many related to outcomes. Discriminability can easily be increased by increasing the sample size. Using mortality as an example of primary outcome, the sample size in a clinical trial comparing two therapies is based on the baseline event rate, the expected incremental benefit, the level of significance (α), and the power to detect differences (1-β). Establishing the anticipated incremental benefit of a new therapy is vitally important because of the enormous sample size repercussions. A sample size calculation for an RCT requires that the investigators establish the minimum therapeutic effect detectable within the trial. This difference in outcomes between interventions is referred to as the minimally important difference or minimal clinically important difference. The minimally important difference is essentially establishing the level of discrimination in the study population exposed to the interventions given acceptable levels of type I and type II error and the baseline event rate. Too often, investigators calculate a sample size based on very large and unrealistic expected differences in outcomes. To determine a plausible effect size, investigators should ask themselves the following questions: (1) what difference or incremental benefit can be realistically expected of the experimental therapy (anticipated biological effect of therapy); (2) are the required number of patients available to participate in the clinical trial (feasibility); and (3) how much of a survival benefit, given the added costs and expected side effects of therapy, would be required for clinicians, patients and administrators to adopt a new therapy (overall benefit of therapy)?

As a concrete example, let us assume that a given study population has an expected mortality rate of 25% in the standard-therapy group while the exper-

---

Table 3—Guides to the Choice of Outcome Measure in an RCT

<table>
<thead>
<tr>
<th>Question</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the outcome causally related to the consequences of the disease?</td>
<td>This difference in outcomes between interventions is referred to as the minimally important difference.</td>
</tr>
<tr>
<td>Is the outcome clinically relevant to the health-care providers and/or patients?</td>
<td>Establishing the anticipated incremental benefit of a new therapy is vitally important because of the enormous sample size repercussions.</td>
</tr>
<tr>
<td>Has the validity of the outcome (for complex outcomes such as scoring systems or composite outcomes) been established?</td>
<td>A sample size calculation for an RCT requires that the investigators establish the minimum therapeutic effect detectable within the trial.</td>
</tr>
<tr>
<td>Is the outcome measure potentially able to discriminate between patients who benefit from a therapy and those in the control group?</td>
<td>This difference in outcomes between interventions is referred to as the minimally important difference or minimal clinically important difference.</td>
</tr>
</tbody>
</table>
Experimental therapy is expected to decrease mortality by an absolute difference of 12.5% (a 50% relative risk reduction). The total number of patients required would approximate 250. Most therapies used in the ICU would not be expected to decrease mortality so dramatically. More realistic expectations may be in the range of a 5% absolute decrease (a 20% relative risk reduction), which would require a total sample size of 2,200 patients, respectively, if the baseline mortality was 25%. Investigators need to consider whether an absolute incremental benefit in the range of 5 to 10% is attainable using the experimental therapy. If not, another more discriminating outcome should be sought.

As an alternate approach, discriminability may be improved by altering the ascertainment period. In other words, mortality rates may be determined at 24 h, 7 days, or at ICU discharge, rather than longer time intervals such as 30 days or 6 months. The timing of the ascertainment will have opposing influences on its ability to discriminate and its clinical relevance. As the ascertainment period of mortality is lengthened, the clinical importance of the outcome is increased. However, as the time from the administration of the therapy to the assessment of mortality is increased, the relationship between the effects of the therapy and the outcome may be confounded by extraneous factors and intervention. Therefore, the ability of an intervention to discriminate between groups on the basis of mortality may decrease as time progresses (Fig 2).

Investigators have almost exclusively used mortality rates to quantify survival following critical illness. However, continuous outcomes such as health-related quality of life might be incorporated into the design of the RCT in order to improve discrimination. Quality-of-life assessments attempt to quantify a subjective sense of patient functioning and well-being. Several authors have described quality of life following admission to ICU. Although health-related quality of life has been appropriately incorporated in RCTs dealing with chronic illness, quality-of-life assessments have not been widely used by critical-care trialists. Although interesting, quality-of-life measures may not discriminate between groups receiving different study interventions because of an inability to assess patients prior to being critically ill and therefore inadequate comparison of quality of life before the trial begins. The MODS score tabulates the degree of abnormality in six major organ systems following critical illness. Scores have been developed, in part, as a means of improving discrimination between interventions in the ICU. MODS scores may be combined with mortality by assigning all patients who die the maximum allowable weighting or score. This approach improved discrimination between transfusion interventions in the TRICC trial.

**Conclusion**

In this article on RCTs in the critical care setting, several major RCT design characteristics were discussed.
cussed. We outlined issues of special interest to critical care investigators related to RCT design approaches, disease definitions, patient selection, study interventions, and outcome measures. Although RCTs provide the most unbiased and accurate assessment of the efficacy and effectiveness of therapeutic and preventive interventions, they remain challenging and expensive to conduct. As more research groups form to address unanswered therapeutic questions in critical care, investigators will invariably better understand strengths and limitations of different RCT design characteristics in critical-care trials (Table 4).

ACKNOWLEDGMENT: We thank our students, teachers, and colleagues who contributed many of the ideas outlined in this article. We also thank Drs. Peter Tugwell, Andreas Laupacis, and Arthur Slutsky for reviewing this article, and Christine Piché for secretarial support. We also thank our collaborators, with whom many of the studies used as examples in this series would not have been possible.

Table 4 — Suggestions When Planning an RCT in Critical Care

Ensure that similar studies have not been conducted elsewhere or are currently underway (a systematic review and discussion with experts in the field are essential). Although self-evident, multiple systematic reviews have revealed that researchers often proceed with a study when the answer is already available.

Explicitly determine whether you are primarily interested in establishing therapeutic efficacy or effectiveness. Efficacy can be evaluated in a pilot study, and effectiveness should be the focus of a larger trial.

Whenever possible, undertake an RCT as part of a broader research program.

If the study intervention is complex (or risky) or if other aspects of study feasibility are questionable, a pilot study should be considered.

Whenever possible, investigators should adopt simple rather than complex designs (two-group parallel design as opposed to factorial design).

The study population should be tailored to the intervention and design approach.

Ideally, the study intervention and treatment protocols should not aim to substantially modify or affect usual clinical practice.

Given the complexity of RCTs in ICU, data collection should aim to clearly describe the study population, and describe co-interventions and all major study outcomes. Minimal data collection strategies are not advocated in this field.

In choosing primary study outcome, investigators should focus on patient-oriented outcomes rather than surrogate or biological markers.

If you are planning a seminal RCT, you may only have one chance to get it right. Therefore, when making compromises, always opt to answer question that most clinicians consider important.

In establishing the minimally important difference (necessary for sample size determinations), select a potentially achievable goal.

Always involve experienced clinical trials methodologists, biostatisticians, and representatives from clinical disciplines involved in the intervention as part of your research team.

The support of a research network will greatly enhance your chances of successful completion of a large RCT.

REFERENCES

3. Guyatt GH, Sackett DL, Cook DJ. Users’ guides to the medical literature. II: How to use an article about therapy or prevention; A. Are the results of the study valid? JAMA 1993; 270:2598–2601
10. van der Linden W. Pitfalls in randomized surgical trials. Surgery 1980; 87:258–262


