Randomized Controlled Trials for Evaluating Surgical Questions

Eric K. Fung, MD; John M. Lore´, Jr, MD

Objective: To discuss some of the obstacles inherent in the design of the randomized controlled trial (RCT) that the surgeon must confront and options to minimize these obstacles.

Data Sources: The literature was searched for articles discussing RCTs using MEDLINE from 1966 to 1998.

Study Selection: Studies relevant to the general use of RCTs for evaluating surgical questions were selected.

Synthesis: Several problems inherent in RCTs were noted: (1) ethical considerations, (2) difficulties in patient accrual, (3) patient preferences, and (4) variability in surgical proficiency/technique. Some means of minimizing these problems are (1) the concept of clinical equipoise, (2) multicenter trials, and (3) stratified sampling of patients. Alternatives to the classic RCT are discussed, namely, the randomized consent design and the patient preference design.

Conclusions: The nature of the RCTs is that they are difficult to use to evaluate surgical techniques. Some options are available to minimize these difficulties. Designing and conducting RCTs to evaluate surgical interventions require careful planning and some compromises. Unless the previously mentioned criteria are applied, the validity of the RCT can be considered no greater than that of other trials.


Following a recent article regarding endoscopic esophagodiverticulostomy, a reviewer remarked, “Had the authors randomized their patients to either this approach, or an open approach, the issue could have been put beyond scientific doubt.”

This statement raises the questions (1) why do we do randomized controlled trials (RCTs)? (2) why don’t we always do RCTs (ie, what are the obstacles that prevent surgical RCTs)? and (3) what solutions are available for minimizing these obstacles?

DEFINITION OF AN RCT

The term RCT is in some ways self-explanatory. First, investigators randomly assign subjects to 1 of 2 or more treatment options. Second, the trial must be controlled, in the sense that 1 of the treatment options serves as the basis for comparison with the other—it is usually designated as the control. Along with these 2 fundamental characteristics, RCTs are also by definition, prospective, usually double-blinded, and evaluated based upon statistical conditions established prior to initiation of the trial (Figure 1).

RATIONALE FOR RCTs

Surgeons are often interested to determine the efficacy of one technique compared with another. Although the case-controlled series is more often used, many consider RCTs to be the most scientific means of investigation to this end. There are 2 major reasons why the medical community considers the RCT more scientifically rigorous than other methods. First, because the subjects are randomized to different treatments, the investigator is unable to exert a selection bias upon the study. Second, due to the random allocation of patients, confounding factors are likely to be randomly distributed among the different groups, thereby minimizing their influence. For these reasons, the RCT provides the strongest evidence that the result was due to the intervention.

OBSTACLES TO RCTs

Recently, an editor of Lancet published a commentary entitled “Surgical Research or
Comic Opera: Questions, but Few Answers. As evidenced by the title, the suggestion of the author was that the level of scientific rigor in surgical research was lacking. In particular, the author suggested that surgical investigators do not use the RCT frequently enough. The numerous, exasperated responses to this editorial serve as testimony to the frustration of surgeons who appreciate the merits of the RCT, but find that it is often an unwieldy tool. Some of the more vexing problems of RCTs follow.

Ethical Considerations

One of the requirements for conducting an RCT is that a null hypothesis be formed. The null hypothesis is a statement of equivalency between the different treatments that the investigator seeks to prove or disprove. That is, both treatment arms yield the same outcome. If the investigator believes that the null hypothesis is invalid because one of the treatments is superior, then the investigator would be subjecting a group of patients to a treatment known to be suboptimal. As such, the trial cannot proceed.

The null hypothesis presents other dilemmas to practicing physicians. One of these is that they must admit to the patient that they do not know which treatment is superior. While truthful, this is a difficult admission for most physicians to make. Furthermore, the declared lack of understanding may undermine the patient’s confidence in his or her physician. Another problem is that the randomization of patients to either of these “equivalent” treatments prevents physicians from using their personal judgment in determining the treatment for their patients. This detachment from the decision-making process is uncomfortable for both patient and physician as it subverts the interpersonal relationship to some extent.

A different ethical argument against the RCT is that it detracts from the quality of care received by the individual patient. The fear is that the necessity of generating good data for their study will distract the physicians from the patient’s needs. The physician may thus be sacrificing the good of the individual patient for that of future patients.

Patient Accrual

Perhaps nothing is so disheartening to an investigator than to invest significant amounts of time and energy into a study only to find that it lacks sufficient statistical power to provide conclusive results. The prospect of enrolling sufficient numbers of patients for a valid RCT is daunting for several reasons. First, some of the illnesses that we aim to treat surgically are relatively rare. It may be so difficult to find patients with certain conditions that the investigator is hesitant to use a control technique on these patients rather than the new technique.

A similar problem is illustrated in an article attempting to determine the role of hyperfractionated radiation therapy in the treatment of head and neck cancer with or without concurrent chemotherapy for locally advanced disease. The authors state, “despite randomization, there were some imbalances within the two treatment groups.” Though the authors have formulated an excellent study design, the relatively small numbers of patients who fit their specifically defined clinical scenario leads to difficulty when trying to achieve statistical balance. If given an infinite number of patients who fit their criteria, randomization would theoretically lead to balanced numbers between the 2 groups.

Another reason surgeons have difficulty generating large numbers of subjects is the constraint of operating schedules. Whereas a medical investigator may administer a pharmacologic agent to several patients simultaneously, the surgeon is restricted by the availability of the operating rooms, anesthesiologists, and ancillary staff. As such, the time required to perform a technique on sufficient numbers of patients (especially in the case of lengthy operations), as well as the time required in providing adequate follow-up, may prove to be a prohibitive factor.

Patient Preferences

Related to the problem of accruing subjects is the problem of patient preferences. Often, patients are referred to specific surgeons who are pioneers of new techniques, and are unwilling to settle for the older methodology. For example, the patient who travels a great distance to be healed by an expert in endoscopic techniques would likely have little interest in being a control subject who receives the more conservative open procedure. Conversely, the patient may prefer to have a more conventional procedure performed, rather than to be a “guinea pig” for a new technique. In either event, the randomization process precludes the patient from making a choice in the matter.

Variability in Surgical Proficiency/Technique

One means of increasing the number of patients enrolled in a study is to increase the number of surgeons who perform the procedure. Unfortunately, this strategy often leads to great variability in the sample, as the individual abilities of different surgeons varies. This difference is most pronounced when a new technique is being evaluated, as the learning curve is often steep.

Since experience plays such a large role in the ability of the surgeon, the evaluation of new techniques may produce unrealistically poor results as surgeons learn the procedure. This situation is unlike that for medical investiga-
tion, in that as drugs become more widely used, we generally become aware of an increased side-effect profile, whereas when a surgery becomes more well known, the resultant familiarity leads to a decreased complication rate.\textsuperscript{13} Furthermore, during the course of the study, surgeons may refine the original technique with subtle technical advances. These changes in technique may detract from the consistency of the study and threaten its validity.

POTENTIAL SOLUTIONS

Despite these factors, some physicians are proponents of forced randomized trials.\textsuperscript{14,15} That is, only those trials that are in the randomized controlled format would be funded. The reasoning is that the RCT remains the best means for examining those questions that remain to be answered in medicine and surgery. Trying to answer questions through other forms of study can lead to misleading results. As such, we must find ways to minimize the difficulties inherent in RCTs.

Clinical Equipoise

The ethical question surrounding the null hypothesis was clarified by Freedman\textsuperscript{16} who postulated that the individual examiner can believe one treatment to be superior given limited evidence, so long as supporters of the alternative treatment have some evidence to support their position. He termed this condition “clinical equipoise.” This solves the dilemma for the investigator who has a few cases suggesting that the new therapy is superior.

Parallel Care

To solve the problem of the diminished physician-patient relationship due to the randomization process, it has been proposed that a second physician, not affiliated with the study, can also follow-up with the patient. This parallel care will help ensure that the patient is not seen only within the context of the study.

Historical Controls

One means for effectively increasing the sample size, as well as avoiding the ethical dilemma of the null hypothesis, is to use historical controls. In these types of trials, the patients in the study are compared against published results of previous, similar trials. The use of historically controlled trials is discouraged from a scientific standpoint, as historical controls are not as closely matched with the subjects as are randomized controls. Furthermore, it has been shown that historically controlled trials are much more likely to show positive results than are RCTs. This difference is largely due to the poorer clinical results for historical control patients than those for randomized controls.\textsuperscript{37} Thus, the results of a historically controlled trial must be considered less compelling than those of a RCT.

Multicenter Trials

A relatively obvious solution to insufficient patient accrual is to use multiple centers to draw upon a larger sample population. An important caveat in using the large-scale multicenter approach is that the study must be made relatively simple. One analyst suggests, “When you make it 10 times bigger you have to make it 10 times simpler because there isn’t 10 times the money to do them.”\textsuperscript{12} Another caveat is that more variability in surgery and surgical techniques is thereby introduced.

Stratified Randomization

With multicenter trials comes the problem of differing skill levels of surgeons. This discrepancy between surgeons can be accounted for by stratified sampling.\textsuperscript{1} That is, the surgeon treating the patient is recorded with the patient’s data. This stratified analysis will allow the investigators to adjust for the experience of the surgeon with a particular technique. In doing so, the results can be analyzed separately to look for trends in patient outcomes. Similarly, in multicenter trials, results may be stratified by institution.

Randomized Consent Design (Prerandomization)

To avoid the discomfort of telling patients that their clinical fate is dependent upon random allotment, Zelen\textsuperscript{18} devised the randomized consent design for clinical trials (Figure 2), also known as prerandomization. In this design, the patient is randomized prior to the informed consent stage. Thus, the physician presents the treatment that has been selected and asks for informed consent, fully knowing to which treatment branch the patient has been randomized. A further variation upon this design is the “single consent design” in which only those patients assigned to the experimental treatment are consented to be part of the study. The randomized consent design is known to have some ethical and scientific problems. However, this design may be quite useful in increasing patient accrual in the case of physicians who prefer to tell their patients definitively which treatment they are consenting to receive.\textsuperscript{19}

Patient Preference Trial Design

Some have argued that the classic RCT is not a realistic representation of the clinical setting in which both the
patient and physician have free choice. They propose a study design in which the patient is offered the treatment options (Figure 3). Patients who have a clear preference are given their preferred treatment. Patients without a preference are randomized. The proponents of this study design point out that those study participants who chose their treatment are not only more realistic representations of future patients, but they may give additional information regarding their decision-making process.

The RCT stands as the most rigorous means for evaluating new treatments. Because of the randomization of patients into different treatments, these trials can demonstrate a clear effect from a treatment. Unfortunately, several barriers to RCT sexist, including ethical dilemmas, difficulties in patient accrual, patient preferences, and variability in surgical proficiency/technique. Some ways of confronting these problems include the principle of clinical equipoise, large-scale multicenter trials, and stratified sampling of participating surgeons or institutions. Variations on the RCT include the randomized consent design and the patient preference design. Both variations are flawed in comparison with the RCT. They offer, however, a strong scientific grounding without as many obstacles as the RCT.

Since all methods of sampling a population are an attempt to create a statistical abstraction of the truth, it should be remembered that no trial method is guaranteed to provide absolutely factual information. Although the RCT is the "gold standard" of trial design since it is the most scientifically rigorous method, the conclusions drawn from an RCT can, nevertheless, be erroneous. Inversely, though nonrandomized trial designs may introduce undesired biases into the data, the conclusions drawn from these studies may be quite useful.

Accepted for publication November 13, 2001.
This study was presented at Eastern Great Lakes Head and Neck Oncology Conference, Toronto, Ontario, November 7, 1998.

Our appreciation to Sol Kaufman, PhD, for his assistance in writing and editing the manuscript.
Corresponding author and reprints: John M. Loré, Jr, MD, Room 208, Sisters of Charity Hospital, 2121 Main St, Buffalo, NY 14214.

REFERENCES