Randomized clinical trials for new surgical operations: Square peg in a round hole?

Joel D. Cooper, MD

In 1995, *Time* magazine published an article entitled “Are Surgeons Too Creative?” that examined the question of whether new operations should be regulated. The public and other interested parties are somewhat sympathetic to this article’s premise, namely, that new surgical procedures should require a similar level of evidence as is required for introduction of new drugs.

In the 1990s, there was a new impetus in the movement for “evidence-based medicine.” This phrase appeared for the first time on a keyword search of PubMed in 1992 when there were 2 citations using this phrase. By 2001 a similar search revealed 2000 such citations.¹ Clinical trials became the basis for evidence-based medicine, and the randomized clinical trial (RCT) was regarded as the highest standard of all. Evidence from clinical trials was increasingly used to establish clinical guidelines, defined as systematically developed statements to help practitioners and patients decide on appropriate healthcare for specific clinical conditions or circumstances.

The first RCT, reported in 1948, was designed to study the effects of streptomycin in patients with tuberculosis. Such trials have primarily been applied to analyzing the risks and benefits of a new medical treatment. The application of RCTs to surgical procedures has been much more limited. However, RCTs are subject to many limitations and deficiencies in general, and for surgical procedures in particular.

There are obvious differences between drugs and operations. A new drug does not change over time, is associated with a variable biologic response, and often requires large numbers of patients and considerable follow-up for assessment. A new drug may manifest unforeseen late side effects and toxicities far removed from the time of initial use. In contrast, none of these characteristics apply to most surgical procedures. A surgical intervention changes over time as technique and experience evolve and as refinements are made in patient selection and preoperative and postoperative management. With this evolution comes a change in risk over time. RCTs require a relatively long time frame and may not yield results for years. This may restrict the evolution of a new procedure or incorporation of new technology during the course of the trial.

RCTs in surgery, especially multicenter trials, are extremely expensive. Small trials cost hundreds of thousands of dollars, and larger ones may cost millions of dollars. The results are often not generalizable because strict inclusion and exclusion criteria, and a steep learning curve for some procedures, may make results of the trial not applicable to a wider range of patients or surgeons. The ease of blinding or double-blinding an RCT, routine for evaluating medical treatments, rarely can be duplicated for most surgical procedures. Thus, the provision of a placebo for the control group, which requires a sham operation for surgical trials, is a major problem. In the absence of a sham surgical procedure, the large placebo effect often associated with the surgical procedures may be misconstrued as confirming the value of the procedure. This is particularly true when the outcome measure relates to subjective improvement without predefined, objective measurable end points.

A major limitation of RCTs in surgery is the difficulty, if not impossibility, of standardizing the procedure being evaluated. There is surgeon to surgeon variation in terms of both surgical approach and technical ability and experience. The preoperative and postoperative care may vary from center to center. Poor-quality surgery or care represents failure to deliver the intended treatment, and the trial may then measure the deliverability and not the efficacy of the treatment. Evolution in technical modification, risk, and selection criteria is likely to occur in a course of a prolonged clinical trial. Surgical procedures typically progress via such modifications that individually are unlikely to produce detectable benefits but that collectively may do so.

One of the most difficult issues to address is the timing of an RCT for new surgical procedures. If the trial is done too soon during the evolution of the procedure, or before the trial surgeons have mastered the technique, emerging procedures that have ultimately proved successful may have been discarded because of poor or questionable early results.² If done too late, the new procedure may have gained widespread acceptance, to the point where an RCT between a surgical and medical therapy, or between 2 surgical procedures, may be impossible to conduct. On the surgeon’s side, there may be selection bias so that patients are diverted from entering the trial because one arm is perceived to be better than the other. On the patient’s side there may be refusal to accept the assigned arm if results from nonrandomized trials to date have seemingly confirmed the benefits of the one arm of the trial.

Other weaknesses of the RCTs include ethical limitations on the types of comparison that can be undertaken; the

From the Hospital of the University of Pennsylvania, Philadelphia, Pa.

Address for reprints: Joel D. Cooper, MD, Hospital of the University of Pennsylvania, Division of Thoracic Surgery, 3400 Spruce Street, 6 Silverstein, Philadelphia, PA 19104 (E-mail: joel.cooper@uphs.upenn.edu).

J Thorac Cardiovasc Surg 2010;140:743-6

0022-5223/\$36.00

Copyright © 2010 Published by Elsevier Inc. on behalf of The American Association for Thoracic Surgery

doi:10.1016/j.jtcvs.2010.06.037

central conflict between best practice for an individual and trial protocols; the problems of applicability to the general population; type 2 errors and other statistical design faults; publication bias; misuse and limitations of statistics; fraud; maintenance of blinding; asking the wrong question; and a simplistic, reductionist view of clinical management. The concentration on the randomized controlled trial devalues information from other sources, such as natural history studies, clinical experience, and case reports.³

Although the inability to apply RCTs to many surgical procedures may be regrettable, the misinformation derived from a flawed RCT may promote and perpetuate an inappropriate procedure for some time. Inappropriate conclusions may be due to flaws in design, inherent limitations in the trial, erroneous interpretation of results, or even chance occurrence, which dictates that 1 of the 20 “significant” study findings with a *P* value less than .05 between the 2 arms may be nothing more than random occurrence.

An oft-cited example of a flawed RCT is the Mayo lung cancer screening trial, which failed to demonstrate survival benefit from lung cancer in a randomized trial that assigned half of the patients to regularly scheduled chest x-rays and the other half to serve as controls. The study was underpowered because it could detect only a 50% reduction in mortality. Some 25% of patients assigned to the radiologic screening group did not comply, and 50% of the control group happened to have an annual chest x-ray unrelated to the trial. However, the results of this trial have helped perpetuate the notion that radiologic screening for detection of early lung cancer has no value for the cancer that causes more deaths in the United States than the next 3 most common cancers combined.

The opposite result, namely, the approval and application of a highly controversial procedure, transmyocardial laser revascularization for intractable angina in patients with unreconstructable coronary circulation, was based on the results of RCTs.^{4,5}

The procedure, done under general anesthesia with an open chest, involves multiple applications of a laser beam to create small channels in the myocardium. The control group was treated with medical management. The surgical group demonstrated significant reduction in angina but no demonstrable benefit in terms of measurable cardiac function, survival, or other relevant objective data that might be correlated with symptomatic improvement. It is illustrative to review the 2 reported randomized trials from the 1950s, comparing ligation of the internal thoracic artery with a sham surgical procedure in which the artery was exposed but not ligated. These trials were prompted by the recognition that medical trials for angina pectoris showed that many patients responded remarkably to placebo medication. After the introduction of internal thoracic ligation in Italy, the procedure was enthusiastically accepted by some American surgeons and even hailed in a *Reader's Digest* article.⁶

A proposed physiologic explanation for the benefit was that blood from the thoracic arteries proximal to the site of ligation would be diverted to the coronary circulation through collateral channels. In 1958 Dimond and coworkers⁷ presented an abstract reporting on 18 patients with exertional angina, with 13 undergoing the ligation procedure and 5 undergoing a sham surgical procedure in a blinded fashion. Marked improvement occurred in 10 of the ligated patients and all 5 of the sham patients. A significant increase in exercise tolerance occurred in 10 of the ligated patients and 5 of the sham patients. Overall, the use of nitroglycerin was diminished in both groups.

Presentation of this abstract was followed by a publication in *The New England Journal of Medicine* by Cobb and coworkers,⁸ in which 17 patients with severe effort-induced angina were randomized to internal thoracic artery ligation or a sham operation. Neither the patient nor the physician evaluating the patient's postoperative experience was aware of which procedure had been performed. Results showed that the number of nitroglycerin tablets used was reduced by 42% in those whose arteries were not ligated and by 34% in those whose arteries were ligated. Five patients in each group reported significant improvement after surgery. This landmark study is considered the first randomized surgical trial to show the powerful placebo effect of surgical procedures.

Randomized trials are poorly suited for evaluating surgical procedures, and this is one of the reasons why less than 7% of published surgical trials are randomized control trials, and of these, the quality of the trial is very low.^{1,9-11} Pride, independence, economic incentive, recognition, and academic advancement may play a role in diverting surgeons from RCTs to individual reports of case series. “Cynics might even claim that the personal attributes that go to make a successful surgeon differ from those needed for collaborative multi-center research.”¹²

ALTERNATIVES TO RANDOMIZED CLINICAL TRIALS

Undoubtedly, the evidence that has led to a change in practice or adoption of a new procedure has come primarily from observational data and reports of case series. This is particularly true if the natural history of a disease or condition is well established and a series of cases are operated on and produce a better clinical course. The effects of some surgical procedures may be large and unlikely to be the result of bias or other confounding factors. There is ample evidence that nonrandomized trials can produce results that correlate quite well with those of RCTs. Results from well-designed observational studies have been shown to correlate well with those from more expensive and impractical RCTs for a wide variety of interventions.^{8,13,14}

An example of a landmark case series was the demonstration of the value of the internal thoracic artery graft for coronary bypass procedures. This publication by Loop and

co-authors,¹⁵ including 10-year survival data, has had a major impact on current practice.

The value of a case series is particularly appropriate when all alternative therapy has been exhausted, the natural history of the condition is well documented and the impact of the surgical intervention is obvious, and the magnitude of the effect is measurable, significant, and expected.

RANDOMIZED TRIALS VERSUS CASE SERIES

Although RCTs may have a limited role to play in surgery, there are circumstances in which they may be absolutely essential to provide the evidence needed for the role and value of a surgical procedure.

When an operation is done to reduce the potential for a future adverse event, a randomized trial is required. An excellent example is the trial of endarterectomy for asymptomatic carotid artery stenosis. For patients with a 60% or greater carotid stenosis, patients were randomized to either carotid endarterectomy or aspirin and medical management. The 5-year stroke rate was 5.1% for endarterectomy and 11% for the control arm, thus confirming the value of the surgical procedure.

On the opposite side, arthroscopic surgery for the relief of symptoms due to chronic arthritis of the knee was demonstrated to be no better than a sham operation. This study illustrates the importance of a blinded, sham-controlled surgical procedure when the outcome to be measured is primarily subjective, that is, relief of symptoms.

When a new surgical procedure is proposed as an alternative to an existing procedure, and both are designed to produce the same result, an RCT should be encouraged. The blinded RCT of laparoscopic versus open cholecystectomy reported by Majeed and colleagues¹⁶ in 1996 demonstrated no reduced pain, hospital stay, or narcotic use in the laparoscopic group compared with the open group. Despite this, the tidal wave of minimally invasive surgical procedures, often based on the unproven presumption that small incisions are invariably and de facto better than a larger incision, is testimony to the many and complex factors that drive the new technology. Perhaps this is one of the best arguments for the need of more evidence-based guidance for surgical procedures where surgeons' eagerness to provide a new operation may seem more related to commercial concerns than to concerns for patients.

For oncology trials, where the anticipated benefit relates to long-term results, such as survival or time to recurrence, RCTs may be essential especially when comparisons of surgery alone versus surgery plus adjuvant chemotherapy or radiotherapy are to be made. The small but proven incremental improvement in outcomes over a period of time have formed the basis for significant progress in the field of surgical oncology. Indeed, surgeons more readily recruit patients for oncology trials than they do for trials of surgical technique.

Case series should be used whenever possible when a new surgical procedure is introduced for a condition having no alternative surgical option. Case series are also appropriate when the natural history of the underlying condition is well documented, the impact of the intervention is obvious, and objective data can be used to compare patients' preoperative and postoperative status. Case series are also ideal after animal studies demonstrating that the proposed procedure is feasible, has a physiologic rationale, and has demonstrable efficacy.

CONCLUSIONS

Surgery remains an art as well as a science. Evidence-based practice is the judicious use of available evidence to assist in decision making about the care of individual patients. It is not prescriptive or "cookbook" medicine and can never replace the surgeon's experience and judgment as to the value of the evidence and its relevance to the care of any individual patient. RCTs are only one such form of evidence, along with case reports, well-designed clinical studies, expert panel consensus, and other types of scientific evidence. The RCT has an important but limited role to play, but is not the tool we need to support guidelines for surgical practice. To invest the RCT with dogma that it is the only valid method for confirming the value of new procedures is to disguise it with the emperor's new clothes. This, however, does not free us up to abandon the need to obtain scientific evidence to support our decisions whenever possible. Rather, we must seek to find alternative, practical, and acceptable alternatives if we are to retain our academic credibility. We must work with epidemiologists and statisticians to design appropriate trials, recognize that many clinical trials are badly flawed, resist the tyranny of the RCT, and recognize that at best the conclusions are only guidelines and at worst the result of a flawed trial rather than of a flawed intervention.

How then should new surgical procedures be introduced into general practice?

It used to be that new surgical procedures, often touted as the latest "cure," soon proved of little value and were discarded by a process of natural selection. One of my mentors, Ronald Belsey, used to note that "the follow-up clinics are the shoals upon which founder many attractive theories in surgery." Thus, procedures designed to improve subjective symptoms in patients with chronic obstructive pulmonary disorder, such as glomectomy, radical hilar stripping, costochondrectomy, phrenicectomy, pleural abrasion, and others, were soon discarded into the wastebasket of history when objective evidence was lacking and clinical experience failed to confirm their value. I am concerned that today the life span of unproven technologies and procedures may be artificially prolonged by factors, including results from flawed trials, competition for patients, economic consequences, marketing and publicity often driven by industry

sponsors, and confusing and often misleading information on the Internet.

Now more than ever, I believe that we require methods to study and validate the role of new and existing surgical interventions. An RCT is by and large not a suitable instrument with which to achieve this goal. As noted by Lantos,¹⁷ from The University of Chicago, "...our statisticians encourage clinical investigators to design trials that make their (statisticians) lives a little too easy by making our lives more difficult. Rather than constraining human choices to meet the needs of statisticians, I think we should encourage the statisticians to devise methods of evaluating the source of choices that free and intelligent citizens make all the time."

I propose the following for the introduction of new surgical procedures:

1. Health care providers should restrict the application of promising new procedures to a limited number of centers of excellence with appropriate resources and experience.
2. Centers should be required to document and report at regular intervals specified information regarding morbidity, mortality, and objective measures of income.
3. These data should be periodically reviewed and evaluated by an independent scientific panel.

By these means, new procedures can be offered to appropriate patients, insurers and patients can be protected against abuse, and the necessary data can be acquired for objective analysis before widespread promulgation of the procedure can occur. Then too, we as surgeons can take increased pride in the remarkable accomplishments that our profession has achieved.

References

1. Treasure T. A brief history of evidence. In: Treasure T, Hunt I, Keogh B, Pagano D, eds. *The Evidence for Cardiothoracic Surgery*. Shrewsbury: tfm Publishing Limited; 2005:5.
2. Berger RL, Celli BR, Meneghetti AL, Bagley PH, Wright CD, Ingenito EP, et al. Limitations of randomized clinical trials for evaluating emerging operations: the case of lung volume reduction surgery. *Ann Thorac Surg*. 2001;72:649-57.
3. Ellis SJ, Adams RF. The cult of the double-blind placebo-controlled trial. *Br J Clin Pract*. 1997;51:36-9.
4. Schofield PM, Sharples LD, Caine N, Burns S, Tait S, Wistow T, et al. Transmyocardial revascularization in patients with refractory angina: a randomized controlled trial. *Lancet*. 1999;353:519-24.
5. Frazier OH, March RJ, Horvath KA. Transmyocardial revascularization with a carbon dioxide laser in patients with end-stage coronary artery disease. *N Engl J Med*. 1999;341:1021-8.
6. Ratcliff JD. New surgery for ailing hearts. *Reader's Digest*. 1957;71:70-3.
7. Dimond EG, Kittle CF, Crockett JE. Evaluation of internal mammary artery ligation and sham procedure in angina pectoris. *Circulation*. 1958;18:712-3.
8. Cobb LA, Thomas GI, Dillard DH, Merendino KA, Bruce RA. An evaluation of internal-mammary-artery ligation by a double-blind technic. *N Engl J Med*. 1959;260:1115-8.
9. McLeod R. Randomized, controlled trials: is there a role for them in surgery? [Editorial]. *Ann Surg*. 2006;244:684-5.
10. Cook RC, Alscher KT, Hsiang YN. A debate on the value and necessity of clinical trials in surgery. *Am J Surg*. 2003;185:305-10.
11. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ*. 2002;324:1448-51.
12. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet*. 1996;347:984-5.
13. Baum M. Reflections on randomized controlled trials in surgery. *Lancet*. 1999;353(Suppl 1):S16-8.
14. Sanchez PG, Kucharczuk JC, Su S, Kaiser LR, Cooper JD. NETT REDUX—accentuating the positive. *J Thorac Cardiovasc Surg*. 2010;140:564-72.
15. Loop FD, Lytle BW, Cosgrove DM, Stewart RW, Goormastic M, Williams GW, et al. Influence of the internal-mammary-artery graft on 10-year survival and other cardiac events. *N Engl J Med*. 1986;314:1-6.
16. Majeed AW, Troy G, Nicholl JP, Smythe A, Reed MW, Stoddard CJ, et al. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. *Lancet*. 1996;347:989-94.
17. Lantos J. Ethics, randomization, and technology assessment. *Cancer*. 1994;74(9 Suppl):2653-6.