



RANDOMIZED TRIALS OF NEW PROCEDURES HAVE UNIQUE PROBLEMS AND PITFALLS

LAWRENCE I. BONCHEK, M.D., F.A.C.S., F.A.C.C.
Editor-in-Chief

The use of randomized trials was introduced 60 years ago to assess the effectiveness of streptomycin for the treatment of tuberculosis.¹ Though neither double-blinded nor placebo controlled, that trial introduced a new era in the assessment of therapeutic innovations. Today, new drugs cannot gain clinical acceptance or regulatory approval without randomized trials. Within the medical community, an insistence on the stringent criterion of randomized studies has been extended *de facto* to include new operations, procedures, and techniques, all of which must now be validated by such studies before they can gain wide acceptance. Though the FDA does not (yet) license operations per se, it does approve devices, which affects many new procedures that employ them.

When coronary bypass surgery was first introduced, Hiatt suggested that “randomized trials should precede widespread dissemination, as is done to a considerable extent for drugs.”² Spodick went further, vehemently suggesting that such studies be performed “from the very first clinical trial” of all new operations” and should precede widespread use not only of surgical procedures but of “costly procedures of all kinds.”³

The subject requires less heated debate. Such opinions reveal a failure to understand the many differences between randomized trials of new drugs and trials of new procedures. Randomized trials that compare the effectiveness of drugs or combinations of drugs are comparing apples with apples. Those that compare medical treatment with surgical or procedural therapy are comparing apples with oranges and are not only difficult to interpret, but are subject to unavoidable misinterpretation. Although I first wrote about this topic many years ago,⁴ non-surgical interventions are so common that this discussion is more pertinent than ever.

DRUGS VERSUS PROCEDURES

Table 1 lists several differences between drugs and procedures. Drugs have an unchanging composition, and as drug usage increases, additional side effects and complications become apparent. When used in a trial, a

drug’s effectiveness is unrelated to the physician’s skill. Not only are results generally consistent among collaborating institutions, but they are also applicable to nonparticipating institutions. Finally, a placebo is usually available, and crossovers between treatment groups are exceptional.

In contrast, new procedures are introduced while they are imperfect. The indications are uncertain, and the risks are high. As the procedure becomes more widespread, refinements occur and the risks decline, often dramatically. In early trials, results vary considerably in the hands of different operators. Placebos are obviously not available for any invasive procedure, and crossovers from medical to procedural treatment groups are commonplace.

BIAS IN RANDOMIZED STUDIES

It is commonly assumed that randomization eliminates bias. Though this is mostly true for trials that compare different drug regimens, when new procedures are compared with medical therapy, some types of bias cannot be avoided.

First is the possibility of *pre-selection bias* which can occur before patients are referred to the study. When two different drug protocols or two surgical protocols are being compared with each other, referring physicians rarely have reason to prefer one to the other and many are willing to accept randomization of their patients. In contrast, when a new procedure (think of coronary bypass in its early years) is being evaluated, physicians are more likely to either want it or to distrust it for their high risk patients, and they refer such patients for the therapy they prefer. The result is a diversion of such patients away from randomization, while the lower risk patients, about whom physicians feel less strongly, are randomized.

It is commonly observed that patients in both limbs of randomized trials fare better than historical controls. This result is generally attributed to both the stringent selection criteria that define the cohort eligible for the

TABLE I

	Drug		Procedure
1.	Unchanging compound	1.	Evolves Continuously
2.	Complications increase with use	2.	Complications decrease with use
3.	Results unrelated to physician skill	3.	Results vary with operator
4.	Placebo usually available	4.	No placebo
5.	Crossovers rare	5.	Crossovers common

trial, and the more attentive, more standardized, and thus more effective therapy that trial patients receive. Undiscussed is the important influence of referral bias, which yields a low risk cohort. This issue merits concern, because in low risk patients, it is difficult to prove that a new therapy is more effective.

An example of this phenomenon occurred during early efforts to assess the efficacy of coronary artery bypass (CABG). In the 1970s a randomized trial of CABG for stable angina was initiated at the Oregon Health Sciences University. I had just joined the faculty after completing my cardiac surgery training there. It was striking to observe the relentless diversion of referrals for CABG away from the University and to private surgical teams who readily accommodated the referring physicians' growing preference for the new option of surgical therapy. The trial terminated with much smaller patient subgroups than originally planned, and with statistically inconclusive results.⁵

A second cause of unintended bias is the fact that even participating investigators have human instincts and subtle biases that can influence the selection or exclusion of patients for randomization. In the U.S. Veterans Administration study of coronary bypass surgery, patients could be excluded because of undefined and subjective criteria such as "poor left ventricular function." Such vague definitions led to a tendency at many participating hospitals to include only the most stable and willing patients.⁶

A third source of bias is non-adherence to the assigned therapy, or "crossing over." In drug studies, crossovers are usually infrequent, they can occur in both directions, and they have little effect on the statistical analyses. In contrast, when an operation or procedure is being studied, crossovers can occur only in one direction – from medical to procedural therapy. If they occur frequently and freely,

the medical group eventually comes to be composed only of patients in whom medical therapy is successful. Even the most elaborate statistical manipulations cannot overcome some of the confounding statistical effects of large numbers of crossovers. This was a particular issue in early studies that compared CABG with medical therapy, since patients with persistent symptoms opted overwhelmingly for bypass surgery, and were no longer "available" to experience complications such as death or myocardial infarction on late follow-up of the medical group.

This latter criticism has usually been met with the argument that patients who do well after crossing over to surgery after a period of medical therapy demonstrate that there is no harm in reserving surgical therapy for patients who fail a trial of initial medical therapy. This conclusion is valid, but the favorable experience with such crossovers must not then be used to compare the results of initial surgical with initial medical therapy. Unfortunately, it is unusual for any later citation of the results of such trials to mention the incidence and problem of crossovers.

WHEN SHOULD RANDOMIZED TRIALS OF PROCEDURES BE DONE?

Let's get back to Spodick's suggestion mentioned earlier that randomized studies should be performed from "the very first clinical trial...of...costly procedures of all kinds."³ Accompanying this well-intentioned if somewhat naive proposal was the more fevered suggestion that editors of scientific journals should reject reports of non-randomized trials! Both of these ideas fail to acknowledge that the rapid evolution of surgical and other procedural techniques usually invalidates "early" trials. The U.S. Veterans' Administration study of coronary bypass surgery began in the early 1970's, and when results were finally reported it was intensely criticized for high mortality and poor graft patency rates in comparison with contemporary outcomes at that time.⁵ Progress was so dramatic and the decline in morbidity

and mortality was so precipitous that initial entries (1970–1972) were actually discarded. Ultimately, the V.A. study had little impact on the practice of coronary surgery as it was outdated on arrival. In a more recent example, randomized trials of intracoronary stents had no sooner been reported than the results were superseded by dramatic changes in anticoagulant therapy.^{7,8} The story is being repeated yet again with drug-eluting stents, for which antiplatelet regimens continue to evolve, and clinical studies of their results lose impact soon after publication.

Another disadvantage of early trials of operations and procedures is their tendency to stifle technical progress, which comes from many investigators. By their nature, trials are confined to a restricted list of institutions, which limits innovation by others.

On the other hand, if a randomized trial is delayed until a procedure is “mature”, and good results are the norm, it will be difficult or impossible to randomize a broad spectrum of patients. In the Coronary Artery Surgery Study (CASS) begun in the United States after coronary bypass was well established, some institutions did not randomize any patients and participated only in a registry.⁹

The incentive to participate in early randomized trials is provided by grants that are often vital to the participating institutions. Even when a proposed trial is premature, a decision not to participate in a multi-institutional study, or – perish the thought – to abandon an ongoing trial that is becoming obsolete and pointless, is a form of bureaucratic and academic self-mutilation. Large collaborative studies generate their own central bureaucracy and peripheral constituency, with many clinicians, statisticians, data coordinators, etc. being supported by grant funds. When the U.S. Veterans’ Administration proposed a randomized study of the Bjork-Shiley mechanical heart valve versus the Hancock porcine valve in 1976, I declined to participate. (I was then in Milwaukee, where the Zablocki Veterans Administration Hospital was part of the academic teaching program at the Medical College of Wisconsin.) The study was doomed to premature obsolescence and irrelevance by the predictably rapid development of prosthetic heart valves. Patients were enrolled between 1977 and 1982, and the study plodded stubbornly onward for many years at great public expense until late results were reported in 1993. By then, no

models of the Bjork-Shiley valve were even sold in the U.S., and the standard Hancock model initially used in the study had long since been supplanted by hydraulically superior models.¹⁰

CONCLUSIONS

Randomized trials that compare new procedures with established ones should guard against pre-randomization bias and must allocate patients to treatment groups based on objective or quantitative criteria, not on subjective clinical judgment. Risk, length of followup, and sample size must be used to calculate the statistical power of the study, so that a significant difference between treatments does not remain undetected (a Type II error). There should already be sufficient experience with the new procedure so that complication rates have stabilized, and so that participating operators are equally comfortable with all procedures being studied.

Even with the above stipulations, however, randomized trials that compare medical with procedural therapy pose many additional problems, and few such studies have had a major impact on clinical practice. (I have discussed these problems and examples at greater length elsewhere.¹¹) The most useful randomized studies of procedures are those that compare one procedure with another, or those that assess a specific refinement in an established procedure or with an established device, such as the use of different anticoagulation regimens for coronary stents.

Finally, randomized trials look at defined (and usually low risk) cohorts that rarely reflect the full spectrum of patients seen in clinical practice, and the results of such trials usually cannot be generalized. Non-randomized retrospective studies and prospective registries can provide clinically useful information that is complementary. The introduction of propensity scoring to derive comparable groups from large registries, and meta-analysis of multiple clinical series, can substitute for the slavish use of randomized studies even when they are unlikely to be helpful. Though sophisticated statistical manipulation is needed to compare outcomes in registry patients, the results can be a useful guide to managing the full spectrum of patients in a typical clinician’s practice. Large registries with a much broader spectrum of patients from the general population can therefore be vital companions to randomized clinical trials.

REFERENCES

1. MRC (Medical Research Council) Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948;2:769-783.
2. Hiatt HH: Lessons of the coronary-bypass debate. *N Engl J Med.* 1977;297:1462.
3. Spodick DH, Aronow W, Barber B, et al. Standards for surgical trials. *Ann Thor Surg.* 1979;27:284.
4. Bonchek LI: Are Randomized Trials Appropriate for Evaluating New Operations? *N Engl J Med.* 1979;301:44.
5. Kloster FE, Kremkau EL, Ritzman LW, et al. Coronary bypass for stable angina: a prospective randomized study. *N Engl J Med.* 1979; 300:149-157.
6. Lawrie GM, Morris GC Jr, Howell JE, et al. A debate on coronary bypass. *N Engl J Med.* 1977;297:1464-1470.
7. Serruys PW, DeJaegere P, et al. A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. *N Engl J Med.* 1994;331:489-495.
8. Fischman DL, Leon MB, et al. A randomized comparison of coronary-stent placement and balloon angioplasty in the treatment of coronary artery disease. *N Engl J Med.* 1994;331:496-501.
9. CASS Principal Investigators and their Associates. Myocardial infarction and mortality in the coronary artery surgery study (CASS) randomized trial. *N Engl J Med.* 1984;310:750-758.
10. Hammermeister KE, Sethi GK, Henderson WG, et al. A comparison of outcomes eleven years after heart-valve replacement with a mechanical valve or prosthesis. *N Engl J Med.* 1993;328:1289-1296.
11. Bonchek LI. The role of the randomized clinical trial in the evaluation of new operations. *Surgical Clinics of North America* 1982; 62:761-769.



*Rufous-throated Hummingbird (sp. Selasphorus rufus); Camelback Mountain, Scottsdale, AZ
David L. Newcomer, M.D.*