

General Principles for Evaluation of New Interventional Technologies and Devices

David Sacks, MD, Committee Chairman, David L. Marinelli, MD, Louis G. Martin, MD, James B. Spies, MD, and the Members of the Society of Interventional Radiology Technology Assessment Committee¹

J Vasc Interv Radiol 2003; 14:S391-S394

INTERVENTIONAL therapies and devices have become increasingly accepted in recent years. It is the purpose of the Technology Assessment Committee of the Society of Interventional Radiology (SIR) to propose guidelines for the scientific evaluation of new technology. This article is a statement of general principles of research and is intended to be the background for subsequent standards produced by the Technology Assessment Committee, such as the reporting standards for clinical evaluation of new peripheral arterial revascularization devices that follow this article. While these principles are necessary to evaluate new technologies and devices, they also represent an optimal way to evaluate current technologies. The goal of these guidelines is to create reliable evidence of device and technique effectiveness (1-5).

GENERAL PRINCIPLES

There are several important considerations in the evaluation of a new interventional device. Fundamentally, the purpose of the study should be to

answer a well-defined question (hypothesis) about the device. This requires appropriate study design, including appropriate selection of study subjects and controls; clear and consistent definitions used in describing the results of the intervention; interpretation of the results including the statistical treatment of the results; and control of bias (6,7). Only then can one develop valid conclusions regarding the efficacy of the device.

Validity refers to the degree to which the results reflect the true direction and magnitude of the treatment effect (8). There are two types of validity: internal validity, which refers to the results of the particular trial reflecting the true effect on that particular group of patients, and external validity, which refers to the applicability of the results to the general population. The primary measures of internal and external validity are discussed in the following sections.

Study Design

A clinical trial is a prospective comparative study of a group of subjects with controls, with the two groups treated identically except for the intervention.

Often the most difficult task in the evaluation of a new interventional device is the design of a comparative trial. Most commonly, a new device is initially evaluated without controls to determine if the device is safe and potentially effective. While ideally this would be a small pilot study that would be followed by a comparative trial, often the pilot study is reported and no comparative trial is ever done.

An attempt may be made to draw conclusions based on comparison of the results with previously completed studies of the standard therapy (historical controls). However, without the direct comparison with contemporaneous controls, too many uncontrolled variables and potential biases are introduced to make valid conclusions possible. The use of historical controls also allows bias in the selection of the comparison study. For example, laser therapy of peripheral arterial disease was portrayed in a favorable light by comparison to selected historical studies that had poor results from balloon angioplasty (9). Usually, there are a number of previous studies that have varying results because each uses different definitions of success and failure, different patient selection criteria, operators with different experience or equipment, and a host of other variables. Depending on the circumstances, a comparison may be made that makes the new technique appear either more or less favorable than the standard method.

Some of the advantages of a randomized controlled trial over a study using historical controls can be obtained with a nonrandomized contemporaneous trial. For example, a contemporaneous trial would control for advances in medical care that may have occurred which might enhance the results of a new therapy over standard therapy as compared to historical controls. In addition, the same selection and outcome criteria can be applied to both the study and the control groups. This eliminates the incompatibility of the definitions of success,

This article first appeared in J Vasc Interv Radiol 1997; 8:133-136.

Address correspondence to SIR, 10201 Lee Hwy, Suite 500, Fairfax, VA 22030.

¹ Curtis W. Bakal, MD, Gary J. Becker, MD, Dana R. Burke, MD, Patricia E. Cole, MD, Michael D. Dake, MD, Richard J. Gray, MD, Margaret E. Hansen, MD, Ziv J. Haskal, MD, Robert W. Holden, MD, Michael D. Katz, MD, Lindsay S. Machan, MD, Nilesh H. Patel, MD, and Richard Shlansky-Goldberg, MD.

© SIR, 2003

DOI: 10.1097/01.RVI.0000094614.61428.67

failure, complication, and so forth. However, even in a well-controlled prospective nonrandomized trial, there may be subtle selection and evaluation biases. For example, when choosing the new therapy over the standard one for a particular patient, a new device could be used in patients with slightly less advanced disease or slightly more favorable prognostic factors, which might alter the results even when stratification is used. Lack of comparability of patient populations can only be controlled by randomizing the subjects (8,10) and blinding the researcher to the allocation sequence (11). The use of randomization allows for valid statistical treatment of variation between the controls and subjects, and that analysis serves as a check on the homogeneity of the two groups and, therefore, on the "quality" of the randomization.

There are reservations about the use of randomized studies for evaluation of new technologies. Investigators often are concerned that the new therapy is so far superior that it would be unethical to randomize patients to the alternative therapy. In addition, because randomized trials usually require a greater number of patients, contemporaneous controls, and careful follow-up, they are more time consuming and expensive than alternative study designs. In a field in which technology changes quickly, a rapidly completed study more accurately reflects the state of the art of current practice. Some useful devices might not be evaluated if a more expensive study were needed (12,13). Unfortunately, the history of medical innovation reveals that many therapies that once were considered far superior and were accepted into practice were subsequently shown to provide no improvement when scrutinized in randomized clinical trials (14). For example, femoro-popliteal atherectomy was portrayed as a significant advance over balloon angioplasty in several large case series (15). A subsequent randomized trial found a higher restenosis rate and no clinical advantage with atherectomy (16,17). Regarding the ethical concerns, all trials should be monitored on an ongoing basis. If a very large treatment effect (or failure) becomes evident during the ongoing analysis of the data, the trial can be stopped. Also a random-

ized protocol can be formulated in which patients are randomized to the new therapy in a 2:1 ratio rather than 1:1. This reduces the power of the study to detect a difference between treatment groups by only a small amount, equivalent to having 9% fewer patients in the trial, and allows one to treat the majority of patients with the expected better treatment without markedly jeopardizing the rigor of the study (10).

The use of case series for the evaluation of new therapies is discouraged. Although it is possible that very large therapeutic effects can be predicted in some case series without controls, it is usually impossible to estimate the true magnitude of the improvement. Most therapeutic advances are not that dramatic, and there is considerable potential for understating or overstating the true effect. Valid information can be obtained from case series with regard to the potential for complications or technical issues related to the feasibility of the treatment, but the potential improvement of the new method over the old cannot be accurately estimated. Therefore, if a clinical series is presented, the conclusions drawn must be very limited and the procedure should not be recommended as standard therapy without a comparative trial.

Another alternative to a randomized trial is to use data acquired from a registry. The registry specifies the type of information that will be collected and the definitions of success, failure, and complications that will be used. A registry could be used as a source of contemporaneous controls for a clinical trial if critical study variables are comparable (6). Because monitoring the generation of the data is difficult and the patient outcomes may not be accurately reported, bias can be introduced and this practice is not recommended.

Registries are most useful in identifying the indications and techniques that could be evaluated in subsequent randomized trials, evaluating technical aspects of the procedure, and documenting complications. They also give an indication of the effectiveness of an intervention across a variety of practice types. However, because participation in registries is voluntary and the accuracy of the data entered may

vary, their findings must be interpreted with caution (18).

The choice of the control procedure is also an important part of the study design. For certain groups of patients, such as those with intermittent claudication from peripheral vascular disease, the control procedure could range from standard angioplasty, surgical bypass, conservative therapy with exercise, to no therapy. The choice is possible because of the relatively benign course of peripheral vascular disease (19). For patients with severe peripheral vascular disease or other potentially lethal conditions, comparison with conservative management may not be feasible or ethical. The control procedure should be a standard, currently accepted method of management. This will allow the best evaluation of the potential role of the new therapy in the scheme of current care.

Study Population and Patient Selection

The subjects of any trial are drawn from a patient population. The population might be all the patients in a given hospital, all the members of a health maintenance organization, or all patients in a region. A description of this population should be included in the published report so that reviewers may judge how the study subjects were then selected and how applicable the study is to the general population. This aids in the assessment of the general applicability or external validity of the study.

The selection criteria for choosing the study subjects and controls from the patient population should be explicitly stated in the protocol. Exclusion criteria should also be described. This again allows the reader to make an estimate of the applicability of the results to the general population and also allows one to judge if systematic bias has occurred in the design of the study that might favor either the intervention or the control procedure. Each patient who meets the selection criteria and who agrees to participate in the trial should then be randomized to the treatment or control group. Randomization should be done by means of either a random number table or a computer-generated random assignment to avoid bias (11). Informed con-

sent must be obtained from all participants.

When randomizing patients, it is important that each subject be included in the analysis of the group to which they were originally randomized, regardless of whether they received that therapy or any therapy. This is called the principle of "intention to treat." This is important because the two therapies that are being studied may have characteristics that prevent some patients from receiving the therapy or cause them to decline the therapy. These may include cost, potential complications, or length of hospital stay or recovery time. For example, a randomized trial of surgery versus angioplasty for treatment of claudication may give misleading results if a patient who develops chest pain prior to the procedure does not undergo surgery but a similar patient randomized to angioplasty is treated under the assumption that angioplasty is a less stressful procedure. Unless patient data are analyzed based on the intention to treat, a higher periprocedural myocardial infarct rate could be falsely attributed to the angioplasty procedure rather than to the fact that high cardiac risk patients had been withdrawn from the surgical group. By including all patients in the analysis, a more accurate estimate of the true treatment effect is possible and this improves both the internal and external validity of the study.

Data Collection and Statistical Analysis

The type of data and how it is to be collected should be decided before the study begins. This will ensure that the data are collected in a uniform manner throughout the study. Important definitions related to disease severity and extent, success and failure, and type of follow-up should be decided ahead of time and included in the protocol. Suggested definitions are given in the sections that follow. These efforts will ensure that bias in the evaluation of the outcomes will be limited and the objective evaluation of the intervention enhanced.

Each participant's pertinent demographic and clinical information should be obtained. These characteristics can be used to test the randomization of the group. For example, the

age, sex, presence of diabetes, chronic renal failure, or other co-morbidities of one group can be compared with the other with a χ^2 test or an unpaired *t* test, giving an estimate of the homogeneity of the two groups. If both the treatment and control groups are very similar, this suggests that there has been successful randomization. However, it may be that certain subgroups of the study population are suspected of having considerably different responses to the proposed therapies. This may be handled in one of two ways. Those selected for the study may be prospectively stratified into groups by age or specific co-morbidity. After stratification, the subgroups can be randomized. Alternatively, the subgroups can be retrospectively analyzed after the study is completed. If a specific subgroup is very important in the evaluation of the device, prospective stratification ensures that the desired proportion of this subgroup will be in each study group. If the analysis is done without stratification, the proportion of the subgroup in each group is left to chance. Therefore, retrospective stratification is a much less desirable strategy.

The validity of the conclusions regarding the outcome of an intervention over time is directly related to the quality of the follow-up. The greater the number of subjects lost, the greater the chance of bias entering the study since patients lost to follow-up often have a different prognosis from those not lost. The follow-up rate must be greater than 80% (6).

The statistical tests to be applied to the data should also be decided when writing the protocol, prior to the initiation of the trial. The statistical test to be applied will be dictated by the type of data collected. For example, categorical data such as success versus failure are analyzed with a nonparametric test such as χ^2 . Continuous data such as residual pressure gradient are analyzed with a parametric test such as the *t* test or analysis of variance. An estimation of the sample size necessary to detect the expected difference in the two interventions is also needed before starting the trial. This may be done by estimating the degree to which the new intervention is an improvement over the comparison therapy. This is the expected effect size. With this and selection of tolerable

levels of type I and type II errors, the necessary sample size can be calculated. The investigators should enroll more than the number of subjects calculated to compensate for any that may be lost to follow-up. This exercise is important in determining the power of the trial that is being performed. Too small a study may lack the power to detect a real difference between two therapies and result in the conclusion that there is no difference when one exists (type II error). Consultation with a biostatistician or other professional skilled in study design prior to commencing the trial will prevent design errors or misapplication of statistical tests that could fatally flaw the results of the trial.

Use of a checklist of essential elements of a randomized trial will improve the quality of the study and should be referred to during study design and reporting (20,21).

Written Protocol

A protocol should be written prior to the commencement of the study detailing the essential elements of the study design and the intended conduct of the trial. This should include a copy of the consent form that will be used. The protocol should be submitted to the institutional review board of the facilities involved and approved prior to the commencement of the study.

References

1. Oxman AD, Sackett DL, Guyatt GH. Users' guide to the medical literature. I. How to get started. *JAMA* 1993; 270:2093-2095.
2. 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; 270:2598-2602.
3. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1994; 271:59-63.
4. Woolf SH. Practice guidelines, a new reality in medicine. II. Methods of developing guidelines. *Arch Intern Med* 1992; 152:946-952.
5. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of anti-thrombotic agents. *Chest* 1992; 102 (suppl):305S-311S.

6. Interventional Cardiology Devices Branch, Division of Cardiovascular, Respiratory and Neurology Devices, Office of Device Evaluation of the Food and Drug Administration. Guidance for the submission of research and marketing applications for interventional cardiology devices. May 1993.
7. Begg CB, McNeil BJ. Assessment of radiologic tests: Control of bias and other design considerations. *Radiology* 1988; 167:565-569.
8. Guyatt G, Drummond M, Feeny D, et al. Guidelines for the clinical and economic evaluation of health care technologies. *Soc Sci Med* 1986; 22:393-408.
9. McLean GK, Burke DR, Marinelli DL. Comment on the clinical appropriateness of an emerging technology. *Radiology* 1989; 172:941-942.
10. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976; 34: 585-612.
11. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-412.
12. Grunkemeier GL, Starr A. Alternatives to randomization in surgical studies. *J Heart Valve Dis* 1992; 1:142-151.
13. Bonchek LI. Are randomized trials appropriate for evaluating new operations? *N Engl J Med* 1979; 301:44-45.
14. Grimes DA. Technology follies: the uncritical acceptance of medical innovation. *JAMA* 1993; 269:3030-3033.
15. McLean GK. Percutaneous peripheral atherectomy. *J Vasc Interv Radiol* 1993; 4:465-480.
16. Vroegindeweij D, Tielbeek AV, Buth J, Schol FPG, Hop WCJ, Landman GHM. Directional atherectomy versus balloon angioplasty in segmental femoropopliteal artery disease: two-year follow-up with color-flow duplex scanning. *J Vasc Surg* 1995; 21:255-269.
17. Tielbeek AV, Vroegindeweij D, Buth J, Landman GHM. Comparison of balloon angioplasty and Simpson atherectomy for lesions in the femoropopliteal artery: angiographic and clinical results of a prospective randomized trial. *J Vasc Interv Radiol* 1996; 7:837-844.
18. Baim DS, Dent KM, King SB, et al. Evaluating new devices: acute (in-hospital) results from the new approaches to coronary intervention registry. *Circulation* 1994; 89:471-481.
19. Hertzner NR. The natural history of peripheral vascular disease: implications for its management. *Circulation* 1991; 83(suppl):I12-I19.
20. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272:1926-1931.
21. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting in randomized trials: the CONSORT statement. *JAMA* 1996; 276:637-639.