

Guest Editorial

A Primer on Randomized Controlled Trials

Donald S. Likosky, PhD

Departments of Surgery and Community and Family Medicine, Dartmouth Medical School, Hanover, New Hampshire

Abstract: Randomized Clinical Trials are held as the gold standard for quantifying the effect of an intervention across two or more groups. In such a trial an intervention is randomly allocated to one of two groups. The benefit of such a trial lies in its ability to establish nearly comparable groups of subjects in all manner except for the effect of the intervention. As such, the effect of a given intervention may be attributed solely to the intervention and not to any other extraneous factor. In the following editorial,

we will discuss several issues that are important for understanding how to conduct and interpret randomized trials: choosing the study population, choosing the comparison group, choosing your outcome, study design, data analysis, and issues of inference. This editorial is intended to make the reader an educated consumer of such trial designs. **Keywords:** randomized clinical trial, methodology, study design. *JECT. 2006;38:10-13*

Not all evidence is created equal. This is a succinct yet important notion. In reading the peer-reviewed literature, the validity of this statement becomes self-evident. The gold standard for proving or disproving the value of a given medical intervention is the randomized controlled trial. Fortunately or unfortunately, this study design may not always be the most appropriate methodology for testing a given research question. In future issues of this journal, I will describe different study designs, exploring how to conduct them, how to draw inferences based on data rather than supposition, and the inherent limitations related to each. It is my hope not to describe each design in infinite detail or to transform you into an expert, but simply provide you with sufficient information to make you more erudite and discerning in your reading of the medical literature.

RANDOMIZED TRIALS—THE EARLY YEARS

On April 21, 1601, Captain James Lancaster departed on a voyage with four ships. Unintentionally, he brought with him bottles of lemon juice on one of these ships and asked that each man drink three spoonfuls of the juice every morning. By August 1 of that year, scurvy appeared rampant on the other three ships. By the time of his arrival



Donald S. Likosky, PhD

on September 9, the sailors on Lancaster's ship were required to assist the men on the three other ships in the docking of their vessels to shore. Lancaster inadvertently became the principal investigator on one of the first randomized trials, because the use of the intervention (lemon juice) was randomly allocated to the ships.

WHY DO WE RANDOMIZE?

The major concern for most research is the issue of confounding. Confounding is the distortion in an observation (surgery is associated with increased mortality) that is brought about by a third factor (age). This third factor,

Address correspondence to: Donald S. Likosky, PhD, Department of Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756. E-mail: donald.likosky@dartmouth.edu

called a confounder, by definition must be associated independently with the exposure (surgery) and the outcome (mortality). The issue of confounding is an important one, because it prevents true associations from revealing themselves. Confounding is dealt with in a number of ways, both in the design and analysis of a study.

Researchers may remove issues of confounding through randomization (Figure 1). In a randomized trial, subjects are randomly allocated to a study group (those undergoing coronary revascularization with conventional coronary artery bypass grafting) and a comparison group (those undergoing coronary revascularization without cardiopulmonary bypass). Allocation is based through chance alone rather than based on patient or surgical preference. In this case, we would imagine that all known (sex, age) as well as unknown (age of patient's parents) variables would be equally distributed to both the study and comparison groups. As such, the issue of confounding has been removed through the design (random allocation of subjects to study and comparison groups) of the study. Successful randomization may be defined as that which creates two or more groups that are similar in every way except for their group allocation. For this reason, randomized controlled trials are thought of as the gold standard for studying relationships between exposure and outcomes, because it removes potentially confounding factors. As such, any differences in observed rates (such as mortality) are thus attributed to differences in the exposure (otherwise defined as an intervention).

Analytic approaches to addressing confounding will be discussed in future issues when we deal with other study designs.

There are several issues that are important for understanding how to conduct and interpret randomized trials: choosing the study population, choosing the comparison group, choosing your outcome, study design, data analysis, and issues of inference.

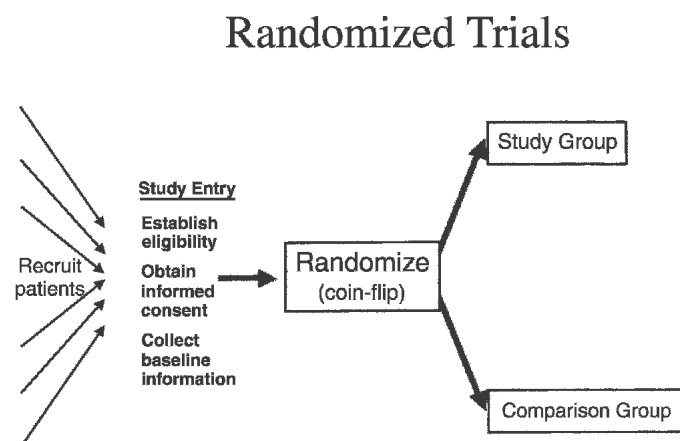


Figure 1. Design of a randomized trial. Acknowledgment: H. Gilbert Welch, MD, MPH.

Choosing the Study Population

Inferences gleaned from randomized trials are dependent, in part, on the study population, namely those receiving the intervention of interest. When choosing your study population, there are really at least two issues to consider.

To Whom Do You Want the Results to Apply: To isolate the association between an exposure and an outcome, researchers often place restrictive study inclusion (to include only adult patients) and exclusion (to exclude patients undergoing emergency surgery) criteria. One advantage of this approach is to create a homogenous study population. Unfortunately, as we increase the number of these criteria, our ability to generalize our findings to other patient populations diminishes. Generalizability may be defined as the extent to which findings obtained from a given study may be applied to the target population. For instance, a study of endovascular vs. open abdominal aortic aneurysm (AAA) repair had the following criteria—entry criteria: maximum external diameter in any plane greater than or equal to 5 cm and AAA greater than or equal to 4.5 cm and the AAA has increased by greater than or equal to 1 cm in diameter in 12 months; exclusion criteria: patient has had a previous AAA repair procedure, AAA is not elective, there is a likelihood of poor compliance to the protocol.

You can imagine that as the number of these criteria increase, the study population becomes less representative of patients seen at a typical physician's office or medical center. In the above example, these findings would not pertain to medical centers that have a majority of urgent or emergent cases. Additionally, studies often restrict their subjects to those that are elective in part because of the amount of time required to recruit and perform baseline screening on potential subjects. Unfortunately, especially in the setting of coronary revascularization, the number of elective patients has been diminishing over time. As such, it is increasingly important for readers to take into consideration the differences between the study and target populations when drawing inferences from a given manuscript.

Additionally, issues of efficacy and effectiveness impact application of the results from such studies. Efficacy deals with a given intervention working under the strict conditions of a trial. In this case, we are interested in maximizing compliance to the treatment (all patients who are scheduled to have coronary revascularization with a bypass machine actually receive this intervention). Effectiveness, on the other hand, deals with whether the intervention works among individuals who are offered this treatment. While efficacy pertains to the idealized setting of a trial that likely has strict protocols, effectiveness pertains to "real life" situations, namely in the application of trial results in your medical center.

Who Is at Highest Risk: Clinical trials often wish to select subjects based on their predicted risk of an adverse event. There may be several reasons for this choice. For instance, these subjects are most likely to receive the greatest benefit, if any, from the intervention. Consequently, you might imagine that these same subjects might be more apt to comply with study protocols.

Choosing the Comparison Population

While ample time should be focused on choosing an appropriate study population, an equal amount of time should be devoted to identifying the appropriate comparison population. Although some studies will have more than one comparison population, researchers should strongly think twice about choosing more than two.

Consider this example: a researcher wishes to measure the benefit in terms of reduced restenosis rates of a new drug-eluting stent among patients with two-vessel coronary disease. He enrolls patients into one of three arms: a new drug-eluting stent, conventional coronary artery bypass grafting (CCAB), and bare metal stents.

So, what is the comparison that the researcher is making? A reasonable argument may be developed for the comparison of a new drug-eluting stent to CCAB for coronary restenosis. However, bare metal stents are likely not the comparison that affords this researcher with the greatest opportunity for new knowledge. These stents are no longer used commonly among most interventional cardiologists. If anything, the researcher likely would want to compare the new drug-eluting stent to the standard practice, which may be defined as either CCAB and/or an existing drug-eluting stent. As new devices are marketed, a standard practice likely would be to compare these devices to the best alternative.

Of note, a researcher will need to enroll considerably more subjects as he/she chooses multiple comparison populations.

Choosing the Outcome

The choice of outcome is another critical component of a randomized trial, or for that matter, any type of study design. Some of the key facets to keep in mind are choice of definition and method for assessing the outcome.

Choice of Definition: Irrespective of study design, one of the most important choices you will make is to choose the definition of your outcome. You might have more than one outcome that you are interested in tracking. The outcome most important to you is often called the primary outcome. Other outcomes of interest to you, but not of primary importance, are called secondary outcomes. In developing your definition, you should review the relevant literature and identify other studies that have focused on similar research questions and review their operational definitions. Other avenues to pursue might be to review

the definitions used by large registries. In the case of cardiovascular research, this includes the Society for Thoracic Surgery (<http://www.sts.org/>), American College of Cardiology (<http://www.acc.org/>), and the Northern New England Cardiovascular Disease Study Group (<http://www.nnedsg.org/>). If you are in agreement with the choices made by these registries, you might want to adopt their definitions for your study.

Some of you, however, might ask why definitions are so important. Let's say you are interested in neurologic injuries, specifically strokes, after coronary revascularization procedures. You might think a reasonable definition for stroke would be "a new fixed neurologic deficit." However, others might wish to know "how long does it have to last to be a stroke," "does a neurologist need to diagnosis the outcome," or "does the patient need to have the diagnosis verified by a brain imaging modality such as computed tomography." All of these questions are important, because they speak to why, in part, differences in rates of such outcomes exist. Some differences across studies thus may be attributed to variations in definitions and not solely based on "real" differences in surgical performance.

Method for Assessing the Outcome: The assessment of outcomes should be made in an unbiased fashion. As such, individuals responsible for assessing outcomes should be blinded to treatment assignment. Additionally, this assessment should be conducted by individuals who have no vested interest in the trial's results. As such, the outcomes assessor likely minimizes any bias, intentional or unintentional.

Study Design

Blinding: We are often told or taught that, in a well-conducted trial, all participants should be blinded to the treatment assignment. Blinding refers to the masking of certain aspects of a trial, whether in the treatment allocation or outcomes assessment. In reality, blinding may not always be feasible, such as in surgery. We may blind the patient to treatment assignment (whether their surgery was performed with or without cardiopulmonary bypass), called a single blind trial, but we could not feasibly blind both the patient and surgeon, called a double blind trial.

Why might we wish to blind? As academics, we think that the knowledge of treatment allocation may influence the patient's response. In the case of the patient, he/she might imagine that the off-pump alternative would provide a better outcome. As such, this bias might affect his/her presentation of symptoms. As a rule of thumb, blinding may be a worthwhile alternative for subjective outcomes, such as pain response. It may not be as valuable for more objective outcomes, such as atrial fibrillation.

In trials studying pharmaceutical agents, researchers may blind patients through the use of identical pharmaceutical capsules (e.g., those that taste and look similar in every way to capsules containing the active agents). Ad-

ditionally, if the study was double blinded, you would want to ensure that the practicing clinicians could not identify any laboratory slips containing this piece of information.

Randomization: Randomization is a methodology for ensuring equivalency of known and unknown potential confounders in both the study and comparison populations. On interviewing the patient and identifying that he/she would be eligible for the trial, the study coordinator would ask the subject if he/she would be willing to be randomized to either treatment arm. If the subject agrees, the study coordinator would open an envelope that would contain treatment assignment information. That assignment information would be recorded in a password-protected file and hidden from the principal investigator and other study personnel. The treatment assignment is often generated using a random number generator. These generators are available with any number of computer programs, such as Microsoft Excel (<http://www.microsoft.com/>), or with most statistical packages, such as Stata (<http://www.stata.com/>).

To know whether the randomization process worked, we require the collection of baseline information. Baseline information includes those factors that are most likely potential confounders of the association between the exposure and outcome. Investigators typically will create a “Table 1,” which is the first table in a research report. We would expect that the prevalence of key variables would be equally balanced between both treatment arms. If they are, as designated with a test of significance ($p < .05$), we would be confident that the randomization was successful. If not, we would question the method used to randomize subjects, as well as worry about the reporting of unadjusted findings.

Data Analysis

The two most common ways of analyzing the results from a randomized trial are intention-to-treat and treatment-received. The major difference in the findings resulting from these two analytic methods stems from what is termed cross-over. Subjects who are cross-overs change their initial treatment assignment, either through their choosing or based on the choice of the clinician.

In intention-to-treat, we conduct the analysis according to the treatment subjects were supposed to receive. This analytic method takes advantage of the control of confounding provided for by randomization. Additionally, intention-to-treat assesses the efficacy of the intervention under the idealized conditions of the trial. A potential disadvantage of this methodology is the underestimation of the true effect of the treatment, because not all subjects may have received the intervention, whereas we allocate the effect of the intervention based on the randomization.

In treatment-received, we conduct the analysis according to the treatment subjects actually receive, regardless of

initial treatment assignment. In this scenario, the effect of random allocation may be compromised for some subjects. This type of analysis may reflect actual care more than the intention-to-treat analysis. However, subjects who adhere to protocols may differ from those who don't adhere to protocols. For instance, in a study of the effect of chemotherapy vs. surgery for invasive cancer, deviations from protocol may not be attributed to the effect of the intervention itself but perhaps to adverse side effects secondary to the chemotherapy.

Regardless of a given trial's findings, researchers should conduct and report results from an intention-to-treat analysis, because it provides an unbiased estimate of the effect of the trial. There may be some value to also providing results from a treatment-received analysis, because it may reflect more closely a “real-life” scenario.

Issues of Inference

While there are many potential factors that may affect the findings of a trial, I will discuss two of these: placebo effect and the social desirability effect. Typically, the placebo effect is attributed to trials involving identifying the effect of a pharmacologic agent. The placebo effect involves a measurable or observed improvement in a subject's health that is not attributed to the inert agent itself but a perceived effect of the treatment. The social desirability effect, which may have application in cardiac surgery, relates to a subject's strong wish to respond favorably to a treatment to please society (e.g., the investigator).

Randomized trials are uniquely situated to allow readers to draw inference regarding causality. Because of their ability to remove issues of confounding by isolating the effect of the treatment, randomized trials allow one to attribute differences in outcomes to a subject's treatment allocation. In all other study designs, readers must take caution not to infer causation, but simply reflect on observed associations between a given exposure and outcome.

CONCLUSION

While it may be tempting for us to wish to always perform randomized trials, they are difficult to perform. Trials require much oversight, both in ensuring the safety of the subjects and in administration and conduct of the trial itself. Nonetheless, I hope that this manuscript has provided information useful for interpreting trials.

SUGGESTED READING

1. Hulley SB. Designing clinical research: an epidemiologic approach. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001.
2. Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998.