The practice of cardiology is increasingly driven by evidence-based medicine centering around the results of clinical trials. In this inaugural article of Clinical Cardiology: Physician Update, I will focus on the design, analysis, and interpretation of clinical trials in cardiovascular medicine. Armed with the tools described herein, clinicians will be better equipped to understand and synthesize the results of the multitude of clinical trials appearing in the cardiology literature. A major goal of this effort is to shorten the delay between the publication of trial results and the translation of their findings into clinical practice.

**Design of Clinical Trials**

Trials of new therapies in cardiology typically compare the new treatment to a control. The control group receives the treatment against which the test intervention is being compared. Control and test treatments must be both medically justifiable and compatible with the healthcare needs of study patients. Either treatment must be acceptable to study patients and to the physicians administering them; there must be a reasonable doubt regarding the efficacy of the test treatment; and there should be reason to believe that the benefits will outweigh the risks of treatment. When the control treatment is a placebo, the trial is referred to as a placebo-controlled trial.2

Given the burgeoning supply of new treatments in the cardiovascular armamentarium, more and more trials compare the test therapy to a standard therapy. This is referred to as an active controlled trial.3,4

Randomized controlled trials typically involve the randomization of patients to either the control or test treatment (ie, randomized concurrent control). These trials are the gold standard for evaluating new therapies, and they form the foundation for the highest level of recommendations in practice guideline documents that stress evidence-based medicine5 (Figure 1). Randomization has 3 important influences that explain why it is considered the standard for trial design: (1) it reduces the likelihood of patient selection bias that may occur either consciously or unconsciously; (2) it enhances the likelihood that comparable groups of subjects are compared, especially if the sample size is sufficiently large; and (3) it validates the use of common statistical tests such as the $\chi^2$ test for comparison of proportions and Student’s $t$ test for comparison of means.6

A stratified randomization scheme is typically used when it is considered important to achieve a balance of key baseline characteristics (eg, location of infarction) among the treatment groups. The ratio of randomization to the control and test therapies may be equal or unbalanced, allowing investigators to acquire more information about the test therapy (eg, risk of intracranial hemorrhage with a new fibrinolytic) while still maintaining a comparison with a control group. Because many clinical trials in cardiology are multicenter in nature, an attempt is made to achieve balance between the treatment assignment at each enrolling center by constraining the randomization to the treatment groups so that the desired ratio occurs in blocks of patients (eg, every 6 or 8 patients) at each center.

When the investigator selects the subjects to be allocated to the control or treatment groups, the study is referred to as a nonrandomized, concurrent control trial.7 Unlike randomized controlled trials, such trials are subject to bias because it may be difficult for investigators to match the test and control groups adequately. Clinical trials using historical controls compare a test intervention to data obtained earlier in a nonconcurrent, nonrandomized control group.7 The use of historical controls allows clinicians to offer potentially beneficial therapies to all subjects, thereby reducing the sample size for the study. The major drawbacks are bias in the selection of the control population and potential failure of historical controls to reflect contemporary diagnostic criteria and current treatment regimens for the disease under study.

Other trial designs useful for evaluating cardiovascular treatments are the crossover design and withdrawal study. The crossover design is actually a special case of the randomized controlled trial in which each subject serves as
his or her own control. Comparisons are made between the treatment response seen during the first treatment to which the patient is allocated and during subsequent treatments. In withdrawal studies, patients with a chronic cardiovascular condition are taken off therapy and a comparison is made between the treatment effect observed on therapy with that observed off therapy. Potential limitations of withdrawal studies are that only patients who have tolerated the test intervention for a period of time are eligible for enrollment and changes in the natural history of the disease may influence the response to withdrawal of therapy.

Given the multitude of drugs administered simultaneously to patients with cardiovascular diseases, an increasingly important area of investigation is drug interactions. Trials that test 2 therapies simultaneously typically use a factorial design. Such trials are most appropriate when there is thought to be no interaction between the test treatments. If such interactions are known to exist or are discovered to exist, it is important to evaluate each test treatment against a control treatment.

To minimize the possibility of bias, blinding (sometimes referred to as masking) of treatment assignment is used. When only the patient is unaware of the treatment assignment, the trial is single blind. If the investigator is also blinded, the trial is double blind.

The responsibility for monitoring evolving efficacy and safety on an interim basis during the conduct of the trial rests with a Data Safety Monitoring Board or Committee. It is usual practice to prespecify the level of statistical evidence the Data Safety Monitoring Board might observe at an interim analysis that might lead to a recommendation of premature discontinuation of the trial because of overwhelming evidence of benefit or harm from the test therapy (eg, stopping boundaries). Investigators and sponsors sometimes also prespecify a calculation of the conditional probability of the trial achieving its objective(s) given the observations at a particular interim look (eg, when half the expected number of patients have been enrolled or half the expected number of events have occurred). Potential recommendations that the Board might make include increasing the sample size if the event rate is lower than expected to increase the power of the study or discontinuation of the study for futility.

**Statistical Considerations**

A formal statement of the objective of a clinical trial is contained within the prespecified null and alternative hypotheses. In a superiority trial, the null hypothesis states that the test and control therapies are equally effective. The alternative hypothesis states that the event rate is lower in patients receiving the test therapy than in those receiving control therapy. The type I error ($\alpha$=false-positive statement that erroneously declares there is a difference between the treatments) is typically 2-sided and set at the 5% level. The type II error ($\beta$=false-negative statement that erroneously declares there is no difference between the treatments) is typically set between 10% and 20%, such that the power of the trial (1-$\beta$) is between 90% and 80%, respectively. The combination of $\alpha$ and $\beta$ errors determines the sample size of the trial.

Fortunately, the mortality rate from cardiovascular illnesses continues to decline. The implication of a low event rate in the control group is that tens of thousands of patients must be randomized to show a difference between treatments. Many trials in cardiology therefore use a composite end point, such as death plus nonfatal myocardial infarction, to maintain a practical size for the trial. When interpreting the results of trials with composite end points, it is important for clinicians to note whether the direction and magnitude of the treatment effect is similar for each of the elements of the end point.

In cardiovascular therapeutics, several efficacious treatments may coexist for a given condition. Potential differences in clinically important areas such as tolerability, ease of administration, and cost may lead investigators to perform a trial demonstrating therapeutic equivalence of 2 treatments. Because it is not possible to show that 2 active therapies are completely equivalent without a trial of infinite sample size, investigators specify a value ($\Delta$) and consider the test therapy equivalent to the standard therapy if, with a high degree of confidence, the true difference in treatment effects is less than $\Delta$. In this case, the null hypothesis states that the rate of events in patients receiving the test therapy exceeds the rate in patients receiving control therapy by $\Delta$. The alternative hypothesis states that the rate of events in patients receiving the test therapy is less than the rate in patients receiving control therapy plus $\Delta$. In a classical equivalence trial, if the effects of the 2 treatments differ by more than the equivalence margin (ie, $\Delta$) in either direction, then equivalence is said not to be present. In practical terms, the usual objective in clinical trials of 2 active therapies is to establish that the new therapy is not worse than the standard therapy by more than $\Delta$. Such one-sided comparisons are referred to as noninferiority trials (Figure 2). The new therapy may satisfy the definition of noninferiority but, depending on the results, may or may not actually show superiority compared with the standard ther-
The event rates in the groups are compared using a difference in event rates. Fisher's exact test to determine the statistical significance of the difference in event rates between the test drug and standard drug is plotted with a depiction of the point estimates (black squares labeled A through F) and the 2-sided 95% confidence intervals. Trial A has confidence intervals that fall entirely to the left of zero and are consistent with superiority of the test drug compared with the standard drug. Although the point estimates for trials B and C are different, the upper bound of the confidence interval (1-sided) falls within the zone of non-inferiority, allowing investigators to claim that the test drug is "not inferior" to the standard drug and, for practical purposes, may be considered therapeutically similar. Trials D and E do not satisfy the criteria for non-inferiority because the upper bound of the confidence intervals falls to the right of the right edge of the zone of non-inferiority. The findings of trial E are actually suggestive of inferiority of the test drug compared with the standard drug. Trial F is an inconclusive study due to its small sample size and extremely wide confidence intervals, preventing investigators from claiming either superiority or non-inferiority of the test drug compared with the standard drug.

The α and β errors of the noninferiority trial determine the sample size just as for superiority trials.

It is important that investigators prespecify the noninferiority margin before learning the trial results to avoid the bias that might be introduced by retrofitting a noninferiority margin such that the test therapy satisfies the definition of noninferiority. Specification of the appropriate margin or Δ is a challenging area involving the desire of regulatory authorities to be assured that the test therapy is at least superior to placebo and the desire of clinicians to set Δ at a clinically meaningful difference between treatments.

Critical Readings of Clinical Trials

By asking 3 main sets of questions, such as those in the Table, physicians can integrate the information in articles describing clinical trials into their own practice. 

- Were patients analyzed in the groups to which they were randomized?
- Were patients, their clinicians, and study personnel "blind" to treatment?
- Were the groups similar at the start of the trial?
- Aside from the experimental intervention, were the groups treated equally?

What were the results?

- How large was the treatment effect?
- How precise was the treatment effect?

Will the results help me in caring for my patients?

- Does my patient fulfill the enrollment criteria for the trial? If not, how close is my patient to the enrollment criteria?
- Does my patient fit the features of a subgroup in the trial report? If so, are the results of the subgroup analysis in the trial valid?
- Were all the clinically important outcomes considered?
- Are the likely treatment benefits worth the potential harm and costs?

Adapted from the data in References 12 and 13.

Figure 2. Example of design and interpretation of noninferiority trials. The prespecified zone of noninferiority is usually based on prior trials comparing the standard drug to placebo. In the example shown, the difference in event rates between the test drug and standard drug is plotted with a depiction of the point estimates (black squares labeled A through F) and the 2-sided 95% confidence intervals. Trial A has confidence intervals that fall entirely to the left of zero and are consistent with superiority of the test drug compared with the standard drug. Although the point estimates for trials B and C are different, the upper bound of the confidence interval (1-sided) falls within the zone of non-inferiority, allowing investigators to claim that the test drug is "not inferior" to the standard drug and, for practical purposes, may be considered therapeutically similar. Trials D and E do not satisfy the criteria for non-inferiority because the upper bound of the confidence intervals falls to the right of the right edge of the zone of non-inferiority. The findings of trial E are actually suggestive of inferiority of the test drug compared with the standard drug. Trial F is an inconclusive study due to its small sample size and extremely wide confidence intervals, preventing investigators from claiming either superiority or non-inferiority of the test drug compared with the standard drug.

By asking 3 main sets of questions, such as those in the Table, physicians can integrate the information in articles describing clinical trials into their own practice.

1. Was the assignment of patients to treatment randomized?
2. Were all patients who entered the trial properly accounted for at its conclusion?
3. Was follow-up complete?
4. Were patients analyzed in the groups to which they were randomized?

Questions to Ask When Reading and Interpreting the Results of a Clinical Trial

Primary guides

1. Were all the clinically important outcomes considered?
2. Are the likely treatment benefits worth the potential harm and costs?

Secondary guides

1. Does my patient fulfill the enrollment criteria for the trial? If not, how close is my patient to the enrollment criteria?
2. Does my patient fit the features of a subgroup in the trial report? If so, are the results of the subgroup analysis in the trial valid?
3. Were all the clinically important outcomes considered?
4. Are the likely treatment benefits worth the potential harm and costs?

Against the benefits associated with a test therapy, clinicians must weigh the risks associated with its use. Terms that describe the harmful effects of test therapies include the absolute risk increase (the absolute increase in events with the test therapy compared with control therapy) and the number needed to harm (1/absolute risk increase). The composite of benefit and harm has sometimes been expressed as net clinical benefit. An example is the composite of lives saved with fibrinolytic therapy for ST elevation myocardial infarction plus the number of patients who survive but suffer a severe disabling stroke from intracranial hemorrhage.

When weighing the evidence from a single clinical trial for a treatment decision in an individual patient, physicians must consider more than the level of significance of the findings. A judgment must be made about whether the patient is...
Randomized Controlled Trials

Summary Measures of Treatment Effect

<table>
<thead>
<tr>
<th>Patients meeting enrollment criteria</th>
<th>NO EVENT</th>
<th>EVENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment A: N = 6000</td>
<td>4400</td>
<td>5000</td>
</tr>
<tr>
<td>Treatment B: N = 6000</td>
<td>4250</td>
<td>5000</td>
</tr>
</tbody>
</table>

**Figure 3.** Evaluation of a clinical trial. In this example, 10,000 patients meeting enrollment criteria for the randomized controlled trial are randomized such that 5000 patients receive treatment A and 5000 patients receive treatment B. A total of 600 patients in group A experience an event (eg, mortality), yielding an event rate of 12%; 750 patients in group B experience an event, yielding an event rate of 15%. The 2×2 table on the right is then constructed, and various statistical tests are performed to evaluate the significance of the difference in event rates between groups A and B (R₁ and R₂, respectively). Common statements describing the treatment effect are relative risk (of events in group A versus group B), odds ratio (for development of events in group A versus group B), or the absolute risk difference (of events in group A versus group B) using the formula shown. A clinically useful method of expressing the results is to calculate the number of patients that need to be treated to prevent one event. Adapted from Figure 10-6 of Antman E. Medical therapy. In: Califf R, ed. Guidelines (Committee on the Management of Patients With Unstable Angina). J Am Coll Cardiol. 2000;36:970–1062.

References


Key Words: drugs • statistics • cost-benefit analysis • trials • myocardial infarction