

# A Primer on the Design, Conduct, and Interpretation of Clinical Trials

Lawrence J. Appel

Departments of Medicine, Epidemiology, and International Health (Human Nutrition), Johns Hopkins University, Baltimore, Maryland

Clinical trials are an especially powerful study design that often guides health care policy and clinical practice. Indeed, well-designed and rigorously conducted trials can establish the etiologic relevance of modifiable risk factors and the benefits (and risks) of candidate therapies. Contemporary schema that classify evidence place results from randomized trials at the pinnacle of evidence. The primary objective of this article is to provide an overview of the design, conduct, and interpretation of trials with an emphasis on aspects that are relevant to nephrology.

*Clin J Am Soc Nephrol* 1: 1360–1367, 2006. doi: 10.2215/CJN.02850806

Clinical trials are experiments that are conducted in humans and are designed to test the effects of a treatment or prevention strategy on prespecified outcomes. As an experiment, a trial attempts to establish and maintain a contrast in therapies between randomly assigned groups. The types of treatments that can be tested are vast and include drugs (e.g., erythropoietin), procedures (e.g., types of vascular access), levels of risk factor modification (e.g., lower BP goal), lifestyle intervention (e.g., weight loss), dietary modification (e.g., reduced protein intake), and health education (e.g., self-monitoring techniques). Clinical trials have special importance because appropriately designed and rigorously conducted trials influence health care policy and provider recommendations. Less well appreciated are the numerous scientific and logistical challenges that are encountered when one designs and implements such studies. Table 1 summarizes the major advantages and disadvantages of clinical trials.

Critical to the proper interpretation and evaluation of trial results is adequate publication. In response to widespread deficiencies, several initiatives have taken place to improve the quality of publications. The Consolidated Standards of Reporting Trials (CONSORT) initiative (1) has had a particularly prominent impact by setting standards for the types of information that must be included in publications of completed trials. A checklist of required items (Table 2) and a flow diagram (Figure 1) are particularly useful tools for readers of trial reports.

## Basic Features of Trials

All trials, including trials designed to prevent or treat kidney disease, share common features that are described next.

Published online ahead of print. Publication date available at [www.cjasn.org](http://www.cjasn.org).

Address correspondence to: Dr. Lawrence J. Appel, Johns Hopkins University, 2024 East Monument Street, Suite 2-618, Baltimore, MD 21205-2223. Phone: 410-955-4156; Fax: 410-955-0476; E-mail: [lappel@jhmi.edu](mailto:lappel@jhmi.edu)

## Core Design

The most common type of design is a parallel-arm trial with two or more groups: A comparison group and at least one new (or active) therapy group (Figure 2). In such a design, participants remain assigned to their randomized group until the end of follow-up. Parallel-arm trials are relatively easy to understand, especially the basic two-group (arm) trial.

An alternative is the crossover design, in which individuals are randomly assigned to a sequence (comparison treatment followed by new treatment, or *vice versa*). Crossover trials can be performed when the outcome is reversible, such as BP. An example of a crossover study is the Dietary Approaches to Stop Hypertension Sodium trial, which tested the effects of three different sodium intakes on BP in people with prehypertension or hypertension (2). A key issue for crossover trials is differential carryover; that is, does the effect of the first-period therapy affect the outcome measurement in the second period. Before the conduct of a trial, data on carryover effects rarely are known. After the trial, carryover effects rarely can be detected because of inadequate statistical power. In view of these considerations, most trialists will attempt to separate the interval between outcome data collection in each period by inserting a washout period.

A common type of parallel-arm trial is the factorial trial (Figure 3). This type of trial tests at least two therapies (e.g., treatment X and treatment Y). Participants receive only treatment X, only treatment Y, both treatments, or neither treatment. Factorial designs provide tremendous efficiency by testing two therapies in the same population simultaneously. The sample size often is no larger than a corresponding parallel-arm trial that tests only one treatment. If there is no interaction between therapies, then one can test the effect of treatment X by combining the results across groups, regardless of whether they receive treatment Y (Figure 3, cells b + d *versus* cells a + c). Likewise, one can test the effect of treatment Y by combining the results across groups, regardless of whether they receive treatment X (Figure 3, cells c + d *versus* cells a + b). However,

Table 1. Advantages and disadvantages of clinical trials

<p><b>Advantages</b></p> <ul style="list-style-type: none"> <li>often provides the strongest evidence in support of cause–effect relationships</li> <li>basis for clinical and public health policy</li> <li>minimize/eliminate bias and confounding</li> </ul> <p><b>Disadvantages</b></p> <ul style="list-style-type: none"> <li>many (if not most) research questions cannot be tested in trials because:                     <ul style="list-style-type: none"> <li>exposure is fixed (e.g., gender)</li> <li>contrast between randomized groups is unobtainable (e.g., pre-ESRD patients unwilling to be randomly assigned to hemodialysis or peritoneal dialysis)</li> <li>the disease is rare; therefore, participants cannot be enrolled in a timely manner</li> <li>practical issues</li> <li>expensive</li> <li>logistically complex</li> <li>recruitment</li> </ul> </li> <li>innumerable ethical issues</li> <li>relevance rapidly diminishes in setting of secular trends</li> <li>rapidly evolving treatment</li> </ul>
---

when unanticipated treatment interactions are detected, factorial design trials can be underpowered because of the inability to combine results for one treatment across treatment groups of the other treatment. Using the notation in Figure 3, cell a would be compared with cell b separately from the comparison of cell c with cell d. A factorial design has been used in several kidney disease trials, including the Modification of Diet in Renal Disease (MDRD) (3), African American Study of Kidney Disease and Hypertension (AASK) (4), and Hemodialysis (HEMO) (5) trials. For example, AASK was a 3 × 2 factorial trial in which participants were randomly assigned to one of three medications that are used as initial antihypertensive therapy (ramipril, metoprolol, and amlodipine) and also randomly assigned to one of two levels of BP control (usual mean arterial pressure goal of 102 to 107 mmHg versus a low mean arterial pressure goal of <92 mmHg) (4).

*Superiority versus Equivalence Trials*

Most trials test whether a new treatment is superior to a control (placebo) group or conventional standard of care. Such a trial design, termed a *superiority* trial, is appropriate when there is no effective therapy or when the new or test treatment plausibly may be more effective than conventional therapy. In contrast, some trials are designed to show that a new treatment is not inferior to standard therapy by a predefined acceptable amount (6). If noninferiority is established, then the new therapy might be preferred because of other considerations (e.g., lower cost, better safety, greater convenience). A host of methodologic issues surround the design and interpretation of equivalence trials (also termed *noninferiority* trials). A key issue

Table 2. Checklist of items to be included in reports of randomized trials: CONSORT recommendations

Title and abstract	Blinding (masking)
Introduction	Statistical analysis
background	Results
Methods	participant flow
participants	recruitment
interventions	baseline data
objectives	numbers analyzed
outcomes	outcomes and estimation
sample size	ancillary analyses
Randomization	adverse events
sequence generation	Comment
allocation	interpretation
concealment	generalizability
Implementation	Overall evidence

that can limit the interpretation of equivalence trials is the assessment of whether the sample size is large enough to rule out a small difference that nonetheless might be clinically relevant. An example of an equivalence trial is Controlled Onset Verapamil Investigation of Cardiovascular End Points (CONVINCE), a trial that tested whether initial therapy with controlled-onset extended-release verapamil is equivalent to a physician’s choice of atenolol or hydrochlorothiazide in preventing cardiovascular disease (CVD) (7).

*Target Population*

The target population represents the set of individuals to whom the results of the trial might be applied. Ideally, the target population corresponds to the type of individuals who are enrolled in the trial as specified by the inclusion and exclusion criteria. Inclusion criteria typically are set in a “positive” terminology (e.g., patients with diabetes, hemodialysis patients). Exclusion criteria are a list of specific criteria for which the investigators will exclude people, even if interested. To maximize the generalizability of trial results, contemporary trials tend to minimize exclusion criteria. Legitimate exclusion criteria are those that are related to safety and the ability to collect the primary outcome variable.

*Recruitment*

Recruitment of participants is one of the most challenging aspects of conducting clinical research. Investigators must identify an accessible population, which is a subset of the target population (Figure 4). They also must secure approval from their institutional review board for their recruitment plans. Common recruitment strategies include laboratory-based and medical record–based screening, mass mailings, physician referrals, and mass media advertisements (newspapers, television, radio, and web sites). In most trials, the number of contacted people who actually enroll is small. Hence, people who do participate are with rare exception a nonrandom subset of the accessible population (and target population).

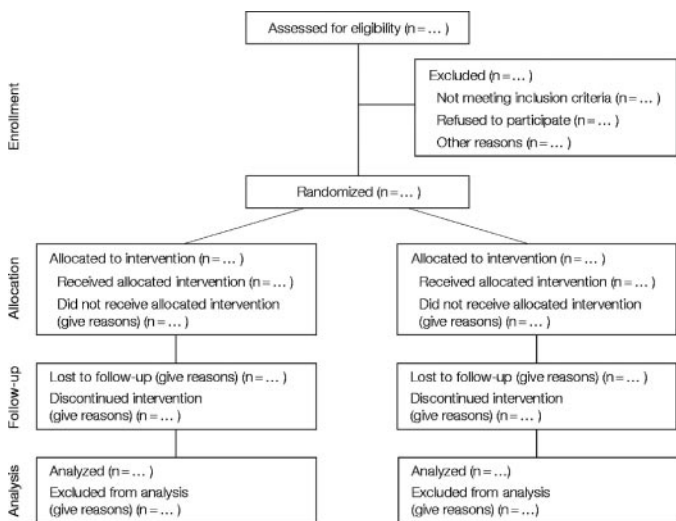


Figure 1. Flow diagram recommended by the CONSORT initiative. Reprinted from reference (1), with permission.

**Randomization**

A hallmark of contemporary trials is randomization of group assignment. The goal of randomization is to allocate participants in an unbiased manner to the possible treatment groups (a process that is independent of the characteristics and preferences of the participants). Randomization also facilitates blinding of study investigators and participants, because there should be no apparent reason for assignment, which, if random, cannot be predicted easily.

Although randomization should distribute evenly factors, both known and unknown, that influence trial outcomes, there is no guarantee. Indeed, by chance alone, baseline levels of important prognostic variables might differ by randomized group. Although this problem can occur in studies of all sizes, it is a particular concern in small trials. Strategies to address the problem include stratification of randomization on the basis of a major prognostic variable. For example, in a trial of therapies to delay progression of chronic kidney disease (CKD), stratification of randomization on the basis of urine albumin excretion might be desirable to ensure that this powerful risk factor is distributed evenly in randomized groups. For logistical reasons, it is difficult to stratify on more than one or two variables.

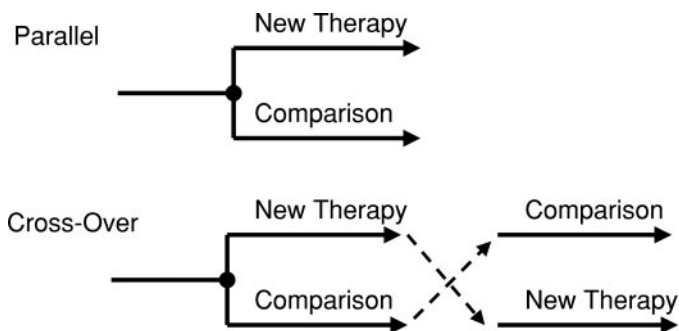


Figure 2. Basic types of trial designs with two groups: A new therapy group and comparison (control) group.

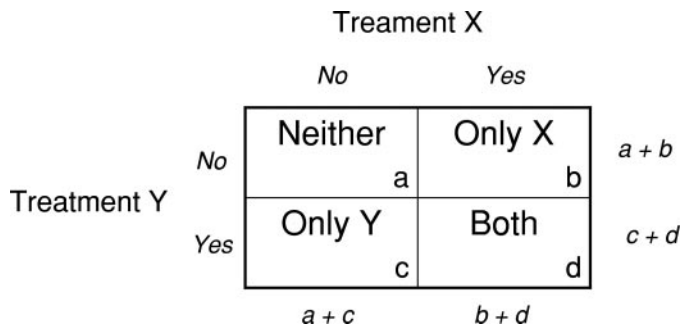


Figure 3. The 2 × 2 factorial design.

An alternative is statistical adjustment for known prognostic variables that are collected before randomization.

Alternatives to randomization have been proposed, but each can lead to biased assignment. These include systematic allocation (e.g., allocation based on birth month, odd-even allocation based on last digit of a hospital identification number or national identification number), nonrandom concurrent controls, and historical controls (8).

**Selection of Active Treatment and Dosage**

A critical issue is selection of active treatments and dosage. Practical considerations, some obvious and some not so obvious, often drive the decision to select a particular therapy. A pharmaceutical company conducts a trial to test a specific product. In other instances, a government sponsor that is testing different types of antihypertensive agents on clinical outcomes may select an agent, in part, on the basis of the willingness of a company to provide active drug and matching placebo according to the timeline of the trial.

Selection of dosage is an inexact science, often driven by phase 1 or 2 dose-response studies followed subsequently by results of phase 3 trials. Some trials use fixed dosages when it is not possible to titrate dosage in response to some indicator variable. Other trials adjust dosage to achieve some predefined goal (e.g., systolic BP <140 mmHg) or a maximally tolerated dosage.

**Control (Comparison) Group**

One of the most contentious aspects of trial design is selection of the comparison group. When there is no effective ther-

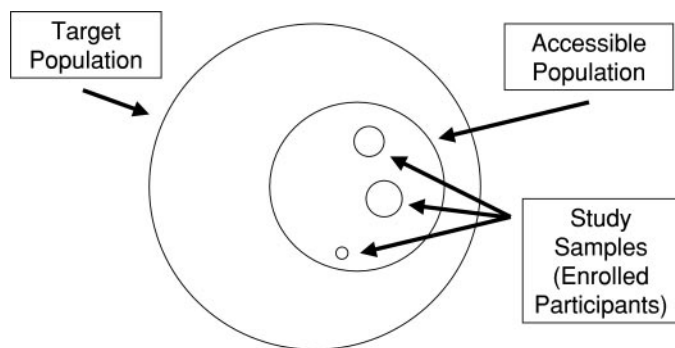


Figure 4. Relationship among target population, accessible population, and study samples (enrolled participants).

apy, a pure control or placebo group is appropriate. The decision to use a placebo *versus* conventional therapy often is controversial (9). A placebo (inert substance) is preferable to “no treatment” without a pill for two reasons. First, the act of taking a pill or receiving a treatment may exert some benefit apart from the active ingredient. This effect is termed *the placebo effect*. Second, use of a placebo pill should minimize the risk for unblinding participants and data collectors if the placebo is identical or very similar in appearance, taste, and smell to pills with active therapy.

To determine whether a “well-accepted” therapy exists, one typically relies on promulgated, evidence-based guidelines or recommendations from professional societies or government bodies. If there is an accepted therapy or recommendation, then the comparison group is an “active” control group that receives recommended therapy. For instance, early antihypertensive therapy trials were placebo controlled (10), whereas more recent trials, such as Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) (11), have used active control groups.

### Types of Data

The two most critical variables in a randomized trial are the outcome variable and the randomization assignment. For this reason, considerable effort is placed on ensuring collection of high-quality outcome data and on ensuring that the randomization process is implemented in an unbiased manner and that participants receive their assigned therapy.

**Baseline Variables.** Baseline data pertain to data that are collected before randomization and include sociodemographic information, relevant comorbidities, and key prognostic variables that influence the outcome variable of interest. Such data are used to describe the participants, compare randomized groups, and, more recently, develop algorithms for imputation of missing outcome data.

**Process Variables.** Process data are postrandomization data that are collected for the purpose of monitoring adherence, determining the potential for confounding, and monitoring safety. Some of these data also can be considered outcomes in secondary trial analyses. Adherence data include pill counts, biologic markers (*e.g.*, blood levels of a drug or biologic marker), and visit attendance. Adherence data, if available during the trial, can be used to modify trial procedures to improve compliance and achieve the intended experimental contrast.

Postrandomization data also are collected to understand the nature of possible confounding. Although trials are designed to test a specific intervention, there is the risk for confounding; that is, the intervention is associated with some other intervention or process that influences the health outcome. For instance, most experts believed that the benefits of statin therapy on CVD prevention resulted from cholesterol reduction; however, this class of drugs has pleiotropic effects, including reductions in BP, inflammation, and proteinuria, effects that also could reduce CVD risk (12).

**Outcome Variables.** Outcome variables are a set of variables that are likely to be affected by the study treatments.

Typically, one variable is specified as the primary outcome variable of the trial. Outcome variables should be prespecified. The primary outcome should be stated explicitly in the trial hypothesis and is typically the variable that is used to determine sample size. Ideally, the outcome variable is specified with considerable detail. For example, rather than stating BP, a fairly general term, the protocol should specify key aspects related to its timing and measurement (*e.g.*, systolic BP, determined from the average of three sets of resting measurements obtained 12 mo after randomization).

Trial outcomes used to guide policy are well-recognized clinical outcomes that are not subject to biased ascertainment. Total mortality is perhaps the least biased variable; in contrast to morbid events, such as stroke, it is not subject to misclassification. However, trials rarely have adequate power to detect differences in total mortality. Often, trials will use cause-specific, morbid events that have obvious clinical relevance, such as onset of renal replacement therapy.

### Masking (Blinding)

Masking (also termed blinding) is a common and important feature of clinical trials. The principal objective of masking is to avoid biases that might occur should an individual change his or her behavior in response to knowledge of randomization assignment (*e.g.*, the participant who is disappointed with assignment to control and then drops out or seeks alternative means to obtain the active treatment, the data collector who interprets a clinical outcome such as a symptom differently because the person was assigned to a presumably effective therapy). A single-blind study keeps outcome assessors (data collectors and clinicians) blinded to assignment. A double-blind study keeps participants blinded, as well as the outcome assessors. In certain studies (*e.g.*, trials that compare surgery with nonsurgery, types of dialysis [hemodialysis *versus* peritoneal dialysis], dietary change such as protein restriction, lifestyle modification such as weight loss), it is extremely difficult, if not impossible, to keep participants blinded to assignment. Although there could be exceptions (*e.g.*, sham surgery), such extraordinary procedures to achieve blinding of study participants are rare.

### Sample Size

Sample size estimation, although a prerequisite to conducting a trial, is largely an educated guess with substantial implications. A large sample size increases costs, may increase the number of participating centers, and may even place a trial in jeopardy if the expected cost exceeds available resources. Regrettably, the precision of these estimates is suboptimal. To determine the sample size for trials with clinical outcomes, one needs to

- Estimate the number of participants in the comparison group who will experience the primary trial outcome
- Estimate a plausible reduction in outcomes that might occur as a result of the new treatment
- Set the type 1 error (also termed the  $\alpha$  level), which is the probability of falsely rejecting the null hypothesis (when the

null hypothesis is true). Commonly, the  $\alpha$  level is set at 0.05, which is equivalent to falsely rejecting the null hypothesis 5% of the time.

- Set the type 2 error, which is the probability of falsely not rejecting the null hypothesis (when the alternative hypothesis is true). Commonly, this is set at 0.20, which is equivalent to falsely not rejecting the null hypothesis 20% of the time when the alternative hypothesis is true.

### Statistical Analyses

The trial protocol should prespecify the primary analysis plan, which includes a description of the statistical model used to test the trial hypothesis. An appropriate analysis plan also will describe an approach to participants with missing data. Typically, the approach to missing data requires an imputation strategy in which missing values are replaced with values that are imputed from a model that used data from similar people with available data. The analysis plan also should prespecify subgroups of interest and document the approach, if any, to multiple comparisons.

## Issues Related to the Design and Interpretation of Clinical Trials

### Subgroup Analyses

Trials typically are designed and powered to determine the overall effects of treatment in all enrolled participants, yet clinicians commonly are interested in the effects of treatment in their patients, as defined by some characteristic such as gender, age, race, genotype, or stage of disease. When treatment differences are present, an “interaction” exists. There are two possible types of interactions: Qualitative and quantitative. A qualitative interaction occurs when an intervention has benefit in one group and harm in another. Qualitative interactions are extremely rare. A more common type of interaction is a quantitative interaction, in which there is a difference in the extent of benefits (*i.e.*, more benefit in one group than another). A quantitative interaction is more likely but also more difficult to detect than a qualitative interaction.

Subgroup analyses are prone to false-negative and false-

positive results (13). For this reason, readers should focus on main effects (*i.e.*, in the whole population), rather than results in subgroups, even if prespecified.

### Surrogate Outcomes

The most persuasive trials are ones that use clinical events or well-accepted surrogate variables as their outcomes. Trials with surrogate outcomes typically are smaller in size and therefore much less costly. Surrogate outcomes often are a measure of the underlying disease process (*e.g.*, C-reactive protein), a measurement of preclinical disease (*e.g.*, coronary artery calcifications), or an etiologically relevant, well-accepted risk factor (*e.g.*, systolic BP, LDL cholesterol). The list of candidate surrogate outcomes is huge, but only a few are so well accepted that trials that use these variables actually influence policy. Trials of CKD progression have used proteinuria, albuminuria, and serum creatinine as surrogate outcomes. The presumption is that change in these surrogate outcomes is tightly linked or at least highly predictive of clinical CKD outcomes such as ESRD. However, policy-making committees and bodies have not always been influenced by the results of trials with surrogate outcomes, because the clinical relevance of most surrogate outcomes is uncertain (14).

Results from the AASK highlight uncertainties about the validity of proteinuria as a surrogate outcome. In AASK, assignment to a lower BP goal (*versus* usual BP goal) did not lead to improved clinical outcomes as indicated by the primary outcome (GFR) (Figure 5A) or the secondary outcome (clinical renal events). However, the low-BP group did reduce proteinuria substantially and significantly in comparison with the usual-BP group (Figure 5B). This divergence between clinical outcomes and proteinuria provides evidence that proteinuria is suboptimal as a surrogate marker of CKD progression.

### Internal and External Validity

Ideally, trials are designed and conducted both to minimize bias (*i.e.*, have high internal validity) and to be relevant to a wide but defined population (*i.e.*, have high external validity, also termed generalizability). In reality, a tradeoff between

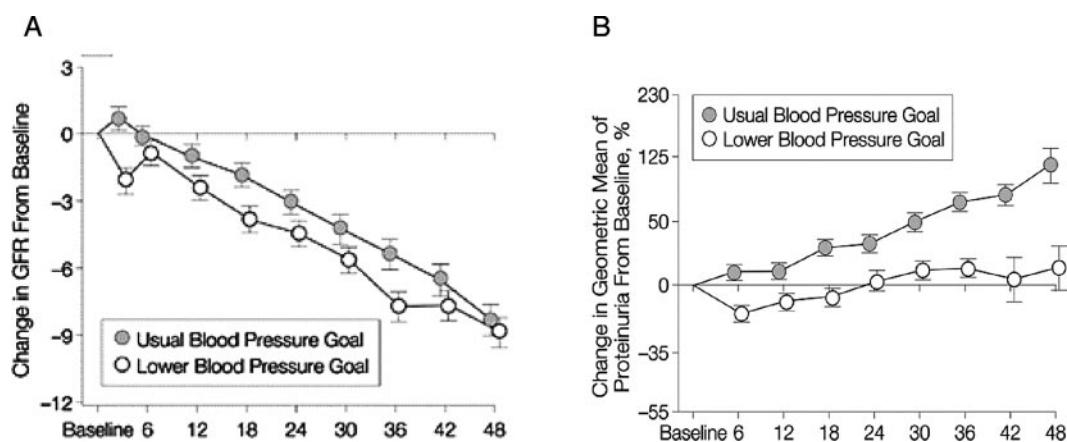


Figure 5. Mean change in GFR (A) and proteinuria (B) in usual and low mean arterial pressure groups in the AASK trial. Reprinted from reference (4), with permission.

internal and external validity is commonplace. The determinants of internal validity are easily appreciated and include appropriately implemented randomization procedures, complete follow-up with few missing data, and high adherence with low drop-in and dropout rates. In contrast, the determinants of external validity are less intuitive and rely on a clinical judgment that is related to the therapy under study and the types of candidate patients (15). Trials with high external validity tend to be inclusive and enroll people who might not adhere to the trial protocol (e.g., refusing follow-up, dropping out of intervention). As such, trials with high external validity may sacrifice internal validity. In general, because a biased or poorly executed trial is unlikely to influence policy or practice, even if it has apparent external validity, investigators place a greater emphasis on internal validity than external validity (Figure 6).

*Intention-to-Treat (versus “On Treatment” Analyses)*

Intention-to-treat is the preferred approach to the analysis of clinical trials. According to this paradigm, individuals are analyzed according to randomized assignment, independent of whether they remain on assigned therapy. Drop-ins are those who are assigned to a control group and adopt the new therapy, whereas dropouts are those who are assigned to a new therapy but discontinue it (Figure 7). A second important aspect of intention-to-treat requires investigators to impute missing outcome data. Missing data commonly occur in clinical studies. Missing data can occur at random; that is, “missingness” is unrelated to assigned therapy; examples are missing data from individuals who move out of the country or who die in an accident. Alternatively, missing data can occur in a non-random manner; that is, the reason for missingness is related to the therapy (e.g., a participant stops participating in a trial because of toxicity related to an intervention, such as adverse effects of steroid therapy). A valid intention-to-treat analysis will impute outcome data for individuals with missing measurements.

An “on treatment” approach is a suboptimal approach to analysis in which participants are analyzed according to ther-

apy that they actually received. Some people stop active therapy, whereas others who are assigned to the control condition start taking active therapy. Despite the intuitive appeal of an on-treatment analysis, substantial biases are associated with this approach. Specifically, in an on-treatment analysis, active treatment is confounded with high adherence, an attribute that commonly is associated with better outcomes (16).

*Efficacy versus Effectiveness*

Efficacy pertains to the effects of a therapy or intervention under optimal conditions, which often exceed the quality of medical care that is available routinely. Effectiveness pertains to the effects of a therapy under more typical conditions in which the intervention or therapy is likely to be implemented. For example, in efficacy trials, extraordinary efforts are made to encourage adherence to therapy, including reimbursement for transportation costs, easy access to skilled nurses, advice from highly trained behavioral counselors, and even financial incentives. A common criticism is that the provision of therapies in efficacy trials does not conform to what is done or even achievable in routine practice. A counterargument is that efficacy trials should precede effectiveness trials to understand the maximum likely benefit that might be achieved. The DASH trial, which was a randomized feeding study that tested the effects of dietary patterns on BP, is one example of an efficacy study (17). In contrast, the PREMIER trial, which tested the effects of a lifestyle program that is designed to change dietary patterns, can be considered an effectiveness trial (18).

*Clinical versus Statistical Significance*

Clinical significance pertains to the magnitude of benefit that would be sufficient to change practice or public health policy. It is not a fixed number and depends on the potential application; for example, a large BP reduction of approximately 10 mmHg would be expected for a new drug therapy for established hypertension, whereas a more modest BP reduction of 2 to 3 mmHg would be deemed effective as a public health intervention applied to a whole population (19). Statistical significance

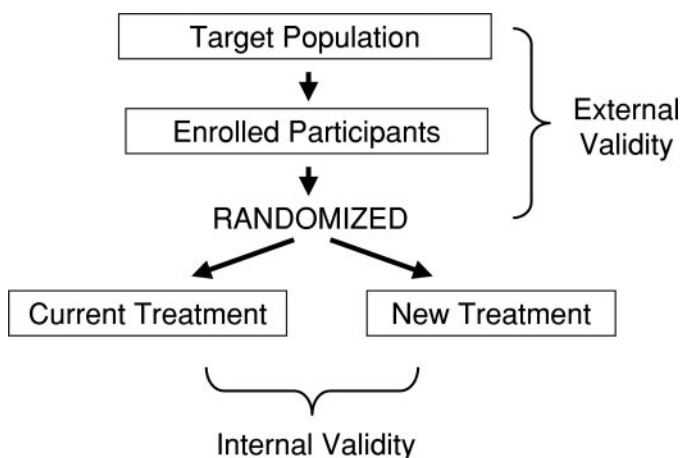


Figure 6. Internal and external validity.

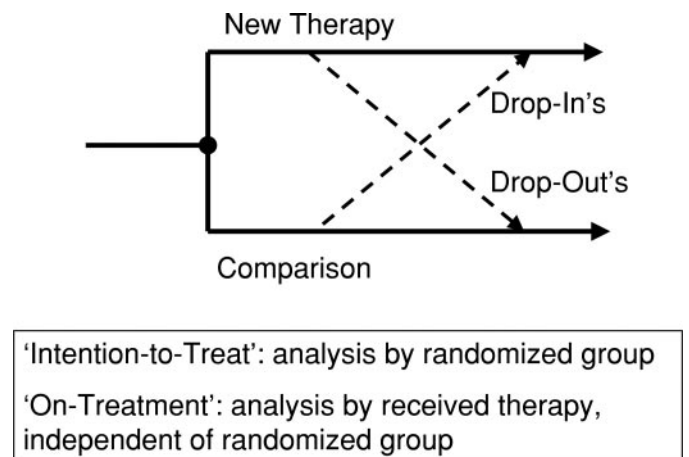


Figure 7. Adherence: Drop-ins, dropouts, intention-to-treat analysis, and on-treatment analysis.

pertains to the probability that the observed result occurred by chance alone; it commonly is expressed as a *P* value. Statistical significance depends on both the magnitude of benefit and the sample size. Trials with very large sample sizes theoretically can detect statistically significant but tiny benefits (*e.g.*, 2% relative reduction in ESRD incidence) that are of marginal or no clinical relevance. To avoid confusion, authors should apply the term *significant* to the results of statistical tests and use the term *clinically relevant* instead of clinically significant.

### Multicenter versus Single-Center Trials

A hallmark of contemporary trials is the use of several clinical centers that conduct the research. Advantages of multicenter studies are the potential for increased recruitment and generalizability from conduct of the trial in several regions of the country (or world). Other benefits are the extensive quality control and routine oversight needed to standardize procedures across centers and the contributions of multiple investigators with complementary expertise. The disadvantages of multicenter trials relate primarily to cost and to logistical aspects of coordinating research across multiple units. In contrast, single-center trials tend to be less expensive and cumbersome than multicenter trials, but the center must assume responsibility for all aspects of trial conduct, including quality control, data management, and data analysis. An example of a well-done, single-center trial is the study by Contreras *et al.* (20), who tested the effects of various maintenance therapies for lupus nephritis. However, even if single-center studies are well performed, their results often are perceived as less robust than those of multicenter trials.

### Conclusion

Well-designed and rigorously conducted trials provide extremely useful information that can inform, if not guide, clinical practice. The conduct of such research is characterized by numerous design, methodologic, and logistic issues. An understanding of these issues is useful for practitioners as they interpret the results of published research and consider the applicability of trial findings to their patients.

### References

1. Moher D, Schulz KF, Altman D; CONSORT Group (Consolidated Standards of Reporting Trials): The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 285: 1987–1991, 2001
2. Sacks FM, Svetkey LP, Vollmer WM, Appel LJ, Bray GA, Harsha D, Obarzanek E, Conlin PR, Miller ER 3rd, Simons-Morton DG, Karanja N, Lin PH; DASH-Sodium Collaborative Research Group: Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. *DASH-Sodium Collaborative Research Group. N Engl J Med* 344: 3–10, 2001
3. Klahr S, Levey AS, Beck GJ, Caggiula AW, Hunsicker L, Kusek JW, Striker G: The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. Modification of Diet in Renal Disease Study Group. *N Engl J Med* 330: 877–884, 1994
4. Wright JT Jr, Bakris G, Greene T, Agodoa LY, Appel LJ, Charleston J, Cheek D, Douglas-Baltimore JG, Gassman J, Glassock R, Hebert L, Jamerson K, Lewis J, Phillips RA, Toto RD, Middleton JP, Rostand SG; African American Study of Kidney Disease and Hypertension Study Group: Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: Results from the AASK trial. *JAMA* 288: 2421–2431, 2002
5. Eknoyan G, Beck GJ, Cheung AK, Daugirdas JT, Greene T, Kusek JW, Allon M, Bailey J, Delmez JA, Depner TA, Dwyer JT, Levey AS, Levin NW, Milford E, Ornt DB, Rocco MV, Schulman G, Schwab SJ, Teehan BP, Toto R; Hemodialysis (HEMO) Study Group: Effect of dialysis dose and membrane flux in maintenance hemodialysis. *N Engl J Med* 347: 2010–2019, 2002
6. Kaul S, Diamond GA: Good enough: A primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 145: 62–69, 2006
7. Black HR, Elliott WJ, Grandits G, Grambsch P, Lucente T, White WB, Neaton JD, Grimm RH Jr, Hansson L, La-courciere Y, Muller J, Sleight P, Weber MA, Williams G, Wittes J, Zanchetti A, Anders RJ; CONVINCe Research Group: Principal results of the Controlled Onset Verapamil Investigation of Cardiovascular End Points (CONVINCE) trial. *JAMA* 289: 2073–2082, 2003
8. Chalmers TC, Celano P, Sacks HS, Smith H Jr: Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 309: 1358–1361, 1983
9. Emanuel EJ, Miller FG: The ethics of placebo-controlled trials: A middle ground. *N Engl J Med* 345: 915–919, 2001
10. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the systolic hypertension in the elderly program (SHEP). SHEP cooperative research group. *JAMA* 265: 3255–3264, 1991
11. ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *JAMA* 288: 2981–2997, 2002
12. Sandhu S, Wiebe N, Fried LF, Tonelli M: Statins for improving renal outcomes: A meta-analysis. *J Am Soc Nephrol* 17: 2006–2016, 2006
13. Lagakos SW: The challenge of subgroup analyses: Reporting without distorting. *N Engl J Med* 354: 1667–1669, 2006
14. Fleming TR, DeMets DL: Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med* 125: 605–613, 1996
15. Rothwell PM: Factors that can affect the external validity of randomised controlled trials. Available: [www.plosclinicaltrials.org](http://www.plosclinicaltrials.org). Accessed May, 2006
16. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 303: 1038–1041, 1980
17. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey

- LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH, Karanja N: A clinical trial of the effects of dietary patterns on blood pressure. DASH collaborative research group. *N Engl J Med* 336: 1117–1124, 1997
18. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Stevens VJ, Vollmer WM, Lin PH, Svetkey LP, Stedman SW, Young DR; Writing Group of the PREMIER Collaborative Research Group: Effects of comprehensive lifestyle modification on blood pressure control: Main results of the PREMIER clinical trial. *JAMA* 289: 2083–2093, 2003
  19. Whelton PK, He J, Appel LJ, Cutler JA, Havas S, Kotchen TA, Roccella EJ, Stout R, Vallbona C, Winston MC, Karimbakas J; National High Blood Pressure Education Program Coordinating Committee: Primary prevention of hypertension: Clinical and public health advisory from the national high blood pressure education program. *JAMA* 288: 1882–1888, 2002
  20. Contreras G, Pardo V, Leclercq B, Lenz O, Tozman E, O’Nan P, Roth D: Sequential therapies for proliferative lupus nephritis. *N Engl J Med* 350: 971–980, 2004