4 Clinical Trials in Cardiovascular Medicine

SCOTT D. SOLOMON AND JOHN MCMURRAY

Modern cardiovascular medicine prides itself on being evidence based. Virtually all the therapeutic advances that have informed the treatment of patients with cardiovascular disease have resulted from the findings of randomized clinical trials. Randomized trials are generally considered to provide the highest level of evidence, and this principle is reflected in the approach that all major guidelines use to support the strength of therapeutic recommendations.[1] While clinical trials are used for evidence generation in virtually all disciplines, there are few fields in which clinical trials have been so impactful as in cardiovascular medicine.

This chapter will review the basic principles of clinical trials in cardiovascular medicine, the approach to designing and executing clinical trials, and an introduction to the interpretation of clinical trials results to inform clinical practice.

CLINICAL TRIALS VERSUS OTHER TYPES OF STUDIES

Observation has always been the key to the generation of medical evidence. For centuries, astute physicians have observed patients' responses to various remedies and occasionally made insightful inferences about the benefit of new treatments. Observational studies (see also Chapter 5) can assess the natural history of disease, demonstrate relationships between risk factors and outcomes, and generate hypotheses for more definitive experimental testing. Yet observational studies are almost always biased and limited when it comes to assessing the merits of new therapies. This single-greatest inherent problem with attempting to infer effects of therapies from observational studies is termed "confounding by indication" and refers to biases, known or unknown, that influence which therapies are used for which patients and which conditions. These biases can be overcome to some extent by taking account of, *or adjusting for*, all the other factors that might have influenced the decision to use that medication and the outcomes in those patients. Although several novel statistical methods have been developed to attenuate indication bias in observational studies,[2] adjustment is rarely able to overcome all the potential biases because all such factors cannot be known or accounted for. Indeed, many therapies that had initially been based on observational data, such as hormone replacement therapy in postmenopausal women to reduce cardiovascular risk,[3] have been refuted by subsequent randomized trials.

In contrast to observational studies, randomized clinical trials are prospective human experiments in which an intervention (which could be a pharmacologic or device therapy or an interventional strategy) are compared with a control and in which randomization is used to eliminate the potential biases related to administration of a therapy (Fig. 4.1). In a large enough study, randomization ensures that patients in both the experimental group and the control group are similar in every respect excepting the randomly allocated therapy. While single arm studies are sometimes referred to as trials, we will in this chapter limit our discussion to multiple arm studies in which treatment allocation is randomized.

CLINICAL TRIAL PHASES

Developmental programs for drugs and devices are categorized in phases (Table 4.1). *Phase I studies* assess the safety and tolerability in the first human experience of a novel therapy typically using healthy volunteers. These studies can be open label and even single arm and collect information that can be helpful in identifying a maximally tolerated dose (dose escalation studies).

Phase II studies are designed to confirm the biologic activity of the experimental therapy in patients with the disease of interest and, in some cases, to determine the likely optimal dose for both efficacy and tolerability. The results of these studies are typically used to determine whether to proceed to a *pivotal*, or *phase III trial*, which is used for regulatory assessment. Safety and tolerability are also assessed along with other secondary and exploratory measures of efficacy that might inform further development. Phase II trials often use *surrogate endpoints* rather than clinical endpoints (see later).

Phase III, or *pivotal* studies, are designed to provide enough information on efficacy and safety for regulatory evaluation and hopefully approval. Pivotal trials require assessment of "approvable" endpoints—that is, endpoints that have been previously agreed upon by regulatory authorities such as the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA). Approval of a new antihypertensive agent may require only demonstration of blood pressure lowering, and trials of cholesterol-lowering medication may require only demonstration of serum cholesterol lowering for approval. In contrast, other indications, such as for treatment of heart failure, may require demonstration of benefit for clinical outcomes, such as reducing death, hospitalizations for heart failure, or myocardial infarction. Phase III trials are sometimes performed for the primary purpose of determining safety for a therapy—a concern regarding cardiovascular safety for previously approved diabetes therapies prompted the FDA in 2008 to issue guidance requiring all diabetes registration programs to assess cardiovascular safety by assessing and adjudicating adverse cardiovascular events

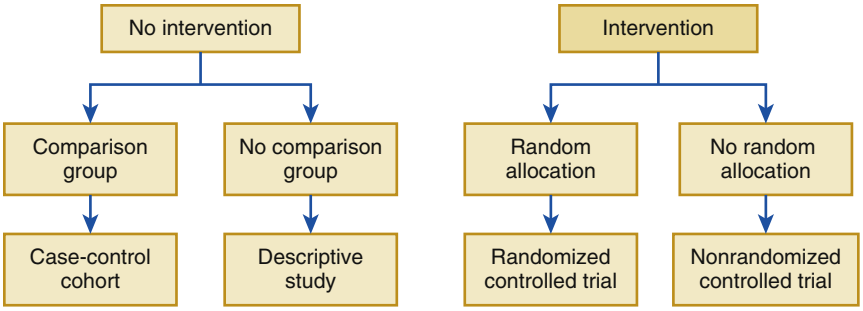


FIGURE 4.1 Types of clinical studies. Studies without intervention are considered case-control studies or descriptive studies, depending on whether or not they have a comparison group. Interventional trials can be either randomized or nonrandomized.

TABLE 4.1 Phases of Clinical Trials

PHASE	FEATURES	PURPOSE
I	First administration of new treatment	Safety and biologic plausibility
II	Early trial in patients with the disease to be studied	Efficacy—dose finding, adverse events, pathophysiologic insights
III	"Pivotal" trial large enough to test safety and efficacy	Designed to allow for regulatory approval
IV	Mechanistic, additional safety	Elucidate mechanisms, assess safety in novel populations, postmarketing surveillance

such as cardiovascular mortality, myocardial infarction, and stroke.[4] What is needed for registration is usually negotiated with regulatory authorities prior to initiation of a clinical trial. Guidance from health authorities regarding what is needed for registration has evolved over time. Recently, the FDA has indicated willingness to consider functional endpoints, such as 6-minute walk or patient-reported outcomes (PROs), for initial approval for a heart failure therapy.[5]

Phase IV trials, sometimes referred to as *post-marketing trials*, are designed to add mechanistic or other support for an indication, to extend a previous indication to a new population, or to meet a regulatory requirement such as providing additional safety information, perhaps in a specific patient population. The EVALUATE trial,[6] for example, was a phase IV trial examining the effect of sacubitril/valsartan compared with enalapril on aortic stiffness and ventricular remodeling to provide mechanistic support for the findings in PARADIGM-HF, a positive phase III outcomes trial.[7]

COMPONENTS OF CLINICAL TRIALS

Randomized clinical trials and clinical trial reports should include all the following components: *rationale, inclusion and exclusion criteria, study design, study execution, study endpoints, and analytic approach*. These components are typically codified in the study protocol, which serves as the principal documentation of the study background, objectives, design, organization, execution, and preliminary outline analysis plan.

Rationale and Study Background

Because all clinical trials are human experiments, they need to be justifiable to investigators, institutional review boards or ethics committees, and participants; a well-thought-out, clinically relevant, scientifically and ethically valid rationale is the essential first step in a clinical trial design. In short, the question should be one for which the answer is not known and for which the result would either directly inform clinical care or would provide crucial information that would inform the continued development of a particular therapy. The scientific rationale for conducting a trial can be in the form of basic research that supports a particular pathway or mechanism that may be affected by the therapy, preclinical data involving animal experiments in which a therapy was tested in a manner similar to a human trial, or early clinical or "pilot" studies that may provide some evidence that a therapy might be efficacious. Because virtually all interventional therapies can be associated with risk, the use of a particular therapy must at least have the potential to be beneficial in a particular disease state, although early-phase trials do not necessarily need to demonstrate that benefit to be successful.

Study Design

Clinical trial designs vary, and each have distinct advantages and disadvantages. The most commonly used is a *parallel group design* (Fig. 4.2A) in which patients are randomized to two or more groups and endpoints are compared between groups. These trials can be placebo controlled or active controlled and can have multiple arms (e.g., a placebo, active comparator and study drug, or multiple doses of a study drug). In this design, patients are randomized to receive one of these therapies for the duration of the trial. This type of design can be used for either clinical outcomes trials or phase II trials in which the primary endpoint is a surrogate (e.g., cholesterol or a natriuretic peptide).

In contrast, in *crossover trials* (Fig. 4.2B), patients receive one therapy for a period of time and then are "crossed over" to receive placebo or another therapy. In this design, individual patients act as their own control, and these designs are typically used for phase II studies in which the endpoint is a measured surrogate such as a biomarker. The advantage of crossover trials is that fewer patients are needed because each patient serves as his or her own control, reducing variability. The disadvantage is that effects of a therapy from the first phase can carry over and contaminate the second phase. This issue is typically mitigated with a *washout* period, a time between therapies during which the effect of the first phase would be expected to wear off. Crossover designs are not suitable for long-acting therapies or to outcomes trials (where a clinical outcome, such as a death or hospitalization, might influence whether the patient would join the second phase).

Factorial design (Fig. 4.2C) trials are essentially parallel group studies in which there are two consecutive randomizations within the same patient population so that an individual patient would be randomized to treatment A versus B, and also to C versus D, leaving four distinct treatment groups (A+C, A+D, B+C, B+D). In a factorial design trial, each randomization is essentially treated as its own trial. Factorial trials are best when the therapies are distinct enough that there will be no "interaction" between therapies. Assuming there is no or minimal interaction between therapies, factorial designs can be executed with a modest increase in the sample size required for a single intervention. If interaction between the two therapies is suspected, sample sizes need to be increased to allow for formal interaction testing. Examples of factorial design trials include the ISIS-2 trial,[8] which randomized patients to both streptokinase or placebo and additionally randomized the same patients to aspirin or placebo, and the DREAM trial, which compared the effects of ramipril versus placebo, and rosiglitazone versus placebo on the incidence of diabetes.[7,9]

Superiority trials (Fig. 4.3) test whether therapy A is superior to therapy B, which can be either an active comparator or placebo. Superiority trials aim to reject the *null hypothesis* that there is no difference between the therapies (see statistical considerations later). In contrast, *noninferiority trials* are designed to determine whether one therapy is *noninferior* to (loosely translated to *not worse than*) another therapy. In the case of noninferiority trials, rejecting the null hypothesis requires that therapy A be *not inferior* to therapy B within a certain margin of error; this requires setting a prespecified noninferiority margin and requiring an upper 95% confidence interval to be within that margin. Noninferiority trials are typically used when it is necessary to show only that a novel therapy is "as good as" an established therapy, which may be clinically important if the novel therapy has a better side effect profile, is less expensive, or may be easier to administer. Trials can be designed to test for both noninferiority and superiority, and a particular therapy can be noninferior even if not superior (see Fig. 4.3). The VALIANT trial[10] compared the angiotensin

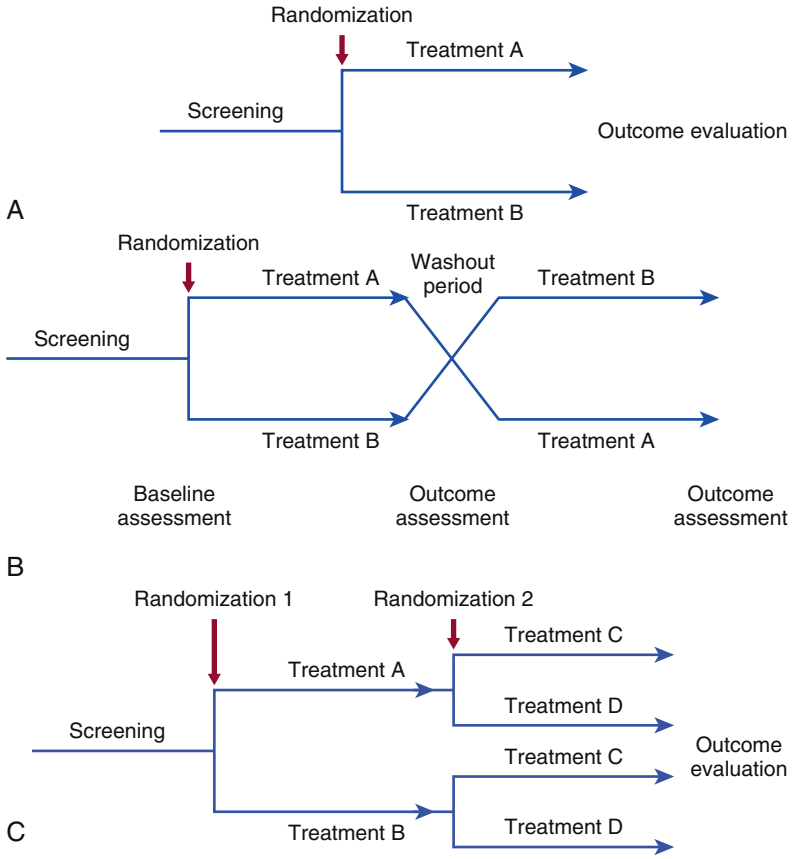


FIGURE 4.2 Clinical trial designs. Three types of clinical trial designs are illustrated. **A,** Parallel group design. **B,** Crossover design. **C,** Factorial design. Note that "treatment" can refer to an active treatment or placebo, and in factorial designs, two of four treatments could be placebo.

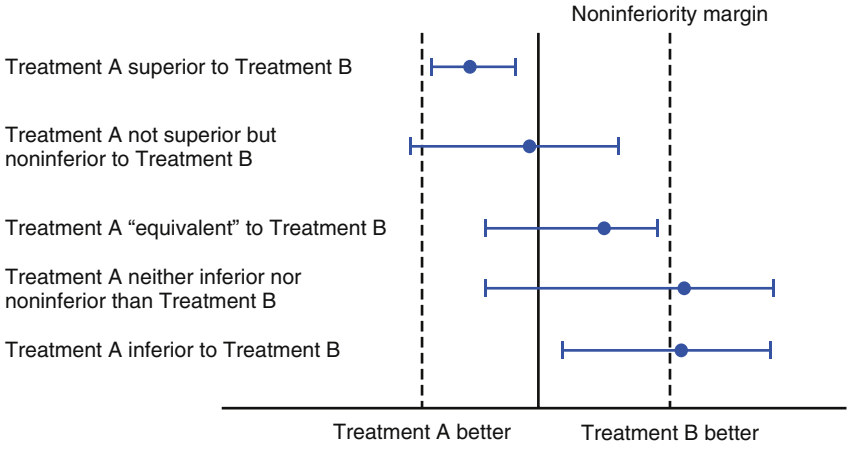


FIGURE 4.3 Superiority, noninferiority, and equivalence in clinical trials (see text).

receptor blocker (ARB) valsartan to the angiotensin-converting enzyme (ACE) inhibitor captopril (and the combination of the two) in post–myocardial infarction (MI) patients; while valsartan was not *superior* to the ACE inhibitor, it was *noninferior*, leading to an indication in post-MI patients.

Study Execution

Randomization

Randomization in a clinical trial can be as simple as a coin-toss (or the electronic equivalent) or considerably more complex, and there are a variety of approaches for randomly allocating treatment in clinical trials. Although randomization should lead to balanced groups in large trials, in smaller trials randomizing by a simple coin-toss (or random number) method can lead to imbalances at any time during the trial. For example, in a 100-patient trial, there would be a 5% risk of 60% of participants being allocated to one therapy. The commonly used *blocked* or *permuted block* randomization scheme mitigates this risk by ensuring equal number of participants assigned to each randomized group within each *block* of *x*, ensuring that the maximum imbalance at any given time is essentially the size of the permuted block.

While randomization in theory should lead to balanced groups in which characteristics known to be important in the disease being studied are balanced, in reality baseline imbalances are common in even relatively large trials and can influence results. For variables for which balance is especially desired, there are a variety of methods to *stratify* either at the randomization stage or the analysis stage. In stratified randomization, a participant is placed into a *stratum* (e.g., men or women) and then randomized within that stratum (ensuring balanced randomization within the stratum). This approach is more complicated than stratifying at the analysis stage, in which patients are *compared* within each stratum, an approach that is equally effective when trials are large.

Cluster randomization is a design in which groups of individuals, rather than actual individuals, are randomized. For example, trials testing specific strategies might randomly allocate clinics to one approach or another, as was done in the HOOPS trial, which randomly allocated 174 practices to pharmacist intervention or usual care to optimize use of guideline-directed therapy in patients with left ventricular dysfunction.[11] This avoids the risk of investigators applying the new strategy under investigation to "control" patients in the same clinic or practice because all patients in each practice or clinic receive one or other strategy. Cluster randomized trials require a slightly larger sample size than noncluster randomized trials.

Blinding (Masking)

Randomized trials can be blinded or unblinded, and blinded trials can be *single-blind*, *double-blind*, or *triple bind*. Blinding is designed primarily to avoid bias by allowing either participants or investigators to know which therapy a patient is on. In an open-label, or unblinded, trial, the participants and investigators will know which therapy is offered. The bias associated with this design can be mitigated using a blinded endpoint approach, often called a *prospective open-label blinded endpoint* (PROBE) design, in which the assessment is performed by individuals who are not aware of treatment assignment. An example of a PROBE design trial is one assessing cholesterol lowering where the laboratory making the cholesterol measurements was not aware of the treatment assignment and therefore could not be biased by this knowledge. It would, in contrast, be nearly impossible to eliminate bias for a patient-reported outcome in an unblinded clinical trial. Unblinded trials are less expensive and simpler to execute than blinded trials and have become an important approach to *pragmatic trial* design (see later).

A *single-blind* study is one in which the investigator, but not the participant, is aware of the study assignment, and it is also simpler to execute than fully blinded trials. If the investigator is involved, however, in collection of data and decisions about the care of the patient that might be influenced by his or her knowledge of the treatment assignment, the integrity of the trial could be compromised.

A *double-blind* study is one in which neither the participants nor the investigators are aware to which therapy a participant is assigned. Double-blind studies are considered the "gold standard" of clinical trial designs. Nevertheless, blinding can be difficult in practice, especially when investigators or patients may get "clues" about which therapy they have been assigned to (e.g., the taste of an experimental compound has unblinded participants to their therapy, and specific laboratory abnormalities, such as elevation in serum potassium, have the potential to unblind investigators). A *triple-blind* (i.e., triple-masking) study is a randomized experiment in which the treatment or intervention is unknown to (1) the research participant, (2) the individual(s) who administer the treatment or intervention, and (3) the individual(s) who assess the outcomes.

Blinding is typically accomplished by matching an experimental therapy to placebo. There are a variety of approaches used to ensure that experimental therapies are matched to placebo, including using dyes to ensure similarity in appearance of medication, overencapsulation, or various ingredients to mask taste. In device trials, there are several ways to accomplish blinding, although this is often impossible. While sham procedures can be performed, they are often impractical. Devices can be implanted but not turned on or can be programmed differently. It can even be challenging to blind the endpoints, because various diagnostic procedures (x-rays, ECGs) can unblind clinicians, investigators, and even endpoint adjudicators.

Inclusion and Exclusion Criteria

Properly defining the patient population is key to a successful clinical trial. Inclusion and exclusion criteria need to be tailored to ensure that the patients enrolled in the trial have the disease being studied and are likely to benefit from the therapy being tested if that therapy has actual benefit. For example, in a lipid-lowering therapy trial, for which the primary endpoint was degree of cholesterol lowering, patients would be required to have elevated levels of cholesterol at baseline. In a study of a similar therapy in which the primary endpoint was reduction in major adverse cardiovascular events (MACE, see later), patients enrolled need to be at risk for those events (e.g., assessment of MACEs in a primary prevention population of young adults might be impractical because of the very low event rate in that population). Often, enrichment criteria are used to ensure patients have sufficient risk—for example, in a heart failure outcomes trial, it is common to include a requirement for elevation in natriuretic peptides to ensure that patients have a high enough event rate. Exclusion criteria are based on ensuring patient safety; typical exclusions might include patients who are pregnant or may become pregnant during the course of the trial (pregnancy tests are often mandatory) if a therapy may be harmful to pregnancy or the fetus. Other exclusions might be specific to the therapy being tested. For example, a specific upper limit of serum potassium might be set when testing a drug that elevates serum potassium, such as a mineralocorticoid receptor antagonist (MRA) or renin-angiotensin system (RAS) inhibitor; alternatively, a lower blood pressure limit is typically used in heart failure trials testing drugs that tend to lower blood pressure, but this threshold may be much lower when testing an inotropic agent. Of note, specific inclusion and exclusion criteria limit the generalizability of a population—a common criticism of clinical trials, and often result in labels, guidelines, or payment decisions that reflect the specifics of those criteria.

Endpoints or "Response Variables"

Clinical trials are generally designed to evaluate both efficacy and safety. The metrics by which efficacy is assessed depend on the disease being studied, the mechanism of action of the therapy, and where the trial fits in the development lifecycle of the therapy. Measures of efficacy in cardiovascular medicine are numerous and include a variety of biomarkers including those that can be measured in the blood, such as cholesterol levels or natriuretic peptides, physical examination measures such as blood pressure or heart rate, or clinical outcomes, such as hospitalization (all cause or cause specific), or death (all cause or cause specific). Evaluation of efficacy requires a predetermined analysis plan with prespecified statistical approaches to determining whether a therapeutic benefit is met (see later). Measured endpoints—such as blood pressure—can be compared directly between treatment groups and may be measured at several time points during the course of a trial, although typically at baseline and at least once during follow-up. Usual analyses would include between-group comparisons adjusted for baseline level, although there are a variety of statistical techniques to handle multiple measures. These types of evaluation are particularly sensitive to subject drop-out leading to missing data, which can occur for a number of reasons, including subject death. For example, assessment of the effect of a drug on ejection fraction over time in a heart failure trial in which there is a high death rate, leading to many patients with missing data, can be problematic—especially if there is *differential* drop-out such that patients in one arm drop out at a greater rate, as might happen if the therapy resulted in fewer deaths, and the patients remaining alive in the placebo group were those whose ejection fraction were least likely to worsen. Such a scenario might lead to underestimation of a true treatment effect. *Surrogate endpoints* are measured endpoints that are thought to be directionally related to clinical outcomes and are often used in phase II trials. Good surrogate endpoints can usually be measured earlier than clinical outcomes, are indicative of disease progression, and are directionally related to the clinical outcome (changes in the surrogate endpoint correlate with clinical outcomes). Natriuretic peptides, for example, are often used as a surrogate in heart failure trials, and reduction in natriuretic peptides has been shown to correlate with improvement in clinical outcomes.[5] Although implanted devices have long recorded data that could be used as endpoints in trials (e.g., arrhythmia endpoints), novel endpoints from data acquired from wearable devices or smart phones are being used with greater frequency.

Clinical outcomes, such as death or hospitalization for heart failure, are typically counted and expressed as a *proportion* (i.e., percentage of patients dying over the course of the trial in each arm) or a *rate* (i.e., number of deaths per 100 patient-years). While clinical outcomes can be expressed as the proportion of patients who have an event at a certain time point (e.g., 30 days post randomization), this approach is best reserved for studies with relatively short-term outcomes. For longer outcomes trials, the *time to event* is usually incorporated by comparing the time from randomization to the event between treatment groups, thus accounting for the difference between a patient who died on the 30th day of a trial and a patient who died on the 300th day of a trial.

Clinical outcomes can be grouped into *composites* in which an "event" is said to occur if any of the several components of the composite occur, and the time to that event is based on the first occurrence of one of the component events. The designation MACE, or three-point MACE, is typically used to describe a composite of cardiovascular death, MI, or stroke. Similarly, a typical composite in heart failure trials is the combination of cardiovascular death or heart failure hospitalization[12] (or more recently, cardiovascular death or heart failure

hospitalization or urgent heart failure visit[13]). Including a fatal and nonfatal component in a composite addresses the issue of *competing risk*. Patients who die in a trial are clearly not at risk for a subsequent nonfatal event; thus, assessing only nonfatal events in trials where fatal events are likely can artificially deflate the risk of the nonfatal event in the group with a higher mortality rate, because this will likely deplete higher risk individuals. The number of composite events will not simply be the sum of all the component events for a given patient because only the first event that occurs is being counted. For example, a composite event of cardiovascular death or heart failure hospitalization would not count a death event if that event occurred after a heart failure hospitalization. Similarly, a second heart failure hospitalization would not be counted in that composite either (see later for alternative approaches that incorporate multiple events).

Patient-reported outcomes have become particularly important in cardiovascular trials because they provide meaningful insight into how specific therapies truly affect how patients feel and their quality of life. PROs are assessed through instruments (or questionnaires) that have been previously validated, although the type and extent of validation can vary. Examples of specific instruments typically used in cardiovascular medicine include the Kansas City Cardiomyopathy Questionnaire (KCCQ) and the Minnesota Living with Heart Failure (MLHF) instruments commonly used in heart failure studies, and the European Quality of Life Group-5D (EQ-5D) instrument, typically used for health economic assessment. Recently, the FDA has indicated that PROs may be considered *approvable endpoints* for certain conditions.[14,15]

STATISTICAL CONSIDERATIONS IN TRIALS

Analysis of Primary and Secondary Endpoints

Clinical trials are considered *hypothesis testing*, although that initially applies only to the primary question or endpoint of the trial. The statistical methods used are designed to determine if the null hypothesis can be rejected. The null hypothesis might be that a particular blood pressure–lowering medication does not affect blood pressure or that a heart failure medication does not impact cardiovascular death or heart failure hospitalizations. The analytic methods used depend on the specific questions being asked. Assessments of measured variables (e.g., blood pressure or ejection fraction) are most often comparisons of *between-group* differences in measures at baseline and follow-up, typically adjusting for baseline values, expressed as a between-group difference in the measure (with the confidence interval). Outcome measures are typically analyzed as time to event data using Kaplan-Meier curves and Cox proportional hazards models, and results are typically expressed as a hazard ratio with a 95% confidence interval. For example, a hazard ratio of 0.80 (0.65, 0.97) for a composite of cardiovascular death or heart failure hospitalization would indicate an estimated 20% reduced hazard of this endpoint in patients treated with the experimental therapy. However, the point estimate is only an estimate of benefit (or harm), and the confidence interval provides the range of potential true treatment effects that are consistent with the data; if the experiment were repeated an infinite number of times, the confidence interval represents the range that would contain the true treatment effect in 95% of those experiments. A narrow confidence interval (CI) implies a precise effect size, while a wide CI suggests greater uncertainty about the true effect of the treatment; in an outcomes study this can be due to an insufficient number of events. Based on traditional methods of significance testing, an upper confidence bound of 1.0 would signify a p-value of 0.05. Additional information about the efficacy of a therapy can be gleaned from the Kaplan-Meier curves directly. For example, some therapies might not show any evidence of benefit for a period of time, which is typical of studies that look at the effect of cholesterol-lowering therapies on MACEs, but for others, the curves might diverge immediately, suggesting very rapid benefit (Fig. 4.4).

The primary endpoint of a trial represents the primary hypothesis that is being tested. Trials usually have several *secondary* endpoints to answer additional questions and can further have *exploratory* endpoints. In registration trials, all or most of the statistical power (see later) is typically allocated to the primary endpoint such that it is the only hypothesis that can be tested at the $p = 0.05$ level. Some trials use co-primary endpoints, in which the statistical power is allocated to more than one endpoint (referred to as *splitting alpha*). Secondary endpoints are tested after the primary and typically are considered only *hypothesis testing* if the primary endpoint is positive, yet might still be considered *hypothesis generating* if not. Statistical power can be allocated to secondary endpoints in a number of ways. For example, they can be assessed hierarchically, such that if a primary endpoint is positive, a first secondary endpoint is tested; if this is positive, then a second secondary endpoint is tested and so forth. Alternatively, alpha can be "split" and allocated to multiple secondary endpoints (similar splits can occur for primary endpoints). Endpoints that are considered *exploratory* do not have any alpha allocated to them and are thus always considered *hypothesis generating* rather than *hypothesis testing*.

ALTERNATIVE METHODS OF ANALYSES

There are several limitations of standard time to first event analyses of clinical outcomes. First, the outcome statistic for a time to first event analyses using Cox proportional hazards models is expressed as a hazard ratio (point estimate and 95% confidence interval). Within each group, the hazard is a measure of the instantaneous rate of an event occurring. This model assumes that these group-level hazards remain *proportional (their ratio remains constant)* during the course of the trial (proportional hazards assumption). When this is not the case, and the hazard ratio varies over the course of the trial, the proportional hazards assumptions are violated, and this method may not be accurate (and may indeed underestimate a true treatment effect). Second, time to first event analysis of a composite outcome has the limitation of counting only the first outcome of a composite, which might be a less important component than another outcome, rendering the results less clinically meaningful. Several alternative methods have been developed to assess benefit in clinical trials to mitigate some of these limitations. Finally, the concept of a reduced hazard may be difficult for both clinicians and patients to understand because this metric contains no implicit information about the absolute magnitude of benefit, or absolute risk reduction, which is dependent on the event rate in the population.

Restricted Mean Survival Time

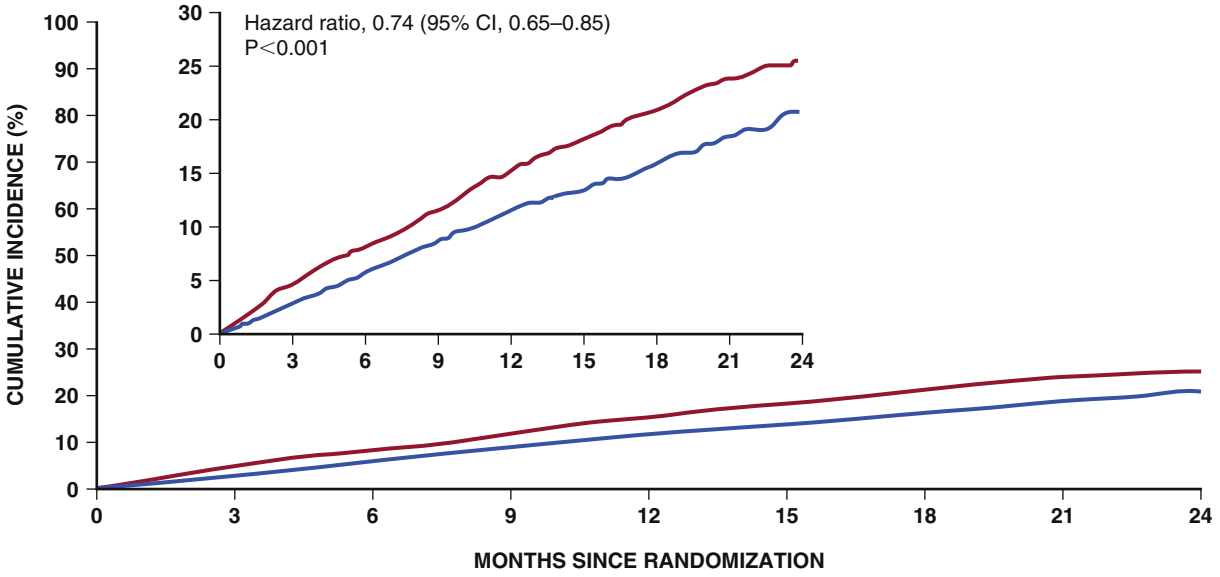
Whereas a hazard ratio represents the reduced (or increased) hazard associated with a therapy, this method is subject to the assumption that the hazard ratio be relatively uniform throughout the duration of the trial and is not a metric that is readily understandable. Restricted mean survival time is another approach that essentially compares the area under the curve of the two Kaplan-Meier survival curves and presents an average event-free survival for each treatment arm.[16] In this method, no assumptions are made regarding the proportionality of hazards over time. Because the mean observation time of individual trials may be relatively short compared with the residual life span of the patient, other methods using actuarial (age-based) approaches have been developed to project the lifetime benefit of therapies.[17]

Win Ratio

The win ratio is a relatively simple method in which patients in the treatment and control group are matched based on their risk profile.[18,19] A hierarchy of events is established, and the first event in the hierarchy (e.g., death) is compared between the matched patients, and a "winner" is determined. If there is no winner for that endpoint, then the approach is extended to the next component of the endpoint (e.g., heart failure hospitalization). The win ratio is calculated as the total number of winners divided by the total numbers of losers. An alternative form of this is the method described by Finkelstein and Schoenfeld in which all possible pairs are compared.[20] This method was used successfully in the ATTR-ACT trial comparing tafamidis and placebo in ATTR-amyloid heart disease.[21]

Recurrent Event Analysis

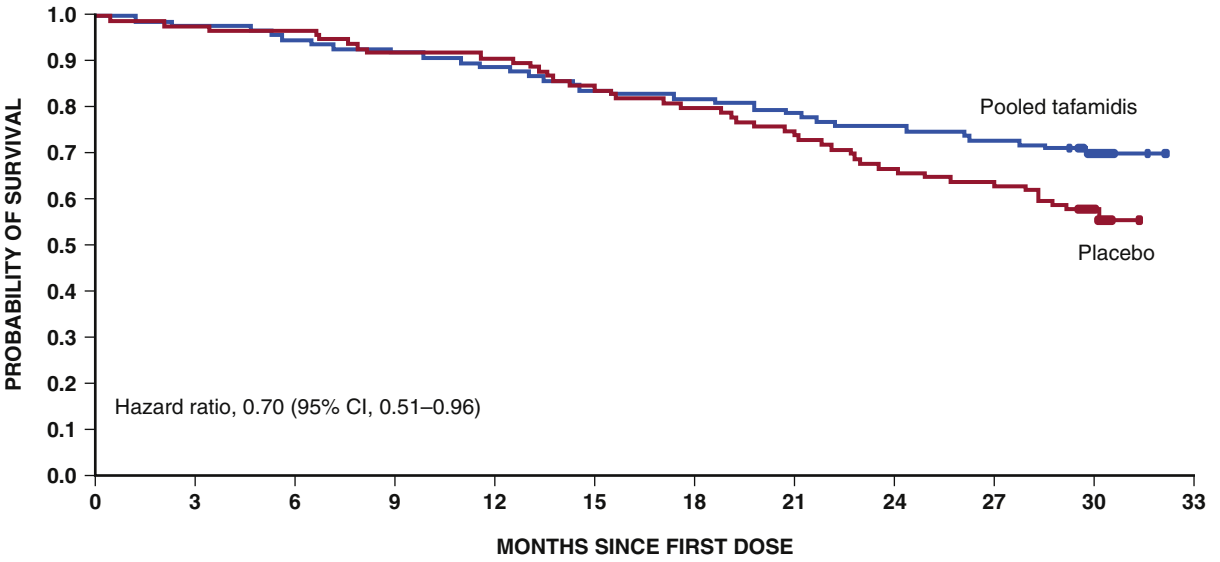
Time to first event analysis of clinical outcomes by definition ignores all outcomes that occur following the first event of a composite; a patient with a single heart failure hospitalization is counted similarly as a patient with multiple heart failure hospitalizations. Recurrent event analyses take into account not just the first event but the subsequent events and thus may more accurately assess the burden of disease in a patient in which multiple events are likely to occur. There are numerous approaches to recurrent event analyses, with a variety



No. at risk

Placebo	2371	2258	2163	2075	1917	1478	1096	593	210
Dapagliflozin	2373	2305	2221	2147	2002	1560	1146	612	210

A



No. at risk (cumulative no. of events)

Pooled tafamidis	264 (0)	259 (5)	252 (12)	244 (20)	235 (29)	222 (42)	216 (48)	209 (55)	200 (64)	193 (71)	99 (78)	0 (78)
Placebo	177 (0)	173 (4)	171 (6)	163 (14)	161 (16)	150 (27)	141 (36)	131 (46)	118 (59)	113 (64)	51 (75)	0 (76)

B

FIGURE 4.4 Kaplan-Meier curves. **A,** The *x* axis represents time since randomization, and the *y* axis represents cumulative incidence of events. Early divergence of treatment group (dapagliflozin) and placebo. **B,** The *x* axis represents time since initial dosage, and the *y* axis represents survival probability. Late divergence in the ATTR-ACT study, likely reflecting time required for treatment to be effective. The number of participants at risk at any time point is listed on the bottom for both treatment groups. (**A** from McMurray JJV, Solomon, SD, Inzucchi SE, et al. DAPA-HF Trial Committees and Investigators. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med*. 2019;381[21]:1995–2008. **B** from Maurer MS, Schwartz JH, Gundapaneni B, et al. ATTR-ACT Study Investigators. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med*. 2018;379[11]:1007–1016.)

of statistical methods. Rather than simply counting the number of events, which would fail to account for the fact that individual patient events are highly correlated (i.e., patients with one event are more likely to have multiple events) as well as differences in observation time, these events use *robust variance estimation* to account for the correlation between events. Post-hoc recurrent event analyses in the CHARM-Preserved trial suggested benefit comparing candesartan to placebo in heart failure with preserved ejection fraction,[22] and the recent PARAGON-HF used a recurrent event analysis as its primary between group comparison.[23]

Power and Sample Size

Prior to embarking on a trial, investigators determine an appropriate sample size based on several assumptions. For a study assessing a "measured" endpoint, such as left ventricular ejection fraction, a sample size can be calculated from knowledge of the expected standard deviation of the endpoint, and an expected effect size. Although the former can often be estimated from prior studies, the latter needs to be assumed on the basis of biologic plausibility and clinical relevance.

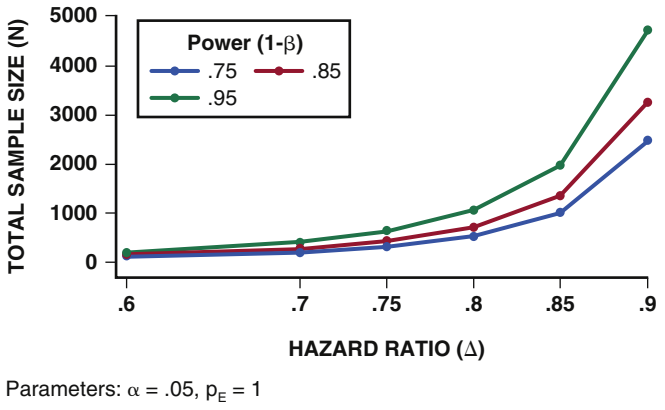


FIGURE 4.5 Relationship between hazard ratio, power, and sample size in a hypothetical clinical trial.

Another factor to consider is the desired *power,* which is defined as the probability that the test will reject a false null hypothesis—higher power increases the likelihood of success but requires greater sample size. Most cardiovascular trials are powered between 80% and 95%.

For outcomes trials, in which the primary endpoint might be time to all-cause mortality, or time to a composite of fatal and nonfatal endpoints, the analytic principles are similar regardless of the type of event or whether the event is an individual event or a composite. To determine the sample size for these types of trials, investigators need to make assumptions about the expected event rate and the expected treatment effect, as well as the desired power. The event rate can often be estimated from prior studies of similar populations. The effect size is never known beforehand, and a clinically meaningful effect size needs to be assumed. While a study can be powered to detect a very small effect size, this will require a larger sample size, which will also grow with a lower event rate (Fig. 4.5). An assumption of a large effect size (due to extreme confidence in a therapy) will lead to a small sample size but runs the risk of underpowering a trial.

Ultimately the power of an outcomes study is based not on the sample size but on the number of events. The event rates expected in outcomes trials determine the number of patients required to test a hypothesis. In diseases with very high event rates, such as advanced heart failure, effective therapies can demonstrate benefit with relatively small sample sizes; in contrast, for primary prevention, with relatively low event rates, very large sample sizes are usually required. Estimation of event rates can be difficult and is usually based on prior knowledge from previous trials or studies of similar populations. Nevertheless, these can be erroneous for a number of reasons (including due to improvement in standard of care from one trial to the next). Sample sizes can be reestimated during the course of the trial to adjust for this uncertainty (see *adaptive designs* later).

ETHICAL CONSIDERATIONS AND INFORMED CONSENT

Equipoise in Clinical Trials

There are several ethical considerations that need to be incorporated when conducting and interpreting the results of clinical trials. The primary ethical principle for conducting clinical trials is *equipoise*, which states that there is sufficient uncertainty about the value of the therapeutic interventions being tested in a trial that it is ethical to randomize a patient to any of the experimental arms.[24] Simply stated, a trial is ethical only if the trial question is worth asking and if the answer is not currently known. However, the determination about whether equipoise exists can often be quite subjective and will not always be agreed upon by individual investigators or the community at large. For example, individual clinicians may believe that one therapy is better than another although not all clinicians or even experts might agree. Many established therapies have not been subject to rigorous controlled trials, and clinicians may believe that they are nevertheless beneficial. Such therapies can be particularly difficult to test in randomized trials. When the TREAT trial was designed to assess the role of darbepoetin in reducing cardiovascular risk in patients with diabetes and chronic kidney disease, several major health authorities refused to participate because they believed that it would be unethical to deny patients this therapy, despite the fact that no rigorous assessment had been previously undertaken.[25]

Both physicians and patients have inherent biases that interfere with conduct of clinical trials. A clinician, investigator, or participant who does not believe that there is equipoise, but that one of the arms in a clinical trial is superior to the other, should not take part in a trial. This is why, in part, it can be so hard to enroll patients in trials involving surgical interventions where clinicians may believe strongly that specific procedures are beneficial. Although randomization ensures the lack of bias for the allocated treatment, it cannot remove intrinsic biases on the part of the investigators or participants which can impede the conduct of a trial by limiting the patients enrolled or the centers enrolling patients. These types of biases are particularly problematic in open-label trials, which, even when randomized, allow the investigators and participants to know which therapeutic arm they have been allocated to, which can lead to patients *crossing over* into or out of a therapeutic arm, a problem that plagued an open-label coxib safety trial in which undue concern about safety in one arm resulted in substantial crossovers.[26]

An interventional trial always has the potential to harm patients. This harm can be explicit, such as when a particular drug or therapy results in an adverse event, or implicit, such as when enrollment in a clinical trial prevents them from getting standard of care therapy that has already proven beneficial, or otherwise delays their access to care. These considerations need to be taken into account by both investigators and institutional review boards that need to approve all interventional studies.

Randomization to Placebo and Standard of Care

One common ethical question in the design of clinical trials is whether it is ethical in specific cases to randomize patients to placebo. If a condition has no proven therapy, then randomization to active therapy or placebo is usually considered ethical. If a therapy is already proven, then there are two potential approaches—to test the new therapy specifically in a head-to-head comparison with the old one, or to test the new therapy against placebo on top of *standard of care* therapy, which would typically include the old one. A head-to-head comparison requires sufficient rationale for believing the new therapy might be as good as the old one; denying those patients randomized to the new therapy from the previously accepted one does not itself present an ethical dilemma. Trials comparing ARBs to ACE inhibitors,[10,27] or sacubitril/valsartan to an ACE inhibitor in heart failure[7] are examples of active comparator trials in which a new active therapy was compared with a well-established therapy. In some cases, the goal of these trials is to prove noninferiority, which is ethically acceptable so long as there is some rationale other than commercial reasons that the second therapy may be superior in some way to the first (in the case of ARBs, there was both rationale that they would provide greater benefit than ACE inhibitors and they were known to have greater tolerability). Superiority trials comparing a novel therapy with a standard of care therapy are ethical so long as there is reasonable likelihood that the experimental therapy will be superior to, and minimal likelihood that the experimental therapy will be worse than, the standard therapy. Alternatively, studies that test novel therapies against placebo in conditions where standard of care therapies exist typically require patients to be on "optimal guideline-directed therapy." In DAPA-HF,[10] for example, which randomized patients to the sodium-glucose cotransporter-2 (SGLT-2) inhibitor dapagliflozin or placebo, participants were virtually all on ACE inhibitor/ARBs or angiotensin receptor neprilysin inhibitors (ARNIs), beta blockers, and a substantial number of patients were on MRAs—all considered standard of care background therapy in heart failure with reduced ejection fraction. Trials in which patients are not

getting optimal background therapy both raise ethical dilemmas and complicate the interpretation of the results.

Potential for Harm in Clinical Trials

While there is always potential for individual benefit to participants in clinical trials, randomization ensures that some patients will not receive the experimental therapy—if it is efficacious, they will be denied the benefit. This is especially true in early-phase trials, including those with normal volunteers. This arrangement is considered ethically acceptable because of the potential benefit to the community at large, and the potential dangers of administering unproven therapies. However, the extent to which an experimental question has sufficient equipoise may change during the course of a trial due to either external evidence (e.g., information to suggest that one arm is more efficacious than another), or because of data emerging from the trial itself as assessed by unblinded individuals on the data safety monitoring committee (see later). In both cases when the equipoise calculus changes, the original rationale for the trial may become moot, or the answer sufficiently known, that the trial would no longer be ethical to continue.

MONITORING OF DATA AND DATA SAFETY

Trials are subject to several types of monitoring to ensure the integrity of the data and the safety of the participants. During the execution of the study, the data-coordinating centers typically review the incoming data to assess data accuracy and quality. These can be as simple as range checks to exclude implausible values or can be more complex to catch potential fraud. Data entered into case report forms (CRFs) which can be either on paper or, increasingly, electronic are often verified with source documentation (source verification), although more and more trials are using *risk-based monitoring* which uses statistical methods to identify data inconsistencies and discrepancies.

Virtually all clinical outcomes trials (and many phase II trials) use an independent data safety monitoring board (DSMB) or data-monitoring committee (DMC) to review the incoming unblinded data for the purpose of ensuring participant safety, maintaining trial integrity, and determining if any factors external or internal to the trial affect the equipoise that presumably existed before the trial began. They use a combination of statistical methods designed specifically to assess both safety and efficacy during the course of a trial and clinical judgment. While their principal mission is to ensure safety of participants, they are also charged with determining whether the therapy under investigation demonstrates benefit sufficient to overturn equipoise and justify discontinuation of the control group, that is, discontinuation of the trial for efficacy. Data-monitoring boards are typically completely independent of the study execution and the sponsor, should have neither financial nor intellectual conflicts of interest, and should thus be unbiased regarding the results of the trial.

DSMBs can make recommendations to alter the conduct of a trial, although these should be done in ways to avoid unblinding investigators, or can recommend stopping trials entirely. Trials can be stopped for either safety concerns or efficacy, and often DSMBs are charged with determining if proceeding with a trial is futile. DSMBs follow specific guidelines for stopping for efficacy and futility, and trials typically have built-in interim analyses in which efficacy assessments are made. The robustness of a benefit, however, typically has to be higher earlier in a trial to avoid a type 1 error. Moreover, the more times the DSMB makes this determination, the greater the chance for type I error; thus greater stringency is required when stopping a trial early, and this can be codified by various stopping boundary approaches.[28–30]

NOVEL APPROACHES TO CLINICAL TRIAL DESIGN AND EXECUTION

Over the past 25 years, clinical trials have become larger, more complex, and more expensive. Typical industry-funded trials in cardiovascular medicine cost between $10,000 and $50,000 per enrolled patient. As such, there has been growing interest in finding ways to conduct trials more efficiently and less costly. Several innovations in trial methodology and execution are being used more frequently in cardiovascular trials.[31]

Pragmatic (Large Simple) Trials

As cardiovascular therapies have shifted from the sickest patients to patients with less severe disease, and effective therapies have lowered risk, sample sizes in clinical trials have grown to ensure the event rates necessary to test hypotheses in relatively low-risk individuals. *Pragmatic trials*, also called "large, simple" trials, are designed to evaluate the effectiveness of interventions in real-life routine practice conditions, whereas traditional trials aim to test whether an intervention works under optimal situations. Pragmatic trials produce results that can be generalized and applied in routine practice settings. Since most results from traditional trials fail to be broadly generalizable, the "pragmatic design" has gained momentum. Some questions are particularly suited to these types of approaches, which include simplification of inclusion and exclusion criteria, streamlining the amount of data obtained, reducing frequency of visits, and simplifying approaches to endpoint ascertainment and adjudication. Trials in primary prevention may be particularly suited to this approach because of the large number of patients needed and the simplified logistics of outpatient trials. One approach to simplifying trials is to remove the need for blinding of therapies (e.g., providing patients with a prescription that can be filled at a local pharmacy following randomization). While simpler to administer, open-label trials are more subject to potential differential drop-out and crossover as participants' perceptions about therapies evolve.[25]

Use of Electronic Medical Records in Clinical Trials

The use of electronic medical records (EMRs) for both identifying patients who fulfill inclusion and exclusion criteria to aid recruitment and for endpoint ascertainment is becoming more commonplace in clinical trials. EMR-based approaches can be used more effectively as large health care systems consolidate medical records into single large EMR systems. Utilization of EMR platforms to identify potential participants in clinical trials can be an extremely effective approach if inclusion and exclusion criteria are captured by EMR data fields. Unfortunately, many EMRs still rely on a substantial amount of free text, and while natural language approaches have been attempted to parameterize data stored in free text, these approaches remain in early stages. Use of EMRs for endpoint ascertainment is appealing but remains problematic because endpoints are captured only when patients interact with the healthcare system and many traditional clinical trials outcomes, including death, will be incompletely captured through the EMR. It is essential to remember that EMR systems were developed primarily for clinical care, not clinical research, and are often not optimized for collecting data in the rigorous manner required by trials. Nevertheless, convergence of clinical and research data collection is happening and will greatly optimize data collection in ways that can be leveraged by trials. Several recent trials, including the SCOT trial[25] and long-term follow-up of the WOSCOPS trial,[32] have successfully used EMR-based ascertainment of events.[33]

Adaptive Designs

Traditional clinical trials are designed based on assumptions about the patient population and the efficacy of the therapy being tested, and traditionally protocols are adhered to rigorously from design to completion with minimal changes occurring along the way. *Adaptive designs* are a way of mitigating the risk associated with potential incorrect assumptions made during the design phase of trials. Simple adaptive approaches might include sample size reestimation based on observed aggregate (*blinded*) event rates during the course of a trial that deviate from the expected event rates.[34] More complex adaptive approaches involve review of *unblinded* data to make adaptations in doses used, or alterations in inclusion/exclusion criteria that might identify patients

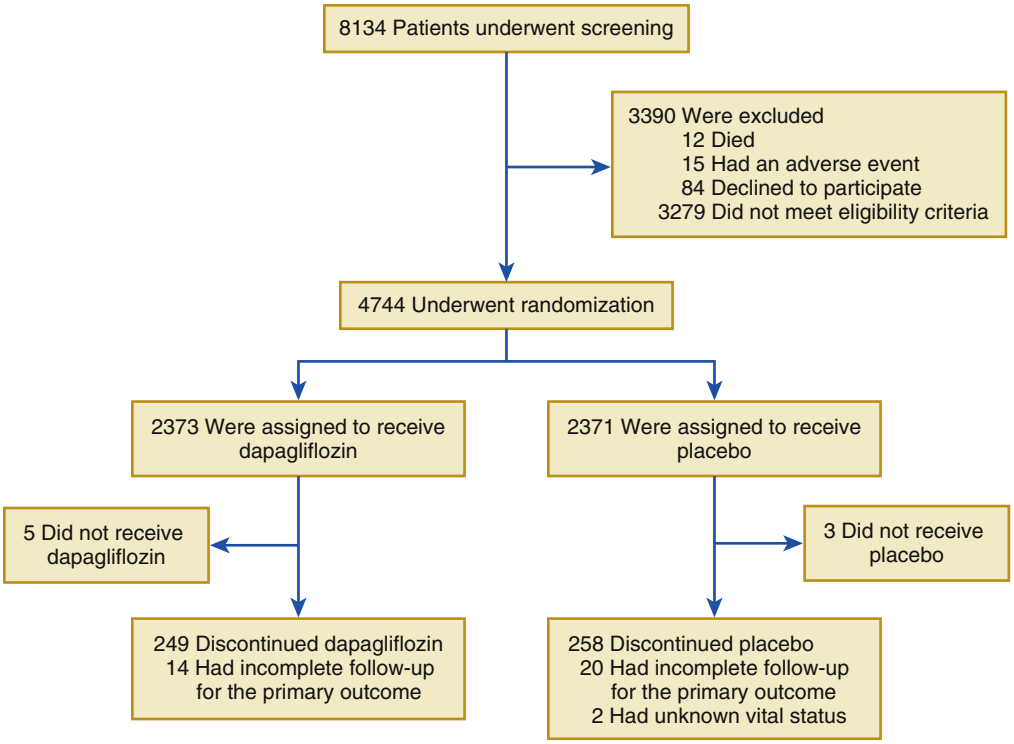


FIGURE 4.6 Typical consort diagram in a clinical trial. (From McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med*. 2019;381[21]:1995–2008.)

more likely to benefit from a therapy. These approaches require appropriate safeguards to ensure that maintenance of trial integrity and not unblinding or biasing investigators—typically handled by using an unblinded group—such as a DSMB—that is firewalled from the other individuals involved in the trial, or even by computer so that no one involved in the trial conduct, is aware of the adaptations.[35]

INTERPRETATION OF CLINICAL TRIALS

Not all involved in the practice of cardiovascular medicine will design or take part in clinical trials, but every practitioner will use the results of clinical trials in their clinical decision making. Thus, understanding the statistical principles underlying the results and limitations of the analyses and reporting of trials is essential to the modern practice of medicine.

Registration and Reporting of Clinical Trials

Since 2005 there has been a requirement that all clinical trials be registered on a clinical trials registration server such as ClinicalTrials.gov prior to enrollment of the first patient, with information about the trial design, inclusion/exclusion criteria, and primary and secondary endpoints listed and publicly available. This is to ensure that trials are conducted according to the protocol and that changes to the protocol are documented, and this is to mitigate publication bias, selective reporting of results. In addition to requiring registration prior to enrollment of the first patient, high-quality journals typically require strict adherence to the protocol in reporting the primary and secondary endpoints of trials and follow and enforce the CONSORT guidelines.[36] The CONSORT statement comprises a 25-item checklist focusing on the trial design, analysis, and interpretation and also suggests a comprehensive flow diagram displaying the disposition of all trial participants (Fig. 4.6). The presentation of the primary results of a clinical trial is relatively formulaic and includes at least a description of the study population usually broken down into the randomized treatment groups (usually the first table in an article), the primary and secondary endpoint results, the primary safety results, and description of the results for prespecified subgroups. All high-quality journals also require authors to report their financial conflicts of interests.

Understanding the Primary Results

The primary results of clinical trials, whether phase II, III, or IV, are subject to the basic principles of hypothesis testing in which the primary analysis determines whether the null hypothesis has been rejected or not, and subsequent analyses are dependent on that outcome. This principle is especially sacrosanct in studies being undertaken for regulatory approval.

In phase II trials, however, where the primary goal of the studies is to decide whether to proceed to more definitive trials, this principle can be more flexible, and investigators and sponsors often use the totality of the evidence rather than simply the results of a single primary endpoint to make decisions about next steps in the investigative plan. In some cases, this is done with the help of formal statistical approaches that allow for incorporation of multiple endpoints,[37] but in other cases this is done more informally.

For clinical outcomes trials, results are typically presented as a proportion of patients within each treatment group who achieve prespecified endpoints and, in most cases, a hazard ratio with a 95% confidence interval representing the results of a Cox model, with a Kaplan-Meier curve showing either event-free survival in the treatment groups or the accumulation of events (see Fig. 4.4). A hazard ratio of 0.80 with a 95% confidence interval of 0.72, 0.91 is interpreted as a 20% reduced hazard of the primary endpoint in the treatment group compared with the control group. The confidence interval suggests that the result could be as great as 38% reduced hazard or as little as a 9% reduced hazard. When the upper bound of the 95% confidence interval crosses 1, the result is considered no longer statistically significant at the 0.05 level. The point estimate is considered a measure of the magnitude of the result and the *p*-value (or how far the upper 95% confidence interval is from 1) a measure of the robustness of the result. Thus, a hazard ratio of 0.95 with confidence interval of 0.92 to 0.98 is statistically significant, although the magnitude of the benefit might be relatively small (and possibly clinically meaningless). Conversely, a study with a primary endpoint hazard ratio of 0.70 with confidence intervals between 0.46 and 0.98 needs to be interpreted with caution because the result suggests that the benefit might be as low as a 2% difference between therapies.

Secondary Endpoints

Secondary endpoints in trials are typically considered only hypothesis testing if the primary endpoint is significant and are otherwise considered hypothesis generating. While a secondary endpoint might be significant even if a primary endpoint is not, these results need to be interpreted with caution. This is especially true for results that were not expected and for which the study may not have been powered. In the ELITE trial, which compared losartan with captopril in just over 700 heart failure patients, the primary endpoint was increase in serum creatinine, which was not different between groups. However, there was a seemingly dramatic difference in all-cause mortality (17 vs. 32 deaths, hazard ratio 0.46, 95% confidence interval 0.05, 0.69; $p = 0.035$). The trial of course was not powered for all-cause mortality, and this result was not confirmed in the properly powered and much larger ELITE II trial.

Interpretation of Subgroups

Subgroup results in clinical trials are often subject to misinterpretation.[38] All outcomes trials, and many nonoutcomes trials, prespecify certain subgroups in which the data will be assessed—typical subgroups include sex, age (often dichotomized at a particular cut point), ejection fraction (often dichotomized at a particular cut point), diabetes status, etc. (Fig. 4.7). The primary reason for identifying and analyzing subgroups is to assess for *consistency* in the treatment response, not, as many believe, to assess for differences. The primary statistical analysis of subgroups is the test for *interaction* or *heterogeneity* between the subgroups with respect to the treatment effect. This asks the question, "Does the subgroup status modify the treatment effect?" Within subgroups, there may be differences between the point estimates and even the *p*-value for a result within a subgroup may be significant, but if the *interaction p*-value is NOT significant, we cannot state that the two

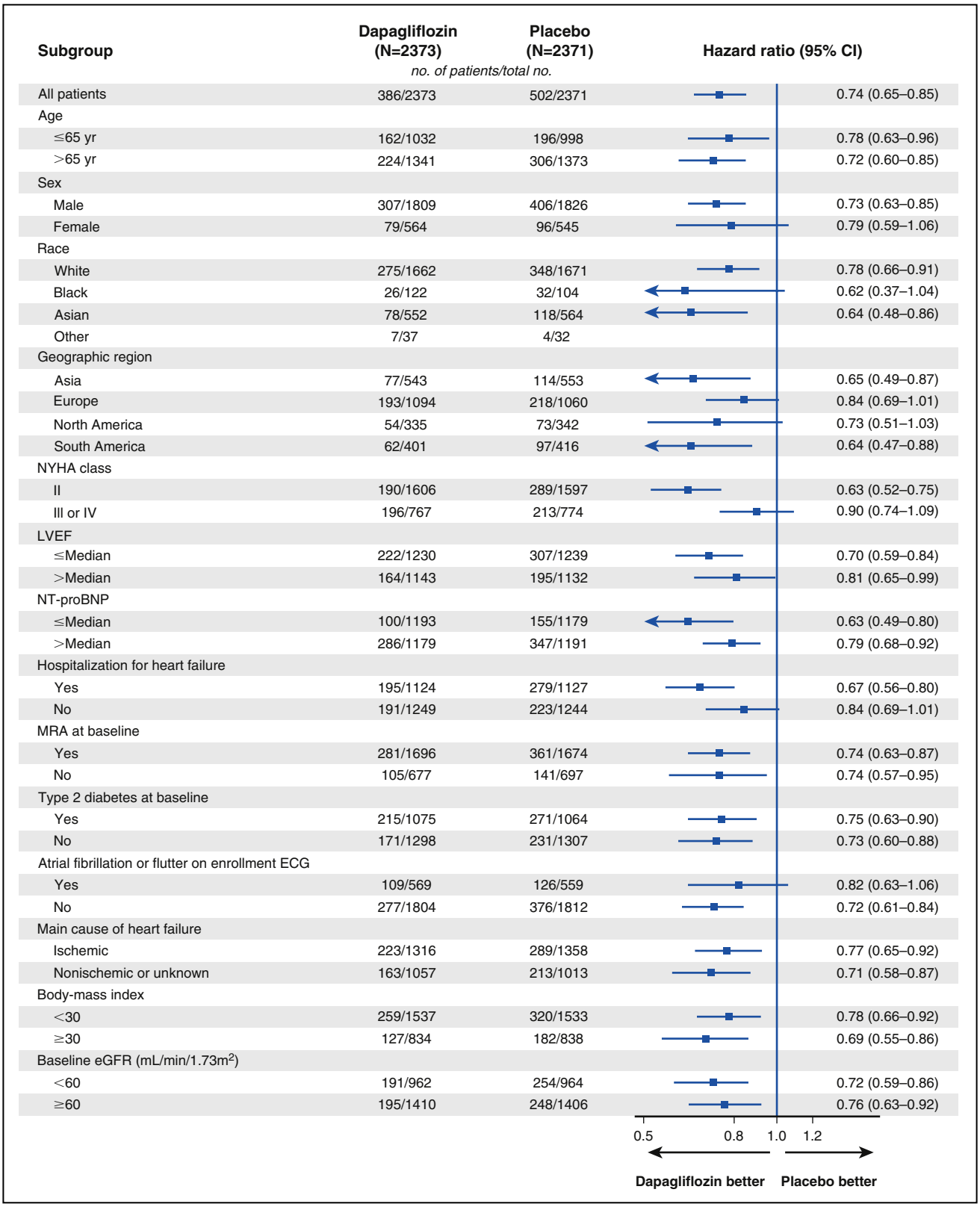


FIGURE 4.7 Subgroup "forest-plot" showing the point estimate and 95% confidence intervals within each prespecified subgroup in the DAPA-HF trial. (From McMurray JJV, Solomon SD, Inzucchi SE, et al. DAPA-HF Trial Committees and Investigators. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med*. 2019;381[21]:1995–2008.)

subgroups are different with respect to treatment. In general, the power to assess interactions is generally lower than the power to assess main effects. Moreover, the more subgroups we assess, the more likely it is to find a significant interaction. This concept was illustrated famously in the ISIS trial in which, as an example, they showed the primary results by astrologic sign and found that patients born under Gemini or Libra had an increased risk for mortality![8]

Subgroup findings can, however, identify individuals who may respond differentially to the treatment being studied in a trial, but the bar for accepting a subgroup finding as true should be quite high. Several factors increase the plausibility of a subgroup finding. The likelihood that a subgroup finding is real is increased if a subgroup is prespecified (not post-hoc); if the subgroup is large, because smaller subgroups are less reliable and always underpowered; if the subgroup is tested for interaction and the interaction *p*-value is adjusted for multiplicity; if the analysis is not just univariate but multivariate (accounting for the correlation between subgroups); if there is evidence from external data (even other trials) that the subgroup findings are plausible (even from adjacent populations); and if there is biologic plausibility—the subgroup finding makes sense. The differential findings for the patients above and below the median ejection fraction of 57% in the PARAGON-HF[21] trial were considered plausible because they fulfilled the aforementioned criteria.

Post-hoc Analyses

Post-hoc analyses should always be considered hypothesis generating. Nevertheless, some of the most valuable contributions of trials have been the results of post-hoc analyses. For example, the finding of reduced atherosclerotic events in the SAVE trial provided the rationale for the HOPE, PEACE, and EUROPA trials, which tested the hypotheses generated by the post-hoc analysis.

CONCLUSION

Although most cardiovascular practitioners will not design or execute trials, virtually all will use the results of trials to care for patients, and most of the evidence presented in this book is the product of clinical trials. A rudimentary understanding of trial methodology and statistical and analytic techniques is thus essential to the modern practice of cardiovascular medicine. The methodology of trials, from trial designs, methods of recruitment, ascertainment of endpoints, and statistical analysis continues to evolve, innovate, and become more efficient as this type of evidence generation becomes essential to bring new therapies to patients.

REFERENCES

1. Gibbons RJ, Smith S, Antman E. American College of Cardiology/American heart association clinical practice guidelines: Part I. Where do they come from? Circulation. 2003;107:2979–2986.
2. Lodi S, Phillips A, Lundgren J, et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. Am J Epidemiol. 2019;188(8):1569–1577.
3. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med. 2003;349(6):523–534.
4. Vijayakumar S, Vaduganathan M, Butler J. Glucose-lowering therapies and heart failure in type 2 diabetes mellitus: mechanistic links, clinical data, and future directions. Circulation. 2018;137(10):1060–1073.
5. Vaduganathan M, Claggett B, Packer M, et al. Natriuretic peptides as biomarkers of treatment response in clinical trials of heart failure. JACC Heart Fail. 2018;6(7):564–569.
6. Desai AS, Solomon SD, Shah AM, et al. Effect of sacubitril-valsartan vs enalapril on aortic stiffness in patients with heart failure and reduced ejection fraction: a randomized clinical trial. J Am Med Assoc. 2019;322(11):1–10.
7. DREAM (Diabetes REduction Assessment with ramipril and rosiglitazone Medication) Trial Investigators, Gerstein HC, Yusuf S, et al. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomised controlled trial [published correction appears in Lancet. 2006 Nov 18;368(9549):1770]. Lancet. 2006;368(9541):1096–1105.
8. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet. 1988;2(8607):349–360.
9. DREAM Trial Investigators, Bosch J, Yusuf S, et al. Effect of ramipril on the incidence of diabetes. N Engl J Med. 2006;355(15):1551–1562.
10. Pfeffer MA, McMurray JJ, Velazquez EJ, et al. Valsartan, captopril, or both in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both. N Engl J Med. 2003;349(20):1893–1906.
11. Lowrie R, Mair FS, Greenlaw N, et al. Pharmacist intervention in primary care to improve outcomes in patients with left ventricular systolic dysfunction. Eur Heart J. 2012;33(3):314–324.
12. McMurray JJ, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med. 2014;371(11):993–1004.
13. McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med. 2019;381(21):1995–2008.
14. https://www.fda.gov/regulatory-information/searchfda-guidance-documents/treatment-heart-failureendpoints-drug-development-guidance-industry. Accessed August 16th, 2020.
15. https://www.fda.gov/drugs/ddt-coa-000084-kansas-city-cardiomyopathy-questionnaire-kccq.
16. Kim DH, Uno H, Wei LJ. Restricted mean survival time as a measure to interpret clinical trial results. JAMA Cardiol. 2017;2(11):1179–1180.
17. Claggett B, Packer M, McMurray JJ, et al. Estimating the long-term treatment benefits of sacubitril-valsartan. N Engl J Med. 2015;373(23):2289–2290.
18. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J. 2012;33(2):176–182.
19. Ferreira JP, Jhund PS, Duarte K, et al. Use of the win ratio in cardiovascular trials. JACC Heart Fail. 2020;8(6):441–450.
20. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. Statist Med. 1999;18(11):1341–1354.
21. Maurer MS, Schwartz JH, Gundapaneni B, et al. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. N Engl J Med. 2018;379(11):1007–1016.
22. Rogers JK, Pocock SJ, McMurray JJ, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. Eur J Heart Fail. 2014;16(1):33–40.
23. Solomon SD, McMurray JJV, Anand IS, et al. Angiotensin-neprilysin inhibition in heart failure with preserved ejection fraction. N Engl J Med. 2019;381(17):1609–1620.
24. Emanual EJ, Wendler D, Grady C. What makes clinical research ethical? J Am Med Assoc. 2000;283:2701–2711.
25. Mix TC, Brenner RM, Cooper ME, et al. Rationale–Trial to Reduce Cardiovascular Events with Aranesp Therapy (TREAT): evolving the management of cardiovascular risk in patients with chronic kidney disease. Am Heart J. 2005;149(3):408–413.
26. MacDonald TM, Hawkey CJ, Ford I, et al. Randomized trial of switching from prescribed non-selective non-steroidal anti-inflammatory drugs to prescribed celecoxib: the Standard care vs. Celecoxib Outcome Trial (SCOT). Eur Heart J. 2017;38(23):1843–1850.
27. Pitt B, Poole-Wilson PA, Segal R, et al. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial–the Losartan Heart Failure Survival Study ELITE II. Lancet. 2000;355(9215):1582–1587.
28. Wittes J. Stopping a trial early - and then what? Clin Trials. 2012;9(6):714–720.
29. Pocock S, Wang D, Wilhelmsen L, Hennekens CH. The data monitoring experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) program. Am Heart J. 2005;149(5):939–943.
30. Fleming TR, DeMets DL. Monitoring of clinical trials: issues and recommendations. Control Clin Trials. 1993;14(3):183–197.
31. Solomon SD, Pfeffer MA. The future of clinical trials in cardiovascular medicine. Circulation. 2016;133(25):2662–2670.
32. Vallejo-Vaz AJ, Robertson M, Catapano AL, et al. Low-density lipoprotein cholesterol lowering for the primary prevention of cardiovascular disease among men with primary elevations of low- density lipoprotein cholesterol levels of 190 mg/dL or above: analyses from the WOSCOPS (West of Scotland Coronary Prevention Study) 5-year randomized trial and 20-year observational follow-up. Circulation. 2017;136(20):1878–1891.
33. Barry SJ, Dinnett E, Kean S, et al. Are routinely collected NHS administrative records suitable for endpoint identification in clinical trials? Evidence from the West of Scotland Coronary Prevention Study. PloS One. 2013;8(9):e75379.
34. Packer M, Holcomb R, Abraham WT, et al. Rationale for and design of the TRUE-AHF trial: the effects of ularitide on the short-term clinical course and long-term mortality of patients with acute heart failure. Eur J Heart Fail. 2017;19(5):673–681.
35. Bhatt D, Mehta C. Adaptive designs for clinical trials. N Engl J Med. 2016;375:65–74.
36. http://www.consort-statement.org.
37. Sun H, Davison BA, Cotter G, et al. Evaluating treatment efficacy by multiple end points in phase II acute heart failure clinical trials: analyzing data using a global method. Circ Heart Fail. 2012;5(6):742–749.
38. Wittes J. On looking at subgroups. Circulation. 2009;119(7):912–915.