CHAPTER 1

Introduction to randomized clinical trials in cardiovascular disease

Tobias Geisler, Marcus D. Flather, Deepak L. Bhatt, and Ralph B. D’Agostino, Sr

1University Hospital Tübingen, Tübingen Medical School, Tübingen, Germany
2University of East Anglia and Norfolk and Norwich University Hospital, Norwich, UK
3VA Boston Healthcare System; Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA
4Boston University, Boston, MA, USA

What is a randomized clinical trial?

The question “does it work” is common when a treatment is being considered for a patient. How do we know whether treatments “work” and what is the best way to demonstrate the efficacy and safety of new treatments? The main rationale behind a clinical trial is to perform a prospective evaluation of a new treatment in a rigorous and unbiased manner to provide reliable evidence of safety and efficacy. This is done by comparing the new treatment to a comparator or control treatment. Defining the term “clinical trial” is not as straightforward as it seems. In its simplest form, a clinical trial is any comparative evaluation of treatments involving human beings. Randomized clinical trials (RCTs) are the optimal means we use to achieve this demonstration. In this chapter we explore the relevance of RCTs to modern medicine and review strengths and weaknesses of this methodology (Table 1.1). As we will discuss below, RCTs represent the highest form of a clinical trial. Since the results of RCTs inform clinical practice guidelines, it is increasingly important for clinicians to understand their methodology, including their strengths and weaknesses. In this chapter we provide an overview of the main methodological aspects of well-designed RCTs.
The RCT is the most powerful design to prove whether or not there is a valid effect of a therapeutic intervention compared to a control. Randomization is a process of allocating treatments to groups of subjects using the play of chance. It is the mechanism that controls for factors except for the treatments, and allows comparison of the treatment under investigation with the control in an unbiased manner. It is important that information on the process of randomization is included in the trial protocol. The number of subjects allocated to each group, those who actually received the assigned treatment and reasons for non-compliance need to be recorded. In a representative analysis of trials listed in the free MEDLINE reference and abstract database at the United States National Library of Medicine (PubMed) in 2000, an adequate approach to random sequence generation was reported in only 21% of the trials [1]. This increased to 34% for a comparable cohort of PubMed-indexed trials in 2006 [2].

The procedure to assign interventions to trial participants is a critical aspect of clinical trial design. Randomization balances for known and unknown prognostic factors (covariates) allows the use of probability theory to express the likelihood that any difference in outcome between intervention groups merely reflects chance [3]. It facilitates blinding the identity of treatments to the investigators, participants, and evaluators, possibly by use of a placebo, which reduces bias after assignment of treatments [4]. Successful randomization is dependent on two related elements—generation of an unpredictable allocation sequence and concealment of that sequence until assignment takes place [5].

There are many procedures for randomization in the setting of a clinical trial and these will be discussed in detail below [see Study design (bias)]. For now we call attention to its importance in allowing the unbiased comparison of the investigational treatment and a control in a clinical trial.
Clinical trial phases

Preclinical studies
Preclinical studies of potentially useful treatments are usually carried out to understand mechanisms of action, effect of different doses, and possible unwanted effects. There are two main types of preclinical studies—those using whole animal models and those using components of living tissue, usually cells or organs. Preclinical studies help to build up hypotheses about how and why treatments may work. Most of these experiments are not randomized and there may be substantial reporting bias (i.e., only interesting results are reported), but they are an essential step in the development of new treatments.

Phase 1 clinical trials
The first step to evaluate the safety of a new drug or biological substance after successful experiments in animals is to evaluate how well it can be tolerated in a small number of individuals. This phase is intended to test the safety, tolerability, pharmacokinetics (PK), and pharmacodynamics (PD) of a drug. Although it does not strictly meet the definition criteria of a clinical trial, this phase is often termed a phase 1 clinical trial. Usually, if the drug has a tolerable toxicological profile, a small number of healthy volunteers are recruited. If the drug has an increased toxicological profile, often critically ill patients are included in whom standard, guideline-based therapy fails. The design of phase 1 clinical trial is usually simple. In general, drugs are tested at different doses to determine the maximum tolerated dose (MTD) before signs of toxicity occur. The most difficult challenge in the planning of phase 1 trials is finding ways to adequately translate the animal experimental data into a dosing scheme and not to exceed the maximum tolerated dose in humans. Phase 1 clinical trials are dose-ranging studies to identify a tolerable dose range that can be evaluated further for safety in phase 2 trials. There are different ways to adjust doses in a phase 1 clinical trial, e.g., single ascending and multiple ascending dosing schemes. Studies in apparently healthy human volunteers usually involve short exposure to new treatments to understand the effects of different doses on human physiology. Starting at low or subtherapeutic doses, especially with novel immunogenic agents, is essential to ensure that unexpected serious side effects are reduced.

Phase 2 clinical trials
Phase 2 clinical trials refer to the results of phase 1 trials. Once the maximum tolerated dose has been defined and an effective and tolerable dose range has been determined, phase 2 trials are designed to investigate how well a drug works in a larger set of patients (usually 100–600 subjects and sometimes up to 4000 patients, depending on the number of groups to be investigated) and to continue measurements of PK and PD in a more global population. Some
Phase 2 trials are designed as case series where selected patients all receive the drug or as randomized trials where candidate doses of a drug are tested against placebo. Usually, different doses of a pharmacological treatment will be compared against placebo in a randomized study design with outcomes based on the mechanistic action of the treatment being evaluated. For example, phase 2 trials of anticoagulants will usually document laboratory measures of anticoagulant effect, incidence of major and minor bleeding, and effects on relevant clinical outcomes. Minimizing risk to patients is essential as most treatments evaluated in phase 2 trials will never be approved for human use. Strategy-based treatments such as new methods for percutaneous coronary intervention (PCI) or surgical procedures also have their equivalent “phase 2” trials in which the new techniques are systematically tested in smaller number of patients to ensure safety and feasibility before being tested in larger trials. For obvious reasons these trials cannot be “placebo controlled,” but should compare the new strategy with an established one. Sometimes “phase 2” trials of treatment strategies are not randomized, which often makes it difficult to draw conclusions about safety and feasibility, and to plan further larger trials.

As an example, in the phase 2 trial Anti-Xa Therapy to Lower cardiovascular events in Addition to standard therapy in Subjects with Acute Coronary Syndrome–Thrombolysis in Myocardial Infarction 46 (ATLAS-1-TIMI 46 trial), the oral factor Xa inhibitor rivaroxaban was tested in several doses (5 mg, 10 mg, or 20 mg total daily dose, given either once or twice daily) in a total of 3491 patients with acute coronary syndromes (ACS) being treated with aspirin or aspirin and clopidogrel and compared with placebo. There was a dose-related increase in bleeding and a trend toward a reduction in ischemic events with the addition of rivaroxaban to antiplatelet therapy in patients with recent ACS. The researchers found that patients assigned to 2.5 mg and 5.0 mg twice-daily rivaroxaban in both the aspirin alone and aspirin plus clopidogrel groups had the most efficacious results versus placebo [6]. These results led to a selection of these dosing groups for transition into a large phase 3 trial that enrolled 15526 patients (ATLAS-2-TIMI-51) [7].

Phase 3 clinical trials
Phase 3 trials are usually RCTs, often multicenter, and including up to several thousand patients (the sample size depending upon the disease and medical condition being investigated). Due to the study size and duration, phase 3 trials are the most expensive, time-consuming, and complex trials to design and run, especially in therapies for chronic medical conditions, and are usually the “pivotal” trials for registration and marketing approval. Other possible motives for conducting phase 3 trials include plans to extend the label by the sponsor (i.e., to demonstrate the drug is effective for subgroups of patients/disease conditions beyond the use for which the drug was originally approved); to collect additional safety data; or to secure marketing claims for the drug. Trials at this stage are sometimes classified as “phase 3B trials” in contrast to “phase 3A trials,” denoting RCTs performed before marketing
approval [8]. Once a drug has proved acceptable in phase 3 trials, the trial results are usually combined into a large comprehensive document describing the methods and results of animal (preclinical) and human (clinical studies), manufacturing processes, product characteristics (e.g., formulation, shelf-life). This document serves as a “regulatory submission” to be reviewed by the appropriate regulatory authorities in different countries before providing approval to market the drug.

**Phase 4 clinical trials**
In phase 4 trials, post-marketing studies delineate additional information, including the drug’s risks, benefits, and optimal use. They also aim to see if a treatment or medication can be used in other circumstances beyond the originally approval indications. Phase 4 clinical trials are done after a treatment has gone through all the other phases and is already approved by the regulatory health authorities. Phase 4 clinical trials may not necessarily be RCTs. A large body of phase 4 trials is made up of registries and observational studies.

The following discussion about the methodology will mainly focus on phase 3 confirmatory RCTs.

**Study objective**
The search for new treatments is an evolutionary process, starting with a series of questions and eventually providing answers through a complex route that involves epidemiology (pattern and impact of disease in the population), basic science (cellular, mechanical, and genetic nature of the disease), and clinical trials to understand the response of patients to the new treatment. Trials that show clear benefits of treatments are usually followed by an assessment of cost and “affordability” to understand if the new treatment can actually be used in clinical practice. Some of these pathways are illustrated in Figure 1.1.

The quest to find effective and safe treatments arises from the needs of patients who present with illness and suffering. Thus, most clinical research is responsive in nature; we are not trying to improve on the healthy human but rather to treat and prevent illness and disease. However, in order to find an effective treatment, it is essential to understand the cause and pathology of the disease. Once specific causes are identified, whether they are protein deficiencies, transport errors, metabolic problems or genetic defects, it becomes possible to identify potential treatments that can then be tested in clinical trials. The challenge is that clinical trials take time and are costly to run, which means that they should be reserved for clinically important questions. Most clinical trials are set up and run by industry for commercial gain—often as industry/academic partnerships—but it should be emphasized that important health issues should be supported by the major healthcare providers, including governments and insurance agencies as part of their programs to
Cardiovascular Clinical Trials

improve health [1]. At present, most independent, non-commercial medical research is funded by competitive grants from governments or charities. While the competitive process helps to maintain high standards, it is an unpredictable method of funding and can lead to delays in carrying out important clinical trials. Lastly, well-intentioned but bureaucratic regulations applied to medical research are actually leading to substantial delays in important and effective treatments reaching patients in a timely manner. Thus, randomized trials are needed as the final pathway to test the hypothesis “Does it work?”. To answer this question reliably, large trials involving many patients from many centers are needed, which means that trial procedures including data collection and analysis need to be as simple and streamlined as possible [9,10].

Given all the above, when a specific phase 3 clinical trial is being designed, the first question is “What is the specific objective?”. For example, with the ATLAS-2 trial mentioned above, the objective was to establish the safety and effectiveness of rivaroxaban with both aspirin alone and aspirin and clopidogrel in reducing ischemic events in patients with ACS. The study objective must be explicitly stated in the study protocol (see below) and drives the study design, implementation, and analysis.

Study populations

The characteristics and features of the subjects to be enrolled in the clinical trial becomes the next issue and should be defined beforehand, using unequivocal inclusion (eligibility) criteria. A complete report of the eligibility criteria used to enrol the trial participants is required to assist readers in the interpretation of the study. In particular, a clear knowledge of these criteria
is needed to evaluate to whom the results of a trial apply, i.e., the trial’s generalizability (applicability) and importance for clinical or public health practice [11,12]. Since eligibility criteria are applied before randomization, they do not have an impact on the internal validity of a trial, but they are central to its external validity. It is important to differentiate between sample population and target population with regard to generalizability of results. The sample population is the population from which study subjects will be enrolled. The target population is the population to which the clinical trial results will be generalized. These are not necessarily the same. The eligibility criteria create a sample population that might significantly deviate from the target population. Thus, eligibility criteria should be kept as general and as realistic as possible. Ideally, study subjects should correspond to those to whom the product will be marketed. Demographic factors (age, gender, and race) and, when appropriate, socioeconomic status should be representatively covered. In addition, there is a sentiment that the study conditions should be realistic. For example, for over-the-counter drugs, regulatory authorities often require, before a drug is approved, the performance of clinical trials in settings similar to those in which the drug will actually be taken. These studies are called “actual use” studies.

Typical selection criteria include the nature and stage of the disease being studied, the exclusion of persons who may be harmed by the study treatment, and issues required to ensure that the study satisfies legal and ethical norms. Informed consent by study participants, for example, is a mandatory inclusion criterion in all clinical trials. The information about the number of patients being screened and meeting the eligibility criteria should be provided in flow diagrams (an example according to the CONSORT statement is shown in Figure 1.2).

Efficacy variables

Clinical trials can have numerous efficacy variables. However, it is essential that the primary efficacy variables should be kept to a minimum. The study objectives and efficacy variables should relate clearly and sharply to each other. Since large amounts of data can be collected and stored electronically, weighting their importance and relevance to the study objectives is crucial, and excess data collection is an important cause of poor trial performance. The primary efficacy variable should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. Ideally, there should only be one or a small number of primary variables. Multiple primary efficacy variables, however, are sometimes used in clinical trials with the hope of increasing the statistical power while keeping the sample size low. These can be counterproductive and increase the chance of producing inconclusive results. Careful consideration of how to deal with “multiple testing” or “alpha spending” is recommended [13,14]. The latter term describes how to distribute the type I or alpha error associated with testing the primary efficacy variables. Other efficacy
variables are classified as secondary and usually summarize variables that further support the primary variables and/or provide more information on the study objectives. Quality of life scales are an example of standard secondary efficacy variables in many clinical trials.

Remarkable effort has been made to solve the multiple testing problems associated with the primary variables. Exclusive testing of individual variables is one approach. The development of composite variables has been shown to be very helpful. These range from the combinations of endpoints, such as combining ischemic stroke, fatal and non-fatal coronary events, and hospitalizations in cardiovascular studies, to scoring scales developed by sophisticated psychometric techniques. Global assessment variables are also used to measure an overall composite.

Another issue of focus concerns the allocation of the alpha error to secondary variables, especially when the effects on the primary variables are not statistically significant [15–17]. For example, in a cardiovascular disease trial, how should the results be interpreted when the primary outcome variable (e.g., exercise testing or improvement of NYHA classification) is not signifi-

---

**Figure 1.2** Flow diagram showing the progress through different stages of a parallel randomized trial of two groups (i.e., enrolment, intervention allocation, follow-up, and data analysis). (According to http://www.consort-statement.org/consort-statement and Moher et al. [106].)
cant at the 0.05 level, but the significance level for a secondary variable related to overall mortality is highly significant at 0.001? [18]. It is hard to ignore such a finding when it refers to a hard clinical endpoint such as mortality. A prior allocation of alpha may need to be applied to major secondary endpoints. Future clinical trials in the same field should have the latter variables as the primary variables.

**Surrogate variables**
A surrogate endpoint is an intermediate endpoint that serves as a surrogate for a true endpoint if it can be used in lieu of the true endpoint to assess treatment benefit (i.e., reliable predictor of the clinical benefit). A surrogate variable should also be able to capture adverse effects. More specifically, it is a laboratory parameter or a physical sign used as a substitute for a clinically meaningful endpoint (e.g., measures of brain natriuretic peptide or 6-minute walking distance as surrogate for worsening heart failure; blood pressure or cholesterol levels as surrogates for coronary events; cardiac necrosis marker levels, Holter-detected ischemia, or microvascular obstruction detected on MRI as surrogates for severity of ischemic heart disease). As a surrogate variable usually represents an intermediate endpoint, it is obtained much sooner than the clinical endpoint of interest. It is usually much cheaper to obtain and has a more frequent incidence than the original endpoint. Surrogate variables have received increasing attention [19,20]. The challenge is to choose a surrogate variable that correlates strongly with the desired clinical endpoint. As an example, a commonly proposed intermediate surrogate variable for stroke is common carotid artery intima–media thickness (IMD) progression as measured by carotid ultrasound [21]. The progression of IMD occurs much earlier than stroke. The question is how well this relates to later development of the event. The value of measuring surrogate variables has been questioned, e.g., regulatory agencies claim that if the surrogate parameter has an effect on a “hard” clinical outcome (e.g., death or myocardial infarction), then the surrogate outcome should be a direct measurement of these. Additionally, history tells us that surrogate outcomes are not always related to the desired clinical outcome [25]. In the classic examples of the Cardiac Arrhythmia Pilot Study (CAPS) and the Cardiac Arrhythmia Suppression Trial (CAST), a combination of encainide/flecainide showed a reduction of the surrogate endpoint of ventricular extrasystoles and arrhythmias, but total mortality and arrhythmic deaths were significantly increased in the treatment arm [22,23]. More recently, in the Heart and Estrogen/Progestin Replacement Study (HERS), estrogen use in post-menopausal women with coronary disease was associated with a modest reduction in cholesterol, but this was not associated with any reduction in cardiovascular deaths or myocardial infarction [24]. Finally, in the Antihypertensive and Lipid-Lowering Treatment to prevent Heart Attack Trial (ALLHAT), of a total of 44,000 patients, 9,067 were randomized to doxazosin and 15,268 to chlorthalidone. Blood pressure was lowered by both treatments. However, treatment with doxazosin was significantly associated with a higher incidence of congestive heart failure, whereas chlorthalidone had
beneficial effects on heart failure incidence [25]. Analysis of the data suggests that chlorthalidone may have some beneficial effect beyond the blood pressure effect. If blood pressure reduction, a surrogate endpoint, had been the primary endpoint variable, this conclusion would not have been reached.

**Control groups**

In principle, there are two ways to show that a therapy is effective. One can demonstrate that a new therapy is better or roughly equivalent to a known effective treatment, or better than a placebo. In many RCTs, one group of patients is given an experimental drug or treatment, while the control group receives either a standard treatment for the illness or a placebo. Control groups in clinical trials can be defined using two different classifications: the type of treatment allocated and the method of determining who will be in the control group. The type of treatment can be categorized as followed: placebo or vehicle; no treatment; different dose or regimen from the study treatment, or different active treatment. The principal methods of creating a control group are by randomized allocation of a prospective control group or by selection of a control population separate from the investigated population (external or historical control) [26].

**Placebo-controlled trials**

A placebo-controlled trial is a way of testing a therapy against a separate control group receiving a sham “placebo” treatment, which is specifically designed to have no real pharmacological effect, and is a key strategy to reduce bias by avoiding knowledge of treatment allocation. Placebo treatment is usually a characteristic of blinded trials, where subjects and/or investigators do not know whether they are receiving a real or placebo treatment. The main purpose of the placebo group is to take account of the “placebo” effect, which consists of symptoms or signs that occur through the taking of a placebo treatment.

**Active-control trials**

In an active-control (also called positive-control) trial, subjects are randomly assigned to the test treatment or to an active-control drug. Such trials are usually double blind, but this is not always possible due to different treatment regimens, routes of administration, monitoring of drug effects, or obvious side effects. Active-control trials can have different objectives with respect to demonstrating efficacy.

The ability to conduct a placebo-controlled trial ethically in a given situation does not necessarily mean that placebo-controlled trials should be conducted when effective therapy exists. Patients and treating physicians might still favor a trial in which every participant receives an active treatment. Still, placebo-controlled trials are frequently needed to demonstrate the effectiveness of new treatments and often cannot be replaced by active-control trials that show that a new drug is equivalent or non-inferior to an established
agent. The limitations of active-control equivalence trials that are intended to show the effectiveness of a new drug have long been recognized [27–29], but are perhaps not as widely appreciated as they should be.

**Study design (bias)**

Bias can be loosely defined as “any influence that causes the results of a trial to deviate from the truth.” This broad definition implies that any element of study design or conduct (including analysis of results) could contribute to bias. In practice, we are particularly concerned about the method of randomization, compliance with treatment, systematic differences in concomitant treatments after randomization (especially in unblinded trials), completeness of follow-up, quality of data, and reporting of outcome measures. Systematic bias occurs when there is a difference in the treatment groups that does not occur by chance, and therefore the measurement of treatment effect may be unduly influenced. Systematic biases are mainly observed in non-randomized comparisons of treatment effects, such as those carried out in observational studies. Randomization, if performed correctly, can balance group differences and minimize systematic bias, to enable the quantification of the true effects of the interventions. Random allocation does not, however, protect RCTs against other types of bias.

**Methods of randomization**

Several methods exist to generate allocation sequences. Besides true random allocation, the sequence may be generated by the process of minimization, a non-random but generally acceptable method (see Table 1.2).

**Simple (unrestricted) randomization**

This method is the most basic of allocation approaches. Analogous to repeated fair coin-tossing, this method is associated with complete unpredictability of each intervention assignment. No other allocation generation approach, irrespective of its complexity and sophistication, surpasses the unpredictability and bias prevention of simple randomization.

**Restricted randomization**

Restricted randomization procedures control the probability of obtaining an allocation sequence with an undesirable sample size imbalance in the intervention groups. In other words, if researchers want treatment groups of equal sizes, they should use restricted randomization.

<table>
<thead>
<tr>
<th>Table 1.2 Methods of sequence generation [30].</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Simple (unrestricted) randomization</td>
</tr>
<tr>
<td>• Restricted randomization</td>
</tr>
<tr>
<td>• Stratified randomization</td>
</tr>
<tr>
<td>• Minimization</td>
</tr>
</tbody>
</table>
Stratified randomization
Randomization can create chance imbalances on baseline characteristics of treatment groups. Investigators sometimes avert imbalances by using prerandomization stratification on important prognostic factors, such as age or disease severity. In such instances, researchers should specify the method of restriction (usually blocking). To reap the benefits of stratification, investigators must use a form of restricted randomization to generate separate randomization schedules for stratified subsets of participants defined by the potentially important prognostic factors.

Minimization
Minimization is a dynamic randomization algorithm designed to reduce disparity between treatments by taking stratification factors into account. Important prognostic factors are identified before the trial starts and the assignment of a new subject to a treatment group is determined in order to minimize the differences between the groups regarding these stratification factors. In contrast to stratified randomization, minimization intends to minimize the total imbalance for all factors together, instead of considering only predefined subgroups [31]. Concerns over the use of minimization have focused on the fact that treatment assignments may be anticipated in some situations and on the impact on the analysis methods being used [32].

The practicality of randomization in a clinical trial can be complicated [33]. The conventional method is for a random number list to be generated by computer and a then treatment allocation list drawn up using the last digit (even or odd) to determine the treatment group. Patients entering the trial are then allocated according to the preprepared randomization list. It is essential that investigators do not have access to this list as they will of course then know the next allocation which can lead to a range of biases. Most trials use a method of central randomization using a telephone- or internet-based system for investigators to randomize patients. This method ensures that all patients are registered in the trial database and that prior knowledge of treatment allocation is not possible. Trials of double-blind pharmacological treatments (i.e., those in which the “active” and “placebo” treatments appear identical) have additional practical issues as the randomization list is used in the production and labeling process. Drug supplies must be provided to centers in “blocks” usually consisting of even amounts of active and placebo in identical packages, except for unique study identification numbers that can be used in emergencies to link the drug pack to the original randomization list for unblinding purposes.

The term “random” is often misused in the literature to describe trials in which non-random, deterministic allocation methods were applied, such as alternation or assignment based on date of birth, case record number, or date of presentation. These allocation techniques are sometimes referred to as “quasi-random.” A central weakness with all systematic methods is that concealing the allocation is usually impossible, which allows anticipation of intervention and biased assignments. The application of non-random methods in clinical trials likely yields biased results [4,34,35].
Readers cannot judge adequacy from terms such as “random allocation,” “randomization,” or “random” without further elaboration. Thus, investigators should clarify the method of sequence generation, such as a random-number table or a computerized random number generator.

In some trials, participants are intentionally allocated in unequal numbers to each intervention and control: e.g., to gain more experience with a new procedure or to limit the size and costs of the trial. In such cases, the randomization ratio (e.g., 2:1 or two treatment participants per each control participant) is reported.

**Random and systematic error**

When the clinical trial results are produced, the differences observed between treatments may represent true outcome differences. However, it is essential that the investigator (and the reader) consider the chance that the observed effects are due to either random error or systematic error. Random error is the result of either biological or measurement variation, whereas systematic error is the result of a variety of biases that can affect the results of a trial (Table 1.3). The process of analyzing the outcomes of a study for random error includes both estimation and statistical testing. Estimates describing the distribution of measured parameters may include point estimates (such as means or proportions) and measures of precision (such as confidence intervals).

**Study design issues to overcome systematic bias**

As stated above, the most important design techniques to overcome bias in clinical trials are blinding and randomization. Most trials follow a double-blind approach in which treatments are prepacked in accordance with a suitable randomization schedule, and supplied to the trial center(s) labeled only with the subject number and the treatment period: no one involved in the

<table>
<thead>
<tr>
<th>Table 1.3</th>
<th>Potential sources of systematic bias at different stages in the course of a trial.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Planning phase</strong></td>
<td></td>
</tr>
<tr>
<td>• Choice of research question</td>
<td></td>
</tr>
<tr>
<td>• Type of research study</td>
<td></td>
</tr>
<tr>
<td><strong>Recruitment phase</strong></td>
<td></td>
</tr>
<tr>
<td>• Allocation of participants to study groups</td>
<td></td>
</tr>
<tr>
<td>• Selection bias (eligible individuals are excluded, because the investigator knows the allocation to treatment group)</td>
<td></td>
</tr>
<tr>
<td>• Delivery of interventions</td>
<td></td>
</tr>
<tr>
<td>• Measurement of outcomes</td>
<td></td>
</tr>
<tr>
<td><strong>Post-recruitment phase</strong></td>
<td></td>
</tr>
<tr>
<td>• Loss to follow-up</td>
<td></td>
</tr>
<tr>
<td>• Analysis</td>
<td></td>
</tr>
<tr>
<td>• Dissemination of results</td>
<td></td>
</tr>
<tr>
<td>• Interpretation of the results by the study group or external persons (e.g., reviewer)</td>
<td></td>
</tr>
</tbody>
</table>
Cardiovascular Clinical Trials

conduct of the trial is aware of the specific treatment allocated to any particular subject, not even as a code letter. Bias can also be reduced at the design stage by specifying procedures in the protocol aimed at minimizing any anticipated irregularities in trial conduct that might impair a satisfactory analysis, including various types of protocol violations, withdrawals, and missing values. The study design should consider ways both to minimize the frequency of such problems, and also to handle the problems that do occur in the analysis of data.

Blinding
Blinding or masking is used in clinical trials to curtail the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial, caused by the impact that the insight into treatment may have on the enrolment and allocation of subjects, their subsequent care, the compliance of subjects with the treatments, the evaluation of endpoints, the handling of drop-outs, the analysis of data, etc.

A double-blind trial is a trial in which neither the investigator nor the study participant or sponsor who is involved in the treatment or investigation of the subjects is aware of the treatment received. This includes anyone who evaluates eligibility criteria or analyses endpoints, or assesses protocol. The principle of blinding is maintained throughout the whole course of the trial, and only when the data are cleaned to an appropriate level, can particular personnel can be unblinded. If unblinding to the allocation code to any staff who are not involved in the treatment or clinical evaluation of the subjects is required (e.g., bioanalytical scientists, auditors, those involved in serious adverse event reporting), adequate standard operating procedures should exist to guard against inappropriate publication of treatment codes. In a single-blind trial, the investigator and/or his/her staff are conscious of the treatment but the subject is not, or vice versa. In an open-label trial, the identity of treatment is known to all participants/study personal. Double-blind trials are the optimal approach, but are associated with greater complexity in providing placebo and the process of drug supply and packaging.

Difficulties in pursuing a double-blind design can be caused by: the different nature of treatments, e.g., surgery compared to drug therapy, or comparison of different drug formulations (e.g., an oral drug compared to an intravenous one). Additionally, the daily pattern of administration of two treatments and the method used to monitor pharmacological effects may differ. A possible way of achieving double-blind conditions despite these circumstances is to apply a “double-dummy” technique. This technique may sometimes imply an administration scheme that is unusual and thus adversely influences the motivation and compliance of the subjects. Ethical difficulties may also arise, e.g., if dummy operative procedures are performed. Nevertheless, it is recommended to make extensive efforts to implement methods to maximize blinding. The double-blind nature of some clinical trials may be jeopardized by obvious treatment-induced effects. In these cases, blinding
may be improved by blinding investigators and relevant sponsor staff to particular test results (e.g., selected clinical laboratory measures). If a double-blind trial is not possible, then the single-blind option should be considered. In some cases, only an open-label trial is practically or ethically possible, or cost constraints preclude producing and packaging a placebo. Consideration should be given to the use of a centralized randomization method, such as telephone- or internet-based randomization, to administer the assignment of randomized treatment and to ensure that all patients are registered in the trials. Furthermore, clinical assessments should be made by medical staff who are not involved in the treatment of the subjects and who remain blinded to treatment. In single-blind or open-label trials, every effort should be undertaken to minimize the various known sources of bias and primary variables should be as objective as possible. The reasons for the degree of blinding should be explained in the protocol, together with actions taken to reduce bias by other means. The PROBE (prospective, randomized, open-label, blinded endpoint) was developed to adopt a more “real-world” principal. By using open-label therapy, the drug intervention and its comparator can be clinically titrated, as would occur in every day clinical practice. Blinding is maintained for the outcome assessment. In a meta-analysis of PROBE trials and double-blind trials in hypertension [36], changes in mean ambulatory blood pressure from double-blind controlled studies and PROBE trials were statistically equivalent; however, the impact of the PROBE design on clinical trial design is still being evaluated.

Unblinding of a single subject should be considered only when knowledge of the treatment assignment is necessary to provide information to the subject’s physician for further therapeutic actions. Any unintended breaking of the blinding should be reported and explained at the end of the trial, irrespective of the reason for its occurrence. The procedure for and timing of unmasking the treatment allocations should be documented.

**Study design (samples)**

Major study designs in RCTs are:

- **Parallel group design**: each study subject is randomly assigned to a treatment or an intervention
- **Crossover design**: within a certain period of time each study subject receives all study treatments in a random sequence (possibly separated by a washout period in case of delayed offset of the study drug action)
- **Factorial**: each study subject is randomly assigned to a fixed combination of treatment (e.g., 2 x 2 factorial design: study drug A + study drug B, study drug A + placebo B, placebo A + study drug B, placebo A + placebo B).

The parallel group design is the preferred design in RCTs with two treatment arms. In a representative analysis of published RCTs, the parallel group design was the most frequently chosen design—more than two-thirds of trials [37]. In case of more than one treatment arm, the parallel group design requires a larger sample size and does not allow for investigation of effects
and interactions of study drug combinations of interest; a factorial design might be a good choice of study design to answer this question. A crossover design may be considered as it may yield more efficient comparison of treatments, e.g., fewer patients required for the same statistical power since every patient serves as his/her own control. However, there are problems with crossover designs in clinical outcome trials because the effects of treatment B are dependent on treatment A, meaning that if treatment A heals the patients or prevents cardiovascular events then treatment B might not have the opportunity to show its effectiveness or the prognostic effects may not be specifically attributable to treatment B. Crossover designs are mainly used for assessing responses to treatment, e.g., blood pressure, blood values or exercise capacity.

Besides the adequate choice of study design to avoid bias, careful selection of sample composition, types of control, and sequence of different treatments (or exposures) for samples are essential to ensure the quality of a clinical trial. In detail, this includes:

- Recruitment, patient population studied, and number of patients to be included
- Eligibility (inclusion and exclusion)
- Measurements of treatment compliance
- Prophylaxis at baseline
- Administration of treatment(s) (specific drugs, doses, and procedures)
- Level and method of blinding/masking (e.g., open, double-blind, single-blind, blinded evaluators, and unblinded patients and/or investigators)
- Type of control(s) (e.g., placebo, no treatment, active drug, dose-response, historical) and study configuration (parallel, crossover, factorial design)
- Method of assignment to treatment (randomization, stratification)
- Sequence and duration of all study periods, including prerandomization and post-treatment periods, baseline periods, therapy withdrawal/washout periods, and single and double-blind treatment periods. When patients were randomized should be specified. It is usually helpful to display the design graphically with a flow chart that includes timing of assessments
- Any safety, data monitoring, or special steering or evaluation committees
- Any interim analyses.

In the past, many clinical trials were restricted to two treatments only, and the choice between parallel samples or a crossover study design was the major decision. In most cases, a parallel-group design was chosen in most RCTs. Nowadays, there is an increasing trend toward using factorial approaches that may allow more than one major question to be answered. For example, when comparing the effects of two antihypertensive treatments in those who also have cholesterol problems, a comparison of the effect of lipid-lowering drugs could also be performed. Accurate use of a factorial design allows for independent assessment of both of these comparisons. Additionally, clinical trials are increasingly designed as large multicenter and often multinational studies to ensure generalizability, and also, for regulatory issues, to justify the need for only one study for approval.
Comparisons

Trials to show superiority
Scientifically, efficacy is established by demonstrating superiority to placebo in a placebo-controlled trial, by demonstrating superiority to an active-control treatment or by proving a dose–response relationship. This type of trial is referred to as a “superiority” trial. When a therapeutic treatment that has been shown to be efficacious in superiority trial(s) exists for treatment of serious illnesses, a placebo-controlled trial may be considered unethical. In that case, the scientifically sound use of an active treatment as a control should be considered. The appropriateness of placebo control versus active control should be considered on a trial-by-trial basis.

Trials to show equivalence or non-inferiority
This type of trial design might be the preferred strategy of the sponsor when there is the suspicion that an experimental treatment is not superior in terms of efficacy but may offer safety or compliance advantages compared to the active control [38]. According to its objective, two major types of trial are described: “equivalence” trials and “non-inferiority” trials. Bioequivalence trials belong to the first category. Sometimes, clinical equivalence trials are also undertaken for the purpose of other regulatory issues, such as proving the clinical equivalence of a generic product to the marketed product. In a non-inferiority trial, putative placebo comparisons are essential:
- (Historical) Effect of control drug versus placebo is of a specified size and there is a belief that this would be maintained in the present study if the placebo were included as a treatment.
- The trial has the ability to recognize when the test drug is inferior to the control drug.
- There is sufficient belief that the test drug would be superior to a placebo by a specified amount.

Many active-control trials are designed to show that the efficacy of an investigational product is not worse than that of the active comparator. Another possibility is a trial in which various doses of the investigational drug are compared with the recommended dose. Active-control equivalence or non-inferiority trials may also incorporate a placebo treatment arm, thus pursuing multiple goals in one trial; e.g., they may establish superiority to placebo and hence simultaneously validate the trial design and evaluate the degree of similarity of efficacy and safety to the active comparator. There are well-known difficulties connected with the use of an active-control equivalence (or non-inferiority) trial that does not include a placebo or does not incorporate multiple doses of the new drug. These relate to the inherent lack of any measure of internal validity (in contrast to superiority trials), thus making external validation necessary. A particularly important issue is establishing a credible non-inferiority “margin” to decide the usefulness of the new treatment and estimate the sample size, which should be discussed with a
statistician. Equivalence (or non-inferiority) trials are not robust in nature, making them particularly susceptible to flaws in the design of a trial or its conduct, thus leading to biased results and the conclusion of equivalence. For these reasons, the design aspects of non-inferiority trials deserve particular recognition and their conduct needs special attention. For example, it is especially important to minimize the incidence of violations of the entry criteria, non-compliance, withdrawals, losses to follow-up, missing data, and other deviations from the protocol, and also to reduce their impact on subsequent analyses. Active comparators should be carefully chosen. A suitable active comparator would be a widely applied therapy whose efficacy for the same indication has been clearly established and measured in well-designed and well-reported superiority trial(s), and which can be reliably anticipated to exhibit similar efficacy in the planned active-control trial. As a consequence, the new trial should have the same important features (primary variables, dose of the active comparator, eligibility criteria, etc.) as the previously conducted superiority trials in which the active comparator clearly demonstrated clinically relevant efficacy, taking into consideration relevant advances in medical or statistical practice.

It is crucial that the protocol of an equivalence or non-inferiority trial contains an explicit statement about its intention. An equivalence (non-inferiority) margin should be specified in the protocol; this margin is the largest difference between the test treatment and active control that can be judged as being clinically tolerable, and it should be smaller than differences observed in superiority trials between the active comparator and the placebo. For the active-control equivalence trial, both the upper and lower equivalence margins are needed, while only the lower margin is needed for the active-control non-inferiority trial. The choice of equivalence margins should be justified clinically. For equivalence trials, two-sided confidence intervals should be used. Equivalence can be concluded when the entire confidence interval lies within the equivalence margins. There are also special issues regarding the choice of analysis sets. Subjects who withdraw consent or drop out of any treatment or comparator group will be predisposed to have a lower treatment response, and hence the results of using the full analysis set may be biased toward showing equivalence. This is discussed further below.

**Trials to show a dose–response relationship**

Dose–response trials may serve several objectives, most importantly: confirmation of efficacy; investigation of the shape and location of the dose–response curve; evaluation of an optimal starting dose; definition of strategies for individual dose adjustments; and determination of a maximal dose beyond which surplus benefit would be unlikely to occur. For these purposes the use of procedures to estimate the relationship between dose and response, including the calculation of confidence intervals and the use of graphical methods, is as important as the use of statistical tests. The hypothesis tests that are used may need to be tailored to the natural ordering of doses or to particular questions regarding the shape of the dose–response curve (e.g., monotonicity).
The details of the applied statistical methods should be provided in the protocol.

**Study protocol**

In the above we have discussed a number of features and considerations necessary to mount a clinical trial. The study protocol pulls it all together.

The protocol is the “recipe” for a clinical trial, describing in detail the scientific rationale, patient eligibility, trial treatments, study investigations, outcome measures, sample size, statistical analysis, and management of safety issues [39]. The protocol should be understandable to investigators and research staff taking part in clinical trials, so brevity and simplicity are key objectives when preparing the protocol. As stated above (see Study populations) one of the most important sections is eligibility (inclusion and exclusion criteria), since this governs how many patients can be entered and is the main driver for enrolment (or lack of it). Inclusion criteria should provide a simple guide to the population that should be screened for eligibility, and exclusion criteria should explain which patients should not be enrolled for safety reasons. Exclusion criteria “rule out patients” and generally make trials less applicable to clinical practice [40]. The usual justification for an extensive list of exclusion criteria is that a “homogeneous” population is needed to test the hypothesis. Since this is not the situation in clinical practice (i.e., patients with a particular disease are often heterogeneous in terms of age, gender, and comorbidities), there seems little logic in supporting this practice. We propose a simple “rule” that no trial should have more than 10 exclusion criteria; this should allow better enrolment and greater generalizability of results. All of the features and issues listed in Table 1.1 also have importance and must be considered and dealt with seriously in the protocol.

**Trial monitoring**

In the past, the monitoring of a trial’s progress was essentially the function of the study sponsor. Today, this is less and less common. Monitoring the quality of the study is still the responsibility of the sponsor; however, this function is often shared with another institution, the Independent Data Monitoring Committee (IDMC), also called the Data and Safety Monitoring Board (DSMB). This committee usually consists of at least three people, and includes no less than two clinicians and one biostatistician [41,42]. Other persons, including epidemiologists, ethicists, and patient advocates, might also belong to the DSMS, depending on the type of clinical trial.

In contrast to monitoring for quality, monitoring for safety and efficacy has shifted from the sponsor towards the IDMC. If unblinded interim analyses are performed for evaluation of efficacy or safety, they are done solely for the IDMC. These interim analyses affect alpha spending [43], and the IDMC has to take into account this issue. Also, usually, the IDMC alone can view unblinded data and its reports may form the basis of the safety information
provided to regulatory authorities. Interim analyses for efficacy and safety have become standard features of clinical trials, and the role of the IDMC has steadily increased. Other activities, such as decisions whether to adjust sample size or study length, may be recommended by the IDMC.

**Data monitoring**

The process of evaluating and analyzing data accruing from a clinical trial can be prone to systematic bias and/or type I error. Therefore, all interim analyses, formal or informal, preplanned or unplanned, by any study participant, sponsor staff member, or data monitoring group need to be explained in full, even if the treatment groups were not identified. The requirements for statistical adjustments due to such additional analyses should be adequately mentioned. Any operating instructions or procedures used for such analyses should be described in detail. Data monitoring without code breaking should also be described, even if this type of monitoring is considered to cause no increase in type I error.

**Data analysis sets**

Ultimately, the data from the clinical trial will be analyzed. Ideally all randomized subjects will be analyzed. The data set that includes all subjects as randomized is called the intention-to-treat (ITT) data set or analysis set. This data set retains all the optimal features of the randomization and permits valid statistical analysis. In practice, the complete ITT analysis data set is not achieved due to premature termination, safety concerns, drop-outs, lack of adherence, etc. Below and in the section Analysis methods, we elaborate more on the implication of the lack of the ideal ITT data set.

**Intention-to-treat analysis and data set**

The particular strength of RCTs is the avoidance of various sources of bias. In order to safeguard the full protective effect of randomization against bias, inclusion of all randomized patients regardless of further study adherence is the preferred strategy [44]. Thus, analysis of the ITT data set or, equivalently, the ITT analysis refers to the analysis performed exactly according to initial randomization on all those randomized. Some trialists prefer to exclude patients in the analysis for whom outcome data are missing. This is not a rare problem since in half of RCTs more than 10% of outcome data are missing [45]. Sometimes this approach is reasonable, but strictly speaking, this kind of analysis does not represent an ITT analysis and, instead, should be claimed as a “complete case” (or “available case”) analysis. Alternatively, it is possible to impute missing outcome data and perform the analysis on available cases plus the imputed data. The validity of this analysis depends upon the extent of the missing data and the reasons for missing data (see section on Missing data below). The modified ITT (mITT) analysis data set is a subset of the ITT data set and allows for the post-randomization exclusion of some subjects in a justified manner, e.g. those who never receive the treatment drugs. Analysis
of this data set is at times reasonable, but can contain biases and has been recently reported to be associated with industry funding and authors’ conflicts of interest [46].

**Per protocol analysis and on-treatment analysis data sets**
The per-protocol (PP) data set is restricted only to those “ideal” study subjects who fulfil the protocol in terms of the eligibility criteria, study interventions, and outcome measures. On-treatment analysis refers to an approach stratified according to the real allocated treatment regardless of randomization. Though a per-protocol and on-treatment analysis may be reasonable in some settings, it should be emphasized that this analysis approach represents a non-randomized, observational comparison.

**Subgroup analysis data sets**
Whether the main results of a clinical trial do or do not support the null hypothesis, it is inevitable that the sample will be analyzed based on clinical subgroups of interest to determine the response to treatment [33,34]. Typical subgroups include gender (binary), age (continuous but may be divided into categories), and diabetes (categorical). Subgroups of interest may be specified prior to the study being analyzed (prospectively defined or *a priori* subgroups) or after the main results have been analyzed (*post hoc* or retrospective). Subgroups can be exploratory in post-hoc analysis of clinical trials and then usually do not have sufficient power to provide reliable estimates of treatment effect, but can be hypothesis generating. Important subgroups of interest (like diabetic or ST-elevation myocardial infarction subgroups in cardiovascular trials) are often predefined per protocol in large phase 3 RCTs, and sample size and power calculations can be performed considering these predefined subgroup analyses. Subgroups that have more than 500 events may have greater reliability, e.g. in large trials with extremely favorable p-values (<0.001) or subgroups in a meta-analysis of large trials. However, the unreliability of even large subgroups has been demonstrated in the ISIS-2 trial, which showed that aspirin was highly effective in reducing mortality after myocardial infarction in the group overall, but when data for patients with the star signs Libra and Gemini were analyzed, there was no apparent benefit of aspirin in contrast to the other star signs [47]. Using such a non-medical subgroup emphasizes the unreliability of subgroup analysis, so in general subgroups should be considered hypothesis generating, not hypothesis testing.

**Unit of analysis**
It should be precisely stated which patients are to be included in each efficacy analysis. For example, as discussed above, depending on the analysis set the included patients can be those with any efficacy observation or with a minimum number of observations; all patients receiving any test drugs/investigational products; only patients completing the trial; all patients with
an observation during a particular time window; or only patients with a specified degree of compliance.

In all of the above cases, the subject (patient) is the unit of analysis. Recognition of this is important in applying statistical analyses procedures to the data. Often there is a single outcome associated with the patient (e.g., does or does not develop a myocardial infarction). The analysis considers this outcome as coming from a single subject. In contrast, there are situations where an outcome is measured repeatedly in a subject (e.g., blood pressure on monthly visits). A proper statistical analysis must consider these multiple, repeated measurements on a single subject. The unit of analysis is still the patient, but now there are correlated measures on him/her. Confusion has accrued over how to deal with multiple measurements collected either at a single time point or at several time points during the course of a trial. Methods such as generalized estimating equations and random-effects models can effectively deal with the problem of correlation within subjects and across time [48–50].

**Missing data**

As mentioned above, numerous factors may influence drop-out rates: study duration, efficacy and toxicity of the study drug, disease nature, and other individual factors. Ignoring those patients who “dropped out” (i.e., who did not complete the study for whatever reason) and only analyzing patients who completed the study can be misleading. A large number of drop-outs, however, even if included in the analysis, may cause bias, particularly if there are differences in the timing of drop-outs between the treatment groups or the reasons for dropping out are related to outcome. Although the effects of early drop-outs can be difficult to evaluate, any possible impact should be investigated as fully as possible. In case the drop-out rate is high, it might be helpful to concentrate on analyses at time points when most of the patients were still under observation and when the full effect of the drug could be expected. It may also be helpful to consider modeling approaches that have been developed for analysis of such incomplete data sets. The results of a clinical trial should be analyzed not only for the subgroup of subjects who completed the study, but also for the entire patient population as randomized (*intention-to-treat*) or at least for all those with any on-study measurements (*on-treatment*). Several factors should be taken into account and compared for the treatment groups in analyzing the effects of drop-outs. These include the reasons for the drop-outs, the time to drop-out, and the proportion of drop-outs among treatment groups at various time points. Procedures for dealing with missing data, e.g., use of estimated or derived data, should be described. Detailed explanation should be given as to how such estimations or derivations were performed and what underlying assumptions were made.

**Sensitivity analysis**

Sensitivity analyses refer to broad-spectrum of computational analyses in which certain inputs are modified to see how this will affect the outcome.
They address the following questions: How confident are the results? How great will be the impact on the results if the basic data are slightly wrong? Will this give a completely different outcome? Unused categories like missing data, conflicting data derived from various sources (e.g., different results reported by the general practitioner and the investigator), or uncertain cases (no disease versus disease) might have a major impact on the general results if included in the analysis. Advanced statistical imputation methods can be applied to check the impact of these unused categories by substituting uncertain cases and missing values, e.g., on the basis of best guess estimates.

**Analysis methods**

For data analysis, the patients included in each efficacy analysis should be precisely described, e.g., all patients receiving any test drugs/investigational products; all patients with any efficacy observation or with a certain minimum number of observations; only patients completing the trial; all patients with an observation during a particular time window; or only patients with a specified degree of compliance. It should be clear, if not defined in the study protocol, when (relative to study unblinding) and how inclusion/exclusion criteria for the data sets analyzed were developed. Even if it was proposed that primary analysis would be based on a reduced subset of the patients with complete follow-up data, there should also be, for any trial aimed at establishing efficacy, an additional analysis using all randomized (or otherwise entered) patients with any on-treatment data. Ideally, there should be a figure or table listing all patients, visits, and observations excluded from the efficacy analysis. The reasons for exclusion should also be analyzed for the whole treatment group over time. The Consort diagram (with extensions as needed) in Figure 1.2 is a useful device for this.

**Demographic and other baseline characteristics**

Group data for the relevant demographic and baseline characteristics of the patients, as well as other factors identified during the study that could affect response, should be presented, and the comparability of the treatment groups for all relevant characteristics should be displayed using graphs or tables. Analysis of the ITT data set is the only analysis justified by the randomization and should be the primary analysis. This may be followed by data on other groups used in principal analyses, such as the “per-protocol” analysis, “complete case,” or other analyses, e.g., groups defined by compliance, concomitant disease/therapy, or demographic/baseline characteristics. When such groups are analyzed, data for the complementary excluded cohort should also be provided. In a multicenter study, where appropriate, comparability between centers should be assessed. A diagram showing the relationship between the entire sample and any other analysis groups should be provided. The critical variables will usually depend on the specific nature of the disease and on the protocol, but should usually cover:
• **Demographic variables:**
  - Age
  - Sex
  - Race.

• **Disease factors:**
  - Specific entry criteria (if not uniform), duration, stage, and severity of disease, and other clinical factors and subgroups of known prognostic significance
  - Baseline values for critical clinical measurements performed during the course of a study or identified as important indicators of prognosis or response to therapy
  - Concomitant illness at trial initiation, such as renal disease, diabetes, heart failure
  - Relevant previous illness
  - Relevant previous treatment for the illness treated in the study
  - Concomitant treatment maintained, even if the dose was changed during the study, including oral contraceptive and hormone replacement therapy; treatments stopped at entry into the study period (or changed at study initiation)
  - Other factors that might affect response to therapy (e.g., weight, renal and hepatic status, antibody levels, metabolic status)
  - Other possibly relevant variables (e.g., smoking, alcohol intake, special diets) if pertinent to the study.

**Measurements of treatment compliance**
Any measurements of compliance of individual patients with the treatment regimen under study and drug concentrations in body fluids should be summarized, and analyzed by treatment group and time interval.

**Analysis of efficacy**
Treatment groups should be compared for all relevant measures of efficacy (primary and secondary endpoints; any pharmacodynamic endpoints studied), as well as benefit/risk assessment(s) in all patients where these are utilized. In general, the results of all analyses implemented in the protocol and an analysis including all patients with on-study data should be performed in studies aimed at establishing efficacy. The analysis should provide results to estimate the size (point estimate) of the difference between the study treatments, the associated confidence interval, and the results of the predefined hypothesis testing. Analyses based on continuous variables (e.g., mean blood pressure, levels of cardiac necrosis markers at time point X, etc.) and categorical factors (e.g., worsening of heart failure, hospital admission, or occurrence of event) can be equally valid. For a multicenter study, where appropriate, data display and analysis of larger individual sites or more sites per geographic region should be included for critical variables to give a clear
picture of trends between the recruitment sites and to evaluate possible differences in study conduct.

It is important for clinicians to avoid an obsession with a “P value less than 0.05” and develop an understanding of how statistical methods can be applied in a valid way to clinical trials. Experienced statisticians will emphasize that many of the assumptions in the design and interpretation of a clinical trial are actually clinical rather than statistical, and it is surprising how often clinical assumptions are carried out in a sloppy manner. For example, the key driver for sample size estimation is the expected difference between the two groups and this is determined by careful scrutiny of the literature. Similarly, subgroup analysis and interpretation can be heavily influenced by seeing analyses from a completed trial. Analyses that are triggered by seeing “interesting results” are often called “data derived” and experience tells us that data-derived analyses often lead to biased conclusions, in contrast to analyses that are prespecified prior to seeing the results.

An understanding of basic statistical issues applied to clinical trials is essential for clinicians. We provide a short introduction to these issues below.

**Distribution of data**

Understanding how to classify measurements in a clinical trial according to accepted statistical methods is an important starting point. Results obtained in a clinical trial can be expressed in two main ways: “continuous” or “categorical.” Measurement of blood pressure, weight, height, plasma cholesterol, serum creatinine, and left ventricular ejection fraction are examples of “continuous” variables. Continuous variables can be expressed at different levels of precision depending on the scale of the measurement methods (e.g., centimeters or millimeters). Common biological data such as height, weight, and cholesterol have a “normal” or Gaussian distribution. For example, in a sample population where height is measured there will be a clustering of measurements around the average (“mean”) and fewer measurements further away from the mean. The shape of the normal distribution curve can be described mathematically and the key variables are the mean, variance, and standard deviation (discussed below). Data that are not normally distributed are commonly found in clinical trials of biomarkers, especially those that are found at low levels in individuals without disease, such as C-reactive protein and troponin. Analysis of these data requires statistical tests suitable for non-parametric analysis.

Measurements of death, hospital admission, stroke, and myocardial infarction are examples of categorical variables. When there are only two possibilities, categorical variables are called “dichotomous” or “binary.” Thus in a trial evaluating the effects of a new treatment on mortality, patients can be classified as having one of two outcomes: dead or alive. Most measurements in clinical trials can be classified as continuous or dichotomous. Other ways of expressing measurements include assigning grades or “ranks.” For example, when asked “Do you feel depressed?” a patient may be asked to select the most appropriate answer from the following: “never, sometimes, often,
always.” The patient’s response has four possible categories that can be ranked from mild to severe. The New York Heart Association classification of heart failure also has four possible categories. When managing patients, it is helpful to know whether a treatment is indicated or not, even when there is uncertainty about the benefits and risks of treatment, which may require dichotomizing continuous variables, e.g., a persistent systolic blood pressure above 140 mmHg may be an indication for pharmacological treatment for hypertension, implying that pressures below 140 mmHg do not require treatment. The reality is more complex, but it is common practice to convert a continuous variable into a dichotomous variable, although this often leads to loss of statistical power.

The importance of understanding the classification of measurements is that different statistical assumptions and tests are used for continuous data and categorical data. When data from a clinical trial have been collected, it is important to present them in a logical and clear manner, which requires the use of tables and figures. Data showing key variables (or “covariates”), which describe the population sample enrolled into a clinical trial (including age, gender, medical history, disease severity, etc.), are called “descriptive” or “summary” statistics.

**Statistical analysis of clinical trials (elementary concepts)**

Figure 1.3 shows a simplified, but helpful, decision tree for statistical comparisons depending on the type of variable measured in a RCT.

The comparison of categorical variables from a two-group clinical trial (parallel sample design with treatment versus control) using the Chi-square test is relatively simple and can be performed on a hand calculator. The first step is to populate the “2 x 2” table, which has the number of subjects with the outcome of interest (yes or no) in the columns, and treatment and control in the rows. The “observed minus expected” is a common method for estimating the Chi-square value in which the average values of the events rates are calculated for each of the four cells (estimated or E). These are then subtracted from the actual value (observed or O), squared, and divided by the expected value for that cell. The four values are added and this provides the Chi-square statistic. Reference to standard tables will provide the associated p-value. This Chi-square test is equivalent to the test of differences in proportions of outcomes for the two treatments, where the proportion of events in the test treatment is compared to the proportion of events in the control.

When treatment effects on changes of a categorical variable are evaluated (i.e., comparing before and after treatment effects), the McNemar test is useful. It is applied to 2 x 2 contingency tables with a dichotomous distribution and matched pairs, and tests the hypothesis whether or not the row and column marginal frequencies are equal (also called marginal homogeneity) [51].

For normally distributed data the key measurements for comparison are means and their associated standard deviations. Standard deviation describes the “spread” of the data, which includes the difference between highest and lowest values, and the variability around the mean. Data which are clustered
around the mean will have smaller standard deviations. Standard deviation is the square root of the variance, which is the sum of all the differences between the mean and actual value divided by the number of observations minus one. The appropriate tests are indicated in Figure 1.3 and are called t-tests.

For non-normally distributed data there are equivalent tests called non-parametric test; they are the Wilcoxon Rank Sum and Mann–Whitney tests.

The above discussion may give the impression that for categorical and continuous data, rigid selection of Chi-square test, t-tests, Mann–Whitney or Wilcoxon tests is imperative. This is not the case and there is a substantial literature showing that, for example, the t-test is valid (robust) when used on simple five-point scale data. Here the five points may measure, for example, pain ranging from no pain (score 1) to maximum pain (score 5) [52–57]. The above discussion was given to indicate the need to select an appropriate statistical test that is valid for the data.

More advanced statistical concepts applied to trials include: when to stop a trial early (“stopping rules”), exploring the strength of association between variables (logistic regression), comparison of events over time (survival analysis), and introducing new treatments which may be safer and more

---

Figure 1.3 Decision tree for analysis of comparative data from a clinical trial.
convenient to use but which are not more effective than existing treatments (“non-inferiority” testing). For a fuller discussion of these concepts the reader is referred to standard statistical texts.

**Adjustments for covariates**
Selection of, and adjustments for, demographic or baseline characteristics, concomitant medical therapy, or any other covariates or prognostically relevant factors should be explained in the analysis report, and methods of adjustment, results of analyses, and supportive information (e.g., ANCOVA, linear regression models) should be addressed in the detailed description of statistical methods. Although not a mandatory part of the trial report, comparisons of covariate adjustments and prognostic factors can have informative value in the summary of clinical efficacy data.

**Time to event (“survival”) analysis**
In many clinical trials the time to the occurrence of an event represents the main outcome measure of interest. This is the case for a binary outcome where the time at which an outcome occurs is built into the analysis. Examples are when the events are adverse, like cardiovascular death, non-fatal myocardial infarction or stent thrombosis, or positive, such as improvement of left ventricular function or heart failure symptoms, or neutral, such as freedom from arrhythmia. In all these cases, it is suitable to apply survival analysis for evaluation of endpoints; the term survival can be misleading as this method could be applied to any event, although originally it was planned for mortality evaluation. Events in survival analysis are defined by a “conversion” from one state (in its simplest form, no event) to another (i.e., occurrence of the event) at an instantaneous time point. Typically, not all the study participants will have experienced the event at the end of the follow-up period. This feature is defined as censoring, meaning that the observation period ended without occurrence of the predefined event, either because the subject has not experienced the event within the follow-up period, has dropped out event-free prior to termination of follow-up, or experiences a different type of event that makes further follow-up impossible (e.g., the patient died from a non-cardiovascular cause and the survival analysis was planned for analysis of cardiovascular death as event). In case of censoring, it is not feasible to know whether and when the subject would have experienced the event. Right censoring means that a subject’s follow-up terminates before occurrence of the event; while left censoring refers to when the event has taken place before the subject has entered the study. Classical statistical tests, such as t-test, Mann–Whitney test or Chi-square test, are not suitable to analyze such survival data because they are not designed to take censoring into account. All trial subjects (including those with censored follow-up) can provide important information for event analysis, and should therefore not be excluded from the analysis. However, considering the censoring time as an equivalent survival time might lead to biased estimate of survival time and event probability. Therefore, specific statistical methods to handle survival data need to be applied. In
clinical practice, all the subjects are not enrolled at the same time and thus the follow-up period can vary from one subject to another.

To estimate the proportion of subjects surviving the event (i.e., being event free) at a specified time point, and hence to calculate the survival probability to that time in relation to the generic population from which the sample derives, the Kaplan–Meier method \[58,59\] is commonly used, which permits censored information to be dealt with. In some situations it is not possible to know or to record the precise time when the event occurs; the only available information is that the event occurred in a certain time window. If this is the case, the survival probability cannot be updated every time an event happens, and the derived survival curve and probability are called life table or actuarial estimates. Comparisons of survival distribution between two samples are commonly applied in clinical trials to establish the efficacy of a new treatment compared to a control treatment. These can be analyzed using a non-parametric test method, the Mantel–Cox test or log-rank test. Sophisticated models based on the linear regression analysis (like the Cox regression analysis) are applied to compare survival data from different samples, mostly in non-randomized clinical trials to adjust for differences in group compositions, but these analyses require certain assumptions for data distribution \[60\].

**Sample size and power**

For scientific, ethical, and practical reasons, the sample size for a trial needs to be planned carefully, with a balance between statistical, medical, and resource considerations. Ideally, a clinical trial should be large enough to have a high probability (statistical power) of detecting with sufficient significance any clinically relevant difference of a given size. For clinical trials with a categorical outcome measure and two comparison groups, we need to know the expected event rate in the control group, the expected difference between groups, and the assumed values for alpha and beta (see below for definitions of these). Obtaining reliable estimates for the expected treatment difference and event rate is probably the most important and difficult aspect of the sample size calculation. Clinical trials, for illustration say superiority trials, start with the null hypothesis that states there is no difference in event rates between the two groups (this refers to the true state of nature if the null hypothesis is true) and attempt to prove the alternative hypothesis (the research hypothesis) that there will be a difference. The hypothesis test is attempting to decide which is the more plausible hypothesis based on the trial’s data. The alpha value, which equates with the p-value and often set at 0.05 or less, is the probability that we will commit a type 1 (alpha) error when the null hypothesis is true. That is, it is the probability of saying the null hypothesis is false, when in fact it is true. In other words, if the null hypothesis is true, we have a 0.05 chance of the data and statistical test telling us to reject the null hypothesis. The alpha error is also known as the “false positive” rate. The type II (beta) error rate is the chance of saying the null hypothesis is true when, in fact, it is not. This error rate is often set at 0.2 or 20% (i.e., the false
negative rate). The “power” of a study is loosely defined as the ability of a study to detect the expected difference in terms of the estimated number of events during the trial and is described by the function $1 - \beta$, which is conventionally 0.8 (or 80%). Formally the power of the statistical test is the probability that the statistical test will tell us to reject the null hypothesis when the null hypothesis is false. Alpha and power values can be modified to be more stringent than 0.05 and 0.8, respectively, and are sometimes set at 0.01 or 0.9, respectively, but the consequence of this additional precision is that sample sizes are much larger. When calculating sample size, the values for alpha and power are converted to function values, which appear as constants in the sample size equation. In addition to calculating the statistical power of a study to detect an expected numeric difference, it is important to determine a “minimal clinically significant difference” or “minimal important difference,” defined as the smallest difference that clinicians and patients would care about. This parameter should be considered in the sample size estimation and should be stated in the clinical trial report, so that any statistically significant difference in outcome can be judged for clinical relevance.

Some trialists have postulated that “underpowered” trials may be acceptable as they could be used later in combined analyses in systematic reviews and meta-analyses [61–63]. Of note, this implies important caveats—the trial should be reported adequately, follow principles of minimizing bias, and published irrespective of the results. Underpowered trials with indeterminate results often remain unpublished. Ideally, all trials should individually have sufficient power to address the primary hypothesis. Irrespective of the power, authors need to properly report their intended sample size and expected difference between treatment and control with all their methods and assumptions. That gives transparency to readers and a measure by which to assess whether the trial achieved its planned sample size.

**Safety evaluation**

Analysis of safety-related data can be performed at three levels:

1. **Extent of exposure (dose, duration, number of patients)** should be explored to determine the degree to which safety can be assessed from the study.
2. **Typical adverse events and laboratory test changes** should be identified, compared for treatment groups, and analyzed, as appropriate, for factors that may affect the frequency of adverse reactions/events, such as time dependence, relation to demographic characteristics, and relation to dose or drug concentration.
3. **Serious adverse events and other relevant adverse events** should be identified, usually by detailed examination of patients who died or who left the study prematurely due to an adverse event, whether or not this was identified as drug related.

The International Conference on Harmonisation (ICH) guideline “Clinical Safety Data Management: Definitions and Standards for Expedited Reporting” defines serious adverse events as follows: “A serious adverse event
Introduction to randomized clinical trials in cardiovascular disease

(experience) or reaction is any untoward medical occurrence that at any dose: results in death, is life-threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity, or is a congenital anomaly/birth defect.”

Safety data have always been important, but are often not consequently analyzed. Most clinical trials designed to show efficacy of a treatment (difference between treatment and a control) are not large enough, either in terms of sample size or follow-up, to identify serious safety concerns. Recent developments regarding health regulatory issues have led to more rigorous analyses. For instance, data mining techniques have been generated and successfully applied to safety data from previously unrevealed structures [64,65]. Drug approval does not only involve efficacy studies, but also complete analyses of safety studies. Furthermore, because there are so many effective treatments in cardiovascular disease, the choice among them may ultimately rest upon the safety issues.

Subsets and more

Evaluation of any “efficacy subset” of patients should be devoted to the effects of dropping patients with available data from analyses because of their poor compliance, missed visits, ineligibility, or any other reasons. As noted above, an analysis using all available data should be performed for all studies that aim to establish efficacy, even if it is not the primary analysis planned by the investigators. In general, it is of advantage to confirm the robustness of the principal trial results with respect to alternative subsets of patient populations. Any relevant differences deriving from the choice of patient population for analysis should be the subject of explicit discussion. If the size of the study allows, any subgroups based on important demographic or baseline characteristics should be examined for unusually large or small responses and the results presented, e.g., comparison of effects by age, gender, race, disease condition, or cardiovascular risk groups. These analyses are not intended to retrieve an otherwise negative study, but may reveal hypotheses worth examining in other trials or be helpful in refining labeling information, or patient or dose selection. Where there is a preceding hypothesis of a differential effect in a particular subgroup, this hypothesis and its assessment should be part of the planned statistical analysis and considered in the sample size calculation.

Types of significance

Statistical significance

Statistical significance tends to be used in the context of null hypothesis significance testing. The null hypothesis (e.g., for a superiority trial) states that there is no effect (or in other words the effect is zero) in a given population. Rejection of the null hypothesis means that we have a statistically significant
reason to believe the null hypothesis is false. If the null hypothesis were true, there is only a 0.05 (the alpha value is set usually at 0.05 and p-value at ≤0.05) chance of reaching this conclusion. Testing of the null hypothesis is often misunderstood in some way: that the p-value is the likelihood that the null hypothesis is wrong; it has to emphasized that null hypothesis testing only yields information about whether results are statistically likely given some assumption about the population (the truth of the null hypothesis). In terms of testing clinical treatments, if a treatment is actually ineffective, statistical significance can only provide an answer to the question, “How likely is it that the statistical test of the treatment would falsely indicate that the treatment is effective?” It does not give any information about practical or clinical significance.

**Practical or “clinical” significance**
In broader terms, “practical significance” answers the question, “How effective is the intervention or treatment, or how much change does the treatment cause?”. In terms of testing clinical treatments, practical significance ideally provides quantified information about the importance of a finding, using metrics such as effect size, number needed to treat (NNT), and preventive fraction.

**Calculation of practical significance**
Effect size is one type of practical significance. It is a measure between two variables in a statistical population and quantifies the extent to which a trial result deviates from expectations. Effect sizes have their own sources of bias, are subject to change based on population variability of the dependent variable, and tend to focus on group effects, not individual changes. Effect size can provide important information about the results of a study, and are recommended for inclusion in addition to statistical significance [66].

Clinical significance usually refers to two related but different concepts either by (1) being of a magnitude of effect that conveys practical (in this case there is an interchangeability with practical significance) or (2) more technically and restrictively, addressing whether an intervention or treatment may or may not fully correct previous findings. There are a variety of statistical methodologies to calculate clinical significance. For detailed description of these analysis methods the reader is referred to the statistical literature.

**How have trials contributed to medical care?**
Many trials have influenced practice in cardiovascular disease and we provide some examples. It is important to note that it is unusual for a single trial to change practice—usually a series of trials taken together provide clear evidence of benefit of using a new treatment paradigm. In addition, good trials build on a wealth of epidemiology, basic science, and clinical studies, which have elucidated the nature of the disease and provided preliminary evidence of the benefit of a potential new treatment. The modern era of clinical trials
started in the late 1970s with investment from the National Institutes of Health to identify and combat major health threats, including cardiovascular disease and cancer. In the 1980s the need for larger studies and cooperation between hospitals was established, leading to many large trials which have provided reliable evidence about benefits and risks of promising treatments.

**Clinical trials for treatment of high blood pressure** *(see also Chapter 12)*

Many major trials of antihypertensive drug therapy have been performed in the last 10–15 years: CAPPP (Captopril Prevention Project) [67]; STOP Hypertension 2 and NORDIL [68,69], ALLHAT [70], Intervention as a Goal in Hypertension Treatment (INSIGHT) [71], INVEST [72], LIFE [73], LIFE-ISH [74], Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT) [75], ONTARGET [76], and ACCOMPLISH [77]. All of them belong to the category of active-control RCTs. Among these trials, two different concepts are followed in the comparison with the active control. In the first group of trials, the aim was to show the preservation of a clinically meaningful efficacy and safety margin (equivalence or non-inferiority trials). Examples are CAPPP, INSIGHT, INVEST, LIFE and LIFE-ISH, STOP Hypertension 2 and NORDIL, and ONTARGET. In the second group of studies using an active control, the margin of clinically meaningful effect has been exceeded, as seen in superiority trials. The ACCOMPLISH trial demonstrated superiority of initiating antihypertensive therapy with an angiotensin converting enzyme (ACE) inhibitor combined with a calcium channel blocker (CCB) over initiating therapy with a thiazide-type diuretic, and the ALLHAT trial showed that the diuretic chlorthalidone, as the active control standard treatment, was superior to amlodipine, lisinopril or doxazosin.

All recent trials assessing the efficacy and safety of antihypertensive drugs have been designed and conducted as randomized, blinded trials. Such trials not only minimize experimental bias, but also implement many important aspects of controlled clinical trials. Many large antihypertensive trials, such as INVEST and ALLHAT, were designed to include a prospective, randomized, open, blinded endpoint evaluation (PROBE) in a large number (thousands) of patients in multiple countries. In some studies, there is also a focus on demonstrating reduction of mortality and/or cardiovascular morbidity rather than simply measuring the blood-lowering effects.

Hypertension is associated with a higher stroke risk, and several early trials, including the MRC Blood Pressure trials, showed that beta-blockers and thiazides could effectively lower blood pressure and reduce the risk of stroke [78]. Several subsequent trials in elderly patients showed that these benefits were not confined to “younger” hypertensives, and that the elderly could also be protected from stroke and death with effective antihypertensive treatment [79]. More recently, combinations of ACE inhibitors and calcium antagonists have been shown to be more effective than the more traditional combinations of beta-blockers and thiazides in the ASCOT trial [80]. Thus,
larger randomized trials have the ability to reliably evaluate the efficacy and safety of treatments.

**Clinical trials of atherosclerosis, lipid lowering, and statins (see also Chapter 12)**

Understanding the main components of atherosclerosis, including the biochemistry of cholesterol and, low density lipoprotein (LDL) oxidation, role of monocytes and activated macrophages, and uptake of oxidized LDL into the arterial wall, took four to five decades, with several groups providing important insights into these processes [81]. Epidemiological studies including Keys’ “7 countries study” also provided evidence of the link between cholesterol and coronary heart disease [82]. Early studies in the 1970s and 1980s of strategies to lower cholesterol (including diet and fibrates) did not show clear benefit because the amount of cholesterol lowering was modest (in the order of 10–15% proportional reduction) and the studies were too small to detect moderate but important benefits [83]. Synthesis of agents that could reduce cholesterol production in the liver by inhibiting a key enzyme, HMG Co-A reductase, led to the development of statins, which were found to reduce cholesterol by 20% or more [84]. The first trial to show that cholesterol lowering could reduce mortality and major morbidity was the 4S study, which randomized 4444 patients to simvastatin or placebo, and demonstrated a proportional reduction of 25% in death over a mean 4 years of follow-up in patients with prior myocardial infarction and cholesterol above 6.5 mmol/L (260 mg/dL) [85]. This led to the widespread uptake of statin use, with other trials confirming these results in a wider range of patients [86].

**Clinical trials of acute myocardial infarction, coronary thrombosis, aspirin, and fibrinolysis (see also Chapter 4)**

Acute coronary syndromes (ACS), including myocardial infarction (MI) and unstable angina, are responsible for about 15% of all deaths globally per year, and complications related to ACS cause 20–30% of all hospital admissions globally each year [87]. The main pathophysiological cause of ACS is plaque rupture leading to activation of platelets and intracoronary thrombus formation. Prior to 1985 standard treatments for MI were pain relief, glyceryl trinitrate (GTN), bed rest, monitoring, and usually heparin to prevent deep venous thrombosis. The ISIS-1 (first International Study of Infarct Survival) and MIAMI (Metoprolol In Acute Myocardial Infarction) trials demonstrated the benefits of beta-blockers early in MI [88,89], and were followed by the GISSI (Gruppo Italiano per to studio delta streptochinasi nell’infarto miocardico)-1 and ISIS-2 trials, which showed the benefits of early fibrinolysis; the latter trial confirmed that aspirin given early in MI could reduce all-cause mortality by about one-fifth [90,91]. Subsequent trials have added to knowledge about the benefits of antiplatelet therapy in ACS and the role of primary PCI in acute ST-segment elevation MI. Primary PCI has replaced fibrinolytic therapy in centers that can deliver it in a timely manner on a 24-hour basis [92].
Conducting reliable trials: issues of size and simplicity

The most common design issues in clinical trials have been discussed in the literature [93–95] and there is wide agreement on many key aspects. Practical issues are a major challenge, including cost, regulation (and its associated bureaucracy which can impede progress in trials), and complex project management. By definition, clinical trials involve patients in a clinical setting, so it is essential to keep trial procedures streamlined and simple and data collection to a minimum. It is noteworthy that clinical trials operate in somewhat artificial environments that do not entirely reflect real-world behavior. Most trials are complicated and not run efficiently: excessive and unnecessary tests (increasing patient discomfort and the time needed to administer and report tests) and excess data collection waste time and resources. A simple rule of thumb for data collection is to limit the number of data points to 300 per study subject (excluding repetitions that may occur with multiple follow-up visits). This approach provides a discipline to the length and complexity of case report forms by ensuring there is clear justification for all data points collected. Clinical trials are costly (small trials cost hundreds of thousands of dollars and large phase 3 trials cost millions of dollars) and this is a major limitation to the number of patients that can be enrolled. Several features of clinical trials require similar time and effort irrespective of trial size, including preparing the protocol, programming the database, and statistical analysis. Large multicenter trials will require more time and effort to set up and run than trials involving fewer centers, but much of this can be streamlined by simplifying the protocol and data collection. Trials that are too small to reliably answer the question are usually a waste of time and money, and making trials larger is an essential priority. “Simple” streamlined trials do not need to be complicated and can be cost-effective to carry out. The term “simple” refers to the outcome measure and not to the conduct of the study; thousands of patients from hundreds of centers and different countries may be followed for long periods of time (up to 5 or more years).

Costs, cost-effectiveness, and health economics

If a large clinical trial has shown that a certain treatment has a beneficial action with an acceptable safety profile, “How much does it cost” is usually the next question. Of course “cost” and “price” are different but related issues, where the upfront “cost” of a treatment is hopefully offset by downstream reduced costs of healthcare, improved quality of life, and better productivity. Health economics is a growing specialty that evaluates the costs and healthcare impact of introducing new treatments. It is a key part of modern clinical trials that provides accurate assessments of cost and cost-effectiveness within the trial [96,97]. This information forms the basis of sensible extrapolations beyond the trial in terms of longer lengths of treatments and management of larger populations with the same disease. Health economics attempts to model the costs and benefits of introducing new and effective treatments and
Cardiovascular Clinical Trials

is used by many healthcare funders to help make decisions about which treat-
ments should and should not be provided.

**Practical issues: managing a trial, funding, and regulation**

Trial management, including trial design, securing funds, obtaining approv-
als, coordinating sites, and data management, is a complex and time-
consuming activity. The current view is that all clinical trials should be
managed by specialist clinical trial units (CTUs) with the appropriate expert-
tise and capacity. Academic clinical trial units usually specialize in a disease
area like cancer or cardiology, but there is a growing trend for larger units to
undertake trial management in broader disease areas. Academic CTUs are
made up of multidisciplinary teams, including trial managers, administrators,
data managers, clinicians, and statisticians. A well-conducted trial will address
an important question reliably and, given sufficient resources, most experi-
enced CTUs can carry out trials to a high standard. The main barriers to
conducting good trials are insufficient funds, unrealistic timelines, and excess
bureaucracy and “red tape.” The other issue facing academic CTUs is the cost
of preparing grants and funding applications that may take several years, and
this activity usually is not funded through grants. Thus, academic CTUs need
to be hosted in institutions that have an understanding of the complex and
long-term nature of clinical trials, and, of course, the enormous health benefits
from the results of well-conducted trials. Most host institutions receive major
benefits from CTUs through methodological and statistical expertise, which
are often undervalued, and the long-term financial security of most academic
CTUs remains uncertain. There has been a longstanding collaboration between
industry and academic CTUs in the evaluation of new and promising cardio-
vascular drugs and devices, and many of these trials are discussed in this
book.

In the 1980s most of the methodological expertise for trials was located in
academic CTUs. During the 1990s and 2000s, increasing expertise and capac-
ity for running trials was developed in the industry sector. Contract research
organizations (CROs) work with companies developing devices and drugs to
carry out the practical aspects of clinical trials, but are usually not involved
in developing the scientific rationale for trials or publishing results in peer-
reviewed journals. The CRO market was estimated to turn over more than 20
billion dollars per annum during 2011, indicating that this is an important
and growing commercial activity [98].

The academic research organization (ARO) is an alternative model that
often employs university faculty members as clinical trial investigators. Since
they must comply with university regulatory requirements, AROs are also
likely to ensure that investigators own the right to publication of findings,
whereas publication rights for RCTs conducted by CROs usually belong to
the drug company by contract. AROs are considered to be more independent,
especially with regard to presentation of trial results and communication with
health regulatory authorities during drug approval processes [99]. Academic–
industry collaboration in trials has many advantages, including the sharing
of expertise to improve the quality of trials, but also raises issues of conflicts
of interest if academic groups or individuals stand to gain disproportionate
financial rewards in the form of direct payments, intellectual property rights,
or stock options [100–102]. There is a growing trend for universities and large
hospitals to develop long-term partnerships with industry for financial gain,
which is a sensible way forward if these partnerships lead to measurable
health improvements and are openly declared. However, the “ground rules”
for such collaborations are still in development.

What does the future hold?

Clinical trials are an essential tool of modern healthcare evaluation and the
demand for new trials both in the commercial and non-commercial sector is
growing year on year. There are major challenges ahead, in particular the
increasing bureaucratic burden of conducting clinical trials, which increases
costs and decreases efficiency and sample size [103,104]. In addition, obtain-
ing approvals for new drugs and devices is becoming more difficult, even
when trials have shown important benefits, because of a reluctance to pay for
more expensive but more effective treatments [105]. This latter issue decreases
the incentive for industry to invest in clinical trials. There is also an urgent
need to ensure that clinicians have the appropriate methodological training
and expertise to understand key design issues in clinical trials, and we have
attempted to cover many of these issues in this book with the help of a range
of expert authors.

References

1. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in
2. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of ran-
2010;340:c723.
1:421–429.
of methodological quality associated with estimates of treatment effects in controlled
oxaban versus placebo in patients with acute coronary syndromes (ATLAS ACS-TIMI
366:9–19.
Retrieved 2007-03-27


87. Data on Cardiovascular disease 2012. www.who.int
