Study Designs

LEARNING OBJECTIVES

By the end of this chapter, the reader will be able to

- List and define the components of a good study design
- Compare and contrast observational and experimental study designs
- Summarize the advantages and disadvantages of alternative study designs
- Describe the key features of a randomized controlled trial
- Identify the study designs used in public health and medical studies

Once a study objective or research question has been refined—which is no easy task, as it usually involves extensive discussion among investigators, a review of the literature, and an assessment of ethical and practical issues—the next step is to choose the study design to most effectively and efficiently answer the question. The *study design* is the methodology that is used to collect the information to address the research question. In Chapter 1, we raised a number of questions that might be of interest, including: How is the extent of a disease in a group or region quantified? How is the rate of development of a new disease estimated? How are risk factors or characteristics that might be related to the development or progression of a disease identified? How is the effectiveness of a new drug determined? To answer each of these questions, a specific study design must be selected. In this chapter, we review a number of popular study designs. This review is not meant to be exhaustive but instead illustrative of some of the more popular designs for public health applications.

The studies we present can probably be best organized into two broad types: observational and randomized studies. In *ob-servational* studies, we generally observe a phenomenon, whereas in *randomized* studies, we intervene and measure a response. Observational studies are sometimes called descriptive or associational studies, nonrandomized, or historical studies. In some cases, observational studies are used to alert the medical community to a specific issue, whereas in other instances, observational studies are used to generate hypotheses. We later elaborate on other instances where observational studies are used to assess specific associations. Randomized studies are sometimes called analytic or experimental studies. They are used to test specific hypotheses or to evaluate the effect of an intervention (e.g., a behavioral or pharmacologic intervention).

Another way to describe or distinguish study types is on the basis of the time sequence involved in data collection. Some studies are designed to collect information at a point in time, others to collect information on participants over time, and others to evaluate data that has already been collected.

In biostatistical and epidemiologic research studies, we are often interested in the association between a particular exposure or risk factor (e.g., alcohol use, smoking) and an outcome (e.g., cardiovascular disease, lung cancer). In the following sections, we discuss several observational study designs and several randomized study designs. We describe each design, detail its advantages and disadvantages, and distinguish designs by the time sequence involved. We then describe in some detail the Framingham Heart Study, which is an observational study and one of the world's most important studies of risk factors for cardiovascular disease.[1] We then provide more detail on clinical trials, which are often considered the gold standard in terms of study design. At the end of this chapter, we summarize the issues in selecting the appropriate study design. Before describing the specific design types, we

present some key vocabulary terms that are relevant to study design.

2.1   VOCABULARY

- *Bias*—A systematic error that introduces uncertainty in estimates of effect or association
- *Blind/double blind*—The state whereby a participant is unaware of their treatment status (e.g., experimental drug or placebo). A study is said to be double blind when both the participant and the outcome assessor are unaware of the treatment status (*masking* is used as an equivalent term to blinding).
- *Clinical trial*—A specific type of study involving human participants and randomization to the comparison groups
- *Cohort*—A group of participants who usually share some common characteristics and who are monitored or followed over time
- *Concurrent*—At the same time; optimally, comparison treatments are evaluated concurrently or in parallel
- *Confounding*—Complex relationships among variables that can distort relationships between the risk factors and the outcome
- *Cross-sectional*—At a single point in time
- *Incidence (of disease)*—The number of new cases (of disease) over a period of time
- *Intention-to-treat*—An analytic strategy whereby participants are analyzed in the treatment group they were assigned regardless of whether they followed the study procedures completely (e.g., regardless of whether they took all of the assigned medication)
- *Matching*—A process of organizing comparison groups by similar characteristics
- *Per protocol*—An analytic strategy whereby only participants who adhered to the study protocol (i.e., the specific procedures or treatments given to them) are analyzed (in other words, an analysis of only those assigned to a particular group who followed all procedures for that group)
- *Placebo*—An inert substance designed to look, feel, and taste like the active or experimental treatment (e.g., saline solution would be a suitable placebo for a clear, tasteless liquid medication)
- *Prevalence (of disease)*—The proportion of individuals with the condition (disease) at a single point in time
- *Prognostic factor*—A characteristic that is strongly associated with an outcome (e.g., disease) such that it could be used to reasonably predict whether a person is likely to develop a disease or not

- *Prospective*—A study in which information is collected looking forward in time
- *Protocol*—A step-by-step plan for a study that details every aspect of the study design and data collection plan
- *Quasi-experimental design*—A design in which subjects are not randomly assigned to treatments
- *Randomization*—A process by which participants are assigned to receive different treatments (this is usually based on a probability scheme)
- *Retrospective*—A study in which information is collected looking backward in time
- *Stratification*—A process whereby participants are partitioned or separated into mutually exclusive or non-overlapping groups

2.2   OBSERVATIONAL STUDY DESIGNS

There are a number of observational study designs. We describe some of the more popular designs, from the simplest to the more complex.

2.2.1   The Case Report/Case Series

A *case report* is a very detailed report of the specific features of a particular participant or case. A *case series* is a systematic review of the interesting and common features of a small collection, or series, of cases. These types of studies are important in the medical field as they have historically served to identify new diseases. The case series does not include a control or comparison group (e.g., a series of disease-free participants). These studies are relatively easy to conduct but can be criticized as they are unplanned, uncontrolled, and not designed to answer a specific research question. They are often used to generate specific hypotheses, which are then tested with other, larger studies. An example of an important case series was one published in 1981 by Gottlieb et al., who reported on five young homosexual men who sought medical care with a rare form of pneumonia and other unusual infections.[2] The initial report was followed by more series with similar presentations, and in 1982 the condition being described was termed Acquired Immune Deficiency Syndrome (AIDS).

2.2.2   The Cross-Sectional Survey

A *cross-sectional survey* is a study conducted at a single point in time. The cross-sectional survey is an appropriate design when the research question is focused on the prevalence of a disease, a present practice, or an opinion. The study is non-randomized and involves a group of participants who are identified at a point in time, and information is collected at that point in time. Cross-sectional surveys are useful for

estimating the prevalence of specific risk factors or prevalence of disease at a point in time. In some instances, it is of interest to make comparisons between groups of participants (e.g., between men and women, between participants under age 40 and those 40 and older). However, inferences from the cross-sectional survey are limited to the time at which data are collected and do not generalize to future time points.

Cross-sectional surveys can be easy to conduct, are usually ethical, and are often large in size (i.e., involve many participants) to allow for estimates of risk factors, diseases, practices, or opinions in different subgroups of interest. However, a major limitation in cross-sectional surveys is the fact that both the exposure or development of a risk factor (e.g., hypertension) and the outcome have occurred. Because the study is conducted at a point in time (see Figure 2–1), it is not possible to assess temporal relationships, specifically whether the exposure or risk factor occurred prior to the outcome of interest. Another issue is related to non-response. While a large sample may be targeted, in some situations only a small fraction of participants approached agree to participate and complete the survey. Depending on the features of the participants and non-participants, non-response can introduce bias or limit generalizability.

In Figure 2–1, approximately one-third of the participants have the risk factor and two-thirds do not. Among those with the risk factor, almost half have the disease, as compared to a much smaller fraction of those without the risk factor. Is there an association between the risk factor and the disease?

FIGURE 2–1 The Cross-Sectional Survey
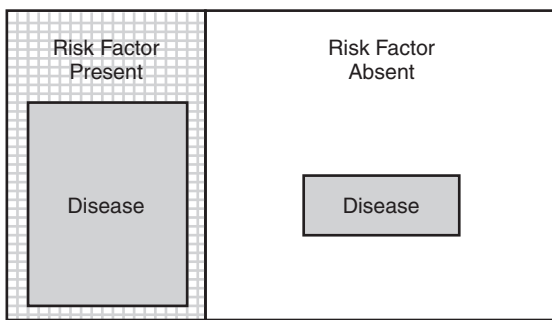
2.2.3 The Cohort Study

A *cohort study* involves a group of individuals who usually meet a set of inclusion criteria at the start of the study. The cohort is followed and associations are made between a risk factor and a disease. For example, if we are studying risk factors for cardiovascular disease, we ideally enroll a cohort of individuals free of cardiovascular disease at the start of the study. In a prospective cohort study, participants are enrolled and followed going forward in time (see Figure 2–2). In some situations, the cohort is drawn from the general population, whereas in other situations a cohort is assembled. For example, when studying the association between a relatively common risk factor and an outcome, a cohort drawn from the general population will likely include sufficient numbers of individuals who have and do not have the risk factor of interest.

When studying the association between a rare risk factor and an outcome, special attention must be paid to constructing the cohort. In this situation, investigators might want to enrich the cohort to include participants with the risk factor (sometimes called a special exposure cohort). In addition, an appropriate comparison cohort would be included. The comparison cohort would include participants free of the risk factor but similar to the exposed cohort in other important characteristics. In a *retrospective cohort study*, the exposure or risk factor status of the participants is ascertained retrospectively, or looking back in time (see Figure 2–3 and the time of study start). For example, suppose we wish to assess the association between multivitamin use and neural tube defects in newborns. We enroll a cohort of women who deliver live-born infants and ask each to report on their use of multivitamins before becoming pregnant. On the basis of these reports, we have an exposed and unexposed cohort. We then assess the outcome of pregnancy for each woman. Retrospective cohort studies are often based on data gathered from medical records where risk factors and outcomes have occurred and been documented. A study is mounted and records are reviewed to assess risk factor and outcome status, both of which have already occurred.

The prospective cohort study is the more common cohort study design. Cohort studies have a major advantage in that they allow investigators to assess temporal relationships. It is also possible to estimate the incidence of a disease (i.e., the rate at which participants who are free of a disease develop that disease). We can also compare incidence rates between groups. For example, we might compare the incidence of cardiovascular disease between participants who smoke and participants who do not smoke as a means of quantifying the association between smoking and cardiovascular disease. Cohort studies can be difficult if the outcome

FIGURE 2-2 The Prospective Cohort Study
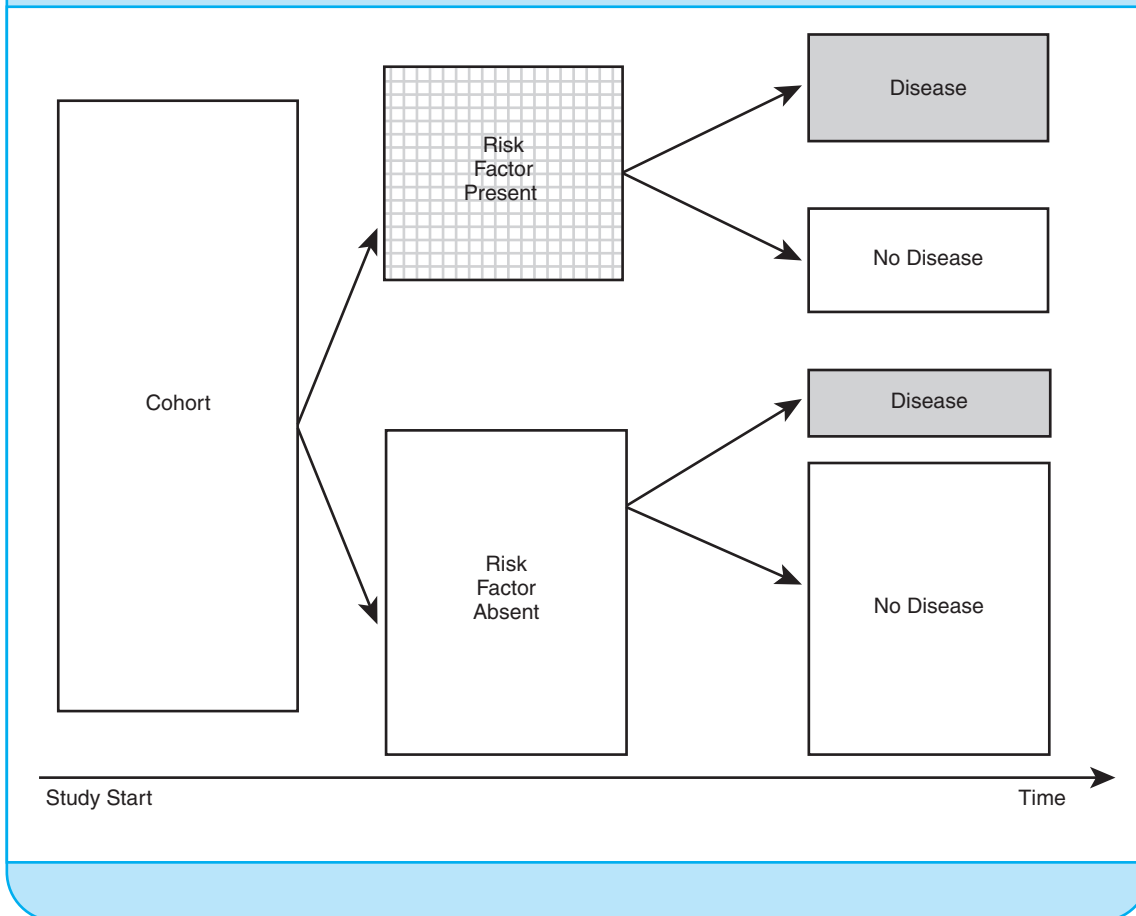
or disease under study is rare or if there is a long latency period (i.e., it takes a long time for the disease to develop or be realized). When the disease is rare, the cohort must be sufficiently large so that adequate numbers of events (cases of disease) are observed. By "adequate numbers," we mean specifically that there are sufficient numbers of events to produce stable, precise inferences employing meaningful statistical analyses. When the disease under study has a long latency period, the study must be long enough in duration so that sufficient numbers of events are observed. However, this can introduce another difficulty, namely loss of participant follow-up over a longer study period.

Cohort studies can also be complicated by confounding. *Confounding* is a distortion of the effect of an exposure or risk factor on the outcome by other characteristics. For example, suppose we wish to assess the association between smoking and cardiovascular disease. We may find that smokers in our cohort are much more likely to develop cardio-

vascular disease. However, it may also be the case that the smokers are less likely to exercise, have higher cholesterol levels, and so on. These complex relationships among the variables must be reconciled by statistical analyses. In Chapter 9, we describe in detail the methods used to handle confounding.

2.2.4   The Case-Control Study

The *case-control study* is a study often used in epidemiologic research where again the question of interest is whether there is an association between a particular risk factor or exposure and an outcome. Case-control studies are particularly useful when the outcome of interest is rare. As noted previously, cohort studies are not efficient when the outcome of interest is rare as they require large numbers of participants to be enrolled in the study to realize a sufficient number of outcome events. In a case-control study, participants are identified on the basis of their outcome status. Specifically, we select a set of *cases*, or

FIGURE 2–3 The Retrospective Cohort Study
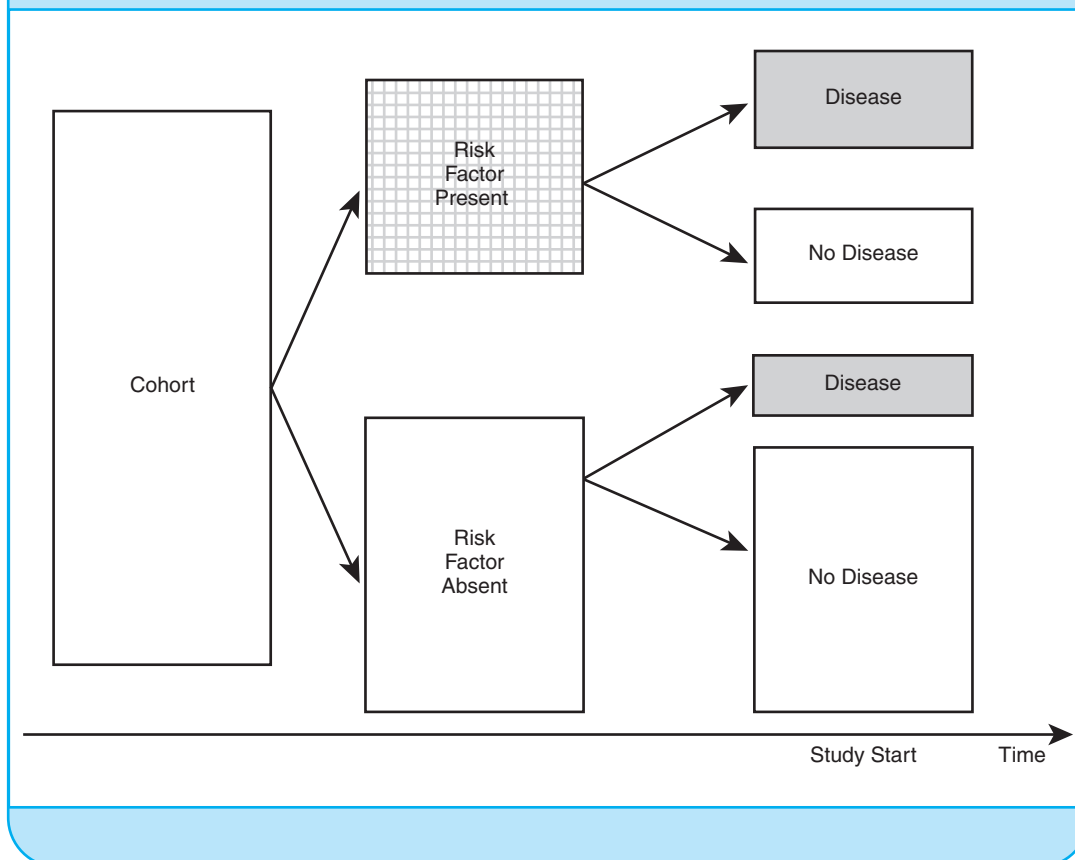
persons with the outcome of interest. We then select a set of *controls*, who are persons similar to the cases except for the fact that they are free of the outcome of interest. We then assess exposure or risk factor status retrospectively (see Figure 2–4). We hypothesize that the exposure or risk factor is related to the disease and evaluate this by comparing the cases and controls with respect to the proportions that are exposed; that is, we draw inferences about the relationship between exposure or risk factor status and disease. There are a number of important issues that must be addressed in designing case-control studies. We detail some of the most important ones.

First, cases must be selected very carefully. An explicit definition is needed to identify cases so that the cases are as homogeneous as possible. The explicit definition of a case must be established before any participants are selected or data collected. Diagnostic tests to confirm disease status should be included whenever possible to minimize the possibility of incorrect classification.

Controls must also be selected carefully. The controls should be comparable to the cases in all respects except for the fact that they do not have the disease of interest. In fact,

the controls should represent non-diseased participants who would have been included as cases if they had the disease. The same diagnostic tests used to confirm disease status in the cases should be applied to the controls to confirm non-disease status.

Usually, there are many more controls available for inclusion in a study than cases, so it is often possible to select several controls for each case, thereby increasing the sample size for analysis. Investigators have shown that taking more than four controls for each case does not substantially improve the precision of the analysis.[3] (This result will be discussed in subsequent chapters.) In many instances, two controls per case are selected, which is denoted as a 2:1 ("two to one") control to case ratio.

The next issue is to assess exposure or risk factor status, and this is done retrospectively. Because the exposure or risk factor might have occurred long ago, studies that can establish risk factor status based on documentation or records are preferred over those that rely on a participant's memory of past events. Sometimes, such data are not documented, so participants are queried with regard to risk

FIGURE 2-4 The Case-Control Study
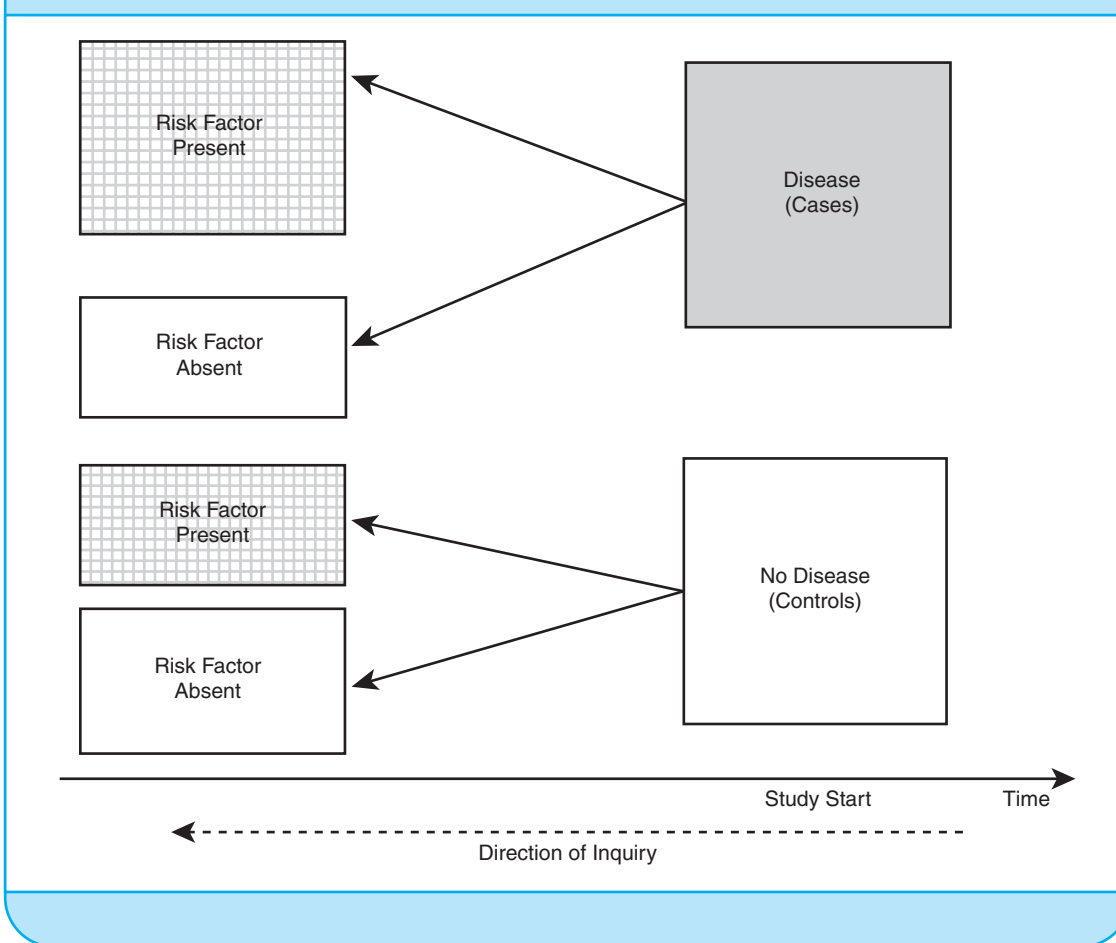
factor status. This must be done in a careful and consistent manner for all participants, regardless of their outcome status—assessment of exposure or risk factor status must be performed according to the same procedures or protocol for cases and controls. In addition, the individual collecting exposure data should not be aware of the participant's outcome status (i.e., they should be blind to whether the participant is a case or a control).

Case-control studies have several positive features. They are cost- and time-efficient for studying rare diseases. With case-control studies, an investigator can ensure that a sufficient number of cases are included. Case-control studies are also efficient when studying diseases with long latency periods. Because the study starts after the disease has been diagnosed, investigators are not waiting for the disease to occur during the study period. Case-control studies are also useful when there are several potentially harmful exposures under

consideration; data can be collected on each exposure and evaluated.

The challenges of the case-control study center mainly around bias. We discuss several of the more common sources of bias here; there are still other sources of bias to consider. *Misclassification bias* can be an issue in case-control studies and refers to the incorrect classification of outcome status (case or control) or the incorrect classification of exposure status. If misclassification occurs at random—meaning there is a similar extent of misclassification in both groups—then the association between the exposure and the outcome can be dampened (underestimated). If misclassification is not random—for example, if more cases are incorrectly classified as having the exposure or risk factor—then the association can be exaggerated (overestimated). Another source of bias is called *selection bias*, and it can result in a distortion of the association (over- or underestimation of the true association) between ex-

posure and outcome status resulting from the selection of cases and controls. Specifically, the relationship between exposure status and disease may be different in those individuals who chose to participate in the study as compared to those who did not. Yet another source of bias is called *recall bias*, and again, it can result in a distortion of the association between exposure and outcome. It occurs when cases or controls differentially recall exposure status. It is possible that persons with a disease (cases) might be more likely to recall prior exposures than persons free of the disease. The latter might not recall the same information as readily. With case-control studies, it is also not always possible to establish a temporal relationship between exposure and outcome. For example, in the present example both the exposure and outcome are measured at the time of data collection. Finally, because of the way we select participants (on the basis of their outcome status) in case-control studies, we cannot estimate incidence (i.e., the rate at which a disease develops).

2.2.5   The Nested Case-Control Study

The *nested case-control study* is a specific type of case-control study that is usually designed from a cohort study. For example, suppose a cohort study involving 1000 participants is run to assess the relationship between smoking and cardiovascular disease. In the study, suppose that 20 participants develop myocardial infarction (MI, i.e., heart attack), and we are interested in assessing whether there is a relationship between body mass index (measured as the ratio of weight in kilograms to height in meters squared) and MI. With so few participants suffering this very specific outcome, it would be difficult analytically to assess the relationship between body mass index and MI because there are a number of confounding factors that would need to be taken into account. This process generally requires large samples (specifics are discussed in Chapter 9). A nested case-control study could be designed to select suitable controls for the 20 cases that are similar to the cases except that they are free of MI. To facilitate the analysis, we would carefully select the controls and might match the controls to cases on gender, age, and other risk factors known to affect MI, such as blood pressure and cholesterol. Matching is one way of handling confounding. The analysis would then focus specifically on the association between body mass index and MI.

Nested case-control studies are also used to assess new biomarkers (measures of biological processes) or to evaluate expensive tests or technologies. For example, suppose a large cohort study is run to assess risk factors for spontaneous preterm delivery. As part of the study, pregnant women provide demographic, medical, and behavioral information through self-administered questionnaires. In addition, each woman submits a blood sample at approximately 13 weeks gestation, and the samples are frozen and stored. Each woman is followed in the study through pregnancy outcome and is classified as having a spontaneous preterm delivery or not (e.g., induced preterm delivery, term delivery, etc.). A new test is developed to measure a hormone in the mother's blood that is hypothesized to be related to spontaneous preterm delivery. A nested case-control study is designed in which women who deliver prematurely and spontaneously (cases) are matched to women who do not (controls) on the basis of maternal age, race/ethnicity, and prior history of premature delivery. The hormone is measured in each case and control using the new test applied to the stored (unfrozen) serum samples. The analysis is focused on the association between hormone levels and spontaneous preterm delivery. In this situation the nested case-control study is an efficient way to evaluate whether the risk factor (i.e., hormone) is related to the outcome (i.e., spontaneous preterm delivery). The new test is applied to only those women who are selected into the nested case-control study and not to every woman enrolled in the cohort, thereby reducing cost.

2.3   RANDOMIZED STUDY DESIGNS

Cohort and case-control studies often address the question: Is there an association between a risk factor or exposure and an outcome (e.g., a disease)? Each of these observational study designs has its advantages and disadvantages. In the cohort studies, we compare incidence between the exposed and unexposed groups, whereas in the case-control study we compare exposure between those with and without a disease. These are different comparisons, but in both scenarios, we make inferences about associations. (In Chapter 6 and Chapter 7, we detail the statistical methods used to estimate associations and to make statistical inferences.) As we described, observational studies can be subject to bias and confounding. In contrast, randomized studies are considered to be the gold standard of study designs as they minimize bias and confounding. The key feature of randomized studies is the random assignment of participants to the comparison groups. In theory, randomizing makes the groups comparable in all respects except the way the participants are treated (e.g., treated with an experimental medication or a placebo, treated with a behavioral intervention or not). We describe two popular randomized designs in detail.

2.3.1   The Randomized Controlled Trial (RCT) or Clinical Trial

The *randomized controlled trial* (RCT) is a design with a key and distinguishing feature—the randomization of participants to one of several comparison treatments or groups. In pharmaceutical trials, there are often two comparison groups;

FIGURE 2-5 The Randomized Controlled Trial
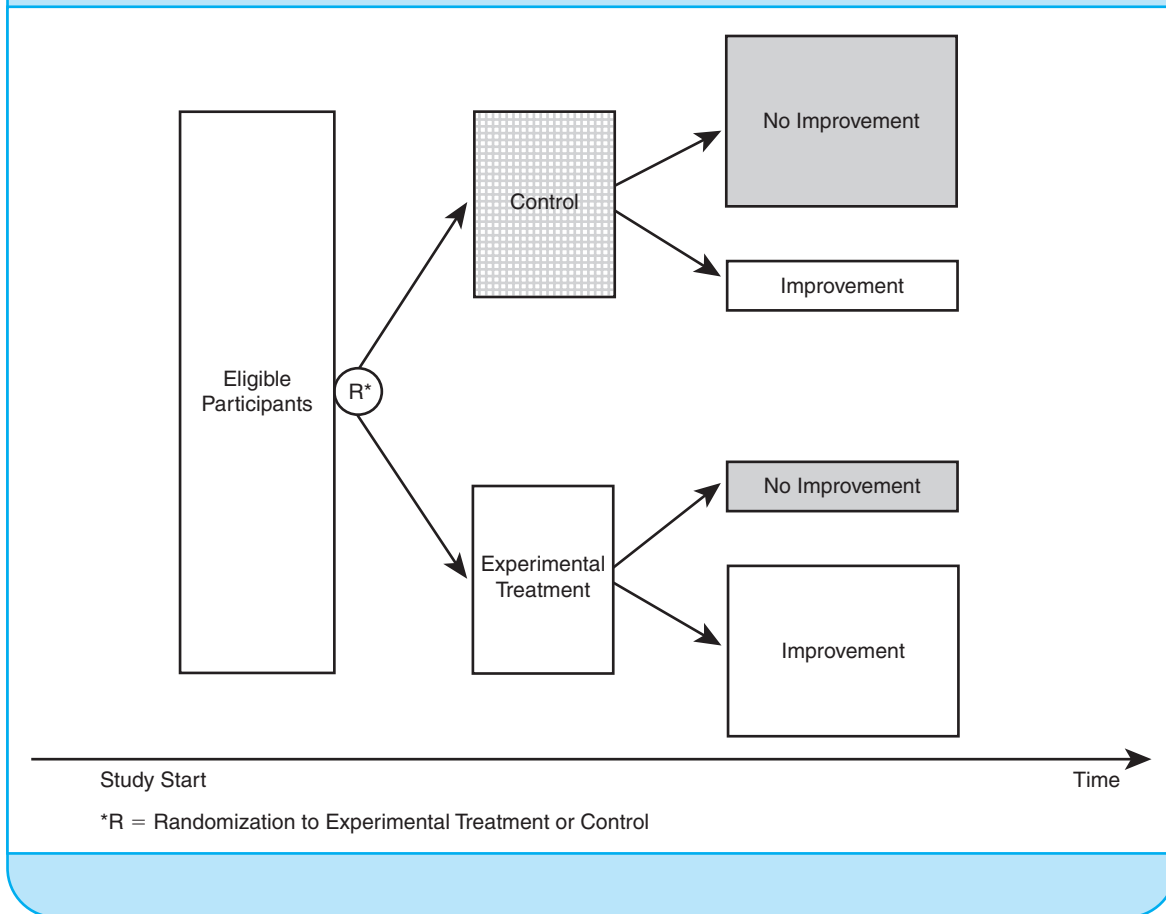
one group gets an experimental drug and the other a control drug. If ethically feasible, the control might be a placebo. If a placebo is not ethically feasible (e.g., it is ethically inappropriate to use a placebo because participants need medication), then a medication currently available and considered the standard of care is an appropriate comparator. This is called an *active-controlled trial* as opposed to a *placebo-controlled trial*. In clinical trials, data are collected prospectively (see Figure 2–5).

The idea of randomization is to balance the groups in terms of known and unknown prognostic factors (i.e., characteristics that might affect the outcome), which minimizes confounding. Because of the randomization feature, the comparison groups—in theory—differ only in the treatment received. One group receives the experimental treatment and the other does not. With randomized studies, we can make much stronger inferences than we can with observational studies. Specifically, with clinical trials, inferences are made with regard to the effect of treatments on outcomes, whereas with

observational studies, inferences are limited to associations between risk factors and outcomes.

It is important in clinical trials that the comparison treatments are evaluated concurrently. In the study depicted in Figure 2–5, the treatments are administered at the same point in time, generating parallel comparison groups. Consider a clinical trial evaluating an experimental treatment for allergies. If the experimental treatment is given during the spring and the control is administered during the winter, we might see very different results simply because allergies are highly dependent on the season or the time of year.

It is also important in clinical trials to include multiple study centers, often referred to as *multicenter trials*. The reason for including multiple centers is to promote generalizability. If a clinical trial is conducted in a single center and the experimental treatment is shown to be effective, there may be a question as to whether the same benefit would be seen in other centers. In multicenter trials, the homogeneity of the effect across centers can be analyzed directly.

Ideally, clinical trials should be double blind. Specifically, neither the investigator nor the participant should be aware of the treatment assignment. However, sometimes it is impossible or unethical to blind the participants. For example, consider a trial comparing a medical and a surgical procedure. In this situation, the participant would definitely know whether they underwent a surgical procedure. In some very rare situations, sham surgeries are performed, but these are highly unusual, as participant safety is always of the utmost concern. It is critical that the outcome assessor is blind to the treatment assignment.

There are many ways to randomize participants in clinical trials. Simple randomization involves essentially flipping a coin and assigning each participant to either the experimental or the control treatment on the basis of the coin toss. In multicenter trials, separate randomization schedules are usually developed for each center. This ensures a balance in the treatments within each center and does not allow for the possibility that all patients in one center get the same treatment. Sometimes it is important to minimize imbalance between groups with respect to other characteristics. For example, suppose we want to be sure we have participants of similar ages in each of the comparison groups. We could develop separate or stratified randomization schedules for participants less than 40 years of age and participants 40 years of age and older within each center. There are many ways to perform the randomization and the appropriate procedure depends on many factors, including the relationship between important prognostic factors and the outcome, the number of centers involved, and so on.

The major advantage of the clinical trial is that it is the cleanest design from an analytic point of view. Randomization minimizes bias and confounding so, theoretically, any benefit (or harm) that is observed can be attributed to the treatment. However, clinical trials are often expensive and very time-consuming. Clinical trials designed around outcomes that are relatively rare require large numbers of participants to demonstrate a significant effect. This increases the time and cost of conducting the trial. There are often a number of challenges in clinical trials that must be faced. First, clinical trials can be ethically challenging. Choosing the appropriate control group requires careful assessment of ethical issues. For example, in cancer trials it would never be possible to use a placebo comparator, as this would put participants at unnecessary risk. Next, clinical trials can be difficult to set up. Recruitment of centers and participants can be difficult. For example, participants might not be willing to participate in a trial because they cannot accept the possibility of being randomly assigned to the control group. Careful monitoring of participants is also a crucial aspect of clinical trials. For example, investigators must be sure that participants are taking the assigned drug as planned and are not taking other medications that might interfere with the study medications (called concomitant medications). Most clinical trials require frequent follow-up with participants—for example, every 2 weeks for 12 weeks. Investigators must work to minimize loss to follow-up to ensure that important study data are collected at every time point during the study. Subject retention and adherence to the study protocol are essential for the success of a clinical trial.

In some clinical trials, there are very strict inclusion and exclusion criteria. For example, suppose we are evaluating a new medication hypothesized to lower cholesterol. To allow the medication its best chance to demonstrate benefit, we might include only participants with very high total cholesterol levels. This means that inferences about the effect of the medication would then be limited to the population from which the participants were drawn. Clinical trials are sometimes criticized for being too narrow or restrictive. In designing trials, investigators must weigh the impact of the inclusion and exclusion criteria on the observed effects and on their generalizability.

Designing clinical trials can be very complex. There are a number of issues that need careful attention, including refining the study objective so that it is clear, concise, and answerable; determining the appropriate participants for the trial (detailing inclusion and exclusion criteria explicitly); determining the appropriate outcome variable; deciding on the appropriate control group; developing and implementing a strict monitoring plan; determining the number of participants to enroll; and detailing the randomization plan. While achieving these goals is challenging, a successful randomized clinical trial is considered the best means of establishing the effectiveness of a medical treatment.

2.3.2 The Crossover Trial

The *crossover trial* is a clinical trial where each participant is assigned to two or more treatments sequentially. When there are two treatments (e.g., an experimental and a control), each participant receives both treatments. For example, half of the participants are randomly assigned to receive the experimental treatment first and then the control; the other half receive the control first and then the experimental treatment. Outcomes are assessed following the administration of each treatment in each participant (see Figure 2–6). Participants receive the randomly assigned treatment in Period 1. The outcome of interest is then recorded for the Period 1 treatment. In most crossover trials, there is then what is a called a *wash-out period* where no treatments are given. The wash-out period is included so that any

FIGURE 2–6 The Crossover Trial
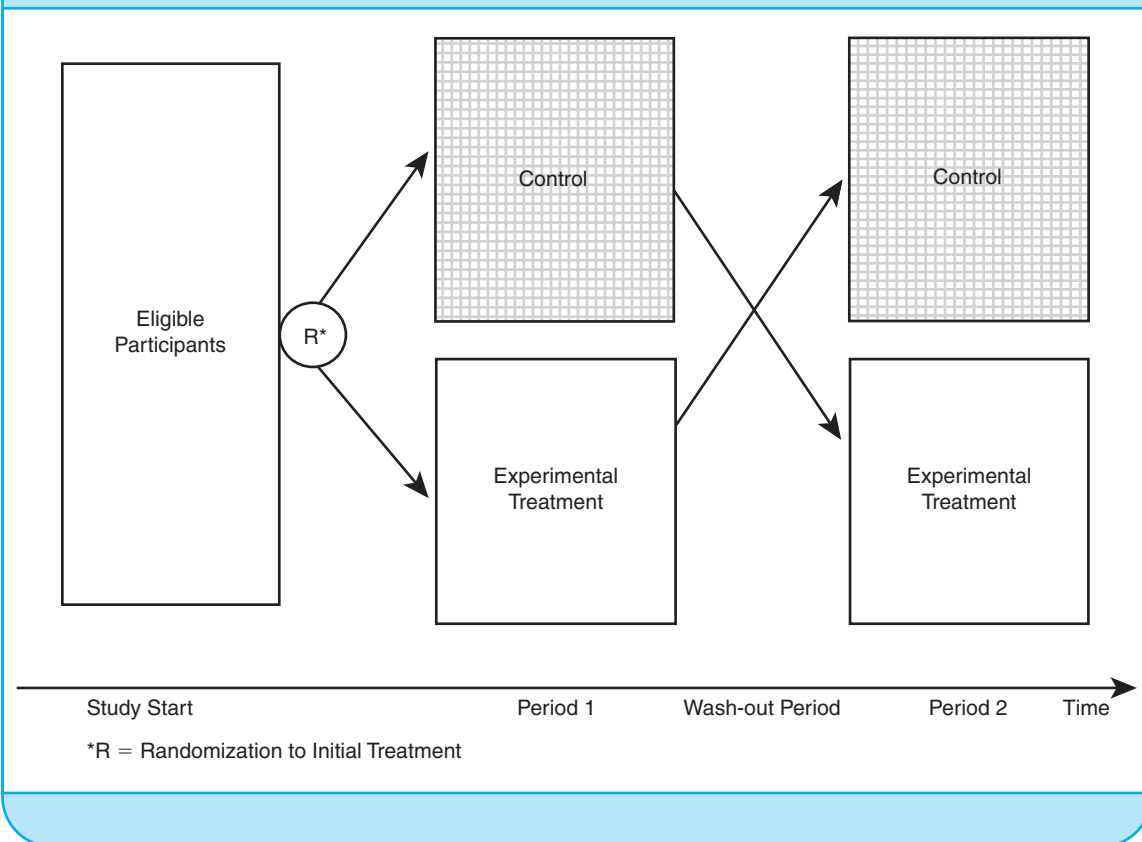
therapeutic effects of the first treatment are removed prior to the administration of the second treatment in Period 2. In a trial with an experimental and a control treatment, participants who received the control treatment during Period 1 receive the experimental treatment in Period 2 and vice versa.

There are several ways in which participants can be assigned to treatments in a crossover trial. The two most popular schemes are called *random* and *fixed* assignment. In the random assignment scheme (already mentioned), participants are randomly assigned to the experimental treatment or the control in Period 1. Participants are then assigned the other treatment in Period 2. In a fixed assignment strategy, all participants are assigned the same treatment sequence. For example, everyone gets the experimental treatment first, followed by the control treatment or vice versa. There is an issue with the fixed scheme in that investigators must assume that the outcome observed on the second treatment (and subsequent treatments, if there are more than two) would be equivalent to the outcome that would be observed if that treatment were assigned first (i.e., that there are no carry-over effects). Randomly

varying the order in which the treatments are given allows the investigators to assess whether there is any order effect.

The major advantage to the crossover trial is that each participant acts as their own control; therefore, we do not need to worry about the issue of treatment groups being comparable with respect to baseline characteristics. In this study design, fewer participants are required to demonstrate an effect. A disadvantage is that there may be carry-over effects such that the outcome assessed following the second treatment is affected by the first treatment. Investigators must be careful to include a wash-out period that is sufficiently long to minimize carry-over effects. A participant in Period 2 may not be at the same baseline as they were in Period 1, thus destroying the advantage of the crossover. In this situation, the only useful data may be from Period 1. The wash-out period must be short enough so that participants remain committed to completing the trial. Because participants in a crossover trial receive each treatment, loss to follow-up or dropout is critical because losing one participant means losing outcome data on both treatments.

Crossover trials are best suited for short-term treatments of chronic, relatively stable conditions. A crossover trial would not be efficient for diseases that have acute flare-ups because these could influence the outcomes that are observed yet have nothing to do with treatment. Crossover trials are also not suitable for studies with death or another serious condition considered as the outcome.

Similar to the clinical trial described previously, adherence or compliance to the study protocol and study medication in the crossover trial is critical. Participants are more likely to skip medication or drop out of a trial if the treatment is unpleasant or if the protocol is long or difficult to follow. Every effort must be made on the part of the investigators to maximize adherence and to minimize loss to follow-up.

2.4   THE FRAMINGHAM HEART STUDY

We now describe one of the world's most well-known studies of risk factors for cardiovascular disease. The Framingham Heart Study started in 1948 with the enrollment of a cohort of just over 5000 individuals free of cardiovascular disease who were living in the town of Framingham, Massachusetts.[1] The Framingham Heart Study is a longitudinal cohort study that involves repeated assessments of the participants approximately every two years. The study celebrated its fiftieth anniversary in 1998 and it still continues today. The original cohort has been assessed nearly 30 times. At each assessment, complete physical examinations are conducted (e.g., vital signs, blood pressure, medication history), blood samples are taken to measure lipid levels and novel risk factors, and participants also have echocardiograms in addition to other assessments of cardiovascular functioning. In the early 1970s, approximately 5000 offspring of the original cohort and their spouses were enrolled into what is called the Framingham Offspring cohort (the second generation of the original cohort). These participants have been followed approximately every four years and have been assessed over eight times. In the early 2000s, a third generation of over 4000 participants was enrolled and are being followed approximately every four years.

Over the past 50 years, hundreds of papers have been published from the Framingham Heart Study identifying important risk factors for cardiovascular disease, such as smoking, blood pressure, cholesterol, physical inactivity, and diabetes. The Framingham Heart Study also identified risk factors for stroke, heart failure, and peripheral artery disease. Researchers have identified psychosocial risk factors for heart disease, and now, with three generations of participants in the Framingham Study, investigators are assessing genetic risk factors for obesity, diabetes, and cardiovascular disease. More details on the Framingham Heart Study, its design, investigators,

research milestones, and publications can be found at *http://www.nhlbi.nih.go v/about/framingham* and at *http://www.bu.edu/alumni/bostonia/2005/summer/pdfs/heart.pdf*.

2.5   MORE ON CLINICAL TRIALS

Clinical trials are extremely important, particularly in medical research. In Section 2.3, we outlined clinical trials from a design standpoint, but there are many more aspects of clinical trials that should be mentioned. First, clinical trials must be conducted at the correct time in the course of history. For example, suppose we ask the research question: Is the polio vaccine necessary today? To test this hypothesis, a clinical trial could be initiated in which some children receive the vaccine while others do not. The trial would not be feasible today because it would be unethical to withhold the vaccine from some children. No one would risk the consequences of the disease to study whether the vaccine is necessary.

As noted previously, the design of a clinical trial is extremely important to ensure the generalizability and validity of the results. Well-designed clinical trials are very easy to analyze, whereas poorly designed trials are extremely difficult, sometimes impossible, to analyze. The issues that must be considered in designing clinical trials are outlined here. Some have been previously identified but are worth repeating.

The number of treatments involved. If there are two treatments involved, statistical analyses are straightforward because only one comparison is necessary. If more than two treatments are involved, then more complicated statistical analyses are required and the issue of multiple comparisons must be addressed (these issues are discussed in Chapter 7 and Chapter 9). The number of treatments involved in a clinical trial should always be based on clinical criteria and not be reduced to simplify statistical analysis.

The control treatment. In clinical trials, an experimental (or newly developed) treatment is compared against a control treatment. The control treatment may be a treatment that is currently in use and considered the standard of care, or the control treatment may be a placebo. If a standard treatment exists, it should be used as the control because it would be unethical to offer patients a placebo when a conventional treatment is available. (While clinical trials are considered the gold standard design to evaluate the effectiveness of an experimental treatment, there are instances where a control group is not available. Techniques to evaluate effectiveness in the absence of a control group are described in D'Agostino and Kwan.[4])

Outcome measures. The outcome or outcomes of interest must be clearly identified in the design phase of the clinical trial. The primary outcome is the one specified in the planned

analysis and is used to determine the sample size required for the trial (this is discussed in detail in Chapter 8). The primary outcome are usually more objective than subjective in nature. It is appropriate to specify secondary outcomes, and results based on secondary outcomes should be reported as such. Analyses of secondary outcomes can provide important information and, in some cases, enough evidence for a follow-up trial in which the secondary outcomes become the primary outcomes.

Blinding. Blinding refers to the fact that patients are not aware of which treatment (experimental or control) they are receiving in the clinical trial. A *single blind* trial is one in which the investigator knows which treatment a patient is receiving but the patient does not. *Double blinding* refers to the situation in which both the patient and the investigator are not aware of which treatment is assigned. In many clinical trials, only the statistician knows which treatment is assigned to each patient.

Single-center versus multicenter trials. Some clinical trials are conducted at a single site or clinical center, whereas others are conducted—usually simultaneously—at several centers. There are advantages to including several centers, such as increased generalizability and an increased number of available patients. There are also disadvantages to including multiple centers, such as needing more resources to manage the trial and the introduction of center-specific characteristics (e.g., expertise of personnel, availability or condition of medical equipment, specific characteristics of participants) that could affect the observed outcomes.

Randomization. Randomization is a critical component of clinical trials. There are a number of randomization strategies that might be implemented in a given trial. The exact strategy depends on the specific details of the study protocol.

Sample size. The number of patients required in a clinical trial depends on the variation in the primary outcome and the expected difference in outcomes between the treated and control patients.

Population and sampling. The study population should be explicitly defined by the study investigators (patient inclusion and exclusion criteria). A strategy for patient recruitment must be carefully determined and a system for checking inclusion and exclusion criteria for each potential enrollee must be developed and followed.

Ethics. Ethical issues often drive the design and conduct of clinical trials. There are some ethical issues that are common to all clinical trials, such as the safety of the treatments involved. There are other issues that relate only to certain trials. Most institutions have Institutional Review Boards (IRBs) that are responsible for approving research study protocols. Research protocols are evaluated on the basis of scientific accuracy and with respect to potential risks and benefits to participants. All participants in clinical trials must provide informed consent, usually on consent forms approved by the appropriate IRB.

Protocols. Each clinical trial should have a protocol, which is a manual of operations or procedures in which every aspect of the trial is clearly defined. The protocol details all aspects of subject enrollment, treatment assignment, data collection, monitoring, data management, and statistical analysis. The protocol ensures consistency in the conduct of the trial and is particularly important when a trial is conducted at several clinical centers (i.e., in a multicenter trial).

Monitoring. Monitoring is a critical aspect of all clinical trials. Specifically, participants are monitored with regard to their adherence to all aspects of the study protocol (e.g., attending all scheduled visits, completing study assessments, taking the prescribed medications or treatments). Participants are also carefully monitored for any side effects or adverse events. Protocol violations (e.g., missing scheduled visits) are summarized at the completion of a trial, as are the frequencies of adverse events and side effects.

Data management. Data management is a critical part of any study and is particularly important in clinical trials. Data management includes tracking subjects (ensuring that subjects complete each aspect of the trial on time), data entry, quality control (examining data for out-of-range values or inconsistencies), data cleaning, and constructing analytic databases. In most studies, a data manager is assigned to supervise all aspects of data management.

The statistical analysis in a well-designed clinical trial is straightforward. Assuming there are two treatments involved (an experimental treatment and a control), there are essentially three phases of analysis:

- Baseline comparisons, in which the participants assigned to the experimental treatment group are compared to the patients assigned to the control group with respect to relevant characteristics measured at baseline. These analyses are used to check that the randomization is successful in generating balanced groups.
- Crude analysis, in which outcomes are compared between patients assigned to the experimental and control treatments. In the case of a continuous outcome (e.g., weight), the difference in means is estimated; in the case of a dichotomous outcome (e.g., development of disease or not), relative risks are estimated; and in the case of time-to-event data (e.g., time to a heart attack), survival curves are estimated. (The specifics of these analyses are discussed in detail in Chapters 6, 7, 10, and 11.)

• Adjusted analyses are then performed, similar to the crude analysis, which incorporate important covariates (i.e., variables that are associated with the outcome) and confounding variables. (The specifics of statistical adjustment are discussed in detail in Chapters 9 and 11.)

There are several analytic samples considered in statistical analysis of clinical trials data. The first is the Intent to Treat (ITT) analysis sample. It includes all patients who were randomized. The second is the Per Protocol analysis sample, and it includes only patients who completed the treatment (i.e., followed the treatment protocol as designed). The third is the Safety analysis sample, and it includes all patients who took at least one dose of the assigned treatment even if they did not complete the treatment protocol. All aspects of the design, conduct, and analysis of a clinical trial should be carefully documented. Complete and accurate records of the clinical trial are essential for applications to the Food and Drug Administration (FDA).[5]

Clinical trials are focused on safety and efficacy. Safety is assessed by the nature and extent of adverse events and side effects. Adverse events may or may not be due to the drug being evaluated. In most clinical trials, clinicians indicate whether the adverse event is likely due to the drug or not. Efficacy is assessed by improvements in symptoms or other aspects of the indication or disease that the drug is designed to address.

There are several important stages in clinical trials. Preclinical studies are studies of safety and efficacy in animals. Clinical studies are studies of safety and efficacy in humans. There are three phases of clinical studies, described here.

Phase I: First Time in Humans Study. The main objectives in a Phase I study are to assess the toxicology and safety of the proposed treatment in humans and to assess the pharmacokinetics (how fast the drug is absorbed in, flows through, and is secreted from the body) of the proposed treatment. Phase I studies are not generally focused on efficacy (how well the treatment works); instead, safety is the focus. Phase I studies usually involve 10 to 15 patients, and many Phase I studies are performed in healthy, normal volunteers to assess side effects and adverse events. In Phase I studies, one goal is to determine the maximum tolerated dose (MTD) of the proposed drug in humans. Investigators start with very low doses and work up to higher doses. Investigations usually start with three patients, and three patients are added for each elevated dose. Data are collected at each stage to assess safety, and some Phase I studies are placebo-controlled. Usually, two or three separate Phase I studies are conducted.

Phase II: Feasibility or Dose-Finding Study. The focus of a Phase II study is still on safety, but of primary interest are side effects and adverse events (which may or may not be directly related to the drug). Another objective in the Phase II study is efficacy, but the efficacy of the drug is based on descriptive analyses in the Phase II study. In some cases, investigators do not know which specific aspects of the indication or disease the drug may affect or which outcome measure best captures this effect. Usually, investigators measure an array of outcomes to determine the best outcome for the next phase. In Phase II studies, investigators determine the optimal dosage of the drug with respect to efficacy (e.g., lower doses might be just as effective as the MTD). Phase II studies usually involve 50 to 100 patients who have the indication or disease of interest. Phase II studies are usually placebo-controlled or compared to a standard, currently available treatment. Subjects are randomized and studies are generally double blind. If a Phase II study indicates that the drug is safe but not effective, investigation cycles back to Phase I. Most Phase II studies proceed to Phase III based on observed safety and efficacy.

Phase III: Confirmatory Clinical Trial. The focus of the Phase III trial is efficacy, although data are also collected to monitor safety. Phase III trials are designed and executed to confirm the effect of the experimental treatment. Phase III trials usually involve two treatment groups, an experimental treatment at the determined optimal dose and a placebo or standard of care. Some Phase III trials involve three groups: placebo, standard of care, and experimental treatment. Sample sizes can range from 200 to 500 patients, depending on what is determined to be a clinically significant effect. (The exact number is determined by specific calculations that are described in Chapter 8.) At least two successful clinical trials performed by independent investigators at different clinical centers are required in Phase III studies to assess whether the effect of the treatment can be replicated by independent investigators in at least two different sets of participants. More details on the design and analysis of clinical trials can be found in Chow and Liu.[6]

Investigators need positive results (statistically proven efficacy) in at least two separate trials to submit an FDA application for drug approval. The FDA also requires clinical significance in two trials, with clinical significance specified by clinical investigators in the design phase when the number of subjects is determined (see Chapter 8).

The FDA New Drug Application (NDA) contains a summary of results of Phase I, Phase II, and Phase III studies. The FDA reviews an NDA within six months to one year after submission and grants approval or not. If a drug is approved, the sponsor may conduct Phase IV trials, also called postmarketing trials, that can be retrospective (e.g., based on medical record review) or prospective (e.g., a clinical trial involving many patients to study rare adverse events). These

studies are often undertaken to understand the long-term effects (efficacy and safety) of the drug.

2.6   SAMPLE SIZE IMPLICATIONS

Biostatisticians have a critical role in designing studies, not only to work with investigators to select the most efficient design to address the study hypotheses but also to determine the appropriate number of participants to involve in the study. In Chapter 8, we provide formulas to compute the sample sizes needed to appropriately answer research questions. The sample size needed depends on the study design, the anticipated association between the risk factor and outcome or the effect of the drug (e.g., the difference between the experimental and control drugs) and also on the statistical analysis that will be used to answer the study questions. The sample size should not be too small such that an answer about the association or the effect of the drug under investigation is not possible, because in this instance, both participants and the investigators have wasted time and money. Alternatively, a sample size should not be too large because again time and money would be wasted but, in addition, participants may be placed at unnecessary risk. Both scenarios are unacceptable from an ethical standpoint, and therefore careful attention must be paid when determining the appropriate sample size for any study or trial.

2.7   SUMMARY

To determine which study design is most efficient for a specific application, investigators must have a specific, clearly defined research question. It is also important to understand current knowledge or research on the topic under investigation. The most efficient design depends on the expected association or effect, the prevalence or incidence of outcomes, the prevalence of risk factors or exposures, and the expected duration of the study. Also important are practical issues, costs, and—most importantly—ethical issues.

Choosing the appropriate study design to address a research question is critical. Whenever possible, prior to mounting a planned study, investigators should try to run a pilot or feasibility study, which is a smaller-scale version of the planned study, as a means to identify potential problems and issues. Whereas pilot studies can be time-consuming and costly, they are usually more than worthwhile.

2.8   PRACTICE PROBLEMS

1. An investigator wants to assess whether smoking is a risk factor for pancreatic cancer. Electronic medical records at a local hospital will be used to identify 50 patients with pancreatic cancer. One hundred patients who are similar but free of pancreatic cancer will also be selected. Each participant's medical record will be analyzed for smoking history. Identify the type of study proposed and indicate its specific strengths and weaknesses.

2. What is the most likely source of bias in the study described in Problem 1?

3. An investigator wants to assess whether the use of a specific medication given to infants born prematurely is associated with developmental delay. Fifty infants who were given the medication and 50 comparison infants who were also born prematurely but not given the medication will be selected for the analysis. Each infant will undergo extensive testing at age 2 for various aspects of development. Identify the type of study proposed and indicate its specific strengths and weaknesses.

4. Is bias or confounding more of an issue in the study described in Problem 3? Give an example of a potential source of bias and a potential confounding factor.

5. A study is planned to assess the effect of a new surgical intervention for gall bladder disease. One hundred patients with gall bladder disease will be randomly assigned to receive either the new surgical intervention or the standard surgical intervention. The efficacy of the new surgical intervention will be measured by the time a patient takes to return to normal activities, recorded in days. Identify the type of study proposed and indicate its specific strengths and weaknesses.

6. An investigator wants to assess the association between caffeine consumption and impaired glucose tolerance, a precursor to diabetes. A study is planned to include 70 participants. Each participant will be surveyed with regard to their daily caffeine consumption. In addition, each participant will submit a blood sample that will be used to measure their glucose level. Identify the type of study proposed and indicate its specific strengths and weaknesses.

7. Could the study described in Problem 6 be designed as a randomized clinical trial? If so, briefly outline the study design; if not, describe the barriers.

8. A study is planned to compare two weight-loss programs in patients who are obese. The first program is based on restricted caloric intake and the second is based on specific food combinations. The study will involve 20 participants and each participant will follow each program. The programs will be assigned in random order (i.e., some participants will first follow the restricted-calorie diet and then follow the food-combination diet, whereas others will first follow the food-combination diet and then follow the restricted-calorie

diet). The number of pounds lost will be compared between diets. Identify the type of study proposed and indicate its specific strengths and weaknesses.

9. An orthopedic surgeon observes that many of his patients coming in for total knee replacement surgery played organized sports before the age of 10. He plans to collect more extensive data on participation in organized sports from four patients undergoing knee replacement surgery and to report the findings. Identify the type of study proposed and indicate its specific strengths and weaknesses.

10. Suggest an alternative design to address the hypothesis in Problem 9. What are the major issues in addressing this hypothesis?

11. In 1940, 2000 women working in a factory were recruited into a study. Half of the women worked in manufacturing and half in administrative offices. The incidence of bone cancer through 1970 among the 1000 women working in manufacturing was compared with that of the 1000 women working in administrative offices. Thirty of the women in manufacturing developed bone cancer as compared to 9 of the women in administrative offices. This study is an example of a
   a. randomized controlled trial
   b. case-control study
   c. cohort study
   d. crossover trial

12. An investigator reviewed the medical records of 200 children seen for care at Boston Medical Center in the past year who were between the ages of 8 and 12 years old, and identified 40 with asthma. He also identified 40 children of the same ages who were free of asthma. Each child and their family were interviewed to assess whether there might be an association between certain environmental factors, such as exposure to second-hand smoke, and asthma. This study is an example of a
   a. randomized controlled trial
   b. case-control study
   c. cohort study
   d. crossover trial

13. A study is designed to evaluate the impact of a daily multivitamin on students' academic performance. One hundred sixty students are randomly assigned to receive either the multivitamin or a placebo and are instructed to take the assigned drug daily for 20 days.

On day 20, each student takes a standardized exam and the mean exam scores are compared between groups. This study is an example of a
   a. randomized controlled trial
   b. case-control study
   c. cohort study
   d. crossover trial

14. A study is performed to assess whether there is an association between exposure to second-hand cigarette smoke in infancy and delayed development. Fifty children with delayed development and 50 children with normal development are selected for investigation. Parents are asked whether their children were exposed to second-hand cigarette smoke in infancy or not. This study is an example of a
   a. prospective cohort study
   b. retrospective cohort study
   c. case-control study
   d. clinical trial

15. A study is planned to investigate risk factors for sudden cardiac death. A cohort of men and women between the ages of 35 and 70 are enrolled and followed for up to 20 years. As part of the study, participants provide data on demographic and behavioral characteristics; they also undergo testing for cardiac function and provide blood samples to assess lipid profiles and other biomarkers. A new measure of inflammation is hypothesized to be related to sudden cardiac death. What study design is most appropriate to assess the association between the new biomarker and sudden cardiac death? Describe its strengths and weaknesses.

REFERENCES

1. D'Agostino, R.B. and Kannel, W.B. "Epidemiological background and design: The Framingham Study." *Proceedings of the American Statistical Association, Sesquicentennial Invited Paper Sessions*, 1989: 707–719.
2. Gottlieb, M.S., Schroff, R., Scganker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A., and Saxon, A. "*Pneumocystis carinii* pneumonia and *mucosal candidiasis* in previously healthy homosexual men: Evidence of a new acquired cellular immunodeficiency." *New England Journal of Medicine* 1981; 305(24): 1425–1431.
3. Schelesselman, J.J. *Case-Control Studies: Design, Conduct, Analysis*. New York: Oxford University Press, 1982.
4. D'Agostino, R.B. and Kwan, H. "Measuring effectiveness: What to expect without a randomized control group." *Medical Care* 1995; 33(4 Suppl.): AS95–105.
5. United States Food and Drug Administration. Available at *http://www.fda.gov*.
6. Chow, S.C. and Lie, J.P. *Design and Analysis of Clinical Trials: Concepts and Methodologies*. New York: John Wiley & Sons, 1998.

© Jones & Bartlett Learning, LLC. NOT FOR SALE OR DISTRIBUTION.