

Chapter 2. Study Designs in Medical Research

Key Concepts

-  Study designs in medicine fall into two categories: studies in which subjects are observed, and studies in which the effect of an intervention is observed.
-  Observational studies may be forward-looking (cohort), backward-looking (case-control), or looking at simultaneous events (cross-sectional). Cohort studies generally provide stronger evidence than the other two designs.
-  Studies that examine patient outcomes are increasingly published in the literature; they focus on specific topics, such as resource utilization, functional status, quality of life, patient satisfaction, and cost-effectiveness.
-  Studies with interventions are called experiments or clinical trials. They provide stronger evidence than observational studies.
-  The single best way to minimize bias is to randomly select subjects in observational studies or randomly assign subjects to different treatment arms in clinical trials.
-  Bias occurs when the way a study is designed or carried out causes an error in the results and conclusions. Bias can be due to the manner in which subjects are selected or data are collected and analyzed.
-  Clinical trials without controls (subjects who do not receive the intervention) are difficult to interpret and do not provide strong evidence.
-  Each study design has specific advantages and disadvantages.

Study Designs in Medical Research: Introduction

This chapter introduces the different kinds of studies commonly used in medical research. Because we believe that knowing how a study is designed is important for understanding the conclusions that can be drawn from it, we have chosen to devote considerable attention to the topic of study designs.

If you are familiar with the medical literature, you will recognize many of the terms used to describe different study designs. If you are just beginning to read the literature, you should not be dismayed by all the new terminology; there will be ample opportunity to review and become familiar with it. Also, the glossary at the end of the book defines the terms we use here. In the final chapter of this book, study designs are reviewed within the context of reading journal articles, and pointers are given on how to look for possible biases that can occur in medical studies. Bias can be due to the manner in which patients are selected, data are collected and analyzed, or conclusions are drawn.

Classification of Study Designs



There are several different schemes for classifying study designs. We have adopted one that divides studies into those in which the subjects were merely observed, sometimes called **observational studies**, and those in which some intervention was performed, generally called **experiments**. This approach is simple and reflects the sequence an investigation sometimes takes. With a little practice, you should be able to read medical articles and classify studies according to the outline in [Table 2–1](#) with little difficulty.

Table 2–1. Classification of Study Designs.

I. Observational studies
A. Descriptive or case–series
B. Case–control studies (retrospective)
1. Causes and incidence of disease
2. Identification of risk factors
C. Cross-sectional studies, surveys (prevalence)
1. Disease description
2. Diagnosis and staging
3. Disease processes, mechanisms
D. Cohort studies (prospective)
1. Causes and incidence of disease
2. Natural history, prognosis
3. Identification of risk factors
E. Historical cohort studies
II. Experimental studies
A. Controlled trials
1. Parallel or concurrent controls
a. Randomized
b. Not randomized
2. Sequential controls
a. Self-controlled
b. Crossover
3. External controls (including historical)
B. Studies with no controls
III. Meta-analyses

Each study design in [Table 2–1](#) is illustrated in this chapter, using some of the studies that are presenting problems in upcoming chapters. In observational studies, one or more groups of patients are observed, and characteristics about the patients are recorded for analysis. Experimental studies involve an **intervention**—an investigator-controlled maneuver, such as a drug, a procedure, or a treatment—and interest lies in the effect the intervention has on study subjects. Of course, both observational and experimental studies may involve animals or objects, but most studies in medicine (and the ones discussed most frequently in this text) involve people.

Observational Studies



Observational studies are of four main types: case–series, case–control, cross-sectional (including surveys), and cohort studies. When certain characteristics of a group (or series) of patients (or cases) are described in a published report, the result is called a **case–series study**; it is the simplest design in which the author describes some interesting or intriguing observations that occurred for a small number of patients.

Case–series studies frequently lead to the generation of hypotheses that are subsequently investigated in a case–control, cross-sectional, or cohort study. These three types of studies are defined by the period of time the study covers and by the direction or focus of the research question. Cohort and case–control studies generally involve an extended period of time defined by the point when the study begins and the point when it ends; some process occurs, and a certain amount of time is required to assess it. For this reason, both cohort and case–control studies are sometimes also called **longitudinal studies**. The major difference between them is the direction of the inquiry or the focus of the research question: Cohort studies are forward-looking, from a risk factor to an outcome, whereas case–control studies are backward-looking, from an outcome to risk factors. The cross-sectional study analyzes data collected on a group of subjects at one time. [Kleinbaum and colleagues \(1997\)](#) describe a number of hybrids or combinations of these designs if you are interested in more detail than we give in this chapter. If you would like a more detailed discussion of study designs used in medicine, see the companion text on epidemiology by [Greenberg and coworkers \(2000\)](#). A book by [Hulley and Cummings \(2001\)](#) is devoted entirely to the design of clinical research. [Garb \(1996\)](#) and [Burns and Grove \(2002\)](#) discuss study design in medicine and nursing, respectively.

Case–Series Studies

A case–series report is a simple descriptive account of interesting characteristics observed in a group of patients. For example, [Alexandrov and coworkers \(1997\)](#) presented information on a series of 40 patients who had been referred for evaluation of stroke, transient ischemic attack, or carotid bruit. The authors wanted to compare two methods to see which better predicted peak systolic velocity. They concluded that the relationship between both methods and peak systolic velocity was very strong.

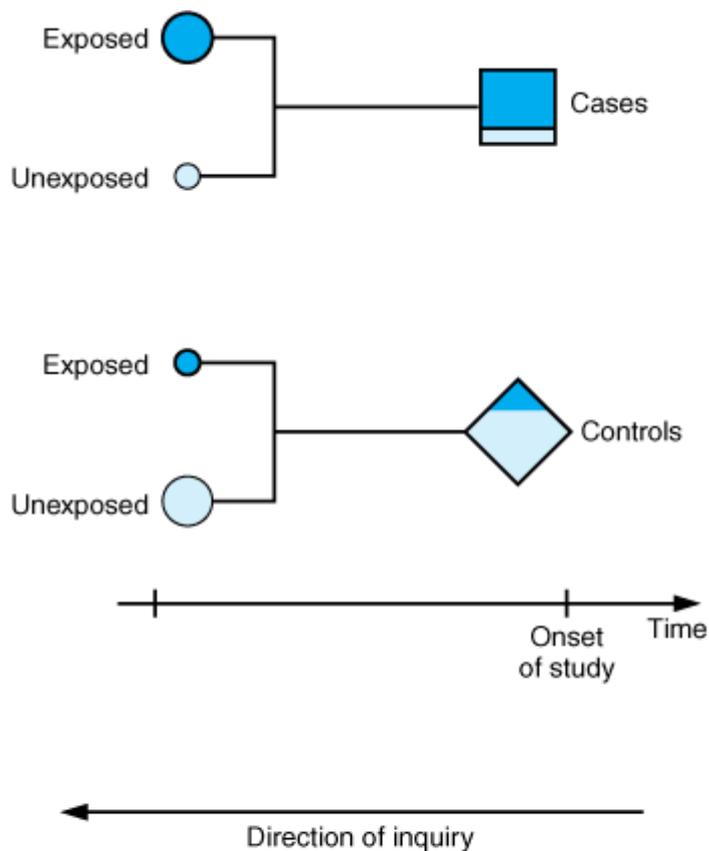
Case–series reports generally involve patients seen over a relatively short time. Generally case–series studies do not include **control subjects**, persons who do not have the disease or condition being described. Some investigators would not include case–series in a list of types of studies because they are generally not planned studies and do not involve any research hypotheses. On occasion, however, investigators do include control subjects. We mention case–series studies because of their important descriptive role as a precursor to other studies.

Case–Control Studies

Case–control studies begin with the absence or presence of an outcome and then look backward in time to try to detect possible causes or risk factors that may have been suggested in a case–series report. The *cases* in case–control studies are individuals selected on the basis of some disease or outcome; the *controls* are individuals without the disease or outcome. The history or previous events of both cases and controls are analyzed in an attempt to identify a characteristic or risk factor present in the cases' histories but not in the controls' histories.

[Figure 2–1](#) illustrates that subjects in the study are chosen at the onset of the study after they are known to be either cases with the disease or outcome (squares) or controls without the disease or outcome (diamonds). The histories of cases and controls are examined over a previous period to detect the presence (shaded areas) or absence (unshaded areas) of predisposing characteristics or risk factors, or, if the disease is infectious, whether the subject has been exposed to the presumed infectious agent. In case–control designs, the nature of the inquiry is backward in time, as indicated by the arrows pointing backward in [Figure 2–1](#) to illustrate the backward, or retrospective, nature of the research process. We can characterize case–control studies as studies that ask “What happened?” In fact, they are sometimes called **retrospective studies** because of the direction of inquiry. Case–control studies are longitudinal as well, because the inquiry covers a period of time.

FIGURE 2–1.



Question: "What happened?"

Source: Dawson B, Trapp R.G: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of case–control study design. Shaded areas represent subjects exposed to the antecedent factor; unshaded areas correspond to unexposed subjects. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest. (Adapted and reproduced, with permission, from Greenberg RS: Retrospective studies. In Kotz S, Johnson NL [editors]: *Encyclopedia of Statistical Sciences*, Vol 8. Wiley, 1988.)

[Olsen and colleagues \(2003\)](#) compared patients who had a surgical site infection following laminectomy or spinal fusion (cases) with patients who developed no infection (controls). The investigators found that length of hospital stay and readmission rates were greater with patients with infections. Furthermore, postoperative incontinence was one of the risk factors associated with the development of infection.

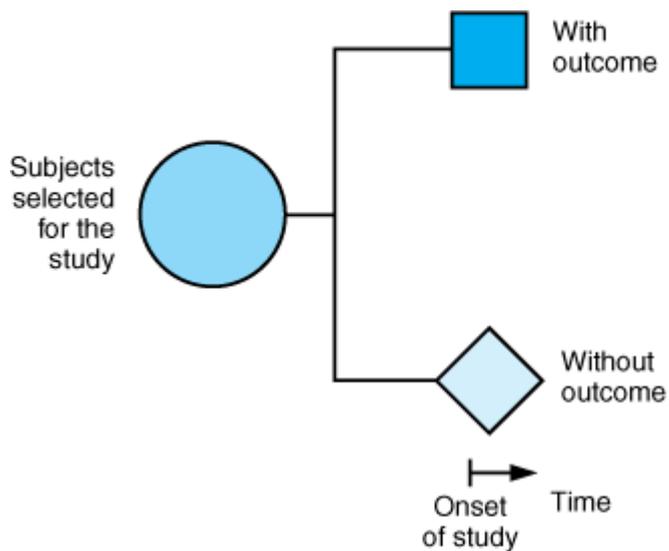
Investigators sometimes use **matching** to associate controls with cases on characteristics such as age and sex. If an investigator feels that such characteristics are so important that an imbalance between the two groups of patients would affect any conclusions, he or she should employ matching. This process ensures that both groups will be similar with respect to important characteristics that may otherwise cloud or confound the conclusions.

Deciding whether a published study is a case–control study or a case–series report is not always easy. Confusion arises because both types of studies are generally conceived and written after the fact rather than having been planned. The easiest way to differentiate between them is to ask whether the author’s purpose was to describe a phenomenon or to attempt to explain it by evaluating previous events. If the purpose is simple description, chances are the study is a case–series report.

Cross-Sectional Studies

The third type of observational study goes by all of the following names: cross-sectional studies, surveys, epidemiologic studies, and prevalence studies. We use the term “cross-sectional” because it is descriptive of the time line and does not have the connotation that the terms “surveys” and “prevalence” do. **Cross-sectional studies** analyze data collected on a group of subjects at one time rather than over a period of time. Cross-sectional studies are designed to determine “What is happening?” right now. Subjects are selected and information is obtained in a short period of time ([Figure 2–2](#); note the short time line). Because they focus on a point in time, they are sometimes also called **prevalence** studies. Surveys and polls are generally cross-sectional studies, although surveys can be part of a cohort or case–control study. Cross-sectional studies may be designed to address research questions raised by a case–series, or they may be done without a previous descriptive study.

FIGURE 2–2.



No direction of inquiry

Question: "What is happening?"

Source: Dawson B, Trapp R.G: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of cross-sectional study design. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest.

Diagnosing or Staging a Disease

In a presenting problem in [Chapter 10, Soderstrom and his coinvestigators \(1997\)](#) were interested in learning more about the relationship between demographic measures that might be helpful in identifying trauma patients who have an elevated blood alcohol concentration. They wanted to develop a simple scoring system that could be used to detect these patients when they come to an emergency department. These patients could be targeted for assessment of alcohol abuse and dependence and other possible substance abuse. They chose to look at the time of day (day or night), the day of the week (weekday or weekend), race (white or nonwhite), and age (40 years or older versus younger than 40). Using these four simple measures, the investigators were able to construct four models: for men whose injury was intentional, men whose injury was not intentional, women whose injury was intentional, and women whose injury was not intentional.

Evaluating Different Methods of Doing the Same Thing

A presenting problem in [Chapter 5](#) is a cross-sectional study designed to examine the relationship between histology slides and magnetic resonance imaging (MRI) to study characteristics of diseased carotid arteries ([Yuan et al, 2001](#)). The histology slides were evaluated by a pathologist who was **blinded** to the imaging results. It is important to establish the level of agreement between the MRI findings and histology, and the level of agreement was found to be relatively high. Cross-sectional studies are used in all fields of medicine, but they are especially common in examinations of the usefulness of a new diagnostic procedure.

Establishing Norms

Knowledge of the range within which most patients fit is very useful to clinicians. Laboratories, of course, establish and then provide the normal limits of most diagnostic tests when they report the results for a given patient. Often these limits are established by testing people who are known to have normal values. We would not, for example, want to use people with diabetes mellitus to establish the norms for serum glucose levels. The results from the people known to have normal values are used to define the range that separates the lowest 2½% of the values and the highest 2½% of the values from the middle 95%. These values are called normal values, or norms.

Outside of the laboratory there are many qualities for which normal ranges have not been established. This was true for two measures of the autoimmune nervous system function. These two measures, heart variation to deep breathing and the Valsalva ratio, are noninvasive tests that can help clinicians evaluate patients with diabetes mellitus and other neuropathic disorders. [Gelber and colleagues \(1997\)](#) analyzed data from subjects recruited from 63 centers throughout North America to develop normative values for these two measurements. After comparing certain demographic groups, such as males versus females, the investigators established the normative values for heart rate variation to deep breathing and the Valsalva.

Surveys

Surveys are especially useful when the goal is to gain insight into a perplexing topic or to learn how people think and feel about an issue. Surveys are generally cross-sectional in design, but they can be used in case-control and cohort studies as well.

[Caiola and Litaker \(2000\)](#) wanted to know the factors that influence fellows to select a specific general internal residency fellowship program. Because they did not know

the names and addresses of the fellows, the authors sent a questionnaire to the program directors and asked them to distribute the questionnaires to the fellows. We examine this study in more detail in [Chapter 11](#) and illustrate how the authors asked the questions on the survey.

Many times investigators use preexisting surveys rather than creating their own, especially if good questionnaires already exist. [Patenaude and colleagues \(2003\)](#) asked medical students at a Canadian medical school to complete a questionnaire on moral reasoning (the Kohlberg Moral Judgment Interview). They wanted to learn how moral reasoning progressed over time, so they gave the questionnaire at the beginning of medical school and again at the end of the third year. They learned that the stage of moral development did not change in about 70% of the students, whereas it either decreased or increased in 15%. The authors had expected the level of moral reasoning to increase, and the results of the study prompted them to raise questions about the possible features of medical education that might inhibit its development.

Interviews are sometimes used in surveys, especially when it is important to probe reasons or explanations more deeply than is possible with a written questionnaire. [Kendler and colleagues \(2003\)](#) wanted to investigate the role of genetic and environmental risk factors for substance abuse. They studied six classes of illicit substances to learn whether substance use disorders are substance-specific. After interviewing almost 1200 sets of adult male twins, they concluded that environmental experiences unique to a given individual are primarily responsible for whether the person misuses one class of psychoactive substances over another. Increasingly, surveys are performed using existing databases of information. As an illustration, [Huang and Stafford \(2002\)](#) used survey data from the National Ambulatory Medical Care Survey to examine the relationship between demographics and clinical characteristics of women who visit primary care physicians and specialists for urinary tract infection. Using preexisting databases can have a number of advantages, such as saving time and effort, but many national surveys use complicated designs; and it is important to know what these are, as we discuss when we explore this study in more detail in [Chapter 11](#).

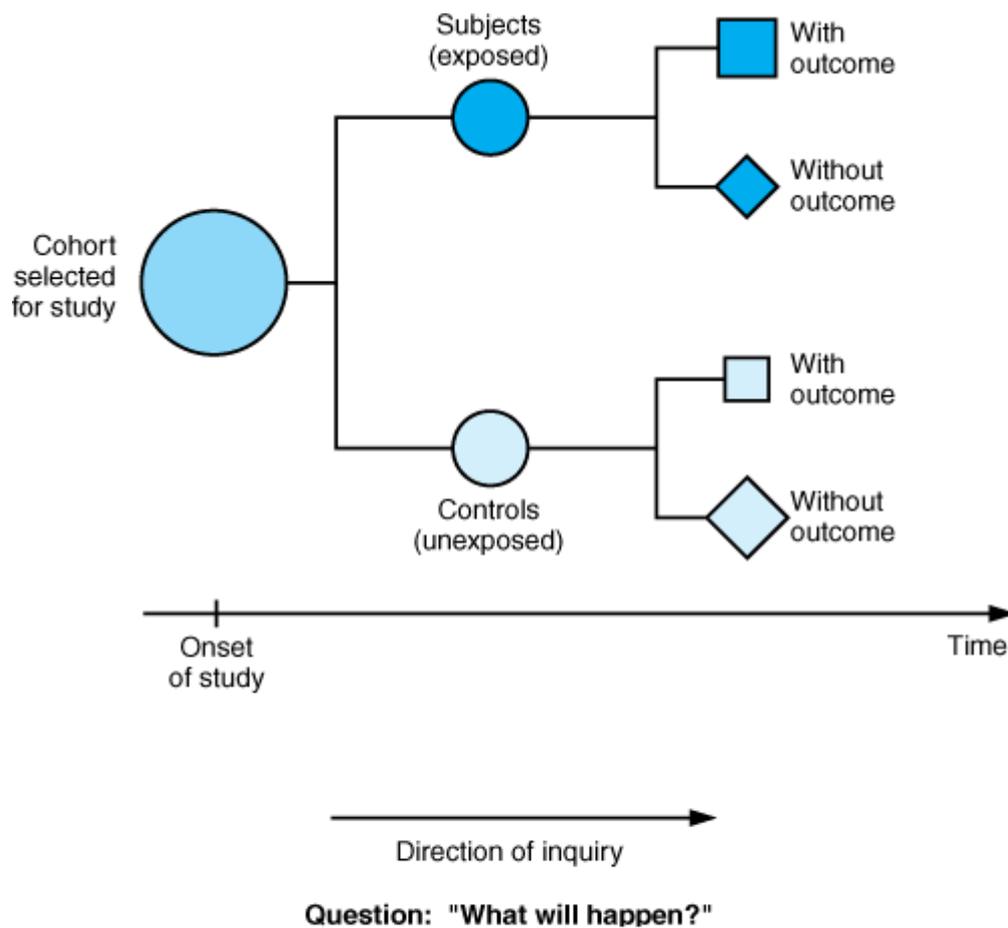
Many countries and states collect data on a variety of conditions to develop tumor registries and databases of cases of infectious disease. [Diermayer and colleagues \(1999\)](#), a presenting problem in [Chapter 4](#), analyzed epidemiologic surveillance data from the State of Oregon and reported an increase in the overall incidence rate of meningococcal disease from 2 cases/100,000 population during 1987–1992 to 4.5

cases/100,000 in 1994. Epidemiologists from Oregon and the Centers for Disease Control in Atlanta, Georgia, wanted to know if the increased number of cases of meningococcal disease indicated a transition from endemic to epidemic disease. They also sought these other features of an epidemic: the predominance of a single bacterial strain rather than a heterogeneous mix of strains and a shift in age distribution of cases toward older age groups.

Cohort Studies

A **cohort** is a group of people who have something in common and who remain part of a group over an extended time. In medicine, the subjects in **cohort studies** are selected by some defining characteristic (or characteristics) suspected of being a precursor to or risk factor for a disease or health effect. Cohort studies ask the question “What will happen?” and thus, the direction in cohort studies is forward in time. [Figure 2–3](#) illustrates the study design. Researchers select subjects at the onset of the study and then determine whether they have the risk factor or have been exposed. All subjects are followed over a certain period to observe the effect of the risk factor or exposure. Because the events of interest transpire after the study is begun, these studies are sometimes called **prospective studies**.

FIGURE 2–3.



Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of cohort study design. Shaded areas represent subjects exposed to the antecedent factor; unshaded areas correspond to unexposed subjects. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest. (Adapted and reproduced, with permission, from Greenberg RS: Prospective studies. In Kotz S, Johnson NL [editors]: *Encyclopedia of Statistical Sciences*, Vol 7. Wiley, 1986.)

Typical Cohort Studies

A classical cohort study with which most of you are probably familiar is the Framingham study of cardiovascular disease. This study was begun in 1948 to investigate factors associated with the development of atherosclerotic and hypertensive cardiovascular disease, for which [Gordon and Kannel \(1970\)](#) reported a comprehensive 20-year follow-up. More than 6000 citizens in Framingham, Massachusetts, agreed to participate in this long-term study that involved follow-up

interviews and physical examinations every 2 years. Many journal articles have been written about this cohort, and some of the children of the original subjects are now being followed as well.

Cohort studies often examine what happens to the disease over time—the natural history of the disease. Many studies have been based on the Framingham cohort; hundreds of journal articles are indexed by **MEDLINE**. Many studies deal with cardiovascular-related conditions for which the study was designed, such as blood pressure and pulse pressure as predictors of congestive heart failure ([Haider et al, 2003](#)), but this very rich source of data is being used to study many other conditions as well. For instance, two recent articles examined the life expectancy of adults who are obese ([Peeters et al, 2003](#)) and the relation of bone mass to development of prostate cancer ([Zhang et al, 2002](#)).

Although the Framingham Heart Study is very long term, many cohort studies follow subjects for a much shorter period. A presenting problem in [Chapter 5](#) describes a cohort study to determine the effect of cholecystectomy on bowel habits and bile acid absorption ([Sauter et al, 2002](#)). Fifty-one patients undergoing cholecystectomy were evaluated before, 1 month after, and 3 months after surgery to detect changes such as abdominal pain, flatulence, and dyspepsia.

Outcome Assessment



Increasingly, studies that assess medical outcomes are reported in the medical literature. Patient outcomes have always been of interest to health care providers; physicians and others in the health field are interested in how patients respond to different therapies and management regimens. There continues to be a growing focus on the ways in which patients view and value their health, the care they receive, and the results or outcomes of this care. The reasons for the increase in patient-focused health outcomes are complex, and some of the major ones are discussed later in this chapter. [Kane \(1997\)](#) provides information on reading outcomes research articles.

Interest in outcome assessment was spurred by the Medical Outcomes Study (MOS), designed to determine whether variations in patient outcomes were related to the system of care, clinician specialty, and the technical and interpersonal skill of the clinician ([Tarlov et al, 1989](#)). Many subsequent studies looked at variations in outcomes in different geographic locations or among different ethnic groups that might result from access issues. In a cross-sectional study, [Santora and colleagues \(2003\)](#) studied variations in breast cancer screening among primary care clinicians

by geographic location. They found that written breast cancer guidelines were used less in suburban and urban areas than in rural areas. [Lurie and colleagues \(2003\)](#) reported over five-fold variation in rates of advanced spinal imaging across geographic areas. Different rates of spinal imaging, in turn, accounted for a significant proportion of geographic variation in spine surgery. Other studies focus on variation in resource use among different medical specialties and systems of health care. Specific focus on the health care organizations reported that poor and elderly patients with chronic illnesses had worse outcomes in health maintenance organizations (HMO) systems than with fee-for-service systems and recommended that health care plans carefully monitor patient outcomes ([Ware et al, 1995](#)). There are many kinds of patient outcomes: economic, functional status, and satisfaction, among others.

Functional status refers to a person's ability to perform his or her daily activities. Some researchers subdivide functional status into physical, emotional, mental, and social components ([Gold et al, 1996](#)). The 6-min walk test (how far a person can walk in 6 min) was studied by [Enright and colleagues \(2003\)](#), and they recommended that the standards be adjusted for age, gender, height, and weight. Many instruments used to measure physical functional status have been developed to evaluate the extent of a patient's rehabilitation following injury or illness. These instruments are commonly called measures of activities of daily living (ADL). [Kretser and colleagues \(2003\)](#) used the activities of daily living (ADL) to compare with models of nutritional intervention. Subjects eligible for Meals-on-Wheels were randomized to receive either the traditional program of five hot meals per week, or a new program of three meals and two snacks every day of the week. The group receiving the new program gained significantly more weight from baseline at both the 3-month and 6-month measurements.

Quality of life (QOL) is a broadly defined concept that includes subjective or objective judgments about all aspects of an individual's existence: health, economic status, environmental, and spiritual. Interest in measuring QOL was heightened when researchers realized that living a long time does not necessarily imply living a good life. QOL measures can help determine a patient's preferences for different health states and are often used to help decide among alternative approaches to medical management ([Wilson and Cleary, 1995](#)).

Patient satisfaction has been discussed for many years and has been shown to be highly associated with whether patients remain with the same physician provider and the degree to which they adhere to their treatment plan ([Weingarten et al, 1995](#)).

Patient satisfaction with medical care is influenced by a number of factors, not all of which are directly related to quality of care. Examples include time spent in the office waiting for the doctor and waiting for resolution after being seen; ease of access to the doctor, including phone contact; appointment desk activity; parking; building directions; waiting room setting; and friendliness of the staff in general ([Lledo et al., 1995](#)).

Cost-effectiveness and **cost–benefit analysis** are methods used to evaluate economic outcomes of interventions or different modes of treatment. [Brown \(2002\)](#), a [Chapter 12](#) presenting problem, investigated the costs and benefits of housing policy strategies to prevent childhood lead poisoning. Using standard methods, she compared the number of children identified with lead poisoning where limited building code enforcement occurred with children living where enforcement was strict. She found that children living in the former environment had a four-fold increase in lead poisoning and that \$46,000 could be saved per building if these structures were brought into compliance. Cost-effectiveness analysis gives policy makers and health providers critical data needed to make informed judgments about interventions ([Gold et al., 1996](#)). A large number of questionnaires or instruments have been developed to measure outcomes. For quality of life, the most commonly used general-purpose instrument is the Medical Outcomes Study MOS 36-Item Short-Form Health Survey (SF-36). Originally developed at the RAND Corporation ([Stewart et al., 1988](#)), a refinement of the instrument has been validated and is now used worldwide to provide baseline measures and to monitor the results of medical care. The SF-36 provides a way to collect valid data and does not require very much time to complete. The 36 items are combined to produce a patient profile on eight concepts in addition to summary physical and mental health measures. Another instrument that focuses specifically on QOL is the EuroQol Questionnaire developed and widely used in Europe and the UK ([Kind, 1996](#)).

Many instruments are problem-specific. [Cramer and Spilker \(1998\)](#) provide a broad overview of approaches to QOL assessment, evaluations of outcomes, and pharmacoeconomic methods—both general purpose and disease-specific.

Some outcome studies address a whole host of topics, and we have used several as presenting problems in upcoming chapters. As efforts continue to contain costs of medical care while maintaining a high level of patient care, we expect to see many additional studies focusing on patient outcomes. The journal *Medical Care* is devoted exclusively to outcome studies.

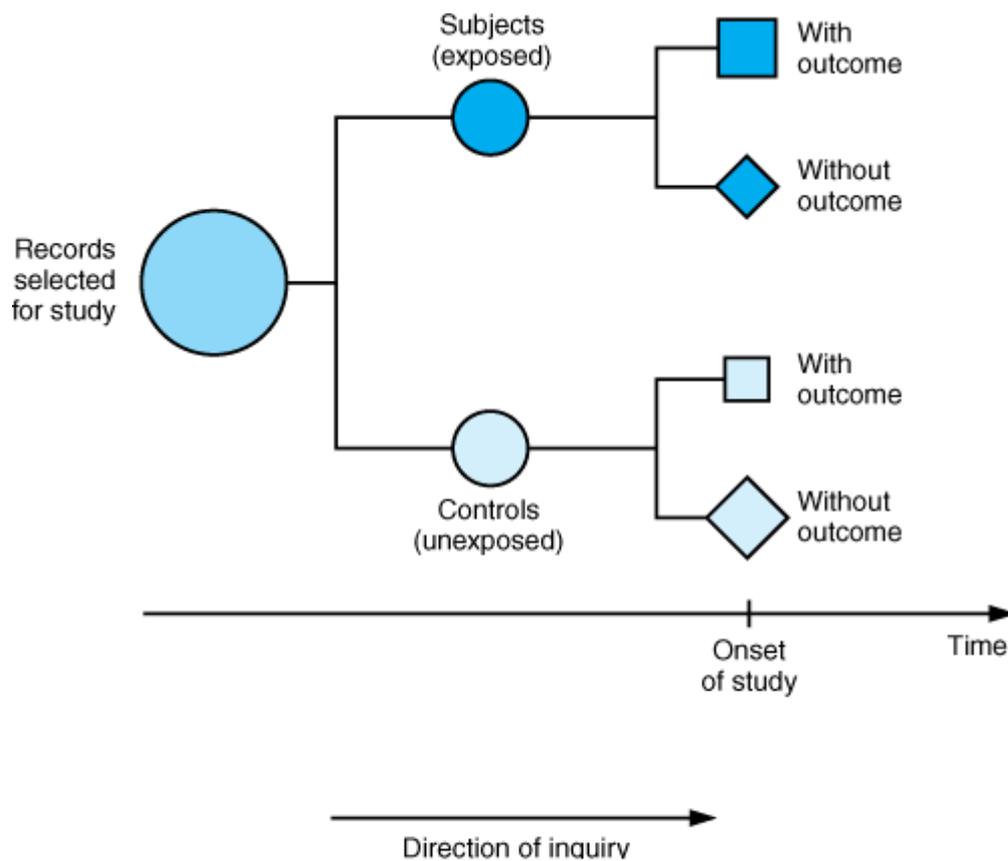
Historical Cohort Studies

Many cohort studies are prospective; that is, they begin at a specific time, the presence or absence of the risk factor is determined, and then information about the outcome of interest is collected at some future time, as in the two studies described earlier. One can also undertake a cohort study by using information collected in the past and kept in records or files.

For example, [Shiple and his coinvestigators \(1999\)](#) wanted to assess study outcomes in men with prostate cancer treated with a specific type of radiation therapy (see [Chapter 4](#)). Six medical centers had consistently followed a group of patients who had previously been treated with this therapy. Shipley used existing records to look at survival and tumor recurrence in 1607 men who were treated between 1988 and 1995 and had had at least four prostate-specific antigen measurements after radiation. This approach to a study is possible if the records on follow-up are complete and adequately detailed and if the investigators can ascertain the current status of the patients.

Some investigators call this type of study a **historical cohort study** or **retrospective cohort study** because historical information is used; that is, the events being evaluated actually occurred before the onset of the study ([Figure 2-4](#)). Note that the direction of the inquiry is still forward in time, from a possible cause or risk factor to an outcome. Studies that merely describe an investigator's experience with a group of patients and attempt to identify features associated with a good or bad outcome fall into this category, and many such studies are published in the medical literature.

FIGURE 2-4.



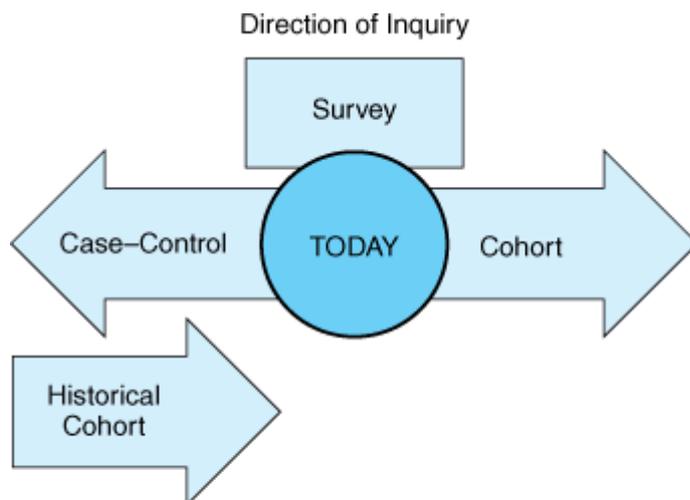
Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of historical cohort study design. Shaded areas represent subjects exposed to the antecedent factor; unshaded areas correspond to unexposed subjects. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest. (Adapted and reproduced, with permission, from Greenberg RS: Retrospective studies. In Kotz S, Johnson NL [editors]: *Encyclopedia of Statistical Sciences*, Vol. 8. Wiley, 1988.)

The time relationship among the different observation study designs is illustrated in [Figure 2-5](#). The figure shows the timing of surveys, which have no direction of inquiry, case-control designs, which look backward in time, and cohort studies, which look forward in time.

FIGURE 2–5.



Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of the time relationship among different observational study designs. The arrows represent the direction of the inquiry.

Comparison of Case–Control and Cohort Studies

Both case–control and cohort studies evaluate risks and causes of disease, and the design an investigator selects depends in part on the research question.

[Henderson and colleagues \(1997\)](#) undertook a cohort study to look at the risk factors for depression in the elderly. After an initial interview to collect information on potential risk factors, the investigators reinterviewed the subjects 3–6 years later to reassess their status. The investigators could have designed a case–control study had they asked the research question as: “Among elderly people exhibiting dementia or cognitive decline, what are the likely precursors or risk factors?” They would need to ascertain the patients’ mental status in the past and any other potential reasons that might be associated with their present condition. As this illustration shows, a cohort study starts with a risk factor or exposure and looks at consequences; a case–control study takes the outcome as the starting point of the inquiry and looks for precursors or risk factors.

Generally speaking, results from a well-designed cohort study carry more weight in understanding a disease than do results from a case–control study. A large number of possible biasing factors can play a role in case–control studies, and several of them are discussed at greater length in [Chapter 13](#).

In spite of their shortcomings with respect to establishing causality, case–control studies are frequently used in medicine and can provide useful insights if well designed. They can be completed in a much shorter time than cohort studies and are correspondingly less expensive to undertake. Case–control studies are especially useful for studying rare conditions or diseases that may not manifest themselves for many years. In addition, they are valuable for testing an original premise; if the results of the case–control study are promising, the investigator can design and undertake a more involved cohort study.

Experimental Studies or Clinical Trials



Experimental studies are generally easier to identify than observational studies in the medical literature. Authors of medical journal articles reporting experimental studies tend to state explicitly the type of study design used more often than do authors reporting observational studies. Experimental studies in medicine that involve humans are called **clinical trials** because their purpose is to draw conclusions about a particular procedure or treatment. [Table 2–1](#) indicates that clinical trials fall into two categories: those with and those without controls.

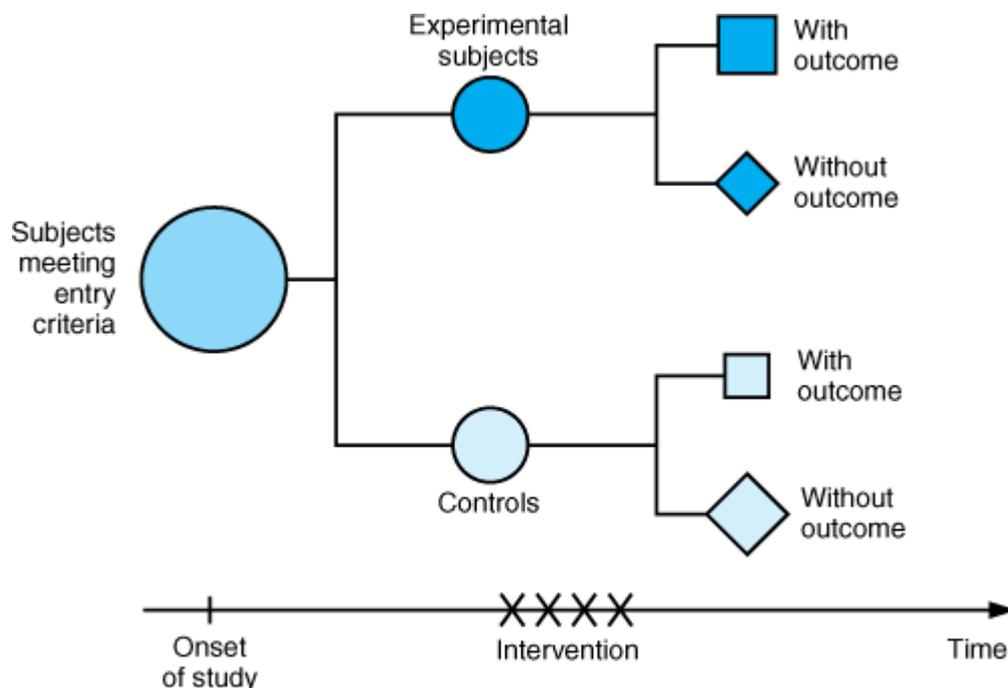
Controlled trials are studies in which the experimental drug or procedure is compared with another drug or procedure, sometimes a placebo and sometimes the previously accepted treatment. Uncontrolled trials are studies in which the investigators' experience with the experimental drug or procedure is described, but the treatment is not compared with another treatment, at least not formally. Because the purpose of an experiment is to determine whether the intervention (treatment) makes a difference, studies with controls are much more likely than those without controls to detect whether the difference is due to the experimental treatment or to some other factor. Thus, controlled studies are viewed as having far greater validity in medicine than uncontrolled studies. The consolidated standard of reporting trials (CONSORT) guidelines reflect an effort to improve the reporting of clinical trials. A comprehensive discussion and illustration of the standard is given by [Altman and colleagues \(2001\)](#).

Trials with Independent Concurrent Controls

One way a trial can be controlled is to have two groups of subjects: one that receives the experimental procedure (the experimental group) and the other that receives the placebo or standard procedure (the control group) ([Figure 2–6](#)). The experimental and control groups should be treated alike in all ways except for the procedure itself so that any differences between the groups will be due to the procedure and not to

other factors. The best way to ensure that the groups are treated similarly is to plan interventions for both groups for the same time period in the same study. In this way, the study achieves **concurrent control**. To reduce the chances that subjects or investigators see what they expect to see, researchers can design **double-blind trials** in which neither subjects nor investigators know whether the subject is in the treatment or the control group. When only the subject is unaware, the study is called a **blind trial**. In some unusual situations, the study design may call for the investigator to be blinded even when the subject cannot be blinded. Blindedness is discussed in detail in [Chapter 13](#). Another issue is how to assign some patients to the experimental condition and others to the control condition; the best method of assignment is random assignment. Methods for randomization are discussed in [Chapter 4](#).

FIGURE 2–6.



Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition; <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of randomized controlled trial design. Shaded areas represent subjects assigned to the treatment condition; unshaded areas correspond to subjects assigned to the control condition. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest.

Randomized Controlled Trials



The **randomized controlled trial** is the epitome of all research designs because it provides the strongest evidence for concluding causation; it provides the best insurance that the result was due to the intervention.

One of the more noteworthy randomized trials is the Physicians' Health Study ([Steering Committee of the Physicians' Health Study Research Group, 1989](#)), which investigated the role of aspirin in reducing the risk of cardiovascular disease. One purpose was to learn whether aspirin in low doses reduces the mortality rate from cardiovascular disease. Participants in this clinical trial were over 22,000 healthy male physicians who were randomly assigned to receive aspirin or placebo and were followed over an average period of 60 months. The investigators found that fewer physicians in the aspirin group experienced a myocardial infarction during the course of the study than did physicians in the group receiving placebo. We discuss several randomized trials as presenting problems. For instance, [Borghi and colleagues \(2002\)](#) compared a traditional low-calcium diet with a diet containing a normal amount of calcium but reduced amount of animal protein and salt for the prevention of recurrent kidney stone formation. The primary outcome was the time to the first recurrence of a symptomatic or presence of a radiographically identified stone. Results indicated that a diet with a normal amount of calcium but reduced animal protein and salt is more effective than the traditional low-calcium diet in reducing the risk of recurrent stones in men with hypercalciuria.

Nonrandomized Trials



Subjects are not always randomized to treatment options. Studies that do not use randomized assignment are generally referred to as **nonrandomized trials** or simply as clinical trials or comparative studies, with no mention of randomization. Many investigators believe that studies with nonrandomized controls are open to so many sources of bias that their conclusions are highly questionable. Studies using nonrandomized controls are considered to be much weaker because they do nothing to prevent bias in patient assignment. For instance, perhaps it is the stronger patients who receive the more aggressive treatment and the higher risk patients who are treated conservatively. An example is a nonrandomized study of the use of a paracervical block to diminish cramping and pain associated with cryosurgery for cervical neoplasia ([Harper, 1997](#); [Chapter 6](#) presenting problem). This investigator enrolled the first 40 women who met the inclusion criteria in the group treated in the usual manner (no anesthetic block before cryosurgery) and enrolled the next 45 women in the group receiving the paracervical block. This design is not as subject to bias as a study in which patients are treated without regard to any plan; however, it

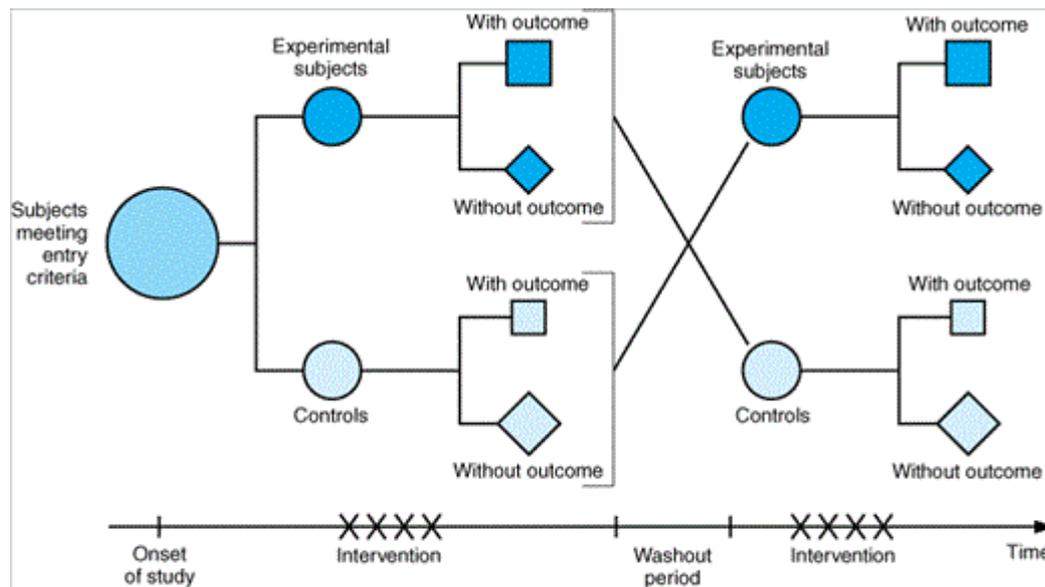
does not qualify as a randomized study and does present some potential problems in interpretation. Whenever patients are assigned to treatments within big blocks of time, there is always the possibility that an important event occurred between the two time periods, such as a change in the method used for cryotherapy. Although that may not have been true in this study, a randomized design would have been more persuasive.

Trials with Self-Controls

A moderate level of control can be obtained by using the same group of subjects for both experimental and control options. The study by [Sauter and colleagues \(2002\)](#) involved patients who underwent cholecystectomy. Follow-up occurred 1 and 3 months after cholecystectomy to detect changes such as abdominal pain, flatulence, and dyspepsia. This type of study uses patients as their own controls and is called a **self-controlled study**. Studies with self-controls and no other control group are still vulnerable to the well-known Hawthorne effect, described by [Roethlisberger and colleagues \(1946\)](#), in which people change their behavior and sometimes improve simply because they receive special attention by being in a study and not because of the study intervention. These studies are similar to cohort studies except for the intervention or treatment that is involved.

The self-controlled study design can be modified to provide a combination of concurrent and self-controls. This design uses two groups of patients: One group is assigned to the experimental treatment, and the second group is assigned to the placebo or control treatment ([Figure 2-7](#)). After a time, the experimental treatment and placebo are withdrawn from both groups for a “washout” period. During the washout period, the patients generally receive no treatment. The groups are then given the alternative treatment; that is, the first group now receives the placebo, and the second group receives the experimental treatment. This design, called a **crossover study**, is powerful when used appropriately.

FIGURE 2–7.



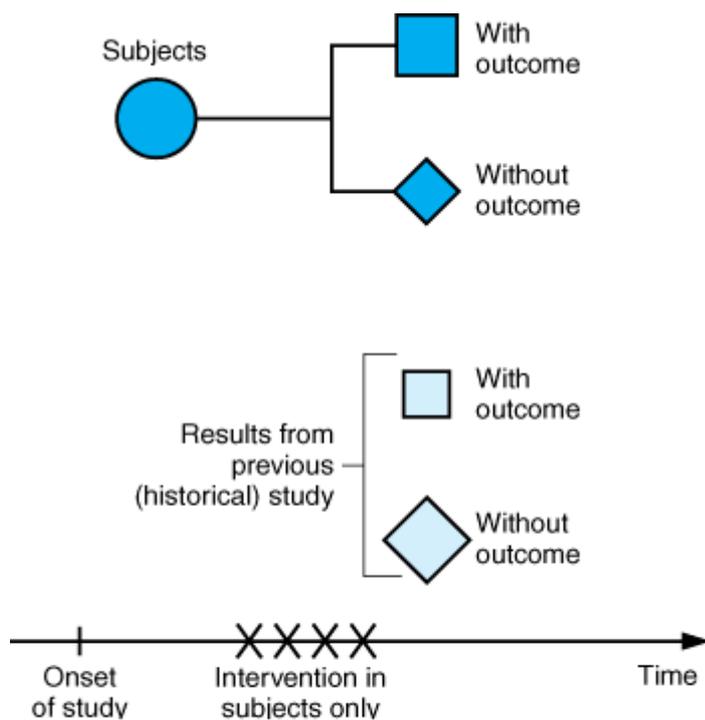
Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of trial with crossover. Shaded areas represent subjects assigned to the treatment condition; unshaded areas correspond to subjects assigned to the control condition. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest.

Trials with External Controls

The third method for controlling experiments is to use controls external to the study. Sometimes, the result of another investigator's research is used as a comparison. On other occasions, the controls are patients the investigator has previously treated in another manner, called **historical controls**. The study design is illustrated in [Figure 2–8](#).

FIGURE 2–8.



Source: Dawson B, Trapp RG: *Basic & Clinical Biostatistics*, 4th Edition: <http://www.accessmedicine.com>

Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

Schematic diagram of trial with external controls. Shaded areas represent subjects assigned to the treatment condition; unshaded areas correspond to patients cared for under the control condition. Squares represent subjects with the outcome of interest; diamonds represent subjects without the outcome of interest.

Historical controls are frequently used to study diseases for which cures do not yet exist and are used in oncology studies, although oncologic studies use concurrent controls when possible. In studies involving historical controls, researchers should evaluate whether other factors may have changed since the time the historical controls were treated; if so, any differences may be due to these other factors and not to the treatment.

Uncontrolled Studies



Not all studies involving interventions have controls, and by strict definition they are not really experiments or trials. For example, [Crook and associates \(1997\)](#) (a presenting problem in [Chapter 9](#)) reported the results of a trial of radiotherapy for prostate carcinoma in which patients were followed for at least 12 and for as long as 70 months. The investigators wanted to determine the length of time a patient had no

recurrence of the tumor as well as how long the patients survived. They found some differences in the probability of long-term survival in patients who had different tumor classification scores (scores that measure the severity of the tumor). This study was an **uncontrolled study** because there were no comparisons with patients treated in another manner.

Uncontrolled studies are more likely to be used when the comparison involves a procedure than when it involves a drug. The major shortcoming of such studies is that investigators assume that the procedure used and described is the best one. The history of medicine is filled with examples in which one particular treatment is recommended and then discontinued after a controlled clinical trial is undertaken. One significant problem with uncontrolled trials is that unproved procedures and therapies can become established, making it very difficult for researchers to undertake subsequent controlled studies. Another problem is finding a significant difference when it may be unfounded. [Guyatt and colleagues \(2000\)](#) identified 13 randomized trials and 17 observational studies in adolescent pregnancy prevention. Six of eight outcomes they examined showed a significant intervention effect in the observational studies, whereas the randomized studies showed no benefit.

Meta-Analysis & Review Papers

A type of study that does not fit specifically in either category of observation studies or experiments is called **meta-analysis**. Meta-analysis uses published information from other studies and combines the results so as to permit an overall conclusion. Meta-analysis is similar to review articles, but additionally includes a quantitative assessment and summary of the findings. It is possible to do a meta-analysis of observational studies or experiments; however, a meta-analysis should report the findings for these two types of study designs separately. This method is especially appropriate when the studies that have been reported have small numbers of subjects or come to different conclusions.

[Veenstra and colleagues \(1999\)](#) (a presenting problem in [Chapter 10](#)) performed a meta-analysis of infection and central venous catheters. The investigators wanted to know whether catheters impregnated with antiseptic were effective in preventing catheter-related bloodstream infection, compared with untreated catheters. They found 12 randomized trials that had addressed this question and combined the results in a statistical manner to reach an overall conclusion about their effectiveness—mainly that the impregnated catheters appear to be effective in reducing the incidence of infection in high-risk patients.

Advantages & Disadvantages of Different Study Designs



The previous sections introduced the major types of study designs used in medical research, broadly divided into experimental studies, or clinical trials, and observational studies (cohort, case–control, cross-sectional, and case–series designs). Each study design has certain advantages over the others as well as some specific disadvantages, which we discuss in the following sections.

Advantages & Disadvantages of Clinical Trials

The randomized clinical trial is the gold standard, or reference, in medicine; it is the design against which others are judged—because it provides the greatest justification for concluding causality and is subject to the least number of problems or biases. Clinical trials are the best type of study to use when the objective is to establish the efficacy of a treatment or a procedure. Clinical trials in which patients are randomly assigned to different treatments, or “arms,” are the strongest design of all. One of the treatments is the experimental condition; another is the control condition. The control may be a placebo or a sham procedure; often, it is the treatment or procedure commonly used, called the standard of care or reference standard. For example, patients with coronary artery disease in the Coronary Artery Surgery Study ([CASS Principal Investigators and Associates, 1983](#)) were randomized to receive either surgical or medical care; no patient was left untreated or given a placebo.

A number of published articles have shown the tendency for nonrandomized studies, especially those using historical controls, to be more likely to show a positive outcome, compared with randomized studies. In some situations, however, historical controls can and should be used. For instance, historical controls may be useful when preliminary studies are needed or when researchers are dealing with late treatment for an intractable disease, such as advanced cancer. Although clinical trials provide the greatest justification for determining causation, obstacles to using them include their great expense and long duration. For instance, a randomized trial comparing various treatments for carcinoma requires the investigators to follow the subjects for a long time. Another potential obstacle to using clinical trials occurs when certain practices become established and accepted by the medical community, even though they have not been properly justified. As a result, procedures become established that may be harmful to many patients, as evidenced by the controversy over silicone breast implants and the many different approaches to

managing hypertension, many of which have never been subjected to a clinical trial that includes the most conservative treatment, diuretics.

Advantages & Disadvantages of Cohort Studies

Cohort studies are the design of choice for studying the causes of a condition, the course of a disease, or the risk factors because they are longitudinal and follow a group of subjects over a period of time. Causation generally cannot be proved with cohort studies because they are observational and do not involve interventions. However, because they follow a cohort of patients forward through time, they possess the correct time sequence to provide strong evidence for possible causes and effects, as in the smoking and lung cancer controversy. In well-designed cohort studies, investigators can control many sources of bias related to patient selection and recorded measurements.

The length of time required in a cohort study depends on the problem studied. With diseases that develop over a long period of time or with conditions that occur as a result of long-term exposure to some causative agent, many years are needed for study. Extended time periods make such studies costly. They also make it difficult for investigators to argue causation because other events occurring in the intervening period may have affected the outcome. For example, the long time between exposure and effect is one of the reasons it is difficult to study the possible relationship between environmental agents and various carcinomas. Cohort studies that require a long time to complete are especially vulnerable to problems associated with patient follow-up, particularly patient attrition (patients stop participating in the study) and patient migration (patients move to other communities). This is one reason that the Framingham study, with its rigorous methods of follow-up, is such a rich source of important information.

Advantages & Disadvantages of Case–Control Studies

Case–control studies are especially appropriate for studying rare diseases or events, for examining conditions that develop over a long time, and for investigating a preliminary hypothesis. They are generally the quickest and least expensive studies to undertake and are ideal for investigators who need to obtain some preliminary data prior to writing a proposal for a more complete, expensive, and time-consuming study. They are also a good choice for someone who needs to complete a clinical research project in a specific amount of time.

The advantages of case–control studies lead to their disadvantages. Of all study methods, they have the largest number of possible biases or errors, and they

depend completely on high-quality existing records. Data availability for case–control studies sometimes requires compromises between what researchers wish to study and what they are able to study. One of the authors was involved in a study of elderly burn patients in which the goal was to determine risk factors for survival. The primary investigator wanted to collect data on fluid intake and output. He found, however, that not all of the existing patient records contained this information, and thus it was impossible to study the effect of this factor.

One of the greatest problems in a case–control study is selection of an appropriate control group. The cases in a case–control study are relatively easy to identify, but deciding on a group of persons who provide a relevant comparison is more difficult. Because of the problems inherent in choosing a control group in a case–control study, some statisticians have recommended the use of two control groups: one control group similar in some ways to the cases (eg, having been hospitalized during the same period of time) and another control group of healthy subjects.

Advantages & Disadvantages of Cross-Sectional Studies

Cross-sectional studies are best for determining the status quo of a disease or condition, such as the prevalence of HIV in given populations, and for evaluating diagnostic procedures. Cross-sectional studies are similar to case–control studies in being relatively quick to complete, and they may be relatively inexpensive as well. Their primary disadvantage is that they provide only a “snapshot in time” of the disease or process, which may result in misleading information if the research question is really one of disease process. For example, clinicians used to believe that diastolic blood pressure, unlike systolic pressure, does not increase as patients grow older. This belief was based on cross-sectional studies that had shown mean diastolic blood pressure to be approximately 80 mm Hg in all age groups. In the Framingham cohort study, however, the patients who were followed over a period of several years were observed to have increased diastolic blood pressure as they grew older ([Gordon et al, 1959](#)).

This apparent contradiction is easier to understand if we consider what happens in an aging cohort. For example, suppose that the mean diastolic pressure in men aged 40 years is 80 mm Hg, although there is individual variation, with some men having a blood pressure as low as 60 mm Hg and others having a pressure as high as 100 mm Hg. Ten years later there is an increase in diastolic pressure, although it is not an even increase; some men experience a greater increase than others. The men who were at the upper end of the blood pressure **distribution** 10 years earlier and

who had experienced a larger increase have died in the intervening period, so they are no longer represented in a cross-sectional study. As a result, the mean diastolic pressure of the men still in the cohort at age 50 is about 80 mm Hg, even though individually their pressures are higher than they were 10 years earlier. Thus, a cohort study, not a cross-sectional study, provides the information leading to a correct understanding of the relationship between normal aging and physiologic processes such as diastolic blood pressure.

Surveys are generally cross-sectional studies. Most of the voter polls done prior to an election are one-time samplings of a group of citizens, and different results from week to week are based on different groups of people; that is, the same group of citizens is not followed to determine voting preferences through time. Similarly, consumer-oriented studies on customer satisfaction with automobiles, appliances, health care, and so on are cross-sectional.

A common problem with survey research is obtaining sufficiently large response rates; many people asked to participate in a survey decline because they are busy, not interested, and so forth. The conclusions are therefore based on a subset of people who agree to participate, and these people may not be **representative** of or similar to the entire population. The problem of representative participants is not confined to cross-sectional studies; it can be an issue in other studies whenever subjects are selected or asked to participate and decline or drop out. Another issue is the way questions are posed to participants; if questions are asked in a leading or emotionally inflammatory way, the responses may not truly represent the participants' feelings or opinions. We discuss issues with surveys more completely in [Chapter 11](#).

Advantages & Disadvantages of Case–Series Studies

Case–series reports have two advantages: They are easy to write, and the observations may be extremely useful to investigators designing a study to evaluate causes or explanations of the observations. But as we noted previously, case–series studies are susceptible to many possible biases related to subject selection and characteristics observed. In general, you should view them as hypothesis-generating and not as conclusive.

Summary

This chapter illustrates the study designs most frequently encountered in the medical literature. In medical research, subjects are observed or experiments are

undertaken. Experiments involving humans are called trials. Experimental studies may also use animals and tissue, although we did not discuss them as a separate category; the comments pertaining to clinical trials are relevant to animal and tissue studies as well.

Each type of study discussed has advantages and disadvantages.

Randomized, controlled clinical trials are the most powerful designs possible in medical research, but they are often expensive and time-consuming. Well-designed observational studies can provide useful insights on disease causation, even though they do not constitute proof of causes. Cohort studies are best for studying the natural progression of disease or risk factors for disease; case–control studies are much quicker and less expensive. Cross-sectional studies provide a snapshot of a disease or condition at one time, and we must be cautious in inferring disease progression from them. Surveys, if properly done, are useful in obtaining current opinions and practices. Case–series studies should be used only to raise questions for further research.

We have used several presenting problems from later chapters to illustrate different study designs. We will point out salient features in the design of the presenting problems as we go along, and we will return to the topic of study design again after all the prerequisites for evaluating the quality of journal articles have been presented.

Exercises

Read the Descriptions of the Following Studies and Determine the Study Design Used.

1. Cryptosporidiosis is an enteric illness that is frequently waterborne. [Khalakdina and colleagues \(2003\)](#) could find no published studies of the risk factors for cryptosporidiosis in immunocompetent adults. Patients with cryptosporidiosis were recruited from a surveillance system, and age-matched controls were recruited by random-digit dialing. Subjects in both groups were interviewed by telephone to obtain information about previous exposures.

2. [Brown and coworkers \(2003\)](#) designed a study to determine the efficacy of immunotherapy with ant venom for treating ant stings. The study involved a group of 68 adults who were allergic to ant stings; each subject was randomly assigned to receive either venom immunotherapy or a placebo. After a sting challenge in which any reactions were recorded, the group originally on the placebo was given the

venom immunotherapy, and after a sufficient time, they too were given a sting challenge.

3. The Prostate Cancer Outcomes Study was designed to investigate the patterns of cancer care and effects of treatment on quality of life. [Clegg and coworkers \(2001\)](#) identified eligible cases from pathology facilities within 6 months of diagnosis. A random sample of eligible cases were contacted and asked to complete a questionnaire on their initial treatment and to provide permission to the investigators to abstract their medical records to obtain information on their initial care.

4. Factors contributing to medical students' self-perceived competency in cancer screening examinations were studied at the UCLA Medical School ([Lee et al, 2002](#)). Students were asked to assess their competency in performing several cancer screening examinations, and multiple regression analysis (see [Chapter 10](#)) was used to identify predictors of competency.

5. A study to determine whether treatment with a calcium channel block or an angiotensin-converting enzyme inhibitor lowers the incidence of coronary heart disease when compared with a diuretic included over 33,000 patients ([ALLHAT 2002](#)). The primary outcome was fatal coronary heart disease or myocardial infarction.

6. [Grodstein and colleagues \(2000\)](#) reported on the relationship between duration, dose, and type of postmenopausal hormone therapy and the risk of coronary heart disease in women. Subjects in the study were selected from the Nurses' Health Study originally completed in 1976; the study included 120,000 married female registered nurses, aged 30–55. The original survey provided information on the subjects' age, parental history of myocardial infarction, smoking status, height, weight, use of oral contraceptives or postmenopausal hormones, and history of myocardial infarction or angina pectoris, diabetes, hypertension, or high serum cholesterol levels. Follow-up surveys were every 2 years thereafter.

7. [Thomas and coworkers \(2002\)](#) designed a study to examine the diagnostic accuracy of three physical signs (Kernig's sign, Brudzinski's sign, nuchal rigidity) for diagnosing meningitis. A total of 297 adults with suspected meningitis underwent lumbar puncture, and the results were compared with the three physical signs.

8. [Kreder and colleagues \(2003\)](#) studied the effect of provider volume on complication rates after total knee arthroplasty in patients. Subjects were in a

national database, and it was used to obtain information about complications, infection rates, and mortality. Low provider volume was related to length of stay in hospital but not to other complications.

9. [Sagawa and colleagues \(2003\)](#) were interested in the efficacy of sputum cytology in a mass screening program for the early detection of lung cancer. The results from an earlier screening program were compared for patients with lung cancer and subjects without lung cancer.

10. **Group Exercise.** The abuse of phenacetin, a common ingredient of analgesic drugs, can lead to kidney disease. There is also evidence that use of salicylate provides protection against cardiovascular disease. How would you design a study to examine the effects of these two drugs on mortality due to different causes and on cardiovascular morbidity?

11. **Group Exercise.** Select a study with an interesting topic, either one of the studies referred to in this chapter or from a current journal. Carefully examine the research question and decide which study design would be optimal to answer the question. Is that the study design used by the investigators? If so, were the investigators attentive to potential problems identified in this chapter? If not, what are the reasons for the study design used? Do they make sense?