CHAPTER 7

Designing Cross-Sectional and Cohort Studies

Stephen B. Hulley, Steven R. Cummings, and Thomas B. Newman

Observational studies have two primary purposes: **descriptive**, examining the distributions of predictors and outcomes in a population, and **analytic**, characterizing associations between these predictor and outcome variables. In this chapter we present two basic observational designs, which are categorized according to the **time frame** for making the measurements.

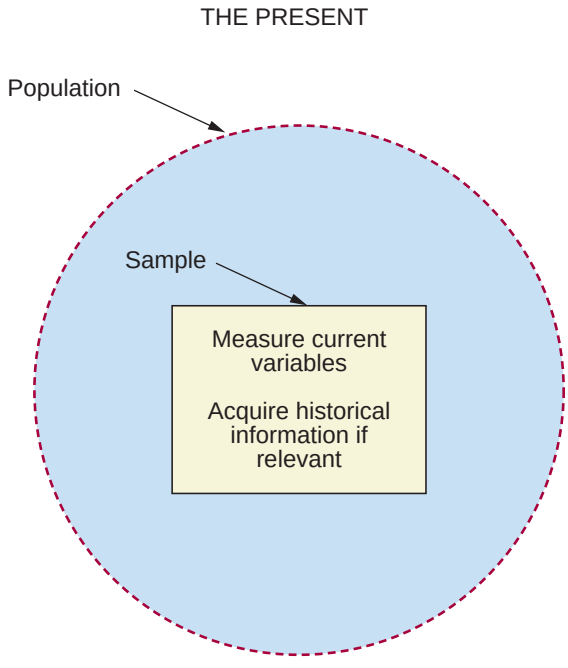
In a **cross-sectional** study, the investigator makes all of her measurements on a single occasion or within a short period of time. She draws a sample from the population and looks at distributions of variables within that sample, sometimes designating them as predictors and outcomes based on biologic plausibility and historical information. For example, if she is interested in studying the relationship between body weight and blood pressure she could measure these variables at a single clinic visit for each study subject and examine whether subjects with higher body weights were more likely to have hypertension.

In a **cohort study** measurements take place over a period of time in a group of participants who have been identified at the beginning of the study ("the cohort"). Thus, the defining characteristic of cohort studies is that a group **assembled at the outset** is followed **longitudinally**. For example the investigator could measure body weight and blood pressure on a cohort of study subjects at an initial clinic visit and then follow them for 5 years to determine the relationship between baseline weight and the incidence of hypertension. In this chapter we discuss **prospective** and **retrospective cohort** designs and **multiple-cohort** designs. We also address **statistical analysis** approaches, and the importance of optimizing **cohort retention** during follow-up.

CROSS-SECTIONAL STUDIES

In a cross-sectional study all the measurements are made at about the same time, with no follow-up period (Figure 7.1). Cross-sectional designs are well suited to the goal of describing variables and their distribution patterns. In the National Health and Nutrition Examination Survey (NHANES), for example, a sample designed to represent the entire U.S. population aged 1–74 was interviewed and examined in the early 1970s. This cross-sectional study was a major source of information about the health and habits of the U.S. population in the year it was carried out, providing estimates of such things as the prevalence of smoking in various demographic groups. Subsequent cross-sectional NHANES surveys have been carried out periodically, and all NHANES data sets are available for public use (www.cdc.gov/nchs/nhanes.htm).

Cross-sectional studies can be used for examining associations, although the choice of which variables to label as predictors and which as outcomes depends on the cause-and-effect hypotheses of the investigator rather than on the study design. This choice is easy for constitutional factors such as age, race, and sex; these cannot be altered by other variables and therefore are always predictors. For other variables, however, the choice can go either way. For example, in NHANES III there was a cross-sectional association between childhood obesity and hours spent



■ **FIGURE 7.1** In a cross-sectional study, the steps are to:

- Define selection criteria and recruit a sample from the population.
- Measure current values of predictor and outcome variables, often supplemented by historical information.

watching television (1). Whether to label obesity or television-watching as the predictor and the other as the outcome depends on the causal hypothesis of the investigator.

Unlike cohort studies, which have a longitudinal time dimension and can be used to estimate **incidence** (the proportion who *develop* a disease or condition over time), cross-sectional studies provide information about **prevalence**, the proportion who *have* a disease or condition at one point in time. Prevalence matters to a clinician, who must estimate the likelihood that the patient sitting in her office has a particular disease; the greater the prevalence, the greater the "prior probability" of the disease (the probability before the results of various diagnostic tests are available; see Chapter 12). That's why more patients with knee pain have osteoarthritis than palindromic rheumatism. Prevalence is also useful to health planners who want to know how many people have certain diseases so that they can allocate enough resources to care for them. When analyzing cross-sectional studies, the prevalence of the outcome can be compared in those with and without an exposure, yielding the **relative prevalence** of the outcome, the cross-sectional equivalent of relative risk (see Appendix 8A for examples).

Sometimes cross-sectional studies describe the prevalence of ever having done something or ever having had a disease or condition. In that case, it is important to make sure that follow-up time is the same in those exposed and unexposed.This is illustrated in Example 7.1, in which the prevalence of ever having tried smoking was studied in a cross-sectional study of children with differing levels of exposure to movies in which the actors smoke. Of course, children who had seen more movies were also older, and therefore had longer to try smoking, so it was important to adjust for age in multivariate analyses (see Chapter 9).

Strengths and Weaknesses of Cross-Sectional Studies

A major advantage of cross-sectional studies is that there is no waiting around for the outcome to occur. This makes them fast and inexpensive, and avoids the problem of loss to follow-up. Another advantage is that a cross-sectional study can be included as the first step in a cohort

EXAMPLE 7.1 Cross-Sectional Study

Sargent et al. (2) sought to determine whether exposure to movies in which the actors smoke is associated with smoking initiation. The steps in performing the study were to:

1. **Define selection criteria and recruit the population sample.** The investigators did a random-digit-dial survey of 6,522 U.S. children aged 10 to 14 years.
2. **Measure the predictor and outcome variables.** They quantified smoking in 532 popular movies and for each subject asked which of a randomly selected subset of 50 movies they had seen. Subjects were also asked about a variety of covariates such as age, race, gender, parent smoking and education, sensation-seeking (e.g., "I like to do dangerous things"), and self-esteem (e.g., "I wish I were someone else"). The outcome variable was whether the child had ever tried smoking a cigarette.

The prevalence of ever having tried smoking varied from 2% in the lowest quartile of movie smoking exposure to 22% in the highest quartile. After adjusting for age and other confounders, these differences were statistically significant; the authors estimated that 38% of smoking initiation was attributable to exposure to movies in which the actors smoke.

study or clinical trial at little or no added cost. The results define the demographic and clinical characteristics of the study group at baseline and can sometimes reveal cross-sectional associations of interest.

However, as previously noted, it's often difficult to establish causal relationships from cross-sectional data. Cross-sectional studies are also impractical for the study of rare diseases, unless the sample is drawn from a population of diseased patients rather than the general population. A **case series** of this sort is better suited to describing the characteristics of the disease than to analyzing differences between these patients and healthy people, although informal comparisons with prior experience can sometimes identify very strong risk factors. In a case series of the first 1,000 patients with AIDS, for example, 727 were homosexual or bisexual males and 236 were injection drug users (3). It did not require a formal control group to conclude that these groups were at increased risk. Furthermore, within a sample of persons with a disease there may be associations of interest, e.g., the higher risk of Kaposi's sarcoma among patients with AIDS who were homosexual than among those who were injection drug users.

Because cross-sectional studies measure only prevalence, rather than incidence, it is important to be cautious when drawing inferences about the causes, prognosis, or natural history of a disease. A factor that is associated with prevalence of disease may be a cause of the disease but could also just be associated with duration of the disease. For example, the prevalence of chronic renal failure is affected not only by its incidence, but also by survival once it has occurred. Given the observation that obesity is associated with greater survival among dialysis patients (4), a cross-sectional study of the predictors of chronic renal failure might overestimate the association between obesity and renal failure.

Serial Surveys

Occasionally, investigators perform a series of cross-sectional studies in the same population, say every 5 years. This design can be used to draw inferences about changing patterns over time. For example, Zito et al. (5), using annual cross-sectional surveys, reported that the prevalence of prescription psychotropic drug use among youth (<20 years old) increased more than

threefold between 1987 and 1996 in a mid-Atlantic Medicaid population. **Serial cross-sectional surveys** have a longitudinal time frame, but they are not the same as a cohort study, because a new sample is drawn each time. As a result, changes within individuals cannot be assessed, and findings may be influenced by people entering or leaving the population (and, thus, the samples) due to births, deaths, and migration.

COHORT STUDIES

Prospective Cohort Studies

Cohort was the Roman term for a group of soldiers that marched together, and in clinical research a cohort is a group of subjects, specified at the outset and followed over time. In a **prospective cohort study**, the investigator begins by assembling a sample of subjects (Figure 7.2). She measures characteristics in each subject that might predict the subsequent outcomes, and follows these subjects with periodic measurements of the outcomes of interest (Example 7.2).

Strengths and Weaknesses of Prospective Cohort Studies

A major advantage of the **cohort design** is that, unlike cross-sectional designs, it allows the calculation of **incidence**—the number of new cases of a condition occuring over time (Table 7.1). Measuring levels of the predictor before the outcome occurs establishes the time sequence of the variables, which strengthens the process of inferring the causal basis of an association. The prospective approach also prevents the predictor measurements from being influenced by the

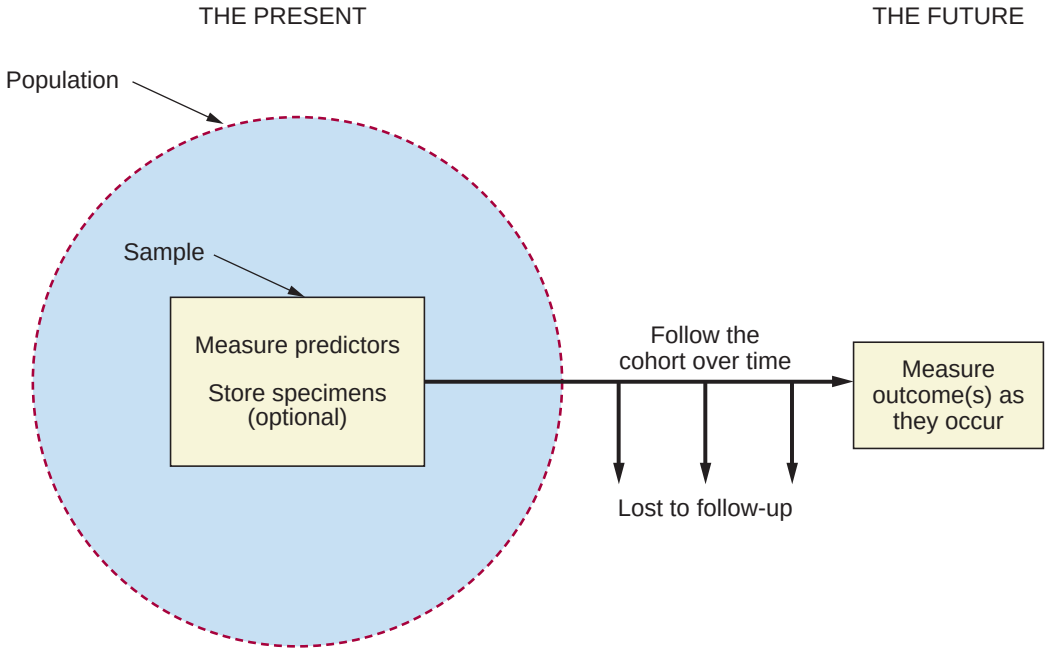


FIGURE 7.2 In a prospective cohort study, the steps are to:

- Define selection criteria and recruit a sample from the population ("the cohort").
- Measure the predictor variables and, if appropriate, the baseline level of the outcome variable.
- Consider the option to store specimens, images, etc. for later analysis of predictors.
- Follow the cohort over time, minimizing loss to follow-up.
- Measure the outcome variable(s) during follow-up.

EXAMPLE 7.2 Prospective Cohort Study

The classic Nurses' Health Study examines incidence and risk factors for common diseases in women. The steps in performing the study were to:

1. **Define selection criteria and assemble the cohort.** In 1976, the investigators obtained lists of registered nurses aged 25 to 42 in the 11 most populous states and mailed them an invitation to participate in the study; those who agreed became the cohort.
2. **Measure predictor variables, including potential confounders.** They mailed a questionnaire about weight, exercise, and other potential risk factors and obtained completed questionnaires from 121,700 nurses. They sent questionnaires periodically to ask about additional risk factors and update the status of some risk factors that had been measured previously.
3. **Follow-up the cohort and measure outcomes.** The periodic questionnaires also included questions about the occurrence of a variety of disease outcomes, which were validated by the investigators.

The prospective approach allowed investigators to make measurements at baseline and collect data on subsequent outcomes. The large size of the cohort and long period of follow-up provided substantial statistical power to study risk factors for cancers and other diseases.

For example, the investigators examined the hypothesis that gaining weight increases a woman's risk of breast cancer after menopause (6). The women reported their weight at age 18 in an early questionnaire, and follow-up weights in later questionnaires. The investigators succeeded in following 95% of the women, and 1,517 breast cancers were confirmed during the next 12 years. Heavier women had a higher risk of breast cancer after menopause, and those who gained more than 20 kg since age 18 had a twofold increased risk of developing breast cancer (relative risk = 2.0; 95% confidence interval, 1.4 to 2.8). Adjusting for potential confounding factors did not change the result.

outcome or knowledge of its occurrence and it allows the investigator to measure variables more completely and accurately than is usually possible retrospectively. This is important for predictors such as dietary habits that are difficult for a subject to remember accurately. When fatal diseases are studied retrospectively, predictor variable measurements about the decedent can only be reconstructed from indirect sources such as medical records or friends and relatives.

All cohort studies share the general disadvantage of observational studies (relative to clinical trials) that causal inference is challenging and interpretation often muddied by the influences of confounding variables (Chapter 9). A particular weakness of the prospective design is its

TABLE 7.1 STATISTICS FOR EXPRESSING DISEASE FREQUENCY IN OBSERVATIONAL STUDIES

TYPE OF STUDY	STATISTIC	DEFINITION
Cross-sectional	Prevalence	$\frac{\text{Number of people who } \textit{have} \text{ a disease or condition at a given point in time}}{\text{Number of people at risk}}$
Cohort	Incidence rate	$\frac{\text{Number of people who } \textit{get} \text{ a disease or condition}}{\text{Number of people at risk} \times \text{time period at risk}}$

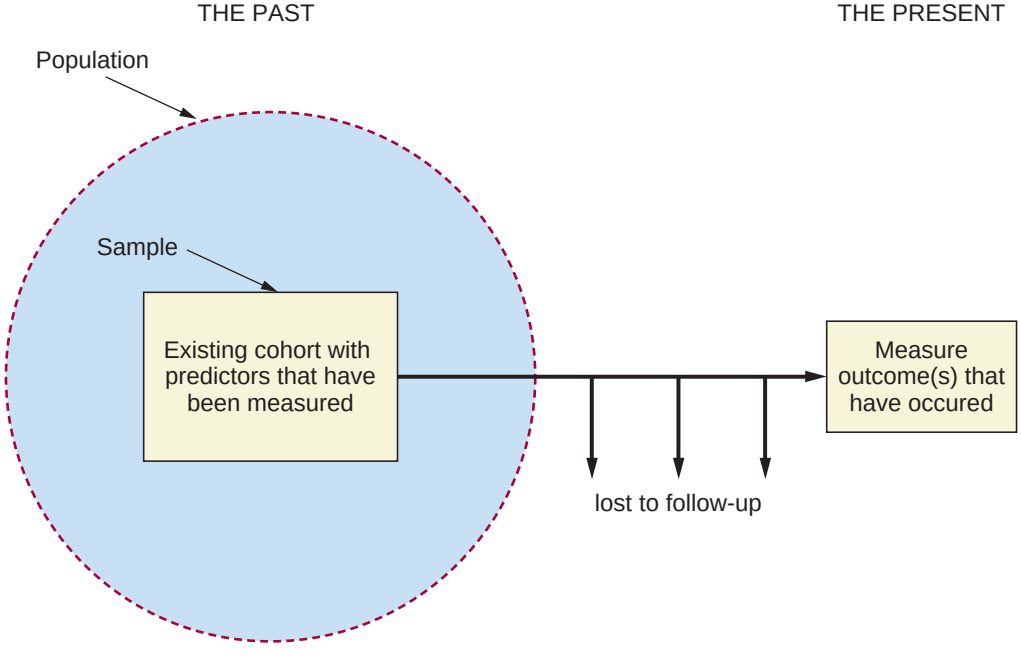
expense and inefficiency for studying rare outcomes. Even diseases we think of as relatively common, such as breast cancer, happen at such a low rate in any given year that large numbers of people must be followed for long periods of time to observe enough outcomes to produce meaningful results. Cohort designs are more efficient for dichotomous outcomes that are more common and immediate, and for continuous outcomes.

Retrospective Cohort Studies

The design of a **retrospective cohort** study (Figure 7.3) differs from that of a prospective one in that the assembly of the cohort, baseline measurements, and follow-up have all happened in the past. This type of study is only possible if adequate data about the predictors are available on a cohort of subjects that has been assembled for other purposes, such as an electronic clinical or administrative database (Example 7.3).

Strengths and Weaknesses of Retrospective Cohort Studies

Retrospective cohort studies have many of the strengths of prospective cohort studies, and they have the advantage of being much less costly and time-consuming. The subjects are already assembled, baseline measurements have already been made, and the follow-up period has already taken place. The main disadvantages are the limited control the investigator has over the approach to sampling and follow-up of the population, and over the nature and the quality of the baseline measurements. The existing data may be incomplete, inaccurate, or measured in ways that are not ideal for answering the research question.



■ **FIGURE 7.3** In a retrospective cohort study, the cohort selection and follow-up have occured in the past, so the steps are to:

- Identify an existing cohort that has some predictor information already recorded.
- Assess loss to follow-up that has occurred.
- Measure the outcome variable(s) that have already occurred.

EXAMPLE 7.3 Retrospective Cohort Study

Pearce et al. used UK National Health Service Central Registry data to describe the risk of leukemia and brain tumors associated with head CT scans in childhood (7). The steps in performing the study were to:

1. **Identify a suitable existing cohort.** The cohort consisted of 178,604 children and young adults aged <22 who received head CT scans between 1985 and 2002.
2. **Collect predictor variable data.** The investigators reviewed the records to collect gender, age, numbers, and types of radiology procedures and estimated radiation dose.
3. **Collect outcome data.** To avoid inclusion of CT scans related to cancer diagnosis, the investigators recorded leukemia occurring at least 2 years after the first CT, and brain tumors at least 5 years after the first CT, through 2008.

Childhood CT scans significantly increased the risk of leukemia and brain cancer, and the increase was dose-related; cumulative doses of 50–60 mGy tripled the risk of both leukemia and brain cancer. However, the absolute increase in risk was low, one excess case of each outcome per 10,000 head scans. The investigators, while noting that the benefits of the CT scans likely outweighed these risks, urged that radiation doses from CT scans be kept as low as possible in children, and that alternative procedures that avoid ionizing radiation be considered whenever appropriate.

Multiple-Cohort Studies and External Controls

Multiple-cohort studies begin with two or more separate samples of subjects: typically, one group with **exposure** to a potential risk factor and one or more other groups with no exposure or a lower level of exposure (Figure 7.4). After defining suitable cohorts with different levels of exposure to the predictor of interest, the investigator measures other predictor variables, follows up the cohorts, and assesses outcomes as in any other type of cohort study (Example 7.4).

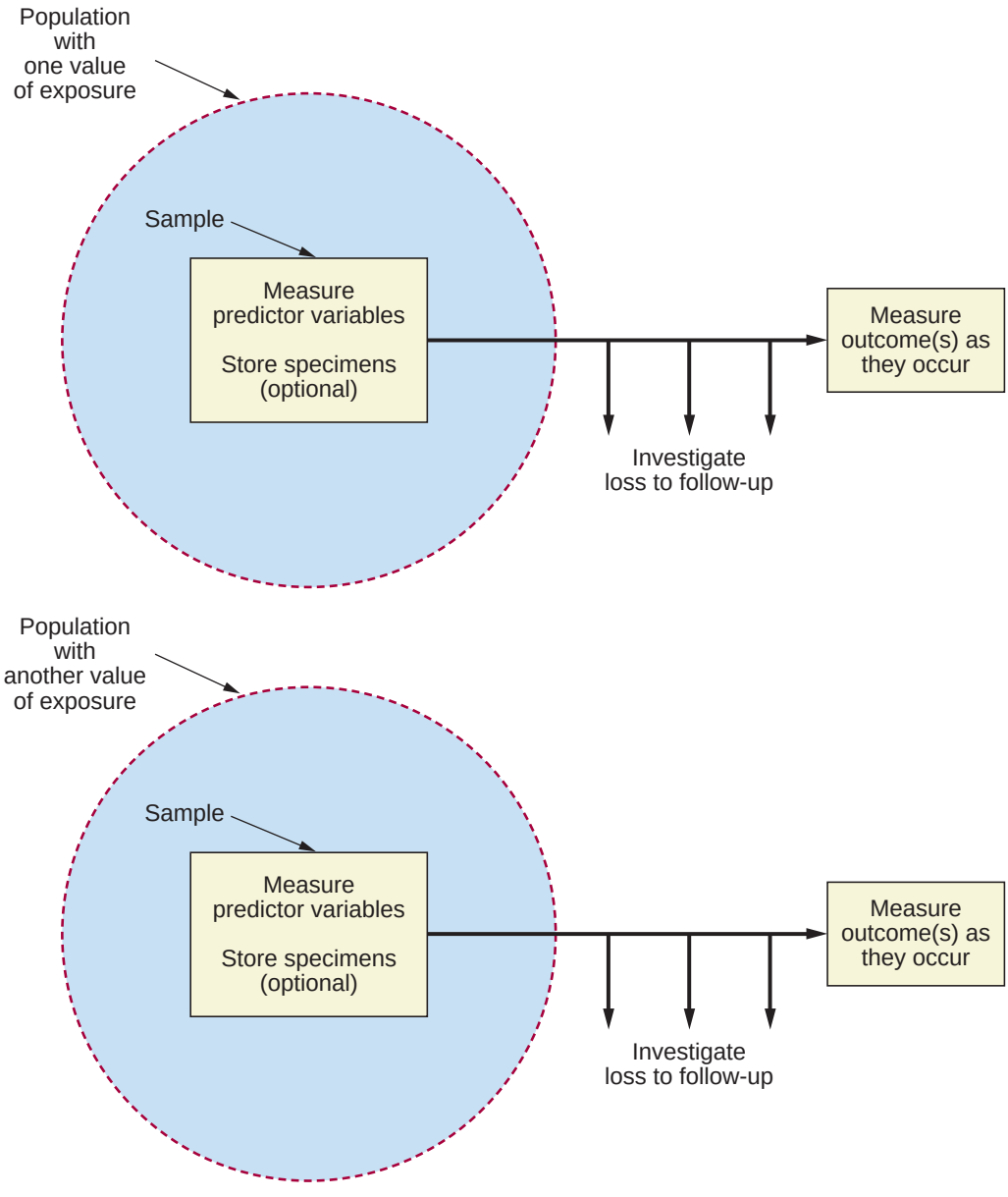
The use of two different samples of subjects in a **double-cohort design** should not be confused with the use of two samples in the case–control design (Chapter 8). In a double-cohort

EXAMPLE 7.4 Multiple-Cohort Design

To determine whether substantial neonatal jaundice or dehydration has adverse effects on neurodevelopment, investigators from UCSF and Kaiser Permanente of Northern California (8, 9) undertook a triple-cohort study. The steps in performing the study were to:

1. **Identify cohorts with different exposures.** The investigators used electronic databases to identify term and near-term newborns who
   1. had a maximum total serum bilirubin level of ≥ 25 mg/dL, or
   2. were readmitted for dehydration with a serum sodium of ≥ 150 mEq/L or weight loss of ≥ 12% from birth, or
   3. were randomly selected from the birth cohort
2. **Collect outcome data.** The investigators used electronic databases to search for diagnoses of neurological disorders and did full neurodevelopmental examinations at the age of 5 for consenting participants (blinded to which of the three cohorts the participant belonged to).

Neither hyperbilirubinemia nor dehydration was associated with adverse outcomes.



■ **FIGURE 7.4** In a double-cohort study (which can be conducted either prospectively or retrospectively) the steps are to:

- Select two or more cohorts from populations with different levels of the exposure (main predictor).
- Measure other predictors.
- Measure outcome variables during follow-up.

study the two groups of subjects are chosen based on the level of a predictor, whereas in a case–control study the two groups are chosen based on the presence or absence of an outcome.

In a variation on the multiple-cohort design, the outcome rate in a cohort can be compared with outcome rates in **census or registry** data from different populations. For example, in a classic study of whether uranium miners had an increased incidence of lung cancer, Wagoner et al. (10) compared the incidence of respiratory cancer in 3,415 uranium miners with that of

white men who lived in the same states. The increased incidence of lung cancer observed in the miners helped establish occupational exposure to ionizing radiation as an important cause of lung cancer.

Strengths and Weaknesses of Multiple-Cohort Designs

The multiple-cohort design may be the only feasible approach for studying rare exposures to potential occupational and environmental hazards. Using data from a census or registry as the external control group has the additional advantage of being population-based and economical. Otherwise, the strengths of this design are similar to those of other cohort studies.

The problem of **confounding** is accentuated in a multiple-cohort study because the cohorts are assembled from separate populations that can differ in important ways (besides exposure to the predictor variable) that influence the outcomes. Although some of these differences, such as age and race, can be matched or used to adjust the findings statistically, other characteristics may not be measurable and create problems in the interpretation of observed associations.

STATISTICAL APPROACH TO COHORT STUDIES

Risks, odds, and **rates** are estimates of the frequency of a dichotomous outcome in subjects who have been followed for a period of time. These three measures are closely related, sharing the same numerator—the number of subjects who develop the dichotomous outcome. Implicit in these three measures is the concept of being *at risk,* which means that the subject did not already have the outcome of interest at the beginning of the study. In a prospective study of the predictors of diabetes, a woman who had diabetes at baseline would not be at risk, since she already had the outcome of interest. On the other hand, there are episodic diseases, like heart failure requiring admission to a hospital, in which the outcome of interest may be the "incident" occurrence of a new episode, even if it occurs in someone who already has the disease.

Consider a study of 1,000 people who were followed for 2 years to see who developed lung cancer, and among whom eight new cases occurred each year. Risk, odds, and rate are shown in Table 7.2.

Of the three measures, risk is the easiest to understand because of its everyday familiarity—the risk of getting lung cancer in two years was 16 out of a thousand. Odds are harder to grasp intuitively—the odds of getting lung cancer were 16 to 984; fortunately, for rare outcomes (as in this case) the odds are quantitatively similar to risk and have no particular advantage. In studies comparing two groups the **odds ratio** is also similar to the **risk ratio** *when the outcome is*

TABLE 7.2 CALCULATION OF RISK, ODDS, AND RATE FOR A STUDY OF 1,000 PEOPLE FOLLOWED FOR TWO YEARS, WITH EIGHT NEW CASES OF LUNG CANCER EACH YEAR

STATISTIC	FORMULA	EXAMPLE
Risk	$\frac{N \text{ who develop the outcome}}{N \text{ at risk}}$	$\frac{16}{1{,}000} = 0.016$
Odds	$\frac{N \text{ who develop the outcome}}{N \text{ who do not develop the outcome}}$	$\frac{16}{984} = 0.0163$
Rate*	$\frac{N \text{ who develop the outcome}}{\text{Person-time at risk}}$	$\frac{16 \text{ cases}}{1{,}992 \text{ person-years}} = 0.008 \text{ cases / Person-year}$

*The denominator for the rate is the number at risk in the first year (1,000), plus the number at risk in the second (992).

rare, and this fact has unique importance in two situations: It is the basis for logistic regression calculations, and it is used to approximate relative risk in case–control studies (Appendix 8B). **Rates**, which take into account the accumulation of events over the course of time, are expressed as numbers of events divided by **person-time** at risk—the total amount of follow-up for each of the study subjects so long as that individual is alive, remains in the study, and has not yet had the outcome.

In some cohort studies significant **loss to follow-up**, unequal follow-up, or deaths or other events that preclude ascertainment of the outcome may occur. In these cases it is helpful to compare **incidence rates** between the groups—the number of outcomes divided by the person-time at risk. Each subject in the study contributes months or years of person-time from entry into the cohort until she either develops the outcome of interest or is "**censored**" due to loss to follow-up or death. The incidence rate in any group in the study is the number of outcomes in that group divided by the sum of that group's person-time at risk. As is true for the **risk ratio** (also known as relative risk), the **rate ratio** can be estimated as the quotient of rates in people who do and do not have a particular risk factor. The Cox proportional hazard model provides a method for multivariate analysis of data of this form (sometimes called "time to event" data); it allows estimation of **hazard ratios**, which are similar to rate ratios and have come into widespread use as the measure of association in **Cox regression analyses**.

Other Cohort Study Issues

The hallmark of a cohort study is the need to define the cohort of subjects at the *beginning* of a period of follow-up. The subjects should be appropriate to the research question and available for follow-up. They should sufficiently resemble the population to which the results will be generalized. The number of subjects should provide adequate power.

The quality of the study will depend on the precision and accuracy of the measurements of predictor and outcome variables (Chapter 4). The ability to draw inferences about cause and effect will depend on the degree to which the investigator has measured all **potential confounders** (Chapter 9), and the ability to generalize to subgroups of the population will depend on the degree to which the investigator has measured all **sources of effect modification**. Predictor variables may change during the study; whether and how frequently measurements should be repeated depends on cost, how much the variable is likely to change, and the importance to the research question of observing these changes. Outcomes should be assessed using standardized criteria, and when their assessment could be influenced by awareness of key risk factors, it is helpful if those making the assessments can be blinded to that predictor.

Follow-up of the entire cohort is important, and prospective studies should take a number of steps to achieve this goal (Table 7.3). Subjects who plan to move out of reach during the study or who will be difficult to follow for other reasons should be excluded at the outset. The investigator should collect information early on that she can use to find subjects who move or die, including the address, telephone number, and e-mail address of the subject, her personal physician, and at least two close friends or relatives who do not live in the same house. Mobile telephone numbers and personal e-mail addresses are particularly helpful, as they often remain unchanged when subjects, friends, or family move or change jobs. If feasible, obtaining the social security number will help in determining the vital status of those lost to follow-up, and obtaining hospital discharge information from the Social Security Administration for subjects who receive Medicare. Periodic contact with the subjects once or twice a year helps in keeping track of them, and may improve the timeliness and accuracy of recording the outcomes of interest. Finding subjects for follow-up assessments sometimes requires persistent and repeated efforts by mail, e-mail, telephone, or even house calls.

TABLE 7.3 STRATEGIES FOR MINIMIZING LOSSES DURING FOLLOW-UP

During enrollment

1. Exclude those likely to be lost:
   a. Planning to move
   b. Uncertainty about willingness to return
   c. Ill health or fatal disease unrelated to research question
2. Obtain information to allow future tracking:
   a. Address, telephone number (mobile phone numbers are particularly useful), and e-mail address of subject
   b. Social Security/Medicare number
   c. Name, address, telephone number, and e-mail address of close friends or relatives who do not live with the subject
   d. Name, address, telephone number, and email address of physician(s)

During follow-up*

1. Periodic contact with subjects to collect information, provide results, and be supportive:
   a. By telephone: may require calls during weekends and evenings
   b. By mail: repeated mailings by e-mail or with stamped, self-addressed return cards
   c. Other: newsletters, token gifts
2. For those who are not reached by phone or mail:
   a. Contact friends, relatives, or physicians
   b. Request forwarding addresses from postal service
   c. Seek address through other public sources, such as telephone directories and the Internet, and ultimately a credit bureau search
   d. For subjects receiving Medicare, collect data about hospital discharges from the Social Security Administration
   e. Determine vital status from state health department or National Death Index

At all times

1. Treat study subjects with appreciation, kindness, and respect, helping them to understand the research question so they will want to join as partners in making the study successful.

*This assumes that participants in the study have given informed consent to collect the tracking information and for follow-up contact.

SUMMARY

1. In a **cross-sectional** study, the variables are all measured at a single point in time, with no structural distinction between predictors and outcomes. Cross-sectional studies yield **weaker evidence for causality** than cohort studies because the predictor variable is not shown to precede the outcome.
2. Cross-sectional studies are valuable for providing descriptive information about **prevalence**, and have the advantage of **avoiding the time**, **expense**, and **dropout problems** of a follow-up design; they are often useful as the first step of a cohort study or experiment, and can be linked in independently sampled **serial surveys** to reveal population changes over time.
3. Cross-sectional studies require a large sample size when studying uncommon diseases and variables in the general population, but can be useful in a **case series** of an uncommon disease.
4. In **cohort studies**, a group of subjects identified at the outset is followed over time to describe the **incidence** or natural history of a condition and to discover **predictors** (risk factors) for various outcomes. The ability to measure the predictor before the outcome occurs establishes the sequence of events and controls bias in that measurement.
5. **Prospective cohort** studies begin at the outset of follow-up and may require large numbers of subjects followed for long periods of time. The latter disadvantage can sometimes be

overcome by identifying a **retrospective cohort** in which measurements of predictor variables have already occurred.

6. The **multiple-cohort** design, which compares the incidence of outcomes in cohorts that differ in the level of a predictor variable ("the **exposure**"), is useful for studying the effects of rare and occupational exposures.
7. **Risks**, **odds**, and **rates** are three ways to estimate the frequency of a dichotomous outcome during follow-up; among these, **incidence rates**, which take into account person-time of participants who remain alive and event-free in the study, are the basis for modern approaches to calculating **multivariate hazard ratios** using Cox proportional hazard models.
8. Inferences about **cause and effect** are strengthened by measuring and adjusting for all conceivable potential confounding variables. Bias in the assessment of outcomes is prevented by **standardizing** the measurements and **blinding** those assessing the outcome to the predictor variable values.
9. The strengths of a cohort design can be undermined by incomplete **follow-up** of subjects. Losses can be minimized by **excluding subjects** at the outset who may not be available for follow-up, by collecting **baseline information** that facilitates tracking, and by **staying in touch** with all subjects regularly.

REFERENCES

1. Andersen RE, Crespo CJ, Bartlett SJ, et al. Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. *JAMA* 1998;279(12):938–942.
2. Sargent JD, Beach ML, Adachi-Mejia AM, et al. Exposure to movie smoking: its relation to smoking initiation among US adolescents. *Pediatrics* 2005;116(5):1183–1191.
3. Jaffe HW, Bregman DJ, Selik RM. Acquired immune deficiency syndrome in the United States: the first 1,000 cases. *J Infect Dis* 1983;148(2):339–345.
4. Kalantar-Zadeh K, Abbott KC, Salahudeen AK, et al. Survival advantages of obesity in dialysis patients. *Am J Clin Nutr* 2005; 81: 543–554.
5. Zito JM, Safer DJ, DosReis S, et al. Psychotropic practice patterns for youth: a 10-year perspective. *Arch Pediatr Adolesc Med* 2003;157(1):17–25.
6. Huang Z, Hankinson SE, Colditz GA, et al. Dual effect of weight and weight gain on breast cancer risk. *JAMA* 1997;278:1407–1411.
7. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukemia and brain tumors: a retrospective cohort study. *Lancet* 2012;380:499–505.
8. Newman TB, Liljestrand P, Jeremy RJ, et al. Outcomes of newborns with total serum bilirubin levels of 25 mg/dL or more. *N Engl J Med* 2006;354:1889–1900.
9. Escobar GJ, Liljestrand P, Hudes ES, et al. Five-year neurodevelopmental outcome of neonatal dehydration. *J Pediatr* 2007;151(2):127–133, 133 e1.
10. Wagoner JK, Archer VE, Lundin FE, et al. Radiation as the cause of lung cancer among uranium miners. *N Engl J Med* 1965;273:181–187.