chapter 4

FREQUENCY

In Chapter 1, we outlined the central questions facing clinicians as they care for patients. In this chapter, we will build a foundation for the evidence that clinicians use to guide their diagnostic and therapeutic decisions. Let us introduce the subject with a patient.

A 22-year-old man presents with sore throat, fever, and malaise of 2 days duration. Further history indicates no exposure to sick persons and no prior history of significant illness. Physical examination reveals a temperature of 38°C, an erythematous pharynx with whitish exudate and tonsillar enlargement, tender anterior cervical lymph nodes, and no other positive findings.

In planning further diagnosis and treatment, the clinician must deal with several questions:

1. How likely is the patient to have streptococcal pharyngitis?
2. If the patient has streptococcal infection, how likely is he to develop a serious complication, such as acute rheumatic fever or acute glomerulonephritis?
3. How likely is penicillin treatment to prevent rheumatic fever or glomerulonephritis?
4. If the patient is treated with penicillin, how likely is an important allergic reaction?

Depending on the answers to these questions, the physician may treat with penicillin right away, obtain a throat culture and await the result, or offer only symptomatic treatment.

Each of these questions concerns the likelihood or commonness of a clinical event under certain circumstances. The questions could all be recast so as to ask—How frequently do streptococcal pharyngitis or rheumatic fever or penicillin allergic reactions occur under particular circumstances?

The evidence required to manage this patient rationally—the likelihood or frequency of disease or outcomes—is, in general, the kind of evidence

needed to answer most clinical questions. Decisions are guided by the commonness of things. Usually, they depend on the relative commonness of things under alternative circumstances: in the presence of a positive test versus a negative test or after treatment A versus treatment B. Because the commonness of disease, improvement, deterioration, cure, or death forms the basis for answering most clinical questions, this chapter will examine measures of clinical frequency.

ASSIGNING NUMBERS TO PROBABILITY STATEMENTS

Physicians often communicate probabilities as words—"usually," "sometimes," "rarely," etc.—rather than as numbers. Substituting words for numbers is convenient and avoids making a precise statement when one is uncertain about a probability. However, it has been shown that there is little agreement about the meanings of commonly used words for frequency.

Example—Physicians were asked to estimate the likelihood of disease for each of 30 expressions of probability found by reviewing radiology and laboratory reports. There was great difference of opinion for each expression. Probabilities for "consistent with" ranged from .18 to .98; for "unlikely," the range was .01 to .93. These data support the authors' assertion that "difference of opinion among physicians regarding the management of a problem may reflect differences in the meaning ascribed to words used to define probability" (1).

Patients also assign widely varying values for expressions of probability. In another study, highly skilled and professional workers thought "usually" referred to probabilities of .35 to 1.0 (± 2 standard deviations from the mean); "rarely" meant to them a probability of 0 to .15 (2).

Thus, substituting words for numbers diminishes the information conveyed. We advocate using numbers whenever possible.

PREVALENCE AND INCIDENCE

In general, clinically relevant measures of the frequency of events are fractions in which the numerator is the number of patients experiencing the outcome (cases) and the denominator is the number of people in whom the outcome could have occurred. Such fractions are of course proportions, but by common usage, are often referred to as "rates." As ex-students of physics, we recognize the incorrectness of this use of rate, but there seems to be little chance that it will disappear.

Clinicians encounter two measures of commonness—prevalence and incidence.

A *prevalence* is the fraction (proportion) of a group possessing a clinical condition at a given point in time.[a] Prevalence is measured by surveying a

[a] There are two kinds of prevalence. *Point prevalence* is measured at a single point in time for each person, although not necessarily for all the people in the defined population. *Period prevalence* is a count of the proportion of cases that were present at any time during a period of time.

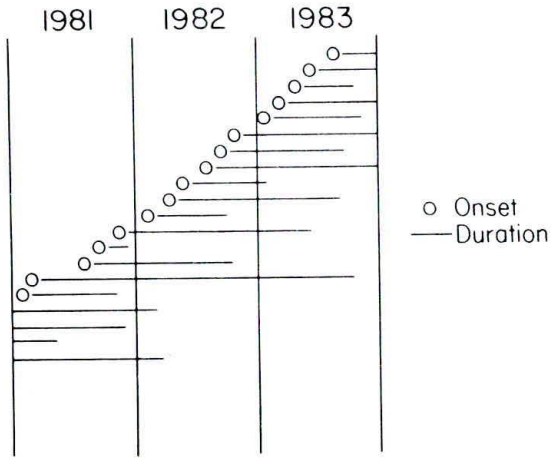

Figure 4.1. Occurrence of disease in 100 people at risk from 1981–1983.

defined population containing people with and without the condition of interest, at a single point in time.

An *incidence* is the fraction or proportion of a group initially free of the condition that develops it over a given period of time. As described later in this chapter and in greater detail in Chapter 5, incidence is measured by identifying a susceptible group of people (i.e., people free of the disease or the outcome) and examining them periodically over an interval of time so as to discover and count new cases that develop during the interval.

To illustrate the differences between prevalence and incidence, Figure 4.1 shows the occurrence of disease in a group of 100 people over the course of 3 years (1981, 1982, 1983). As time passes, individuals in the group develop the disease. They remain in this state until they either recover or die. In the 3 years, 16 people suffer the onset of disease and 4 already had it. Eighty do not develop disease and do not appear on the figure.

At the beginning of 1981 there are four cases, so the prevalence at that point in time is 4/100. If all 100 individuals, including prior cases, are examined at the beginning of each year, one can compute the prevalence at those points in time. At the beginning of 1982, the prevalence is 5/100 because two of the pre-1981 cases lingered on into 1982 and two of the new cases developing in 1981 terminated (hopefully in a cure) before the examination at the start of 1982. Prevalences can be computed for each of the other two annual examinations, and assuming that none of the original 100 people died, moved away, or refused examination, these prevalences are 7/100 at the beginning of 1983 and 5/100 at the beginning of 1984.

To calculate the incidence of new cases developing in the population, we consider only the 96 individuals free of the disease at the beginning of 1981 and what happens to them over the next 3 years. Five new cases

developed in 1981; six new cases developed in 1982, and five additional new cases developed in 1983. The 3-year incidence of the disease is all new cases developing in the 3 years (16) divided by the number of susceptible individuals at the beginning of the follow-up period (96), or 16/96 in 3 years. What would be the annual incidences for 1981, 1982, and 1983, respectively? Remembering to remove the previous cases from the denominator, the annual incidences would be 5/96 for 1981, 6/91 for 1982, and 5/85 for 1983.

Every measure of disease frequency of necessity contains some indication of time. With measures of prevalence, time is assumed to be instantaneous, as in a single frame from a motion picture. Prevalence depicts the situation at that point in time for each patient even though it may, in reality, have taken several weeks or months to collect observations on the various people in the group studied. For incidence, time is the essence because it defines the interval during which susceptible subjects were monitored for the emergence of the event of interest. Two distinct approaches to the assessment of incidence are encountered in the medical literature and are described below.

Table 4.1 summarizes the characteristics of incidence and prevalence. Although the distinctions between the two seem clear, the literature is replete with misuses of the terms, particularly incidence (3).

Why is it important to know the difference between prevalence and incidence? Because they are answers to two different questions: (1) What proportion of a group of people have a condition? and (2) at what rate do new cases arise in a group of people as time passes? The answer to one question cannot be obtained directly from the answer to the other.

MEASURING PREVALENCE AND INCIDENCE

Prevalence Studies

The prevalence of disease is measured by surveying a group of people, some of whom are diseased at that point in time while others are healthy.

Table 4.1
Characteristics of Incidence and Prevalence

	Incidence	Prevalance
Numerator	New cases occurring during a period of time among a group initially free of disease	All cases counted on a single survey or examination of a group
Denominator	All susceptible people present at the beginning of the period	All people examined, including cases and noncases
Time	Duration of the period	Single point
How measured	Cohort study (see Chapter 5)	Prevalence (cross-sectional) study

The fraction or proportion of the group who are diseased (i.e., cases) constitutes the prevalence of the disease.

Such one-shot examinations or surveys of a population of individuals including cases and noncases are called *prevalence* studies. Another term is *cross-sectional studies* because people are studied at a point (cross-section) in time. They are among the more common types of research designs reported in the medical literature, constituting approximately one-third of original articles in major medical journals.

The following is an example of a typical prevalence study.

Example—What is the prevalence of rheumatoid arthritis in the general population? To answer this question, O'Sullivan and Cathcart surveyed all 4552 of the people over age 15 living in a small town in Massachusetts. Each participant completed a questionnaire and underwent an examination that included a medical history, physical examination, and blood tests. The presence of rheumatoid arthritis was defined by explicit criteria in general use: the New York and the American Rheumatology Association (ARA) criteria.

Of the 77% of the defined population who participated, the prevalence of rheumatoid arthritis was about 4 cases per 1000 by the New York criteria and 26 per 1000 by the ARA criteria (4).

Incidence Studies

In contrast to prevalence, incidence is measured by first identifying a population free of the event of interest and then following them through time with periodic examinations to determine occurrences of the event. This process, also called a cohort study, will be discussed in detail in Chapter 5.

Up until now, we have defined incidence as the rate of new events in a group of people of fixed size, all of whom are observed over a period of time. This is called *cumulative incidence* because new cases are accumulated over time.

Example—The death rate after acute respiratory failure complicating chronic respiratory disease was studied by observing the survival of 145 patients. After 1 year, 90 patients had died, for a death rate (incidence of death) of 90/145/year. After 5 years, the death rate was 122/145/5 years (5).

A second approach to incidence is to measure the number of new cases emerging in an ever-changing population, where people are under study and susceptible for varying lengths of time. Typical examples are clinical trials of chronic treatment in which eligible patients are enrolled over several years so that early enrollees are treated and followed longer than late enrollees. In an effort to keep the contribution of individual subjects commensurate with their follow-up interval, the denominator of the incidence measure in these studies is not persons at risk for a specific time period but person-time at risk of the event. An individual followed for 10 years without becoming a case contributes 10 person-years, whereas an individual followed for 1 year contributes only one person-year to the

denominator. An incidence of this type is expressed as the number of new cases per total number of person-years at risk and is sometimes called an *incidence density*.

The person-years approach is also useful for estimating the incidence of disease in large populations of known size when an accurate count of new cases and an estimate of the population at risk are available—for example, a population-based cancer registry.

A disadvantage of the incidence density approach is that it lumps together different lengths of follow-up. A small number of patients followed for a long time can contribute as much to the denominator as a large number of patients followed for a short time. If these long-term follow-up patients are systematically different from short-term follow-up patients, the resulting incidence measures may be biased.

INTERPRETING MEASURES OF CLINICAL FREQUENCY

In order to make sense of prevalence and incidence, the first step is a careful evaluation of the numerator and denominator. Two questions serve to guide this evaluation: What is a case, and what is the population?

What is a "Case"?—Defining the Numerator

Up to this point, the general term "case" has been used to indicate a disease or outcome the frequency of which is of interest. Classically, prevalence and incidence refer to the frequency of a disease among groups of people. However, clinical decisions often depend on information about the frequency of disease manifestations, such as symptoms, signs, or laboratory abnormalities, or the frequency of disease effects, such as death, disability, symptomatic improvement, etc.

To interpret rates, it is necessary to know the basis upon which a case is defined, because the criteria used to define a case can strongly affect rates.

Example—One simple way to identify a case is to ask people whether they have a certain condition. How does this method compare to more rigorous methods? In the Commission on Chronic Illness study, the prevalences of various conditions, as determined by personal interviews in the home, were compared to the prevalences as determined by physician examination of the same individuals. Figure 4.2 illustrates the interview prevalences and the clinical examination prevalences for various conditions.

The data illustrate that these two methods of defining a case can generate very different estimates of prevalence and in different directions, depending on the condition (6).

For some conditions, broadly accepted, explicit diagnostic criteria are available. The American Rheumatism Association criteria for rheumatoid arthritis (Table 4.2) are an example (7). These criteria demonstrate the extraordinary specificity required to define reliably so common a disease as rheumatoid arthritis. They also illustrate a trade-off between rigorous

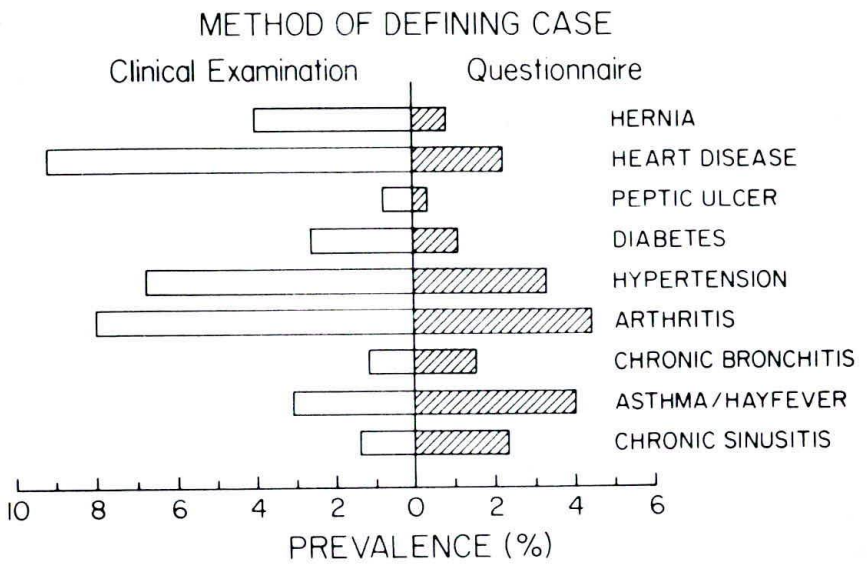

Figure 4.2. Prevalence depends on the definition of a case. The prevalence of diseases in the general population based on people's opinions (survey) and clinical evaluation. (Data from Sanders BS: Have morbidity surveys been oversold? *Am J Public Health* 52:1648–1659, 1962.

definition and clinical reality. If only "classic" cases were included in a rate, most patients who would ordinarily be considered to have the disease would not be included. On the other hand, including "probable" cases could overestimate the true rate of disease.

What is the Population?—Defining the Denominator

In order to make sense out of the number of cases, we must have a clear picture of the size and characteristics of the group of individuals in which the cases arose. A rate is useful only to the extent that the individual practitioner can decide to which kinds of patients the rate applies.

Customarily, the group indicated in the denominator of a rate is referred to as the population or, more particularly, the *population at risk*, where "at risk" means susceptible to the disease or outcome counted in the numerator. For example, it is not meaningful to describe the incidence or prevalence of cervical cancer in a population that includes women who have had hysterectomies or includes men.

Ideally, the denominator of a rate would include all people who could have the condition or a representative sample of them. But what is relevant depends on one's perspective. For example, if we wanted to know the true prevalence of rheumatoid arthritis in Americans, we would prefer to include in the denominator all people in the United States, rather than

Table 4.2
Rheumatoid Arthritis Diagnostic Criteria (American Rheumatism Association 1958 Revision)[a]

1. Morning stiffness.
2. Pain on motion or tenderness in at least one joint.[b]
3. Swelling (soft tissue thickening or fluid, not bony overgrowth alone) in at least one joint.[b]
4. Swelling of at least one other joint.[b]
5. Symmetrical joint swelling with simultaneous involvement of the same joint on both sides of the body.[b] Terminal phalangeal joint involvement will not satisfy the criterion.
6. Subcutaneous nodules over bony prominences, on extensor surfaces, or in juxta-articular regions.[b]
7. Roentgenographic changes typical of rheumatoid arthritis (which must include at least bony decalcification localized to or greatest around the involved joints and not just degenerative changes).
8. Positive agglutination (anti-gammaglobulin) test.
9. Poor mucin precipitate from synovial fluid (with shreds and cloudy solution).
10. Characteristic histologic changes in synovial membrane.
11. Characteristic histologic changes in nodules.

CATEGORIES	NUMBER OF CRITERIA REQUIRED	MINIMUM DURATION OF CONTINUOUS SYMPTOMS
Classic	7 of 11	6 weeks (Nos. 1–5)
Definite	5 of 11	6 weeks (Nos. 1–5)
Probable	3 of 11	6 weeks (1 of Nos. 1–5)

[a] Adapted from Ropes MW, Bennett CA, Cobb S, Jacox R, Jessar RA: 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bull Rheum Dis* 9:175–176, 1958.
[b] Observed by physician.

patients in office practice. But if one wanted to know the prevalence of rheumatoid arthritis in office practice—perhaps in order to plan services—the relevant denominator would be patients seen in office practice, not people in the population at large. In one survey, only 25% of adults found to have arthritic and rheumatic complaints (not necessarily rheumatoid arthritis) during a community survey had received services for such complaints from any health professional or institution (8).

It is customary for epidemiologists to think of a population as consisting of all individuals residing in a geographic area. And so it should be for studies of cause-and-effect in the general population. But in studies of clinical questions, the relevant populations generally consist of patients suffering from certain diseases or exhibiting certain clinical findings, and who are found in clinical settings that are similar to those in which the information will be used. Commonly, such patients are assembled at a limited number of clinical facilities where academic physicians see patients. In these instances, the population includes all patients with the appropriate

findings from the hospitals or clinics involved. They may be a small and peculiar subset of all patients with the findings in some geographic area, and even unusual for office practice in general.

What difference might the choice of a population make? What is at issue is the generalizability of observed rates. As discussed in Chapter 1, the incidence of further seizures in children who have had one febrile seizure varied from about 5% in the general population to as high as 75% in some clinics. Knowing which incidence is appropriate to one's patients is critical because it will influence the decision whether to begin chronic anticonvulsant treatment. The appropriate incidence depends upon the location and nature of the reader's practice. If the reader is an academic pediatric neurologist, referral center experience is more relevant. If the reader is a family physician or pediatrician providing community-based primary care, referral center experience may be irrelevant. Some of the authors reporting high incidences of subsequent seizures in children seen in referral centers argued that their high rate indicated that all such children should receive long-term anticonvulsant treatment. Such a conclusion may not be justified for the clinician in primary care practice, where the incidence of subsequent seizures is less than 5%.

Sampling

It is rarely possible to study all the people who have or might develop the condition of interest. Usually one takes a sample, so that the number studied is of manageable size. This raises a question: Is the sample representative of the population?

In general, there are two ways to sample. In a *random sample*, every individual in the population has an equal probability of being selected. The more general term *probability sample* is used if every person has a known (not necessarily equal) probability of being selected. On the average, the characteristics of people in probability samples are similar to those of the population from which they were selected, particularly if a large number are chosen.

Other methods of selecting samples may well be biased and so do not necessarily represent the parent population. Most groups of patients described in the medical literature, and found in most clinicians' experience, are based on biased samples. Typically, patients are included in studies because they are under care in an academic institution, available, willing to be studied, and perhaps also particularly interesting and/or severely affected. There is nothing wrong with this practice—as long as it is understood to whom the results do (or do not) apply.

RELATIONSHIP AMONG INCIDENCE, PREVALENCE, AND DURATION OF DISEASE

As described previously anything that increases the duration of the clinical findings in a patient will increase the chance that that patient will be identified in a prevalence study. The relationship among incidence and

Table 4.3
The Relationships Among Incidence, Prevalence and Duration[a] of Disease: Asthma in the United States[b]

AGE	ANNUAL INCIDENCE	PREVALENCE	DURATION = $\frac{\text{PREVALENCE}}{\text{ANNUAL INCIDENCE}}$
0–5	6/1000	29/1000	4.8 years
6–16	3/1000	32/1000	10.7 years
17–44	2/1000	26/1000	13.0 years
45–64	1/1000	33/1000	33.0 years
65+	0	36/1000	33.0 years
	3/1000	30/1000	10.0 years

[a] Duration = $\frac{\text{Prevalence}}{\text{Annual Incidence}}$.

[b] Approximated from several sources.

prevalence and duration of disease in a steady state—that is, where none of the variables is changing much over time—is approximated by the expression:

$$\text{Prevalence} \approx \text{Incidence} \times \text{Average Duration of the Disease}$$

Example—Table 4.3 shows approximate annual incidence and prevalence rates for asthma. Incidence falls with increasing age, illustrating the fact that the disease arises primarily in childhood. But prevalence stays fairly stable over the entire age span, indicating that asthma tends to be chronic and is especially chronic among older individuals. Also, because the pool of prevalent cases does not increase in size, about the same number of patients are recovering from their asthma as new patients are acquiring it.

If we use the formula (Prevalence ÷ Incidence = Average Duration), we can determine that asthma has an average duration of 10 years. When the duration of asthma is determined for each age category by dividing the prevalences by the incidences, it is apparent that the duration of asthma increases with increasing age. This reflects the clinical observation that childhood asthma often clears with time, whereas adult asthma tends to be more chronic.

BIAS IN PREVALENCE STUDIES

Prevalence studies can be used to investigate potentially causal relationships between risk factors and a disease. For this purpose, they are quick but inferior alternatives to incidence studies. Two biases are particularly troublesome: temporal sequence and old versus new cases.

Interpreting Temporal Sequences

In prevalence studies, disease and the possible factors responsible for the disease are measured simultaneously, and so it is often unclear which came before the other. The time dimension is lost, and if it is included in the interpretation it must be inferred. In contrast, studies of incidence do have

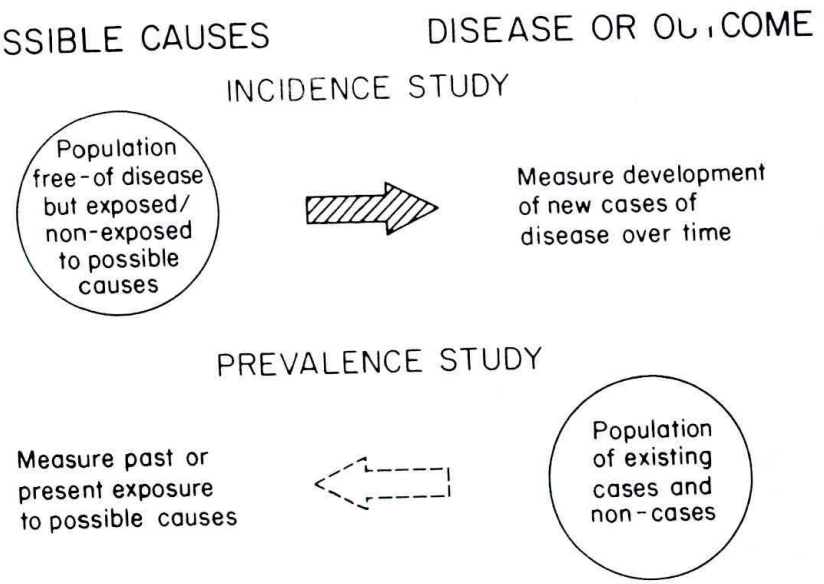

Figure 4.3. Temporal relationship between possible causal factors and disease for incidence and prevalence studies.

a built-in sequence of events because possible causes of disease are measured initially, before disease has occurred. These relationships are illustrated in Figure 4.3.

Old Versus New Cases

The difference between cases found in the numerator of incidences and of prevalences is illustrated in Figure 4.4. In a cohort study, most new cases can be ascertained if a susceptible population is followed carefully through time. On the other hand, prevalence surveys include old as well as new cases, and they include only those cases that are available at the time of a single examination—that is, they identify only cases that happen to be both active (i.e., diagnosable) and alive at the time of the survey. Obviously, prevalences will be dominated by those patients who are able to survive their disease without losing its manifestations.

In many situations, the kinds of cases included in the numerator of an incidence are quite different from the kinds of cases included in the numerator of a prevalence. The differences may influence how the rates are interpreted.

Prevalence is affected by the average duration of disease. Rapidly fatal episodes of the disease would be included in an incidence, but most would be missed by a prevalence survey. For example, 25–40% of all deaths from coronary heart disease occur within 24 hours of the onset of symptoms in

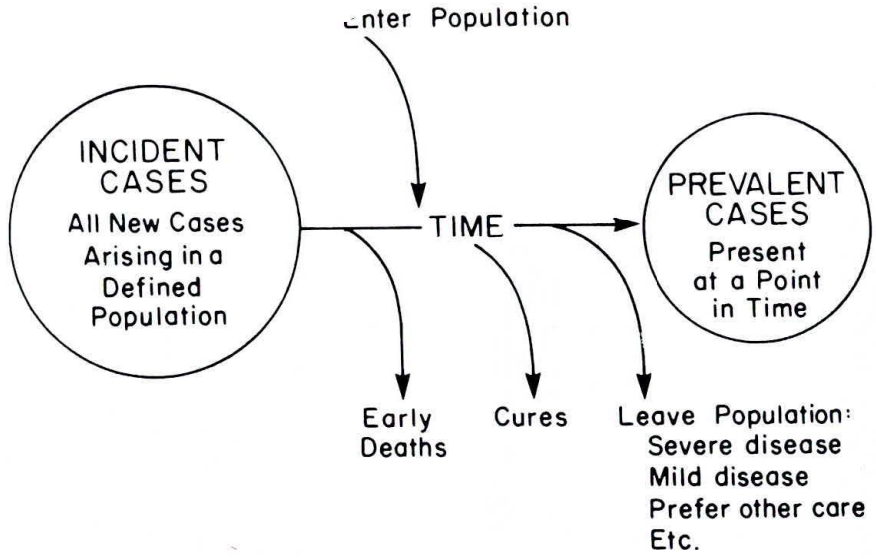

Figure 4.4. The difference in cases for incidence and prevalence studies.

individuals with no prior evidence of disease. A prevalence survey would, therefore, underestimate cases of coronary heart disease. On the other hand, diseases of long duration are well represented in prevalence surveys, even if their incidence is low. For example, although the incidence of Crohn's disease is only about 2–7/100,000/year, its prevalence is over 100/100,000, reflecting the chronic nature of the disease (9).

Prevalence surveys can also selectively include more severe cases of disease, ones that are particularly sustained and obtrusive. For example, patients with rheumatoid arthritis who are not currently active would not be included in a survey based on current symptoms and physical findings. Similarly, patients with recurrent but controllable illnesses, such as congestive heart failure or depression, may be well at a given point in time and therefore might not be discovered on a single examination. Unremitting disease, on the other hand, is less likely to be overlooked and, therefore, would contribute disproportionately to the pool of cases assembled by a prevalence survey.

USES OF INCIDENCE AND PREVALENCE

What purposes do incidence and prevalence serve? Clinicians use them in three different ways: predicting the future, describing things as they are, and making comparisons.

Predicting the Future

Incidence is a description of the rate at which a disease has arisen over

time in a group of people assembled in the past. It can also be used to predict the probability that similar people will develop the condition in the future. For incidence, the sequence of events is clear because the population is known to be free of the outcome at the outset and all cases are assessed.

On the other hand, as pointed out above, prevalence describes the situation among a group of individuals at a given point in time; it offers no sound basis for predicting the future. If 30% of patients with stroke are depressed, this does not mean that 30% of stroke patients will become depressed in the future. It may be that depression predisposes to stroke or that nondepressed stroke patients are more likely to recover quickly. Because of the way in which they are measured, prevalences often reveal little about the sequence of events and only include a fraction of all possible cases. Thus, they are treacherous grounds for predicting the future.

The Probability that a Patient Has the Condition

Prevalence is particularly useful in guiding decisions about whether or not to use a diagnostic test, as pointed out in Chapter 3 because prevalence is a determinant of predictive value. Knowing that a patient with a combination of demographic and clinical characteristics has a given probability of having the disease not only influences the interpretation of a diagnostic test result but also may affect powerfully the selection among various treatment options.

The patient with pharyngitis, presented at the beginning of this chapter, illustrates how variations in prevalence can influence the approach to a clinical problem.

Example—Three approaches to the treatment of pharyngitis were compared. Their value was judged by weighing the potential benefits of preventing rheumatic fever against the costs of penicillin allergy. The three options were to obtain a throat culture and treat only those patients with throat cultures positive for β-hemolytic Group A streptococci, treat all patients without obtaining a culture, and neither culture nor treat any patient.

The analysis revealed that the optimal strategy depended upon the likelihood that a patient would have a positive culture, which can be estimated from the prevalence of streptococcal infection in the community at the time and the presence or absence of such clinical findings as fever. It was concluded that, if the probability of a positive culture for an individual patient exceeds 20%, the patient should be treated; if it is less than 5%, the patient should not be cultured or treated; and if the probability lies between 5% and 20%, the patient should be cultured first and treated based on the result (10).

This study represents a rational approach to the use of prevalences as indicators of individual probabilities of disease in guiding clinical decision making.

Making Comparisons

Although isolated incidences and prevalences serve useful functions, as described previously, they become much more powerful tools in support of clinical decisions when used to make comparisons. It is the comparison between the frequencies of disease among individuals exposed to a factor and individuals not exposed to the factor that provides the best evidence suggesting causality, not just the commonness of the disease among those exposed. For example, the risk (incidence) of lung cancer among males who smoke heavily is of the order of 0.17% per year, hardly a common event. Only when this incidence is contrasted with the incidence in nonsmokers (approximately 0.007% per year) does the devastating effect of smoking emerge. Clinicians use measures of frequency as the ingredients in comparative measures of the association between a factor and the disease or disease outcome. Ways of comparing rates will be described in more detail in Chapter 5.

SUMMARY

Most clinical questions are answered by reference to the commonness of events under varying circumstances. The commonness of clinical events is indicated by proportions or fractions, the numerators of which include the number of cases and the denominators of which include the number of people from whom the cases arose.

There are two measures of commonness—prevalence and incidence. Prevalence is the proportion of a group who have the disease at a single point in time. Incidence is the proportion of a susceptible group who develop new cases of the disease over an interval of time.

Prevalence is measured by a single survey of a group containing cases and noncases, whereas measurement of incidence requires examinations of a previously disease-free group over time. Thus, prevalence studies identify only those cases who are alive and diagnosable at the time of the survey, whereas cohort (incidence) studies ascertain all new cases. Prevalent cases, therefore, may be a biased subset of all cases because they do not include those who have already succumbed or been cured. Additionally, prevalence studies frequently do not permit a clear understanding of the temporal relationship between a causal factor and a disease.

To make sense of incidence and prevalence, the clinician must understand the basis upon which the disease is diagnosed and the characteristics of the population represented in the denominator. The latter is of particular importance in trying to decide if a given measure of incidence or prevalence pertains to patients in one's own practice.

Incidence is the most appropriate measure of commonness with which to predict the future. Prevalence serves to quantitate the likelihood that a patient with certain characteristics has the disease at a single point in time and is used for decisions about diagnosis and screening. The most powerful use of incidence and prevalence, however, is to compare different clinical alternatives.

POSTSCRIPT

Counting clinical events as described in this chapter may seem to be the

most mundane of tasks. It seems so obvious that examining counts of clinical events under various circumstances is the foundation of clinical science. It may be worth reminding the reader that Pierre Louis introduced the "numerical method" of evaluating therapy less than 200 years ago. Dr. Louis had the audacity to count deaths and recoveries from febrile illness in the presence and absence of blood-letting. He was excoriated for allowing lifeless numbers to cast doubt on the healing powers of the leech, powers that had been amply confirmed by decades of astute qualitative clinical observation.

REFERENCES

1. Bryant GD, Norman GR: Expressions of probability: words and numbers. *N Engl J Med* 302:411, 1980.
2. Toogood JH: What do we mean by "usually"? *Lancet* 1:1094, 1980.
3. Friedman GD: Medical usage and abusage, "prevalence" and "incidence." *Ann Intern Med* 84:502–503, 1976.
4. O'Sullivan JB, Cathcart ES: The prevalence of rheumatoid arthritis. Follow-up evaluation of the effect of criteria on rates in Sudbury, Massachusetts. *Ann Intern Med* 76:573–577, 1972.
5. Asmundsson T, Kilburn KH: Survival after acute respiratory failure. *Ann Intern Med* 80:54–57, 1974.
6. Sanders BS: Have morbidity surveys been oversold? *Am J Publ Health* 52:1648–1659, 1962.
7. Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA: 1958 revision of diagnostic criteria for rheumatoid arthritis. JJ Bunim (Ed). *Bull Rheum Dis* 9:175–176, 1958.
8. Spitzer WO, Harth M, Goldsmith CH, Norman GR, Dickie GL, Bass MJ, Newell JP: The arthritic complaint in primary care: prevalence, related disability, and costs. *J Rheum* 3:88–99, 1976.
9. Sedlack RE, Whisnant J, Elveback LR, Kurland LT: Incidence of Crohn's disease in Olmsted County, Minnesota, 1935–1975. *Am J Epidemiol* 112:759–763, 1980.
10. Tompkins RK, Burnes DC, Cable WE: An analysis of the cost-effectiveness of pharyngitis management and acute rheumatic fever prevention. *Ann Intern Med* 86:481–492, 1977.

SUGGESTED READINGS

Ellenberg JH, Nelson KB: Sample selection and the natural history of disease: Studies of febrile seizures. *JAMA* 243:1337–1340, 1980.

Friedman GD: Medical usage and abusage, "prevalence" and "incidence." *Ann Intern Med* 84:502–503, 1976.

Morgenstern H, Kleinbaum DG, Kupper LL: Measures of disease incidence used in epidemiologic research. *Int J Epidemiol* 9:97–104, 1980.

APPENDIX 4.1. MAIN QUESTIONS FOR DETERMINING THE VALIDITY OF STUDIES OF PREVALENCE[a]

1. What are the criteria for being a *case*?
2. What is the defined *population*?
3. Are cases and noncases from *an unbiased sample* of the population ?

[a] These questions are not meant to be all-inclusive nor to replace independent, critical thinking. They are a rough guideline, including only the most basic elements of a sound study.

chapter 5

RISK

Risk generally refers to the probability of some untoward event. In this chapter, the term "risk" is used in a more restricted sense to describe the likelihood that people who are without a disease, but are exposed to certain factors ("risk factors"), will acquire the disease.

Many people in our society have a strong interest in their risk of disease. Their concern has spawned many popular books about risk reduction and is reflected in newspaper headlines about the risk of cancer from exposure to toxic chemicals or nuclear accidents, of cardiovascular disease after use of birth control pills and of AIDS from sexual behavior or transfusion.

In this chapter, we will consider how estimates of risk are obtained by observing the relationship between exposure to possible risk factors and the subsequent incidence of disease. Then we will describe several ways of comparing risks, both as they affect individuals and populations.

RISK FACTORS

Factors that are associated with an increased risk of becoming diseased are called *risk factors*. There are several kinds of risk factors. Some, such as toxins, infectious agents, and drugs, are found in the physical environment. Others are part of the social environment. For example, disruption of family (e.g., loss of a spouse), daily routines, and culture has been shown to increase rates of disease—not only emotional but physical illness as well. Other risk factors are behavioral; among them are smoking, inactivity, and driving without seat belts. Risk factors are also inherited. For example, having the haplotype HLA B27 greatly increases the risk of acquiring the spondylarthropathies.

Exposure to a risk factor means that a person has, before becoming ill, come in contact with or has manifested the factor in question. Exposure can take place at a single point in time, as when a community is exposed to radiation during a nuclear accident. More often, however, exposure to risk factors for chronic disease takes place over a period of time. Cigarette

smoking, hypertension, sexual promiscuity, and sun exposure are examples. There are many different ways of characterizing the dose of chronic exposure: ever exposed, current dose, largest dose taken, total cumulative dose, years of exposure, years since first exposure, etc. (1). Although the various measures of dose tend to be related to each other, some may show an exposure-disease relationship, whereas others do not. For example, cumulative dose of sun exposure is a risk factor for nonmelanoma skin cancer, whereas episodes of severe sunburn is a better predictor of melanoma. Choice of an appropriate measure of exposure is usually based on all that is known about the biologic effects of the exposure and the pathophysiology of the disease.

INFORMATION ABOUT RISK

Large and dramatic risks are easy for anyone to appreciate. Thus, it is not difficult to recognize the relationship between exposure and disease for such conditions as chickenpox, sunburn, or aspirin overdose because they follow exposure in a relatively rapid, certain, and obvious way. But much of the morbidity and mortality in our society is caused by chronic diseases. For these, the relationships between exposure and disease are far less obvious. It becomes virtually impossible for individual clinicians, however astute, to develop estimates of risk based on their own experiences with patients. This is true for several reasons, which are discussed below and summarized in Table 5.1.

- **Long latency**

Many chronic diseases have long latency periods between exposure to risk factors and the first manifestations of disease. Patients exposed during one time in a clinician's professional life may experience the consequences in another, years later, when the original exposure is all but forgotten. The link between exposure and disease is thereby obscured.

- **Frequent exposure to risk factors**

Many risk factors—such as cigarette smoking or driving when intoxicated—occur so frequently in our society that they scarcely seem dangerous. Only by comparing patterns of disease in other populations, or

Table 5.1
Situations in which Personal Experience is Insufficient to Establish a Relationship Between Exposure and Disease

Long latency period between exposure and disease
Frequent exposure to risk factor
Low incidence of disease
Small risk from exposure
Common disease
Multiple causes of disease

investigating special subgroups within our own (e.g., Mormons who neither smoke nor drink), can we recognize risks that are in fact rather large.

- **Low incidence of disease**

Most diseases, even ones thought to be "common," are actually quite rare. Thus, although lung cancer is the most common kind of cancer in Americans, the yearly incidence of lung cancer even in heavy smokers is less than 2/1,000. In the average physician's practice, years may pass between new cases of lung cancer. It is difficult to draw conclusions about such infrequent events.

- **Small risk**

If a factor confers only a small risk, a large number of "cases" are required to observe a difference in disease rates between exposed and unexposed people. This is so even if both the risk factor and the disease occur relatively frequently. It is still uncertain whether coffee and diabetes are risk factors for carcinoma of the pancreas, because estimates of risk are all small and, therefore, easily discounted as resulting from bias or chance. In contrast, it is not controversial that hepatitis B infection is a risk factor for hepatoma, because people with evidence of hepatitis B infection are hundreds of times more likely to get liver cancer than those without it.

- **Common disease**

If the disease is one of those ordinarily occurring in our society—heart disease, cancer, or stroke—and some of the risk factors for it are already known, it becomes difficult to distinguish a new risk factor from the others. Also, there is less incentive to look for a new risk factor. For example, the syndrome of sudden, unexpected death in adults is a common way to die. Many cases seem related to coronary heart disease. However, it is entirely conceivable that there are other important causes, as yet unrecognized because an adequate explanation for most cases is available.

On the other hand, rare diseases invite efforts to find a cause. Phocomelia is such an unusual congenital malformation that the appearance of just a few cases raised suspicion that some new agent (as it turned out, the drug, thalidomide) might be responsible. Similarly, physicians were quick to notice when several cases of carcinoma of the vagina, a very rare condition, began appearing. A careful search for an explanation was undertaken, and maternal exposure to diethylstilbestrol was found.

- **Multiple causes and effects**

There is usually not a close, one-to-one, relationship between a risk factor and one particular disease. Some people with hypertension develop congestive heart failure and many do not. Many people who do not have hypertension develop congestive heart failure as well. The relationship between hypertension and congestive heart failure is obscured by the fact that there are several other causes of the disease, and hypertension causes several diseases. Thus, although people with hypertension are about three times more likely to develop congestive heart failure and hypertension is

the leading cause of that condition, physicians were not particularly attuned to this relationship until recently, when adequate data became available.

For these reasons, individual clinicians are rarely in a position to confirm associations between exposure and disease, though they may suspect them. For accurate information, they must turn to the medical literature, particularly studies that are carefully constructed and involve a large number of patients.

USES OF RISK

Information about risk serves several purposes.

Prediction

Risk factors are used, first and foremost, to predict the occurrence of disease. The quality of predictions depends on the similarity of the people on whom the estimate is based to the people for whom the prediction is made.

Although risk factors may signify an individual's increased risk of disease, relative to an unexposed person, their presence does not mean that an individual is very likely to get the disease. Most people, even those with many strong risk factors, are unlikely to get a disease—at least over several years' time. Thus, a heavy cigarette smoker, who has a twenty-fold increase in the risk of lung cancer compared to nonsmokers, nevertheless has only a one in a hundred chance of getting lung cancer in the next 10 years.

In individual patients, risk factors usually are not as strong predictors of disease as are clinical findings of early disease. As Rose put it:

> Often the best predictor of future major diseases is the presence of existing minor disease. A low ventilatory function today is the best predictor of its future rate of decline. A high blood pressure today is the best predictor of its future rate of rise. Early coronary heart disease is better than all of the conventional risk factors as a predictor of future fatal disease (2).

Cause

It is often assumed that any excess incidence of disease in exposed versus nonexposed persons is because of exposure to a risk factor. However, risk factors need not be causes. A risk factor may mark a disease outcome indirectly, by virtue of an association with some other determinant(s) of disease—that is, it may be confounded with a causal factor. For example, lack of maternal education is a risk factor for low birth weight infants. Yet, other factors related to education, such as poor nutrition, less prenatal care, cigarette smoking, etc., are more directly the causes of low birth weight.

A risk factor that is not a cause of disease is called a *marker*, because it "marks" the increased probability of disease. Not being a cause does not diminish the value of a risk factor as a way of predicting the probability of disease. But it does imply that removing such a risk factor might not remove the excess risk associated with it.

Diagnosis

The presence of a risk factor increases the probability that a disease is present. Knowledge of risk, therefore, can be used in the diagnostic process, inasmuch as increasing the prevalence of disease among patients tested is one way of improving the performance (positive predictive value) of a diagnostic test.

However, the presence of a risk factor usually increases the probability of disease very little for any one individual at one point in time, compared to other aspects of the clinical situation. For example, age and sex are relatively strong risk factors for coronary artery disease, yet the prevalence of disease in the most at-risk age and sex group, old men, is only 12.3% compared to 0.4% for the least at-risk group, young women. When specifics of the clinical situation, such as type of chest pain and results of an electrocardiographic stress test, are considered as well, the prevalence of coronary disease can be raised to 99.8% for old men and 93.1% for young women (3).

More often, it is helpful to use the absence of a risk factor to help rule out disease, particularly when one factor is strong and predominant. Thus, it would be reasonable to consider mesothelioma in the differential diagnosis of a pleural mass if the patient were an asbestos worker; but mesothelioma would be considerably less likely if the patient had never worked with asbestos. Knowledge of risk factors is also used to improve the efficiency of screening programs by selecting subgroups of patients at increased risk.

Prevention

If a risk factor is also a cause of disease, its removal can be used to prevent disease whether or not the mechanism by which the disease takes place is known. Some of the classic events in the history of epidemiology are illustrations. For example, before bacteria were identified Snow found an increased rate of cholera among people drinking water supplied by a particular company and controlled an epidemic by cutting off that supply. The concept of cause and its relationship to prevention will be discussed in Chapter 11.

PROBABILITY AND THE INDIVIDUAL

The best available information for predicting disease in an individual is past experience with a large number of similar people. For example, an observed incidence of 2/1000/year for the occurrence of lung cancer in heavy smokers becomes an estimate of the probability, 0.002, that an individual heavy smoker will get lung cancer in a year. In practical terms,

incidence is used to estimate the probability that an individual will experience the event of interest. If our knowledge of human disease were more complete, we would not need to resort to probability. But we do not have that luxury.

However, there is a basic incompatibility between the incidence of a disease in groups of people and chances that an individual will contract that disease. Quite naturally, both patients and clinicians would like to answer questions about the future occurrence of disease as precisely as possible. They are uncomfortable about assigning a probability, such as the chances that a person will get lung cancer or stroke in the next 5 years. Moreover, any one person will, at the end of 5 years, either have the disease or not. So in a sense, the average is always wrong for the individual, because it is expressed in different terms.

Nevertheless, probabilities can guide clinical decision making. Even if a prediction does not come true in an individual patient, it will usually be borne out in many such cases. After all, weather forecasts are not always accurate either, but they do help us decide whether to carry an umbrella.

STUDIES OF RISK

There are several scientific strategies for determining risk. In general, there is a trade-off between scientific rigor and feasibility.

Observational Studies

The most satisfactory way of determining whether exposure to a potential risk factor results in an increased risk of disease would be to conduct an experiment. People currently without disease would be divided into groups of equal susceptibility to the disease in question. One group would be exposed to the purported risk factor and the other would not, but the groups would otherwise be treated the same. Later, any difference in observed rates of disease in the groups could be attributed to the risk factor.

Unfortunately, the effects of most risk factors cannot be studied in this way. Consider some of the questions of risk that concern us today. Are inactive people at increased risk for cardiovascular disease, everything else being equal? Does heterosexual exposure lead to AIDS? Do seat belts decrease the risk of dying from an auto accident? For such questions as these, it is usually not possible to conduct an experiment. People become exposed or not to risk factors for reasons that have nothing to do with the scientific value of the information their exposure may provide. As a result, it is usually necessary to study risk in less obtrusive ways.

Clinical studies in which the researcher gathers data by simply observing events as they happen, without playing an active part in what takes place, are called *observational studies*. On the other hand, in *experimental studies*, the researcher determines who is exposed. Although experimental studies are more scientifically rigorous, observational studies are the only feasible way of studying most questions of risk.

Observational studies are subject to a great many more potential biases than are experiments. When people become exposed or not exposed to a certain risk factor in the natural course of events, they are also likely to differ in a great many other ways. If these ways are also related to disease they could account for any association observed between risk factors and disease.

This leads to the main challenge of observational studies: to deal with extraneous differences between exposure groups in order to mimic as closely as possible an experiment. The differences are considered "extraneous" from the point of view of someone trying to determine cause-effect relationships. The following example illustrates one approach to handling such differences.

Example—Although the presence of sickle-cell trait (HbAS) is generally regarded as a benign condition, several studies have suggested that it is associated with defects in physical growth and cognitive development. A study was undertaken, therefore, to see if children born with HbAS experienced problems in growth and development more frequently than children with normal hemoglobin (HbAA), everything else being equal. It was recognized that a great many other factors are related both to growth and development and also to having HbAS. Among these are race, sex, birth date, birth weight, gestational age, 5-minute Apgar score, and socioeconomic status. If these were not taken into account, it would not be possible to distinguish the effects of HbAS, in and of itself, from the effects of the other factors. The authors chose to deal with these other factors by matching. For each child with HbAS, they selected a child with HbAA who was similar with respect to the seven other factors. Fifty newborns with HbAS and 50 with HbAA were followed from birth to 3–5 years old. No differences in growth and development were found (4).

Other ways of dealing with differences between groups will be described in the next chapter (Chapter 6).

Cohorts

The term *cohort* is used to describe a group of people who have something in common when they are first assembled, and who are then observed for a period of time to see what happens to them. Table 5.2 lists some of the ways in which cohorts are used in clinical research.

Whatever members of a cohort have in common, observations of them should fulfill two criteria if they are to provide sound information.

First, cohorts should be observed over a meaningful period of time in the natural history of the disease in question. This is so there will be sufficient time for the risk to be expressed. If we wish to learn whether neck irradiation during childhood results in thyroid neoplasms, a 5-year follow-up would not be a fair test of the risk associated with irradiation, because the usual time period between exposure and the onset of this disease is considerably longer.

Second, all members of the cohort should be observed over the full

Table 5.2
Cohorts and their Purposes

CHARACTERISTIC IN COMMON	TO ASSESS EFFECT OF	EXAMPLE
Age	Age	Life expectancy for people age 70 (regardless of when born)
Date of birth	Calendar time	Tuberculosis rates for people born in 1910
Exposure	Risk factor	Lung cancer in people who smoke
Disease	Prognosis	Survival rate for patients with breast cancer
Preventive intervention	Prevention	Reduction in incidence of pneumonia after pneumococcal vaccination
Therapeutic intervention	Treatment	Improvement in survival for patients with Hodgkin's disease given combination chemotherapy

period of follow-up. To the extent that people drop out of the cohort and their reasons for dropping out are related in some way to the outcome, the information provided by an incomplete cohort can be a distortion of the true state of affairs.

Cohort Studies

In a *cohort study*, a group of people (a cohort) is assembled, none of whom has experienced the outcome of interest. On entry to the study, people in the cohort are classified according to those characteristics that might be related to outcome. These people are then observed over time to see which of them experience the outcome. It is then possible to see how initial characteristics relate to subsequent outcome events. A cohort study is diagrammed in Figure 5.1. Other names for such studies are *longitudinal* (emphasizing that patients are followed over time), *prospective* (implying the forward direction in which the patients are pursued), and *incidence* (calling attention to the basic measure of new disease events over time).

The following is a description of a classical cohort study, which has made an extremely important contribution to our understanding of cardiovascular disease.

Example—The Framingham Study was begun in 1949 to identify factors associated with an increased risk of coronary heart disease (CHD). A representative sample of 5209 men and women, aged 30–59, was selected from approximately 10,000 persons of that age living in Framingham, a small town near Boston. Of these, 5127 were free of CHD when first examined and, therefore, were at risk of

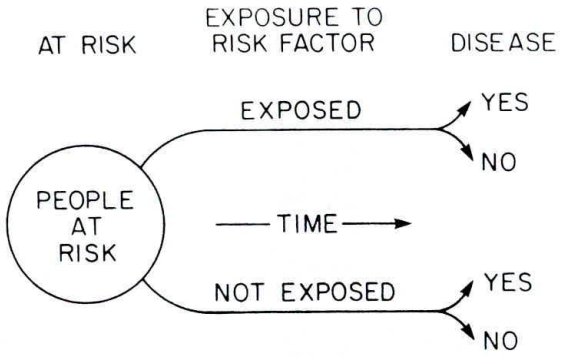

Figure 5.1. Design of a cohort study of risk.

developing CHD subsequently. These people have been re-examined biennially for evidence of coronary disease. The study has run for 30 years and has demonstrated that risk of developing CHD is associated with blood pressure, serum cholesterol, cigarette smoking, glucose intolerance, and left ventricular hypertrophy. There is a large difference in risk between those with none and those with all of these risk factors (5).

Historical Cohort Studies

Cohort studies can be conducted in two ways (Fig. 5.2). The cohort can be assembled in the present and followed into the future (a *concurrent cohort study*); or it can be identified from past records and followed forward from that time up to the present (a *historical cohort study*).

Most of the advantages and disadvantages of cohort studies, as a strategy, apply whether they are concurrent or historical. However, the potential for difficulties with the quality of data is different for the two. In concurrent studies, data can be collected specifically for the purposes of the study and with full anticipation of what is needed. It is thereby possible to avoid biases that might undermine the accuracy of the data. On the other hand, data for historical cohorts are often gathered for other purposes—usually as part of medical records for patient care. These data may not be of sufficient quality for rigorous research.

Disadvantages of Cohort Studies

Cohort studies of risk are the best available substitute for a true experiment, when experimentation is not possible. However, they present a considerable number of practical difficulties of their own. Some of the advantages and disadvantages of cohort studies, for the purpose of describing risk factors, are summarized in Table 5.3.

The principal disadvantage is that, if the outcome is infrequent, and most are, a large number of people must be entered in a study and remain under observation for a long time before results are available. For example,

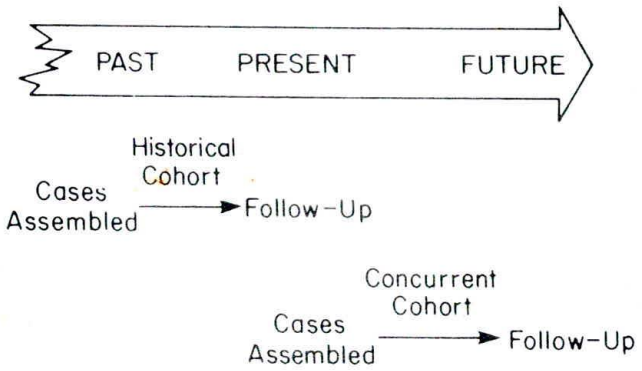

Figure 5.2. Historical and concurrent cohort studies.

Table 5.3
Advantages and Disadvantages of Cohort Studies

ADVANTAGES	DISADVANTAGES
The only way of establishing incidence (i.e., absolute risk) directly	Inefficient, because many more subjects must be enrolled than experience the event of interest; therefore, cannot be used for rare diseases
Follow the same logic as the clinical question: if persons exposed, then do they get the disease?	Expensive because of resources necessary to study many people over time
Exposure can be elicited without the bias that might occur if outcome were already known	Results not available for a long time
Can assess the relationship between exposure and many diseases	Can only assess the relationship between disease and of exposure to relatively few factors (i.e., those recorded at the outset of the study)

the Framingham Study of coronary heart disease was the largest of its kind and studied one of the most frequent of the chronic diseases in America. Nevertheless, over 5000 people had to be followed for several years before the first, preliminary conclusions could be published. Only 5% of the people had experienced a coronary event during the first 8 years!

A related problem with cohort studies results from the fact that the people being studied are usually "free living" and not under the control of researchers. A great deal of effort and money must be expended to keep track of them. Cohort studies, therefore, are expensive, sometimes costing millions of dollars.

Because of the time and money required for cohort studies, this approach cannot be used for all clinical questions about risk. For practical reasons, the cohort approach has been reserved for only the most important questions. This has led to efforts to find more efficient, yet dependable ways of assessing risk. One of these ways, case control studies, will be discussed in Chapter 10.

COMPARING RISKS

The basic expression of risk is incidence, defined in Chapter 4 as the number of new cases of disease arising in a defined population during a given period of time. But usually we want to compare the incidence of disease in two or more cohorts, which have different exposures to some possible risk factor. To compare risks, several measures of the association between exposure and disease, called *measures of effect*, are commonly used. They represent different concepts of risk and are used for different purposes. Four measures of effect are discussed below, summarized in Table 5.4, and illustrated by an example in Table 5.5.

Attributable Risk

First, one might ask, "What is the additional risk (incidence) of disease following exposure, over and above that experienced by people who are not exposed?" The answer is expressed as *attributable risk*, the incidence of disease in exposed persons minus the incidence in nonexposed persons. Attributable risk is the additional incidence of disease related to exposure, taking into account the background incidence of disease, presumably from other causes. Note that this way of comparing rates implies that the risk factor is a cause and not just a marker. Because of the way it is calculated, attributable risk is also called *risk difference*.

Table 5.4
Measures of Effect

EXPRESSION	QUESTION	DEFINITION[a]
Attributable risk (risk difference)	What is the incidence of disease attributable to exposure?	$AR = I_E - I_{\bar{E}}$
Relative risk (risk ratio)	How many times more likely are exposed persons to become diseased, relative to nonexposed?	$RR = \frac{I_E}{I_{\bar{E}}}$
Population attributable risk	What is the incidence of disease in a population, associated with the occurrence of a risk factor?	$AR_p = AR \times P$
Population attributable fraction	What fraction of disease in a population is attributable to exposure to a risk factor?	$AF_P = \frac{AR_p}{I_T}$

[a] Where:
I_E = incidence in exposed persons
$I_{\bar{E}}$ = incidence in nonexposed persons
P = prevalence of exposure to a risk factor
I_T = total incidence of disease in a population

Table 5.5
Calculating Measures of Effect: Cigarette Smoking and Death from Lung Cancer[a]

Simple Risks	
Death rate from lung cancer in cigarette smokers	0.96/1000/year
Death rate from lung cancer in nonsmokers	0.07/1000/year
Prevalence of cigarette smoking	56%
Total death rate from lung cancer	0.56/1000/year

Compared Risks

Attributable risk = 0.96/1000/year − 0.07/1000/year
= 0.89/1000/year

$$\text{Relative risk} = \frac{0.96/1000/\text{year}}{0.07/1000/\text{year}} = 13.7$$

Population attributable risk = 0.89/1000/year × 0.56
= 0.50/1000/year

$$\text{Population attributable fraction} = \frac{0.50/1000/\text{year}}{0.56/1000/\text{year}} = 0.89$$

[a] Estimated data from Doll R, Hill AB: *Br Med J* 1:1399–1410, 1964.

Relative Risk

On the other hand, one might ask, "How many times more likely are exposed persons to get the disease relative to nonexposed persons?" To answer this question, we speak of *relative risk* or *risk ratio*, the ratio of incidence in exposed persons to incidence in nonexposed persons. Relative risk tells us nothing about the magnitude of absolute risk (incidence). Even for large relative risks, the absolute risk might be quite small if the disease is uncommon. It does tell us the strength of the association between exposure and disease and so is a useful measure of effect for studies of disease etiology.

Interpreting Estimates of Individual Risk

The clinical meaning attached to relative and attributable risk is often quite different, because the two expressions of risk stand for entirely different concepts. The appropriate expression of risk depends upon the question being asked.

Example—The Royal College of General Practitioners has been conducting a study of the health effects of oral contraceptives. During 1968 and 1969, over 23,000 women taking oral contraceptives and an equal number of women who had never taken the pill were entered into the study by 1400 physicians. These physicians subsequently reported oral contraceptive use, morbidity, and mortality twice a year. The use of oral contraceptives was updated regularly. After 10 years of follow-

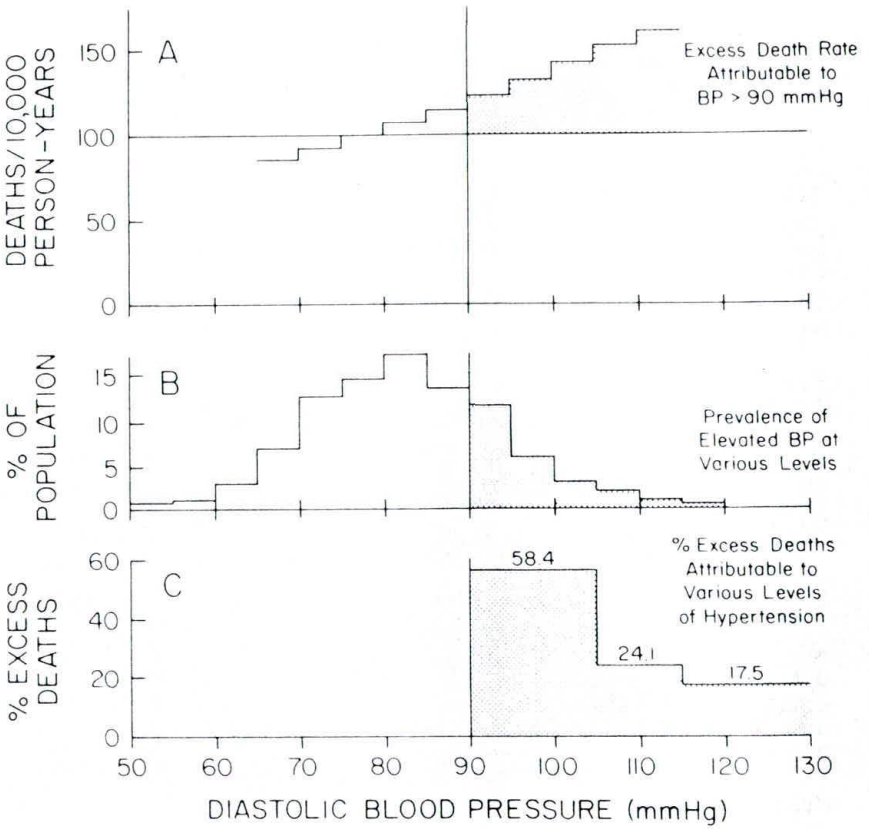

Figure 5.3. Relationships among attributable risk, prevalence of risk factor, and population risk for hypertension. (Adapted from The Hypertension Detection and Follow-up Cooperative Group. *Ann NY Acad Sci*, 304:254–266, 1978.)

up, it was reported that oral contraceptive users had a risk of dying from circulatory diseases that was 4.2 times greater than for nonusers. But the risk of dying was increased by only 22.7/100,000 women-years. An individual woman, weighing the risks of oral contraceptives, must deal with the two concepts of risk very differently. On the one hand, a four-fold greater risk of dying might loom large. On the other, two chances in 10,000 is a very remote possibility (6).

In general, because attributable risk represents the actual, additional probability of disease in those exposed, it is a more meaningful expression of risk in most clinical situations.

Population Risk

Another way of looking at risk is to ask, "How much does a risk factor contribute to the overall rates of disease in groups of people, rather than

individuals?" This information is useful for deciding which risk factors are particularly important and which are trivial to the overall health of a community, and so it can inform those in policy positions how to choose priorities for the deployment of health care resources.

To estimate population risk, it is necessary to take into account the frequency with which members of a community are exposed to a risk factor. A relatively weak risk factor (in terms of relative risk) that is quite prevalent could account for more of the overall incidence of disease in a community than a stronger risk factor that is rarely present.

Population attributable risk is a measure of the excess incidence of disease in a community that is associated with the occurrence of a risk factor. It is the product of the attributable risk and the prevalence of the risk factor in a population.

One can also describe the fraction of disease occurrence in a population that is associated with a particular risk factor, the *population attributable fraction*. It is obtained by dividing the population attributable risk by the total incidence of disease in the population.

Figure 5.3 illustrates how the prevalence of a risk factor determines the relationship between individual and population risk. *A* shows the attributable risk of death according to diastolic blood pressure. Risk increases with increasing blood pressure. However, few people have extremely high blood pressure (*B*). When hypertension is taken to be a diastolic blood pressure $\geq$ 90 mm Hg, most hypertensive people are just over 90 and very few are in the highest category, 115 mm Hg. As a result (*C*), the greatest percentage of excess deaths in the population (58.4%), is attributable to relatively low-grade hypertension, 90–105 mm Hg. Paradoxically, then, physicians could save more lives by effective treatment of mild hypertension than severe hypertension.

Measures of population risk are less frequently encountered in the clinical literature than are measures of individual risk, e.g., attributable and relative risk. But a particular practice is as much a population for its health care providers as is a community for health policy makers. Also, the concept of how the prevalence of exposure affects risk in groups can be important in the care of individual patients. For instance, when patients cannot give a history or when exposure is difficult for them to recognize, we depend on the usual prevalence of exposure to estimate the likelihood of various diseases. When considering treatable causes of cirrhosis in a North American patient, for example, it would be more profitable to consider alcohol than schistosomes, inasmuch as few North Americans are exposed to *Schistosoma mansoni*. Of course, one might take a very different stance in the Nile Delta, where people rarely drink alcohol and schistosomes are prevalent.

SUMMARY

Risk factors are characteristics that are associated with an increased risk of becoming diseased. Whether or not a particular risk factor is a cause of disease, its presence allows one to predict the probability that disease will occur.

Most suspected risk factors cannot be manipulated for the purposes of an experiment, so it is usually necessary to study risk by simply observing people's experience with risk factors and disease. One way of doing so is to select a cohort of people who are and are not exposed to a risk factor and observe their subsequent incidence of disease.

When disease rates are compared, the results can be expressed in several ways. Attributable risk is the excess incidence of disease related to exposure. Relative risk is the number of times more likely exposed people are to become diseased, relative to nonexposed. The impact of a risk factor on groups of people takes into account not only the risk related to exposure but the prevalence of exposure as well.

Although it is scientifically preferable to study risk by means of cohort studies, this approach is not always feasible because of the time, effort, and expense they entail.

REFERENCES

1. Weiss NS, Liff JM: Accounting for the multicausal nature of disease in the design and analysis of epidemiologic studies. *Am J Epidemiol* 117:14–18, 1983.
2. Rose G: Sick individuals and sick populations. *Int J Epidemiol* 14:32–38, 1985.
3. Diamond GA, Forrester JS: Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med* 300:1350–1358, 1979.
4. Kramer MS, Rooks Y, Pearson HA: Growth and development in children with sickle-cell trait. *N Engl J Med* 299:686–689, 1978.
5. Dawber TR: *The Framingham Study. The Epidemiology of Atherosclerotic Disease.* Cambridge, Harvard University Press, 1980.
6. Royal College of General Practitioners' Oral Contraception Study: Further analysis of mortality in oral contraceptive users. *Lancet* 1:541–546, 1981.

SUGGESTED READINGS

Dawber TR: *The Framingham Study. The Epidemiology of Atherosclerotic Disease.* Cambridge, Harvard University Press, 1980.
Morganstern H, Kleinbaum DG, Kupper LL: Measures of disease incidence used in epidemiologic research. *Int J Epidemiol* 9:97–104, 1980.
Relative or attributable risk? Editorial. *Lancet* 2:1211–1212, 1981.
Rose G: Sick individuals and sick populations. *Int J Epidemiol* 14:32–38, 1985.

5. Source:

Lilienfeld DE, Stolley PD. (1994). Foundations of epidemiology
Ch.10: Observational studies: I. Cohort studies.
Ch.11: Observational studies: II. Case-control studies.
New York: Oxford University Press.

10 OBSERVATIONAL STUDIES: I. COHORT STUDIES

In the experimental method, an investigator studies the effect of a change in the genetic composition or environment of a cell, an organ, or an organism and makes a comparison with a similar cell, organ, or organism that has not been subjected to that change. This ideal is the basis of both experimental and observational epidemiologic studies. In an experimental epidemiologic study, the investigator assigns the treatment; however, in the observational study, the investigator can only *observe* the outcomes associated with the individual exposures experienced by participants in the study. The investigator does not control the assignment of that exposure experience.

The data collected in an observational study can be tabulated in the form of a fourfold table, as shown in Table 10–1. If two similar groups can be identified that differ only by being exposed to a given environmental factor, e.g., oral contraceptives, or by possessing a particular characteristic, e.g., a specific blood group, the epidemiologist can follow these two groups and observe the incidence of disease in each. This type of investigation is known as a "cohort" study and is the subject of this chapter. In many situations, however, it is impractical for the epidemiologist to identify groups of individuals based upon their exposure histories or characteristics. One can more readily identify those individuals who have ("cases") or do not have ("controls") the disease of interest; the individuals' histories of past exposure to the factor or characteristic of interest can then be obtained and compared. This type of investigation is known as a "case-control" study, and will be discussed in Chapter 11.

The general concept of a cohort study is relatively simple, although such studies can be conducted in several ways. A sample of the population is selected and information is obtained to determine which persons either have a particular

Table 10–1. The Distinction between Cohort and Case-Control Studies

	ETIOLOGICAL CHARACTERISTIC OR EXPOSURE	CASE–CONTROL STUDY ↓ DISEASED GROUP (CASES)	NONDISEASED GROUP (CONTROLS)
COHORT STUDY →	Present (exposed)		
	Absent (not exposed)		

characteristic (such as a behavior or physiological trait) that is suspected of being related to the development of the disease being investigated, or have been exposed to a possible etiological agent. These individuals are then followed for a period of time to observe who develops and/or dies from that disease or physiological condition (such as decline in a pulmonary function test). The necessary data for assessing the development of the disease can be obtained either directly (by periodic examinations of everyone in the sample) or indirectly (by reviewing physician and hospital records, disease registration forms, and death certificates). Incidence or death rates for the disease are then calculated, and the rates are compared for those with the characteristic of interest and those without it. If the rates are different (either absolutely or relatively), an association can be said to exist between the characteristic and the disease. It is important to obtain information on other general characteristics of the study groups, such as age, gender, ethnicity, and occupation, in addition to the specific characteristic of interest, in order to account for the influence of any factors that are known to be related to the disease. Statistical methods are available for such analyses (Breslow and Day, 1987; Kelsey et al., 1986; Kahn and Sempos, 1989; Fleiss, 1981).

This type of study has been described by a variety of terms: "prospective," "incidence," "longitudinal," "forward-looking," and "follow-up," but "cohort study" will be used in this book. A distinction should be noted between cohort studies, described in this chapter, and cohort analyses, discussed in Chapter 5. In cohort studies individuals are followed or traced, whereas in cohort analyses there is no actual follow-up of persons; the follow-up is *artificially constructed* by the analysis of mortality (or morbidity) in successive age groups over a series of time periods (see p. 94).

MEASURING ASSOCIATION IN COHORT STUDIES

The data collected in a cohort study consist of information about the exposure status of the individual and whether, after that exposure occurred, the individual developed a given disease. These data may be tabulated into a 2 × 2 table (Table

RES-13

Table 10–2. Framework of a Cohort Study

ETIOLOGICAL CHARACTERISTIC OR EXPOSURE	DEVELOPED DISEASE	DID NOT DEVELOP DISEASE	TOTAL
Present (exposed)	a	b	a+b
Absent (not exposed)	c	d	c+d

10–2). The incidence rate among those persons exposed to the factor being investigated is a/(a + b), while the rate for those not so exposed is c/(c + d). The epidemiologist is interested, then, in determining whether the incidence rate for those exposed is greater than the rate for individuals not exposed, i.e., is a/(a + b) greater than c/(c + d)? If it is, then an association is said to exist between the factor and the subsequent development of disease. The question then asked by the epidemiologist is: How strong is the association?

Relative Risk

The **relative risk** (''RR'') is used to measure the strength of an association in an observational study (Cornfield, 1951):

$$\text{Relative Risk (RR)} = \frac{\text{Incidence rate of disease in exposed group}}{\text{Incidence rate of disease in unexposed group}}$$

The variance, confidence limits, and statistical tests for the relative risk may be found in the Appendix (p. 317). In a cohort study, if the incidence of myocardial infarction among cigarette smokers was 3 per 1,000 and that for nonsmokers was 1 per 1,000, then the relative risk of myocardial infarction for smokers compared to nonsmokers would be:

$$\text{RR} = \frac{(3/1{,}000)}{(1/1{,}000)} = 3.0$$

This value of the relative risk means that a cigarette smoker is three times as likely to develop a myocardial infarction as is a nonsmoker.

The magnitude of the relative risk reflects the strength of the association; i.e., the greater the relative risk, the stronger the association. A relative risk of 3.0 or more indicates a strong association; for cigarette smoking and lung cancer, for instance, it is greater than 10.0, signifying a very strong relation (United States Surgeon General, 1982). In contrast, the relative risk for a family history of breast

cancer (sister or mother) and female breast cancer is about 2.0, indicating a moderate association (Kelsey, 1979). A relative risk between 1.0 and 1.5 indicates a weak association.

Relative risks may also be less than 1.0 in value, suggesting a protective effect from exposure to a factor. For example, in a cohort study in Mali of meningococcal vaccine efficacy conducted during an epidemic of meningococcal meningitis, Binken and Bond (1982) found that the incidence rate of the disease among those vaccinated was 0.7 per 10,000 persons and among those not vaccinated 4.7 per 10,000 persons over the 5-week period following the vaccination campaign. Hence, the relative risk of meningitis for those vaccinated compared to those not vaccinated was 0.15, meaning that the risk of developing meningitis for someone who was vaccinated is only 15 percent of that for someone who was not vaccinated. This relative risk suggests a strong association between vaccination and protection from developing the disease.

Inferences about the association between a disease and exposure to a factor are considerably strengthened if information is available to support a gradient in the relationship between the degree of exposure (or ''dose'') to the factor and the disease. Relative risks can be calculated for each dose of the factor. The general approach is to treat the data as a series of 2 × 2 tables, comparing those exposed at various levels of the factor with those not exposed at all. An example of this type of analysis is the study by Vessey and his colleagues (1989) of the relationship between oral contraceptive use and ovarian cancer.

In the early 1970s, the possibility of a relation between oral contraceptive use and gynecologic cancer occurrence was suggested. During the period 1968-1974, 17,032 white married women, aged 25 to 39 years, were recruited at the Oxford Family Planning Association clinics in England and Scotland (Vessey et al., 1976). Of those enrolled, 6,838 were parous women who used oral contraceptives and 3,154 were parous women who used an intrauterine device (IUD). Some of these women were followed for up to 20 years (from 1968) and deaths were recorded by specific cause. The risk of mortality from ovarian cancer for different duration levels of oral contraceptive use are shown in Table 10–3 compared with those who had no exposure, i.e., women who used an IUD. The relative risks of death from ovarian cancer for oral contraceptive users relative to nonusers were:

RR (less than 48 months of oral contraceptive use) = 12.1/9.2 = 1.32

RR (48–95 months of oral contraceptive use) = 1.8/9.2 = 0.20

RR (more than 96 months of oral contraceptive use) = 1.5/9.2 = 0.16

Table 10–3. Mortality Rates per 100,000 Women-Years and Relative Risk of Ovarian Cancer by Duration of Use of Oral Contraceptives

TOTAL DURATION OF USE	OVARIAN CANCER MORTALITY RATE	RELATIVE RISK[a] OF OVARIAN CANCER
Never[b]	9.2	1.00
≤ 47 months	12.1	1.32
48–95 months	1.8	0.20
96+ months	1.5	0.16

[a]Compared to "Never" users

[b]Intra-uterine device users

Source: Vessey et al. (1989).

This pattern of declining relative risk of ovarian cancer with increased duration of oral contraceptive use suggests that these pharmaceuticals might protect against this disease. A statistical significance test to determine whether such relative risks are different from 1.0 was developed by Cochran (1954), and a method for calculating an overall (pooled) relative risk for all categories was developed by Mantel and Haenszel (1959) (see Appendix, p. 320). If several studies of the same epidemiologic problem have been carried out at different times and in different places, it may be useful to scrutinize the estimates and then determine whether they are similar (Breslow and Day, 1987; Greenland, 1987; Kahn and Sempos, 1989).

Attributable Fraction

A measure of association that is influenced by the frequency of a characteristic in a population is the **attributable fraction** (also known as the "attributable risk"). Levin (1953) originally defined it in terms of lung cancer and smoking as the "maximum proportion of lung cancer attributable to cigarette smoking." Attributable fraction can also be defined as the maximum proportion of a disease in a population that can be attributed to a characteristic or etiologic factor. Another way of using this concept is to think of it as the proportional decrease in the incidence of a disease if the entire population were no longer exposed to the suspected etiological agent. Although we are discussing attributable fraction in the context of cohort studies, this measure of association is also useful in the interpretation of case-control investigations (see Chapter 11).

As an example of the calculation of the attributable fraction, suppose that the incidence of lung cancer in the overall population is 120 cases per 100,000 persons; among nonsmokers in that population, it is 30 cases per 100,000 persons:

and among smokers, it is 330 cases per 100,000 persons. The relative risk of lung cancer among smokers compared to nonsmokers would then be 11.0 (330 per 100,000 / 30 per 100,000). Also assume that 30 percent of the population smokes. If the 30 percent of the population that smokes were to stop, then the incidence of lung cancer in that group would be reduced from 330 cases per 100,000 persons to 30 cases per 100,000 persons. The attributable fraction of lung cancer for cigarette smoking would then be:

$$\text{Attributable Fraction (AF)} = \frac{0.3\ (330 \text{ per } 100{,}000 - 30 \text{ per } 100{,}000)}{120 \text{ per } 100{,}000}$$

$$= \frac{0.3\ (300 \text{ per } 100{,}000)}{120 \text{ per } 100{,}000}$$

$$= \frac{90 \text{ per } 100{,}000}{120 \text{ per } 100{,}000}$$

$$= 75\%$$

An alternative way to calculate the attributable fraction is:

$$\text{Attributable Fraction (AF)} = \frac{P\ (RR - 1)}{P\ (RR - 1) + 1} \times 100\%$$

where RR = the relative risk and P = proportion of the total population that has the characteristic; the derivation of this formula can be found in the Appendix (p. 319). In the lung cancer example, P is 30 percent and RR is 11.0. The attributable fraction would therefore be:

$$AF = \frac{0.3\ (11.0 - 1)}{0.3\ (11.0 - 1) + 1} = \frac{3.0}{3.0 + 1} = \frac{3.0}{4.0} = 75\%$$

Standard error and confidence limits have been derived for the attributable fraction by Walter (1975, 1976) (see Appendix).

The effect of various values of the relative risk (RR) and various proportions of those with a characteristic in the population (P) on the values of the attributable fraction is shown in Table 10–4. When the frequency of a characteristic in a population is low (e.g., 10 percent) and the relative risk for that characteristic in a given disease is also low (e.g., 2), only a small proportion (9 percent) of the cases of disease can be attributed to that characteristic (Adams et al., 1989). However, with a high relative risk (e.g., 10) and a high proportion of the population having the characteristic (e.g., 90 percent), a much larger percentage (89

Table 10–4. Attributable Fractions* as a Proportion for Selected Values of Relative Risk and Population Proportion with the Characteristic

P = PROPORTION OF POPULATION WITH CHARACTERISTIC (%)	RR = RELATIVE RISK			
	2	4	10	12
10	.09	.23	.47	.52
30	.23	.47	.73	.77
50	.33	.60	.82	.84
70	.41	.67	.86	.89
90	.47	.73	.89	.91
95	.49	.74	.90	.9[illegible]

*Attributable fraction $= \frac{P\,(RR - 1)}{P\,(RR - 1) + 1}$

percent) of cases can be attributed to it. In these calculations, it is assumed that other etiological factors are equally distributed among those with and without the characteristic.

The measurement of attributable fraction is particularly useful in planning disease control programs (Walter, 1975, 1976; Stellman and Garfinkel, 1989). It enables health administrators to estimate the extent to which a particular disease is due to a specific factor and to predict the effectiveness of a control program in reducing the disease by eliminating exposure to the factor. For example, epidemiologic studies have suggested that throughout the world, the hepatitis B virus is the etiologic agent for 75 percent to 90 percent of primary hepatocellular cancer (Beasley, 1988). A global hepatitis B vaccination campaign could therefore greatly reduce the occurrence of this cancer.

Computations of attributable fraction are also helpful in developing strategies for epidemiologic research, particularly if there are multiple factors. In the United States, for example, it is estimated that in certain age groups, 80 to 85 percent of lung cancer can be attributed to cigarette smoking. Other etiological factors apparently play a relatively minor role, and the investigator interested in ascertaining these factors may decide to limit further studies to nonsmoking lung cancer patients. In general, if close to 100 percent of a disease is attributable to one or more factors, a search for additional etiological factors may not be profitable unless one is interested in studying other characteristics that influence those already exposed to a high-risk factor.

Exposure Assessment

A crucial aspect of the design of cohort studies concer he categorization of subjects into "exposed" and "unexposed" groups that can be compared with respect to disease incidence. If subjects cannot be correctly categorized, a cohort study is not feasible. An example of this inability to correctly classify exposure arose when epidemiologists at the Centers for Disease Control attempted to plan a cohort study of Vietnam veterans in regard to their exposure to Agent Orange, a defoliant that contained the toxic contaminant dioxin (Lilienfeld and Gallo, 1989; Centers for Disease Control Veterans Health Study, 1988). It was hoped that by learning about troop locations each day and comparing them to areas where the defoliant was sprayed the same day, an exposure score could be computed for each subject. However, when this score was compared with serum dioxin levels in a sample of such persons, it was clear that the exposure score would not be valid. Thus, the correct classification of exposure was problematic. The cancellation of the cohort study led to great protest by veterans' organizations who felt that their possible health risks were being ignored. However, conducting a cohort study with this high potential for misclassification might have led to results that underestimated the health risks of exposure to Agent Orange, if such risks actually exist.

Exposure assessment is important in all cohort studies, not only in those of occupational exposures. For example, the possible role of cardiovascular risk factors, such as hypertension and hypercholesterolemia, in pediatric atherosclerosis and adult cardiovascular disease is currently being studied in a cohort study of several thousand children in Bogalusa, Louisiana (Berenson and McMahon, 1980; Berenson, 1986). The exposure to these factors during childhood can be assessed directly, rather than trying to do so later in life.

TYPES OF COHORT STUDIES

Cohort studies can be classified as follows:

1. Concurrent studies
   (a) General population sample
   (b) Select groups of the population
      (i) Special groups—professional, veteran, etc.
      (ii) Exposed groups—occupational, etc.
2. Nonconcurrent studies
   (a) Population census taken in the past—usually special and unofficial
   (b) Select groups of the population
      (i) Special groups—professional, veteran, etc.
      (ii) Exposed groups—occupational, etc.

Concurrent and concurrent cohort studies are contrasted in Figure 10–1. In a **concurrent study,** those with and without the characteristic or exposure are

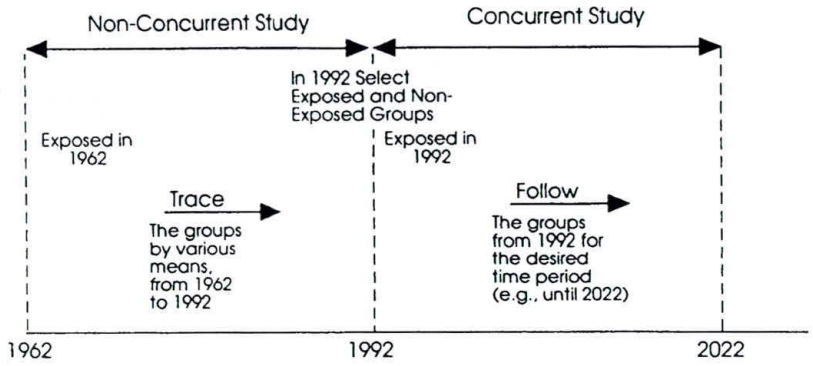

Figure 10–1. Diagrammatic representation of concurrent and nonconcurrent cohort studies.

selected at the start of the study (1992 in Figure 10–1) and *followed* over a number of years by a variety of methods. In a **nonconcurrent study,** the investigator goes back in time (to 1962 in Figure 10–1), selects his or her study groups, and *traces* them over time, usually to the present, by a variety of methods. These two types of cohort studies must be distinguished because they involve different methodological problems.

A simple example of a nonconcurrent cohort study would be an investigation of the safety of silicone breast implants. The epidemiologist might locate a group of plastic surgeons, each of whom used only one brand of silicone breast implant. The patient records of these surgeons would be reviewed for patients who had an implant placed two or three decades ago. Alternatively, if the epidemiologist identified a group of community hospitals in which silicone breast implant procedures were conducted, the medical records of the hospitals could be reviewed to provide information on the patients and the brand of implant used for each procedure. Regardless of the means by which the patients were identified, they would be followed up to the present time by contacting either the patient or the patient's family. For each brand of implant, the morbidity and mortality experience of the patient group would then be compared with that of the general population.

Concurrent Studies

In concurrent studies, the investigator begins with a group of individuals and follows them for a number of years. This was the approach used in the American Cancer Society's Cancer Prevention Study I (CPS I) of the health effects of cigarette smoking (Hammond, 1966; Garfinkel, 1985). The design of this study was similar to that of an earlier, smaller study (Hammond and Horn, 1958). For

this investigation, 68,116 volunteers were recruited between October 1, 1959 and February 15, 1960. Each volunteer was asked to enroll families in which at least one person was 45 years of age or older. All persons in each household were asked to complete forms detailing their smoking histories, family history, medical history, occupational history, and various health habits. Follow-up was conducted every year (through the volunteers), and every two years subjects were asked to complete a follow-up questionnaire. Death certificates were obtained for each reported death. About 1,045,000 completed forms were received from persons residing in 1,121 counties in 25 states. Through September 30, 1962, 97.4 percent of the participants were successfully traced; 971,362 were reported to be alive, 46,212 had died, and 27,513 could not be traced. Age- and cause-specific and age-standardized mortality rates by history of tobacco use were computed from the collected data. Since tobacco use differed so markedly between men and women, the data were analyzed separately by gender. Figures 10–2 and 10–3 illustrate some of the findings for men in this classic study.

Figure 10–2 shows an increasing risk of mortality from bronchogenic (or lung) cancer with increasing number of cigarettes smoked and lower mortality rates among ex-smokers than among current smokers. Figure 10–3 shows that the mortality rates among ex-smokers decrease as the period of time since they had stopped increases, except for those who had stopped smoking within a year of entry into the study. This exception may reflect the fact that some of the men gave up smoking because they had already been diagnosed as having lung cancer. Such findings (the outcomes associated with cessation of exposure) are important

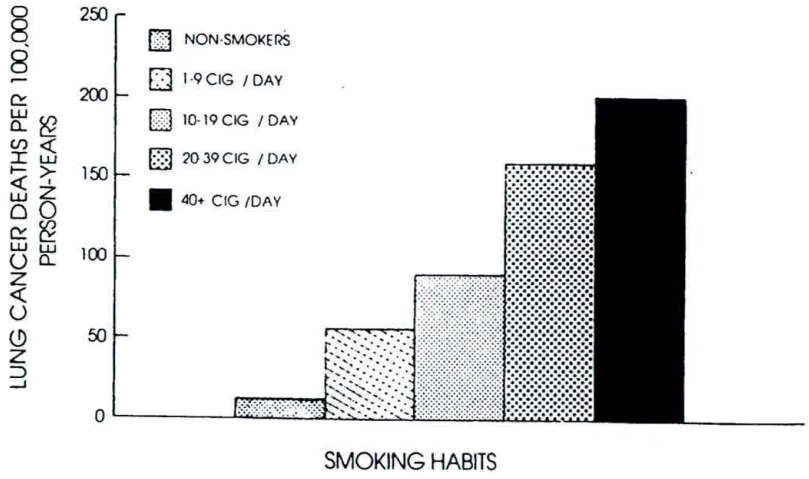

Figure 10–2. Age-adjusted death rates from malignant neoplasm of lung for men by amount of cigarette smoking at beginning of cohort study in 1959–1960. *Source:* Hammond (1966).

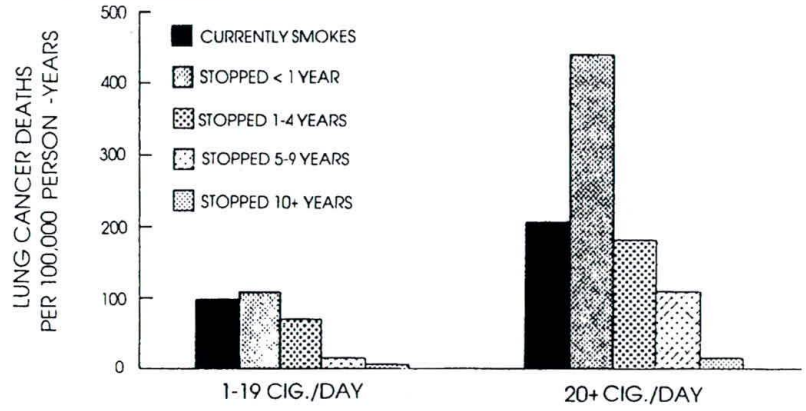

Figure 10–3. Age-adjusted death rates from malignant neoplasm of lung among men who had never smoked, who had stopped smoking, and who were still smoking at beginning of cohort study in 1959–60. *Source:* Hammond (1966).

in deriving etiological inferences from cohort studies (a subject that will be discussed in detail in Chapter 12). The groups in the CPS I study were not probability samples of the general population, which would have been preferable, but a probability sample of the required size would have been impossible to obtain. A similar study, known as the Cancer Prevention Study II, was started by the American Cancer Society in the late 1970s in order to examine more recent exposures of persons who may have been too young to participate in the CPS I study. Data collected in this ongoing investigation are now being analyzed.

A similar approach was used by Hirayama (1981a, b) in his pioneering study of passive smoking and lung cancer. He had collected information on the smoking habits of spouses of 91,540 nonsmoking wives and 20,289 nonsmoking husbands in six prefectures in Japan in 1965. The mortality of these men and women was assessed from death certificates during the 14 years of follow-up. Nonsmoking spouses of smokers had an elevated risk of lung cancer compared with that for nonsmoking couples (Figure 10–4). For nonsmoking men whose wives smoked 20 or more cigarettes daily, the risk was more than twice that of nonsmoking men married to nonsmoking women.

In some situations a cohort study can be conducted in a population selected from a well-defined geographical, political, or administrative area. This is particularly feasible when the disease or cause of death is fairly frequent in the population and does not require recruitment of a large number of persons for the study. The Framingham Heart Study is a good example of this type of cohort study (Dawber, 1980). It was initiated in 1948 by the United States Public Health Service to study the relationship of a variety of factors to the subsequent development

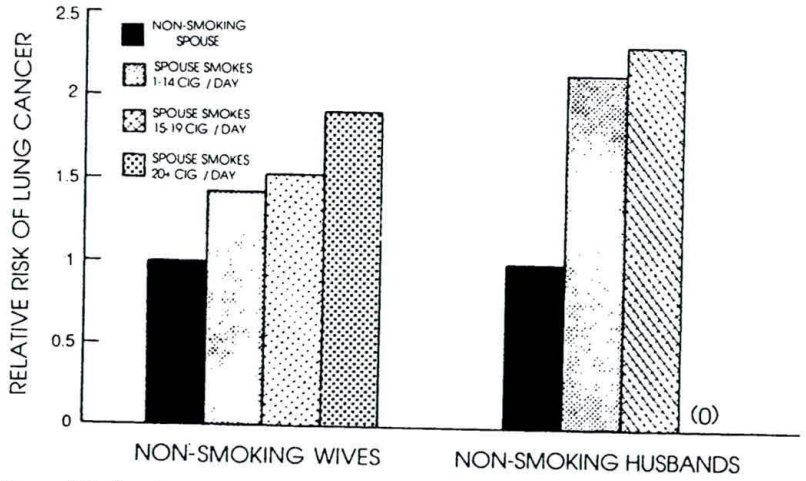

Figure 10–4. Age-adjusted relative risk of lung cancer among nonsmoking husbands and wives, by the smoking habits of their spouses. *Source:* Adapted from Hirayama (1981a).

of heart disease. The town of Framingham, Massachusetts, was chosen for its population stability, cooperation with previous community studies, presence of a local community hospital, and proximity to a large medical center. The initial population sample was a group of persons 30 to 62 years old that, when followed over a period of twenty years, would result in enough new cases or deaths from cardiovascular disease to ensure statistically reliable findings. The town's population in this age group was approximately 10,000. A sample of 6,507 men and women was selected. About 98 percent of the 4,469 respondents were free of coronary heart disease at the initial examination (Feinleib, 1985). Another 740 volunteers were also included in the cohort as part of a community outreach effort to ensure the continued participation in the study by each cohort member. After the first examination, each person was reexamined at two-year intervals for a thirty-year period. Information was obtained on several factors that could be related to heart disease, such as serum cholesterol level, blood pressure, weight, and history of cigarette smoking. Table 10–5 presents the incidence rates of coronary heart disease (CHD) among men and women during the first thirty years of follow-up by initial systolic blood pressure, gender, and age (Stokes et al., 1989). There is an increasing risk of CHD with increasing initial systolic blood pressure in the 35- to 64-year-old age group, a gradient of CHD disease which is slightly steeper in the older male age group and slightly less steep for the women.

The Framingham [illegible] Study also illustrates a strength of the cohort study:

Table 10–5. Average Annual Incidence per 1,000 Persons of Coronary Heart Diseae in Framingham, Massachusetts, 30-Year Follow-Up, by Systolic Blood Pressure and Gender

SYSTOLIC BLOOD PRESSURE (mmHg)	AGE 35–64		AGE 65–94	
	MEN	WOMEN	MEN	WOMEN
<120	7	3	11	10
120–139	11	4	19	13
140–159	16	7	27	16
160–179	23	9	34	35
>180	22	15	49	31

Source: Stokes et al. (1989).

investigating a variety of outcomes associated with a given exposure. For example, in addition to investigating the association between systolic blood pressure and CHD, for example, the Framingham investigators explored the relation between systolic blood pressure and stroke; a strong relationship between increased systolic blood pressure and elevated stroke risk was found.

The Framingham Heart Study became a prototype for similar studies in Tecumseh, Michigan, and other areas (Keys, 1970; McGee and Gordon, 1976; Napier et al., 1970). However, the difficulties in selecting general population samples for such studies tend to make investigators utilize a special group that for one reason or another can be followed more easily; certain professional groups, people enrolled in medical care programs, veterans, and others. In Doll and Hill's (1964) classic cohort study of cigarette smoking and lung cancer, for instance, a questionnaire was sent to all physicians on the British Medical Register who were living in the United Kingdom (see Chap. 1, p. 9). Follow-up was simplified because the subjects were physicians and therefore maintained contact with several professional organizations. Information from death certificates that listed "physician" as occupation was obtained from the Registrar General's Office. Lists were also obtained from the General Medical Council or the British Medical Association for deaths that had occurred abroad or in the military service.

A more recent example of the use of a unique population is the Oral Contraception Study of the Royal College of General Practitioners (1974) in England (Kay, 1984). Between May 1968 and July 1969, 23,000 oral contraceptive users and an equal number of nonusers, matched only for age and marital status, were recruited by physicians from among their patients. The oral contraceptive users selected were the first two women in each calendar month for whom the physicians wrote a prescription for an oral contraceptive. A nonuser was selected by the following procedure: starting with the user's record, returned to its correct place in the doctor's file, each subsequent record was examined in alphabetical order until the next record was found for a woman whose year of birth was within three years either side of that of the user and who had never used an oral contraceptive. Both the user and the nonuser had to be either married or known to be "living as married." These 46,000 women were followed with regard to their morbidity and mortality experience. In 1974, 1977, 1978, 1981, and 1988, progress reports were issued, showing associations between oral contraceptive use and (1) deep venous thrombosis, (2) acute myocardial infarction, and (3) subarachnoid hemorrhage (Table 10–6).

A similar approach was used by Hennekens and his colleagues (1979) in the Nurses' Health Study. These investigators sent questionnaires on possible risk factors (e.g., oral contraceptive use, smoking habits) to 121,700 nurses in 1976. Follow-up questionnaires were sent every two years thereafter to update risk factor information and to ascertain newly diagnosed conditions. Such data allow the epidemiologist to determine the effect of changes in risk factors on subsequent health events.

Concurrent cohort studies are not limited to noninfectious diseases. An example of the application of this method to infectious diseases is the study by Beasley et al. (1981, 1988) implicating the hepatitis B virus in the etiology of primary hepatocellular cancer. These investigators recruited 21,227 male Taiwanese government civil servants between November 1975 and June 1978, and 1,480 from a cohort study of risk factors for cardiovascular disease. Of these 22,707 men, 3,454 were hepatitis B surface antigen (HBsAg) positive, indicating past infection with the hepatitis B virus. By the end of 1986, 161 participants had developed primary hepatocellular cancer. The HBsAg positive group had a significantly higher rate of the disease than did the HbsAg negative group (Table 10–7). The relative risk of death from primary hepatocellular carcinoma among those who were HBsAg positive compared with that for those who were negative

Table 10–6. Age-Adjusted Relative Risks of Oral Contraceptive Users Compared to Nonusers

DISEASE (ICD–9 CATEGORY)	RELATIVE RISK (ORAL CONTRACEPTIVE USER TO NONUSER)
Nonrheumatic heart disease and hypertension (400–429)	5.6
Ischemic heart disease (410–414)	3.9
Subarachnoid hemorrhage (430)	4.0
Cerebrovascular accident (431–433)	2.1
Deep thrombosis of the leg, pulmonary embolism (450–453)	(*)

Source: Adapted from Layde et al. (1981).

*Rate for nonusers was 0.0; no relative risk could be calculated.

Table 10–7. Relation between HBsAg Antibody Status on Entrance to Study and Subsequent Development of Primary Hepatocellular Carcinoma through December 31, 1986

HBSAG STATUS	NUMBER	CASES OF PRIMARY HEPATOCELLULAR CARCINOMA	AVERAGE ANNUAL[1] INCIDENCE RATE OF PRIMARY HEPATOCELLULAR CARCINOMA PER 100,000 POPULATION
Positive	3,454	152	494.5
Negative	19,253	9	5.3

Relative risk of death from primary hepatocellular carcinoma among those who are HBsAg positive compared with those who are negative is (494.5/100,000)/(5.3/100,000) = 98.4.

[1]For 8.9 years of follow-up.

Source: Beasley (1988).

was 98.4, indicating a very strong association between HBsAg status and primary hepatocellular carcinoma.

The concurrent cohort study is particularly useful when the investigator does not know what the specific agent is when the study begins. In early 1984, for example, before the human immunodeficiency virus (HIV–1) had been identified as the etiologic agent for the acquired immunodeficiency syndrome (AIDS), the Multicenter AIDS Cohort Study (MACS) was begun to investigate the etiology and natural history of the disease (Kaslow et al., 1987). In Baltimore, Chicago, Los Angeles, and Pittsburgh, 4,955 homosexual men were recruited between April, 1984 and March, 1985. Each recruit provided blood, urine, feces, saliva, and semen specimens, which were stored for future analyses. The study population is reexamined every six months to determine if the participants have antibodies to the HIV–1 virus and, if so, what AIDS manifestations, if any, have developed. As hypotheses concerning the various manifestations of AIDS are developed, these specimen banks will be used to test those hypotheses.

In the concurrent cohort studies discussed so far, the study population was divided into those with and those without one or more possible etiological factors. The groups were sometimes classified according to different degrees of exposure or to levels of a characteristic such as the presence of the hepatitis B surface antigen. The incidence and mortality rates of these subgroups were then compared. The study groups were selected because they offered particular advantages for follow-up and information about a specific factor was obtainable from them. In a different type of concurrent study, a specific group that has been exposed to a possible etiological factor is selected and followed to determine the effects of this exposure as compared with the experience of a pop on not exposed to

that substance. This method has been especially useful in studies of the effects of exposure to substances in occupational environments. The elucidation of the relation between occupational exposure to asbestos and lung cancer provides an example of this strategy.

In 1955, Doll reported that the relative risk of lung cancer in a group of asbestos factory workers compared to the general population was 10. In 1963, Selikoff and his co-workers began a cohort study of 370 members of the International Association of Health and Frost Insulators and Asbestos Workers (IAHFIAW) (Selikoff et al., 1968). Follow-up of this cohort continued until 1967, when the investigation ended. The study findings suggested that there was an interaction between asbestos exposure and cigarette smoking in the development of respiratory cancer. These investigators initiated a study in 1967 of all U.S. and Canadian members of the IAHFIAW (Selikoff, 1979). The union provided the investigators with a membership list for 1966. Each member was mailed a questionnaire in which he was questioned about his smoking habits and the use of a mask while working. Some 17,800 men were followed from January 1, 1967 until December 31, 1976; 2,271 men died during the nine-year period. A control group, which had not been exposed to asbestos, was selected from the roster of 1,045,000 persons enrolled by the American Cancer Society in 1959 for the CPS I study described earlier. The control group, selected to be similar to the exposed group except for the exposure to asbestos, consisted of "men, not a farmer, no more than a high school education, a history of occupational exposure to dust, fumes, vapors, gases, chemicals, or radiation, and alive as of January 1, 1967." This group numbered 73,763 such persons. Follow-up of the nonexposed individuals was conducted in September, 1972. Official mortality statistics were used to extrapolate the observed mortality through 1976.

One of the major findings of this study is the positive interaction between both cigarette smoking and asbestos in markedly elevating the risk of lung cancer (Table 10–8). This type of relation is indicated by the fact that the death rate for

Table 10–8. Age-Adjusted Lung Cancer Death Rates per 100,000 Man-Years, by Cigarette Smoking Status and Occupational Exposure to Asbestos Dust

	NONSMOKERS	CIGARETTE SMOKERS
Not exposed to asbestos dust	11.3 (1.0)*	122.6 (10.9)
Exposed to asbestos dust	58.4 (5.2)	601.6 (53.2)

*Figure in parentheses is relative risk of lung cancer mortality compared with that for nonsmoking persons not exposed to asbestos dust.

Source: Hammond et al. (19

the combination of cigarette smoking and asbestos exposure was five times that of arette smokers without asbestos exposure and ten times that of nonsmoking persons with asbestos exposure. One might expect the relative risk for smoking workers to be about 15 if no positive interaction were present; however, it was 53, indicating such an interaction.

Nonconcurrent Studies

In nonconcurrent cohort studies, the period of observation starts from some date in the past, as illustrated in Figure 10–1; aside from the observation period, however, all other aspects of a nonconcurrent cohort study are the same as for a concurrent cohort investigation. These studies cannot be conducted with samples of the general population unless the investigator has access to a census of a community, usually unofficial, which was conducted in the past. Samples of the population covered by the census can then be selected and traced from the time of the census (Comstock, Abbey, and Lundin, 1970).

Nonconcurrent studies usually involve specially exposed groups or industrial populations because past census information is often unavailable and employment, medical, or other types of records usually are available. This is illustrated by the study of the relation between polycythemia vera (PV) and leukemia, which had been clinically observed since 1905 (Modan and Lilienfeld, 1965). The increased medical use of radiation treatment for PV and the observations of the leukemogenic effect of ionizing radiation in various studies raised the question as to whether the development of leukemia in patients with PV was part of the disease's natural history or a result of treatment with X-ray and/or P^{32}, a radioactive isotope. A study was undertaken to estimate the risk of developing leukemia among patients with PV and to determine whether it was increased as a result of P^{32} and/or X-ray treatment. Medical records of patients with PV who had been seen during 1947–1955 in seven medical centers were obtained at the same time as those of two comparison groups: (a) patients with polycythemia secondary to lung disease and (b) patients with questionable polycythemia. These groups were then classified by method of treatment into four categories: (1) no radiation treatment, (2) X-ray alone, (3) P^{32} only, and (4) a combination of X-ray and P^{32}. The patients were traced through December 31, 1961. Leukemia occurred predominantly in patients who had received some form of radiation, either X-ray, P^{32}, or a combination of the two (Figure 10–5). This finding has since been confirmed in a randomized clinical trial (Berk et al., 1981).

Nonconcurrent cohort studies of industrial exposures to possible etiological agents of disease can only be carried out by using company records of past and present employees that include information on the date that they begin their employment, age at hiring, the date of departure, and whether they are living or

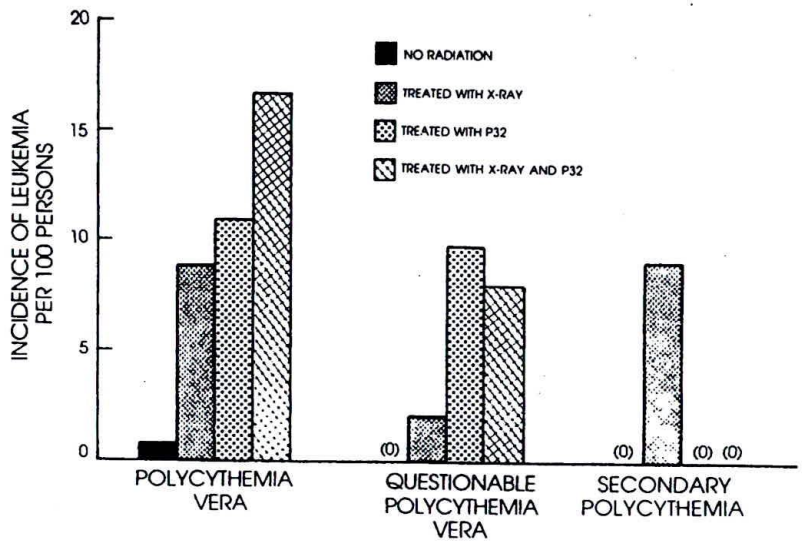

Figure 10–5. Incidence of leukemia among persons with polycythemia vera, questionable polycythemia vera, and secondary polycythemia vera. *Source:* Modan and Lilienfeld (1965).

dead. The mortality experience can be determined and compared with that of another industry, or with the mortality rate of the state where the industry is located, or of the country as a whole. This approach was used by Rinsky and his colleagues (1987) in a study of the relationship between exposure to benzene and leukemia mortality.

The study population consisted of all 1,165 nonsalaried white men employed in a rubber hydrochloride department of any of three plants in Ohio engaged in the manufacture of this natural rubber film for at least one day between January 1, 1940 and December 31, 1965. The cohort was assembled by using company personnel records. The cohort was traced through December 31, 1981, using vital status data from the Social Security Administration, the Ohio Bureau of Motor Vehicles, and a commercial tracing service. Death certificates were obtained for all deceased members. At the same time, an industrial hygienist used company records of benzene exposure to estimate the cumulative occupational exposure to benzene of each person in the cohort. At the time these exposure estimates were developed, the industrial hygienist did not know which of the cohort members had died from leukemia or from other causes.

The observed mortality from leukemia (nine deaths) was then compared with that expected if the cohort had had the same mortality experience as the United States population during the same time period. The results, shown in Figure 10–6, indicated a striking relationship between cumulative occupational exposure to benzene and leukemia mortality.

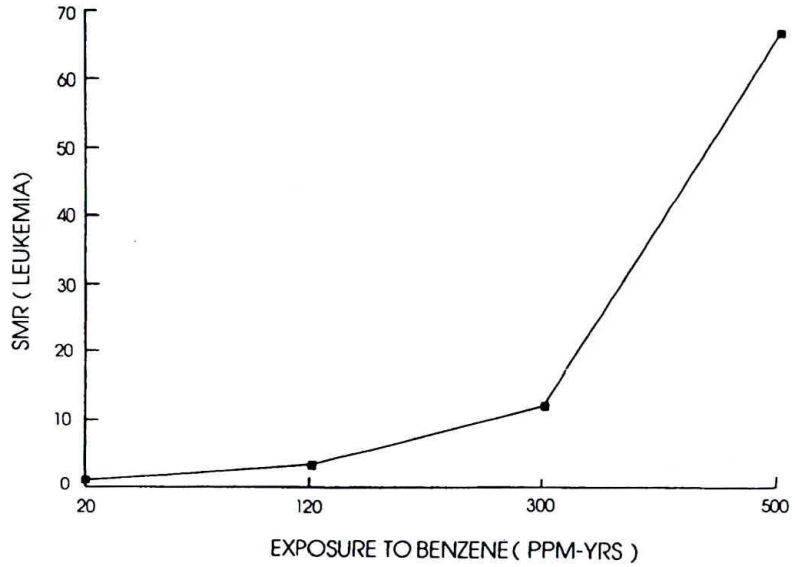

Figure 10–6. Standardized mortality ratio for 1,165 white men with at least one day of exposure to benzene from January 1, 1940 through December 31, 1965, according to cumulative exposure (parts of benzene per million particles × years of exposure). *Source:* Adapted from Rinsky et al. (1987).

STUDY PROCEDURES

A major source of difficulty in carrying out cohort studies is maintaining follow-up of the selected groups of persons. This is least troublesome in concurrent cohort studies for obvious reasons. At the very start of such studies, methods can be adopted for keeping in contact with the population on an annual basis, including periodic home visits, telephone calls, and mailed questionnaires, or even all three. The names and addresses of several friends and relatives can be obtained at the beginning of a study so that they may be contacted if the person moves out of the community. (Geographic mobility of people, particularly in the United States, can pose a problem.) To minimize the difficulties posed by tracing a cohort, cohort studies are often conducted in a health maintenance organization, in which the study population can be relatively easily followed. Another approach is to use a health or disability (for morbidity) or life (for mortality) insurer's clientele, as there is an economic incentive for the study population to inform the insurance company of the outcomes of interest. For deaths in the United States that have occurred since 1979, the National Death Index (administered by the National Center for Health Statistics) will inform investigators of the year and place of death for a given person (User's Manual, Nation eath Index, 1981).

In many countries, national or regional registries for cancer and other diseases can be used to follow up subjects in a cohort study.

Despite the best efforts, a certain number of individuals will likely be lost to follow-up. Even for this group, information on mortality status can often be obtained from state vital statistics bureaus. Their mortality experience can then be compared with that of the individuals not "lost to follow-up" to determine if there are any differences between the two groups. In addition, the successfully traced group can be compared to the "lost" group with respect to several known characteristics. To the extent that they show similar frequencies of a variety of characteristics of interest in the study, one's confidence is increased that no bias has been introduced into the findings by the lost group.

In a nonconcurrent cohort study, when one goes back perhaps twenty or thirty years to select a study group, the problem of tracing becomes more difficult. Every available source of information about subjects in the study should be used. Table 10–9 presents the various means used by Modan (1966) in determining the survivorship status of patients in his study of polycythemia vera and leukemia. In all cohort studies, it is desirable to trace as high a percentage of the study group as possible. Questions are frequently raised about the possibility of bias in the results if the degree of follow-up is less than 95 percent. This issue has been considered in several studies. Modan and Lilienfeld (1965) found that a very good estimate of the total mortality rate was obtained from the first 77 percent of the patients traced, although the group that was reached first had a somewhat higher leukemia mortality rate than those traced later. In a study of the outcome of neurosis, on the other hand, Sims (1973) found considerable differences

Table 10–9. Distribution of Sources of Information on Patient's Survivorship Status in the Study of Polycythemia Vera and Leukemia

SOURCE OF INFORMATION	NUMBER OF PATIENTS	PERCENT
Patient	158	12.9
Local physician	201	16.4
Relative	103	8.4
Hospital	540	44.2
Neighbors	49	4.0
Postmaster	18	1.5
Town-County clerk	20	1.6
Health department	89	7.3
Other	24	2.0
Untraced	20	1.6
Total	1,222	100.0

Source: Modan (1966

between the patients who were easily contacted and those who were traced with more effort. Only three deaths had occurred among the first 110 patients traced (59 percent of the study group), but eighteen additional deaths were discovered in the sixty-six patients (36 percent of the study group) who were found by more intensive tracing. Rimm and his colleagues (1990) have noted that even the type of mail service used during follow-up can affect response rates. Thus, it appears that the pattern varies in different studies and, perhaps, with different diseases, so that a general rule cannot be established about the degree of follow-up necessary to ensure unbiased conclusions. The safest course is to attempt to achieve as complete a follow-up as possible.

ANALYSIS OF RESULTS

General Strategy

It has already been made clear that the results of cohort studies are preferably analyzed in terms of relative risks, which provide a relatively simple expression of the relation between mortality rates from different diseases in the groups being compared. This is particularly true if the follow-up observations are made in the same period for all the study groups.

Many cohort studies, however, whether concurrent or nonconcurrent, involve lengthy and varying periods of observations. Persons are lost to follow-up or die at different times during the course of the study, and consequently they are under observation for different time periods. In some studies, persons are enlisted or enter the study at different times and, if the follow-up is terminated at a specific time, they will have been observed for different lengths of time. Two related methods are available for analyzing the results of such studies:

1. The calculation of person-years or months of observation as the denominator for the computation of incidence or mortality rate.
2. Actuarial, life table, or survivorship analysis (also known as cumulative incidence or mortality analysis).

Person-years of observation are often used as denominators in the computation of rates in cohort studies, as in the Royal College of General Practitioner's Oral Contraceptives Study. They are particularly useful when several factors, such as age, sex, and varying periods of observation (which result from persons entering and leaving the study at different ages and times), make the computation of an actuarial life table difficult or impossible. This analytic approach takes into consideration both the number of persons who were followed and the duration of

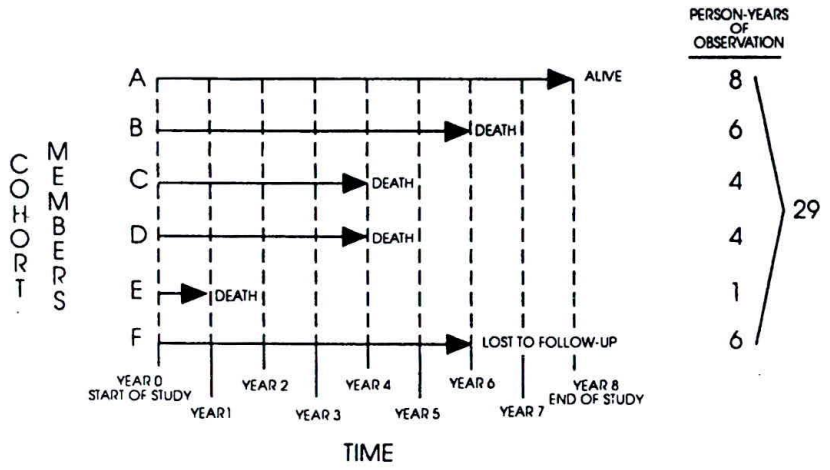

Figure 10–7. Diagrammatic illustration of contribution of person-years observed in a hypothetical eight-year cohort study of six persons (A,B,C,D,E, and F).

observation. In Figure 10–7, six persons are followed during an eight-year concurrent cohort study. Four of these persons (B, C, D, and E) die during the course of the follow-up. One person (A) is alive at the end of the study, and one person (F) is lost to follow-up after six years. The total number of person-years of observation (during which the cohort members were at risk of dying) is 29 years. The death rate in this study is therefore 4 deaths/29 person-years of observation (13.8 per 100 person-years of observation).

The use of person-years of observation makes it possible to express in one figure the time period when a varying number of persons is exposed to the risk of an event such as death or the development of disease. In addition, the age distribution of the groups under observation changes as a study progresses, as do the mortality and morbidity rates over time (Matanoski et al., 1975). The use of person-years is limited by the assumption that the risk of occurrence of an event per unit time is constant during the period of observation for the individual and that that risk is the same among similar persons in the cohort (Sheps, 1966; Breslow and Day, 1987). The overall effect of these limitations is modest and usually acceptable in most cohort studies.

Many regard using life tables (also known as survivorship methods) as the preferred method of analyzing data from cohort studies (see Appendix) (Chiang, 1961; Kahn and Sempos, 1989; Breslow and Day, 1987). They provide direct estimates of the probability of developing or dying from a disease for a given time period, and relative risks can be computed as the ratio of these probabilities. Life table methods can be used when the assumptions for person-years cannot be satisfied.

Latency

Regardless of the technique used to estimate the relative risk of developing a disease, one must also examine the possible effect of different latency or incubation periods. For instance, if a malignancy does not develop for at least a decade after the exposure to the suspected carcinogen began, then persons in the cohort would not be at risk for developing the disease until at least a decade had passed since their first exposure. Only after that decade had passed would those persons begin to accrue person-years of observation or be included in a life table (in the first interval); likewise, only if the disease developed after that first decade would that event be included in the analysis.

Adjustment for Age and Other Factors

The relative risks that are calculated by using either person-years or life table methods are unadjusted for age or other possible **confounding** factors. A confounding factor is one that is related both to the disease of interest and to another factor that is itself associated with the disease. For example, suppose that an epidemiologist conducted a cohort study of cigarette smoking and lung cancer. Many factors related to cigarette smoking (e.g., age, gender, and race) are independently associated with lung cancer. Hence, to measure the true relative risk between lung cancer and cigarette smoking, the epidemiologist would need to adjust the observed relative risk for these and other possible confounding factors. If adjustment for these factors does not change the relative risk, then little or no confounding is said to be present.

The epidemiologist may use two different approaches to adjust (or "control") for possible confounding factors:

1. Stratify the data by the possible confounding factors into multiple 2×2 tables to calculate the stratum-specific relative risk. An adjusted relative risk may then be calculated with Mantel-Haenszel techniques.
2. Use statistical techniques to mathematically model the risk of developing the disease, adjusted for the effects of the possible confounding factors. Examples include the logistic, the log-linear, and the proportional hazards models.

Where the entire study groups was exposed, however, it is necessary to use an external comparison or control group. If none is available, the mortality (or, if such data are available, the morbidity) experience of the exposed group is usually compared with that of the entire population living in the same geographical area as the exposed group, with statistical adjus nts for age, sex, and

calendar time of exposure and follow-up. For mortality, the number of deaths in the exposed group is compared with the expected number, based on the appropriate death rates for that geographical area. This comparison is then expressed as a Standardized Mortality Ratio (SMR) (see Chapter 4). This approach is frequently used in epidemiologic studies of occupational exposures (Monson, 1990). The previously described benzene-leukemia study by Rinsky et al. (1987) provides an example of this type of data analysis (see p. 215).

SUMMARY

In a cohort study, the investigator assembles a group of persons exposed to a possible etiologic factor and another, comparable group not exposed to that factor. These two groups are followed for the development of diseases. The investigator then calculates the incidence rate for a given condition in the exposed and unexposed groups, and a relative risk of developing the disease is calculated from those incidence rates. The stronger the association, the larger the relative risk; relative risks of 3.0 to 4.0 or more are usually indicative of strong associations between the factor and the disease. The proportion of disease in a population that is associated with that factor (assuming an etiologic relation) is the attributable fraction. The larger the attributable fraction of a disease for a given factor, the more difficult it becomes to study other possible agents of that disease.

There are two types of cohort studies: concurrent and nonconcurrent. In a concurrent study, the investigator assembles the exposed and nonexposed groups at the same time that the study is being conducted; these groups are then followed concurrently with the conduct of the study. In a nonconcurrent study the investigator reconstructs the groups in their entirety at some time in the past. This may be done with any set of records that provides information on all members of the population regarding their exposure at the same time in the past. Both groups are then followed to the present for the development of disease.

The process of following up the cohort of persons exposed and not exposed poses the greatest challenge to the epidemiologist in this study design. Inadequate follow-up can result in biased data and either spurious associations or missed relationships. It is also possible that the follow-up conducted in the early phases of a study may provide information on a portion of the cohort that is not reflective of the entire group. Analysis of the data at such a stage might result in different inferences than if one waited until both groups had been followed up completely.

Two methods are available for the analysis of cohort studies: (1) the calculation of incidence rates among those exposed and those not exposed using person-years of observati- , and (2) the calculation of life-tables to provide interval-specific incidence ra of disease among those exposed and those not exposed.

The use of person-years assumes that the risk of developing the disease is the same in each time period of follow-up and also that the risk of developing disease for each member of the cohort is the same. The incidence rates for those exposed and those not exposed are then compared by calculating the relative risk of disease, a measure of the strength of the association between the exposure and the disease. The magnitude of the relative risk may be affected by the presence of confounding factors, which may be related to the exposure, to the disease, or both. The effects of confounding may be adjusted for by stratification (calculating stratum-specific relative risks) or by constructing a statistical model of the data. If the entire cohort was exposed to the factor (e.g., an occupational study), an SMR-based analysis may be used to control for possible confounding factors, such as age and gender.

STUDY PROBLEMS

1. It has often been stated that the Standardized Mortality Ratio (SMR) and the relative risk are equivalent. Are they? Why might such a statement be made?
2. How useful is the attributable fraction to the epidemiologist?
3. A certain virus V is suspected of being the cause of infectious disease D. Design a cohort study to elucidate the relationship between V and D. How does the design change if V is a "slow virus" or if D is currently viewed as a noninfectious disease?
4. Internationally, several medical billing data bases are being developed by health maintenance organizations (HMOs) and national health care systems. How can these systems be used to conduct both concurrent and nonconcurrent cohort studies?
5. A few surgeons seek your advice (as the local epidemiologist) concerning a study they would like to conduct to determine the effect of tonsillectomy on subsequent mortality. What might you recommend?

REFERENCES

Adams, M. J., Khoury, M. J., and James, L. M. 1989. "The use of attributable fraction in the design and interpretation of epidemiologic studies." *J. Clin. Epid.* 42:659–662.

Beasley, R. P., Hwang, L.-Y., Lin, C.-C., and Chien, C.-S. 1981. "Hepatocellular carcinoma and hepatitis B virus." *Lancet* 2:1129–1133.

Beasley, R. P. 1988. "Hepatitis B virus. The major etiology of hepatocellular carcinoma." *Cancer* 61:1942–1956.

Berenson, G. S. 1986. "Bogalusa Heart Study." In *Causation of Cardiovascular Risk Factors in Children: Perspectives on Cardiovascular Risk in Early Life.* G. S. Berenson, ed. New York: Raven Press.

Berenson, G. S., and McMahon, G. A., eds. 1980. *Cardiovascular Risk Factors in Children: The Early Natural History of Atherosclerosis and Essential Hypertension.* New York: Oxford University Press.

Berk, P. D., Goldberg, J. D., Silverstein, M. N., Weinfeld, A., Donovan, P. B., Ellis, J. T., Landau, S. A., Laszlo, J., Njean, Y., Pisciotta, A. V., and Wasserman, L. R. 1981. "Increased incidence of acute leukemia in polycythemia vera associated with chlorambucil therapy." *New Engl. J. Med.* 304:441–447.

Binken, N., and Bond, J. 1982. "Epidemic of meningococcal meningitis in Bamako, Mali: epidemiological features and analysis of vaccine efficacy." *Lancet* 2:315–317.

Breslow, N. E., and Day, N. E. 1987. *Statistical Methods in Cancer Research, Vol. 2. The Design and Analysis of Cohort Studies.* Lyon: International Agency for Research on Cancer.

Centers for Disease Control Veterans Health Study. 1988. "Serum 2,3,7,8-tetrachlorodibenzo-p-dioxin levels in U.S. Army Vietnam-era veterans." *JAMA* 260:1249–1254.

Chiang, C. L. 1961. "A stochastic study of the life table and its applications. III. The follow-up study with the consideration of competing risks." *Biometrics* 17:57–78.

Cochran, W. G. 1954. "Some methods of strengthening the common χ^2 tests." *Biometrics* 10:417–451.

Comstock, G. W., Abbey, H., and Lundin, F. E., Jr. 1970. "The nonofficial census as a basic tool for epidemiologic observations in Washington County, Maryland." In *The Community as an Epidemiologic Laboratory: A Casebook of Community Studies.* I. I. Kessler and M. L. Levin, eds. Baltimore, Md.: The Johns Hopkins Press, pp. 73–99.

Cornfield, J. 1951. "A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast, and cervix." *J. Nat. Cancer Inst.* 11:1269–1275.

Dawber, T. R. 1980. *The Framingham Study: the Epidemiology of Atherosclerotic Disease.* Cambridge, Mass.: Harvard University Press.

Doll, R., and Hill, A. B. 1964. "Mortality in relation to smoking: Ten years' observation of British doctors." *BMJ* 1:1399–1410; 1460–1467.

Doll, R. 1955. "Mortality from lung cancer in asbestos workers." *Br. J. Ind. Med.* 12:81–86.

Feinleib, M. 1985. "The Framingham Study: sample selection, follow-up, and methods of analysis." In *Selection, Follow-up, and Analysis in Prospective Studies: A Workshop.* Garfinkel, L., Ochs, O., Mushinski, M., eds. NCI Monograph No. 67. Washington, D.C.: U.S. Government Printing Office.

Fleiss, J. 1981. *Statistical Methods for Rates and Proportions,* 2nd ed. New York: J. Wiley and Sons.

Garfinkel, L. 1985. "Selection, follow-up, and analysis in the American Cancer Society Prospective Studies." In *Selection, Follow-up, and Analysis in Prospective Studies: A Workshop.* Garfinkel, L., Ochs, O., Mushinski, M., eds. NCI Monograph No. 67. Washington, D.C.: U.S. Government Printing Office.

Greenland, S. 1987. "Quantitative methods in the review of epidemiologic literature." *Epi. Rev.* 9:1–30.

Hammond, E. C. 1966. "Smoking in relation to the death rates of one million men and

women." In *Epidemiological Studies of Cancer and Other Chronic Diseases*. NCI Monograph 19, pp. 127–204.

Hammond, E. C., and Horn, D. 1958. "Smoking and death rates—Report on forty-four months of follow-up of 187,783 men. Part I. Total mortality. Part II. Death rates by cause." *JAMA* 166:1159–1172; 1294–1308.

Hammond, E. C., Selikoff, I. J., Seidman, H. 1979. "Asbestos exposure, cigarette smoking, and death rates." *Ann. N.Y. Acad. Sci.* 330:473–490.

Hennekens, C. H., Speizer, F. E., Rosner, B., Bain, C. J., Belanger, C., and Peto, R. 1979. "Use of permanent hair dyes and cancer among registered nurses." *Lancet* 1:1390–1393.

Hirayama, T. 1981a. "Non-smoking wives of heavy smokers have a higher risk of lung cancer: a study from Japan." *BMJ* 282:183–185.

———. 1981b. "Non-smoking wives of heavy smokers have a higher risk of lung cancer" (letter). *BMJ* 283:1466.

Kahn, H. A., and Sempos, C. T. 1989. *Statistical Methods in Epidemiology*. New York: Oxford University Press.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., Rinaldo Jr., C. R. 1987. "The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants." *Am. J. Epidemiol.* 126:310–318.

Kay, C. R. 1984. "The Royal College of General Practitioners' Oral Contraception Study: some recent observations." *Clinic. Obs. Gyn.* 11:759–781.

Kelsey, J. L. 1979. "A review of the epidemiology of human breast cancer." *Epi. Rev.* 1: 74–109.

Kelsey, J. L., Thompson, W. D., and Evans, A. S. 1986. *Methods in Observational Epidemiology*. New York: Oxford University Press.

Keys, A., ed. 1970. *Coronary Heart Disease in Seven Countries*. Amer. Heart Assoc. Monog. No. 29. New York: The American Heart Association.

Layde, P. M., Beral, V., and Kay, C. R. 1981. "Further analyses of mortality in oral contraceptive users. Royal College of General Practitioners' Oral Contraception Study." *Lancet* 1:541–546.

Levin, M. L. 1953. "The occurrence of lung cancer in man." *Acta Unio In Contra Cancrum* 9:531–541.

Lilienfeld, D.E., and Gallo, M. 1989. "2,4–D, 2, 4, 5–T, and 2, 3, 7, 8–TCDD: an overview." *Epi. Rev.* 11:28–58.

Mantel, N., and Haenszel, W. 1959. "Statistical aspects of the analysis of data from retrospective studies of disease." *J. Nat. Cancer Inst.* 22:719–748.

Matanoski, G. M., Selter, R., Sartwell, P. E., Diamond, E. L., and Elliott, E. A. 1975. "The current mortality rates of radiologists and other physician specialists: deaths from all causes and from cancer." *Am. J. Epidemiol.* 101:188–198.

McGee, D., and Gordon, T. 1976. *The Framingham Study: The Results of the Framingham Study Applied to Four Other U.S.-Based Epidemiologic Studies of Cardiovascular Disease*. Washington, D.C.: U.S. Government Printing Office.

Modan, B. 1966. "Some methodological aspects of a retrospective follow-up study." *Am. J. Epidemiol.* 82:297–304.

Modan, B., and Lilienfeld, A. M. 1965. "Polycythemia vera and leukemia—the role of radiation treatment." *Medicine* 44:305–344.

Monson, R. 1990. *Occupational Epidemiology*, 2nd edition. Boca Raton, Florida: CRC Press.

Napier, J. A., Johnson, B. C., and Epstein, F. H. 1970. "The Tecumseh, Michigan Community Health Study." In *The Community as an Epidemiologic Laboratory: A Casebook of Community Studies*. I. I. Kessler and M. L. Levin, eds. Baltimore, Md: The Johns Hopkins University Press, pp. 25–46.

Rimm, E. B., Stampfer, M. J., Colditz, G. A., Giovannucci, E., and Willett, W. C. 1990. "Effectiveness of various mailing strategies among nonrespondents in a prospective cohort study." *Am. J. Epidemiol.* 131:1068–1071.

Rinsky, R. A., Smith, A. B., Filloon, T. G., Young, R. J., Okun, A. H., and Landrigan, P. J. 1987. "Benzene and leukemia: An epidemiologic risk assessment." *New Engl. J. Med.* 316:1044–1050.

Selikoff, I. J., Hammond, E. C., and Churg, J. 1968. "Asbestos exposure, smoking, and neoplasia." *JAMA* 204:106–112.

———, ———, and Seidman, H. 1979. "Mortality experience of insulation workers in the United States, 1943–1976." *Ann. N.Y. Acad. Sci.* 330:91–116.

Sheps, M. C. 1966. "On the person-years concept in epidemiology and demography." *Milbank Mem. Fund Q.* 44:69–91.

Sims, A.C.P. 1973. "Importance of a high tracing-rate in long-term medical follow-up studies." *Lancet* 2:433–435.

Stellman, S. D., and Garfinkel, L. 1989. "Proportions of cancer deaths attributable to cigarette smoking in women." *Women Health* 15(2):1–29.

Stokes III, J., Kannel, W. B., Wolf, P. A., D'Agostino, R. B., and Cupples, L. A. 1989. "Blood pressure as a risk factor for cardiovascular disease." *Hypertension* 13 (Supplement 1):113–118.

User's Manual: The National Death Index. 1981. U.S. Department of Health and Human Services Publication (PHS)81-1148. Hyattsville, Maryland: National Center for Health Statistics.

United States Surgeon General. 1982. *The Health Consequences of Smoking. Cancer*. Washington, D.C.: U.S. Government Printing Office.

Vessey, M. P., Doll, R., Peto, R., Johnson, B., and Wiggins, P. 1976. "A long-term follow-up study of women using different methods of contraception. An interim report." *J. Biosocial Sci.* 8:373–427.

Vessey, M. P., Villard-Mackintosh, L., McPherson, K., and Yeates, D. 1989. "Mortality among oral contraceptive users: 20-year follow-up of women in a cohort study." *Brit. Med. J.* 299:1487–1491.

Walter, S. D. 1975. "The distribution of Levin's measure of attributable risk." *Biometrika* 62:371–374.

———. 1976. "The estimation and interpretation of attributable risk in health research." *Biometrics* 32:829–849.

11

OBSERVATIONAL STUDIES: II. CASE-CONTROL STUDIES

In case-control studies, comparisons are made between a group of persons that have the disease under investigation and a group that do not. Usually those with the disease are called "cases" and those without the disease are called "controls." Indeed, case-control studies may be viewed as an extension of the case series that a health professional might assemble from his or her practice but with an important addition—the *control group allows for a comparison* to be made with regard to exposure history. Since the exposure history is assessed for some period in the past, case-control studies are also called "retrospective studies."

Whether the characteristic or factor of interest is (or was) present in the two groups is usually determined by interview, review of records, or biological assay. The proportion of cases exposed to the agent or possessing the characteristic (or factor) of etiological interest is compared to the corresponding proportion in the control group. If a higher frequency of individuals with the characteristic is found among the cases than among the controls, an association between the disease and the characteristic may be inferred.

When interested in determining whether prior exposure to an environmental factor is etiologically important, the epidemiologist will attempt to obtain a history of such exposure by interviewing the cases and controls. In practice, information on both current and past characteristics is usually obtained. One must constantly be aware that the derivation of inferences depends upon the temporal sequence between the characteristic and the disease.

The data for a case-control study are generally tabulated in the form of a four-fold table, as shown in Table 11–1. Such a table allows for the comparison

Table 11–1. Framework of a Case-Control Study

CHARACTERISTIC	NUMBER OF INDIVIDUALS WITH DISEASE (CASES)	NUMBER OF INDIVIDUALS WITHOUT DISEASE (CONTROLS)	TOTAL
With	a	b	a + b
Without	c	d	c + d
Total	a + c	b + d	a + b + c + d = N

of the prevalence of exposure among the cases, $a/(a + c)$, with that for the controls, $b/(b + d)$.

In a case-control study the odds ratio is an estimate of relative risk calculated as the cross-product of the entries in Table 11–1, ad/bc (see Appendix). Two assumptions are necessary in making this estimate: (a) the frequency of the disease in the population must be small, and (b) the study cases should be representative of the cases in the population and the controls representative of the noncases in the population. This cross-product estimate can be made with either actual numbers or percentages (Cornfield, 1951). The relative risk (or odds ratio) stays the same whatever the frequency of the exposure in a population. For example, whether smoking is highly prevalent or not, for a mother who smokes, the odds ratio describes her increased risk of delivering a low-birth-weight baby.

A study by Hurwitz and his colleagues (1987) of the relationship between the use of various medications and Reye's syndrome shows how the case-control approach can be used to investigate an etiological hypothesis and how the data can be analyzed with a four-fold table. Reye's syndrome is a rare, acute, and often fatal encephalopathy marked by brain swelling, low blood sugar, and fatty infiltration of the liver. Observations from case-series studies, case reports, and smaller case-control studies had implicated aspirin (salicylate) ingestion during viral illness as a possible cause of this disease of children. Cases deemed eligible for the Hurwitz study had to have received a diagnosis of Reye's syndrome from a physician, reported an antecedent respiratory or gastrointestinal illness or chicken pox within the three weeks before hospitalization, and experienced stage II or deeper encephalopathy. The control group consisted of children who did not have Reye's syndrome but did have chicken pox or a respiratory or gastrointestinal illness within a period of a few weeks before selection for the study. In this study there were four types of controls: emergency room patients (ER controls), inpatients (hospital controls), school children at the same school as the patient (school controls), and children located by the use of random-digit telephone dialing (community controls). The controls were matched to the cases on three patient characteristics: age, race, and the presence of an antecedent illness. The key data from the study are shown in Table 11–2. The percentage of salicylate users among

Table 11–2. Number of Hospitalized Reye's Patients and Number of Pooled Controls with a History of Salicylate Use in Three-Week Period

	CASES OF REYE'S SYNDROME	CONTROLS (POOLED)
Used salicylates	26	53
Did *not* use salicylates	1	87
Total	27	140

Source: Hurwitz et al., 1987.

the cases was 96.3% (26/27) as compared to 37.9% (53/140) among the controls, and the odds ratio (the estimate of relative risk) was calculated as follows:

$$\frac{ad}{bc} = \frac{26 \times 87}{53 \times 1} = \frac{2262}{53} = 42.7$$

The variance, standard error, confidence limits, and significance test for the odds ratio can be computed by the procedures presented in the Appendix.

Children with viral illnesses (chicken pox, upper respiratory, or gastrointestinal) who used salicylates during the illness were 42.7 times more likely to develop Reye's syndrome than were children with the same viral illness who did not use salicylates. Thus, aspirin use during viral illness appeared to be strongly associated with the development of Reye's syndrome, increasing the risk over forty-fold.

While these data provide an estimate of risk, they do not allow one to estimate the incidence of Reye's syndrome in the population of children at risk. To estimate the incidence one would need to know the number of all cases of Reye's syndrome among children (for the numerator) and the number of children who experienced respiratory or gastrointestinal illnesses or chicken pox (the denominator). Most case-control studies do not allow one to estimate incidence because denominator data are not available and numerator data may be incomplete.

THE SELECTION OF CASES AND CONTROLS

Various methods have been used to select cases and controls for case-control studies (Table 11–3) Sometimes investigators select cases from one source and controls from a variety of sources, permitting comparisons with different control groups as in the Reye's syndrome study (see Table 11–4). Consistency of results among studies using different types of control groups increases the validity of inferences that may be derived from the findings.

How many controls should be obtained for each case? Appropriate controls are often scarce or limited. In comparing workers at a factory who were or were not exposed to a substance, for instance, one would be limited to the finite set of workers who worked at the factory. In other situations, appropriate controls are readily available, as when studying normal birth outcomes compared to undesirable birth outcomes. Even when controls are abundant, it may be costly and time consuming to enroll and interview controls; one would want to include only as many as are needed. In studies of rare diseases the number of cases may be so small that the study has insufficient power to detect meaningful differences in exposure. An increased number of controls—up to four per case—may give the study more power (Gail et al., 1976). When the number of cases is large and the power is greater than 0.9 with only one control per case, additional controls cannot add very much to the power.

In selecting cases one may often use all cases occurring in a defined time

Table 11–3. Some Sources of Cases and Controls in Case-Control Studies*

CASES	CONTROLS
All cases diagnosed in the community (in hospitals, other medical facilities including physicians' offices)	Probability sample of general population in a community obtained by various methods including random-digit dialing
All cases diagnosed in a sample of the general population	Noncases in a sample of the general population or subgroup of a sample of general population (e.g., random-digit dialing)
All cases diagnosed in all hospitals in the community	Sample of patients in all hospitals in the community who do not have the diseases being studied
All cases diagnosed in a single hospital	Sample of patients in same hospital where cases were selected
All cases diagnosed in one or more hospitals	Sample of individuals who are residents in same block or neighborhood of cases
Cases selected by any of the above methods	Spouses, siblings, or associates (schoolmates or workmates) of cases Accident victims

*Various combinations of sources are possible.

Table 11–4. Comparison of Salicylate Exposure among Reye's Patients and Four Types of Controls

	CASES	CONTROLS			
		EMERGENCY ROOM	INPATIENT	SCHOOL	COMMUNITY
Exposed to aspirin (%)	96	40	27	44	34
Total N	27	30	22	45	43
Odds ratios	—	39	66	33	44

Source: Hurwitz et al., 1987.

period or geographic area. The researcher then has an idea about the age, race, and gender of the cases, as well as other characteristics. To ensure comparability of cases and controls one may **restrict** the controls to the same age range, race, and gender (or other characteristic) as the cases, or one may **group match** (also known as **frequency match**). For example, the cases can be stratified into different ten-year age groups. The control group can then be similarly stratified. Comparisons can then be made at each factor level between cases and controls with the usual statistical significance tests (Cochran, 1954; Mantel and Haenszel, 1959).

As an alternative to group matching, individual cases and controls can be **pair-matched** for various characteristics so that each case has a pairmate. Ideally, these pairmates should be chosen to be alike on all characteristics except for the particular one under investigation. In practice, if many characteristics are chosen for matching, or if many levels are chosen for each characteristic, it becomes difficult to find matching controls for each of the cases. In epidemiologic studies, there are usually a small number of cases and a large number of potential controls to select (or sample) from. Each case is then classified by characteristics that are not of primary interest, and a search is made for a control with the same set of characteristics. If the factors are not too numerous and there is a large reservoir of persons from which the controls can be chosen, case-control pair matching may be readily carried out. However, if several characteristics or levels are considered and there are not many more potential controls than cases, matching can be difficult. It is quite likely that for some cases, no control will be found; indeed, it may be necessary to either eliminate some of the characteristics from consideration or reduce the number of levels for some of them. With age matching, for example, it is often unlikely that pairs can be formed using one-year age intervals, but five- or ten-year age groups may make matching feasible.

The number of characteristics or levels for which matching is desirable and practical is actually rather small. It is usually sensible to match cases and controls only for characteristics such as age and gender whose association with the disease

under study is already known or has been observed in available mortality statistics, morbidity surveys, or other sources. In addition, when cases and controls are matched on any selected characteristic, the influence of that characteristic on the disease can no longer be studied. Hence, caution should be exercised in determining the number of variables selected for matching, even when feasible. If the effect of a characteristic is in doubt, the preferable strategy is not to match but to adjust for these characteristics in the statistical analysis.

POTENTIAL SOURCES OF BIAS

Selection Bias

A method commonly used in conducting case-control studies is to select the cases of the disease under study from one or more hospitals. The control groups usually consist of patients admitted to the same hospital, with diseases other than the one under study. This is a popular method for the initial studies that explore a suspected relation because the data can generally be obtained quickly, easily, and inexpensively. But several assumptions and sources of bias must be considered in analyzing the findings from such studies.

Selection bias is one of the major methodological problems encountered when hospital patients are used in case-control studies. W. A. Guy (see Chapter 2) was the first to suggest that a spurious association between diseases or between a characteristic and a disease could arise because of the different probabilities of admission to a hospital for those with the disease, without the disease, and with the characteristic of interest (Guy, 1856). This possibility was then demonstrated mathematically by Berkson (1946).

The influence of these differences on the study group in the hospital can be illustrated with a hypothetical example.

Let X = Etiological factor or characteristic

A = Disease group designated as cases

B = Disease group designated as controls

Assume that there is no real association between disease A and X in the group population, as indicated in Table 11–5; that is, the percentage of those with A who have X and the percentage of those with B who have X is equal. Assume also that there are different rates or probabilities of admission to the hospital for persons with X, A, and B, each of which acts independently, as follows: X = 50

Table 11–5. Frequency of Characteristic X in Disease Groups A and B in the General Population

CHARACTERISTIC	NUMBER OF INDIVIDUALS IN DISEASE GROUPS	
	A (CASES)	B (CONTROLS)
With X	200	200
Without X	800	800
Total	1,000	1,000
Percent of total with X	20	20

percent; A = 10 percent; B = 70 percent. Now consider the actual numbers of people in these groups who are admitted to the hospital:

(a) *For those with A and X:*
10 percent of the 200 in this category are admitted because they have A = 20
50 percent of the remaining 180 in this category are admitted because they have X = 90
Total admitted = 110

(b) *For those with A and without X:*
10 percent of the 800 in this category are admitted because they have A = 80

(c) *For those with B and X:*
70 percent of the 200 in this category are admitted because they have B = 140
50 percent of the remaining 60 in this category with B are admitted because they have X = 30
Total admitted = 170

(d) *For those with B and without X:*
70 percent of the 800 in this category are admitted because they have B = 560

These numbers are then inserted into the four cells of Table 11–5, allowing a comparison of disease A (cases) and disease B (controls) with respect to those who do and do not have the characteristic in our hypothetically constructed hospital population, as shown in Table 11–6. The result is that 58 percent of those with disease A have X as compared to 23 percent o se with disease B. This indicates that an association exists between A and X, even though this association

Table 11–6. A Hypothetical Hospital Population Based on Differential Rates of Hospital Admission

CHARACTERISTIC	NUMBER OF INDIVIDUALS IN DISEASE GROUPS	
	A (CASES)	B (CONTROLS)
With X	110	170
Without X	80	560
Total	190	730
Percent of total with X	58	23

is not present in the general population (the source of the hospital population). This spurious association results from the different rates of admission to the hospital for people with the different diseases and X. However, spurious associations such as this will not arise if either (Kraus, 1954):

1. X does not affect hospitalization, that is, no person is hospitalized simply because of X; or
2. the rate of admission to the hospital for those persons with A is equal to those with B.

One can never be absolutely certain that the first condition is met in any given study. For example, if X represents eye color, it might be assumed that this would not influence the probability of hospitalization. It is possible, however, that persons with a particular eye color belong to an ethnic group whose members are mainly of a specific social class, which, in turn, may influence the probability of their hospitalization. The likelihood of a spurious association is greater if the factor under investigation (i.e., X) is another disease rather than a characteristic or an attribute. The second condition is, of course, the exception rather than the rule since persons with different diseases usually have different probabilities of hospitalization. In any event, one cannot assume that these differences do not exist unless it is demonstrated that there are no differences in the hospitalization rates for individuals regardless of the disease.

In hospital studies, the same factors that may produce a spurious association, also termed "Berksonian" or "selection" bias, can have the reverse effect. The differences in hospital admission rates may conceal an association in a study and fail to detect one that actually exists in the population.

Selection bias is not limited to the analysis of hospital patients. It may be present in any situation or type of population where persons with different diseases or characte s enter a study group at different rates or probabilities. For example, in studying an autopsy series from a specified hospital population where

the autopsy rates differ for the diseases and characteristics being studied in the manner described above, the inferred associations will be biased and may result in a spurious association or mask a real association (McMahan, 1962; Mainland, 1953; Waife et al., 1952).

Selection biases, however, do not necessarily invalidate study findings. This issue should be resolved on its own merits for any particular investigation, and the following means are available to increase the likelihood that an observed association is real:

1. The strength of the association can be evaluated to see if it could result from the type of selection bias described above. A strong association is less likely to result from selection bias than a weak one.

2. Depending on the disease and the personal characteristic (such as serum cholesterol level) or the possible etiological factor (such as cigarette smoking), it may be possible to classify the characteristic or factor into a gradient from low to high levels. If the degree of association between the disease and the characteristic or factor consistently increases or decreases with increasing levels of the characteristic or factor, this "dose-response relationship" reduces the likelihood that the association is a result of selection bias. For selection bias to occur, it would be necessary to hypothesize the very unlikely occurrence of a similar gradient of rates of entry into the study group or of hospitalization in a study of hospitalized patients for the characteristic and the disease. This can be illustrated with some data from a recent study of oral contraceptive use and breast cancer among women 45 years old and younger in England (McPherson, et al., 1987). Information was obtained on past oral contraceptive use by women with breast cancer in six London hospitals and two Oxford hospitals during 1980–1984. The same information was obtained from a similarly aged control group (female

Table 11–7. Duration of Oral Contraceptive Use before First Term Pregnancy among Female Breast Cancer Patients and Hospital Controls 45 Years Old and Younger

DURATION OF ORAL CONTRACEPTIVE USE	CASES (%)	CONTROLS (%)
No Use	235 (67%)	273 (78%)
≤ 1 Year	27 (8%)	26 (7%)
1–4 Years	43 (12%)	29 (8%)
> 4 Years	46 (13%)	23 (7%)
Total	351 (100%)	351 (7%)

Source: McPherson et al., 1987.

patients in these hospitals admitted for conditions not related to contraceptive use) during this time period. Table 11–7 presents the results of a comparison of breast cancer patients and controls according to the duration of oral contraceptive use before the first pregnancy. Not only is there a higher proportion of oral contraceptive users among the breast cancer patients than the controls, but the breast cancer patients tended to use oral contraceptives for a longer time period than the control patients. A gradient showing an increase in oral contraceptive use among the cases compared with the controls is evident. Another illustration is provided by Antunes and his colleagues (1979), who examined the possible relationship between estrogen use and endometrial cancer with a case-control research design. Their findings are shown in Figure 11–1. A gradient of duration of postmenopausal estrogen use and endometrial cancer is evident.

3. As a precaution against the influence of selection biases, one may draw controls from a variety of sources. Should the frequency of the study characteristic be similar in each control group and differ from the case group, selection bias would not be a likely explanation for the observed association. The study of Reye's syndrome used controls from an emergency room, in patients, school children, and the community and found consistent results for each group (Table 11–4). In their classic study of lung cancer and smoking Doll and Hill (1952) demonstrated the importance of multiple control groups. They obtained infor-

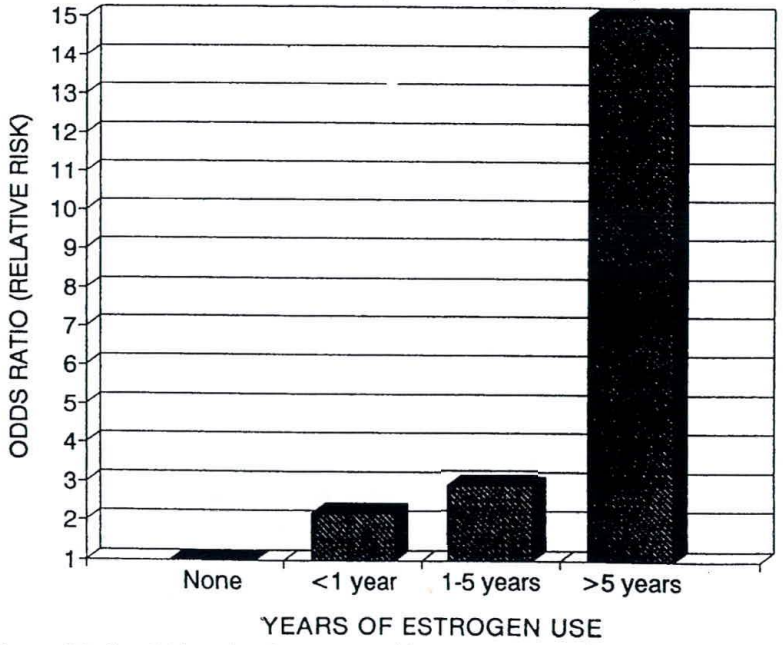

Figure 11–1. Odds ratios for endometrial cancer cases and controls according to duration of use of postmenopausal estrogen. *Source:* Adapted from Antunes, et al. (1979).

mation on the smoking habits of a sample of the general population from a social survey that was conducted in Great Britain during 1951. The smoking habits of patients in their control group were compared with those of persons in the social survey who were residents of Greater London, after adjusting for the age differences between the two groups. Table 11–8 shows the distribution of smoking habits among males in these two groups. The smaller proportion of nonsmokers and the higher proportion of heavy smokers among the controls than in the general population may result from the fact that patients in the control group had diseases that were also related to smoking habits. Thus, the degree of relationship between smoking and lung cancer shown in Table 11–8 was actually underestimated by the use of hospital controls in that investigation.

Representativeness

When cases are drawn from death certificate data bases or centralized registries, it is possible to select a representative sample of cases. This applies to case-control studies of various causes of death or of cancer or other registered illnesses. When cases are drawn from a limited, well-defined population it is also fairly easy to identify all cases. Thus, a case-control study of diarrhea in a day-care center can be designed to interview the parents of every child in the day-care center, or even to examine every child.

Many times it is not easy to identify all the cases of a disease. Even if one canvasses physicians, laboratories, and hospitals to find cases of an illness, there may be people with the illness who are not being treated or who are unaware of their condition. An example might be early miscarriage; a proportion of miscarriages may occur in women who are not aware that they are pregnant (and thus not aware that they miscarried), or women who have not yet been to a physician

Table 11–8. Comparison between Smoking Habits of Male Patients without Cancer of the Lung (Control Group) and of Those Interviewed in the Social Survey: London, 1951

SUBJECT	PERCENTAGE OF NONSMOKERS	PERCENT SMOKING DAILY AVERAGE OF CIGARETTES				NUMBER INTERVIEWED
		1–4	5–14	15–24	25+	
Patient with diseases other than lung cancer	7.0	4.2	43.3	32.1	13.4	1,390
General population sample (Social Survey)	12.1	7.0	44.2	28.1	8.5	199

Source: Doll and Hill (1952).

to begin prenatal care. There is probably no easy way to ensure obtaining a representative sample of women having early miscarriages. Cases of miscarriage drawn from a population of female physicians, for example, would probably select higher educated, higher social class women who are more likely to seek prenatal care earlier in pregnancy. In a study of life style and miscarriages this might introduce a bias, especially if the controls were selected from the general population.

Bias in Obtaining Information

Another bias that may distort the findings from case-control studies develops from the interviewer's awareness of the identity of cases and controls. This knowledge may influence the structure of the questions and the interviewer's manner, which in turn may influence the response. Whenever possible, interviews should be conducted without prior knowledge of the identify of cases and controls, although administrative constraints often prevent such "blind" interviews. In special circumstances, hospital patients may be interviewed at the time of admission so that information of epidemiologic interest is obtained before the patient is seen by a physician and thus before a diagnosis is made establishing the identity of cases and controls. This requires a comprehensive, general-purpose interview routinely administered to all patients admitted. Several epidemiologic studies have utilized a unique set of data from the Roswell Park Memorial Institute, where such a procedure is used (Bross, 1968; Bross and Tidings, 1973; Levin et al., 1950; Levin et al., 1955; Lilienfeld, 1956; Solomon et al., 1968; Winkelstein et al., 1958). Comparing their results with those of studies that depend on more conventional sources of controls provides a means for evaluating possible interviewer bias. A similar approach is used by the Slone Epidemiology Unit which routinely obtains drug histories from patients entering hospitals in the Boston region and other cities.

Patients interviewed as diagnosed cases in studies occasionally have had their diagnoses changed later. If data obtained from the erroneously diagnosed group resemble data from the control rather than the case series, interviewer bias can be discounted (Table 11–9).

The association of a factor and a disease may often be restricted to a specific histologic type or other component of the disease spectrum, as determined by objective means. For example, the fact that oat cell pulmonary carcinoma is more positively related to a history of exposure to bis-chloromethyl ether (BCME) than adenocarcinoma of the lung more firmly established the relationship between the two (Pasternak, et al., 1977). When such diagnostic details and their significance are unknown to the interviewer, another check on possible interviewer bias is provided.

The subjects sponses to an interview can also be directly validated by

Table 11–9. The Smoking Habits of Patients in Different Disease Groups, 45–74 Years of Age, Standardized According to the Age Distribution of the Population of England and Wales as of June 30, 1950

DISEASE GROUP	PERCENTAGE OF NONSMOKERS	PERCENT SMOKING DAILY AVERAGE OF CIGARETTES <5	5–14	15–24	25+	NUMBER INTERVIEWED
Males						
Cancer of lung	0.3	4.6	55.9	35.0	24.3	1,224
Patients incorrectly thought to have cancer of lung	5.3	9.9	35.5	37.8	11.4	102
Other respiratory diseases	1.9	9.9	38.3	38.7	11.2	301
Other cancers	4.6	9.4	47.2	26.0	12.8	473
Other diseases	5.6	9.0	44.8	26.9	13.7	875
Females						
Cancer of lung	40.6	13.7	22.0	9.5	14.2	90
Patients incorrectly thought to have cancer of lung	66.9	16.4	12.7	4.2	0.0	45
Other respiratory diseases	66.5	22.4	0.0	11.1	0.0	25
Other cancers	68.4	14.3	11.0	5.0	1.2	294
Other diseases	55.9	22.1	17.5	3.6	0.9	157

Source: Doll and Hill (1952).

comparison with other records. This was shown in a study of the accuracy of recall of the history of contraceptive use. Case-control studies of the relation between oral contraceptive use and a variety of diseases assumed that women recalled their use of oral contraceptives with reasonable accuracy (Collaborative Group for the Study of Stroke in Young Women, 1973; Mann et al., 1975; Thomas, 1972; Vessey and Doll, 1968). This assumption was tested by comparing oral contraceptive histories of seventy-five women attending family planning clinics with information available in the clinic records. It was found that the type of information obtained in the case-control studies was likely to be remembered with reasonable accuracy (Glass et al., 1974). This finding has been confirmed by Stolley et al. (1978).

Most investigators take great pains to prevent bias by rigorously training study interviewers in proper interview methods. Moreover, it is possible to check the interviewers' technique by video-taping the interview or reinterviewing a sample of the subjects to detect information bias at an early stage of a study when corrective measures are possible.

ANALYZING CASE-CONTROL STUDIES

We described the odds ratio in the beginning of this chapter. The comparison of exposure among cases and controls and the calculation of the odds ratio are the unique features in analyzing data from case-control studies. Odds ratios can be calculated for different amounts of exposure, or for subgroups stratified by other risk factors. Analysis of matched pairs is a special case when each pair is a separate strata. Multivariate methods can be used to estimate the effect of several variables on the odds ratio, and one can consider each variable while controlling for the others.

Odds Ratio for Multiple Levels of Exposure

Inferences about the association between a disease and a factor are considerably strengthened if information is available to support a gradient between the degree of exposure (or "dose") to a characteristic and the disease in question. Odds ratios can be computed for each dose of the characteristic. The general approach is to treat the data as a series of 2 × 2 tables, comparing controls and cases at different levels of exposure, and then calculating the risk at each level. The data from Table 11–7 are presented in Table 11–10, together with the computed odds ratios. The users with different durations of oral contraceptive use are compared with the nonusers, whose risk of breast cancer is set at 1.0. The odds ratios (OR) for users relative to nonusers are:

$$\text{OR } (\leq 1 \text{ year's use}) = \frac{27 \times 273}{26 \times 235} = \frac{7{,}371}{6{,}110} = 1.2$$

$$\text{OR } (1\text{–}4 \text{ years' use}) = \frac{43 \times 273}{29 \times 235} = \frac{11{,}739}{6{,}815} = 1.7$$

$$\text{OR } (>4 \text{ years' use}) = \frac{46 \times 273}{23 \times 235} = \frac{12{,}558}{5{,}405} = 2.3$$

It is possible to employ statistical tests of significance to determine whether or not the obtained relative risks differ from "unity" or 1.0. These tests can be applied to the summary relative risk (Cochran's test) or to all the categories (the Mantel-Haenzel test) (Cochran, 1954; Mantel and Haenszel, 1959) (see Appendix).

Table 11–10. Relative Risk of Breast Cancer for Smokers and Nonsmokers, by Duration of Oral Contraceptive Use (Data from Table 11–7)

DURATION OF ORAL CONTRACEPTIVE USE	BREAST CANCER CASES	HOSPITAL CONTROLS	ODDS RATIO (ESTIMATED RELATIVE RISK)
No use	235	273	1.0
≤ 1 Year	27	26	1.2
1–4 Years	43	29	1.7
> 4 Years	46	23	2.3

Source: McPherson et al., 1987.

Matched Cases and Controls

When cases and controls are matched in pairs in order to make the two groups comparable with regard to one or more factors, the fourfold (2 × 2) table takes a form different from that shown in Table 11–1. The status of the cases with regard to the presence or absence of the characteristic is compared with its presence or absence in their respective controls (Table 11–11). The cell in the upper left-hand corner of Table 11–11 contains *r* number of pairs in which both cases and controls possess the characteristic of interest. The marginal totals (a, b, c, d) represent the entries in the cells of Table 11–11 and the total for the entire table is ½*N* pairs where *N* represents the total number of paired individuals. The calculation of the odds ratio for this table is simple (Kraus, 1958): OR = s/t (provided t is not 0). Both a test of significance and a method of calculating the standard error are presented in the Appendix.

An example of the method of analysis for matched pairs in a case-control study comes from the work of Chow et al. (1990) on the relation between past exposure to *Chlamydia trachomatis* and ectopic pregnancy. Prior *Chlamydia trachomatis* infection had been associated with both tubal infertility and pelvic inflammatory disease, conditions associated with ectopic pregnancy. Chow and

Table 11–11. Symbolic Representation of Matched Cases and Controls with and without the Exposure of Interest

	CONTROLS		
	EXPOSED	UNEXPOSED	TOTAL
Exposed	r	s	a*
Unexposed	t	u	c*
Total	b*	d*	½ N

*a, b, c, and d correspond to the cells of Table 11–1.

Table 11–12. Matched Pair Analysis of a Case-Control Study of the Association between *Chlamydia trachomatis* and Ectopic Pregnancy

		CONTROLS	
		PAST EXPOSURE TO *C. TRACHOMATIS*	NO EXPOSURE TO *C. TRACHOMATIS*
Cases	Past exposure to *C. trachomatis*	72	109
	No exposure to *C. trachomatis*	36	40

Source: Chow, 1990, personal communication.

her colleagues recruited the cases of ectopic pregnancies from admissions and the controls from prenatal clinics. The case-control pairs were matched for age (±1 year), ethnicity, hospital, and restricted to women whose pregnancy was of 12 to 24 weeks duration. Cases with previous bilateral tubal ligation, ectopic pregnancy, or an intrauterine device present at the time of conception were excluded from the study. A total of 257 matched case-control pairs were assembled and each pair was categorized as to past exposure to *Chlamydia trachomatis* (assessed by antibody titer of ≥1:64). Based on Table 11–11, each pair could be categorized in one of four ways:

r. Case exposed and control exposed (++) = 72
s. Case exposed and control not exposed (+−) = 109
t. Case not exposed and control exposed (−+) = 36
u. Case not exposed and control not exposed (−−) = 40

Group *s* is the group where cases were exposed and controls were not (+−); group *t* is the group where cases were not exposed, but controls were exposed (−+). As in the above formula, the odds ratio is estimated as s/t or 109/36 = 3.0 (see Table 11–12). The calculation considers only the discordant pairs, and this can be explained intuitively: One can see that pairs where both were exposed or where both were unexposed would give no information about the relationship of exposure to disease. For example, one could not measure the effect of fluoride on cavities in a group of pairs that had all received fluoride, or that had all been unexposed to fluoride (Schlesselman, 1982).

Interrelationships between Risk Factors

Odds ratios can also be used to determine whether interrelationships exist between various characteristic · risk factors. A case-control study of lung cancer, cigarette smoking, and asbestos exposure among workers in southern Norway exposed

to multiple risk factors provides an example of this (Kjuus et al. 1986). In two neighboring counties in the southern part of Norway, all cases of lung cancer in males during 1979–1983 were ascertained. For each case, a similarly aged control was selected from among the patients in the same geographical area as the case. All men with conditions that would have precluded possible employment in heavy industry were excluded from the study. The 176 cases and 176 controls were interviewed about their history of exposure to asbestos and their smoking habits. The histories were then coded into four categories according to the level of asbestos exposure the person had reported (no exposure, light or sporadic exposure, moderate exposure between 1 and 10 years duration or heavy exposure less than 1 year in duration, and more than 10 years of moderate exposure or more than 1 year of heavy exposure). The relative risks for each category of asbestos exposure and smoking habit are shown in Table 11–13. From these data, it appears that the relative risk increases with an increase in either smoking or asbestos exposure. When the factors are considered together, the odds ratio rises sharply. This suggests that these factors modify and increase each other's effect on the disease.

Effect of Misclassification

Misclassification of both disease and exposure can occur in any type of study. In a case-control study, misclassification of disease would lead to some of the selection biases already discussed; it would alter a person's probability of entering the study. Assuming that selection bias has been dealt with, misclassification of exposure must be addressed in a case-control study. Exposure status usually cannot be measured directly by the researcher in such a study. Instead, the researcher relies on records (e.g., employment records describing work assignments and possible occupational exposures), recall (e.g. employment, residential, smoking, pharmaceutical histories), or even the recall of a close friend or relative, usually a spouse (e.g. diet, smoking, alcohol consumption, exercise). There are two types of misclassification that can occur: (1) differential—where the amount or direction of misclassification is different in the cases and controls, and (2) nondifferential—where the amount and direction of misclassification is the same in cases and controls. Misclassification error can occur in one direction for cases and controls; for example, everyone may underreport their own or their spouse's habitual alcohol consumption. Misclassification can occur in opposite directions; spouses of cirrhosis patients might overreport alcohol consumption, while spouses of other patients might continue to underreport alcohol use. People typically may misreport their abortion histories, smoking histories, number of sexual partners, and income, and this may be all in one direction or not. People may also misreport information because they can't remember their typical breakfast 10 years ago, the number of cigarettes their husbands used to smoke, the length of their menstrual

Table 11–13. Odds Ratio Estimates of the Relative Risks of Lung Cancer for Combined Exposure to Asbestos and Smoking

CIGARETTES SMOKED DAILY	ASBESTOS EXPOSURE			
	NONE	LITTLE	MODERATE	HEAVY
0–4	1.0	1.2	2.7	4.1
5–9	2.9	1.2	7.8	11.9
10–19	9.1	1.9	24.6	37.3
20–29	16.5	19.8	44.6	67.7
≥30	90.3	108.4	243.8	370.2

Source: adapted from Kjuus et al. (1986).

cycle during each decade of life, or how many hours a day they were exposed to silica dust during each year of employment.

Differential misclassification (because the exposure status of cases is more or less likely to be miscategorized than that of the controls) can produce bias in either direction, raising or lowering the estimate of risk (Schlesselman, 1982). Nondifferential misclassification (randomly distributed among cases and controls) generally shifts the odds ratio toward the null hypothesis (OR = 1.0), but exceptions to this can occur (Dosemeci et al., 1990). The effect of misclassification may also depend on how exposure is defined, as a continuous or categorical variable, and if categorical, as a two-level or multilevel variable.

These effects of misclassification emphasize the need to verify the information obtained in a study by every feasible means. Information with respect to previous exposures or characteristics of study individuals may be verified by obtaining records from independent sources (such as hospitals, physicians, schools, military services, and employment records) on either all or a sample of individuals in the study. Disease diagnoses should be verified whenever possible by independent review of medical records, histological slides, electrocardiograms, etc. The degree of verification possible depends upon the factors or characteristics and the diseases being studied. For example, verification of alcohol consumption or of the content of an individual's diet over a period of time poses serious problems of verification. Alternatively, in a health maintenance organization, for instance, records of prior illness or drug prescriptions may be available, eliminating the possibility of misclassification. Another approach is to use antibody titers as an index of past exposure to an infectious agent. This method has been used in case-control studies of hepatitis B infection and primary liver cancer (Szmuness, 1978). Recently, biological markers for some other exposures have been developed. For example, the presence of cotinine, a metabolite of nicotine, in the blood, urine, or saliva can serve as a biomarker of exposure to cigarette smoking; a high level would indicate active smoking, and a low level, exposure to environmental tobacco smoke.

Attributable Fraction

Another measure of association, influenced by the frequency of a characteristic in the population, is the attributable fraction. As noted in Chapter 10, this is the proportion of a disease that can be attributed to an etiological factor; alternatively, it is considered the proportional decrease in the incidence of a disease if the entire population were no longer exposed to the suspected etiological agent. As in cohort studies, the attributable fraction may be estimated in case-control studies as follows:

$$\text{Attributable Fraction (AF)} = \frac{P(OR - 1)}{P(OR - 1) + 1} \times 100\%,$$

where OR = the odds ratio and P = proportion of the total population classified as having the characteristic. The derivation of this formula can be found in the Appendix. Standard errors and confidence limits have been derived for the attributable fraction by Walter (1975, 1978) (see Appendix).

Computations of attributable fraction are also helpful in developing strategies for epidemiologic research, particularly if there are multiple etiological factors (Walter, 1975). In the study of past *Chlamydia trachomatis* infection and ectopic pregnancy, for example, the attributable fraction for past chlamydial infection was 47 percent, while that for douching (an independent risk factor) was 45 percent (Chow et al., 1990). These data suggest the need for further investigation of douching practices in relation to ectopic pregnancy occurrence, while underscoring the need for control of chlamydial infections to prevent ectopic pregnancies.

Regression Models and Adjustment for Confounding Variables

In a case-control study, several variables may be studied as potential risk factors, variables thought to influence the outcome (occurrence of disease). As will be discussed in Chapter 12, it is always possible that these variables may be **confounded** with one another. For example, in a case-control study of lung cancer, exposures of interest may include cigarette smoking, exposure to asbestos, and use of alcohol. Which of these exposures are associated with lung cancer and which are not (but are associated with one another)? The epidemiologist can deal with this problem by using **multivariate analysis,** a set of techniques for studying the effects of several factors simultaneously (Kleinbaum et al., 1982). These techniques range from simple cross-classification and adjustment to more complex methods of statistical regression analysis.

Various models have been used by epidemiol ts, such as "multiple logistic," "log-linear," "multiple linear," and "simple linear" regression. These techniques permit the investigator to determine which of the variables has an independent association with the outcome, to determine which variables interact among themselves, and to quantify the relative contribution of each variable or combination of variables to the risk of the disease. Multivariate analysis does not necessarily distinguish causal from noncausal associations, but it may give indications about the relative strengths of the independent and joint effects of multiple exposures.

ADVANTAGES AND DISADVANTAGES OF CASE-CONTROL STUDIES

Advantages

The case-control study can be used to test hypotheses concerned with the long-term effects of an exposure on a disease, and the study can often be completed quickly. For example, in one to two years data can be collected about 20 or 30 years of exposure to an environmental or occupational hazard.

The case-control study can also be used to test hypotheses about rare diseases or diseases that have long latency periods. The first case-control study estimating the association between diethylstilbestrol (DES) and adenocarcinoma of the vagina in young women used only 8 cases and 32 controls (Herbst et al., 1971). The disease was very rare (about 10 cases in 10 million young women) and 15 to 20 years elapsed between exposure and disease, but the case-control study identified the risk factor and estimated the relative risk. In Table 11–14 one may see how the rareness of disease influences the number of subjects needed in cohort or case-control studies and the advantage of a case-control study for studying rare conditions.

The case-control study is well suited to the study of adverse effects of a drug or treatment, or of a new disease where efficient identification of a risk factor can lead to prompt public health intervention.

The case-control study can be relatively inexpensive because it may use fewer study subjects and take a shorter period of time than some other designs. It also allows examination of several risk factors for a single disease.

Disadvantages

It is sometimes difficult to find an appropriate control group, for theoretical or practical reasons. For example, what is the appropriate control group for auto accident victims? What is the appropriate control group for tennis players with a particular injury? W"" there be enough subjects available for a control group?

It is sometime.. difficult to decide if the exposure preceded the disease. In

Table 11–14. Sample Size Requirements for Cohort and Case-Control Studies*

DISEASE INCIDENCE IN UNEXPOSED GROUP	FREQUENCY OF ATTRIBUTE DETECTABLE IN POPULATION (%)	RELATIVE RISK	SAMPLE SIZE NEEDED IN EACH GROUP	
			COHORT STUDY	CASE-CONTROL STUDY
1/1,000	50	1.2	576,732	2,535
		2.0	31,443	177
		4.0	5,815	48
1/100	50	1.2	57,100	2,535
		2.0	3,100	177
		4.0	567	48
1/10	50	1.2	5,137	2,535
		2.0	266	177
		4.0	42	48

*Power = 90%; alpha = 5%.

Source: Kahn and Sempos (1989).

studying diarrhea among breast-fed or formula-fed babies, one would want to know if diarrhea led to cessation of breast feeding, or if cessation of breast feeding led to an episode of diarrhea. Similarly, in a study of heart disease among letter carriers, one would like to know whether healthy people choose to become letter carriers or whether letter carrying (and walking each day) leads to healthier cardiovascular systems.

Case-control studies are subject to a number of biases, especially survival biases, selection biases, recall biases, and misclassification. Well-designed studies can sometimes minimize the introduction of biases, but the potential for biases must be considered for each study question. Case-control studies frequently rely on information collected from living cases of the disease of interest. If the deceased cases are different from the surviving cases, a bias may be introduced into the study.

Case-control studies do not actually measure incidence of disease in the population at risk, although estimates can sometimes be made (when all cases of the disease are known, and the population at risk is known).

SUMMARY

In a case-control study, the investigator compares the history of past exposure to a factor or presence of a characteristic among those persons with a given disease or condition (cases) and among those who do not have the disease or condition

(controls). The proportion of those exposed among the cases is compared with that among the controls. If these proportions are different, then an association exists between the factor and the disease. Cases can be ascertained from hospitals, clinics, disease registries, or during a prevalence or incidence survey in a population. Controls can likewise be sampled from hospitals, clinics, or a random sample of the population. Care must be exercised in the case and control selection methods, because selection biases can lead to spurious associations. An alternative approach to control selection is to match each control to each case, based on factors thought to be related to the exposure of interest and the disease. In the process of matching, the investigator loses representativeness, i.e., the ability to generalize the findings to the general population, but gains greater comparability among the cases and controls. Unbiased collection of data from both cases and controls is necessary. Biases can occur in recalling past exposures.

The measure of the strength of an association in a case-control study is the odds ratio estimate of the relative risk of developing the disease for those who have been exposed compared with that for those not exposed. Odds ratios can be calculated for both matched and unmatched designs. Misclassification of either the presence or absence of disease, or of exposure status, can affect the estimate of the relative risk. Confounding factors can also affect the estimate of the relative risk. Techniques such as the Mantel-Haenszel test and logistic regression can be used to adjust for confounding factors in the data analysis. However, such statistical techniques cannot make up for errors in study design or data collection. Another measure of association is the attributable fraction, which measures the proportion of disease occurrence that is associated with the factor of interest.

Case-control studies have many advantages and disadvantages compared with cohort studies (Table 11–15). Among the advantages are their lower costs, shorter time to completion, and the ability to examine the association of many

Table 11–15. Advantages and Disadvantages of Case-Control Studies

Advantages

1. Generally a short study period.
2. One may study rare diseases.
3. Inexpensive.
4. One may study several risk factors for a single disease.
5. Useful for studying adverse drug reactions or new diseases.

Disadvantages

1. Sometimes difficult to choose appropriate control.
2. Sometimes difficult to determine if exposure preceded the disease.
3. Prone to biases in selection and information.
4. One is usually unable to calculate incidence rates.

factors with a given disease. Among their disadvantages are the potential for bias in case and control selection, the potential for recall bias during data collection, and the possible bias associated with investigating survivors of a disease.

STUDY PROBLEMS

1. What would be an appropriate control group (or groups) for the following conditions (mention possible exclusions):
   (a) Babies born at very low birth weight (≤1500 grams).
   (b) Infants with chronic ear infections.
   (c) Transplant patients who reject a transplant.
   (d) Russian roulette suicide victims.

2. Marzuk et al. (1992) conducted a case-control study of cocaine and alcohol use as risk factors for suicide by Russian roulette. The controls were handgun suicides. Toxicological analyses were performed and the data below were obtained. The authors did not calculate an odds ratio, but you can. Calculate the odds ratios and write one sentence for each odds ratio explaining its meaning.

(a)	DRUGS OR ALCOHOL PRESENT IN BLOOD	NO DRUGS OR ALCOHOL IN BLOOD	TOTAL
Russian roulette suicide victims	11	3	14
Handgun suicide victims	33	21	54
Total	44	24	68

(b)	COCAINE DETECTED IN BLOOD	NO COCAINE DETECTED IN BLOOD	TOTAL
Russian roulette suicide victims	9	5	14
Handgun suicide victims	19	35	54
Total	28	40	68

3. Name an advantage and a disadvantage of using a case-control study design to test the hypothesis that cocaine use increases the probability of death from Russian roulette.

4. The recent controversy over silicone breast implants began with the observation of breast cancer among women with the implants.
   (a) What is the advantage in using a case-control study to test the hypothesis that silicone breast implants are associated with breast cancer?
   (b) Who should be the cases in such a study?
   (c) What groups would make appropriate controls?
   (d) What variables might one use to select the control group?
   (e) What variables might be useful in group or pair matching?
   (f) What would be the problem in choosing many variables for matching?
   (g) How could one collect information about women's silicone breast implants?
   (h) What problems arise in collecting the women's medical histories?

REFERENCES

Antunes, C.M.F., Stolley, P. D., Rosensheim, N. B., Davies, J. L., Tonascia, J. A., Brown, C., Burnett, L., Rutledge, A., Pokempner, M. and Garcia, R. 1979. "Endometrial cancer and estrogen use." *New Engl. J. Med.* 300:9–13.

Berkson, J. 1946. "Limitations of the application of fourfold table analysis to hospital data." *Biometrics* 2:47–53.

Bross, I.D.J., and Tidings, J. 1973. "Another look at coffee drinking and cancer of the urinary bladder." *Prev. Med.* 2:445–451.

Bross, I.D.J. 1968. "Effect of filter cigarettes on the risk of lung cancer." *Nat. Cancer Inst. Monogr.* 28:35–40.

Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. 1987. "The reduction in risk of ovarian cancer associated with oral contraceptive use." *NEJM* 316:650–655.

Chow, J. M., Yonekura, M. L., Richwald, G. A., Greenland, S., Sweet, R. L., Schachter, J. 1990. "The association between *Chlamydia trachomatis* and ectopic pregnancy." *JAMA* 263(23):3164–3167.

Cochran, W. G. 1954. "Some methods of strengthening the common χ^2 tests." *Biometrics* 10:417–451.

Collaborative Group for the Study of Stroke in Young Women. 1973. "Oral contraception and increased risk of cerebral ischemia or thrombosis." *New Engl. J. Med.* 288:871–878.

Cornfield, J. 1951. "A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix." *J. Natl. Cancer Inst.* 11:1269–1275.

Doll, R. and Hill, A. B. 1952. "A study of the aetiology of carcinoma of the lung." *Brit. Med. J.* 2:1271–1286.

Dosemeci, M., Wacholder, S. and Lubin, J. H. 1990. "Does nondifferential misclassification of exposure always bias a true effect toward the null value?" *Am. J. Epidemiol.* 132(4):746–74

Gail, M., Williams, R., Byar, D. P., and Brown, C. 1976. "How many controls?" *J. of Chronic Disease* 29:723–731.

Glass, R., Johnson, B., and Vessey, M. 1974. "Accuracy of recall of histories of oral contraceptive use." *Brit. J. Prev. Med.* 28:273–275.

Guy, W. A. 1856. "On the nature and extent of the benefits conferred by hospitals on the working classes and the poor." *J. Roy. Stat. Soc.* 19:12–27.

Herbst, A. L., Ulfelder, H. and Poskanzer, D. C. 1971. "Association of maternal stilbestrol therapy with tumor appearance in young women." *NEJM* 284(16):878–881.

Hurwitz, E. S., Barrett, M. J., Bregman, D., et al. 1987. "Public health service study of Reye's Syndrome and medications: Report of the main study." *JAMA* 257(14):1905–1911.

Kahn, H. A. and Sempos, C. T. 1989. *Statistical Methods in Epidemiology*. New York: Oxford University Press.

Kjuus, H., Skjaerven, R., Langard, S., Lien, J. T., Aamodt, T. 1986. "A case-referent study of lung cancer, occupational exposures and smoking. II: Role of asbestos exposure." *Scand. J. Work Environ. Health* 12:203–209.

Kleinbaum, D. G., Kupper, L. L., Morgenstern, H. 1982. *Epidemiologic Research*. Belmont, Calif.: Lifetime Learning Publications.

Kraus, A. S. 1954. "The use of hospital data in studying the association between a characteristic and a disease." *Pub. Health Rep.* 69:1211–1214.

———. 1958. "The Use of Family Members as Controls in the Study of the Possible Etiologic Factors of a Disease." Sc.D. Thesis, Graduate School of Public Health, University of Pittsburgh.

Levin, M. I., Goldstein, H., and Gerhardt, P. R. 1950. "Cancer and tobacco smoking: A preliminary report." *JAMA* 143:336–338.

Levin, M. I., Kraus, A. S., Goldberg, I. D., and Gerhardt, P. R. 1955. "Problems in the study of occupation and smoking in relation to lung cancer." *Cancer* 8:932–936.

Lilienfeld, A. M. 1956. "The relationship of cancer of the female breast to artificial menopause and marital status." *Cancer* 9:927–934.

Mainland, D. 1953. "Risk of fallacious conclusions from autopsy data on incidence of disease with applications to heart disease." *Amer. Heart J.* 45:644–654.

Mann, J. I., Vessey, M. P., Thorogood, M., and Doll, R. 1975. "Myocardial infarction in young women with special reference to oral contraceptive practice." *Brit. Med. J.* 2: 241–245.

Mantel, N., and Haenszel, W. E. 1959. "Statistical aspects of the analysis of data from retrospective studies of disease." *J. Natl. Cancer Inst.* 22:719–748.

Marzuk, P. M., Tardiff, K., Smyth, D., Stajic, M., Leon, A. C. 1992. "Cocaine use, risk taking and fatal Russian Roulette." *JAMA* 267(19):2635–2637.

McMahan, C. A. 1962. "Age-sex distribution of selected groups of human autopsied cases." *Arch. Path.* 73:40–47.

McPherson, K., Vessey, M. P., Neil, A., Doll, R., Jones, L., Roberts, M. 1987. "Early oral contraceptive use and breast cancer: results of another case-control study." *Br. J. Cancer* 56:653–660.

Pasternak, B., Shore, R. E., Albert, R. E. 1977. "Occupational exposure to chloromethyl ethers." *J. Occupational Medicine* 19:741–746.

Schlesselman, J. J. 1982. *Case-Control Studies: Design, Conduct, Analysis*. New York: Oxford University Press.

Snedecor, G. W., and Cochran, W. G. 1967. *Statistical Methods* 6th ed. Ames, Iowa: The Iowa State University Press.

Solomon, H. A., Priore, R. I., and Bross, I.D.J. 1968. "Cigarette smoking and periodontal disease." *J. Amer. Dent. Assoc.* 77:1081–1084.

Stolley, P. D., Tonascia, J. A., Sartwell, P. E., Tockman, M. S., Tonascia, S., Rutledge, A., and Schinnar, R. 1978. "Agreement rates between oral contraceptive users and prescribers in relation to drug use histories." *Am. J. Epid.* 107:226–235.

Szmuness, W. 1978. "Hepatocellular carcinoma and the hepatitis B virus: Evidence for a causal association." *Prog. Med. Virol.* 24:40–69.

Thomas, D. B. 1972. "Relationship of oral contraceptives to cervical carcinogenesis." *Obstet. Gynec.* 40:508–518.

Vessey, M. P., and Doll, R. 1968. "Investigation of relation between use of oral contraceptives and thromboembolic disease." *Brit. Med. J.* 2:199–205.

Waife, S. O., Lucchesi, P. F., and Sigmond, B. 1952. "Significance of mortality statistics in medical research: Analysis of 1,000 deaths at Philadelphia General Hospital." *Ann. Intern. Med.* 37:332–337.

Walter, S. D. 1975. "The distribution of Levin's measure of attributable risk." *Biometrics* 62:371–374.

———. 1978. "Calculation of attributable risk from epidemiological data." *Int. J. Epid.* 7:175–182.

Winkelstein Jr., W., Stenchever, M. A., and Lilienfeld, A. M. 1958. "Occurrence of pregnancy, abortion, and artificial menopause among women with coronary artery disease: A preliminary study." *J. Chron. Dis.* 7:273–286.

COMMENTARY

Third-generation oral contraceptives: how risky?

See pages 1569, 1575, 1582, 1589, 1593

The risk of venous thromboembolic disease in a woman who is taking oral contraceptives (OCs) exceeds that of other women.[1] In view of the results reported in this issue by the World Health Organization (WHO), Jick et al, and Bloemenkamp et al, can we now conclude that women who are taking a "third generation" OC—ie, one that contains, as its progestagen component, either desogestrel or gestodene—are at particularly high risk? We must look to such non-randomised epidemiological studies for an answer because (*a*) differences among OCs in their influence on coagulation variables are generally modest, and may or may not be relevant clinically; and (*b*) it is not feasible to conduct randomised trials of sufficient size to examine the possibility of an adverse effect that may occur in only a tiny fraction of OC users.

The latest results provide reasonably strong evidence that users of third-generation OCs have a higher risk of venous thromboembolic disease than do users of other OCs, and further suggest that the newer OCs are in fact responsible. Each study was large, and each came to the same conclusion: that there was approximately a two-fold difference in risk between current users of third-generation OCs and other OCs. The increases in risk were similar for desogestrel and gestodene. The association persisted after adjustment for several risk factors for venous thromboembolism that might have influenced the choice of preparation—eg, age, weight, smoking, parity, and varicose veins. The hypothesis of a causal relation receives modest support from the previously documented influence of hormonal composition of OCs on a woman's risk of vascular disease. Increases in both the oestrogenic and progestagenic potency of OCs are associated with an increased risk of arterial disease,[2,3] and the risk of venous thromboembolism probably rises with increasing oestrogen dose.[2-6] However, while the work on thromboembolism in relation to progestagen potency of earlier OCs has not been extensive, it does not point to an association.[1,6]

Are there some users of third-generation OCs whose risk of venous thromboembolic disease is unusually high, relative to their risk if they were taking a different OC? The WHO data suggest that the added risk to a woman taking a third-generation OC is roughly the same irrespective of body mass index. Bloemenkamp et al show that the magnitude of the added risk is not influenced by family history of venous thrombosis, but may be especially large in women who have never been pregnant or who are carriers of the factor V Leiden mutation.

Is the evidence of an increased risk of venous thromboembolism in users of third-generation OCs strong enough for health professionals to recommend that women discontinue or not start taking such a preparation and use another OC instead? A recommendation of this sort must take into account the size of both risks and benefits related to different types of OCs. The increased risk of venous thromboembolic disease attributable to use of a third-generation OC, beyond the risk associated with use of an earlier OC, seems to be about 10–15 per 100 000 woman-years of use. If the typical case-fatality was about 1%, the increased rate of fatal venous thromboembolism would be 1–1·5 per million woman-years. Unfortunately, we know very little about the risks and benefits of any serious health outcome other than venous thromboembolic disease. The data of Jick et al provide some reassurance that mortality from vascular disease as a whole among current users of third-generation OCs, about half of which is due to arterial disease, does not differ from that of users of other OCs. However, only very substantial differences in risk would have been detected in that study. Possible differences in the incidence of myocardial infarction or diabetes mellitus—differences that could well be present and favour users of third-generation OCs, if metabolic responses are any guide[8]—have not been examined.

In practical terms, what do women and their health advisers need to know? Certainly women who have been or who are considering using third-generation OCs need to be aware of the probable increased risk of venous thromboembolic disease. However (and putting aside issues such as menstrual cycle control), the actual decision comes down to weighing an increase in this risk, one that would cause about one death in one million users each year, against a possible decrease in the risk of other serious conditions. Until (i) we know more about the relation of incidence and mortality of other important health outcomes between users of third-generation and earlier OCs; or (ii) a subgroup of women can be identified who are at very much higher risk of venous thromboembolic disease with third-generation OCs than with earlier OCs, women will not have a sound basis for making a decision.

Noel Weiss
Department of Epidemiology, University of Washington, Seattle, WA, USA

1 Stadel BV. Oral contraceptives and cardiovascular disease. *N Engl J Med* 1981; **305**: 672–77.
2 Meade TW, Greenbert G, Thompson SG. Progestagens and cardiovascular reactions associated with oral contraceptives and a comparison of the safety of 50- and 30-μg oestrogen preparations. *BMJ* 1980; **280**: 1157–61.
3 Kay CR. Progestogens and arterial disease: evidence from the Royal College of General Practitioners' study. *Am J Obstet Gynecol* 1982; **142**: 762–65.
4 Vessey M, Mant D, Smith A, Yeates D. Oral contraceptives and venous thromboembolism: findings in a large prospective study. *BMJ* 1986; **292**: 526.
5 Helmrich SP, Rosenberg L, Kaufman D, Strom B, Shapiro S. Venous thromboembolism in relation to contraceptive use. *Obstet Gynecol* 1987; **69**: 91–95.
6 Gerstman BB, Piper JM, Freiman JP, et al. Oral contraceptive oestrogen and progestin potencies and the incidence of deep venous thromboembolism. *Int J Epidemiol* 1990; **19**: 931–36.
7 Vandenbroucke JP, Koster T, Briet E, Reitsma PH, Bertina RM, Rosendaal FR. Increased risk of venous thrombosis in oral contraceptive users who are carriers of factor V Leiden mutation. *Lancet* 1994; **344**: 1453–57.
8 Speroff L, DeCherney A, and the Advisory Board for the New Progestins. Evaluation of a new generation of oral contraceptives. *Obstet Gynecol* 1993; **81**: 1034–47.

Waiting for coronary artery bypass surgery: abusive, appropriate, or acceptable?

See page 1605

Waiting for the doctor is as old as the profession itself. In ancient days long waiting times were the signboard of a wise and skilful doctor or signified the presence of