relationship in real life is greater than is revealed by the data. In the above example the misclassifications of patients regarding their blood pressure or stroke status reduced an actual twelvefold increase of stroke in hypertensives to an observed twofold increase.

Nevertheless, the study of large groups allows one to detect important relationships, using poor data that are intolerable in conscientious patient care. This, then, is the explanation to the clinician of the seeming tolerance of epidemiology for inadequate data.

REFERENCES

Chapman, J. M., V. A. Clark, and A. H. Coulson. 1966. Problems of measurement in blood pressure surveys: Inter-observer differences in blood pressure determinations. *Am. J. Epidemiology*, **84**:483–494.

Cluff, L. E., G. F. Thornton, and L. G. Seidl. 1964. Studies on the epidemiology of adverse drug reactions: I. Methods of surveillance. *J. Am. Med. Assoc.*, **188**:976–983.

Collen, M. F., J. L. Cutler, A. B. Siegelaub, R. L. Cella. 1969. Reliability of a self-administered questionnaire. *Arch. Intern. Med.*, **123**:664–681.

Feinstein, A. R. 1968. Clinical epidemiology. II. The identification rates of disease. *Ann. Intern. Med.*, **69**:1037–1061.

Friedman, G. D., C. C. Seltzer, A. B. Siegelaub, R. Feldman, and M. F. Collen. 1972. Smoking among white, black and yellow men and women: Kaiser-Permanente Multiphasic Health examination data, 1964–1968. *Am. J. Epidemiology*, **96**:23–35.

Madow, W. G., Interview data on chronic conditions compared with information derived from medical records. National Center for Health Statistics Report, Ser. 2, No. 23, U.S. Department of Health, Education, and Welfare, 1967.

Reid, D. D., and J. G. Evans. 1970. New drugs and changing mortality from non-infectious disease in England and Wales. *Brit. Med. Bull.*, **26**:191–196.

Wilcox, J. 1961. Observer factors in the measurement of blood pressure. *Nursing Research*, **10**:4–20.

Yerushalmy, J. 1969. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. *Radiol. Clin. N. Amer.*, **7**:381–392.

Chapter 4

Basic Methods of Study

In the two preceding chapters the reader has been introduced to the data employed in epidemiology and the basic measurements that are used to describe groups of persons. It is now appropriate to consider the major types of epidemiological investigation. Each type of study uses these tools in a particular way and has a unique logical framework. In addition, each type of study is especially appropriate for the unique circumstances surrounding any particular investigation—the aims of the investigation, the populations available for study, and the human and financial resources that can be brought to bear on the problem.

Relationships

Much of the effort of medical scientists in understanding the etiology of disease and developing appropriate therapies involves a study of the relationship of one type of event or characteristic or "variable" to another. Consider the following questions as examples:

Does exposure to cold wet weather predispose to the common cold?

What is the influence of the serum potassium concentration on the contractility of the heart?

Is obesity related to the occurrence of gallstones?

What is the effect of vitamin C deprivation on wound healing?

Which part of the hemoglobin molecule carries the oxygen?

Does BCG vaccination provide protection against pulmonary tuberculosis?

In Table 4-1 these questions are listed together with the relationship that should be studied to help answer each. In a two-variable relationship one is usually considered the *independent* variable, which affects the other, or *dependent*, variable.

The relationships that are studied need not be only between one variable and a second. Often, the investigator must be concerned with the interrelationship of three or more variables. For example, in order to better understand the relationship of potassium to the force of cardiac contraction, calcium concentration must also be taken into account. Whether or not obesity is related to gallstone occurrence may depend on racial characteristics and the type of diet eaten, both of which must be considered and assessed as additional independent variables.

Observational versus Experimental Studies

There are two basic approaches to investigating the relationship between variables. In *observational* studies, nature is allowed to take its course and changes or differences in one characteristic are related to changes or differences in the other, if any. In *experimental* studies, the investigator actually intervenes and makes one variable change and then sees what happens to the other. In doing so he tries, as much as possible, not to allow other important variables to affect the outcome. By controlling the experimental situation, he may conclude that the intervention or manipulation of the in-

Table 4-1 Examples of Relationships Studied in Order to Answer Certain Questions

Question	Suggests study of the relationship between variables: Independent variables	Dependent variables
Does exposure to cold wet weather predispose to the common cold?	Daily weather conditions	Incidence of common cold
What is the influence of the serum potassium concentration on the contractility of the heart?	Serum potassium concentration	Stroke output of the heart
Is obesity related to the occurrence of gallstones?	Skinfold thickness	Prevalence of gallstones
What is the effect of vitamin C deprivation on wound healing?	Vitamin C content of the diet	Tensile strength of healing wounds
Which part of the hemoglobin molecule carries the oxygen?	Portion of hemoglobin molecule	Affinity for oxygen
Does BCG vaccination provide protection against pulmonary tuberculosis?	Presence or absence of vaccination	Incidence of tuberculosis

dependent variable actually affected, or caused the change in, the dependent variable.

Epidemiology includes both observational and experimental studies. An example of an epidemiologic experiment was the large-scale field trial of poliomyelitis vaccine in which two large groups of children, comparable in all important respects (e.g., age, health, socioeconomic status, and likelihood of exposure to poliomyelitis virus) received vaccine and placebo, respectively, with follow-up to measure the subsequent incidence of poliomyelitis (described in Chap. 9).

Because of the difficulties of performing well-controlled experiments on human populations and the availability of an abundance of observational data, epidemiologists have tended to concentrate on observational studies. In doing so they have tried to "control" the important extraneous variables by their data-analysis methods. Also, they are always on the lookout for "natural experiments"—spontaneous occurrences which approximate experiments by virtue of a change in only one independent variable that is apparently unaccompanied by changes in other important variables. An example might be the sudden graded exposure to ionizing radiation received in 1945 by the population of Hiroshima, which has permitted the study of the relationship of different doses of radiation exposure to the subsequent development of a variety of diseases.

Such natural experiments are rare (thank goodness) and the observational epidemiologist has to rely on other techniques and criteria for determining the possible effects of additional variables.

Observational studies fall into two main categories, descriptive and analytic. These studies, in turn, may be subdivided into cross-sectional or prevalence studies, case-control studies, and incidence or cohort studies, depending on the groups of persons investigated and the time relationships involved. (Case-control studies are probably best included only in the analytic category.) These will be described subsequently. Attention will also be paid to defining and clarifying the relationship between prospective and retrospective studies due to the confusion that revolves around this distinction.

Descriptive versus Analytic Studies

There are two fundamental objectives of observational epidemiologic studies. One is to *describe* the occurrence of disease or disease-related phenomena in populations. The other is to *explain* the observed pattern of occurrence of disease. Seeking the latter objective involves the identification of causal or etiological factors.

Descriptive studies usually involve the determination of the incidence, prevalence, and mortality rates for diseases in large population groups, according to basic group characteristics such as age, sex, race, and geographic area. In this way, the general distribution of disease in the population is described.

Studies attempting to explain disease are often referred to as *analytic* studies. The starting point for an analytic study is often a descriptive finding that raises certain questions or suggests certain hypotheses that require further investigation. With analytic studies the investigator has a specific question or group of questions in mind that he sets about to answer.

The distinction between descriptive and analytic studies is not clear-cut. A large-scale descriptive study may (perhaps unexpectedly) provide abundant and impressive data that give a clear answer to a specific question. In an analytic study, designed to answer specific questions, data collected incidentally may be of great descriptive interest and raise further questions for investigation.

Despite this fuzziness, it is often useful to categorize epidemiologic studies in this manner. Descriptive studies usually involve a more diffuse, superficial, or general view of a disease problem. Analytic studies narrow down on a specific question and may require a more rigorous study design and data analysis.

Prevalence or Cross-Sectional Studies

Prevalence, or *cross-sectional*, studies examine the relationships between diseases and other characteristics or variables of interest as they exist in a defined population at one particular time. The presence or absence of disease and the presence or absence of the other variables (or, if they are quantitative, their level) are determined in each member of the study population or in a representative sample at one particular time. The relationship between a variable and the disease can be examined in two ways, either (1) in terms of the prevalence of disease in different population subgroups defined according to the presence or absence (or level) of the variables or, conversely, (2) in terms of the presence or absence (or level) of the variables in the diseased versus the nondiseased.

Case-Control Studies

Case-control studies are similar to prevalence studies in that they assess the relationship of *existing* disease to other variables or attributes. After the initial identification of cases, that is, location of persons with the disease of interest, a suitable control group or comparison group of persons without the disease is identified. The relationship of an attribute to the disease is examined by comparing the diseased and nondiseased with regard to how frequently the attribute is present or, if quantitative, what the levels of the attribute are in the two groups.

Incidence or Cohort Studies

Instead of measuring the relationship of attributes to existing disease, as do prevalence and case-control studies, *incidence*, or *cohort*, studies look more directly at attributes or factors related to the *development* of disease. A study population free of the disease under investigation is identified at a particular time. The attributes of interest are measured initially in this group of persons, known as a *cohort*. Then, these persons are followed up over a period of time for the development of the disease being studied. The relationship of an attribute to the disease is examined by dividing the population into subgroups according to the presence or absence (or level) of the attribute initially and comparing the subsequent incidence of disease in each of the subgroups.

An Illustrative Example

Prevalence, case-control, and incidence studies are discussed in detail in Chapters 6, 7, and 8, respectively. At this point an example may help to clarify the distinction among these study plans. Suppose we wish to learn whether obesity predisposes to degenerative arthritis of the knees. In a prevalence study we would x-ray the knees of a defined population, perhaps all the adults in a community, and determine degree of obesity by measuring height and weight or skinfold thickness. We would then compare the prevalence of osteoarthritis in population subgroups showing various degrees of obesity. Or, we may wish to contrast the mean skinfold thickness or other obesity measure in those with osteoarthritis and those without.

In a case-control study of this question, we might collect a group of persons with osteoarthritis of the knees hospitalized at a local hospital during the past year. For a control group, we might select for each osteoarthritis case, a person of the same sex and similar age, admitted to the same hospital during the same week for minor elective surgery such as herniorrhaphy or hemorrhoidectomy. We would then compare the recorded heights and weights of the case group with those of the control group to see if, indeed, the osteoarthritis cases were more obese.

To approach this problem by an incidence study, we would go back to a defined adult population and x-ray their knees to exclude persons with existing osteoarthritis. We would then measure skinfold thickness or height and weight in order to divide the population without osteoarthritis into the obese and nonobese or, preferably, some finer gradations of fatness. We would call them back 10 years later for repeat knee x-rays, which would demonstrate new cases of osteoarthritis. Then we would compare the incidence of osteoarthritis in the various fatness groups.

Remembering our original question, "Does obesity *predispose* to osteoarthritis?" the incidence study approach seems to provide the most direct answer, since we looked for obesity *before* the osteoarthritis developed. The prevalence and case-control studies provided only indirect evidence, since they looked at obesity at the same time as disease. However, the time sequence can often be taken into account in the prevalence and case-control studies. In addition to measuring current weight in persons with and without osteoarthritis we could also have inquired about their weight 10 years ago, or at age 25, or before their knees started to hurt, thus investigating their weight prior to the development of osteoarthritis. The information obtained may not be as accurate as that derived from weighing the subjects initially in an incidence study, but time sequence can be considered in prevalence or case-control studies.

Prospective and Retrospective Studies

The question of time sequence leads naturally into a consideration of the much-discussed *prospective* and *retrospective* studies. It is almost a matter of faith that investigations are unsatisfactory if they are retrospective. One often hears such comments as, "Of course, this study was retrospective, so we can not be confident of the findings."

Before discussing the merits of prospective versus retrospective studies, it is important to clarify their meaning. Actually, much confusion has resulted because the terms are used in two different ways leading to such semantic horrors as "retrospective-prospective" studies.

One of the meanings of prospective versus retrospective has to do with the time period over which the data were recorded in relation to the time the decision was made to do the study. In this sense, retrospective studies involve a decision to carry out an investigation with observations that have been recorded in the past. In contrast, prospective studies involve the collection of observations after the decision is made to carry out the investigation.

The other meaning of prospective versus retrospective studies is related not to the time sequence of the observations and the decision to do the study but, rather, to the time sequence of observations of study variables and the occurrence of disease. In this sense, prospective studies are analogous to incidence studies, and retrospective studies are analogous to prevalence, or case-control, studies. Prospective or incidence studies measure characteristics and wait for disease to develop, while retrospective or prevalence studies measure the characteristics in persons already diseased.

It is strongly suggested that this second set of definitions be discarded, since better terms are available, as noted. The advantages and disadvantages of prevalence, incidence, and case-control studies will be discussed in Chaps. 6, 7, and 8. The following discussion of prospective versus retrospective studies will consider only the first pair of definitions, relating to when the data were collected.

In prospective studies the investigator can plan and control the methods for making and recording observations, keeping in mind their purpose. In retrospective studies the already-recorded data may have been collected for an entirely unrelated purpose. Therefore these data may well be incomplete and recorded in a manner not appropriate for the present study.

Consequently, there often are severe problems involved in retrospective studies. Consider a study of changes in the outcome of treatment of congestive heart failure in a particular hospital over a period of several years. In carrying out a retrospective study, the investigator would be plagued by the fact that the criteria for the diagnosis of congestive heart failure vary over the years and vary from doctor to doctor. The recent advent of central venous pressure measurements may have improved the ability to diagnose the condition. Cases diagnosed many years ago may differ in character and severity from those diagnosed last year. Therefore observed changes in outcome may be related more to differences in initial severity than to the effects of treatment. If one of the criteria for improvement were weight loss, the investigator would find, to his frustration, that admission and discharge weights were not recorded for many patients over the years, ruling them out of this aspect of the study.

If this study were carried out prospectively, the investigator could initially establish criteria for the diagnosis of congestive heart failure and set up objective measures of severity and improvement. In addition, he could establish procedures to ensure that all the needed measurements were made uniformly on all patients. Thus, the superiority of a prospective study of this question is obvious.

Not all retrospective data need be of poor quality. If we again consider retrospective studies using hospital charts, a variety of data come to mind that would probably have been recorded accurately and consistently. Examples are time of admission, number of days spent in the hospital, sex of the patient, whether the patient died, and whether he received any blood transfusions.

REFERENCE

MacMahon, B., and T. F. Pugh, *Epidemiology: Principles and Methods.* (Boston: Little, Brown, 1970), Chap. 3.

Chapter 5

Descriptive Studies

Descriptive epidemiologic studies reveal the patterns of disease occurrence in human populations. They provide general observations concerning the relationship of disease to basic characteristics. These characteristics include such personal items as age, sex, race, occupation, and social class. Also of great importance are geographic location and time of occurrence of disease. Thus, the major characteristics of interest in descriptive epidemiology may be summarized under the categories: person, place, and time.

At first glance, the goal of describing disease occurrence in this way may seem trivial and not worthy of the efforts of medical scientists. However, such studies are of fundamental importance and serve a variety of purposes—chiefly:

1 Alerting the medical community as to what types of persons (e.g., young or old, male or female, "white collar" workers or "blue collar" workers) are most likely to be affected by a disease, where

the disease will occur, and when. This information is of great value to the physician in making a diagnosis, even though he may not be aware that he is using it.

2 Assisting in the rational planning of health- and medical-care facilities (e.g., number of coronary-care-unit beds needed for the cases of myocardial infarction in a particular community).

3 Providing clues to disease etiology and questions or hypotheses for further fruitful study (e.g., low prevalence of tooth decay in certain areas in the United States suggested further studies concerning the value of fluoride in drinking water).

PERSON

Basic demographic and social characteristics of persons constitute the attributes of greatest concern. Among these characteristics are age, sex, race or ethnic group, marital status, social class or socioeconomic status, religion, and occupation.

Age

Age is one of the most important factors in disease occurrence. Some diseases occur almost exclusively in one particular age group, such as hypertrophic pyloric stenosis in young infants or carcinoma of the prostate in the elderly. Other diseases occur over a much wider age span, but tend to be more prevalent at certain ages than others.

The time of life at which an infectious disease predominates is influenced by such factors as the degree of exposure to the agent at various ages, variations in susceptibility with age, and the duration of the immunity developed after infection. The influence of age-related exposure and duration of immunity is illustrated by the contrast between the single occurrence of chicken pox almost exclusively in young children and the repeated occurrence of gonorrhea, predominantly in adolescents and young adults. Chicken pox is readily transmitted among children playing together or gathering in classrooms, and it produces a lifelong immunity. Gonorrhea is transmitted by sexual contact and results in no immunity.

Many chronic or degenerative diseases such as coronary heart disease and osteoarthritis show a progressive increase in prevalence with increasing age. It is tempting to regard a disease with this age pattern as being due merely to aging itself. It should be remembered, however, that increasing age also marks the passage of time, during which the body is accumulating exposure to harmful environmental influences. For example, the wrinkling and loss of elasticity in skin that we associate with aging can be brought about or accelerated by chronic exposure to sunlight.

Instead of adopting the fatalistic view that a disease is an inevitable consequence of aging, a search for other causative factors should be undertaken. One of the great contributions of epidemiology in the past few decades has been to show that atherosclerosis and its consequences are not due merely to aging, as was previously thought, but that a person's habits and manner of living may contribute importantly to this disease process.

To see how age patterns of disease occurrence lead to clues and hypotheses, note the age trend, reported by Lilienfeld (1956), in the incidence of breast cancer among single and married women in New York State (Fig. 5-1). The steady geometric (note the logarithmic scale) increase in incidence with age diminishes sharply in the forties with a lesser continuing increase in the older years. The reduction in the forties of the rate of increase with age suggested the hypothesis that the hormonal changes of the menopause tend to decrease susceptibility to breast cancer. This hypothesis continues to be of great interest to scientists studying the causes of breast cancer.

Current and Cohort Age Tabulations The tabulation of disease rates in relation to age at one particular time, as in Fig. 5-1, is known as a *current*, or *cross-sectional*, presentation. This shows disease rates as they are occurring simultaneously in different age groups; thus, different people are involved in each age group. The other way to tabulate and analyze age relationships is in terms of *cohorts*. A cohort is a specific group of people, identified at one period of time and followed up as they pass through different ages during part or all of their life-span.

The results of cross-sectional and cohort age analysis can differ

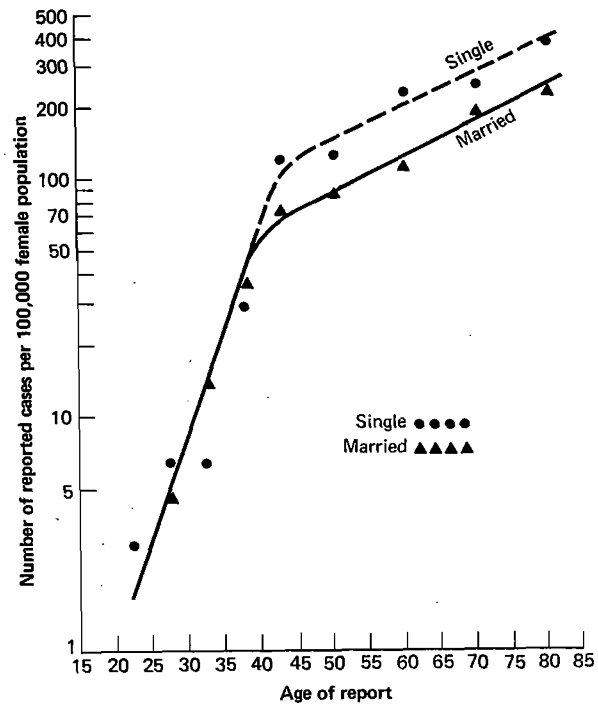


Figure 5-1 Annual age-specific incidence rates of reported cases of female breast cancer, by marital status, New York State exclusive of New York City, 1949. *(Reproduced, by permission, from Lilienfeld, 1956.)*

to a surprising degree, and either approach may be more appropriate for a particular problem. In a classic study, Frost (1939) compared cross-sectional and cohort age analyses of tuberculosis death rates in Massachusetts. Fig. 5-2 shows the cross-sectional curves for males in the years 1880, 1910, and 1930. First of all, note that at all ages the mortality rates decreased between 1880 and 1930. Also, observe that the shapes of the age curves were changing. The 1930 curve was of particular concern to public health workers in the 1930's because it showed tuberculosis mortality rates rising with age

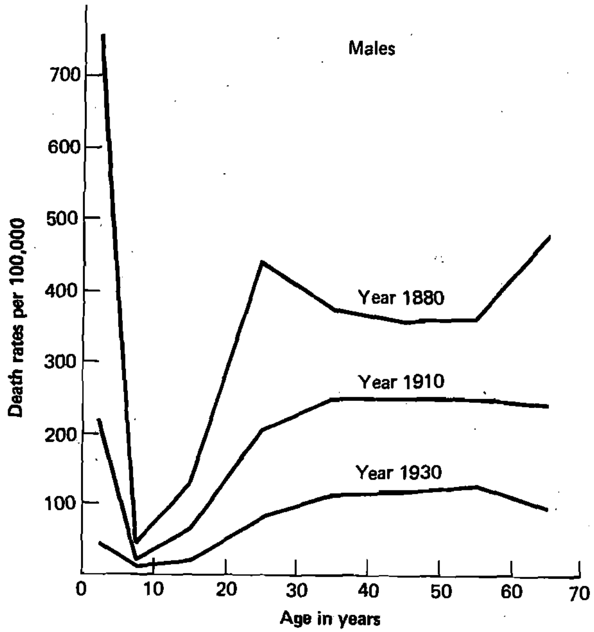


Figure 5-2 Massachusetts death rates from tuberculosis—all forms—by age, 1880, 1910, 1930. *(Reproduced, by permission, from Frost, 1939.)*

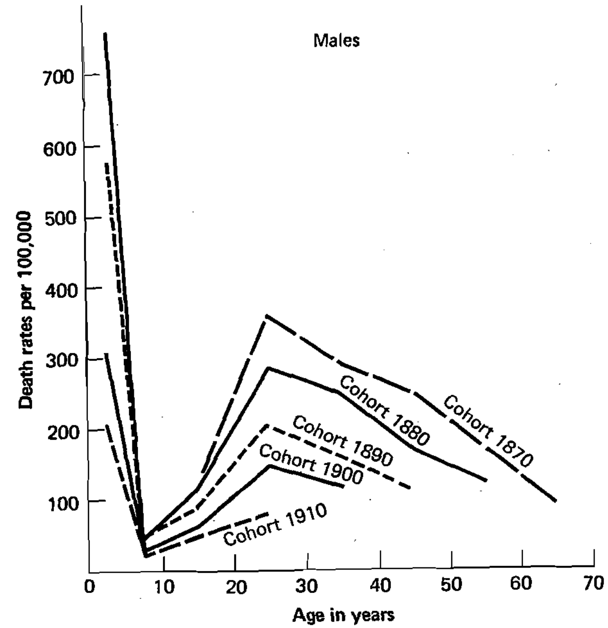
in adult life, reaching a maximum between age 50 and 60. This new age pattern of susceptibility to death from tuberculosis was thought possibly to be due to the failure of many individuals to become infected and acquire immunity during youth.

The matter was clarified, however, by a cohort analysis, shown in Fig. 5-3. Each curve represents the cohort of persons born in the 10-year period leading up to and including the year shown above the curve (e.g., 1870 cohort included individuals born from 1861–1870). In this cohort analysis it is apparent that each group experienced its maximum adult risk of tuberculosis death during the age decade of the twenties with subsequent decline in risk with age. Thus, for any particular group of adults, there was no increase in susceptibility with age, after all. The reason that, in 1930, persons in their fifties

had higher death rates than younger persons was because they belonged to the 1880 cohort, which experienced a greater exposure to tuberculosis than any of the succeeding cohorts.

Note that the frequently quoted calculations of average life expectancy are determined essentially on a cross-sectional rather than a cohort basis, using what are known as "life tables" (Hill, 1971). The presently observed annual mortality rates for each year of age are applied successively to a hypothetical population beginning either at birth or at some other starting point. It is assumed that as this hypothetical population ages, year by year, it will experience the same mortality rates for each year of life as are now observed in the current population at each age. Actually, the cohort of persons born now may be exposed to different risks of death as they go through

Figure 5-3 Massachusetts death rates from tuberculosis—all forms—by age, in successive 10-year cohorts. *(Reproduced, by permission, from Frost, 1939.)*



various age periods of life than are experienced by persons who are *now* in those age periods.

Sex

Some diseases occur more frequently in males, others, more frequently in females. If sex-linked inheritance can be excluded, a sex difference in disease incidence initially brings to mind the possibility of hormonal or reproductive factors that either predispose or protect. For example the greater occurrence of coronary heart disease in young men than in young women cannot be explained entirely by sex differences in the so-called coronary risk factors such as blood lipid concentrations, blood pressure, cigarette smoking, diabetes mellitus, and obesity. There may be some important hormonal factor that contributes to the male-female difference—perhaps protection of the female by estrogens before menopause. Similarly, the greater prevalence of gallstones in women than in men is probably attributable, in part, to the effects of repeated pregnancies and, in addition, to hormonal effects on bile composition.

But men and women differ in many other ways, including habits, social relationships, environmental exposures, and other aspects of day-to-day living. The higher male prevalence of cirrhosis of the liver and chronic bronchitis are at least partly related to the fact that, on the average, men currently drink more alcohol and smoke more cigarettes than women.

Sex differences in disease occurrence are important descriptive findings and often suggest avenues for further research. No disease can be considered to have a well-understood etiology if it manifests a male or female predominance which is not explained.

Race

Racial differences in disease prevalence have often been noted. In the case of some diseases (e.g., black-white differences in sickle-cell anemia and skin cancer), the differences are genetically determined. With other diseases, the explanation may not be so simple, especially when racial differences are accompanied by differences in socioeconomic status.

A case in point is the higher prevalence of hypertension and its complications in blacks than in whites in the United States. Suggested explanations have included (1) increased genetic susceptibility in blacks, (2) increased emotional stress in blacks due to racial discrimination, (3) lower average socioeconomic levels in blacks (since in whites the prevalence of hypertension is higher in lower socioeconomic groups), and (4) less access for blacks to good medical care (Howard and Holman, 1970). It may eventually be shown that some or all of these mechanisms are involved.

Marital Status

Marital status is another important descriptive variable. Married persons have lower mortality rates than single persons, including both overall mortality and mortality from most specific diseases. Whether the married state provides health benefits or whether characteristics favoring long life also predispose to marriage has not been decided.

Of great interest in studies of cancer etiology has been the contrast between cancer of the breast and cancer of the uterine cervix in their relation to marital status. Breast cancer is more apt to develop in single women or women who marry late in life, while cervical cancer is associated with early marriage. Further studies stemming, in part, from these observations suggest that cervical cancer is associated with coital activity at an early age and that having a first pregnancy at an early age may help protect a woman from breast cancer.

The data regarding the relationship of breast cancer incidence to age (see Fig. 5-1) also revealed a higher incidence in single women than in married women in their forties and later age decades. Lilienfeld suggested the hypothesis that early artificial or surgical menopause, occurring more often in married than in single women, might be protective against breast cancer. This hypothesis received some confirmation in an analytic case-control study in which it was found that (1) women with breast cancer less often gave a history of artificial menopause than did control subjects, and (2) married women more often gave a history of artificial menopause than did single women.

Socioeconomic Status

Socioeconomic status or social class is a somewhat nebulous concept, but it can be measured fairly conveniently by the occupation or income of the family head, by his or her educational level, or by residence, in terms of the value and amenities of the home or dwelling unit. The British have used occupation to define five social classes—I. Professional, II. Intermediate, III. Skilled, IV. Partly Skilled, and V. Unskilled. Using this classification system, the Registrar General for England and Wales has provided descriptive data relating social class to a variety of conditions.

As mentioned in our discussion of hypertension, many diseases show a distinct social class gradient, with higher rates in the lower socioeconomic classes. Included are rheumatic heart disease, chronic bronchitis, tuberculosis, stomach ulcer, stomach cancer, and nutritional-deficiency diseases.

On the other hand, low socioeconomic status appears to confer protection against some diseases. In the series of annual poliomyelitis epidemics that began in 1947, the higher social classes were the most severely affected. It is believed that in economically disadvantaged groups, poor sanitary conditions had resulted in widespread subclinical infection in the first few years of life, resulting in immunity. When "higher" living standards prevent this early infection, acquiring poliomyelitis later in childhood is more likely to cause paralytic disease.

A marked socioeconomic gradient in infant mortality has long been noted. Table 5-1 shows social class and rates of infant mortality (at age under one year) per 1,000 live births in England and Wales during two time periods, 1930–1932 and 1949–1953. Note that even though there was a marked improvement by the later time period, Social Class V still had over twice the infant mortality rate observed in Social Class I. Infant mortality rates have often been used as an index both of living standards and of availability of medical services in comparing nations or areas within a nation.

PLACE

Where disease occurs is a matter of great importance. Comparison of disease rates in different places may provide obvious clues to

Table 5-1 Infant Mortality Rates in England and Wales as Related to Social Class during Two Time Periods, 1930–1932 and 1949–1953

Social class	Infant mortality rates* 1930–1932	Infant mortality rates* 1949–1953
I	32.7	18.7
II	45.0	21.6
III	57.6	28.6
IV	66.8	33.8
V	77.1	40.8
All classes	61.6	29.5

*Deaths of infants under one year old per 1,000 live births. Registrar General's data, Taylor and Knowelden (1964).

etiology or serve as a stimulus to further fruitful investigation. The places of concern may be as large as a continent or as small as part of a room. As illustrative examples, descriptive findings will be presented from international comparisons, comparisons of regions within the United States and Canada, and comparisons of areas in a city.

International Comparisons

Because of the problems regarding the validity of mortality statistics, described in Chap. 3, it is difficult to take seriously small differences among nations in mortality rates for specific diseases. However, it is also difficult to explain away very large differences as due to artifact—that is, where the death rate for a disease in one country is two or three times as large as the death rate in another. Large differences are particularly impressive when both countries are known to have reasonably good vital statistics systems.

The Unique Position of Japan Ranking the disease-specific mortality rates of various nations has revealed Japan to be among the highest nations for some diseases and among the lowest for others. Table 5-2 shows some age-adjusted or age-specific mortality rates for stomach cancer, colon cancer, breast cancer, cerebrovascular disease (primarily strokes), and coronary heart disease. Note that among the nations studied, Japan is the highest ranking country for stomach cancer and cerebrovascular disease and the lowest ranking for breast cancer, colon cancer, and coronary heart disease.

Because of Japan's unique position among nations regarding these diseases, consideration of the mode of life in Japan has suggested a number of questions and hypotheses for further study. The relatively frequent practice and long duration of breast feeding of infants in Japan has raised the question of whether lactation may diminish the risk of developing breast cancer. The traditional Japanese diet has come under considerable scrutiny in hopes of finding predisposing factors for stomach cancer and protective factors for colon cancer. Also, the low fat intake in Japan has been thought responsible for the low average serum-cholesterol levels observed there and the low incidence of coronary heart disease. Although some portion of the high cerebrovascular death rate may be due to a known tendency of Japanese to attribute any sudden death to a cerebral hemorrhage, the high salt intake of Japanese has come under suspicion as a possible predisposing factor for hypertension and stroke.

Many traditional practices in Japan are now changing, and it will be of considerable interest to learn whether disease rates will change in ways that are consistent with the above hypotheses. Already, the migration of Japanese to places where they adopt new eating and living habits has permitted comparative studies aimed at identifying environmental factors that predispose to disease. Gordon (1957) compared mortality rates for Japanese in Japan, Hawaii, and the United States mainland and found contrasting trends for cerebrovascular disease and for coronary heart disease. Cerebrovascular disease mortality rates in both sexes decreased, and coronary mortality rates in men increased from Japan to Hawaii to the United States mainland. This suggested that as Japanese were adopting "the American way of life," their susceptibility to the two diseases in question was moving in the direction of that found in other Americans. (The assumption that migrants are genetically similar to those who remain in their native land should always be viewed with caution.) In order to explore in detail the reasons for the

Table 5-2 Ranking of Death Rates of Various Countries, from Highest to Lowest*

Rank	Stomach cancer, males, 1964–1965	Colon cancer (except rectum), males, 1964–1965	Breast cancer, females, 1964–1965	Cerebrovascular disease, males, ages 65–74, 1964	Coronary heart disease, males, ages 45–54, 1964
1	Japan (69)	Scotland (16)	Netherlands (26)	Japan (1,680)	Finland (442)
2	Chile (58)	Denmark (14)	England & Wales (24)	Scotland (901)	Scotland (358)
3	Austria (42)	U.S.A.—white (14)	Denmark (24)	Finland (751)	U.S.A. (354)
4	Finland (40)	Canada (13)	Scotland (24)	West Germany (750)	Australia (324)
5	West Germany (37)	New Zealand (13)	Canada (23)	Italy (708)	Northern Ireland (324)
6	Italy (34)	Northern Ireland (13)	New Zealand (23)	Hungary (706)	Canada (311)
7	Portugal (33)	Ireland (13)	South Africa (23)	Austria (663)	New Zealand (293)
8	Netherlands (28)	Australia (12)	Northern Ireland (22)	Northern Ireland (656)	England & Wales (245)
9	Belgium (27)	Belgium (12)	U.S.A.—white (22)	Israel (614)	Israel (214)
10	Switzerland (26)	England & Wales (12)	Switzerland (22)	England & Wales (612)	West Germany (183)
11	Norway (26)	France (12)	Ireland (22)	Australia (611)	Denmark (181)
12	Scotland (25)	Switzerland (11)	Belgium (21)	Czechoslovakia (560)	Norway (164)
13	South Africa (25)	U.S.A.—nonwhite (11)	Israel (21)	Switzerland (530)	Netherlands (162)
14	Ireland (24)	Netherlands (11)	U.S.A.—nonwhite (20)	France (528)	Austria (159)
15	England & Wales (23)	South Africa (11)	Australia (19)	U.S.A. (495)	Belgium (159)
16	Sweden (22)	Austria (10)	Sweden (19)	Norway (492)	Czechoslovakia (151)
17	Northern Ireland (22)	West Germany (10)	West Germany (18)	Denmark (478)	Hungary (147)
18	Denmark (22)	Sweden (10)	Austria (17)	New Zealand (475)	Switzerland (134)
19	France (21)	Norway (8)	Norway (17)	Netherlands (416)	Italy (133)
20	Israel (18)	Italy (8)	France (16)	Canada (414)	Venezuela (131)
21	U.S.A.—nonwhite (18)	Portugal (8)	Italy (16)	Sweden (394)	Sweden (125)
22	Canada (18)	Israel (7)	Finland (14)	Belgium (334)	France (74)
23	New Zealand (17)	Finland (5)	Portugal (13)	Venezuela (281)	Japan (51)
24	Australia (15)	Chile (4)	Chile (9)		
25	U.S.A.—white (9)	Japan (3)	Japan (4)		

*Death rates per 100,000 shown in parentheses (rounded from the original).
Sources: Stomach, colon, and breast cancer: Segi et al. (1969); cerebrovascular and coronary heart disease: World Health Organization (1967).

above geographic trends in cerebrovascular and coronary disease, parallel data collection methods have been established in three on-going epidemiologic studies of Japanese, located in Hiroshima, Japan; Honolulu, Hawaii; and in the San Francisco Bay area (Belsky et al., 1971).

Comparisons of Regions within Countries

The availability of mortality statistics for states and finer geographic subdivisions in the United States and other nations has permitted the discovery of interesting place-to-place variations in disease occurrence. Differences in mortality rates between urban and rural areas are a common finding. The higher mortality from lung cancer in cities than in farming areas is consistent with an etiologic role of either cigarette smoking or air pollution, since both are more common in cities.

The North-to-South Gradient of Multiple Sclerosis Geographic variation within nations may take the form of a distinct north-to-south gradient, which suggests that climate or other factors related to latitude may be involved. An example is the finding in the United States and Canada of generally higher mortality rates for multiple sclerosis the farther north one looks (Fig. 5-4). Confirmation of the north-to-south trend also comes from other nations and from prevalence rates found in several cities (Fig. 5-5). While hypotheses abound, to date no one has convincingly explained this geographic distribution of multiple sclerosis (Alter, 1968).

Areas within a City

When studying disease occurrence within a city, it is often desirable to plot the occurrence of disease in each census tract, since information about other characteristics of persons in each tract is available.

Rheumatic Fever in Baltimore Figure 5-7 from Gordis et al. (1969) shows the distribution in Baltimore census tracts of the homes of hospitalized rheumatic fever cases in 1960–1964. Most of

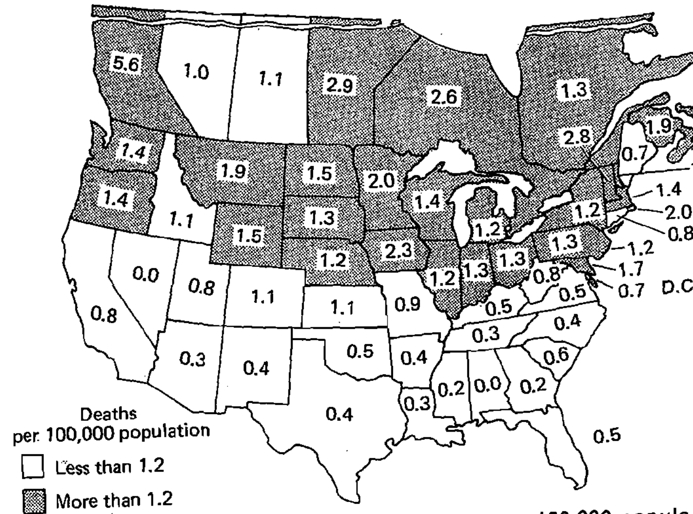
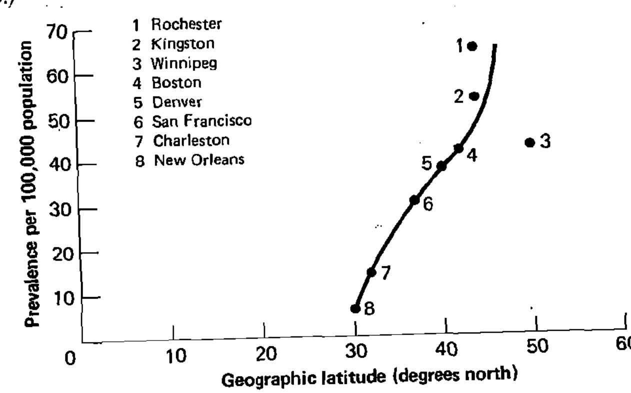


Figure 5-4 Multiple sclerosis mortality rates per 100,000 population in the U.S. states and in Canadian provinces. *(Reproduced, by permission, from Alter, 1968, adapted from Limburg, 1950.)*

Figure 5-5 Multiple sclerosis prevalence rates in cities of the United States and Canada. *(Reproduced, by permission, from Alter, 1968.)*



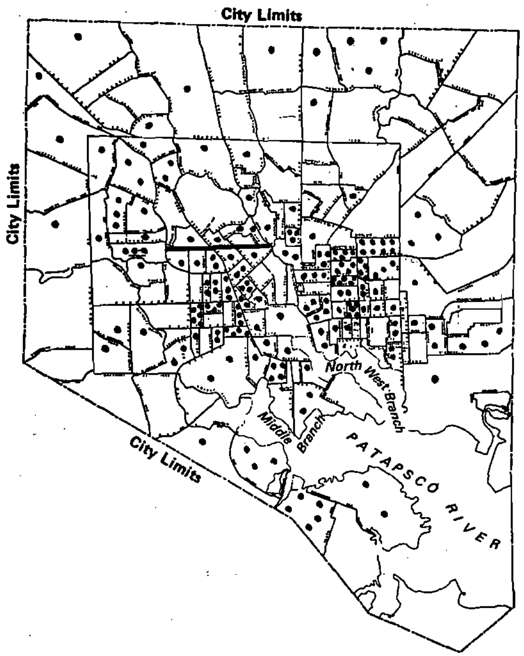


Figure 5-6 Residence distribution of hospitalized rheumatic fever patients, Baltimore 1960–1964. Heavy line shows North Avenue. Each dot represents one patient. *(Reproduced, by permission, from Gordis et al., 1969.)*

the cases occurred in two clusters on either side of the central business district—in the low-income area south of North Avenue, which is shown by a heavy line. Naturally, we cannot form a judgment based only on the "numerator" cases but must relate these to the "denominator" populations in each tract to develop rates.

Thus, annual incidence rates were calculated for groups of census tracts in this area and ranged as high as 40.2 per 100,000. In contrast, the incidence rate for the entire city was only 15.6 per 100,000.

Studying the high and low incidence areas further revealed that nonwhite children suffered a higher incidence of rheumatic fever than did whites. Only among whites was higher socioeconomic

status related to lower incidence of rheumatic fever. When housing characteristics were examined, the degree of crowding was the variable that was most closely related to rheumatic fever occurrence. The authors emphasized the importance of socioeconomic conditions in this disease and showed that the higher incidence in nonwhites might have been due to the crowded living conditions in which most Baltimore nonwhites lived.

TIME

The pattern of disease occurrence in time is often an extremely informative descriptive characteristic. A great variety of time trends may be found in the literature; these involve simple increases or decreases of disease incidence, or more complex combinations of these changes in time. To provide an introduction to this interesting subject, a few examples of short-term, periodic, and long-term trends will be described.

Short-term Increases and Decreases in Disease Incidence

Short-term changes are those increases or decreases in disease incidence that are measured in hours, days, weeks, or months. These are most often observed in the study of epidemics of infectious disease, as will be illustrated below. However, important short-term trends also have been noted in the occurrence of symptoms of, or even deaths due to, chronic noninfectious disease, in relation to both natural phenomena such as heat waves and man-made stresses such as marked increases in air pollution.

Epidemics An *epidemic*, or *outbreak*, is the occurrence of a disease in members of a defined population clearly in excess of the number of cases usually or normally found in *that* population. In investigating epidemics, a careful tabulation of the distribution of disease-onset times of the affected members of the population, in terms either of counts of cases or incidence ("attack") rates, may be very helpful in determining the initiating causes and mechanism of spread.

For a thorough discussion of the propagation of epidemics the reader is referred to Sartwell (1965). A few basic principles should be mentioned, however, before we consider some examples of time patterns. Epidemics only affect susceptible members of the population, of which there may be many or few. Others in the population are immune due to antibodies related to previous disease occurrence, immunization, or passive transfer from mother to infant. Still others may be resistant due to other inherent factors. After a person is exposed to the disease-causing agent, there is an incubation period until the disease first appears. Susceptible persons may also develop inapparent infections, in which no symptoms or signs become evident. The infectious agent may leave the host during the communicable period, which varies in timing and duration from one disease to the next.

Infections are transmitted from one person to another in a variety of ways: by direct personal contact, by touching contaminated objects, or by droplets spraying from one person to another close by, as during talking or sneezing. Evaporation of such droplets may yield "droplet nuclei" which, like certain disease-carrying dusts, may remain airborne for longer periods and travel longer distances. Other modes of transmission include vehicles such as certain foods or water, and *vectors* such as arthropods which carry the infectious agent.

The infection is usually introduced into the population directly or indirectly by one or more persons. If there is a sufficient proportion of susceptibles and the infection spreads rapidly enough, the disease will show a trend of increasing incidence through time to a maximum, followed by a fairly steady diminution until the disease disappears completely or almost completely. The decrease is largely due to the fact that the population begins to run out of susceptible individuals as those who were previously susceptible acquire the disease and become immune. As susceptibles become increasingly scarce, the infectious agent, no matter how well and how rapidly transmitted, finds less and less fertile soil in which to grow, so to speak. The rise through time from a negligible incidence rate to a maximum followed by a fall to low levels again appears graphically as a simple epidemic curve, usually but not always involving short-term trends involving days, weeks, or months.

Fig. 5-7 shows an epidemic of measles (rubeola) that occurred

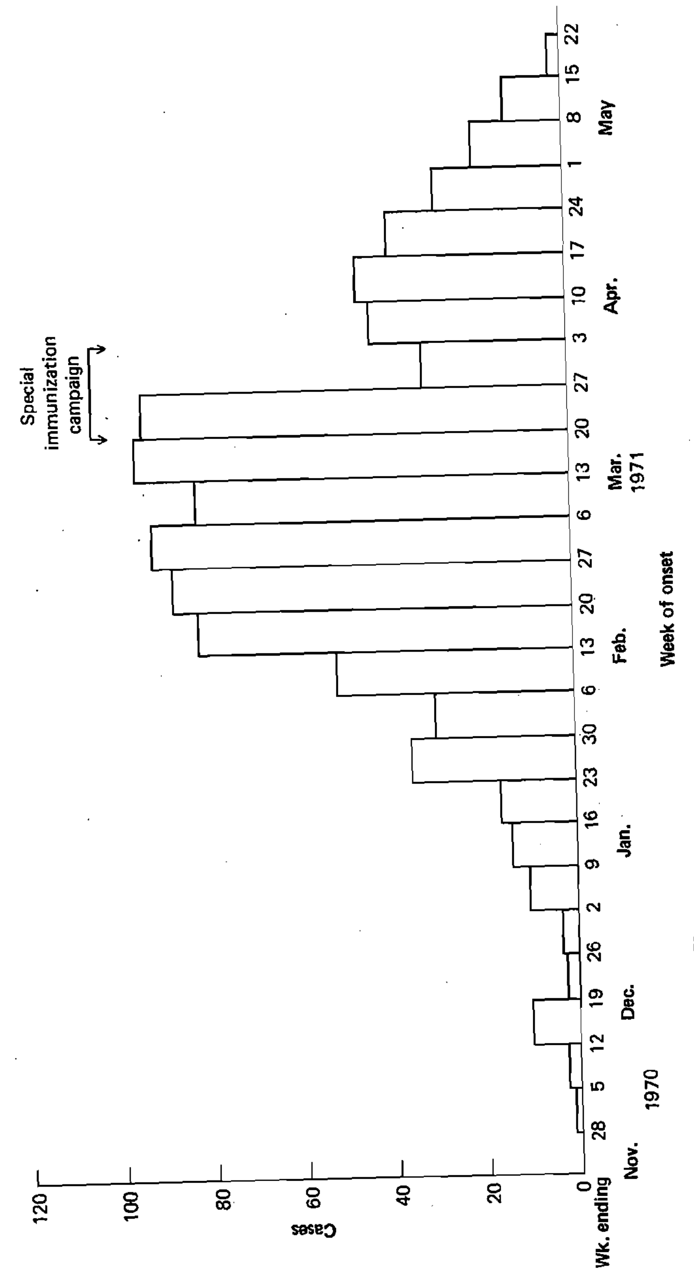


Figure 5-7 Measles cases, by week of onset. Dallas, Texas, December 1, 1970–May 22, 1971. *(Reproduced, by permission, from Luby et al., 1971.)*

among children in Dallas, Texas from late 1970 almost to the middle of 1971 (Luby et al., 1971). By May 1971 there were 1,071 reported cases. The histogram displaying the epidemic's time sequence is of special interest because it shows an abrupt drop after the apparent peak of the epidemic was reached at the end of March. During the 2-week period in which the abrupt fall in case counts began, a special immunization campaign for children was carried out. Although alternative explanations should be considered, it appears that the campaign was helpful in controlling the epidemic by sharply reducing the number of susceptible individuals in the population.

The observed time pattern of an epidemic may provide a strong indication of the mode of initiation and spread. Figs. 5-8 and 5-9 show two different outbreaks of the same disease, infectious hepatitis. Fig. 5-8 depicts the number of cases, by *week* of onset, in an epidemic occurring in Barren County, Kentucky, lasting from June 1970 through April 1971. Fig. 5-9 is drawn on a different time scale and shows the number of cases, by *day* of onset, in an epidemic that occurred in Orange County, California, between August 21, 1971 and September 13, 1971.

The essential distinction between the two epidemics is their duration, particularly in relation to the known incubation period of this disease, which ranges from 15 to 50 days and is commonly about 25 days.

Although the communicable period for this disease has not been clearly defined, the clustering of the Orange County cases within such a narrow time interval—the great bulk appearing within 9 days and all within 24 days—suggested that the outbreak resulted from a "point source," that is, a single common exposure to the virus. With an incubation period measured in weeks, person-to-person spread among the group of cases could not have been a significant factor in an epidemic that ended so soon after it started.

In contrast, the bulk of the Barren County cases occurred over an interval of 4 months, and the total epidemic lasted for 10 months, so there was ample time for direct person-to-person spread. This mode of transmission, however, would not be the only possible mechanism consistent with a hepatitis epidemic of this duration. Prolonged exposure of a population to a contaminated food or water supply could also result in a long-lasting epidemic. However, no such mechanism could be incriminated in Barren County.

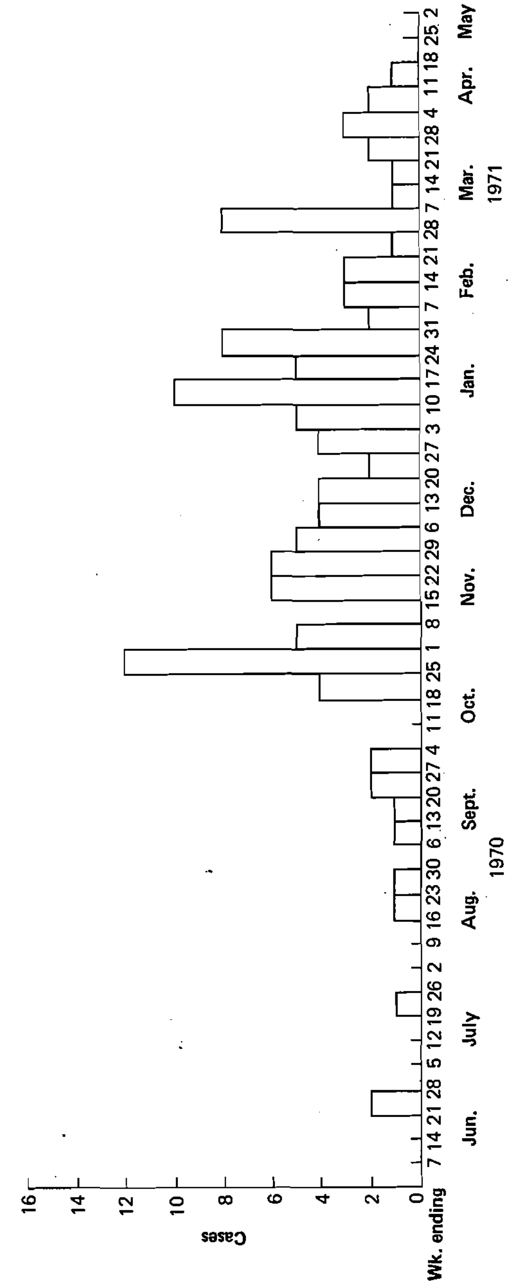


Figure 5-8 Infectious hepatitis cases, by week of onset, Barren County, Kentucky, June, 1970–April, 1971. *(Reproduced, by permission, from Carman et al., 1971.)*

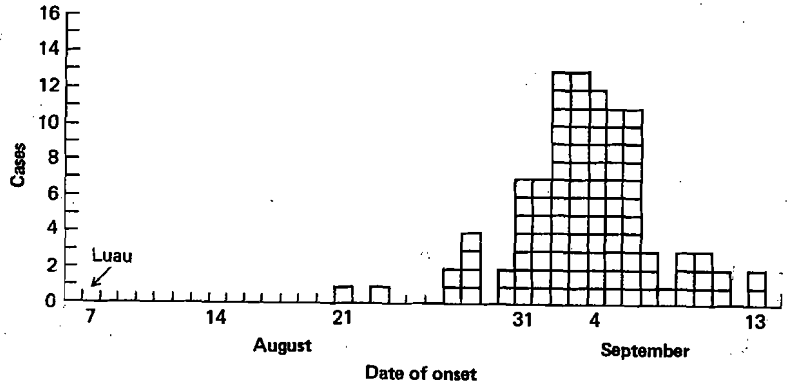


Figure 5-9 Cases of infectious hepatitis in individuals attending luau, by day of onset, Orange County, California, 1971. (Date of onset for one case undetermined.) *(Reproduced, by permission, from Philp et al., 1972).*

Further investigation of each epidemic revealed quite interesting information. In Barren County, most of the 118 cases occurred in children attending two elementary schools. The first observed case in June occurred in a boy whose parents frequently baby-sat for other children in the area. Among the children exposed to his parents was a seven-year-old boy who was the first hepatitis case at one of the schools. The seven-year-old's illness began on September 26. All but one of the cases involving children in that school and their families could be traced directly or indirectly to contact with the seven-year-old. Similarly, at the other school the first illness occurred in a nine-year-old girl on September 21. It was not determined how she became infected but she spread the disease to a total of 34 persons via 5 children with whom she had contact (Carman et al., 1971).

In Orange County almost all the 99 cases were members of a private sailing club. The only time all the cases were together during the prior year was at a club luau at a remote island off the California coast on August 7. One clue to the vehicle of infection was found in the age distribution of incidence. The attack rate was 6 percent in persons under age 15 and was 62 percent in persons age 15 and over.

Detailed analysis of suspect foods revealed only one item regarding which *all* persons *not* consuming it remained free of the disease. This was a mai-tai punch. Sizable proportions of persons *not* eating each of the other foods did become ill, which tends to exonerate these other food items. The description of the food-handling situation at the luau by Philp and his associates (1972) clearly shows how the mai-tai could have served as the vehicle for the fecal-oral transmission of the hepatitis virus.

> All food and beverages were brought to the island by boat. Commercial bottled water for drinking was imported. Little attention, however, was paid to food-handling practices. The only water for handwashing was a single water tap located 85 yards beyond the luau site and 35 yards beyond the two privies available to the group. Handwashing was a rare event. Foods were cooked in an earthen pit for 6 hours to the point of disintegration. Unwashed papaya, pineapple, and oranges were peeled and cut for a fruit mix. Mai-tai punch was prepared from the cut fruits, unwashed fresh strawberries, instant tea, canned pineapple juice, canned lemon-lime drink, bottled water, club soda, and rum. All were mixed in a new plastic garbage can. The punch was served from a 3-tiered fountain made with fiberglass bowls suspended over the garbage can reservoir. Punch was pumped from the can to the top of the fountain, where it flowed over the fiberglass tiers back down into the garbage can. Persons who drank punch filled their cups under one of the streams as it cascaded from one bowl to the next. Many reported punch running over their hands and back into the punch reservoir. Due to the presence of pieces of fruit, particularly strawberries, the pump which forced the punch to the top of the fountain plugged frequently and was unplugged by one or more persons reaching into the bottom of the garbage can and pulling fruit pieces out of the pump. Persons who unclogged the pump also reported sand at the bottom of the can.

Recurrent or Periodic Time Trends

The incidence of certain diseases shows regular recurring increases and decreases. This regular pattern may exhibit cycles which last several years. Many cycles occur annually and represent seasonal variation in disease occurrence. Seasonal variation is a well-known characteristic for many infectious diseases and is usually based on characteristics of the infectious agent itself, the life pattern of its vector or other animal hosts, or changes in the likelihood of person-to-person spread. For example, waterborne gastrointestinal infections often exhibit a peak occurrence in the later summer months when recreational swimming and other factors facilitate their transmission. On the other hand upper respiratory infections frequently show a seasonal rise in or near winter, aided by the concentration of people indoors where virus-containing airborne droplets are readily exchanged.

Shorter-term periodic variations have also been observed. For example, death rates from automobile accidents show weekly cycles with the highest rates occurring on weekends, particularly Saturdays. To date there are no available statistics on the number of passenger-miles driven on each day of the week. Thus it is not possible to state whether the weekend increase in deaths is due merely to an increased exposure of the population to the moving automobile or whether the risk of death per passenger-mile actually increases, possibly due to such factors as more reckless driving or more alcohol consumption on weekends.

Long-term or Secular Trends

Some diseases exhibit a progressive increase or decrease in occurrence that is manifested over years or decades. These long-term time trends are often referred to as *secular* trends.

Figure 5-10 shows the mortality rates in United States males of several leading types of cancer, from 1930 to 1968. A marked secular increase in mortality from lung cancer has occurred, representing about a fifteenfold increase in 1967 as compared to 1930. This increase is believed to be due largely to an increase in the proportion of men who smoked cigarettes. Fig. 5-10 also shows a marked downward trend in stomach cancer mortality. This improvement has not been explained but remains gratifying, nonetheless.

As has been discussed, when a marked increase in incidence occurs in a short period of time, it is quite apparent to the medical community and is referred to as an epidemic, an emotionally loaded

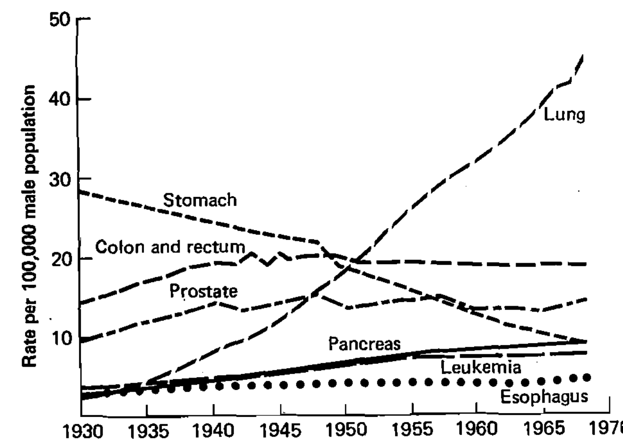


Figure 5-10 Age-adjusted male cancer death rates, by site, United States, 1930–1968. *(Reproduced, by permission, from Silverberg and Holleb, 1972: Data from the U.S. National Vital Statistics Division and Bureau of the Census.)*

term that spurs prompt action. Long-term trends are barely perceptible and might go unnoticed were it not for the study of vital statistics. It would perhaps be well to label as "epidemics," the long-term increases such as that noted for lung cancer. If this term were applied, more action might be taken to investigate the causes and to institute control measures.

REFERENCES

Alter, M. 1968. Etiologic considerations based on the epidemiology of multiple sclerosis. *Amer. J. Epidemiology*, 88:318–332.

Belsky, J. L., A. Kagan, and S. L. Syme. 1971. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii, and California. Research Plan. *Atomic Bomb Casualty Comm. Tech. Rep.*, 12–71.

Carman, M., W. B. Guerrant, and C. Hernandez. 1971. Infectious hepatitis—Kentucky. *Center for Disease Control: Morbidity and Mortality Weekly Report.* 20:136–137.

Frost, W. H. 1939. The age selection of mortality from tuberculosis in successive decades. *Amer. J. Hyg.* 30:91–96.

Gordis, L., A. Lilienfeld, and R. Rodriguez. 1969. Studies in the epidemiology and preventability of rheumatic fever: I. Demographic factors and the incidence of acute attacks, and II. Socio-economic factors and the incidence of acute attacks. *J. Chron. Dis.* 21:645–666.

Gordon, T. 1957. Mortality experience among the Japanese in the United States, Hawaii and Japan. *Public Health Reports* 72:543–553.

Hill, A. B., *Principles of Medical Statistics.* (London: Oxford University Press, 1971), pp. 220–228.

Howard, J., and B. L. Holman. 1970. The effects of race and occupation on hypertension mortality. *Milbank Memorial Fund Quarterly* 48:263–296.

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

Limburg, C. C. 1950. Geographic distribution of multiple sclerosis and its estimated prevalence in the United States. *Proceedings of the Association for Research in Nervous and Mental Diseases,* 28:15–24, Baltimore: Williams and Wilkins.

Luby, J. P., H. J. Dewlett, and M. S. Dickerson. 1971. Measles—Dallas, Texas. *Center for Disease Control: Morbidity and Mortality Weekly Report,* 20:191–192.

Philp, J. R., T. P. Hamilton, T. J. Albert, R. S. Stone, C. F. Pait, R. R. Roberto, N. J. Fiumara, A. R. Hinman, and C. Friedmann. 1972. Hepatitis A Outbreak, Orange County, Calif. *Center for Disease Control: Hepatitis Surveillance,* Report No. 35, pp. 12–13. See also *Amer. J. Epidemiology* 97:50–54, 1973.

Sartwell, P. E., (Ed.): *Maxcy-Rosenow Preventive Medicine and Public Health, Ninth Edition.* (New York: Appleton-Century-Crofts, 1965), Chap. 4.

Segi, M., M. Kurihara, and T. Matsuyama, *Cancer Mortality for Selected Sites in 24 Countries. No. 5 (1964–1965).* Department of Public Health, Tohoku University School of Medicine. Sendai, Japan, 1969.

Silverberg, E., and A. I. Holleb. 1972. Cancer statistics, 1972. *Ca-A Cancer Journal for Clinicians,* 22:2–20.

Taylor, I., and J. Knowelden, *Principles of Epidemiology.* (Boston: Little, Brown, 1964), Chaps. 6, 7 (infectious epidemics); Chap. 12, pp. 318–321 (social class).

World Health Organization: 1967. Mortality statistics: cardiovascular diseases, annual statistics, 1955–1964, by sex and age. Epidemiological Vital Statistics Reports, 20:535–710.

Chapter 6

Prevalence Studies

In going beyond descriptive observations to delve more deeply into disease etiology, there are, as defined in Chap. 4, three basic types of observational investigations:

1 Prevalence or cross-sectional studies
2 Case-control studies
3 Incidence or cohort studies

These will be discussed in greater detail here and in the next two chapters. As will be seen, prevalence studies are, conceptually, quite straightforward, and provide a good basis for subsequent consideration and comparison of the other two study types.

How Prevalence Studies Are Carried Out

Initial Steps The question(s) for study must be clearly defined in terms of the relationship between some possible predisposing factor(s) and the disease under investigation. Then a suitable study population is identified. If this population is small enough to be studied using the human and financial resources available (e.g., students in a school, adults in a small town), the entire population can be included. If the target population is too large (e.g., children in the United States, men in a large city), then a representative sample is selected.

Sampling Methods for selecting an appropriate sample constitute an important and well-developed field of statistical study, and cannot be dealt with comprehensively in this book. The reader should be familiar with a few basic types of samples, since sampling may be necessary in any type of epidemiologic study. For a more complete discussion the reader is referred to Hansen et al. (1953) and Hill (1971).

The most elementary kind of sample is a *simple random sample* in which each person has an equal chance of being selected directly out of the entire population. One way to carry out this procedure is to assign each person a number, starting with 1, 2, 3, and so on. Then, numbers are selected at random, usually from a table of random numbers (see Arkin and Colton, 1963), until the desired sample size is attained.

A *stratified random sample* involves dividing the population into distinct subgroups according to some important characteristic, such as age or socioeconomic status, and selecting a random sample out of each subgroup. If the proportion of the sample drawn from each of the subgroups, or *strata*, is the same as the proportion of the total population comprised by each stratum (e.g., age group 40–59 comprises 20 percent of the population, and 20 percent of the sample comes from this age stratum), then all strata will be fairly represented with regard to numbers of persons in the sample. This proportionality is often desirable and may simplify data analysis. On the other hand, the investigator may have to take a larger proportion of his study sample out of one or a few sparsely populated strata, in order to make available for study adequate numbers of subjects with certain important characteristics.

A *cluster sample* involves (1) dividing the population into subgroups, or *clusters*, that are not necessarily (and preferably not) homogeneous, as are strata, (2) drawing a random sample of the clusters, and (3) selecting all or a random sample of the persons in each cluster. When each cluster comprises persons in a localized geographic area, such as a county, cluster sampling is especially useful for national surveys. It is obvious that many more persons can be studied for the same cost if they live in a few U.S. counties, than if they are scattered all over the country.

Finally, *systematic samples* involve first deciding what fraction of the population is to be studied—for example, one-half or one-tenth—and listing the population in order, perhaps as in a directory or on a series of index cards. Then, starting at the beginning of the list, every second or every tenth (or whatever interval is dictated by the fraction to be chosen) is selected. In order to sample in this manner, the investigator must be quite sure that the intervals do not correspond with any recurring pattern in the population. Consider what would happen if the population were made up of a series of married couples with the husband always listed first. Picking every fourth person would result in a sample of men only, if one started with the first or third subject, or of women only, if one started with the second or fourth.

Sampling can be done in multiple stages, such as sampling within strata which are, in turn, within clusters. In this manner, sampling can become quite involved and require expert assistance in planning. Experience has also revealed subtle problems and biases that might not occur to the novice. Sampling by households is a good example. If there is no one home when the interviewer arrives, he or she should come back again rather than go to the house next door, because households with a person at home in the daytime tend to differ from those without. Similarly, the first house seen as one approaches a new block should not be routinely called upon, since persons in corner houses tend to differ from those in the middle of the block.

Data Collection Once the total study population or sample is defined, the necessary data are collected. Presence of disease may be determined in a variety of ways. For example, in a small town, all or almost all the existing cases of a disease can often be found by contacting all the practicing physicians and reviewing hospital records. Or, the disease can be detected by a special examination of all the residents.

The presence of, or exposure to, the possible causative factors under investigation should also be determined by appropriate tests and measurements. For example, in considering the possible role of inhaled factors, degree of cigarette smoking can be determined by interview, and air pollution levels in various places of living or work can be determined by appropriate measuring devices.

Data Analysis The usual way to tabulate the data in a prevalence study is to subdivide the population according to the suspected predisposing factors being studied and compare the disease prevalence rates in each subgroup. If the relationship of chronic cough to number of cigarettes smoked is to be studied in a group of middle-aged men, then the group may be divided into appropriate smoking categories, such as: none, less than one-half pack per day, one-half pack or more but less than one pack, one pack or more but less than two packs, and two packs or more. The prevalence rate of chronic cough is then determined for each smoking subgroup and the rates in the subgroups are compared. Of course, before the rates are computed, strict criteria must be established for the definition of what constitutes "chronic cough."

Interpretation

In general, the prevalence study will show the presence or absence of a relationship between the study variable(s) and existing disease. *Existing* disease, as contrasted with *developing* disease found in an incidence study, implies a need for caution, since existing cases may not be representative of *all* cases of the disease.

Consider coronary heart disease, for example. One of the important manifestations of coronary heart disease is sudden unexpected death. In a prevalence survey, cases of coronary heart disease showing sudden unexpected death as their first clinical manifestation will be missed because the duration of recognizable disease is so extremely short. It would indeed be remarkable if such a case happened to occur just at the moment the individual was taking the survey examination! From this extreme example it can readily be seen that the shorter the duration of the disease, whether it kills or is cured rapidly, the less chance its victim has of being detected in a one-time prevalence survey. It follows logically, then, that cases of long duration are overrepresented in a prevalence study. The characteristics of these long-duration cases may, on the average, differ in a variety of ways from the characteristics of all cases of the disease being studied.

While we are considering the duration of illness of diseased persons in a prevalence study, it would be worthwhile to digress slightly and point out that there are two basic properties of a disease that are reflected in its prevalence. One is how much disease develops per unit of time, or incidence; the other is how long it lasts, or duration. In fact, under stable conditions, where the incidence and duration of a disease have remained constant over a period of time, the relationship between prevalence, incidence, and duration can be expressed as a simple mathematical equation: Prevalence equals incidence times mean duration ($P = I\bar{d}$). Thus, if any one of the three measures is unknown, it can be computed from the other two, provided that conditions are stable, as mentioned.

Another factor leads to "prevalence cases" being an unrepresentative sample of all cases; that is, if certain types of cases leave the community. Some affected persons may be institutionalized elsewhere or move to another city where there are special facilities for treatment, thus escaping local surveillance procedures.

When interpreting the findings of a prevalence study, care must be taken to avoid assigning an unsubstantiated time sequence to an association between a trait or other factor and the disease. If it is found, for example, that cancer patients exhibit more anxiety or other emotional problems than the unaffected members of the population, it cannot be assumed that the anxiety preceded the cancer. After all, cancer patients may have good reason to be nervous or disturbed. In contrast, there would be no doubt about the cancer being preceded by such traits as eye color, blood type, or maternal exposure to radiation.

Example I: Prevalence Studies of Chronic Respiratory Disease in Berlin, New Hampshire

In 1961, Ferris and Anderson (1962) carried out a prevalence study of chronic respiratory disease in relation to cigarette smoking and air pollution in Berlin, New Hampshire. This industrial town with almost

18,000 inhabitants is located in a valley near the Canadian border and is almost completely surrounded by mountains. The major industry and chief source of air pollution is a paper and pulp mill.

In this study, the investigators planned to diagnose three disease states—chronic bronchitis, bronchial asthma, and irreversible obstructive lung disease—using simple pulmonary function tests and a standardized interview questionnaire about respiratory symptoms. These standardized methods for assessing pulmonary disease, developed and tested in Great Britain and already used in several studies, would permit a comparison of the findings in Berlin, New Hampshire to those in British and other population groups. At that time there was great interest in the apparent disparity in the relative frequency of chronic bronchitis in Great Britain and the United States, and it was believed that differences in diagnostic criteria and fashions might have been at least partially responsible.

The investigators, in cooperation with the local health department, selected the study sample in two stages. First, using the town's tax roll book which listed the adults in alphabetical order, they randomly selected 36 pages (clusters). Second, from the 36 pages they systematically selected every second name of those in the 25–54-year age stratum and all names of persons aged 55–69. Persons aged 70 and over were listed separately in the town records, and a sample of this age stratum was randomly selected.

Before any data were collected, the local physicians and the state Health Department were contacted and the study was publicized by newspaper and radio. The study subjects were invited by letter to take the study examination at a clinic in the Health Department. Failure to respond led to a telephoned invitation, and if this, in turn, failed, the subject was visited at home and, if he agreed, the interview and physiologic testing were carried out there. Through these persistent efforts, over 95 percent of the 1,261 selected subjects were examined, with the only nonparticipants being those who were away from home during the survey and a few who refused.

Respiratory symptoms were elicited by the standardized interview. Smoking habits, occupational exposures, and previous chest illnesses were also assessed in the interview. Forced expiratory volume, both total (FEV) and during the first second ($FEV_{1.0}$), and peak flow were measured with a recording vitalometer.

The presence of disease was defined by strict criteria. For example, the diagnosis of chronic bronchitis required "the report of bringing up phlegm from the chest six times a day on four days a week for three months in a year, for the past three years or more."

Data analysis revealed a greater prevalence of respiratory disease in men than in women. Furthermore, there was a clear relationship of respiratory disease to smoking. For example, in men the age-adjusted prevalence of chronic bronchitis was:

15.0% in those who had never smoked
18.9% in exsmokers
29.8% in smokers of 1–10 cigarettes per day
34.2% in smokers of 11–20 cigarettes per day
42.3% in smokers of 21–30 cigarettes per day
61.1% in smokers of 31–40 cigarettes per day
75.3% in smokers of 41 or more cigarettes per day

The town was divided into three areas with low, intermediate, and high degrees of air pollution, according to independent measurements of air quality. Residence of study subjects in these three areas showed only an equivocal relationship to respiratory disease. However, if only nonsmokers were considered, it appeared that among men, chronic bronchitis was more apt to be found in residential areas having greater air pollution.

The planned United States–British comparison was later reported by Reid et al. (1964). The findings in Berlin, New Hampshire were compared with those derived from a random sample of urban and rural dwelling adults in Britain examined in a comparable fashion. It was found that the British exceeded the Americans very little in the prevalence of simple chronic bronchitis, characterized by chronic cough and sputum production. However, bronchitis complicated with shortness of breath and repeated acute illnesses was more prevalent in Britain, particularly in urban men.

The prevalence survey in Berlin, New Hampshire was repeated in 1967 using comparable methods (Ferris et al., 1971). It was noted that the prevalence of respiratory disease symptoms was lower in 1967 and that, on the average, there was some improvement in pulmonary function. Because there had also been a fall in air pollution between 1961 and 1967, the authors concluded that this was the probable explanation for the observed improvement. In their analysis they were careful to consider other possible explanations for the change, such as observer differences and the increasing use of filter-tip cigarettes.

The second survey in 1967 illustrates the usefulness of repeated prevalence studies in assessing time trends in disease or other population characteristics, provided that comparable measurement methods are used. The effort and expense of keeping a population under continuous long-term surveillance can often be avoided by conducting careful cross-sectional studies at fairly wide intervals.

Example 2: Cardiovascular Disease in Evans County, Georgia

In Chaps. 4 and 5, emphasis has been placed on descriptive epidemiologic findings as a source of hypotheses for further analytic studies. Another very important source of ideas and hypotheses has been clinical observations by astute and concerned health-care professionals. A physician's observations and interest proved to be a major stimulus for the epidemiologic study of cardiovascular disease in Evans County, Georgia, which began in 1960 as a prevalence study (Hames, 1971, Cassel, 1971*a* and *b*, McDonough et al., 1965).

Dr. Curtis Hames, who practiced in the area, was impressed with the difference in frequency with which he found coronary heart disease occurring in whites and blacks. Although coronary heart disease was a common problem in his white patients, he rarely saw it in blacks, despite the fact that many black patients had hypertension and appeared to consume a high animal-fat diet. Furthermore, the male-female difference in susceptibility to coronary heart disease which was so obvious in whites was not apparent in blacks.

In order to confirm and explain these differences in a systematic fashion, Hames encouraged the interest and participation of epidemiologists and other investigators. Largely due to his excellent rapport with the community, there was nearly complete participation of the selected study subjects.

Evans County is located on flat or slightly rolling terrain about 60 miles inland from the coastal port of Savannah, Georgia; its greatest diameter is 19 miles. The county's economy was, in 1960,

primarily agricultural, although its extensive pine forests were a source of lumber, pulpwood, and turpentine. In 1960, the total population was 6,952, of which 66.5% were white and 33.5% were black.

The study sample consisted of a 50 percent random selection of county residents, aged 15 through 39, and all residents aged 40 and over. Of the 3,377 persons chosen, 92 percent underwent the study examination, which consisted of a medical and dietary history, physical examination, urinalysis, serum-cholesterol measurement, electrocardiogram, and chest x-ray. The social class standing of each subject was determined according to the occupation, education, and source of income of the head of the household.

The diagnosis of coronary heart disease required that a subject have either a history of angina pectoris, a history of myocardial infarction, or electrocardiographic evidence of myocardial infarction. Each of these manifestations was defined as definite, probable, possible, or absent according to standard criteria. It is essential for investigators to establish, adhere to, and describe in study reports, strict criteria for the diagnosis of disease so that others may know just what kinds of cases were included or excluded. Strict criteria also permit other investigators to reproduce the study findings, or at least to understand why their own study results might differ.

The major findings of the Evans County prevalence study included confirmation of the initial clinical observations. Coronary heart disease prevalence was indeed lower in blacks than in whites, the difference occurring only in men. Part of this black-white difference could be explained by social class, since white men of lower socioeconomic status had coronary heart disease prevalence rates approaching the low levels in blacks, almost all of whom were in the lower social bracket. The age-adjusted prevalence rates were:

97 per thousand in high-social-class whites
40 per thousand in low-social-class whites
21 per thousand in blacks

The investigators could not explain these racial and social class differences by taking into account differences in other risk factors, including blood pressure, serum-cholesterol levels, body weight, cigarette smoking, and dietary intake. However, it was noted that habitual physical activity, as estimated by type of occupation, was inversely related to coronary heart disease prevalence. Men in occupations involving the most physical exertion (e.g., manual labor, sharecropping) showed the lowest prevalence of coronary heart disease. Since these occupations were primarily engaged in by blacks and low-social-class whites, it appeared that physical activity might explain their relatively low prevalence of coronary heart disease.

As with the Berlin, New Hampshire study, described above, a second examination procedure was carried out several years later, beginning in 1967. However, this was *not* for the purpose of repeating the prevalence study. Rather, the second round of examinations was applied only to the initially examined cohort as part of the follow-up for an incidence study. An initial prevalence survey can be, and often is, used as the first stage of an incidence study, in that it defines and characterizes a *population at risk*—those initially free of the disease being studied. As will be described in Chap. 8, this population at risk can then be followed up for the development of the disease.

The incidence study confirmed the black-white difference in the occurrence of coronary heart disease, but the social class difference in whites was no longer evident. It appeared that this was due to a rapid catching up of the lower-class men to the upper-class men in coronary heart disease risk, during a period when Evans County was changing from an agrarian to an industrial economy. The only subgroup of white men with the same low risk as the blacks were sharecroppers, again suggesting a protective effect of high levels of physical activity.

REFERENCES

Arkin, H., and R. R. Colton, *Tables for Statisticians, 2d ed.* (New York: Barnes and Noble, 1963), pp. 26–27, 158–161.

Cassel, J. C. 1971*a.* Summary of major findings of the Evans County cardiovascular studies. *Arch. Intern. Med.*, **128**:887–889.

Cassel, J. C.: 1971*b.* Review of the 1960 through 1962 cardiovascular disease prevalence study. *Arch. Intern. Med.*, **128**:890–895.

Ferris, B. G., Jr., and D. O. Anderson. 1962. The prevalence of chronic respiratory disease in a New Hampshire town. *Am. Rev. Resp. Dis.*, **86**:165–177.

Ferris, B. G., Jr., I. T. T. Higgins, M. W. Higgins, J. M. Peters, W. F. van Ganse, and M. D. Goldman. 1971. Chronic nonspecific respiratory disease, Berlin, New Hampshire, 1961–1967: a cross-sectional study. *Am. Rev. Resp. Dis.*, **104**:232–244.

Hames, C. G. 1971. Evans County cardiovascular and cerebrovascular epidemiologic study: Introduction. *Arch. Intern. Med.*, **128**:883–886.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow, *Sampling Survey Methods and Theory.* vol. I., Methods and Application. (New York:Wiley 1953), Chaps. 1–3.

Hill, A. B., *Principles of Medical Statistics.* (London: Oxford University Press, 1971), Chaps. 2, 3.

McDonough, J. R., C. G. Hames, S. C. Stulb, and G. E. Garrison. 1965. Coronary heart disease among Negroes and whites in Evans County, Georgia. *J. Chron. Dis.*, **18**:443–468.

Reid, D. D., D. O. Anderson, B. G. Ferris, Jr., and C. M. Fletcher. 1964. An Anglo-American comparison of the prevalence of bronchitis. *Brit. Med. J.*, **2**:1487–1491.