PRIMER OF EPIDEMIOLOGY

Gary D. Friedman
M.D., S.M. in Hyg., F.A.C.P.

Senior Epidemiologist, Department of Medical Methods Research, Kaiser-Permanente Medical Care Program

Assistant Clinical Professor of Medicine and of Ambulatory and Community Medicine, University of California School of Medicine, San Francisco

Lecturer in Epidemiology, University of California School of Public Health, Berkeley

McGRAW-HILL BOOK COMPANY

A Blakiston Publication

New York St. Louis San Francisco Düsseldorf Johannesburg Kuala Lumpur London Mexico Montreal New Delhi Panama Paris São Paulo Singapore Sydney Tokyo Toronto

To Ruth,
Emily,
Justin,
and Richard

PRIMER OF EPIDEMIOLOGY

90MUMU79

This book was set in Helvetica by Black Dot, Inc. The editors were Paul K. Schneider and Shelly Levine Langman; the designer was Joseph Gillians; and the production supervisor was Thomas J. LoPinto. The drawings were done by Eric G. Hieber Associates Inc.
The Murray Printing Company was printer and binder.

Library of Congress Cataloging in Publication Data

Friedman, Gary D date
Primer of epidemiology.

"A Blakiston publication."
1. Epidemiology. I. Title. [DNLM: 1. Epidemiology. WA100 F911p 1974]
RA651.F68 614.4 73-20037
ISBN 0-07-022425-0

CONTENTS

PREFACE

It has seemed to me that many health-care professionals do not have an adequate understanding or appreciation of what epidemiology is all about or how it relates to their own work. Furthermore, one frequently finds a failure in communication between the clinician and the epidemiologist despite their common concern over human health and disease. I believe it is fair to say that most students of medicine and other health sciences regard epidemiology as a boring and irrelevant subject which they study only because they are required to. Another common view of epidemiology among health-care professionals is that it is highly esoteric or mathematical and too complex for them to understand.

With those problems in mind I have attempted to write a concise textbook for physicians, medical students, and other health-care professionals that would explain epidemiologic concepts clearly and simply. I have also tried to bridge the gap in communication between the clinician and epidemiologist in a variety of ways, such as providing a number of clinical examples throughout the book,

explaining to the clinician why the epidemiologic emphasis on the study of groups rather than individuals is necessary, and trying to show the relevance of epidemiology to the major concerns of the clinician such as diagnosis and choice of therapy. Also, I have described several interesting epidemiologic studies to illustrate various methods of investigation. Rather than showing just tables of data to illustrate the results of these studies, I have tried to describe them in sufficient detail so the reader will come away with a real feeling for what it is like to carry out an epidemiologic study. I have attempted, also, to provide some much sought-after practical advice on how to conduct a simple epidemiologic or clinical study and on critical reading of the medical literature. Finally, there is some discussion of epidemiology in relation to the study of problems currently of great social and political importance—the changing health care system and environmental hazards.

Some epidemiologists may be disappointed at the lack of discussion of some of the epidemiologic classics such as Snow's studies of cholera or Goldberger's studies of pellagra. Despite the importance and beauty of these studies, I believe that most students are much more interested in examples that relate to current health and social problems.

Few, if any, of the ideas and concepts in this book are original. I am deeply indebted to those who trained me in epidemiology and related subjects and to the many colleagues and friends with whom I have worked over the past decade for all I have learned from them. A number of the examples, references, and other materials that appear here were suggested to me by colleagues, to whom I am most grateful. It would be impossible for me to name all who, in one way or another, helped me to write this book, but I hope they are aware of my appreciation.

I would like to single out for special thanks Dr. Loring G. Dales, Dr. Mark J. Yanover, and my wife, Ruth, who read the entire manuscript carefully during its preparation and made many valuable suggestions. I am grateful to Mrs. Agnes M. Lewis for carefully typing the manuscript and drawing some of the figures, and to Dr. Morris F. Collen for his advice and encouragement.

Gary D. Friedman

Chapter 1

Introduction to Epidemiology

EPIDEMIOLOGY: DEFINITION, PURPOSE, AND RELATION TO PATIENT CARE

Epidemiology is the study of disease occurrence in human populations. The primary units of concern are *groups* of persons, not separate individuals. Thinking in epidemiologic terms often seems foreign to clinicians and other health-care professionals, who are trained to think of the unique problems of each particular patient.

Whether one focuses on individuals or groups should depend upon what one is trying to accomplish. In caring for a sick patient, the need to individualize the diagnosis and treatment for that unique patient is obvious. However, groups of persons must be studied in order to answer certain important questions. These questions often relate to the etiology and prevention of disease and to the allocation of effort and resources in health-care facilities and in communities.

Some examples of questions that require epidemiologic study of human populations are:

When can we expect the next influenza epidemic?
Why are we seeing so much coronary heart disease these days?
How can cancer of the uterine cervix best be prevented?
How often should healthy patients be given medical checkups and what examinations and tests should these checkups include?

Although they also focus on groups, clinical studies of the natural course of disease or the effects of treatments should be distinguished from epidemiologic studies. In general, epidemiologists are more concerned with disease patterns in natural populations such as communities or nations. Clinical studies, on the other hand, are concerned with groups of *patients* seen in a medical facility. However, the methods of investigation are often quite similar, so that training and experience in epidemiology are useful for the clinical investigator.

In addition to being related to clinical research, epidemiology is intimately involved in clinical practice. Clinicians regularly use epidemiologic knowledge in the diagnosis and treatment of disease. Accordingly, after the elements of epidemiology are presented in subsequent chapters, the relationship of epidemiology to clinical research and to medical care will be described.

How Epidemiology Contributes to Understanding Disease Etiology

Each scientific discipline in medicine is uniquely able to answer certain questions. If our goal is to understand how a particular disease occurs, each discipline can attack the problem at its own level and contribute to our understanding.

It is sometimes implied that the purpose of epidemiology is to provide clues to etiology which can later assist the laboratory scientist in arriving at the real answer. This is a distorted view. There are certain questions that can only be answered outside of the laboratory.

A new vaccine may be developed and prepared by biologists

and biochemists, but epidemiologists will have to answer whether the vaccine is successful in preventing disease.

Similarly, laboratory scientists can identify carcinogenic compounds in tobacco smoke and may even be able to produce lung cancer in experimental animals by forcing them to smoke cigarettes. However, the idea that cigarette smoking causes human lung cancer would be unconvincing unless epidemiologists also showed that lung cancer occurred more often in cigarette smokers than in nonsmokers.

Causation of Disease A moment's thought about any disease reveals that more than one factor contributes to its occurrence. For example, tuberculosis is not merely caused by the tubercle bacillus. Not everyone exposed to the tubercle bacillus becomes ill with tuberculosis. Other factors have been identified which clearly contribute to the occurrence of this disease. These factors include poverty, overcrowding, malnutrition, and alcoholism. Amelioration of these other factors can do much to prevent this disease.

Epidemiologists have organized the complex multifactorial process that leads to disease in various ways. One useful way to view the causation of some diseases, particularly certain infectious diseases, is in tripartite terms of the agent, the environment, and the host. For acute rheumatic fever the agent is the beta-hemolytic streptococcus. However, not all persons infected with this organism develop the disease. Thus, considerations of host susceptibility are important. Constitutional factors appear to play a role not only in whether or not the disease develops but also in the localization of cardiac damage. Important environmental factors include social conditions such as poverty and crowding as well as nonhuman aspects of the environment such as season, climate, and altitude.

Another epidemiologic view of disease etiology is as a "web of causation." This concept of disease causation considers all the predisposing factors of any type and their complex relations with each other and with the disease. One current view of the multiple factors leading to myocardial infarction well illustrates a causal web (Fig. 1-1). (Despite the apparent complexity of this diagram, it is undoubtedly an oversimplification and will certainly be modified by further study.) Note that many interrelated factors ultimately lead to

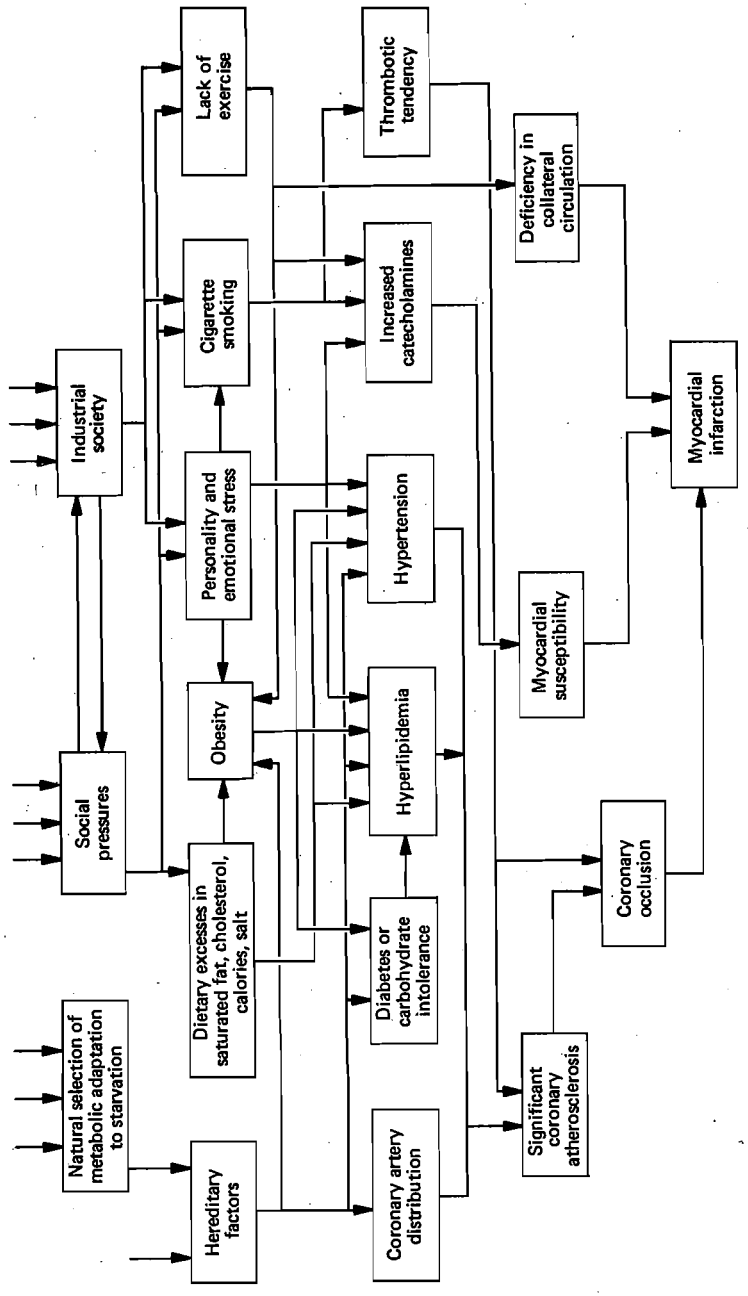


Figure 1-1 The web of causation for myocardial infarction: a current view.

myocardial infarction. Each of these factors mentioned is also influenced by a variety of other factors not shown, leading to as complex a causal web as one chooses to construct. Nevertheless, based on the information presented, it can be seen that a variety of actions could be taken which might reduce the occurrence of myocardial infarction. These actions include dietary modifications, treatment of hypertension, and changing public attitudes toward smoking and exercise.

It is tempting to search for a primary cause, or the most important or most direct of the many causal factors. The benefits of this search are perhaps more philosophical or psychological than practical. In terms of disease prevention it may be most practical to attack a causal web at a spot that seems relatively remote from the disease. To prevent malaria, we do not merely try to destroy the malaria parasite; rather, we drain swamps to control the mosquito population, since this is a practical and effective approach. Similarly, economic development and general improvements in living conditions seem to have done more to reduce mortality from tuberculosis than any chemotherapeutic agent directed specifically at the tubercle bacillus.

Definition and Classification of Diseases

No discussion of disease causation would be complete without some comment about the relatively arbitrary and varying ways in which diseases are defined.

What physicians are faced with are ill persons! However, it has been convenient and valuable to divide the ill persons into categories and give each category a name. We call each category a disease. Ill people do not always fit well into our categories, as any physician will discover if he tries to practice medicine using only the textbooks.

We name diseases to reflect something about our perception or understanding of what the disease entails. Some disease names are merely descriptive of some aspect such as appearance (e.g., erythema multiforme) or subjective sensation (e.g., headache). Some names probe a bit deeper but are still descriptive of pathologic anatomy, often as defined by gross or microscopic appearance (e.g.,

adenocarcinoma of the colon or fracture of the femur). On the other hand, the disease name may focus on some real or supposed causative factor; e.g., pneumococcal pneumonia implies a pulmonary infection by the pneumococcus.

As knowledge about disease causation increases, the disease names are often switched from descriptive terms to terms implying a causal factor. Many ill persons who had been formerly named by a variety of descriptive terms become reclassified under a single causal heading. Similarly, a single descriptive heading may have contained patients with a variety of causally defined diseases. One of the former names for the condition we now call tuberculosis was *phthisis*, meaning "wasting away." Patients in whom wasting dominates the clinical picture constitute only a portion of persons with tuberculosis, and tuberculosis is only one of the causes of wasting.

Causal names for disease are useful in that they immediately imply means for prevention or therapy; in fact, they can drastically change the manner in which a particular health problem is handled. However, causal names can also lead to problems. When the focus on one causal factor such as an infectious agent is reflected in the disease name, we often forget that other factors are operating and tend to regard the infectious or other agent as the only cause.

In summary, disease names are important tools for thought and communication. However they must be viewed in proper perspective. They tend to mask differences among patients, and they have a way of influencing and narrowing our thinking. Disease names may even become "the thing itself," whereas the emphasis should be on the ill person. Furthermore, disease names are transitory. The naming and classifying of ill persons has changed markedly through history and will continue to change.

REFERENCE

MacMahon, B., and T. F. Pugh, *Epidemiology: Principles and Methods.* (Boston: Little, Brown, 1970), Chaps. 1, 2, and 4.

Chapter 2

Basic Measurements in Epidemiology

There is one thing I would be glad to ask you. When a mathematician engaged in investigating physical actions and results has arrived at his conclusions, may they not be expressed in common language as fully, clearly, and definitely as in mathematical formulae? If so, would it not be a great boon to such as I to express them so?

Michael Faraday,
Letter to James Clerk Maxwell

Epidemiology is a quantitative science. Its measured quantities and descriptive terms are used to describe *groups* of persons.

Counts The simplest and most frequently performed quantitative measurement in epidemiology is a count of the number of persons in the group studied who have a particular disease or a particular characteristic. For example, it may be noted that 10 people

in a college dormitory developed infectious hepatitis or that 16 stomach cancer patients were foreign-born.

Proportions and Rates

In order for a count to be descriptive of a group it must be seen in proportion to it; i.e., it must be divided by the total number in the group. The 10 hepatitis cases would have quite a different significance for the dormitory if the dormitory housed 500 students than if it housed only 20. In the first case the proportion would be $^{10}/_{500}$, or 0.02, or 2 percent. (Percentage, or number per one hundred, is one of the most common ways of expressing proportions. Number per 1,000 or 1 million, or any other convenient base may be used.) In the second case the proportion would be $^{10}/_{20}$ or 0.50.

The use of denominators to convert counts into proportions seems almost too simple to mention. However, a proportion is one basic way to describe a group. *One of the central concerns of epidemiology is to find and enumerate appropriate denominators in order to describe and to compare groups in a meaningful and useful way.*

Certain kinds of proportions are used very frequently in epidemiology. These are referred to as *rates*. The various types of rates involve or imply some time relationship. The two most commonly used rates which every physician should understand and remember are the prevalence rate and the incidence rate:

Prevalence Rate

$$\text{Prevalence rate} = \frac{\text{number of persons with a disease}}{\text{total number in group}}$$

Prevalence describes a group at a certain point in time. It is like a snapshot of an existing situation. For example, *the prevalence of electrocardiographic abnormalities at our screening examination was 5 percent*; or, *the prevalence of diarrhea in the children's camp on July 13 was 33 percent.* Or, *the prevalence of significant hyperbilirubinemia in full-term infants on the third postpartum day is 20 percent.* As can be seen by the above examples the point in time is

not necessarily a true geometric point with no length, but is a relatively short time such as a day. Nor does the point have to be in calendar time. It can refer to an event which may happen to different persons at different times, such as an examination or the third postpartum day.

Incidence Rate

$$\text{Incidence rate} = \frac{\text{number of persons developing a disease}}{\text{total number at risk}} \text{ per unit of time}$$

Incidence describes the rate of development of a disease in a group over a period of time, which is included in the denominator. In contrast to a snapshot, incidence describes a continuing process over a given time period. For example, *the incidence of myocardial infarction is about 1 percent per year in men aged 55–59 in our community*; or, *at the height of the epidemic the incidence of chicken pox in the first grade children was 10 percent per day.*

Not everyone in a study population may be at risk for developing a disease. For example, some diseases are lifelong in duration, so that once you have it you cannot develop it again. Persons with such a disease are usually removed from the denominator population at risk.

In the medical literature the word "incidence" is often used to describe prevalence or simple proportion. For example, *the incidence of gallstones is 20 percent in middle-aged women*; or, *in our autopsy series the incidence of liver cirrhosis was 12 percent.* This imprecise use of "incidence" should be avoided, since the specific concept of incidence, defined as a rate of development, is a useful one.

Other Rates Some other rates, often used in epidemiology, are described below.

$$\text{Period prevalence} = \frac{\text{number of persons with a disease during a period of time}}{\text{total number in group}}$$

Sometimes one wishes to have a measure of all the diseases affecting a group during a period of time such as the year, 1970, rather than at a point in time. The period prevalence of a disease in 1970 turns out to be the prevalence at the beginning of 1970 plus the annual incidence during 1970.

$$\text{Mortality, or death, rate} = \frac{\text{number of persons dying (due to a particular cause or due to all causes)}}{\text{total number in group}} \text{ per unit of time}$$

Mortality rate is analogous to incidence but refers to the process of dying rather than the process of becoming ill.

Any rate may refer to a subgroup of a population. For example:

$$\text{Age-specific mortality rate} = \frac{\text{number of persons dying in a particular age group}}{\text{total number in the same age group}} \text{ per unit of time}$$

$$\text{Case fatality rate} = \frac{\text{number of persons dying due to a particular disease}}{\text{total number with the disease}}$$

Case fatality rate refers to the proportion of persons with a particular disease who die. The time element is usually not specified but may be, if desired, as with incidence.

A variety of other disease rates are described by Siegel (1967). In most rates the numerator must include only persons who are derived from the denominator population. The denominator is considered the total population at risk of being or becoming one of the numerator. Thus, these rates can be viewed as a statement of probability that a condition exists (prevalence) or will develop (incidence) in the population at risk.

Some rates depart somewhat from the ideal of having the numerator being derived from the denominator population at risk. This is done for convenience, because of the ready availability of

data that approximate the ideal. Consider the

$$\text{Maternal mortality rate} = \frac{\text{number of deaths from puerperal causes during a year}}{\text{number of live births during the same year}}$$

Actually, the true population of mothers at risk for puerperal death includes those that have had stillbirths as well as those that have had live births. Legally required registration and counting of live births makes this live-birth denominator much more accessible.

Handling Changing Denominators If a denominator population is growing or shrinking during the period of time for which a rate is to be computed, then it is customary to use the population size at the *midpoint* of the time interval as an estimate of the average population at risk. If an incidence rate is to be computed for the year 1973, then the population at risk as of July 1, 1973 is used for the denominator.

Comparison of Rates, Using Differences or Ratios

Differences It is often desired to compare a rate in one group with that in another. One may simply note both rates and observe that one is larger than the other. By subtracting the smaller from the larger, one may obtain the magnitude of the difference.

The difference between two incidence rates is sometimes called "attributable risk" if the two groups being compared differ in some other aspect that is believed to play a causal role in the disease. For example, in Hammond's (1966) study of smoking and mortality the lung cancer mortality rate in nonsmokers ages 55–69 was 19 per 100,000 persons per year as compared to 188 per 100,000 in cigarette smokers. The difference between the two lung cancer mortality rates was 169 per 100,000 per year. This is the lung cancer risk attributable to smoking, *if* smoking is the only important difference between the groups in factors affecting the development of lung cancer. Only the *excess* rate in smokers should be attributed

to smoking—not the entire smokers' incidence rate—since nonsmokers develop some lung cancer, too.

Ratios Another way to compare two rates is by determining the ratio of one to the other, that is, dividing one by the other. In the smoking and lung cancer example, the ratio of the rate in smokers to that in nonsmokers was $^{188}/_{19}$ or 9.9. The smokers had a 9.9 times greater risk of dying from lung cancer than did the nonsmokers. The ratio of two rates is sometimes called the "relative risk," "risk ratio," "morbidity ratio," or, if mortality rates are under consideration, the "mortality ratio."

Ratio Comparisons of Several Groups to a Single Standard When one wishes to compare several different rates, it is often convenient to determine the ratio of all the different rates to a single standard. The standard of comparison may be an actual rate for a particular group that seems appropriate to use. In the study of smoking and lung cancer, smokers were divided according to the number of cigarettes currently smoked per day. Nonsmokers were again used as the standard of comparison, and their mortality rate was arbitrarily designated as 1.0. In comparison, the ratios for male smokers, ages 55–69, were 3.5 for smokers of 1 to 9 cigarettes per day, 8.8 for smokers of 10 to 19 cigarettes per day, 13.8 for smokers of 20 to 39 cigarettes per day, and 17.5 for smokers of 40 or more cigarettes per day.

It may be that the group to be used as a standard differs from the other groups in some important respect, resulting in a biased or unfair comparison. For example, suppose that the men in the different smoking categories not only had different smoking habits but were, on the average, of substantially different ages as well. Then it would not be fair to compare their lung cancer incidence as if differences in smoking were all that mattered, since we know that age is also important—the older one gets the higher the likelihood is of developing lung cancer. In order to eliminate this bias we have to determine as a standard of comparison an *expected rate* instead of an actual rate. To do this, we might calculate, for example, what lung cancer incidence rate would be expected in nonsmokers, as before, but now assuming that they were of the same age composition as

that of each group of smokers. The method for computing this expected rate involves what is called *age adjustment*, or *age standardization*. This will be discussed further in Chapter 11.

An example of a morbidity ratio comparison using an expected rate is shown in Fig. 2-1. In the Framingham Heart Study men and women in five different blood pressure level groups were compared with one another with respect to the subsequent incidence of coronary heart disease during 8 years. Morbidity ratios were used with an expected rate as a standard of comparison, set at 100 percent. The expected rate was that observed in the whole population, but age-adjusted so that it could be applied fairly to the particular blood pressure group under consideration.

In the figure it can be seen, for example, that for those persons with the lowest systolic blood pressure levels, less than 120 mm Hg, the observed incidence was $^{10}/_{629}$. The expected incidence, based on the experience of the whole population, was $^{30.7}/_{629}$. The ratio of these rates is $^{10}/_{30.7}$, or 33 percent. In contrast to the low incidence in the low blood pressure group, the incidence in the highest group, those with a systolic blood pressure of at least 180 mm Hg, was 223 percent of the expected incidence.

Quantitative Attributes

In considering counts, proportions, and rates we have been dealing with qualitative differences between people—presence or absence of disease, or possession of one versus another attribute. Other characteristics of groups that must be considered lie on a quantitative scale. These characteristics include such measures as height, weight, blood pressure, antibody titer, and diameter of tuberculin skin-test reaction. Epidemiology requires appropriate measures so that groups can be described and compared with respect to these quantitative attributes.

In discussing such measures, one must mention some concepts that are usually presented in books or courses on statistics or biostatistics (see Ipsen and Feigl, 1970). In this introduction to epidemiology it is not necessary to present statistical aspects in great detail, but certain basic measures do deserve mention. Parenthetically, it might be well to remark that one need not be highly

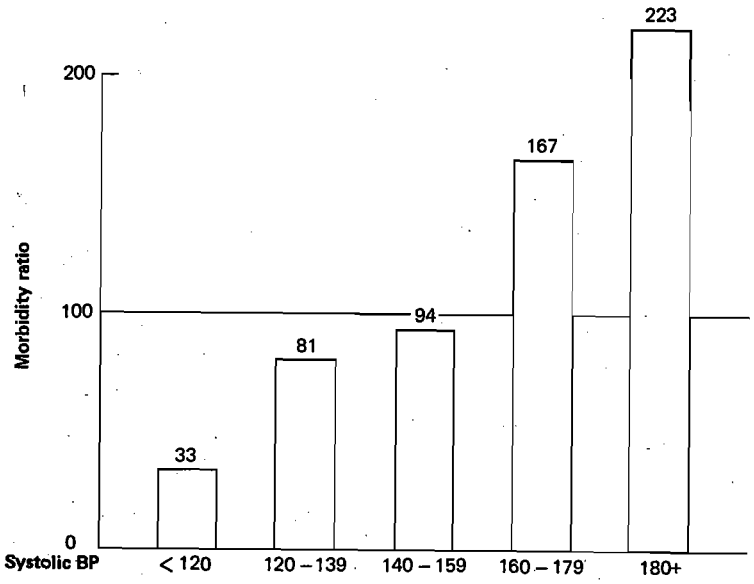


Figure 2-1 Risk of developing coronary heart disease (CHD) in 8 years according to initial systolic blood pressure level. Men and women, ages 30–59 years at entry. Framingham Heart Study. *(Reproduced, by permission, from Kagan et al., 1963.)*

talented in mathematics to understand or carry out epidemiologic studies. While some studies in epidemiology do require sophisticated statistical methods, most problems can be handled well by the simple quantitative measures described here.

Distributions The most complete summary of a quantitative measurement made on a group of persons is the *distribution.* The distribution tells either how many or what proportion of the group were found to have each value (or each small range of values) out of all the possible values that the quantitative measure can have. In addition, the counts or proportions (or percentages) may be cumulated by adding each successive amount to all those that preceded it.

A distribution of serum uric acid values for 1,734 nonsmoking white men, ages 40–49, is shown in Table 2-1. Note that both numbers and percentages are shown for both the distribution and cumulative distribution.

A distribution may be displayed graphically as a histogram, in which bars represent the numbers or proportions of subjects in each "class interval." The uric acid distribution in Table 2-1 is shown in Fig. 2-2 as a histogram. Note that in plotting a histogram the *area* of each bar communicates the number or proportion of subjects represented. If all bars represent class intervals of the same width, then the area is proportional to the height. If some class intervals or bars are wider, as are the extreme right and left bars in Fig. 2-2, their height must be scaled down proportionally.

Another way to display a distribution is by plotting a series of points. Each point shows the midpoint of an interval and the number or proportion of subjects falling into that interval. The points may be connected by straight lines, yielding a polygon, or they may be

Table 2-1 Distribution and Cumulative Distribution of Serum Uric Acid Concentrations: Nonsmoking Men, Ages 40–49

Serum uric acid (mg/100cc)	Distribution		Cumulative distribution	
	Number	Percent	Number	Percent
1.0–2.9	10	0.6	10	0.6
3.0–3.9	68	3.9	78	4.5
4.0–4.9	315	18.2	393	22.7
5.0–5.9	565	32.6	958	55.3
6.0–6.9	431	24.8	1,389	80.1
7.0–7.9	229	13.2	1,618	93.3
8.0–8.9	85	4.9	1,703	98.2
9.0–11.9	31	1.8	1,734	100.0
Total	1,734	100.0		

Mean = 5.93 mg/100 cc
Standard Deviation = 1.31 mg/100 cc
Range = 1.32 to 11.12, or 9.8 mg/100 cc
Median = 5.84 mg/100 cc
Interquartile Range = 5.07 to 6.79, or 1.72 mg/100 cc
Source: Kaiser-Permanente multiphasic examination data, 1964–1968, tabulated by A. B. Siegelaub, M.S.

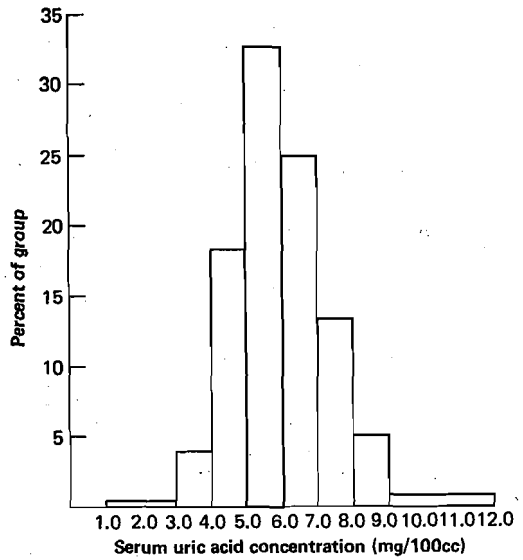


Figure 2-2 Percentage distribution of serum uric acid levels in Table 2-1, displayed as a histogram.

connected so as to form a smooth curve. The uric acid distribution in Table 2-1 is shown as a curve in Fig. 2-3.

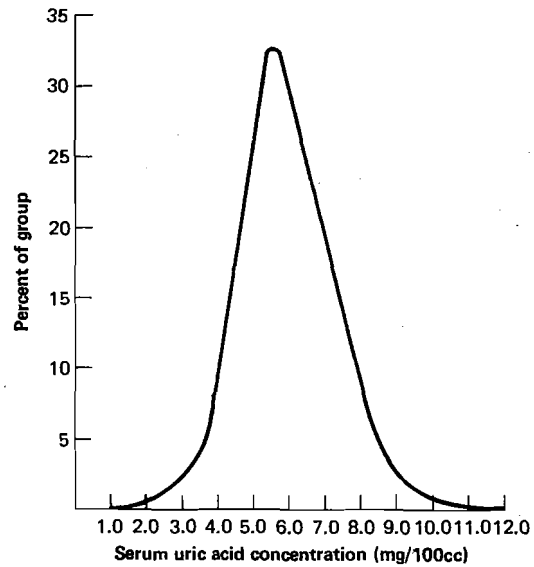
Cumulative distributions are usually shown graphically by curves. Fig. 2-4 shows the cumulative distribution curve for the same uric acid data.

Means The *mean*, or arithmetic average, is one of the so-called measures of central tendency of the values for the whole group. It is computed by adding all the individual values together and dividing by the number in the group. When one wishes to compare two or more groups, it may be cumbersome to compare their entire distributions. Comparing means is much simpler. In many cases, for comparative purposes, the mean is a reasonably good representation of the group's values, and it can be expressed with just one number.

It should always be remembered though that the mean is only one feature of a distribution and that two differently shaped distributions may have the same mean. It is often important to know more about the distribution than just the mean. In some cases we may be most interested in knowing how many people are at one extreme of the distribution.

Standard Deviations A good supplement to the mean in describing a group is the *standard deviation*, which is a measure of dispersion or variation. One way to compute it is to (1) square the difference between each value and the mean, (2) add the squared differences, (3) divide that sum by the total number of values minus one, and (4) find the square root of the result of (3). The mean tells where the values for a group are centered. The standard deviation is a summary of how widely dispersed the values are around this center. The standard deviation is also needed in comparing means of different groups to see how likely it is that a difference between two means could have occurred by chance, using statistical significance tests.

Figure 2-3 Percentage distribution of serum uric acid levels in Table 2-1, displayed as a curve.



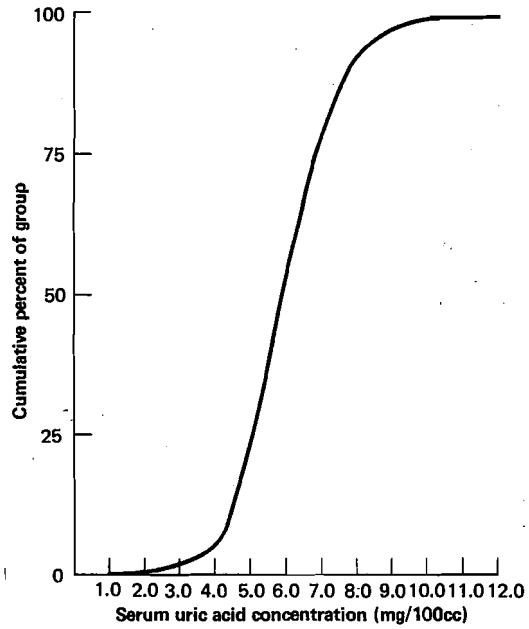


Figure 2-4 Cumulative percentage distribution of serum uric acid levels in Table 2-1, displayed as a curve.

Ranges The *range* of a distribution, the difference between the lowest value and the highest value observed, is, of course, another measure of dispersion. It is often less valuable than the standard deviation, however, since it only tells us about two members of a group. An extremely high or low value may be due to a measurement error.

Quantiles: Values That Divide a Group into Equal Parts Another way to describe a group on a quantitative scale or to classify each member of a group on such a scale is to divide the group into *quantiles*, or equal subgroups, along the scale. The simplest division is into two parts—the lower half and the upper half. The point on the scale that divides the group in this way is called the *median*. In the uric acid distribution shown in Table 2-1 the median value is 5.84 mg/100 cc. (When the median lies within an interval, e.g., between

5.0 and 6.0, we interpolate to estimate just where it lies). One-half of the group has values this high or higher and one-half has values this low or lower. Note that the median value can also be read from the cumulative distribution curve (Fig. 2-4) by seeing what uric acid value corresponds to the 50 percent point on the curve.

Just as one can compare two groups by their means, so one can also compare them by their medians. Medians are less often used than means but they have a few virtues that make them very useful in certain situations. One such situation is when a group has a few members with extreme values. The mean is substantially affected by these extreme values but the median is not. Suppose one wishes to summarize the weights of 22 women attending an obesity clinic. All but one are evenly distributed from 180 to 220 lb (i.e., 180, 182, 184, etc.). One is the fat lady in a traveling circus who weighs 420 lb. When she leaves, the mean weight of the clinic patients will drop by 10 lb, but the median will drop by only 1 lb. Medians are affected little by extreme values.

Another virtue of the median is its usefulness when some values are missing, but known to be above or below a certain level. Suppose one wishes to compare the age at death of two groups of fifty-year-old women exposed to different amounts of ionizing radiation. If one uses the mean age at death, then one must wait until all members of each group die. Conclusions cannot be drawn from the mean age of just some of the deaths, since an early difference between the two groups may be later counterbalanced by a difference in the opposite direction. By the time all the women have died, it is very probable that the investigator will also be dead or no longer interested in the study. Thus it is important to have an earlier answer. The median age at death is one such early measure, since it may be determined when only half the women in each group have died.

Groups may be divided into more than two parts. Three equal parts are known as *terciles*, four equal parts as *quartiles*, five as *quintiles*, ten as *deciles*. The finest division commonly used is into 100 parts, or *percentiles*. Percentiles are often useful for ranking individuals in relation to the total group. (Note that the borderlines between any divisions may be read from the cumulative distribution curve.)

Just as groups can be compared with respect to their medians,

they can also be compared as to their borderlines between quartiles, and so on. Similarly, persons in the upper quartile of a value can be compared with those in each of the other quartiles. Also, one may wish to have a measure of dispersion in a group analogous to the standard deviation. The size of the interval between two percentiles, e.g., the 20th and 80th, can be used. One such measure of spread is the *interquartile range*, the interval between the top of the lowest quartile and the bottom of the highest quartile. Note that the interquartile range can easily be read off of a cumulative distribution curve as in Fig. 2-4.

Quantiles may prove very helpful in determining which of two quantitative variables has a stronger relationship to disease. In a particular population group the incidence of coronary heart disease may increase a certain amount with each 20-mm-Hg increase in systolic blood pressure and a different amount with each 20 mg/100 cc increase in serum cholesterol, but this tells us nothing of the relative importance of the two attributes since the units of measurement for blood pressure and cholesterol are completely different, and not at all comparable. A more appropriate contrast would be to note how much the incidence of coronary heart disease increases as one moves up the scale of each measurement by quantile divisions such as deciles or quartiles.

A good example of such a comparison is shown in Fig. 2-5. In the Framingham Heart Study two serum lipid measures, cholesterol and the cholesterol/phospholipid ratio, were compared to determine which was the better predictor of the subsequent development of coronary heart disease. The study population was divided into quartiles of each of the two lipid values. As shown by the morbidity ratios in the figure, the risk of coronary heart disease was clearly related to cholesterol, the incidence being distinctly higher in each successive quartile. In contrast, the increase in risk with increasing quartile of cholesterol/phospholipid ratio was slight, showing that the latter measure was a distinctly inferior predictor.

Epidemiologic Measurements in Perspective

In summary, epidemiology requires that groups of people be described and compared in a quantitative fashion. However, the

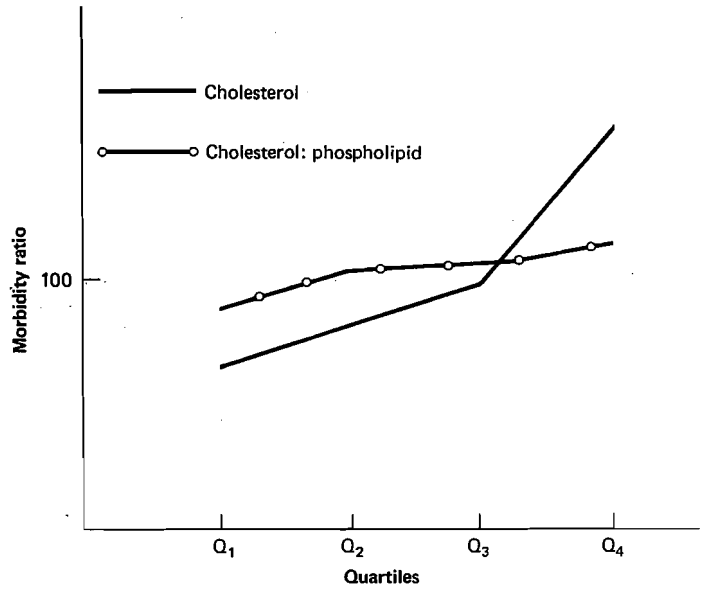


Figure 2-5 Risk of developing coronary heart disease in 10 years in subjects classified into quartiles of cholesterol and cholesterol/phospholipid ratio. Men, ages 30–59 years at entry. Framingham Heart Study. *(Reproduced in modified form, by permission, from Kannel et al., 1964.)*

particular characteristics of interest may be either qualitative or quantitative in nature.

When qualitative attributes are considered, persons with a particular attribute are counted, and the proportion of the total group studied that they constitute is determined. Since disease is the main concern of epidemiology, proportions of groups with disease or rates of disease are given primary attention. Disease rates are usually considered with respect to time. Disease present at one particular time is measured by a prevalence rate. Disease developing over a period of time is measured by an incidence rate.

Comparing disease rates among different groups is of primary importance. These comparisons are often expressed as differences between rates or as ratios of one rate to another.

Quantitative attributes are also important. It is often necessary

to consider the entire distribution of the quantitative measure in a group. However, this distribution may be described in a summary fashion by such measures as the mean and standard deviation. Breaking the group into equal parts according to ranking on a quantitative scale (quantiles) serves many useful purposes.

Obviously, the measurements described in this chapter do not exhaust the repertory of the epidemiologist. Other measurements have been used, and new ones will be invented for specific purposes. The simple measures described are established, time-tested, and widely understood.

REFERENCES

Hammond, E. C. Smoking in relation to death rates of one million men and women, *Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases*, edited by W. Haenszel, National Cancer Institute Monograph 19. U.S. Department of Health, Education, and Welfare, Bethesda, January 1966, pp. 127–204.

Ipsen, J., and P. Feigl, *Bancroft's Introduction to Biostatistics*, 2d ed. (New York: Harper and Row, 1970), Chaps. 2, 4, 5 and 11.

Kagan, A., W. B. Kannel, T. R. Dawber, and N. Revotskie. 1963. The coronary profile. *Ann. N.Y. Acad. Sci.*, **97**:883–894.

Kannel, W. B., T. R. Dawber, G. D. Friedman, W. E. Glennon, and P. M. McNamara. 1964. Risk factors in coronary heart disease: An evaluation of several serum lipids as predictors of coronary heart disease: the Framingham study. *Ann. Intern. Med.*, **61**:888–899.

Siegel, M., Indices of community health, in *Preventive Medicine* edited by D. Clark and B. MacMahon. (Boston: Little, Brown, 1967), pp. 67–79.

Chapter 3

Observations Used in Epidemiology

A wide variety of observations and measurements have been used by epidemiologists in their efforts to *describe* and *explain* the occurrence of disease in human populations. There are so many factors that influence human health and disease that almost any aspect of persons and their environments may be fair game for study. Depending upon what is being explored, epidemiologic studies may require the collaboration of scientists from other medical specialties and a variety of other disciplines. Ophthalmology, psychology, physical anthropology, bacteriology, and meteorology are just a few examples.

While we need not consider all varieties of data that may be used, certain types of observations recur frequently enough to deserve discussion. Health-care professionals must have some appreciation of the nature and limitations of these data sources. Not only are they used in scientific study, but they also provide the basis for vital decisions in day-to-day patient care.

Measures of Data Quality: Validity and Reliability

Observations or measurements, whether made by man or machine, involve some degree of error. Errors affect two important aspects of data quality—*validity* and *reliability.*

Validity Validity, or accuracy, is a measure of how closely the observations correspond to the actual state of affairs. As a clinical illustration, consider a patient with a rapid irregular heartbeat due to atrial fibrillation. Measurement of his heart rate by the radial pulse is considered inaccurate or lacking in validity because some heart beats produce a pulse too weak to be felt at the wrist. Compared to the true heart rate the radial pulse rate is *biased* toward lower values, resulting in what is commonly known as a "pulse deficit."

Reliability Reliability or reproducibility is a measure of how closely a series of observations of exactly the same thing match one another. If the cholesterol concentration of two portions of the same serum specimen is measured in an automated chemical analyzer, the two results should ideally be exactly the same. To the extent that they are not, the analyzer is said to lack reliability.

Effects of Lack of Validity and Reliability

Observations may be highly reliable but invalid. The cholesterol concentration on duplicate specimens may always agree within 5 mg/100 cc. Yet the readings may consistently be about 30 mg/100 cc too high.

This lack of validity does not necessarily rule out the use of the data. In some instances, knowing a person's absolute level of cholesterol may not be as important as knowing how that person ranks in his group. If all the group's values are 30 mg/100 cc too high, each person in the group will still be properly ranked in relation to the others. However, if one wishes to compare the mean cholesterol for that entire group with the mean of another group, for whom serum cholesterol has been measured accurately, the comparison will be unfair or biased.

Now consider the effects of unreliability. If a group of observations is unreliable, most will also be invalid due to departures from

the true values. However, if the unreliability is due to fluctuations that center around the true value, then the average or mean of a large series of observations may be quite a valid measure of the true average or mean. In this case many individuals will be improperly ranked relative to one another, if the ranking is based on one measurement for each. However, a comparison of the mean cholesterol of one large group with that of another may be quite fair and unbiased.

Usual Sources of Variation in Measurements

Not all the fluctuations in measurements or observations are attributable to lack of validity or reliability. The attributes themselves usually vary in a variety of ways.

Consider the distribution of blood pressures found in a community survey in which each subject has two measurements made. The major components of variation in the distribution are as follows:

Differences among subgroups—e.g., blacks have higher blood pressures, on the average, than whites; older persons have higher blood pressures than younger ones.

Differences among individuals within a subgroup—e.g., among black men, age 50, some individuals have higher blood pressures than others.

Differences within each individual—due to a variety of influences, each individual's blood pressure varies from one moment to the next. Some of these intraindividual differences may follow a regular pattern, e.g., diurnal variation.

Measurement errors—even if all blood pressures measured were exactly the same, they would appear to vary because of the failings of the observer, be it human or a mechanical device.

Sampling Variation

Another source of error or variation in data, known as *sampling variation*, is due to chance. It can be overcome by studying groups that are sufficiently large.

When we study the occurrence of a disease in a group of men, aged 50–59, in a community, we would like to think that our findings are applicable to all men of that age decade in that community. The

findings would undoubtedly be true of all 50-to-59-year-old men in the community if we studied all of them, but we usually have to take a sample. If the sample is selected in such a way that all men have an equal chance of being chosen, then we have what is called a *random sample.*

Experience and the laws of probability tell us that the larger the sample that is studied, the more likely are the findings to be representative of the total population. Conversely, the smaller the sample, the more likely we are to be misled. If repeated samples are drawn from a population, the findings in each sample will differ from one another—thus the term, "sampling variation." The larger the sample size, the less the variation, and the less chance of error.

This fact may be readily seen in the classic example of a large bag full of an equal mixture of black and white marbles. If an observer tries to determine what proportion of the marbles are white by pulling out only two marbles, he has a 25 percent probability of picking out two white marbles and concluding erroneously that all the marbles are white. If he pulls out four marbles instead, his chances of getting all white marbles are much less, only 1 in 16, or about 6 percent. One may apply the laws of probability to compute the likelihood of this false conclusion with any size sample; the result corresponds with our intuitive feeling that the more marbles one looks at, the less the chance of concluding that those in the bag are all white.

Thus, the larger the sample or group studied, the less the probability that chance error may occur. Statistical significance tests (such as "*t*" or chi square tests, and a variety of others described in statistics texts) are used to measure the probability of chance errors, given the size and characteristics of the study population and the question that is being asked. The result of a test of statistical significance is a probability level or "*p*" value, as frequently seen in medical journal articles. The expression "$p < 0.05$" means that there is less than a 5 percent probability that the observed result could have occurred by chance error.

Clinical Observations

Clinical observations are the primary basis for decisions as to the presence or absence of a particular disease. The most basic clinical

observations constitute the clinical history and physical examination. These are usually obtained by physicians, nurses, and other specially trained physicians' assistants.

The means for obtaining a history and physical examination need not be described here, but some comment about their limitations is in order. Many physicians have had memorable experiences in the unreliability of the medical history interview when they were medical students. Consider this all-too-familiar example. In preparation for rounds with the professor of cardiology the student devotes 10 minutes to careful questioning of the patient concerning nocturnal dyspnea and convinces himself that the patient indeed becomes short of breath at night and must sit up in bed in order to breathe more easily. After presenting the history during rounds the next day, the embarrassed student hears the patient tell the professor that he has never been short of breath at night.

The physical examination is no more reliable. If the patient is examined by half a dozen physicians, there will often be one or two who will hear (or not hear) a faint diastolic murmur not heard (or heard) by the others. The same degree of disagreement may be expected concerning the palpability of an elusive spleen. Differences in observer skill cannot be denied. Yet the murmur-hearers and spleen-feelers hold the psychological advantage, and objectivity probably suffers as a result.

Blood pressure, measured with a sphygmomanometer, has been a convenient measurement for the study of observer error in clinical medicine. It is a very sobering experience to be among a group viewing a movie prepared by Wilcox (1961), which shows a series of 14 views of a descending column of mercury in a sphygmomanometer accompanied by Korotkov's sounds amplified from a stethoscope. The group is asked to record the systolic and diastolic pressure for each measurement displayed. Even though all observers are seeing the same column of mercury and hearing the same sounds, the differences in the recorded results are striking. The greatest surprise comes when the viewers, learning that some of the early and late scenes are exactly the same, find discrepancies in their own readings for duplicate measurements.

When the results of a series of blood pressure measurements are tabulated, one human source of error that usually comes to light is *digit preference.* Physicians may tend to record values rounded off

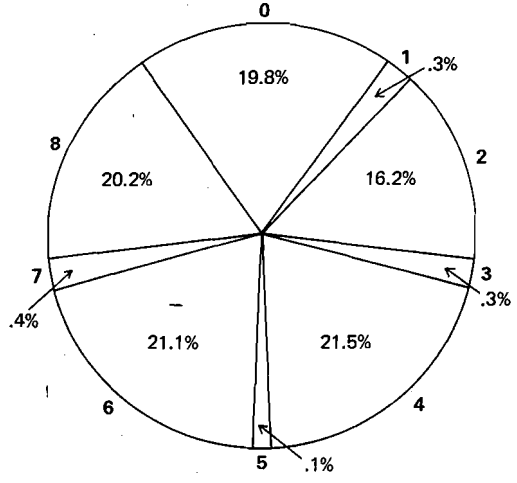
to a last digit of 5 or 0, or a preference for even over odd numbers becomes apparent (Fig. 3-1). Also noted has been a tendency to slant borderline values downward to avoid making unpleasant diagnoses.

Observations of Medical Specialists

Physicians in certain medical specialties make particular observations that are supposed to provide highly objective evidence as to the presence or absence of disease. Radiologists have the x-ray, cardiologists have the electrocardiogram, and pathologists have their stained microscopic sections. Implicit in giving a pathologist the last word in a clinicopathologic conference is perhaps the feeling that his observations will not only shed additional light on difficult problems, but that they are more reliable and valid than those of a bedside clinician.

A few of these specialists have made important contributions to our knowledge of the extent of observer variation in medicine. They

Figure 3-1 Percentage distribution of terminal digits on both systolic and diastolic blood pressure readings by an examining physician in the Los Angeles Heart Study. *(Reproduced, by permission, from Chapman, Clark, and Coulson, 1966.)*



have had the interest and courage to participate in studies to compare observations of the same visual object by different members of the same specialty or to compare duplicate observations by the same individual. The lack of reliability, even in these so-called objective measurements, has been striking.

Perhaps the classic series of studies in this area was carried out by Yerushalmy (1969) and his associates in the field of radiology. In one such study 14,541 entering college students received 70-mm chest photofluorograms. Each film was interpreted twice by two physicians and once by six others. Follow-up study of students with films read as "positive" by more than one reader, was accomplished by 14- by 17-in. chest film interpreted by a group of radiologists. The final interpretation regarding the presence of pulmonary tuberculosis was that 177 students had films that were "roentgenologically positive," 61 were "roentgenologically urgent," and 13 were "clinically active." Each of these cases, of course, had initial films that had been read by eight different readers. The percentages of original readings that were falsely read as negative were as follows:

	False negatives
Roentgenologically positive	26.9%
Roentgenologically urgent	25.4%
Clinically active	25.0%

Thus about one-quarter of all these nontrivial cases were missed the first time by competent x-ray readers.

Another series of 1,256 14- by 17-in. films were interpreted by a group of five competent radiologists and tuberculosis specialists. The number of films read as positive for tuberculosis by each reader was 56, 59, 62, 70, and 109, respectively. "Moreover, the radiologist who selected 109 did not include all those selected by the one who selected only 56." Similarly each reader read a different number as being positive when he read the films a second time. In each case some of the films read as positive once were read as negative by the same reader on another occasion.

The presence or absence of significant disease was not the only subject of inter- and intraobserver disagreement. Commonly accepted descriptive terms for pulmonary lesions such as "active,"

"inactive," "fibrotic," "soft," "hard," and "cavity" showed great differences among readers. After 2 years of work in trying to develop a reliable classification scheme to describe pulmonary lesions, the group of radiologists concluded that they had failed. "It was disappointing to find that many conferences and much practice, together and apart failed to increase reliability and agreement to a useful degree."

Interpretation of serial roentgenograms, the basis for many clinical decisions about tuberculosis patients, was also found to be grossly inconsistent. In making a judgment as to whether two x-ray films taken at different times showed progression, regression, or stability of disease, two readers disagreed with each other in about one-third of cases and a single reader disagreed with himself in about one-fifth of cases.

Clinical Diagnoses

Diagnoses are inferences or conclusions based on clinical and laboratory observations. Not only may these observations be incorrect, but the reasoning leading to the conclusions may also be in error. Yet even if the observations are complete and accurate and the reasoning is sound, differently trained physicians use different criteria for making the same diagnosis. Also leading to observer variations is the fact that one physician may have access to more laboratory tests or other specialized data. Furthermore, different terms may be used to refer to the same clinical condition, and a single term (e.g., "arteriosclerotic heart disease") may have different meanings to different physicians.

Thus, clinical diagnoses by themselves are indeed undependable indicators of disease for scientific study. Whenever possible, specific criteria should be established for making diagnoses. These criteria should be adhered to carefully and described clearly so that the work may be repeated or evaluated by others.

Medical Chart Review

Both epidemiologic studies and patient care frequently rely upon the review and abstracting of information from medical records. Just as

is found for other types of observations, the reading of charts involves substantial amounts of error. Even if the information is relatively complete and the various handwritings are legible, two chart readers will extract differing information. Usually, however, matters are much worse, with missing information and cryptic or illegible physicians' notes.

Disease Reporting

Physicians are legally required to report certain diseases to the local public health authorities at the time of diagnosis. The primary purpose of this is to detect the onset of epidemics of certain serious diseases and to provide information so that appropriate community-wide control measures can be undertaken. In addition to their usefulness in disease control these data may also be used to measure disease incidence in the community.

Despite official requirements, many diseases are underreported. For example it has been estimated that, despite the mounting concern over the recent epidemic of venereal disease, only one-fourth of all cases are reported. Desire to avoid social stigma for patients, the pressures of other work, and laxity are among the reasons that have been given for underreporting.

This is not only the case with regard to certain infectious diseases. In the 1950's and 1960's two large agencies, the American Medical Association and the U.S. Food and Drug Administration carried out special programs to encourage physicians to report instances of suspected adverse drug reactions. The purposes of these programs were to obtain some measure of the frequency of occurrence of various drug reactions and to provide a means of receiving early warnings of as yet unsuspected side effects of drugs. The response of physicians was quite disappointing. By and large, busy physicians do not wish to take the time to fill out the reporting forms. In a study of various approaches to detecting adverse drug reactions in a hospital, Cluff et al. (1964) judged a system whereby physicians were to fill out a drug reaction card at the time of discharge to be "completely unsatisfactory," since intensive daily surveillance of just one service yielded four times as many reactions as were listed by report card from the entire hospital.

Death Certificates and Mortality Statistics

Mortality data for nations, states, and communities, as obtained from death certificates, have played an important role in epidemiologic research for more than a century. Many major problems and inaccuracies are associated with death certificates. (See Feinstein, 1968, for detailed discussion.) Nevertheless, they constitute a widely implemented collection of data about fatal illnesses that can be used to study disease occurrence on a local, national, or international scale.

Death certificate diagnoses are usually clinical diagnoses and are thus subject to all the vagaries described above. In addition, the patient may have had several diseases contributing to his death, but under current procedures, only one underlying cause is to be selected. Before 1949 in the United States, coding rules automatically led to the choice of one underlying cause out of several possibilities. For example, if both diabetes mellitus and heart disease were listed on the death certificate, diabetes was coded as the underlying cause even if the doctor felt that heart disease was more to blame. Starting in 1949, the physician was asked to indicate the underlying cause. While this may have been an improvement, it resulted in some sudden changes in apparent mortality rates (e.g., a drop in diabetes mortality, as would be expected); it also forced physicians to oversimplify many complex situations where multiple causes might have been involved. For this reason, many authorities have urged the adoption of a multiple-cause coding system for death certificates. If mortality statistics are to become more meaningful, it would be helpful if physicians were trained in uniform and proper procedures for filling out death certificates.

Other changes in diagnostic classifications have been made in the *International Classification of Diseases*, now in its eighth revision, leading to abrupt changes in reported mortality rates for the diseases affected. Studies of time trends in disease mortality must take into account these coding changes as well as the technological advances that lead to increased diagnoses of particular conditions and changes in the fashion of allocating deaths to one disease instead of another. Fig. 3-2, from a study by Reid and Evans (1970), shows time trends in mortality rates for nephritis, hypertension, and

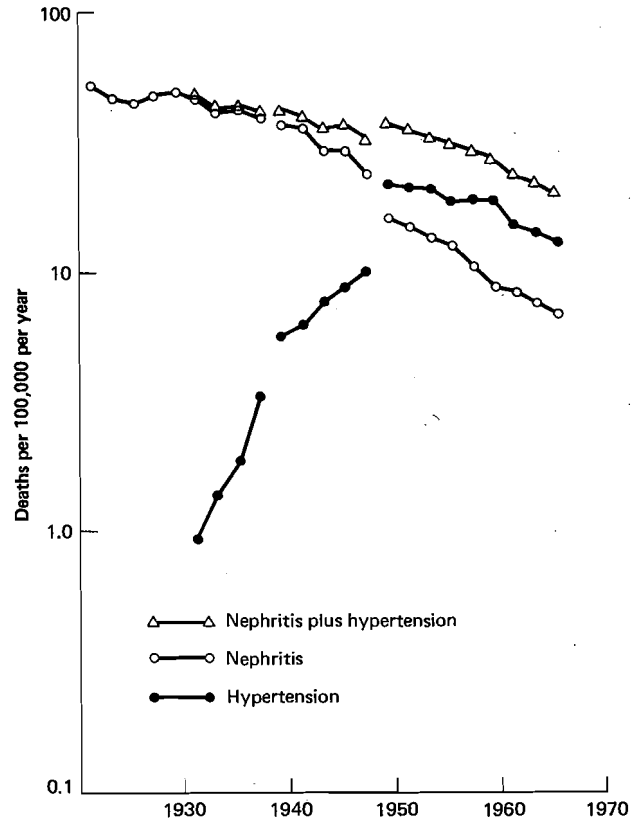


Figure 3-2 Mortality rates from nephritis, hypertension, and both combined, among males aged 45–54 years in England and Wales, from 1931–1966. *(Reproduced, by permission, from Reid and Evans, 1970.)*

both combined, among men ages 45–54 in England and Wales, and illustrates several of these factors. The gaps in the curves reflect changes in disease classification. The sharp rise in the death rate for hypertension between 1931 and 1950 probably reflects the increased use of the sphygmomanometer and an increased awareness of the importance of hypertension. The reciprocal changes in hypertension and nephritis deaths may represent an increasing tendency to

attribute uremia to kidney damage produced by hypertension rather than by inflammation.

Responses to Questionnaires

The clinical history is only one of many kinds of data that may be obtained by questionnaires. Data relating to social status or exposures to environmental hazards can also be obtained in this manner.

It does not seem necessary to belabor the frailties of human observations and their written or oral communications any further, except to encourage again a reasonably skeptical attitude toward the results of questionnaire studies and to show some examples of problems commonly encountered.

Nonresponse If given a choice, a substantial proportion of individuals will not answer questions. In 1971 a questionnaire was mailed to 8,250 Kaiser Foundation Health Plan members participating in a study to evaluate periodic checkups involving multiphasic screening. Because of the incomplete response to one mailing, four subsequent mailings were sent out to nonrespondents. The percentage of the total group responding to each mailing is shown in Table 3-1. The final nonresponse rate of 20.4 percent (100 percent minus 79.6 percent) is not at all unusual for a mailed questionnaire.

Nonresponse can also occur under more controlled or supervised conditions. As part of a multiphasic examination at Kaiser-Permanente, patients are given a self-administered questionnaire containing a series of questions about their smoking habits. The answers to the questions about smoking were used to classify examinees in a study of the characteristics of smokers and nonsmokers (Friedman et al., 1972). In doing so it was found that about 12.7 percent of 111,024 persons did not answer at least one of the crucial questions about present or past smoking habits.

Nonresponse would not constitute a serious problem if it merely reduced the number of subjects available for study; however, it may also lead to a biased study sample if the respondents and nonrespondents differ with respect to health or some other characteristic being studied. Unfortunately, this is frequently the case.

Table 3-1 Response to Five Mailings of a Questionnaire by 8,250 Kaiser Health Plan Members*

Mailing	Percentage of total study group responding
First	43.4
Second	15.4
Third	8.6
Fourth	7.0
Fifth	5.2
Total	79.6

*Data tabulated by Barbara A. Campbell, M.A.

Inconsistent or Otherwise Unusable Responses It is surprising how often persons will answer both "yes" and "no" to the same questionnaire item or provide otherwise inconsistent responses. In the study of smoking just referred to 2.3 percent of subjects did not indicate that they smoked cigarettes, but then gave a positive response to some duration of smoking or quantity of cigarettes smoked. Because of this and other serious inconsistencies, plus the omissions described above, 16.5 percent, or about one-sixth of the total subjects, had to be eliminated.

Overreporting of Disease Symptoms Patients who either deny or exaggerate disease symptoms are well known to physicians. In a study of the reliability of a self-administered questionnaire (Collen et al., 1969) it was found that on the average, one-fifth of persons who answered "yes" to a symptom question the first time, denied the symptom when the questionnaire was administered again at the same examination. Physicians who perform follow-up examinations after patients have answered a symptom questionnaire often find that positive responses to questions about serious symptoms either cannot be substantiated or appear nonsignificant upon careful history-taking. As an example of the likely overreporting of symptoms on a self-administered questionnaire, 15.9 percent of

1,950 girls, ages 15–19, taking Kaiser-Permanente multiphasic examinations, answered "yes" to the following question describing symptoms almost pathognomonic of angina pectoris: "In the past year have you had repeated pain (or pressure or tight feeling) in your chest when you walked fast or uphill and that left after a few minutes rest?"

Presenting Oneself in a Favorable Light This is such a universal trait that it hardly needs to be mentioned except that it can introduce systematic biases into epidemiologic studies. Persons tend to deny venereal disease and drug abuse and underestimate their alcohol consumption.

Household Surveys

Information about medical conditions and other pertinent social and personal characteristics is frequently obtained by household interview survey. Assessments of health and social problems by survey may be the basis of determining priorities for community or national policy. Thus, the limitations of this study method should be well understood.

Problems associated with survey data and the techniques of obtaining representative samples of individuals for questioning have been a major concern of social scientists. In the health field, the *National Health Survey* (NHS) has been authorized by the United States Congress to carry out surveys by the household interview method since 1957. In the course of this work NHS scientists have carried out important methodologic studies to determine the accuracy of interview-acquired information.

In one such study (Madow, 1967), patients' reporting of chronic conditions was compared to the chronic conditions recorded by physicians in their medical records during a 1-year period. Overall, 45.3 percent or almost half the chronic conditions recorded by the physicians were not reported by the patients despite the fact that patients were given a fairly comprehensive checklist of conditions to jog their memories.

Thus, interview data about illness are apt to be incomplete. As might be expected, conditions for which the patient made more

frequent doctor visits were more apt to be reported, as were those for which a doctor was seen more recently. Furthermore, conditions were more likely to be reported in an interview if they affected the person's way of life, for example, by causing pain or worry or limitations in his work or in what he could eat or drink.

Laboratory Data

Mechanical, electrical, and chemical measurements are also subject to error. Well-run clinical laboratories maintain continuing quality control programs to monitor the validity and reliability of their measurements. When significant errors occur, monitoring permits institution of prompt corrective action.

Yet, even with the most careful quality control, significant errors occur, due both to known and unknown factors. Many of these factors cannot be controlled within the laboratory. Only in the past decade, for example, did it become generally known that exposure of a blood specimen to light would cause a breakdown of bilirubin and significantly lower the serum bilirubin concentration measured in the laboratory. Similarly, ingestion of a variety of drugs can affect the measurement of important blood constituents. A well-known example is the effect of iodide-containing drugs on the protein-bound iodine test of thyroid function.

STUDYING RELATIONSHIPS IN IMPERFECT DATA: THE VALUE OF INVESTIGATING LARGE GROUPS

This section is, to the author, one of the most important in this book. It will attempt to bridge a serious gap in understanding and communication between the scientifically minded clinician and the epidemiologist.

As will be developed more fully in the next chapter, one of the primary concerns of the epidemiologist, like other scientists, is the study of *relationships*. The epidemiologist focuses on relationships between diseases and other human or environmental attributes by studying population groups.

The clinician focuses on the individual patient and strives to obtain complete and accurate information, in order to provide the

best possible diagnosis and treatment. In his appropriate concern for the patient's welfare, he can tolerate few avoidable errors in this information. Accustomed to high standards in his pursuit of information and the expenditure, if necessary, of hundreds of dollars per patient in laboratory tests and specialized diagnostic procedures, he becomes intolerant of the use of relatively low-quality data such as questionnaires or death certificates in epidemiologic studies.

A case in point is the difficulty in convincing some neurologists of the validity of epidemiologic studies of stroke that do not include an evaluation of all study subjects by a neurologist. Neurologists spend years learning the subtleties of the neurological examination and the fine points of differentiating strokes from a variety of other neurological conditions (many of which are quite rare). To many physicians with such a background it is inconceivable that one would undertake a scientific study of stroke based, say, on identification of cases simply by asking, "Have you ever had a stroke?"

Yet in a study of a large population, the human and financial resources to provide a neurologist's examination for all subjects are not available now, nor will they be in the foreseeable future. So let's compromise and have any ill persons in whom the attending physician suspects a stroke evaluated by a neurologist. This approach is more workable and can be employed in special intensive population studies such as the Framingham Study (described in Chap. 8). Yet even there, practical difficulties arise; if a person has a stroke which is rapidly fatal or which occurs out of town, he will probably not be seen by a neurologist.

The epidemiologist is not in favor of bad data. He wants the best he can get. But experience has shown that he can discern important relationships, even in data of relatively poor quality because studying large groups provides power to overcome error. With some validity to the data and large enough numbers of study subjects to minimize sampling error, one may still derive some valuable information from poor quality data.

Consider the following numerical example. Suppose that we wish to determine whether there is a relationship of stroke to hypertension and we can only use a questionnaire which asks, "Have you ever had a stroke?" and "Have you ever had high blood pressure?" The questionnaire is administered to 10,000 persons,

ages 65–74. Let us postulate that the true state of affairs for this population happens to be that 200 persons have had a stroke and 2,000 have had high blood pressure. Of the stroke cases, 150 had high blood pressure. The *true* population breakdown is shown in Table 3-2.

A slight digression here may be of value to the reader who is unfamiliar with the presentation of data in a "two-by-two" or "fourfold" table, frequently used in epidemiology and exemplified by Table 3-2. These tables show the relationship of one "yes-or-no," or dichotomous, variable to another. The presence or absence of one disease or characteristic is indicated at the left and the presence or absence of the other is shown at the top.

Table 3-2 shows how the population is divided in the four possible ways according to each of the two characteristics. The number, 150, in the upper left corner indicates that there are 150 persons with both a history of stroke and of hypertension. The number, 1,850, to the right of the 150, represents 1,850 persons with a history of hypertension but no history of stroke. The sum of 150 and 1,850, or 2,000, is shown at the far upper right and represents all persons with a history of hypertension. The 8,000 persons without a history of hypertension are shown in the second row. Fifty of the 8,000, on the left, have a history of stroke. The 7,950, next to them, do not have a history of stroke. The total of 50 plus 7,950, or 8,000, is shown to the right. Totals of the columns are shown below and represent the 200 persons with a history of stroke and the 9,800

Table 3-2 "True" Breakdown of a Population of 10,000 Persons, Ages 65–74, According to the Presence or Absence of a History of Hypertension and a History of Stroke (Fictitious Data)

		Stroke history (True)		
		Present	Absent	Total
Hypertension history (True)	Present	150	1,850	2,000
	Absent	50	7,950	8,000
	Total	200	9,800	10,000

without. The grand total of the population, or 10,000, is shown at the lower right-hand corner.

Returning now to the argument at hand, the prevalence of a history of stroke in those with a history of high blood pressure is $^{150}/_{2,000}$ or 7.5 percent. The prevalence of a stroke history in those without a hypertension history is $^{50}/_{8,000}$ or 0.625 percent. Thus, if one could only know the true situation, one would find that those with high blood pressure in the past had $^{7.5}/_{0.625}$, or 12 times, the likelihood of the nonhypertensives, of having a history of stroke.

Now let us estimate that our questionnaire only elicits positive responses to the stroke question from 160, or four-fifths, of the stroke cases and, in addition, 196, or 2 percent, of the 9,800 nonstroke cases answered "yes" to the stroke question by mistake. Let us also assume that only one-half of hypertensives were aware of, and reported, their elevated blood pressure and that 5 percent of nonhypertensives erroneously reported that they were hypertensive.

As a result of these errors, some of the persons from each "true" category will be moved to each of the four "reported" categories. For example, consider the 150 persons with *true* strokes and *true* hypertension. Only half report their hypertension. Of the 75 reporting either hypertension or nonhypertension one-fifth do not report their stroke. So the 150 "true" stroke cases with hypertension will be distributed into the four "reported" categories as shown in Table 3-3.

Table 3-3 Parceling Out the 150 Persons with a "True" History of Both Stroke and Hypertension into Four Categories According to What They Will Report on the Questionnaire (Fictitious data)

		Stroke history (reported)	
		Present	Absent
Hypertension history (reported)	Present	60	15
	Absent	60	15

One may go through this exercise with each of the other three "true" categories and divide each into the four "reported" categories. If one then adds all the persons in each of the "reported" categories, the (rounded) result is as shown in Table 3-4.

Now the observed prevalence of a history of stroke in prior hypertensives is $^{88}/_{1,400}$, or 6.3 percent. This is about twice the 3.1 percent prevalence ($^{268}/_{8,600}$) in prior normotensives. *Despite the poor quality of the data, the relationship between hypertension and stroke, while not as strong as in reality, may still be perceived.* Thus, the study of relationships in groups of people can, to some degree, overcome certain kinds of error.

This is not an argument for using poor data when better are obtainable. One must always be aware of the limitations of his data and how inaccuracies and biases may affect his results. In the example it was assumed that the failure to report hypertension was equally true of persons with and without stroke. If stroke affected memory so as to further diminish the reporting of hypertension in the stroke case group, then the study might have missed the stroke-hypertension relationship completely, or might even have led to the opposite conclusion. Thus, data can be, and often are so bad as to be unrevealing or even misleading, despite large numbers.

The example given illustrates another epidemiologic principle. Where relationships are observed in data with an appreciable number of misclassified subjects (e.g., persons with a disease classified as not having it), the results are conservative. That is, the

Table 3-4 Findings in the Total Population Based upon What They Report on the Questionnaire (Fictitious Data)

		Stroke history (reported)		
		Present	Absent	Total
Hypertension history (reported)	Present	88	1,312	1,400
	Absent	268	8,332	8,600
	Total	356	9,644	10,000

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