CHAPTER 1

EPIDEMIOLOGY AND ITS BASIC MEASUREMENTS

by

Gary S. FRIEDMAN

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 patient, the need to individualize the diagnosis and threatment 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 epidemiological studies. In general, epidemiologists are 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 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 prensented 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 adistorted view. There are certain questions that can 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 able to produce lung cancer in experimental animals by forcing them to smoke cigarettes. However, the idea that cigarette sloking causes human lung cancer would be unconvincing unless epidemiologists also showed that lung cancer occurred more often in cagerette 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 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 epidemiological view of disease etiology is as a web of caution. This concept of disease causation considers all the predisposing factors to a disease 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. Note that many interrelated factors ultimately lead to myocardial infarction. Each of the factors ultimately lead to myocordial infarction. Each of the factors mentioned is also influenced by a variety of other factors, 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. For 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 who tries to practice medicine using only the textbooks will discover.

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., fracture of the femur or adenocarcinoma of the colon). 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 infectious agent is reflected in the disease name, we often forget that other factors are operating and ten 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. The 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.

BASIC MEASUREMENTS

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 100, is one of the most common ways of expressing proportions. Number per 1000 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 post-partum day.

Incidence Rate

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

Incidence describes the rate of development of a disease in a group over a period of time; this time period is included in the denominator. In contrast to prelevance, which is like a snapshot of all cases, incidence describes the continuing occurrence of new cases of a disease. For example, the incidence of myocardial infarction is about 1 percent per year in men aged 55 to 59 in our community; or, at the height of the epidemic the incidence of chickenpox 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 one 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 usesd in epidemiology, are described below.

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

Sometimes one wishes to have a measure of all the disease 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.

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

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

Any rate may refer to subgroup of a population. An example is the age-specific mortality rate.

Age-specific mortality rate =

number of persons

dying in a particular age group per unit of time
total number in the same age group

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

Case fatality rate refers to the proportion of persons with a particular disease who die. The time period 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 derived from the denominator population at risk. This is done for convenience, because of the ready availability of data that approcimate the ideal. Consider the maternal mortality rate :

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

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 shinking 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 :

Attributable Risk and Relative Risk

Differences It is often desired to compare a rate in one group with that in another. One may simply note noth 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 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 to 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 smoker's 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 to 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 fait 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 hroup in smokers. The method for computing as that of each group of smokers. The method for computing this expected rate involves what is called age adjustment, or age standardization.

Epidemiologic Measurements in Perspective

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

Obviouslu, 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.

BIBLIOGRAPHY

MacMahon, B., and T.F. Pugh : Epidemiology : Principles and Methods Little, Brown, Boston 1970