Chapter 33

EPIDEMIOLOGIC MEASUREMENT: BASIC CONCEPTS AND METHODS

LYNN I. LEVIN, PHD, MPH

INTRODUCTION

DEFINITION OF EPIDEMIOLOGY

BASIC MEASURES OF DISEASE FREQUENCY

DIRECT METHOD OF RATE ADJUSTMENT

TYPES OF VARIABLES

SUMMARY STATISTICS FOR CATEGORICAL AND CONTINUOUS VARIABLES

METHODS FOR DISPLAYING DATA

SAMPLING

MAJOR TYPES OF STUDY DESIGNS AND MEASURES OF ASSOCIATION USED IN EPIDEMIOLOGY

BIAS

HYPOTHESIS TESTING

STATISTICAL ASSOCIATIONS AND CAUSE-AND-EFFECT RELATIONSHIPS

SUMMARY

L.I. Levin; Division of Preventive Medicine, Walter Reed Army Institute of Research, Silver Spring, MD 20910-7500

INTRODUCTION

This chapter will provide the public health professional with the basic tools and concepts of epidemiology. The target audience is the military practitioner of public health who may be deployed or working in a public health setting where surveillance activities or rudimentary research studies are needed. The epidemiologic methods covered in this chapter are at the elementary level and are presented from the more simple concepts to the more complex. Other chapters in this volume specifically address outbreak investigations and surveillance activities (see chapters 31, Disease and Nonbattle Injury Surveillance: Outcome Measures for Force Health Protection, and 32, Outbreak Investigation). The epidemiologic tools that are used to conduct these activities are presented here. The focus is on data collection methods that are necessary to run programs, assess outbreaks, allocate resources, or perform other activities that are related to the formulation of public health policy and practice.

DEFINITION OF EPIDEMIOLOGY

Epidemiology is the study of the distribution and determinants of disease, injury, or health-related factors in human populations.[1] In contrast to clinical medicine, which focuses on the individual, epidemiology evaluates groups of people or populations. A basic premise of this discipline is that diseases, injuries, or other medical conditions are not randomly distributed across populations but rather vary according to factors such as environmental exposures (eg, deployment history) or personal characteristics (eg, smoking status). As early as the 5th century BC, Hippocrates hypothesized that the development of human diseases may be associated with the environment.[2] The goal of the epidemiologist is to determine who becomes ill and why, and this is accomplished by comparing groups with differing characteristics.

BASIC MEASURES OF DISEASE FREQUENCY

Epidemiology is a quantitative science. One of the fundamental measures used by epidemiologists is the rate of disease. By computing rates, it is possible to compare two populations or groups. A rate requires consideration of the population from which the cases are derived:

$$\text{Rate} = \frac{\text{Number of Events}}{\text{Population at Risk of Event}}$$

Typically, the denominator is the total population at risk of disease, and the numerator is the number of people with disease. The rate is usually expressed in some conventional base, such as events per 1,000 individuals. There are several types of rates used in epidemiology to describe morbidity and mortality.

Prevalence

A commonly used measure in epidemiology is the prevalence rate. Although it has historically been called a rate, prevalence is a proportion representing the fraction of the population that has a disease at a single point in time. It is a snapshot of the burden of disease in a defined population and includes both new and existing cases. Thus, prevalence depends on both the incidence of new cases of disease and the duration of disease in those cases. This measure typically is used to assess the need for and costs of health services.

$$\text{Prevalence} = \frac{\text{Number of Existing Cases of Disease at a Given Time}}{\text{Total Population at a Given Time}}$$

For example, a 15% prevalence of rash-associated illness in units of an Army support battalion in Operation Joint Endeavor, Bosnia, was determined in this way[3]:

$$\text{Prevalence} = \frac{69 \text{ Cases of Rash-associated Illness}}{466 \text{ Deployed Unit Members}} = 0.148 = 15\%$$

Incidence Rate

Incidence describes the rate of development of a disease or a medical condition in a population over a period of time, that is, the occurrence of new events in a specified time period. This measure is used to evaluate the causal or etiologic role of risk factors and the development of disease.

$$\text{Cumulative Incidence Rate} = \frac{\text{Number of New Cases During a Specified Period}}{\text{Total Population at Risk of the Disease During that Period}}$$

This measure is called the cumulative incidence rate because it accumulates the number of cases over time that derive from the population at risk determined at the beginning of the time period (ie, fixed population with the assumption that no individuals are lost to follow-up). The denominator contains the count of people. Cumulative incidence is a proportion that ranges from 0 to 1 (assuming only one possible occurrence of the disease per person). When reporting a cumulative incidence, the time frame must be specified. Attack rates are a measure of cumulative incidence.

For example, in a study among US Army soldiers in infantry basic training, 303 individuals were followed for 12 weeks of training to determine the incidence of training-related injuries. Of the 303 soldiers, 112 developed one or more lower extremity musculoskeletal injuries.[4]

$$\text{Cumulative Incidence Over a 12-week Period} = \frac{\text{112 Soldiers Injured}}{\text{303 Soldiers in Basic Training}} = 0.369 = 37\%$$

Another measure of incidence is the incidence density rate. This is a measure of the instantaneous rate of development of a disease in a population.

$$\text{Incidence Density Rate} = \frac{\text{Number of New Cases of a Disease During a Specified Period}}{\text{Total Person-time of Observation During the Given Time Period}}$$

The numerator of the incidence density is the number of new cases occurring in the population (the same as in the cumulative incidence). The denominator, however, is the sum of each person's time at risk or the sum of the time that each person was under observation and susceptible to the disease rather than the count of individuals at the beginning of the follow-up period. Thus, this measure can incorporate data from a dynamic population: individuals who are followed for various periods of time or who are lost to follow-up. An incidence density rate must include the relevant time units in the denominator, such as person-days or person-years. In calculating person-time, following 10 individuals for 3 years (30 person-years) is equivalent to following one individual for 30 years. Incidence density can range from zero to infinity.

For example, in a study of the incidence of HIV-1 infection in the US Army during the period 1 November 1985 to 31 October 1993, a total of 1,061,768 active-duty soldiers were followed for a total of 3,629,688 person-years of follow-up. During the period of the study, 978 soldiers with HIV-1 seroconversion were identified.[5]

$$\text{Incidence Density Rate} = \frac{\text{978 HIV-1 Soldiers}}{\text{3,629,688 Person-years}} = 2.7/10{,}000 \text{ Person-years}$$

Prevalence and incidence measures are interrelated. When the incidence rate has been constant over time (a "steady state" situation), the duration of disease has remained unchanged, and the prevalence of disease overall is low, then the prevalence (P) equals the product of the incidence density (I) and the average duration ($\overline{D}$):

$$P = I \bullet \overline{D}$$

If two of these measures are known, the third can be calculated.

Incidence density rates, like other rates, can be calculated for subpopulations, such as different age, sex, or race-ethnic groups. The following is an age-specific incidence rate:

$$\text{Age-specific Incidence Rate} = \frac{\text{Number of New Cases in a Specified Age Group}}{\text{Total Person-years of Observation for the Age Group}}$$

For example, the incidence rate of HIV-1 infection among 20- to 24-year-olds in the study[5] cited above was determined in this way:

$$\text{Age-specific Incidence Rate for Ages 20 to 24} = \frac{\text{405 HIV-1 Soldiers}}{\text{1,216,125 Person-years Among Soldiers Aged 20 to 24}} = 3.3/10{,}000 \text{ Person-years}$$

Mortality Rate

Incidence rates describe the incidence of disease in a population, whereas mortality rates describe the incidence of death in a population. The crude mortality rate can be determined using this equation:

$$\text{Crude Mortality Rate} = \frac{\text{All Deaths in a Calendar Year}}{\text{Total Population at Risk of Death During Year}} \bullet 10^n$$

For example, the following is the crude non–battle-related mortality rate among deployed servicemembers to the Persian Gulf region, 1 August 1990 to 31 July 1991.[6]

$$\text{Crude Mortality Rate} = \frac{\text{225 Nonbattle-related Deaths}}{\text{264,868 Person-years in the Persian Gulf Region}} = 85/100{,}000 \text{ Person-years}$$

The cause-specific mortality rate is determined in a similar fashion:

$$\text{Cause-specific Mortality Rate} = \frac{\text{Number of Deaths from a Specific Cause in a Year}}{\text{Total Population During the Year}}$$

A cause-specific mortality rate was determined for motor vehicle mortality among deployed servicemembers to the Persian Gulf region for the period 1 August 1990 to 31 July 1991[6]:

$$\text{Cause-specific Mortality Rate} = \frac{\text{62 Deaths Due To Motor Vehicle Accidents}}{\text{264,868 Person-years in the Persian Gulf Region}} = \frac{23}{100{,}000 \text{ Person-years}}$$

All of these measures of morbidity and mortality are used to monitor disease trends in surveillance activities, identify outbreaks, and plan and evaluate health services.

DIRECT METHOD OF RATE ADJUSTMENT

Rate adjustment allows the public health officer to compare rates of disease or injury in two communities that have different demographic characteristics, such as different age distributions or race-ethnic compositions. Because incidence rates or mortality rates typically vary by age and race, it is necessary to adjust or standardize these rates by these factors so that the rates can be compared on equal terms. Adjusted rates are artificial in that they are only used for comparison purposes and do not describe a particular population. The process of adjusting rates requires the use of a standard population with which both communities are compared and data on the factor-specific rates of disease in both communities. If, for example, injury rates in the Army are to be compared with injury rates in the Navy, these rates would have to be adjusted for age and sex, because the composition of the two services differs in these factors. Texts such as Kahn and Sempos[7] or Selvin[8] have examples on how to perform a direct age adjustment.

TYPES OF VARIABLES

Categorical Variables

In addition to obtaining information on rates of disease, epidemiologic studies can also collect information on many variables about individuals.

Variables that are divided into categories or are assigned codes are called categorical variables. Each category is defined and there are a limited number of values that can be measured. Dichotomous variables are categorical variables that assume only two values, such as male and female or inducted into military services and not inducted. Polychomotous variables can be divided into more than two categories, such as race and cause of death. Some discrete variables can be ordered or ranked. Examples of ordered variables are severity of pain (eg, mild, moderate, intense) and military rank.

Continuous Variables

Variables that can be measured on a continuous scale, such as height or 1-mile run times, are considered continuous variables. The values that these variables can assume are only limited by the level of accuracy of the scale on which they are measured.

SUMMARY STATISTICS FOR CATEGORICAL AND CONTINUOUS VARIABLES

Categorical Variables

Discrete data can be summarized by calculating frequencies, proportions, or percentages. A proportion is the number of people with a characteristic divided by the total number of people. The percentage is a proportion multiplied by 100.

Continuous Variables

Measures of Central Tendency

There are three main measures of central tendency to describe the distribution of a continuous variable. The arithmetic mean, is calculated as the sum (Σ) of all of the observed values (x_i) divided by the total number of observations (n):

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The mode is the most frequently observed value. The third measure, the median, is the midway point of a series of numbers such that 50% of the numbers are above the value and 50% are below. For an odd number of entries, the median is the middle number. For an even number of entries, it is the average of the two middle numbers. If the distribution is fairly symmetric, then the median value will be close to the mean value. If the distribution is skewed, then the median value is a better measure of central tendency than the mean.

Measures of Variation or Dispersion

The range is the difference between the highest and lowest values. The standard deviation estimates the amount of variability in a set of numbers. It is defined as the square root of the sum of the squared deviations from the mean, divided by the number of observations minus 1.

$$\text{Standard Deviation} = s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}$$

The larger the standard deviation, the more variable or nonhomogenous the distribution. Statistical theory states that a population with a normal distribution of values will have a characteristic bell-shaped curve. Since the shape of this frequency distribution is symmetrical, the mean, median, and mode are the same. With normally distributed data, the standard deviation describes the width of the curve. The range of values one standard deviation above and below the mean will include 68% of the observations. The range of values two standard deviations above and below the mean will include 95% of the observations. The range of values three standard deviations above and below the mean will encompass over 99% of the observations. Exhibit 33-1 contains an example illustrating how to calculate a standard deviation.

METHODS FOR DISPLAYING DATA

Tables

A concise way to summarize data is to present the data in a table. Tables typically contain frequency data on a range of values for discrete variables or summary statistics for continuous variables. A contingency table is used to summarize counts of people or observations. The number of rows and columns in the table represent the various levels of two variables. A particular type of contingency table, the 2 x 2 table or 4-fold table, is used for dichotomous variables such as exposed or nonexposed and diseased or nondiseased. Examples of 2 x 2 tables are presented later in this chapter.

The title of the table should be clear and contain enough information on the who, when, and where of the study that the reader does not have to refer to the text to understand the table. Rows and columns should be clearly labeled. Footnotes are often used to clarify headings or abbreviations.

Graphs

Displaying data graphically is another effective way to summarize information. The type of data dictates the type of graph that should be used. The x-axis, also known as the horizontal axis or abscissa, typically represents the values of the variable of interest. The y-axis, also known as the vertical axis or ordinate, displays the number of cases, the rate of disease, or some other measure of the frequency of occurrence. As with a table, a graph should be clearly labeled (including a legend if it displays more than one factor) so that the reader does not have to read the accompanying text to interpret the graph.

EXHIBIT 33-1

A HYPOTHETICAL EXAMPLE TO CALCULATE MEASURES OF VARIATION OR DISPERSION

To determine the age as well as other demographic characteristics of female applicants to military service, a survey was conducted at a Military Entrance Processing Station. The ages of the first seven women chosen for the sample were 18, 19, 19, 20, 21, 22, and 24 years. For this example:

Mean = $\bar{x}$ = (18 + 19 + 19 + 20 + 21 + 22 + 24) / 7 = 20.4

Mode = 19

Median = 20

Range is 18 to 24 (6 years)

The standard deviation is calculated in the following way:

x_i	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
18	20.4	–2.4	5.76
19	20.4	–1.4	1.96
19	20.4	–1.4	1.96
20	20.4	–0.4	0.16
21	20.4	0.6	0.36
22	20.4	1.6	2.56
24	20.4	3.6	12.96
			25.72

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{25.72}{6}} = 2.1$$

Pie Graph

A pie graph or chart displays data in a circular fashion, comparing parts or segments of the data to the whole. Data are converted into percentages, adding up to 100%. The percentages are converted to degrees by multiplying the percentage by 3.6 (Figure 33-1).

Bar Graph

Discrete data can be displayed in a bar graph. An important feature of a bar graph is that the bars do not touch. The data are noncontinuous and the only possible values are the ones that are noted by each bar (Figure 33-2).

Histogram

Continuous data can be displayed in a histogram. The bars in a histogram touch because the data on the x-axis are continuous and each bar represents an interval of values and not just one value. If all bars represent intervals of the same width, then the height of each bar represents the relative frequency of each interval. The histogram provides a visual picture of the shape of the frequency distribution (Figure 33-3).

Arithmetic Scale Line Graph

A line graph is typically used when the x-axis represents time and the y-axis represents rates of disease. Each axis is measured in arithmetic units, with equal distances between the units (Figure 33-4).

Semilogarithmic Scale Line Graph

In a semilogarithmic scale line graph, the y-axis is based on a logarithmic scale and the x-axis on an

arithmetic scale. When displaying rates of disease, it is useful to use semilogarithmic graphs because a straight line indicates a constant rate of change, the slope of the line indicates the rate of increase or decrease, and parallel lines represent identical rates of increase or decrease (Figure 33-5).

SAMPLING

A critical component of any study design is the method used for selecting the study population. The target population is that group of individuals from which inferences about disease patterns are to be drawn. Typically, it is not possible or efficient to obtain data on all members of the target population. Instead, sampling procedures are used to select a suitable sample or subset of the target population.

In a probability sample, every individual has a known (usually equal) probability of being included in the sample. Therefore, generalizations to the target population can be made with a measurable amount of precision and confidence. In a nonprobability sample, probability theory may not apply and there is more opportunity for bias in selection of subjects.

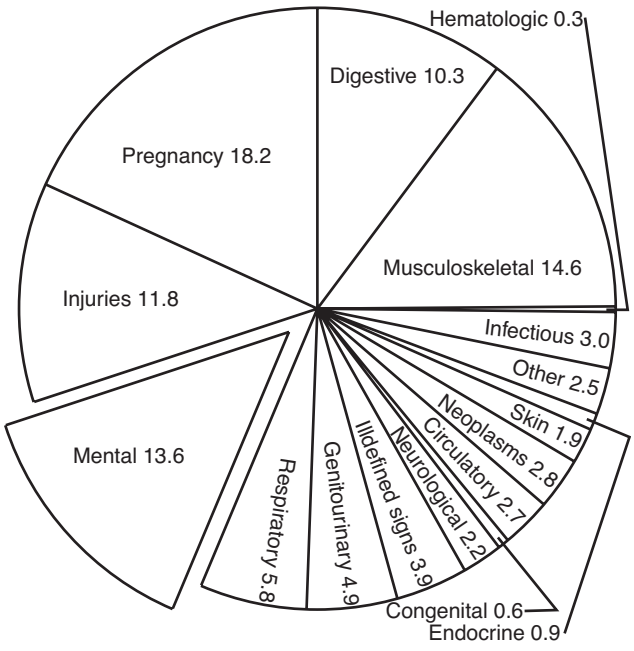

Fig. 33-1. Each wedge in this example of a pie graph represents percent of hospitalizations by 17 major diagnostic categories for active duty soldiers in 1997.
Source: Trends in hospitalizations due to mental disorders, US Army active duty soldiers. *Medical Surveillance Monthly Report*. 1998;4(5):15.

Nonprobability Sampling Designs

Consecutive Sampling

With this design, every individual in a given setting who meets the selection criteria is chosen over a specified time period. A consecutive sample can be drawn, for example, by using every recruit in-processing at a basic training post in a defined time period. If all individuals are studied in a given time period, the sample may be a good representation of the overall population, but there is no known probability of any given individual being included in the sample.

Convenience Sampling

Individuals from a population who are readily available are included in a convenience sample. Examples of convenience samples include sailors in a clinic waiting room or soldiers entering the post exchange. Selecting such a sample is inexpensive and easy. Results from this type of survey may be biased, however, as there is no assurance that individuals in this sample reflect the characteristics of the target population. At times, this method is used to obtain preliminary data or to generate hypotheses for future studies.

Judgmental Sampling

With this technique, individuals are handpicked to be in the study. Specific individuals are chosen for the sample because they are considered to be representative of the population of interest or possess specific selection criteria. The dangers of bias using this method can easily go unrecognized.

Quota Sampling

The composition of the survey population in this method is determined in advance. Then quotas are determined for individuals from various demographic categories, such as age and race. The only requirement is that the specified number of individuals in a given category be recruited. The basis for the quotas and the method of recruitment often

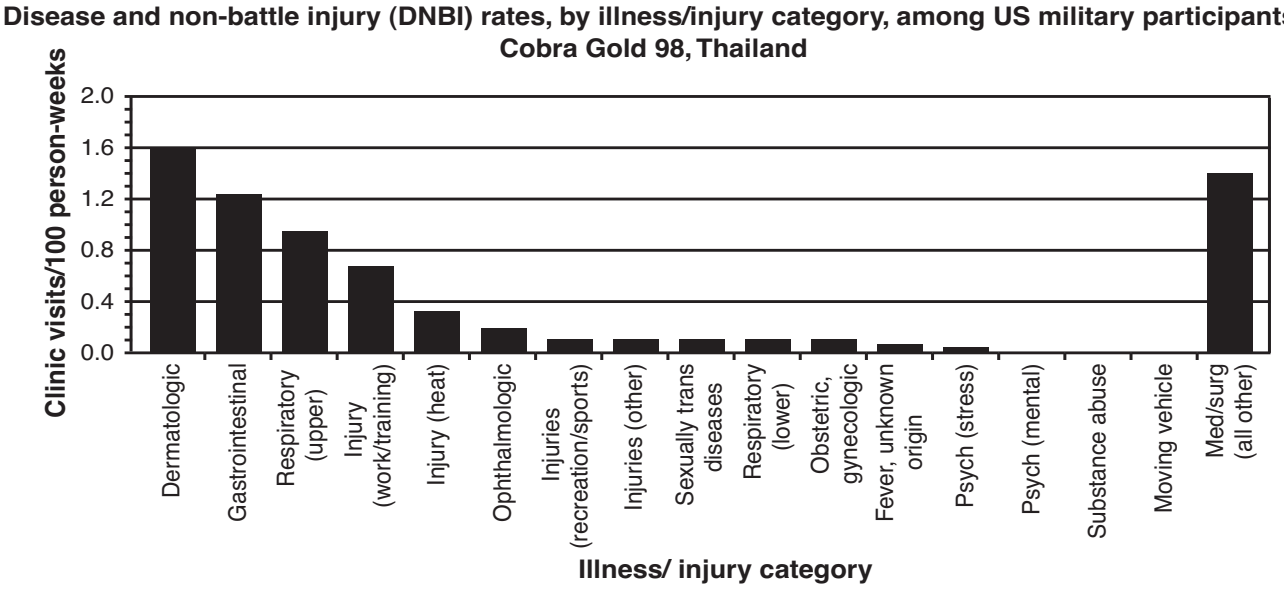

Fig. 33-2. This example of a bar graph shows the disease and nonbattle injury rates for US military participants in the Cobra Gold 98 exercise (Thailand). The data have been divided by illness and injury category.
Source: Morbidity surveillance during a joint multinational field training exercise (Operation Cobra Gold 98), Thailand. *Medical Surveillance Monthly Report*. 1998;4(6):3.

leads to bias since people are excluded from the study because of convenience factors in finding the designated number of subjects.

Probability Sampling Designs

Simple Random Sampling

The most elementary type of probability sample is the simple random sample. In this design, each person has an equal chance of being selected for the sample from the population under study. To select a random sample, the first step is obtaining a list of all individuals in the population. Each individual (or household or other basic unit) is called a sampling unit. A sampling frame is the list of all the sampling units in the study population. At times, it may be very difficult to generate such a list. Telephone directories or personnel lists are examples of lists that may be used. It is important to consider possible inaccuracies in a given list. If the list is incomplete or not updated regularly, then a biased sample may be drawn. Once the list is obtained, each person is assigned a number and the numbers are then selected at random, usually using a table of random numbers. The sample then consists of the sampling units that are selected. By chance, even a simple random sample may not adequately represent the underlying target population.

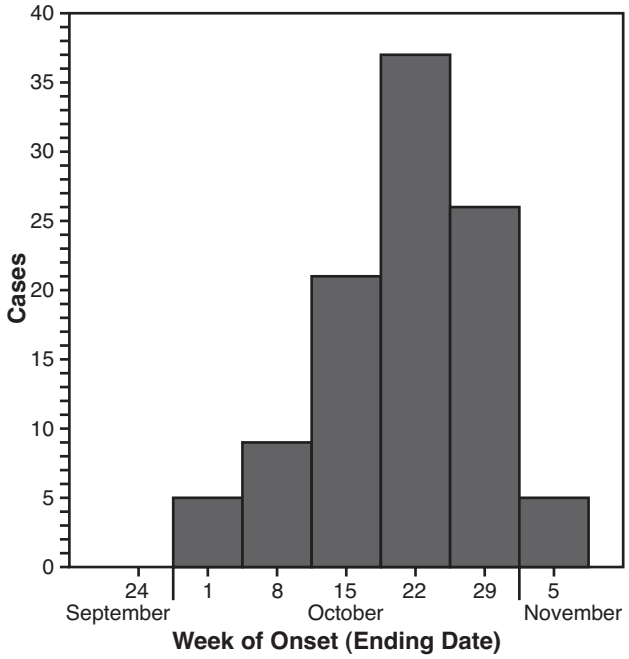

Fig. 33-3. This example of a histogram shows cases of febrile illness among US military personnel in Haiti from September 18, 1994, to November 5, 1994. The height of each bar represents the frequency. (The graph excludes three cases for which dates on onset were unknown.)
Source: Centers for Disease Control and Prevention. Dengue fever among U.S. military personnel—Haiti, September-November, 1994. *MMWR*. 1994;43:845–848.

Fig. 33-4. This example of an arithmetic line graph shows monthly active duty hospitalization rates for disease, injury, battle casualty, and the total rates in Operation Joint Endeavor (Bosnia) from December 1995 to April 1997.
Source: *Medical Surveillance Monthly Report*. 1997;3(2):8.

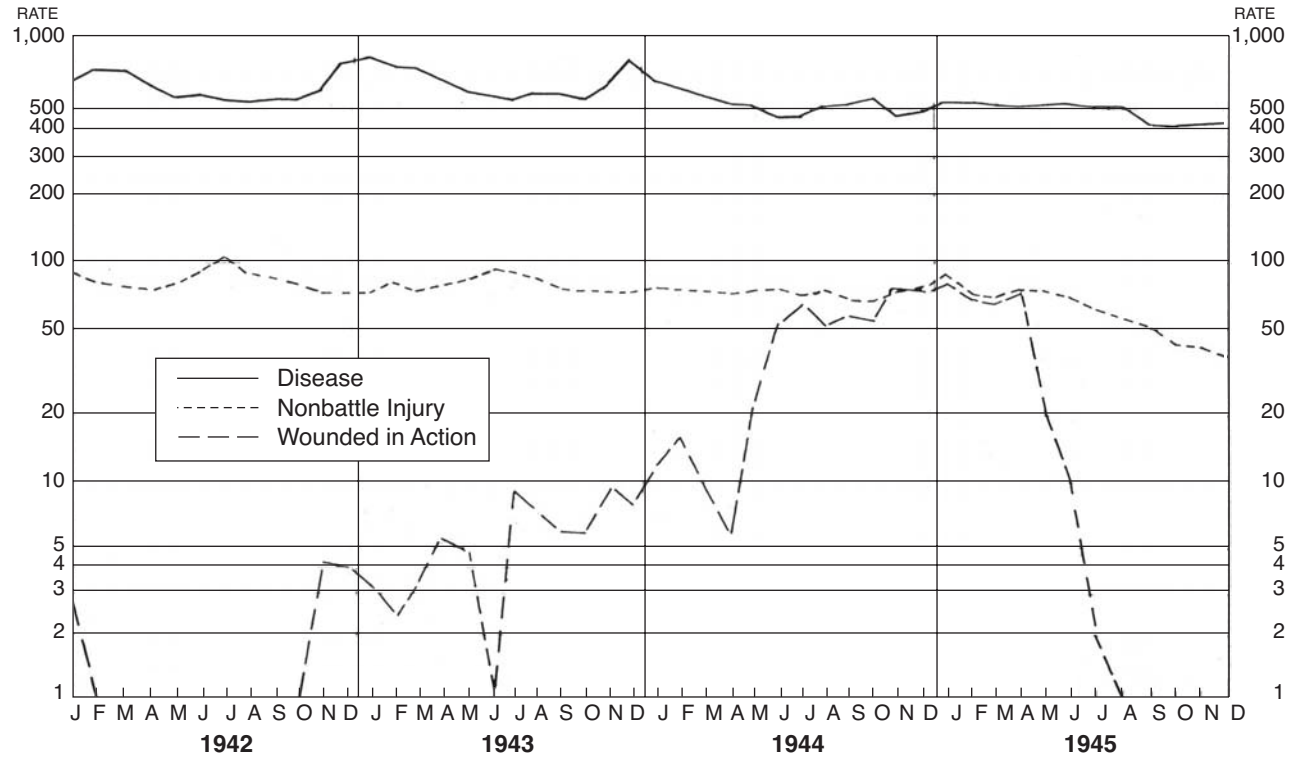

Fig. 33-5. This example of a semi-logarithmic scale line graph shows monthly admission rates for disease, nonbattle injury and wounded in action worldwide from 1942 to 1945 among US Army troops.
Source: Lada J, Reister FA, eds. *Medical Statistics in World War II*. Washington, DC: Office of the Surgeon General, Department of the Army; 1975: frontispiece.

Systematic Sampling

This technique is more widely used than simple random sampling because it does not require a complete listing of the study population or sampling frame. For a systematic sample, a fraction of the population to be studied is chosen. Instead of using a table of random numbers, the investigator chooses every k^{th} individual. If one-tenth of the population is to be surveyed, then every tenth person in a sequence will make up the study population. To begin the selection, a random starting point is chosen. For example, a systematic sample of pregnant soldiers who deliver in a medical treatment facility could be selected such that every 5th woman admitted to the maternity ward over a specified time period is included. Knowledge about all women who are admitted over a given time period is not necessary. In fact, it would be impossible to generate such a listing of these women before the study begins. To obtain an unbiased sample with this method, there must not be any periodic ordering of the sampling units.

Stratified Sampling

In a stratified random sample, the population is divided into distinct subgroups or strata based on important characteristics, such as age or race, and then a simple random sample or systematic sample is selected within each stratum. An individual only appears in one stratum, and each stratum is designed to be homogeneous with respect to the characteristic being studied. This technique is frequently used to increase the numbers of persons from a specific stratum of the population and therefore may improve the efficiency of the sampling design.

Cluster Sampling

In cluster sampling, the population is divided into large subgroups or clusters that are not homogeneous in composition. The clusters then become the sampling unit and a random sample of clusters is obtained. All persons in the cluster are in the study or a random sample of individuals from the cluster may be drawn. The entire population does not have to be enumerated in advance. For example, logistically it may be too difficult to obtain a random sample of soldiers engaged in a field operation. However, it may be possible to select a random sample of Army units and then randomly select soldiers from those units. This technique is used for conducting immunization surveys in developing countries.

Multistage Sampling

Multistage sampling is more complicated than other sampling methods and involves randomly choosing, in stages, a series of clusters or subunits of a population. For example, multistage sampling would be useful when conducting a survey in an Army division on the use of personal protective measures. The first stage in this sampling scheme, or the primary sampling unit, may be a simple random sample of the battalions in the division. Then, the second stage, or the secondary sampling unit, would consist of a simple random sample of the companies within the selected battalions (a smaller cluster). The tertiary sampling unit would be a random sample of the platoons within the chosen companies. Then, within the platoon selected, a random sample of soldiers would be taken. An advantage of this scheme is that complete enumeration is necessary only for each chosen platoon.

Standard Error

A standard deviation is the spread of individual observations around the mean in a single sample. The standard error is the standard deviation of the means of repeated samples randomly drawn from the same population. The standard error (SE) of a mean of the simple random sample is calculated as follows:

$$\text{Standard Error of the Sample Mean} = \text{SE} = \frac{s}{\sqrt{n}}$$

where s is the estimated standard deviation from the sample and n is the sample size.

When a sample is chosen in an unbiased fashion, then the only source of error is random variation. The size of the sample and the heterogeneity of the population influence this variation. In an unbiased sample, as the sample gets larger, it is more likely that the sample estimate will be close to the value obtained for the target population. In other words, the larger the sample, the more precise the estimate will be. A precise estimate gives a value that is likely to be repeated if the sampling were done again and again. A smaller standard error implies greater precision and results from a larger sample.

Sampling can become quite involved and may require expert assistance in the planning stages of a study. The calculation of the standard error for the other probability sampling techniques becomes quite complex. With each technique, the investiga-

tor must be aware of biases that may be introduced in the selection of a study population. A large sample size will not correct for a biased sample, and the standard error reflects only random variation and does not address bias.

The costs involved in obtaining a sample may limit the investigator's choice of sampling method. These costs must be balanced against the efficiency of the design. In general, the method that produces a smaller standard error for a given cost should be used. Further discussion of sampling techniques can be found elsewhere.[7]

MAJOR TYPES OF STUDY DESIGNS AND MEASURES OF ASSOCIATION USED IN EPIDEMIOLOGY

Much of epidemiology assesses the relationship between exposures or risk factors and disease or injury occurrence. In experimental studies, the investigator has control over some factor or exposure that can be altered. This variation in an exposure can then be associated with different outcomes. A specialized form of the experimental approach is the randomized clinical trial. Such studies rarely are feasible during a combat deployment. In contrast, training exercises commonly present opportunities to conduct clinical trials, as noted in a study to determine the efficacy of new treatments for diarrhea.[9]

The vast majority of epidemiologic investigations rely on the observational approach. With this approach, there is no manipulation of a risk factor, but rather the risk factors are evaluated as they vary naturally from one individual to another. Different outcomes are then observed under natural conditions. Control of extraneous variables is accomplished in the design of the study or in data analysis. Observational studies can be descriptive or analytic. Descriptive studies characterize the distribution of disease in terms of attributes of person, place, and time. Analytic studies attempt to evaluate disease associations with specific factors. An important component of these designs is the availability of a comparison or control group.

Case Series

Case series, also known as case reports, are collections of notable cases, which, for example, may present at a medical clinic during deployment. Case series also can result from medical surveillance activities. They do not constitute an analytic study as there is no comparison or control group, nor are they good as descriptive studies since they do not report disease rates. They can, however, provide insight into potentially important characteristics of the disease. For example, several soldiers with acute respiratory disease (ARD) were discovered at Fort Dix in 1976, one of whom subsequently died. Further investigations determined that a new type of influenza, swine influenza A, was the cause of this morbidity and mortality.[10]

Ecological Studies

In a study with an ecological design, data are not collected on specific individuals, but rather aggregate data are collected on groups of individuals. These studies also are known as correlational studies because a characteristic of a group usually is plotted against a characteristic of another group. An international comparison of risk factors and disease based on country-level data is an example of the type of data used for ecological analyses. Such analyses are usually done as a first step in assessing whether a public health problem may exist. Interpreting these data can be difficult as an undefined factor, on which data were not collected, may explain the observed association. Moreover, because the data are grouped, it is not known whether the individuals with the disease are the same individuals exposed to the risk factor. For example, an ecologic study was conducted to evaluate the relationship between alcohol consumption levels and mission readiness indicators among shipboard sailors. Data on alcohol consumption and various medical and legal indicators of mission readiness were not obtained on the same individual. Rather aggregate data based on the platforms ships were used in the analysis.[11]

Descriptive Studies

Descriptive studies describe the prevalence, incidence rate, or mortality rate of disease in a population according to characteristics of person (Who gets the disease?), place (Where do they live or work or travel?), and time (When does the disease occur?). Who gets a disease can be described by factors such as age, sex, rank, military occupational specialty, unit, or immunization status. Where disease is found can be addressed by information on international, national, or local comparisons; urban and rural differences; travel history; and altitude or climate. Information on how the pattern of disease changes with time (secular trends) or the impact of seasonal fluctuations describe when the illness occurs. Results from descriptive studies are used to

assess the need for health services and to generate hypotheses about factors that may cause disease. The etiologic importance of these factors can be further evaluated using other study designs. For example, a medical surveillance system was used to track weekly incidence rates of disease and nonbattle injuries among multinational peacekeepers deployed to Haiti in 1995. Results from the surveillance system were used to direct health care resources to prevent and treat casualties.[12]

Cross-Sectional Studies

Cross-sectional studies, also known as prevalence surveys, examine the relationship at one point in time between risk factors, such as environmental exposures or demographic factors, and existing disease. A problem with the interpretation of findings from a cross-sectional study is that it can be difficult to determine the sequence of events. For example, if results from a cross-sectional survey showed that a measure of stress was associated with duodenal ulcers, it cannot be assumed that the stress preceded the onset of ulcers since both the risk factor and the disease were evaluated at the same time. Conceivably, the ulcers might have been responsible for the stress rather than the reverse. For some exposures that do not vary with time, such as genetic factors, a cross-sectional study can provide meaningful information on an exposure–disease relationship. This study design is also useful for studying chronic conditions, such as arthritis or chronic respiratory disease, where the onset of disease is difficult to determine. A cross-sectional study is the only one that estimates the prevalence of disease, an important measure for health care planning purposes. Because these studies determine prevalence rather than incidence of disease, they can provide only limited information on etiologic factors. Individuals who are surviving with disease are included in cross-sectional studies and individuals who have a rapid recovery or die quickly from the disease are often not included and thus are underrepresented. Results from these studies, however, can be further evaluated in cohort and case-control studies.

In 1990, Smoak and colleagues[13] conducted a cross-sectional study of healthy young adults (404 females and 534 males) at induction into the US Army at Fort Jackson, SC. Serum collected on all individuals was used to determine the seroprevalence of *Helicobacter pylori* infection. Demographic data were abstracted from accession records. The associations between antibody levels and several demographic factors were then assessed.

Case-Control Studies

The defining feature of a case-control study is that study subjects are selected as individuals with disease (cases) and without disease (controls), then data on past exposures that pertain to etiologic factors are collected in both groups. Cases can be ascertained from several sources, including hospitals, outpatient clinics, and disease surveillance activities. Possible sources of controls include hospital patients who do not have the disease of interest, friends or relatives of the case, and a random sample of a population such as servicemembers living in the same barracks or assigned to the same unit.

There are several methodological concerns and potential biases that must be considered when conducting a case-control study. It is assumed that the cases are representative of or include all cases that come from the source population and the controls are representative of all people without disease in the same source population. The definition of a case must be clearly specified with criteria for who is eligible to be included in the study. It is preferable to include only incident cases rather than prevalent cases, so that factors that are associated with the occurrence of disease can be evaluated rather than factors that are associated with surviving with the disease. Because exposure is determined after the identification of disease, there is the potential bias that cases may recall events differently than controls or that an interviewer may query cases differently than controls. To obtain better comparability of cases and controls and to control for potential confounding, cases and controls may be "matched" on characteristics such as age or sex that are already known to be associated with both exposure and disease. Confounding and potential biases are discussed in greater detail later in this chapter.

Although exposure is ascertained after the onset of disease in case-control studies, there are several reasons why case-control studies are desirable. These studies generally require fewer resources, fewer study subjects, and less time to collect data compared with cohort studies, which are described later in the chapter. They also are more feasible for the study of rare diseases (Exhibit 33-2).

Calculation of the Odds Ratio

When the exposure information and disease status are coded as dichotomous variables, the data from a case-control study can be arrayed in a 4-fold or 2 x 2 contingency table. The table in Exhibit 33-3 summarizes the essential data obtained in a case-control

EXHIBIT 33-2

ADVANTAGES AND DISADVANTAGES OF A CASE-CONTROL STUDY

Advantages

- Smaller number of study subjects required compared to cohort studies
- Relatively inexpensive
- Relatively quick results
- Several exposures can be evaluated
- Efficient for studying rare diseases
- Efficient for studying diseases with long latency

Disadvantages

- Temporal relationship between exposure and disease may be difficult to establish
- Possible bias in the selection of cases and controls
- Possible bias in ascertainment of exposure (recall bias)
- Cannot calculate incidence rates individually for exposed and unexposed groups (estimates only the ratio)

EXHIBIT 33-3

HOW TO CALCULATE AN ODDS RATIO

Risk Factor	Disease: Cases	Disease: Controls	
Exposed	a	b	a + b
Nonexposed	c	d	c + d
	a + c	b + d	

a = the number of individuals who are exposed and have the disease

b = the number of individuals who are exposed and do not have the disease

c = the number of individuals who are nonexposed and have the disease

d = the number of individuals who are nonexposed and do not have the disease

study. A relative risk (RR), also known as a rate ratio, is defined as the risk of disease in the exposed divided by the risk of disease in the nonexposed and cannot be directly calculated using data from a case-control study because the incidence of disease cannot be determined. In fact, during the design of a case-control study, the investigator arbitrarily determines the estimated number of cases to be studied.

An estimate of the RR, known as the odds ratio (OR), can be calculated.[14] An odds is defined as the likelihood of an event happening versus the likelihood of the event not happening. According to the notation in Exhibit 33-3, the odds of exposure in cases is given by this formula: $(a/a+c)/(c/a+c) = a/c$. Similarly, the odds of exposure in controls is b/d. Therefore, an odds ratio is defined in this way:

$$OR = \frac{\text{Odds of Exposure Among the Diseased}}{\text{Odds of Exposure Among the Nondiseased}} = (a/c)/(b/d) = ad/bc$$

An RR and an OR are known as "measures of association" as they measure the association between an exposure and disease. An RR and an OR of greater than one implies a positive association of the disease with exposure to the factor (ie, exposure leads to an increased risk of disease). An RR and an OR of less than one implies a negative asso-

EXHIBIT 33-4

AN EXAMPLE OF THE CALCULATION OF AN ODDS RATIO

	Cases	Controls
Ever in Vietnam	45	145
Never in Vietnam	172	454

Odds ratio = ad/bc
= $(45 \times 454)/(145 \times 172)$
= $20{,}430/24{,}940$
= 0.82

Source: Kang H, Enzinger FM, Breslin P, Feil M, Lee Y, Shepard B. Soft tissue sarcoma and military service in Vietnam: a case-control study. *J Natl Cancer Inst*. 1987;79:693–699. Published erratum: *J Natl Cancer Inst* 1987;79:1173.

ciation of the disease with exposure to the factor (ie, exposure leads to a decreased risk of disease or has a protective effect against disease). Finally, an RR and an OR of one implies no association between disease and exposure to the factor. Thus, the RR indicates the strength of the association between a factor and a disease and is an important measure in studies of disease etiology.

For example, a case-control study was conducted to examine the association of soft tissue sarcomas with military service in Vietnam by interviewing 217 men with soft tissue sarcoma and 599 controls.[15] The data collected and the odd ratio that can be calculated from them are shown in Exhibit 33-4.

Based on these data, Vietnam veterans had a lower risk of soft tissue sarcomas than those men who had never served in Vietnam, as the OR is less than one. These data can also be interpreted by stating that there was an 18% lower risk of sarcoma in men who served in Vietnam compared with men who did not serve in Vietnam.

EXHIBIT 33-5

ADVANTAGES AND DISADVANTAGES OF A COHORT STUDY

Advantages

- Ideal time sequence (exposure precedes disease)
- Lack of bias in the measurement of exposure
- Several disease outcomes can be evaluated
- Efficient for studying rare exposures
- Yields incidence rates in the exposed and unexposed populations

Disadvantages

- Often impractical for studying rare diseases
- Long follow-up period may be required for outcomes with long latency
- Relatively expensive
- Problem of loss of subjects to follow-up
- Exposure levels may change over time, requiring sophisticated follow-up and analysis

Cohort Studies

Cohort studies also have been referred to as incidence, prospective, follow-up, or longitudinal studies. Typically a group of individuals (a cohort) that is free of the disease under investigation is assembled, evaluated to determine exposure history and other risk factors, and then followed forward in time to determine the occurrence of disease. Some designs provide for repeated examinations of study subjects over the course of the follow-up period. The development of disease is then observed in the various exposure groups. Incidence of disease can be calculated for those who have been exposed to a risk factor and for those who have not. If the entire cohort has been exposed, then the cohort can be compared with the general population or some other nonexposed comparison group or well-studied cohort. A critical feature of this design is the comparison between a group of individuals defined as exposed and a group defined as nonexposed. A major strength of cohort studies is that the exposure is ascertained before the onset of disease. Other advantages of the cohort design include the ability to calculate incidence rates directly in the exposed and nonexposed populations and to evaluate several disease outcomes in relation to the defined exposures. These studies, however, are often costly, require large numbers of subjects, may require a long follow-up period, and are subject to attrition problems (known as "lost to follow-up") that can bias the generalizability of results (see Exhibit 33-5).

Cohort studies also can rely on historical records. This design is sometimes referred to as a nonconcurrent follow-up study or a historical cohort study. For example, a cohort of Navy shipyard workers in the 1940s could be assembled based on personnel records and then divided into asbestos exposure groups based on job title. These workers could then be followed to the present for outcomes such as lung cancer. An important limitation in this design might be the absence of data on smoking exposure since this information might not have been captured in the historical records.

Calculation of the Risk Ratio and Rate Ratio

In a cohort study, the ratio of two proportions (based on cumulative incidence data) is called a risk ratio (Exhibits 33-6a and 33-6b), while the ratio of two rates (based on incidence density sampling) is

EXHIBIT 33-6a

HOW TO CALCULATE A RISK RATIO

The RR based on cumulative incidence data from a cohort study can be readily calculated using a 2 x 2 table:

Risk Factor	Disease: Yes	Disease: No	
Exposed	a	b	$a+b=N_1$
Nonexposed	c	d	$c+d=N_0$
	$a+c=M_1$	$b+d=M_0$	

Risk Ratio = $[a/(a + b)]/[c/(c + d)]$

EXHIBIT 33-6b

HOW TO CALCULATE A RATE RATIO

The RR for data based on incidence density rates is displayed in the following way (where PT stands for person-time):

Risk Factor	Disease: Yes		
Exposed	a		PT_1
Nonexposed	c		PT_0

Rate Ratio = $(a/PT_1)/(c/PT_0)$

called an incidence rate ratio (IRR). The RR and IRR do not measure the probability that someone with a risk factor will develop disease; they compare the risk or rate of disease in an exposed group with the risk or rate of disease in a nonexposed group. The calculation of the IRR is illustrated in Exhibit 33-7. In a historical cohort study at four Army training centers during a 47-month period, incidence rates of febrile acute respiratory disease (ARD) were compared between basic trainees in modern, energy-efficient barracks and those in old barracks. These data were collected at Fort Jackson, South Carolina. These data show that the crude rate of ARD among basic trainees at Fort Jackson who lived in the modern barracks is 1.45 times greater than the rate of ARD among basic trainees who lived in the old barracks.

Attributable Risk in the Exposed Group

In addition to the RR, the attributable risk in the exposed group (AR_e) also can be calculated from a cohort study. The AR_e, also known as the risk difference or excess risk, is defined as the rate of disease in the exposed group minus the rate in the nonexposed

EXHIBIT 33-7

AN EXAMPLE OF THE CALCULATION OF A RATE RATIO

Risk Factors	Admissions for Acute Respiratory Disease at Fort Jackson, 1982-1986: Yes		Person-time
Modern Barracks	3,355		451,294 trainee-weeks
Old Barracks	3,312		647,056 trainee-weeks
Total	6,667		1,098,350 trainee-weeks

$$\frac{\text{Rate of Disease in Modern Barracks}}{\text{Rate of Disease in Old Barracks}} = \frac{3{,}355/451{,}294}{3{,}312/647{,}056} = \frac{.00743}{.0051} = 1.45$$

Source: Brundage JF, Scott RM, Lednar WM, Smith DW, Miller RN. Building-associated risk of febrile acute respiratory diseases in army trainees. *JAMA.* 1988;259:2108–2112.

group. From the example shown in Exhibit 33-7, the rate difference was calculated in this way:

AR_e = Incidence Rate in Exposed – Incidence Rate in Nonexposed
= 3,355/451,294 – 3,312/647,056
= .00743 – .00512 = 0.0023
= .23 admissions/100 trainee-weeks at Fort Jackson

This is the excess amount of disease over baseline (estimated by using the nonexposed rate) that is attributed to the exposure. Thus, this measure answers the question, "In Army trainees who live in modern barracks, how many of the admissions for ARD are attributed to living in modern barracks?" If this exposure is eliminated, then the attributable risk is the amount of decrease in the disease rate that is expected.

The attributable risk proportion is the proportion of the incidence in the exposed population that is attributed to the exposure. It also can be presented as a percent (AR_e%). This measure defines the percentage of disease in an exposed population that would be prevented by eliminating the exposure.

$$AR_e\% = \left(\frac{\text{Incidence Rate in Exposed Group} - \text{Incidence Rate in Nonexposed Group}}{\text{Incidence Rate in Exposed Group}}\right) \bullet 100$$

Using the data presented above,

$$AR_e\% = \left(\frac{.00743 - .00512}{.00743}\right) \bullet 100 = 31\%$$

This formula can also be calculated using the RR:

$$AR_e\% = ([RR-1]/RR) \bullet 100 = [(1.45-1)/1.45] \bullet 100 = 31\%$$

The attributable risk percent (but not the attributable risk itself) also can be calculated using the OR generated from a case-control study.

Attributable Risk in the Total Population

The population attributable risk (PAR) is the incidence of disease in the total population that is attributed to the exposure. Since this measure is based on the total population, it takes into account both exposed and nonexposed individuals. Using the data obtained from Fort Jackson, the PAR is calculated in this way:

PAR = Incidence in the Total Population – Incidence in the Nonexposed Group
= 6,667/1,098,350 – 3,312/647,056
= .00607 – .00512
= .00095
= .95 admissions per 1,000 trainee-weeks

The incidence in the total population must be known to use this formula. The population attributable risk percent (PAR%) is the percent of the incidence in the total population that is attributed to the exposure and can be calculated by the formula:

$$PAR\% = \left(\frac{\text{Incidence in the Total Population} - \text{Incidence in Nonexposed Group}}{\text{Incidence in the Total Population}}\right) \bullet 100$$

From the above example,

$$PAR\% = \left(\frac{.00005}{.00607}\right) \bullet 100 = 16\%$$

It can also be calculated using the RR and an estimate of the proportion of the population that has the exposure (p_e):

$$PAR\% = p_e(RR-1)/[p_e(RR-1) + 1]$$

$$p_e = \frac{451,294}{1,098,350} = 0.41$$

$$\frac{.41(1.45\text{-}1)}{.41(1.45\text{-}1) + 1} = \frac{.185}{1.185} \bullet 100 = 16\%$$

If the OR is used, then the proportion exposed in the control group is substituted for the proportion exposed in the population at large. Other terms that are used to describe this measure are the attributable fraction in the population or etiologic fraction in the population. This measure answers the question, "What proportion of ARD admissions at Fort Jackson can be attributed to living in modern barracks?"

Measures of attributable risk assume that there is a causal relationship between exposure and disease and, therefore, should be calculated only when there is sufficient evidence to imply causality.

Experimental Studies and Randomized Clinical Trials

The major type of experimental study in medicine is the clinical trial. In a controlled clinical trial, the investigator manipulates or intervenes with one

group (the treatment group) and withholds intervention or gives a placebo to another group. Random allocation of patients to various treatment or nontreatment groups is the central tenet of randomized clinical trials. The randomization of patients attempts to make the groups comparable at the onset of the study, so any differences noted between the two groups can be ascribed to the intervention. This design is commonly used for testing drugs or vaccines, but it can be used to test any intervention against a control. Examples of studies done during US military deployments include the treatment of traveler's diarrhea with ciprofloxin and loperamide[9] and the efficacy trial of doxycycline chemoprophylaxis against leptospirosis.[16] The reader is referred to other texts[17] for further discussion of this study design.

BIAS

Definition

Bias is manifest in a study when study results differ systematically from the true values. It is also known as systematic error, as opposed to random error or chance. Bias can be introduced at any stage of an investigation, including the design, the conduct, or the inferences drawn from the results. Because of this systematic error, the strength of an association can be underestimated or overestimated. Types of bias in epidemiologic studies can be broadly classified into selection bias, information bias, and confounding. As the field of epidemiology has evolved and study designs have become more elaborate, new types of bias have been described.[18] The following is a listing of some of the more common types of selection bias and information bias that can be found in epidemiologic studies.

Types of Selection Bias

Ascertainment Bias

This bias is the systematic error that arises from the method used to identify individuals for a study. In cohort studies, selection bias occurs if individuals are included in the study based on their disease status. In case-control studies, selection bias occurs if the likelihood an individual is selected for study is based on exposure status. For example, an outbreak investigation may base conclusions on cases that are not very ill, as these patients may be more willing to participate and be available for an interview. These milder cases of disease may differ from more serious cases with respect to exposure factors.

Healthy Worker Effect

Individuals who are employed or are inducted into military service are more physically fit and healthier than the general population, resulting in lower disease rates when compared with the total population.

Volunteer Bias

This error is the result of systematic differences between study subjects who volunteer to participate in a study or who return a questionnaire and those who do not.

Types of Information Bias

Detection Bias

This is a systematic error due to methods of diagnosis or verification of cases. For example, a pathologist may make the diagnosis for some cases, while the diagnosis of other cases is based solely on medical records.

Interviewer Bias

This systematic error is the result of interviewers not questioning study subjects in a uniform manner.

Measurement Bias

This bias arises from inaccurate quantification or classification of exposures or outcomes. It can result from the subjectivity of the measurement scale.

Recall Bias

This error is the result of differences in either the truthfulness, accuracy, or completeness of participants' recall of events. Recall bias is due to systematic differences in recall between the groups being compared and often reflects the greater thought a sick person will have given to possible explanations for his or her plight.

Avoiding or Reducing Bias

There are no statistical methods to control for selection or information bias that is introduced into a study. If a bias is present, the results of the study

are very difficult to interpret. There are study design features, however, that can help reduce or avoid some biases. These include blinding the interviewers to the participant's diagnosis; blinding the volunteers in a randomized clinical trial so that they are unaware of their treatment assignment; blinding study staff who review or code data so that they are unaware of the diagnosis or treatment; achieving high response rates in studies (over 80%); comparing the characteristics of individuals who are lost to follow-up with those who remain in the study; establishing explicit criteria for assessing exposures and outcomes; obtaining information about exposures from independent sources that are unaffected by memory; and recognizing potential confounding variables and controlling for them in the design or analysis of the study (as described below). If there is concern that bias is present in a study, it is often possible to estimate the direction of the bias (ie, whether the bias results in an apparent association towards the null value of 1.0 or away from the null). In a case-control study, for example, if some cases do not recall their smoking history with the same completeness as the controls, then the smoking histories of the two groups may appear to be more similar than they are. This problem with recall would result in a bias toward the null.

Confounding

Definition

Confounding bias occurs when the observed relation between the risk factor and the outcome is distorted by the influence of a third variable, the confounder. A confounding variable must be associated with both the risk factor of interest and the outcome. Confounding can be introduced into a study because of the complex relationship between several exposures or demographic factors and the outcome. Thus, as a result of confounding, an apparent association between a specific exposure and disease may be noted when no real association exists. Confounding can also lead to an overestimate or underestimate of the true measure of association between the risk factor and the disease. Statistical tests are not used to assess whether confounding is present in a study; several other techniques (described below) are available to the investigator to assess and control confounding.

A hypothetical example adapted from a study of injury in male and female Army trainees based on cumulative incidence rates will demonstrate these concepts more clearly. Results from a cohort study showed that female trainees were two times more likely to develop a stress fracture as compared with male trainees. The data are presented in Exhibit 33-8.

To evaluate the association between sex and training injury, factors that might confound the relationship need to be assessed. Physical fitness and various anthropometric measurements are possible confounders because these factors are known to be associated with both sex (the exposure) and stress fractures (the disease). To consider the possible confounding effect of fitness, the data shown above are stratified into tables according to the aerobic fitness of the trainee (1-mile run times). For this example, the data are divided into two tables, based on those trainees who had fast 1-mile run times and those who had slow 1-mile run times (Exhibit 33-8). The overall

EXHIBIT 33-8

ASSOCIATION BETWEEN SEX AND THE RISK OF STRESS FRACTURES AMONG ARMY TRAINEES

Crude

Sex	Stress Fracture		
	Yes	No	Total
Women	66	434	500
Men	52	748	800
Total	118	1,182	1,300

RR = (66/500)/(52/800) = .132/.065 = 2.0

Stratified

Trainees with fast 1-mile run times

Sex	Stress Fracture		
	Yes	No	Total
Women	6	94	100
Men	30	570	600
Total	36	664	700

RR = (6/100)/(30/600) = .06/.05 = 1.2

Trainees with slow 1-mile run times

Sex	Stress Fracture		
	Yes	No	Total
Women	60	340	400
Men	22	178	200
Total	82	518	600

RR = (60/400)/(22/200) = .15/.11 = 1.4

or crude RR is 2.0. The RRs within each stratum are both less than the crude RR and are similar to each other. Therefore, the association is uniform between the strata. If the stratum-specific RRs are averaged, then the effect of fitness level is removed (ie, controlled for). As noted above, fitness met the definition of a confounder because it was related to both sex (the exposure) and stress fractures (the disease). Among those without stress fractures, 14% (94/664) of fit trainees were women while 66% (340/518) of nonfit trainees were women. Moreover, men who were fit had a lower rate of stress fractures compared with men who were not fit [5% (30/600) versus 11% (22/200)]. These data are used later in this chapter to illustrate a common method to control for confounding in the analysis of data.

Controlling for Confounding in the Design of a Study

Randomization

Randomization is a method of allocating individuals to groups based on a predetermined plan, such as one based on a table of random numbers. The goal is to make comparison groups similar at the start of an investigation. This technique is used in clinical trials to randomly distribute study subjects between the treatment and placebo group. A major advantage of randomization is that it can control for both known and unknown confounders. The disadvantages include the difficulty in maintaining the randomization scheme once it is established.

Restriction

To control for confounding in the design of a study, participants can be restricted to a particular category of the confounding variable. For example, if smoking is a proposed confounder, then the study participants can be restricted to nonsmokers. Some problems with this approach include a reduction in the number of eligible subjects and a decrease in the generalizability of the results.

Matching

Study participants can be matched or "balanced" with respect to potential confounding variables, thereby removing the effect of confounding. This approach is frequently used in outbreak investigations and case-control studies. If a factor such as age is already known to be related to both exposure and disease, then the cases and controls can be matched on age. Matching in this context controls for the effects of age only if the analysis stratifies on the matching factor. Once a factor is used for matching purposes, its role as an etiologic factor cannot be evaluated because the cases and controls have been chosen to be similar with respect to this factor.

In pair matching, each identified case has a corresponding control chosen from the pool of eligible controls such that the case and control pair will be similar with respect to the matching variables. During the analysis, the pair matches must be maintained. For outcome exposure variables, the McNemar's chi-square test statistic is computed. For continuous outcome variables, the paired t-test is used. Matched pair analysis may increase the statistical power of a study. As noted above, however, once a factor becomes a matching variable, the effect of this variable on disease cannot be analyzed. At times, it is difficult to find matches, and this problem can greatly increase the cost of the study. There also is the danger of overmatching, which occurs when several factors are controlled for, and differences between cases and controls are, therefore, minimized.

Frequency matching refers to matching groups on the basis of the prevalence of confounders. For example, if 25% of the cases are in the age group 20 to 24 years, then controls will be chosen such that 25% of the controls fall in this category. With frequency matching, the data must be stratified by these matching variables during the analysis. The reader is referred to Kahn and Sempos[7] or Selvin[8] for a more-detailed discussion on matching.

Controlling for Confounding in the Analysis of the Study

Stratified Analysis

Controlling for confounding during the analysis of a study can be accomplished using many different techniques. The data are divided into strata based on the confounding variable, and the analysis is done separately for each stratum. As shown in Exhibit 33-8, to assess the relationship between sex and stress fractures, the data were divided into strata based on fitness level, the confounding variable. In this example, there were two strata—fast 1-mile run times and slow 1-mile run times.

Mantel-Haenszel Summary Odds Ratio

The Mantel-Haenszel summary OR (OR_{MH}) can be used to obtain a uniform OR that takes into account the effect of the confounding. To calculate an OR_{MH}, several 2 x 2 tables (*i* tables) are constructed according to the number of levels of the stratified variable. It is assumed that the true OR in each of the tables is uniform and that the variability in the OR that is observed in each stratum-specific table is due entirely to random error. The following is the formula for OR_{MH} with the weight for each table equal to $b_i c_i / T_i$:

$$OR_{MH} = \frac{\sum a_i d_i / T_i}{\sum b_i c_i / T_i}$$

where a_i, b_i, c_i, and d_i are cell counts for the i^{th} stratum, T_i is the total in the i^{th} stratum and the sum is across the strata of the confounder. For cohort studies with count denominators, the RR_{MH} is calculated in this way[19]:

$$RR_{MH} = \frac{\sum a_i(c_i + d_i) / T_i}{\sum c_i(a_i + b_i) / T_i}$$

For cohort studies with person-year denominators, the IRR_{MH} is calculated in this fashion[19]:

$$IRR_{MH} = \frac{\sum a_i(PT_{0i}) / T_i}{\sum c_i(PT_{1i}) / T_i}$$

In Exhibit 33-8, where the data were stratified into two tables based on fitness level, the RR_{MH} is computed as follows:

$$RR_{MH} = \frac{[6(30 + 570)/700] + [60(22 + 178)/600]}{[30(6 + 94)/700] + [22(60 + 340)/600]} = 1.3$$

To assess the role of confounding on the relation between sex and stress fractures, the adjusted RR_{MH} of 1.3 is then compared with the unadjusted RR of 2.0. The investigator must use his or her judgment to determine if the unadjusted or adjusted RR should be reported. As a rule, if there is more than a 10% change in the RR_{MH} compared with the crude RR, then the RR_{MH} should be presented; in this example, the RR_{MH} would be reported as it is the best estimate of the association between sex and stress fractures.

Of note, if the RRs are not similar across the strata, then it is not appropriate to present a summary RR. Rather, RRs for each stratum are reported separately. This phenomenon is called interaction or effect modification. Statistical tests can be performed to determine the heterogeneity of the stratum-specific RRs. See the text by Kahn and Sempos[7] or Selvin[8] for further discussion of these methods.

Multivariate Analysis

Multiple logistic regression is another technique used to control for confounding variables. In addition to controlling for multiple confounders, this approach is also used for predictive modeling of epidemiologic data. The reader is referred to software packages such as EGRET (Statistics and Epidemiology Research Corporation, Seattle, Wash) or SAS (SAS Institute, Inc, Cary, NC) and to texts[7,8] for further discussion of these techniques.

HYPOTHESIS TESTING

Elements of Hypothesis Testing

There are several possible explanations to account for an observed increased (or decreased) RR or OR. The association could have been observed by chance. The findings could be the result of bias or confounding or both. Finally, the association could represent a cause-and-effect relationship.

The purpose of statistical hypothesis testing is to determine the role random variation or chance plays in interpreting the results of a study. Hypothesis testing is a structured process. Before a study begins, an epidemiologic question or hypothesis is formulated. The hypothesis is stated in two forms: the null hypothesis (H_0) and the alternative hypothesis (H_1). The null hypothesis always states that there is no association between the risk factor and disease in the population or no difference in outcomes between the two groups compared. An example of a null hypothesis is that there is no difference in malaria risk between servicemembers who use bednets and those who do not. The alternative hypothesis always states that there is an association between the risk factor and disease in the population: that there is a difference in the risk of malaria between bednet users and nonusers. A two-sided alternative hypothesis implies that the difference may go in either direction. A one-sided alternative hypothesis states that the difference can only be in one direction, (eg, there is an increased risk of ma-

EXHIBIT 33-9

HYPOTHESIS TESTING

Conclusion from hypothesis test using the sample data	True situation in the population	
	Difference Exists (H_1)	No Difference Exists (H_0)
Difference exists (reject H_0)	Correct, Power (I-β)	Type I error, α error
No difference (do not reject H_0)	Type II error, β error	Correct

laria associated with nonuse of bednets). One-sided alternative hypotheses should be used when there is prior evidence that the difference is in only one direction and a finding in the opposite direction would not be believed.

Statistical tests are based on a null hypothesis under which any observed difference between two groups would be attributed to chance in the data. The purpose of statistical testing is to determine whether the null hypothesis can be confidently rejected. Thus, these tests determine the probability that an association as strong or stronger than the one observed could have occurred by chance alone if no association really existed. There is some probability that the null hypothesis will be rejected when, in fact, it is true. This risk is determined by the significance level chosen for the test. For example, if the null hypothesis is rejected at the 5% level, this means that there is a 5% chance of rejecting the null hypothesis when it is true.

Type 1 error occurs when the H_0 is rejected when it is actually true. This is signified by α. An α level of 0.05 or 0.01 is often used in studies. A type II error (or β error) occurs when the study results fail to reject H_0 when it is actually false. The levels set for β typically are 0.10 or 0.20. Power, a statistical term defined as 1-β, is the ability to detect a difference when a difference exists. Thus, if the β error is 0.20, then the power of a study is 80%. Generally speaking, the larger the sample size, the greater the power of a study. Exhibit 33-9 summarizes the relationship between α level and β level. Exhibit 33-10 summarizes the necessary steps in hypothesis testing.

p Value

The *p* value is based on the assumption that the null hypothesis is true. It reflects the probability of obtaining a measure of association such as a RR or OR as large as (or larger than) the one observed in the study if the null hypothesis is correct. A very small *p* value means that the measure of association that is observed is very unlikely if the null hypothesis is true. For example, if a statistical test is significant at a *p* value of less than 0.05, this means that under the null hypothesis less than 5% of the time a difference of the observed magnitude or greater would occur by random variation alone. It should be noted that 5% is an arbitrary cut-off point used to reject the H_0 and many epidemiologists do not place much value on $p < .05$. If the exact *p* value is reported, then the level of significance can be better assessed.

EXHIBIT 33-10

STEPS IN STATISTICAL SIGNIFICANCE TESTING

- State the hypothesis: an association exists between the factor and disease
- Formulate the null hypothesis: no association exists between the factor and disease
- Choose α and β levels
- Collect data
- Choose and apply the correct statistical test
- Determine the probability of obtaining the observed or more extreme data if null hypothesis is true
- Reject or fail to reject the null hypothesis based on observed *p* value

Confidence Interval

A confidence interval is a range of values of a measure of association, such as the RR or OR, that has a defined probability of containing the true measure. Thus, a 95% confidence interval implies that if a study is repeated 100 times, 95 times the true value of a measure such as a RR will fall within the confidence interval. Confidence intervals of 90% and 95% are commonly used. The upper and lower bounds of a confidence interval define the limits of the interval. When the 95% confidence interval does not include 1.0 (eg, the null value for a RR), then the association is considered statistically significant and has similar meaning to rejecting the null hypothesis at the p-value less than 0.05 level. In addition to providing information on the statistical significance of a test, a confidence interval conveys information on the precision of the point estimate as noted by the width of the confidence interval. The width of a confidence interval is determined both by the size of the study and the level of confidence that the investigator chooses. The larger the sample size, the more precision in the study findings. As a sample size increases, the confidence interval becomes narrower.

Confidence Interval Based on a Proportion

A prevalence can be expressed as a proportion or a percentage. The standard error for a proportion is calculated in the following way:

$$SE(p) = \sqrt{\frac{p(1-p)}{n}}$$

and the 100(1-α)% confidence interval (CI) is calculated using this formula:

$$100(1-\alpha)\% = p \pm Z_{\frac{\alpha}{2}} \bullet SE(p)$$

where p is the proportion and n is the sample size. When the sample size is large [np is ≥ 5 and n(1 – p) is ≥ 5], then p approximates a normal distribution. The critical value of the standard normal distribution associated with a specified level of confidence is *z*. Tables of critical values of the standard normal distribution that correspond to desired levels of confidence can be found in standard statistic textbooks. For a 95% CI, $\alpha = 0.05_{\alpha/2} = 0.025$, and $z_{.025}$ corresponds to the critical value of 1.96.

The following example illustrates how to calculate a 95% CI using prevalence data. A serologic survey of vaccine-preventable infections conducted in 1,504 US Army recruits without prior service found that 17.2% lacked measles antibody.[20] The 95% CI for this percentage was calculated in this way:

p = 0.172
n = 1,504
z = 1.96

$$SE(p) = \sqrt{\frac{p(1-p)}{n}} = SE(.172) = \sqrt{\frac{.172(1-.172)}{1{,}504}} = .0097$$

$$95\%\ CI = p \pm 1.96 \bullet SE(p) = 172 \pm 1.96 \bullet .0097$$
$$= (.153, .192) = (15.2\%, 19.2\%)$$

Confidence Interval Based on Rate Ratio

Consider the study of building-associated risk of febrile ARD in Army trainees.[21] The IRR based on the incidence rate of ARD in trainees living in modern barracks versus old barracks in Fort Jackson was 1.45. A 95% CI for IRR can be constructed using the method shown in Exhibit 33-11 based on Woolf's[22] estimate of the standard error of the natural log (ln) of the IRR. This is interpreted to mean that there is a 95% chance that the interval (1.38, 1.52) includes the true value of the IRR in the population. Because the interval does not include the null value 1.0, the result is statistically significant.

Confidence Interval Based on an Odds Ratio

Constructing a CI based on an OR can also be best understood by working through an example. Consider the case-control study of soft tissue sarcoma and military service in Vietnam.[15] The OR

EXHIBIT 33-11

AN EXAMPLE FOR THE CALCULATION OF A 95% CONFIDENCE INTERVAL FOR A RATE RATIO

$$IRR = 95\%CI = \exp\left[\ln(IRR) \pm 1.96 \sqrt{\left(\frac{1}{a} + \frac{1}{c}\right)}\right]$$
$$= \exp\left[\ln(1.45) \pm 1.96 \sqrt{\left(\frac{1}{3{,}355} + \frac{1}{3{,}312}\right)}\right]$$
$$= (1.38, 1.52)$$

Source: Brundage JF, Scott RM, Lednar WM, Smith DW, Miller RN. Building-associated risk of febrile acute respiratory diseases in army trainees. *JAMA*. 1988;259:2108–2112.

comparing men who never served in Vietnam to those who ever served in Vietnam was 0.82 (see Exhibit 33-4). A 95% CI for the OR in the population can be constructed using the following method based on Woolf's[22] estimate of the standard error based on the OR:

$$95\%\ \text{CI} = \exp\left[\ln(\hat{\text{OR}}) \pm 1.96\sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}\right]$$

$$= \exp\left[\ln(0.82) \pm 1.96\sqrt{\left(\frac{1}{45}+\frac{1}{145}+\frac{1}{172}+\frac{1}{454}\right)}\right]$$

$$= (0.56, 1.20)$$

This means that there is a 95% chance that the interval (0.56, 1.20) includes the true value of the OR in the population. Because this interval includes the null value (1.0), the result is not statistically significant.

In addition to the method proposed by Woolf, there are other formulas to construct confidence intervals around a RR or OR, and they are discussed elsewhere.[7,8]

Calculation of the Mantel-Haenszel Chi-square Test Statistic

If study participants are classified by the presence or absence of an exposure and the presence or absence of disease (ie, a 2 x 2 table), then the chi-square (χ^2) test is an appropriate test statistic. In epidemiologic studies, the Mantel-Haenszel summary chi-square (χ^2_{MH}) is often used as it combines information from each table in a stratified analysis resulting in a test statistic that measures the overall association between a risk factor and disease.

For a series of 2 x 2 tables or *i* tables (as shown above for the Mantel-Haenszel summary RR), the Mantel-Haenszel chi-square is calculated in this way:

$$\chi^2_{MH} = \frac{\left(\sum_{i=1}^{N} a_i - \sum_{i=1}^{N}\frac{N_{1i}M_{1i}}{T_i}\right)^2}{\sum_{i=1}^{N}\frac{N_{1i}N_{0i}M_{1i}M_{0i}}{T_i^2(T_i-1)}}$$

The chi-square value obtained from the study is then compared with a table of chi-square distributions with one degree of freedom, or its square root can be looked up in a normal distribution table. These tables are usually included in statistic books. The chi-square test statistic corresponds to a *p* value (explained below). The chi-square test statistic is more likely to be statistically significant as the sample size gets larger or as the difference between the two groups gets bigger.

For example, to test the hypothesis that sex is related to the risk of stress fractures in trainees, the Mantel-Haenszel chi-square would be the appropriate test statistic as it is designed to incorporate data from a stratified analysis. First the hypothesis is posed in terms of the null: there is no difference in the risk of stress fractures among male and female trainees. An alternative hypothesis is then formulated stating that there is a difference in the proportion of male versus female trainees who develop stress fractures during basic training. A Mantel-Haenszel chi-square is always a two-sided test. From the data that were collected (see Exhibit 33-8), the Mantel-Haenszel chi-square would be calculated as follows:

$$\chi^2_{MH} = \frac{[66 - (100\bullet 36/700 + 400\bullet 82/600)]^2}{100\bullet 600\bullet 36\bullet 664/(700^2\bullet 699) + 400\bullet 200\bullet 82\bullet 520/(600^2\bullet 599)}$$

$$= 1.92$$

This Mantel-Haenszel chi-square corresponds to a *p* value of 0.17. Exact probabilities for a given chi-square value can be obtained from software packages such as Epi Info and SAS. Exhibit 33-12 presents a partial table of critical values of the chi-square with one degree of freedom. From this table, the chi-square of 1.92 falls between the *p* values of 0.20 and 0.10. In these data, no statistically significant association exists between sex and the risk of stress fractures once the effect of aerobic fitness has been controlled since the *p* value is greater than the arbitrary cut-off level of 0.05.

Of note, there are many other types of test statistics that can be used for hypothesis testing, depend-

EXHIBIT 33-12

CHI-SQUARE CRITICAL VALUES FOR VARIOUS PROBABILITIES (ONE DEGREE OF FREEDOM)

p^*	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.001
χ^2	0.455	1.642	2.706	3.841	5.412	6.635	7.879	10.827

*Probability of obtaining an observed value as large as or larger than the one observed in the study under the null hypothesis

ing on the type of variables in a study and whether certain assumptions are made about the underlying distribution of the data. These inferential procedures are called nonparametric statistics or distribution-free statistics. The reader is referred to several texts for a discussion of the use of other test statistics.[7,8,23]

Sample Size and Power

The calculation of a sample size depends on the study design and the measure of outcome variables. For example, to calculate a minimum sample size needed for a case-control study, the investigator determines the following parameters: the α level (eg, 0.05), the β level (eg, 0.20), the minimum size of the OR that will be significant at the 0.05 level (eg, OR = 2.0), and the proportion of subjects exposed versus nonexposed (Exhibit 33-13).

The power of a test is the probability of rejecting H_0 if in fact it is false (a correct decision). Power calculations are important in planning a study, as the larger the study population, the greater the probability of detecting a difference if a difference really exists. Power calculations are also important in the evaluation of a negative study. The conclusion that no association was found may be the result of a small sample size or a true lack of association. In a case-control study, multiple controls are commonly obtained for each case to increase the sample size and therefore the power of the study. This is usually done if the cost of accruing controls is low. In calculating the power (I-β) for a case control study, four basic parameters are required: α (Type I error), difference or effect size, the proportion of the population exposed to the risk factor, and sample size.

There are several computer programs now available to determine sample sizes when planning a study and to solve for power in evaluating the conclusions of a negative study. Epi Info is one such program that is useful for epidemiologists in the field.

EXHIBIT 33-13

COMMON ELEMENTS IN CALCULATING SAMPLE SIZE FOR A STUDY

- Determine the type of study design
- State the null hypothesis
- State the alternative hypothesis and determine if a 1- or 2-sided test is needed
- Determine the appropriate measure of association
- Determine type I error (α) and power (I-β)
- Determine the proportion of the population exposed to risk factors
- Determine the magnitude of measures of association

Statistical Significance Versus Clinical Significance

The word *significant* in the expression statistically significant is often misinterpreted as representing the medical or biological significance of an association. For example, a small difference in the mean diastolic blood pressure between two groups may be statistically significant if the groups are very large, but this finding may or may not be of clinical or biological significance. In addition, statistical significance addresses only random error, not bias or confounding. Thus, an incorrect result due to bias may show statistical significance.

STATISTICAL ASSOCIATIONS AND CAUSE-AND-EFFECT RELATIONSHIPS

Statistical associations from well-controlled experimental studies may represent cause-and-effect relationships. In epidemiology, however, most studies are observational (ie, the investigator does not determine who is exposed and who is nonexposed), and important decisions affecting preventive medicine and public health activities must be made on the basis of observational data. Since statistical tests do not provide proof of a causal relationship between an exposure and disease, guidelines have been established over the years to aid epidemiologists in deciding whether a statistical association obtained from an observational study design is "causal."[24]

One important criterion is the strength of the association or the size of the RR. In general, the larger the RR, the greater the likelihood that the association is causal. Even if some uncontrolled or unknown confounding is present, it is unlikely that controlling for this confounder could decrease the RR sufficiently to make the association unimportant. However, uncontrolled confounding plays a more important role when the RR is close to 1.0 and could account for the magnitude of the risk that is reported. Another important criterion to determine if an association is causal is that the exposure must precede the disease. This temporal relationship is not always clear in a

cross-sectional or case-control study and is one reason why these studies are not conclusive. A strength of the cohort design is that the exposure always precedes the disease. Another criterion is biological plausibility. If the relationship between the exposure and disease outcome makes sense in terms of known biological facts and mechanisms, then it is more plausible that the exposure could cause the disease. Although a threshold effect is found for some biological phenomena, in general, a dose-response relationship between exposure and disease is another criterion that may help establish a cause-and-effect relationship. A spurious association, however, may also exhibit a dose-response relationship. Finally, although the design of any given study may have unique features or may introduce bias, findings that are consistent across studies, especially studies with different designs or conducted in different populations, suggest a cause-and-effect relationship. This is one of the most important tenets used by epidemiologists.

SUMMARY

This chapter was written to provide military preventive medicine personnel with a better understanding of the fundamental epidemiologic methods and tools used during outbreak investigations and medical surveillance during deployments. The public health officer on deployment often is faced with assessing whether a public health problem exists among servicemembers; proposing a plan; implementing an appropriate intervention; and then evaluating the impact of an intervention. Using appropriate epidemiologic methods to collect data, to interpret data, and to present data in an understandable format will greatly aid in this mission.

Acknowledgment

The author would like to acknowledge Colonel John W. Gardner and Dr. Yuanzhang Li for their help in the preparation of this chapter.

REFERENCES

1. MacMahon B, Pugh TF. *Epidemiology: Principles and Methods.* Boston: Little, Brown; 1970.
2. Jones WHS, ed. *Hippocrates: Airs, Waters, Places*. Cambridge, Mass: Harvard University Press; 1948.
3. US Army Medical Surveillance Activity. Injury incidence in soldiers attending medical specialist (MOS91B) AIT, Fort Sam Houston, Texas. *Med Surv Month Rep*. 1997;3(Dec):10–11.
4. Jones BH, Cowan DN, Tomlinson JP, Robinson JR, Polly DW, Frykman PN. Epidemiology of injuries associated with physical training among young men in the Army. *Med Sci Sports Exerc*. 1993;25:197–203.
5. Renzullo PO, McNeil JG, Wann ZF, Burke DS, Brundage JF, the United States Military Consortium for Applied Retroviral Research. Human immunodeficiency virus type-1 seroconversion trends among young adults serving in the United States Army, 1985–1993. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1995;10:177–185.
6. Writer JV, DeFraites RF, Brundage JF. Comparative mortality among US military personnel in the Persian Gulf region and worldwide during Operations Desert Shield and Desert Storm. *JAMA.* 1996;275:118–121.
7. Kahn HA, Sempos CT. *Statistical Methods in Epidemiology*. New York: Oxford University Press; 1997.
8. Selvin S. Statistical Analysis of Epidemiologic Data. 2nd ed. New York: Oxford University Press; 1996.
9. Petruccelli BP, Murphy GS, Sanchez JL, et al. Treatment of traveler's diarrhea with ciprofloxacin and loperamide. *J Infect Dis.* 1992;165:557–560.

10. Gaydos JC, Hodder RA, Top FH Jr, et al. Swine influenza at Fort Dix, New Jersey, Jan–Feb 1976: Case finding and clinical study of cases. *J Infect Dis*. 1977;Suppl 13:S356–S362.

11. LCDR Tanis Batsel, Walter Reed Army Institute of Research. Personal Communication, 2000.

12. Gambel JM, Drabick JJ, Martinez-Lopez L. Medical surveillance of multinational peacekeepers deployed in support of the United Nations Mission in Haiti, June–October 1995. *Int J Epidemiol*. 199;28:312–318.

13. Smoak BL, Kelley PW, Taylor DN. Seroprevalence of *Helicobacter pylori* infections in a cohort of US Army recruits. *Am J Epidemiol.* 1994;139:513–519.

14. Cornfield J. A method of estimating eomparative rates from clinical data: applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst*. 1951;11:1269.

15. Kang H, Enzinger FM, Breslin P, Feil M, Lee Y, Shepard B. Soft tissue sarcoma and military service in Vietnam: a case-control study. *J Natl Cancer Inst.* 1987;79:693–699. Published erratum: *J Natl Cancer Inst* 1987;79:1173.

16. Takafuji ET, Kirkpatrick JW, Miller RN, et al. An efficacy trial of doxycycline chemoprophylaxis against leptospirosis. *N Engl J Med*. 1984;310:497–500.

17. Friedman LM, DeMets DL, Furberg CD. *Fundamentals of Clinical Trials*. New York: Springer-Verlag New York, Inc; 1998

18. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979;32:51–63.

19. Rothman KS. Greenland S. *Modern Epidemiology*. 2nd e. Philadelphia: Lippincott Williams & Wilkins; 1998

20. Kelley PW, Petruccelli BP, Stehr-Green P, Erickson RL, Mason CJ. The susceptibility of young adult Americans to vaccine-preventable infections: A national serosurvey of US Army recruits. *JAMA*. 1991;266:2724–2729.

21. Brundage JF, Scott RM, Lednar WM, Smith DW, Miller RN. Building-associated risk of febrile acute respiratory diseases in army trainees. *JAMA.* 1988;259:2108–2112.

22. Woolf B. On estimating the relation between blood groups and disease. *Ann Hum Genet.* 1954;19:251–253.

23. Siegel S. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill Book Company; 1956.

24. Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58:295–300.