1. What is epidemiology?

The word epidemiology consists of the Greek words epi = among, demos = people, and logos = doctrine, and thus means the doctrine of what is among or happening to people. Epidemiology is “the study of the distribution and determinants of disease frequency” in human population (or in the group of population) These three closely interrelated components encompass all epidemiological principles and methods. The measurement of disease frequency involves quantification of the existence or occurrence of disease. The distribution of the disease considers such questions as who is getting disease within a population, where and when the disease is occurring. Knowledge of such distributions is essential to describe patterns of disease as well as to formulate hypotheses concerning possible causal or preventive factors. The determinants of disease, which can be derived from the knowledge of frequency and distribution of the disease, are necessary to test the epidemiological hypotheses. Epidemiological research is based on the systematic collection of observations related to the phenomenon of interest in a defined population. These data then are subjected to quantification, which includes the measurement of random variables, the estimation of population parameters, and the statistical testing of hypotheses.

Epidemiologists study the occurrence of disease or other health-related conditions or events in defined populations. Viewed as the study of the distribution and societal determinants of the health status of populations, epidemiology is the basic-science foundation of public health. Investigation of disease occurrence is not a new phenomenon. The development of epidemiological theory and methods in recent decades, however, has opened new possibilities and stimulated interest within many fields of application. Sometimes the starting point for an epidemiological study is a certain characteristic or exposure rather than a disease. There is a close connection between epidemiology and preventive medicine. Prevention programs are rarely implemented for an entire population, therefore, prevention programs may be planned to enable studies of the effect of the intervention on the disease frequency in the population by comparing disease rates among those receiving the preventive program with rates among those who do not. In this way there is usually an opportunity to evaluate preventive measures that have been undertaken, using experimental epidemiology.
Epidemiological research has often provided information that has formed the basis for public health decisions long before the basic mechanism of a particular disease was understood. In recent years the value of information about disease distribution for planning the delivery of health care has become more apparent. In several studies disease occurrence has been related to health-care need, demand, and supply. There is also an increasing interest in studying the effectiveness of the health-care system and/or of different treatments. The common basis for these different applications of epidemiology is the study of disease occurrence and its relation to various characteristics of individuals or their environment.
2. Epidemiological observations

The quality of the data is commonly described with use of four terms:

- accuracy
- precision
- reliability
- validity

Accuracy is the degree to which a measurement represents the true value of the attribute being measured. A measurement or observation can represent a true value without detail. As with other measurements the accuracy of the study result depends on "validity" and "precision" of the measure.

Validity is the extent to which the study measures what it is intended to measure; lack of validity is referred to as "bias" or "systematic error." An important distinction must be made between the term’s internal validity and external validity. Internal validity concerns inferences about the population of individuals of restricted interest from which a study sample has been drawn; external validity concerns inferences about an external population beyond the study's restricted interest.

Precision is the reproducibility of a study result, that is, the degree of resemblance among study results, were the study to be repeated under similar circumstances: lack of precision is referred to as "random error". A study that is based on information from too few subjects allows for considerable random variation in the study result, since only a few extra cases occurring in one or another category would substantially affect the results. Such a study would have a low precision.

Reliability is a measure of how dependably an observation is exactly the same when repeated; it refers to the measuring procedure rather than to the attribute being measured. Reliability is not synonymous with repeatability or reproducibility; rather, it is a broader term that includes the concept of consistency, which refers to how closely the findings in different samples or populations conform to one another under different conditions or at different times.
3. Diagnosis of disease

Before a disease can be studied, it is necessary to define it. Often, this results in a case definition, which may differ from a clinical definition. Studies in which different definitions are used may lead to different conclusions.

In principle the variables that are the basis for diagnosis may depend upon:

• subjective observations by the patient (symptoms)
• subjective observations by the examiner (signs)
• objective observations (tests)

Symptoms refers to manifestations that only the examined person (patient) may observe, e.g., insomnia, nausea, or fatigue. Of course, symptoms can be perceived and described differently by different individuals and by the same individual in different situations. Accuracy in recording symptoms is influenced by the instrument used to collect the data; standardized interview and questionnaire methods have been developed to increase reproducibility.

Signs refers to manifestations that may be observed by an examiner (usually a physician), for example, weight loss, apathy or agitation. Ascertainment of signs is affected by the subjective judgment of one (or several) examiner(s). The accuracy of such examinations is dependent on:

• inter-observer variation (the degree of agreement among different examiners)
• intra-observer variation (the degree of agreement between different examinations made by one examiner).

In many cases reproducibility can be increased, and hence the importance of this source of error in the diagnosis reduced, by standardization of the conditions under which the observations are made. The examination routines can be standardized and sometimes detailed and standardized criteria for classification of the observations can be formulated.
Tests will refer here to manifestations that can be read from an instrument and hence are less dependent on subjective judgments by the person examined or the examiner. The reproducibility of the test’s result can be examined, by repeating. The difference in the results may be accounted for by the differences in instruments, techniques, or performance.

The manifestations (symptoms, signs, and tests) that are considered to be typical for a certain disease have been used to formulate diagnostic criteria, i.e., conditions that must be fulfilled for the diagnosis to be made. The choice of diagnostic criteria therefore determines whether an examined individual is to be classified as having a certain disease. If stringent criteria are used, there is only a small probability that people who do not have the disease are being classified as having it, but a relatively large probability that some people who have the disease will be classified as not having it. On the other hand, if less stringent criteria are used, the opposite type of misclassification will tend to occur: most people with the disease will be correctly classified as having it, but there is a relatively large probability that people without disease will be classified as having it.

There is no completely accurate method to identify those individuals who have a certain disease. Some observations, though, are considered more accurate than others when making a diagnosis. Naturally, diagnostic accuracy varies from one disease to another, but also from one group of individuals to another. The accuracy also depends on what kind of examination has been performed and on the interpretation of the observations.

The diagnostic criteria can be examined in terms of four parameters:

- sensitivity
- specificity
- positive predictive value
- negative predictive value

"Sensitivity" refers to the probability that the test result is positive, given that the people tested actually have the disease that the test is supposed to detect. "Specificity" to the probability that the test result is negative, given that the people tested do not have the disease that the test is supposed to detect. "Positive predictive value" refers to the probability that the people who suffer the disease tested will have positive test result. "Negative predictive value" refers to the
probability that the actual disease state is negative given that the test result is negative (Table 1).

Table 1. Possible combinations of whether or not a disease is present and whether the result of a diagnostic test is positive or negative.

<table>
<thead>
<tr>
<th>Actual disease status (Dx)</th>
<th>Test results (T)</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
<td>a+b</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>c</td>
<td>d</td>
<td>c+d</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity = Probability (T+, Dx+) = a/(a+c)
Specificity = Probability (T-, Dx-) = d/(b+d)
Predictive value+ = Probability (Dx+, T+) = a/(a+b)
Predictive value- = Probability (Dx-, T-) = d/(c+d)
4. Epidemiological measures of disease

Measures of disease occurrence are central to all epidemiological activity. Such measures can be formulated in a variety of ways:

- absolute numbers and numbers related to size of the population
- incidence and prevalence

Measures of prevalence describe what proportion of the population has the disease in question at one specific point in time. Prevalence is the total number of individuals who have a characteristic or disease at a particular point in time divided by the number who are at risk of having that characteristic or disease at that designated point in time. Prevalence depends on both the number of people who have had the disease or characteristic in the past and the duration of the disease or characteristic.

Measures of incidence, on the other hand, describe the frequency of occurrence of new cases during a time period. Incidence is the number of new cases of a disease in a defined population within a specified time-period divided by the number who are at risk of having that disease or characteristic at that designated time-period.

It is useful to think of each individual as being in one of two "states": diseased or disease-free. In this framework the prevalence measure describes the proportion of the population that is in the diseased state at a specific time. The incidence measure describes the rate of flow from the disease-free state to the diseased state.

The relevant measure of disease occurrence is incidence. Measures of prevalence may be relevant in connection with the planning of health services or in assessing the need for medical care in a population. For example, studies of chronic diseases such as rheumatoid arthritis, in which the point of transition from non-diseased to diseased occurs gradually, and the definition of disease is arbitrary, generally employ prevalence measures, whereas studies on cancer or myocardial infarction generally use incidence measures.
The prevalence measure is called simply the “prevalence” \((P)\) (other terms in use include prevalence rate and prevalence proportion). The prevalence is defined as:

\[
P = \frac{N_d}{N_t}
\]

\(N_d\) is the number of individuals having the disease at a specific time
\(N_t\) is the number of individuals in the population at that point in time

The prevalence is dimensionless and can never take values less than 0 or greater than 1.

**Example:** A sample including 1,500 women age 70-74 years was selected from the population of Helsinki City area. After examination, 120 were classified as having the diagnosis of depression. The prevalence of depression was:

\[
P = \frac{120}{1500} = 0.08
\]

This means that 8% of the women in the age group 70-74 suffer from depression.

There are two commonly used incidence measures: cumulative incidence or incidence proportion (CI) and incidence rate or incidence density (IR).

The definition of cumulative incidence is:

\[
CI = \frac{N_{gd}}{N_{tb}}
\]

\(N_{gd}\) is the number of individuals who get the disease during a certain period
$N_e$ is the number of individuals in the population at the beginning of the period.

Both numerator and denominator include only those individuals who at the beginning of the period are free from the disease and therefore at risk to get it. The cumulative incidence is, therefore, the proportion of individuals in the disease-free state at the beginning of the period that move to the disease state during the period. That is, the numerator is a subset of the denominator. Alternatively, it can be viewed as the average risk for the individuals in the population to get the disease during that period. Being a proportion, the cumulative incidence is dimensionless and can only take numeric values in the range from 0 to 1.

**Example:** The Swedish census from 1960 showed there were 3,076 males age 20-64 who were employed as plastic workers. According to the Swedish Cancer Environment Registry, 11 of those workers developed brain tumors during the period 1961-1973 (National Board of Health and Welfare). The cumulative incidence during the 13-year period therefore is:

\[ CI = \frac{11}{3,076} = 0.004 \]

This means that 0.4% of plastic workers develops a brain tumor in a time span of 13 years.

In some studies different subgroups of the study population are considered at risk of getting the disease during different periods of time and individuals are considered at risk for periods of varying lengths. This variation in the risk period derives from the fact that different individuals enter the study at different points in time or that some migrate during the observation period. In such situations the cumulative incidence may not be directly calculable from the data. The length of the observation period directly affects the cumulative incidence: the longer the period, the greater cumulative incidence.

Incidence rate is a basic measure of disease occurrence, which is defined as

$N_{cd}$ is the number of cases of the disease that occur in a population during a period of time.
\( E_{\text{nr}} \) is the sum for each individual in the population of the length of time at risk of getting the disease.

The sum of the time periods in the denominator is often measured in years and is referred to as "person years," "person time," or "risk time." For each individual in the population the time at risk is the time during which that individual is in the study population and remains free from the disease, and therefore at risk to get it. These time periods at risk are then summed for all individuals. The rationale is that the total number of individuals who move from the disease-free state to the disease state during any period of time is the product of three factors: size of the population, the length of the time period, and the "force of morbidity" that operates on the population. It is this "force of morbidity" that the incidence rate measures. Therefore the incidence rate is obtained by dividing the number of cases by the product of the size of the population and the length of the period, which is equivalent to summing the individual periods of time for each individual in the population. By dividing the number of cases by the time at risk, the length of the observation period is taken into account. Also, individuals entering or exiting the population during the observation period because of initiation, competing mortality, or any other reason are automatically accounted for. Thus, by including time at risk in the definition, the incidence rate accounts for the major drawbacks encountered with the cumulative incidence measure.

**Example:** In 1973 there were 29 cases of myocardial infarction in Stockholm among men age 40-44 years (Ahlbom 1978). The number of person-years was 41,532 for men in that age group. The incidence rate per year is therefore:

\[
I = \frac{29}{41,532} = 0.0007
\]

This means that 7 male persons/10,000 (7 \( \times 10^{-4} \)) per one year, in the age group 40-44 will suffer from a myocardial infarction.
In practical situations it is often not possible to calculate the time at risk for each individual. It may not even be possible to exclude the period during which some of the individuals are no longer at risk because they already have developed the disease. An approximation to the total time at risk that is usually satisfactory can be obtained by multiplying the average of the population size at the beginning and the end of the observation period by the length of the period. The size of the population at the middle of the observation period may also be used for this approximation. In this case the incidence rate can be calculated as:

**Example:** During a 5-year period 270 cases of duodenal ulcer occurred in the male population of a city. The number of men in the city was 18,500 at the beginning of the period and 21,500 at the end. Which is the incidence rate?

Number of cases = 270

Average population size =

\[
\frac{(18,500 + 21,500)}{2} = 20,000
\]

Observation period = 5 years

Total risk period = 20,000 x 5 = 100,000 person-years

Incidence rate per year:

\[
IR = \frac{270}{100,000} = 0.0027
\]
The incidence rate is not a proportion like the two previous measures, since the numerator is the number of cases and the denominator is the number of person-time units. The magnitude of the incidence rate can never be less than zero but there is no upper limit for the numerical value of the incidence rate.

*The interrelation among the three measures*

Prevalence is dependent on the incidence and the duration of disease. In a stable situation this association may be expressed as follows, where D indicates average duration of the disease:

\[
P / (1 - P) = I \times D
\]

The denominator on the left side of the equation reflects the part of the population that is free from the disease. It is included in the formula since only those people who are free from the disease are at risk of getting it. For rare diseases, i.e., diseases where P is low, the following approximation may be used:

\[
P = I \times D
\]

The cumulative incidence depends on the incidence rate and the length of the period at risk. It also is affected by mortality from diseases other than the disease studied, as stated previously. If this mortality from other diseases is disregarded the following relation applies:

\[
CI = 1 - \exp(-I \times t)
\]

where \( t \) is the length of the period and "exp" indicates that the mathematical constant \( e \approx 2.718 \) should be raised to the power given by the expression in parenthesis. For diseases with low incidence rate or when the period is short, the following approximation may be used:

\[
CI = I \times t
\]

Despite the close interrelationship of prevalence and incidence, each provides a somewhat different type of information with differing utility. Prevalence measures are most useful for
health care providers, to assess the public health impact of a specific disease within a community and to project medical care needs for affected individuals, such as numbers of health care personnel or hospital beds that will be required.

The measures of disease occurrence discussed above may be calculated for a whole population or calculated separately for parts of the population. In the first case the measures are called "crude" measures and in the latter case "specific" measures. For example, if incidence rates are calculated for different age groups within a population, they are referred to as age-specific incidence rates. Crude rates represent the actual experience of the population and provide data for the allocation of health resources and public health planning. The magnitude of a crude measure for a population depends not only on the magnitude of the specific measures that apply to subgroups of the population, but also on the way the population is distributed over the different subpopulations. Specific rates provide the most detailed information about the pattern of the disease in a population.
5. Measures of association

In epidemiological studies the frequency of disease occurrence among individuals who have a certain characteristic is generally compared with the corresponding frequency among those who do not have that characteristic. The compared groups are usually referred to as “exposed” and “non exposed”. This comparison constitutes the fundamental way of studying the association between exposure and disease occurrence.

The concept of an association, or statistical dependence, between a factor and a disease is fundamental to ascription of the factor as possibly causal. An exposure or attribute that increases the probability of a disease is a risk factor. Epidemiologists calculate a measure of association as a single summary parameter that estimates the association between the risk of developing a disease and an exposure. These measures of association draw on two concepts: the probability of an event or disease and the odds of an event or disease. The most frequently used measures of association are the attributable risk (AR), relative risk (RR) and the odds ratio (OR).

Attributable risk is the difference between the incidence of the disease in exposed group (I_e) and incidence of the disease in non-exposed group (I_0)

$$AR = I_e - I_0$$

AR indicates on an absolute scale how much greater the frequency of disease is in one group compared with the other.

Relative risk (RR) indicates the average risk of disease that is due to a given exposure in the exposed group. It provides an estimate of the magnitude of an association between an exposure and a disease; it is the ratio of the incidence of the disease in exposed group (I_e) divided by the corresponding incidence of the disease in non-exposed group (I_0).

$$RR = \frac{I_e}{I_0}$$

A relative risk of 1.0 indicates that the incidence rates of disease in the exposed and non-exposed groups are identical and thus that there is no association observed between the
exposure and the disease in the data. The value greater than 1.0 indicates a positive association or an increased risk among those exposed to factor. A relative risk less than 1.0 means that there is an inverse association or a decreased risk among those exposed.

The relative risk (RR) can be calculated on the basis of historical or prospective cohort studies as well as cross-sectional studies. In a case control study, where participants are selected on the basis of disease status, it is usually not possible to calculate the rate of development of disease given the presence or absence of exposure. In this case relative rate can be estimated by calculating the ratio of the odds of exposure among the cases to that among the controls.

To facilitate the calculation of these measures, epidemiological data are presented in a two-by-two table (Table II, Table III).

Table II. Two-by-two table showing the relationship between the presence of a disease and exposure to a risk factor that may contribute to the disease

<table>
<thead>
<tr>
<th>Disease</th>
<th>No disease</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>exposure</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No exposure</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
<td>b+d</td>
</tr>
</tbody>
</table>

*The letter “a” represents people who have the disease and have been exposed to a risk factor under study.

\[
RR = \frac{a}{a+b} \div \frac{c}{c+d}
\]
Table III. Two-by-two table showing the relationship between the case-controls and exposure to a risk factor that may contribute to the disease

<table>
<thead>
<tr>
<th>Exposure</th>
<th>No exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>a</td>
</tr>
<tr>
<td>Controls</td>
<td>c</td>
</tr>
</tbody>
</table>

The letter “a” represents cases, which have been exposed to a risk factor under study.

\[ OR = \frac{ad}{cb} \]

RR is a measure of the strength of the association between an exposure and disease and provides information that can be used to judge whether a valid observed association is likely to be causal. In contrast, AR provides a measure of the public health impact of an exposure, assuming that the association is one of cause and effect. The magnitude of the RR alone does not predict the magnitude of the AR. The odds ratio and risk ratio are very similar in instances of rare disease but are quite dissimilar when the prevalence begins to exceed 5 or 10 per cent.

Example: Of 595 patients who had received blood transfusion and 712 patients who had not, 75 and 16, respectively, developed hepatitis during a 2.5 year follow-up in a study designed to evaluate the relative risk.

The study is a cohort study.
This means that there is a 6.5 time higher risk to develop hepatitis if you have had a blood transfusion compared to if you have not had one.

**Example:** The incidence of myocardial infarction is higher in Finland than in Sweden. In the southern suburbs of Stockholm a study was conducted based on middle-aged men who developed myocardial infarction in the years 1974-76 and on controls from the general male population. Information was compiled on the country of origin and length of stay in Sweden. Calculate the Odds ratio of myocardial infarction

a) for those living in Sweden for 0-19 years

b) for those living in Sweden for 20 years or more

**Table IV. Distribution of cases and controls for those living in Sweden for 0-19 years**

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Cases</td>
<td>22</td>
<td>324</td>
</tr>
<tr>
<td>Controls</td>
<td>31</td>
<td>832</td>
</tr>
</tbody>
</table>

\[
OR_{0-19} = \frac{(22 \times 832)}{(324 \times 31)} = 1.8
\]
Table V. Distribution of cases and controls for those living in Sweden for more than 20 years

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>Cases</td>
<td>10</td>
</tr>
<tr>
<td>Controls</td>
<td>24</td>
</tr>
</tbody>
</table>

\[
OR_{20-} = \frac{(10 \times 832)}{(324 \times 24)} = 1.1
\]

These both measures should be interpreted so that there is almost a twofold risk for middle aged Finnish men to suffer from myocardial infarction if they have lived in Sweden for less than 19 years compared to Swedish middle aged men. Furthermore, the risk for Finnish men is only a little elevated if they lived in Sweden for more than 20 years.
6. Design strategies in epidemiological research

The basic design strategies used in epidemiological research can be broadly categorized according to whether such investigations focus on describing the distributions of disease (descriptive epidemiology) or elucidating its determinants (analytical epidemiology).

Descriptive epidemiology is concerned with describing the general characteristics of the distribution of a disease, particular in relation to person, place in time. The characteristics measured in this type of study are: crude number of e.g. death, or of a specific disease, number adjusted to the size of population at risk (prevalence and incidence), means of studied variables (e.g. mean age, mean alcohol consumption and so on). Descriptive data provide valuable information to enable health care providers and administrators to allocate resources efficiently and to plan effective prevention or education programs. Due to limitation inherent in their design, however, descriptive studies are primarily useful for the formulation of hypotheses that can be tested subsequently using an analytical design.

There are three main types of descriptive studies:

- correlational (ecological) study
- case reports and case series
- cross-sectional surveys

Correlational study uses data, which usually have been routinely collected, from entire population to compare disease frequencies between different groups during the same period of time or in the same population at different points of time. While such studies are useful for the formulation of hypotheses, they cannot be used to test them because the number of limitations inherent in their design. Correlational studies refer to whole populations rather than to individuals, it is not possible to link an exposure to occurrence of disease in the same person.

**Example:** Examination of the relationship between per capita consumption of cigarettes and operative procedures for the treatment of non-unions in five counties in California in 1996. Even if this study showed that counties with a higher rate of consumption of cigarettes had a higher rate of operations for non-unions, the investigator still could not be sure that the individuals who smoked in these counties truly had a higher rate of non-union. This conclusion,
if indeed it is erroneous, is known as an ecological fallacy because the correlation between two ecological variables is often different from the corresponding individual correlation within the same populations.

Case report is the most basic type of descriptive study of individuals, consisting of careful, detailed report by one or more clinicians of the profile of a single patient. The individual case report can be expanded to a case series, which describes characteristics of a number of patients with a given disease. Some people would argue that these types of studies are not epidemiological studies, since they focus on a few individuals. Rather they are clinical studies, which may generate hypotheses to be tested in epidemiologic studies.

Cross-sectional survey (also known as prevalence studies or disease-frequency surveys) is a study in which the status of an individual with respect to the presence or absence of both exposure and disease is assessed at the same point in time. Since exposure and disease are assessed at the same point in time, cross-sectional survey cannot always distinguish whether the exposure preceded the development of the disease or whether presence of the disease affected the individual's level of exposure. This type of studies useful for raising the question of the presence of an association rather than for testing a hypothesis.

**Example:** Silverstein et al., in a cross-sectional study, determined the prevalence of carpal tunnel syndrome among 652 workers in thirty-nine jobs at seven different industrial sites. Specific hand-force and repetitiveness characteristics their were estimated for the different jobs. The prevalence of carpal tunnel syndrome ranged from 0.6 percent among workers in low-force, low-repetition jobs to 5.6 per cent among those in high-force, high-repetition jobs. In order to infer that a statistical association between job-related exposure factors and carpal tunnel syndrome is evidence of etiology, it should be demonstrated that the job exposure occurred before the carpal tunnel syndrome. The temporal relationship between physical load factors and the onset of carpal tunnel syndrome cannot be demonstrated in a cross-sectional study examining the prevalence of carpal tunnel syndrome. Prevalence studies are performed on survivor populations and thus may be affected by selection bias for example, individuals who have more severe carpal tunnel syndrome may have left the workforce and thus may not be accounted for in a cross-sectional survey.
In analytical study designs the comparison is explicit, since the investigator assembles groups of individuals for the specific purpose of systematically determining whether or not the risk of disease is different for individuals exposed or not exposed to a factor of interest. These studies can be divided into two broad design strategies:

- **observational (non-experimental)**
- **intervention (experimental)**

The major difference between the two lies in the role played by the investigator. In observational studies, the investigator simply observes the natural course of events, noting who is exposed and non-exposed and who has and has not developed the outcome of interest. In intervention studies, the investigator themselves allocate the exposure and then follow the subjects for the subsequent development of the disease. For ethical and practical reasons, however, epidemiological investigations are usually non-experimental; that is, they are based on existing exposure conditions.

There are two basic types of observational analytical investigations:

- cohort (also known as a "follow-up") study
- case-control study

The goal of an observational study is to arrive at the same conclusions that would have been derived in experimental trial. In general, the decision to use a particular design strategy is based on features of the exposure and disease, the current state of the knowledge, and logistic consideration such as available time and resources.

**Cohort studies**

The cohort study is the more straightforward approach. All subjects in the study population are assigned to a category of exposure at the start of the time period under study. Study subjects are defined into two groups: people who are exposed become the *exposed cohort*, people who are not exposed become the *non-exposed cohort*. The study subjects are then followed up for a defined observation period and all new cases of the disease under study are identified. All potential subjects must be free of the disease being studied at the time that exposure status is
defined. The combination of the population and time period at risk (i.e., the person-time observed) is sometimes referred to as the study base, indicating the base experience from which the cases arise. The information obtained is used to estimate the incidence rate or cumulative incidence among the exposed and unexposed groups. The comparison of disease incidence among exposed and unexposed may be either absolute (absolute risk) or relative (relative risk).

The unexposed group is intended to provide information about the disease incidence rate that would be expected in the exposed group if the exposure under study did not affect the occurrence of the disease. Therefore, the unexposed group should be selected in such a way that it is similar to the exposed group with regard to other risk indicators for the disease under study.

In principle there are three different approaches to defining the unexposed group:

1. Internal comparison: A single cohort is identified that contains a sufficient number of exposed and unexposed subjects.
2. External comparison: An exposed cohort is identified and efforts are made to find another cohort that is unexposed but is similar in other respects to the exposed cohort.
3. Comparison with the "general" population: An exposed cohort is identified and comparisons are made with the disease incidence in, for example, the total population of a defined geographic region (considered as "unexposed").

Confounding may occur if a risk indicator other than the studied exposure is unequally distributed between the groups. Confounding can to a certain extent be controlled in the data analysis; for example, differences in the distributions of factors such as age and sex between the groups can be corrected in the analysis. There are other problems to be considered, however, when selecting the unexposed group.

The incidence of disease can vary considerably over time (year, season etc.), and by geographic area, ethnic background, and socioeconomic status. Differences between the exposed and unexposed groups with regard to these or other factors can influence the results of the study. To avoid such problems it may be advisable to select exposed and unexposed
subjects with similar distributions for these factors and to follow them during the same observation period.

If information from the total population is used to represent the "unexposed" group, then all study subjects classified as exposed (and other people with the same exposure) are also included among the "unexposed. Unless the proportion of exposed subjects in the total population is low, this dilution of the "unexposed" population may result in an underestimation of the relative morbidity in the exposed group.

The procedures used to identify new cases of the disease under study should be similar for the exposed and unexposed groups. If, for example, certain screening procedures are used to a different extent in the exposed group, case identification may be more complete in the exposed group.

Selecting more than one unexposed group can reveal to what extent the study results are influenced by the selection of an unexposed group. Separate analyses using these groups that yield similar results indicate that the results probably were not influenced by the choice of the unexposed group. Another approach that has sometimes been used to check the comparability between the exposed and unexposed group is to include comparisons of the occurrence of diseases that are not expected to be associated with the exposure under study.

Cohort studies are sometimes based on information about exposure and disease collected in the past (so called "retrospective cohort studies"). By the other words, the investigation is initiated at a point of time after both the exposure and disease have already occurred. The accuracy of such a study depends upon the completeness of disease ascertainment in the registry for the population and time period under study.

In prospective cohort studies, the groups of exposed and unexposed subjects have been assembled at the beginning of the study, but the disease has not yet occurred, so that investigator must conduct follow-up during an appropriate interval to ascertain the outcome of interest.

Case-control studies
For most diseases, the occurrence of new cases is a relatively rare event. Thus, in a cohort study, a large number of study subjects may have to be examined for exposure status and followed up for a long time to obtain a sufficient number of cases. Such a study is often not practical or feasible. The problem can be avoided by using exposure, information from a sample of the study population rather than including the entire study population.

Case-control studies are based on this principle. As in cohort studies, information is obtained on all cases that occur in the study population during a defined observation period. In addition, a comparison group of "controls" are selected as a representative sample of the study population; ideally the control group reflects the exposure distribution in the entire study population. Exposure information is then obtained for cases and controls only, rather than for all members of the population.

In a case-control study, the investigator selects individuals on the basis of whether or not they have the disease and then determines their previous exposure. Exposure status is determined retrospectively for both cases and controls. Exposure status is compared between cases and controls. If the exposure is more common among cases than controls, then the exposure is associated with the disease. The association is expressed as the odds ratio. The control group in a case-control study should be selected in such a way that it reflects the exposure distribution in the study population, i.e., in the population that generated the cases. There are two different approaches to control selection:

1. A *random sample of the study population*. This approach has the advantage that the controls will be representative of the study population in a formal (statistical) sense and therefore the selection process for controls does not introduce any systematic error. This option obviously requires that the population is accessible for random sampling. Potential disadvantages of random sampling for control selection are the possibility of a high non-response rate among healthy population controls, and the possibility of differences in the quality of exposure information between cases and healthy controls.

2. A *sample of the study population that is not randomly selected*. This is the only option when the cases are identified in such a way that the study population is not accessible for random sampling. For example, if the cases are patients diagnosed with the disease being studied in a particular clinic, these patients most likely do not represent all cases of the disease occurring in a population from which a random sample may be drawn, and therefore a nonrandom selection of controls is necessary. Sometimes this approach is used for other
reasons, such as reducing non-response among controls or improving, the comparability of exposure information between cases and controls. While such nonrandom selection schemes may result in greater comparability of cases and controls, controls selected in this way may not reflect the exposure distribution in the overall study population, thus possibly introducing, a systematic error.

When controls are selected non-randomly, it can be difficult to judge to what extent their exposure distribution reflects the exposure distribution in the population that generated the cases. When hospital controls are selected, a single diagnostic group may be used as a source of controls, provided that neither the disease nor the probability for hospitalization is associated with the studied exposure(s). It may be useful to select two or more diagnostic groups sufficiently large to permit separate analyses to minimize the problem, especially when such associations cannot be ruled out at the beginning of the study. Using a sample of all patients (or hospitalization episodes) regardless of diagnosis is less advisable, since many exposures (such as diet, tobacco and alcohol consumption) tend to be related to several common diseases. When there is a choice between these two approaches in selecting controls, random sampling is the preferred method.

In case-control studies, controls are sometimes selected by individual matching to the cases: for every case, one or more controls are selected who are similar to that case in certain respects. Cases and controls are matched on potential confounders, such as age, sex and residence, for which information can be obtained before data collection begins.

The matching is performed with the purpose of reducing random error rather than confounding.

Example: Study of the epidemiology of acute herniation of lumbar intervertebral discs, (Kelsey and Hardy). The investigators compared the characteristics of patients who had such herniation with those of two control groups who were known not to have it and found that driving a motor vehicle was associated with an increased risk of herniated discs. That study provides an excellent demonstration of how the many concerns in case-control studies, including selection of controls, analysis of data with various confounders, and consideration of other biases, should be addressed.
Experimental studies or clinical trials may be viewed as a type of prospective cohort study, because participants are identified on the basis of their exposure status and followed to determine whether they develop the disease. The distinguishing feature of the intervention design is that the exposure status of each participant is assigned by the investigator.

Intervention studies can be divided on:

- “true” experimental (randomized controlled trial, RCT)
- quasi-experimental

In randomized controlled trial the exposure status is determined by random allocation. It is the optimal design to study cause-effect or efficacy of an intervention. Ideally the control group should receive “placebo” treatment and both the patients and the researcher should be “blinded” to the treatment status, that is no-one knows who is being treated and who is not until after the outcome is measured. This is randomized “double blind placebo controlled trial”. When using a placebo in the control group, the expected placebo effect in the intervention group is accounted for when the two groups are compared. If there is a “placebo”-effect of the intervention, this placebo effect will be the same in both groups, and the observed differences between the groups are likely to be due to the “real” effect of the intervention. Unfortunately such studies are not possible for practical, economical, or ethical reasons.

Quasi-experimental study is an experimental study where at least one criteria for RTC is missing: either no random allocation of exposure or/and no control group.

**Example** (quasi-experimental design): Testing a new treatment in one group of ill patients and assessing whether the subjects are better or worse after the treatment (before-after design). No control group, in a sense the group serve as its’ own control in such a study. Or if the treatment is tested in a group of patients, which is compared to another group of patients who do not receive the treatment, but treatment allocation is not random.

Intervention studies are often considered as providing the most reliable evidence from epidemiologic research. This is due to the unique strength of randomization as a means of determining exposure status in a trial. When participants are allocated to a particular exposure group at random, such a strategy achieves, on average, control of all other factors that may affect disease risk. The ability to control both known and unknown influences makes the
randomized trial such a powerful epidemiologic strategy, especially for studying small or moderate effect. There are also particular concerns of costs and feasibility for intervention studies. Nevertheless, when well designed and conducted, intervention studies can indeed provide the most direct epidemiologic evidence on which to judge whether an exposure causes or prevents a disease.

Advantages and disadvantages of the cohort and the case-control designs

The cohort design might seem to be a more appealing choice than the case-control design. It has a simple, logical structure that leads to measurements of disease incidence for the exposed and unexposed groups, or each category of exposure if several are used. This design permits absolute as well as relative comparisons of disease incidence among the exposed and unexposed. The case-control design provides somewhat less information, in that usually only the relative risk can be estimated. Absolute measures of incidence can also be made in case-control studies, but only if the sampling fraction for the controls or the baseline incidence rate for the total study population is known.

Case-control studies, like cohort studies, are based on follow-up of incident cases in a certain study population during a defined observation period. Case-control studies, however, use exposure information from a sample of the study population rather than from the whole study population. Unless the disease incidence is very high, obtaining exposure information for a sample of the study population will be much less expensive and can yield more information on exposure, as fewer subjects need to be studied. The case-control design therefore makes investigations based on large study populations more feasible, an important consideration since large studies is usually needed to reduce random error.

The disadvantage of the case-control design is the difficulty in selecting a satisfactory control group, with the consequent problem of introducing systematic error in the study by the selection of controls. This problem is an issue only for non-random selection of controls. When the control group is defined as a random sample of the study population, control selection is a simple technical procedure and introduces no systematic error beyond what would be present in a cohort study using the entire study population. Using information from sample (controls) rather than from the entire study population does increase, random variability, but when the size of the control group is adequate, this effect is small or negligible. The amount of random
error that can be removed by expanding the control group to include the entire study population is often trivial, whereas the corresponding cost would be great.

In case-control studies involving contact with study subjects or their relatives, questions about previous exposure are usually answered by the study subjects only after the cases have fallen ill. The cases, therefore, may have spent more time thinking about past exposures and causes of their disease, while the controls have no motivation to do so. This difference between cases and controls in the accuracy and completeness of exposure information can introduce a "recall bias" into the study. A similar bias may be seen in cohort studies in which, for some reason, exposure information is obtained only after the cases have been identified. In Table IV the advantages and disadvantages of the different study designs are presented.
Table IV. Advantages and disadvantages of the different study designs.

<table>
<thead>
<tr>
<th>Study design</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecological study</td>
<td>• is quick and cheap if the population based data have been routinely collected and is readily available</td>
<td>not possible to establish the temporal sequence and quality of data routinely collected</td>
</tr>
<tr>
<td></td>
<td>• is a useful first step to explore an association</td>
<td></td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td>- usually good control over selection of study subjects; random sampling easy; representative sample yields prevalence relative risk quick and cheap can study several outcomes and exposures at the same time</td>
<td>- does usually large sample exposures does not yield results often not outcome is time consuming long delay for results ethical concerns</td>
</tr>
<tr>
<td>Cohort study</td>
<td>- logic temporal sequence from exposure to outcome (in most cases we can be certain that the exposure came before the outcome) establishes the absolute risk exposure can be measured without bias because the outcome is not known at the time when exposure is measured can assess multiple outcomes potential confounders can be measured easier than in a case control study</td>
<td>- often not outcome is time consuming long delay for results ethical concerns</td>
</tr>
<tr>
<td>Case control study</td>
<td>- valuable for studying rare conditions</td>
<td>- limited to the exposure area</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
<td>- can be carried out in a short time</td>
<td>- does usual exposure area</td>
</tr>
<tr>
<td></td>
<td>- relatively inexpensive</td>
<td>- does not yield</td>
</tr>
<tr>
<td></td>
<td>- relatively small number of study subjects required</td>
<td>- does not yield</td>
</tr>
<tr>
<td></td>
<td>- can study several exposure variables at the same time</td>
<td>- potential bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- from selection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- in measurability</td>
</tr>
</tbody>
</table>

| Randomized controlled trial            | - with randomization, comparability of groups is very likely, for both known and unknown confounders | - may be exposed to greater situation (p from those) |
|                                        | - the temporal sequence can be established | - may be unstandardized practice |
|                                        | - provides the best evidence for a causal relationship/efficacy of treatments | - this may mean |
|                                        |                                           |                               |
Testing an epidemiological hypothesis first involves consideration of the concept of “association” between a particular exposure and a disease. Association refers to the statistical dependence between two variables, that is, the degree to which the rate of disease in persons with a specific exposure is either higher or lower than the rate of disease among those without that exposure. The presence of an association, however, in one of cause and effect. Making judgments about causality from epidemiological data involves a chain of logic that addresses two major areas:

- whether for any individual study, the observed association between an exposure and disease is valid
- whether the totality of evidence taken from a number of sources supports judgment of causality

The first - related to the question whether the particular study findings reflect the “true” relationship between the exposure and the disease – is a matter of determining the likelihood that alternative explanations such as:

- chance
- bias
- confounding

could account for the findings.

Causal Inference

Contemporary epidemiologists are concerned with determining the etiology of disease. A cause is an event that, either alone or in conjunction with other elements produces a sequence of other events that result in an effect. Patients, referring physicians, employers, and other frequently ask Orthopedists: “What is the cause of this disease?” It is hopeless to attempt to answer that question; however, we may be able to answer other questions, such as "What are some of the causal contingencies in a particular disease?" and "Is there a known factor, a risk factor for this disease?" Epidemiologists prefer to use the term risk factor instead of cause to
indicate an attribute or exposure that is related to an increased probability of a disease. Additionally, for a factor to be considered a risk factor it must precede the occurrence of the disease and the observed association must not be due to problems with the study design or the analysis of the data.

The important, or primary, causal contingency may depend on one's viewpoint. Each observer selects the causal contingency on the basis of his or her perspective. It is compelling to search for a primary, or the most important, causal factor among many. From a practical standpoint, if the goal is to prevent a disease it may be more beneficial to focus on a causal factor somewhat remote from the disease. For instance, general improvement in living conditions and economic development in underdeveloped countries can do more to reduce the incidence of tuberculosis and its orthopedic manifestations than can any vaccination, chemotherapy, or operative procedure.

Epidemiologists, more than other medical scientists, are concerned with philosophical theories rather than the purely technical aspects of the experimental method because experiments play a minor role in the analysis of naturally occurring phenomena. Causal inference is the logical development of a theory based on observations and a series of arguments that attributes the development of a disease to one or more risk factors.

Thus, testing hypotheses and discarding those that fail advances knowledge. The epidemiologist starts with a hypothesis, collecting data about a disease that then are used either to refute or to accept the hypothesis. The alternation between the generation of hypotheses and the collection of data allows many hypotheses to be discarded without any experiment being performed. Thus, experiments are reserved for the testing of hypotheses derived observational studies. Judgments about causality in epidemiology depend on new knowledge.

The probability that an association exists is the first criterion used in causal inference in epidemiology. For an association to be considered causal, the cause must precede the effect (the property of time-order) and there must be an asymmetrical direction such that the cause leads to the effect.

Statistical Inference
Statistical inference is part of the basis of epidemiology because the observations studied by epidemiologists are subject to random fluctuations. The testing of hypotheses is a statistical procedure to determine the probability that the data that are collected are consistent with the specific hypothesis under investigation. Make example of hypothesis formulation. The opposite of $H_0$ is a hypothesis, $H_\alpha$, that states: ….. Conventionally, the investigator seeks to deny or nullify $H_0$; thus, it is known as the null hypothesis. $H_\alpha$ and $H_0$ are mutually exclusive and exhaustive: that is, one or the other must be true, but they cannot both be true. Thus, if $H_0$ is denied, then $H_\alpha$ is affirmed. $H_\alpha$ and $H_0$ refer to the entire population, even though data are available only from one sample of patients.

The significance test is based on the calculation of a test statistic (for example, the t value, z value, or chi-square value) and on some theoretical assumptions. One such assumption is that the null hypothesis is true, and another is that sampling uncertainty is random. On the basis of these assumptions, it is possible to calculate how likely or unlikely the outcome observed in the sample would be. The test statistic, therefore, is a number that compares the observed and expected values of the parameter being measured under the null hypothesis. The significance test encompasses the rationale that there is some range of test-statistic values such that either the assumptions of the test, including the null hypothesis, are true or a rare event has occurred or one of the assumptions is untrue and, specifically, the null hypothesis is false. The point at which a test-statistic value is rare enough to warrant rejection of the null hypothesis is determined by convention but typically is set at the value that would occur no more than 5% or 1% of the time in repeated tests if the null hypothesis were true. This absolute value, which must be exceeded in order for the null hypothesis to be rejected, is called the critical value. The probability that the test-statistic value is too small to be consistent with $H_0$ being true is known as the significance level, or $\alpha$; it is conventionally set at either 0.05 or 0.01 and is commonly called the p value. In summary, if the test statistic represents an occurrence of less than 5% (or 1%) of the time under random sampling if the null hypothesis were true, then the result is considered significant and the null hypothesis is rejected in favor of $H_\alpha$.

Many tests of significance have been used, depending on the nature and distribution pattern (discrete, continuous, categorical, and so on) of the data being analyzed. Often, an investigator selects the appropriate test, calculates the test statistic (for example, the t value, z value, or chi-square value) and the significance level (the p value) with use of the collected data, and either rejects or fails to reject the null hypothesis.
When data are analyzed for all possible statistical associations without a previous specific hypothesis, the likelihood that an investigator will find a significant association by chance alone increases as the number of statistical tests that are performed increases.

There are two types of error associated with the significance test. It is possible to reject the null hypothesis when a rare event has occurred even though the null hypothesis is actually true. This is called a type-I error. In all cases in which a null hypothesis is true, a type-I error will occur \((100 \times \alpha)\) per cent of the time, where \(\alpha\) equals the significance level (usually 0.05 or 0.01). In a type-II error, the null hypothesis is false but the calculated test statistic is not significant and therefore is determined to be consistent with the null hypothesis being true. The null hypothesis is thus accepted in error. The relative frequency with which a type-II error occurs is symbolized by \(\beta\). An experiment usually is designed to control the probability of \(\alpha\) to be less than 0.20.

Power is the complement of a type-II error, or the probability that the null hypothesis will be rejected when it is indeed false and is equal to \((1-\beta)\). Power is a function of the level of significance, the reliability of the sample data (the degree of spread in the data or the standard deviation), and the size of the experimental effect. A power analysis should be considered if no significant difference can be found between the two groups being compared in a study. The groups may not be enough to allow detection of a significant difference (the null hypothesis being that there is a difference), and the power analysis will demonstrate the probability that the null hypothesis has been correctly rejected. The probability of a type-I or type-II error is inversely related for any determined experimental design and fixed sample size. Improving the experimental design and increasing the sample size can increase power.

Estimation is another tool used in statistical inference concerned with calculation of the values of specific population parameters. A point estimate (the sample mean) can be used to estimate the population mean. A weakness in the point estimate is its failure to make a probability statement regarding how close the estimate is to the population parameter. A confidence interval estimate remedies this problem by providing an interval of plausible estimates of the population mean as well as a best estimate of its precise value. The confidence interval that is conventionally chosen is 95 or 99 per cent, similar to the conventional choice of 0.05 or 0.01 for the level of significance. The 95% confidence interval means that, assuming the sample mean will follow an approximately normal distribution, 95% of all sample means based on a given sample size will fall within +1.96 standard errors of the
population mean. This number is derived from tire mathematics, of normal curve distribution. Similarly, 99% of all sample means based on a given sample size will fall within +2.576 standard errors of the population mean (the 99% confidence interval). A confidence interval, therefore, is a range of values for a study variable specifying the probability (usually 95%) that the true value of the variable is included within the range. The size of a confidence interval gives some idea of the precision of the point estimate in a way that is not offered by a p value; the narrower the confidence interval, the more precise the data. As the sample size increases, the size of the confidence interval decreases. As the standard deviation increases, reflecting the increased variability between individual observations, the size of the confidence interval increases. As the desired level of confidence increases (for example, from 95 to 99 per cent), the size of the confidence interval increases.

Statistical reasoning is based on the precepts that natural processes can be described by stochastic models and that the study of random collections of individuals will allow identification of "systematic patterns of scientific import. The truth or falsehood of a hypothesis cannot be inferred from a significance test. The type-I and type-II error rates define a critical region for the summary statistic, which represents a decision rule as to whether the null hypothesis is to be rejected or not. Goodman (1) warned that a decision rule tells nothing about whether a particular hypothesis is true; it says only that, if investigators behave according to such a rule, in the long run they will reject a true hypothesis not more, for example, than once in 100 times and they may have evidence that they will reject the hypothesis sufficiently often when it is false. Poor data are often irrelevantly supported by sophisticated statistical techniques. A significance test makes no assumption about the plausibility of the null hypothesis. Investigators must focus on the substance of the issues being studied and must not become sidetracked by the mechanics of data analysis. Judgment must take precedence over statistical inference when large sources of error are not quantified in the statistical analysis. These sources of error are collectively known as bias and deserve careful consideration in the design, execution, and analysis of all studies.

Bias

Bias is defined as "any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth". (1) Bias can lead to an incorrect estimation of the association between an exposure and the risk of a disease. Such an association is considered real if all attempts to explain it away as due to bias
have failed. There are three broad categories of bias: confounding bias, selection bias, and information bias.

Confounding is a distortion in an effect (such as relative risk) that results from an effect of another variable (the confounder) associated with the disease and exposure being studied. Confounding can lead to an overestimate or an underestimate of the true association between a disease and an exposure and even can change the direction of the observed effect. For a factor to be a confounder, it must, in and of itself, be a risk factor for the disease in the unexposed population and it must be associated with the exposure variable in the population from which the cases were derived. In addition, it must not be an intermediate step in the causal pathway between the exposure and the disease. For example, if a study were to be designed to explore whether smoking is a risk factor for motor-vehicle accidents, consumption of alcohol would be considered a confounder because it is both a risk factor for motor-vehicle accidents and it is associated with smoking. During the introduction of a new operative procedure, the good-risk patients may be selectively managed with the procedure while the poor-risk patients may receive the standard treatment. This is called a confounding by indication bias. Age and gender are well recognized confounders.

An effect modifier is a factor that changes the magnitude of an effect measure (for example, relative risk or odds ratio). Effect modification differs from confounding: the latter is a bias that the investigator tries to prevent or remove from the data, whereas the former is a constant of nature. For example, immunization status against serum hepatitis is an effect modifier for the consequences of being stuck by a needle that was used on a person infected with hepatitis. Immunization status is an effect modifier because people who are immunized are less likely to contract hepatitis than those who are not.

Selection bias (also known as detection bias and unmasking bias) refers to a distortion in the estimation of an effect due to systematic differences in characteristics between subjects who are selected for a study and those who are not. Selection bias can result when a procedure used to identify a disease varies with the exposure status. For example, this bias can be introduced by an examiner who performs a clinical evaluation without being blinded to the disease or exposure status of the subject. The results of the evaluation may differ if the examiner expects the disease to be present in the patients and absent in the controls. A lack of a response to a questionnaire or loss of patients to follow-up would not be serious problems if they merely resulted in a reduction in the number of subjects available for study; however, they
may result in selection bias if the respondents and the non-respondents of the patients being followed and those who have been lost differ with respect to some characteristic being studied.

Information bias (also known as observational bias and misclassification bias) refers to a distortion in the estimation of an effect that results from error in the measurement of either an exposure or a disease or from the misclassification of subjects with regard to at least one variable. In describing inaccuracy of measurement, two types of misclassification bias can occur: non-differential misclassification, when the inaccuracy is the same for the two study groups (for example, the patients and controls in a case-control study), and differential (non-random) misclassification, when the inaccuracy differs between groups (for example, when an exposure measure, such as repetitive work tasks, is determined more accurately among the patients than among the controls). Non-differential misclassification increases the similarity between the exposed and unexposed groups. Any association between exposure and disease will be underestimated, so the observed estimate of effect is said to be biased toward the null value of 1.0, meaning that if exposure and disease had no association the expected relative risk would be equal to 1.0. Differential misclassification leads to a biased risk estimate that may be either away from (an overestimate) or toward (an underestimate) the null value. Whereas confounding bias is generally correctable in the analysis stage of a study, selection and information biases may not be correctable.

Study-Design Strategies for Determining the Relationships between Exposure and Disease

Case-control and cohort studies are the two basic types of observational analytical study designs that have as their objective the testing of hypotheses.

This study design is good for the investigation of rare diseases but is highly susceptible to selection and recall bias. Recall bias occurs when a study relies on the patient's memory to determine exposure status because a patient who has a disease is more likely to remember possible exposures than a healthy person.

Example: Probably the most impressive prospective cohort study is the Framingham study, which began around 1950. Coronary heart disease and its consequences were thought by the medical profession to be inevitable changes of aging; however, clinical observations and descriptive epidemiological studies suggested that preventable environmental factors played a
role. More than 5000 people who did not have coronary artery disease were enrolled in the long-term study, originally planned to last for twenty years, and each participant was given a comprehensive medical examination every two years. The findings of this study have continued to emerge, delineating the risk factors for heart disease and associated atherosclerotic and non-atherosclerotic disorders, and they have led to preventive and treatment paradigms that have resulted in better health care.

Example (Retrospective (historical) cohort study): It would be unethical to test the hypothesis that low levels of exposure to radiation shorten human life expectancy with use of a prospective study; however, if a group of people (a cohort) that already has been exposed can be identified, then even if the exposure was in the past a retrospective cohort study would be feasible. Seltser and Sartwell undertook such a study, comparing members of the Radiological Society of North America with members of other medical specialty societies. These authors demonstrated that the death rate was highest among radiologists, intermediate among internists, and lowest among ophthalmologists and otolaryngologists; improvements in equipment and safety techniques showed a so-called disappearing effect in the latter part of the study.

The advantages of a cohort study include a temporal sequence of exposure and disease that is usually clear, a minimization of observational bias in determining exposure, the ability to examine multiple effects of one exposure, and usefulness when exposure is rare. Cohort studies are time-consuming, expensive, not suitable for the investigation of rare diseases, and potentially biased with regard to loss of subjects to follow-up when subjects must be followed for many years. A case-control study can be tested within a cohort study. In this study design, patients and controls are drawn from the population in a cohort study, as has been done frequently in the Framingham study. Nested case-control studies obviate the issue of recall bias because the data are collected prospectively.

The advantage of intervention study (clinical trials) design is that known and unknown confounders are distributed, on the average, equally among the study groups. Susceptibility bias may occur in a clinical trial if the two groups are dissimilar in terms of their initial state. A randomized clinical trial precludes the possibility of susceptibility bias.
A double-blind randomized trial is the best way to test the hypothesis that Coumadin (warfarin) prevents deep-vein thrombosis after total hip arthroplasty because randomization minimizes confounding and blinding eliminates selection bias.

A common practice in the epidemiological literature is the documentation of unavoidable departures from the ideal study design as well as a discussion of estimates of the magnitude and direction of biases and deduction of the extent to which these errors threaten or do not threaten validity. Gartland reviewed ten articles on the long-term follow-up of patients after primary total hip arthroplasty to determine the strategy that had been used in the design of each study. He concluded that all of the studies were deficient in design, were flawed by confusing data, and contained results of doubtful validity. The quality of orthopaedic studies and literature will improve if sufficient attention is paid to alternative explanations and there is recognition of the central role that epidemiological principles play in advancing knowledge and understanding.

References