Epidemiology is the study of the distribution and determinants of disease or other health-related states or events in specified populations and the application of this study to the control of health problems. The key component of this definition is that epidemiology focuses on populations, an emphasis that distinguishes epidemiology from clinical case studies, which focus on individual subjects. Health events can be characterized by their distribution (descriptive epidemiology) and by factors that influence their occurrence (analytic epidemiology). In both descriptive and analytic epidemiology, health-related questions are addressed using quantitative methods to identify patterns or associations from which inferences can be drawn and interventions developed, applied, and assessed.

**DESCRIPTIVE EPIDEMIOLOGY**

**Surveillance**

The goals of descriptive epidemiology are to define the frequency of health-related events and determine their distribution by person, place, and time. The foundation of descriptive epidemiology is surveillance, or case detection. Retrospective surveillance identifies health events from existing data, such as clinical or laboratory records, hospital discharge data, and death certificates. Prospective surveillance identifies and collects information about cases as they occur, for example, through ongoing laboratory-based reporting.

With passive surveillance, case reports are supplied voluntarily by clinicians, laboratories, health departments, or other sources. The completeness and accuracy of passive reporting are affected by whether reporting is legally mandated, whether a definitive diagnosis can be established, illness severity, interest in and awareness of the medical condition among the public and the medical community, and whether a report will elicit a public health response. Because more severe illness is more likely to be diagnosed and reported, the severity and clinical spectrum of passively reported cases are likely to differ from those of all cases of an illness. Passively collected reports of nationally notifiable diseases are tabulated in the Morbidity and Mortality Weekly Report (http://www.cdc.gov/mmwr/).

In active surveillance, an effort is made to ascertain all cases of a condition occurring in a defined population. Active case finding can be prospective (through routine contacts with reporting sources), retrospective (through record audit), or both. Population-based active surveillance, in which all cases in a defined geographic area are identified and reported, provides the most complete and unbiased ascertainment of disease and is optimal for describing the rate of a disease and its clinical spectrum. By contrast, active surveillance conducted at only one or several participating facilities can yield biased information on disease frequency or spectrum based on the representativeness of the patient population and the size of the sample obtained.

**Case Definition**

Establishing a standard case definition is an important first step for surveillance and description of the epidemiology of a disease or health event. Formulation of a case definition is important, particularly when laboratory diagnostic testing results are not definitive. More restrictive case definitions minimize misclassification, but they can exclude true cases and are most useful when investigating a newly recognized condition, in which the ability to determine etiology, pathogenesis, or risk factors is decreased by inclusion of noncases in the study population. A more inclusive definition can be important in an outbreak setting to detect potential cases for further investigation or when preventive interventions can be applied. Multiple research or public health objectives can be addressed by developing a tiered case definition that incorporates varying degrees of diagnostic certainty for definite and probable cases.

**Sensitivity, Specificity, and Predictive Value**

Sensitivity, specificity, and predictive values can be used to quantify the performance of a case definition or the results of a diagnostic test or algorithm (Table 1.1). Unlike sensitivity and specificity, predictive values vary with the prevalence of a condition within a population. Even with a highly specific diagnostic test, if a disease is uncommon among those people tested, a large proportion of positive test results will be false positive, and the positive predictive value will be low (Table 1.2). If the test is applied more selectively such that the proportion of people tested who truly have disease is greater, the test’s predictive value will be improved. Thus, sensitivity and specificity are characteristics of the test, whereas predictive values depend both on test sensitivity and specificity and on the disease prevalence in the population in which the test is applied. Often, the sensitivity and specificity of a test are inversely related. Selecting the optimal balance of sensitivity and specificity depends on the purpose for which the test is used. Generally, a screening test should be highly sensitive, whereas a follow-up confirmatory test should be highly specific.

**Incidence and Prevalence**

Characterizing disease frequency is one of the most important aspects of descriptive epidemiology. Frequency measures typically include a count of new or existing cases of disease as the numerator and a quantification of the population at risk as the denominator. Cumulative incidence is expressed as a proportion and describes the number of new cases of an illness occurring in a fixed at-risk population over a specified period of time, generally 1 year unless otherwise stated. Incidence density is the rate of new cases of disease in a dynamic at-risk population; the denominator typically is expressed as the population-time at-risk (e.g., person-time).

Because the occurrence of many infections varies with season, extrapolating annual incidence from cases detected during a short observation period can be inaccurate. In describing the risk of acquiring illness during a disease outbreak, the attack rate, defined as the number of new cases of disease occurring in a specified population and time period, is a useful measure. Finally, the case-fatality rate, or proportion of cases of a disease that result in death, is used to quantify the mortality resulting from a disease in a particular population and time period.

Prevalence refers to the proportion of the population having a condition at a specific point in time. As such, it is a better measure of disease.
burden for chronic conditions than is incidence or attack rate, which identify only new (incident) cases. Prevalent cases of disease can be ascertained in a cross-sectional survey, whereas determining incidence requires longitudinal surveillance. When disease prevalence (P) is low, assessing disease burden becomes a challenge. However, in populations with high prevalence, disease burden is caused by an environmental exposure or is vector borne, or during an outbreak with a point source exposure. Time also is a useful descriptor of disease occurrence. Evaluating long-term (secular) trends provides information that can be used to identify emerging health problems or to assess the impact of prevention programs. The timing of illness in outbreaks can be displayed in an epidemic curve and can be useful in defining the mode of transmission of an infection or its incubation period and in assessing the effectiveness of control measures.

Describing Illness by Person, Place, and Time

Characterizing disease by person, place, and time is often useful. Demographic variables, including age, sex, socioeconomic status, and race or ethnicity, often are associated with the risk of disease. Describing a disease by place can help define risk groups, for example, when an illness is caused by an environmental exposure or is vector borne, or during an outbreak with a point source exposure. Time also is a useful descriptor of disease occurrence. Evaluating long-term (secular) trends provides information that can be used to identify emerging health problems or to assess the impact of prevention programs. The timing of illness in outbreaks can be displayed in an epidemic curve and can be useful in defining the mode of transmission of an infection or its incubation period and in assessing the effectiveness of control measures.

**ANALYTIC EPIDEMIOLOGY**

**Study Design**

The goal of analytic studies is to identify predictors of an outcome. This goal can be addressed in experimental or epidemiologic (observational) studies. In addition, ecologic or trend studies can be used to assess predictors when the frequency or distribution of an outcome has changed over time or differs among populations.

In *experimental studies*, hypotheses are tested by systematically allocating an exposure of interest to subjects in separate groups to achieve the desired comparison. Such studies include randomized, controlled, double-blind treatment trials, as well as laboratory experiments. By carefully controlling study variables, investigators can restrict differences among groups and thereby increase the likelihood that the observed differences are a consequence of the specific factor being studied. Because experiments are prospective, the temporal sequence of exposure and outcome can be established, making it possible to define cause and effect.

By contrast, *epidemiologic studies* test hypotheses using observational methods to assess exposures and outcomes among individual subjects in populations and to identify statistical associations from which inferences regarding causation are drawn. Although observational studies cannot be controlled to the same degree as experiments, they are practical in circumstances in which exposures or behaviors cannot be assigned. Moreover, the results often are more generalizable to a real population having a wide range of attributes. The 3 basic types of observational studies are cohort studies, cross-sectional studies, and case-control studies (Table 1.3). Hybrid study designs, incorporating components of these 3 types, also have been developed. In planning observational studies, care must be taken in the selection of participants to minimize the possibility of bias. Selection bias results when study subjects have differing probabilities regarding exposure to the variable under evaluation. In contrast to experimental or observational studies that analyze information about individual subjects, *ecologic studies* draw inferences from data on a population level. Inferences from ecologic studies must be made with caution because populations differ in multiple ways and therefore, inferences from ecologic studies also are made with caution because populations differ in multiple ways and populations are exposed to environments that differ in multiple ways and because relationships observed on a population level do not necessarily apply on the individual level (a problem known as the *ecologic fallacy*).

Because of these drawbacks, ecologic studies are suited best for generating hypotheses that can be tested using other study methods.

**Cohort Studies**

In a *cohort study*, subjects are categorized based on their exposure to a suspected risk factor and are observed for the development of disease. Associations between exposure and disease are expressed by the *relative risk* of disease, or *risk ratio*, in exposed and unexposed groups (Table 1.4). Cohort studies typically are prospective, with exposure defined before disease occurs. However, cohort studies also can be retrospective, in which the cohort is selected after the outcome has occurred. In this case, exposures are determined from existing records that preceded the outcome, and thus the directionality of the exposure-disease relationship is still forward. Characterizing exposures before development of disease is a major benefit of cohort studies because this approach minimizes selection bias and simplifies inference of cause and effect. Another advantage of cohort studies is that they can be used to assess multiple potential outcomes resulting from an exposure. However, in a cohort study it can be difficult to investigate multiple exposures as risk factors for a single outcome. Cohort studies also are impractical for studying
increases from 0 at enrollment (Fig. 1.1B). These 2 approaches are related
alative hazard
ity is initially 1 and declines in a step-function as the outcomes of interest
expressed in a
approach to assess or compare the impacts of preventive or therapeutic
of time-to-event data for outcomes such as death or illness is a powerful
but also, for those experiencing the outcome, on when it occurs. Analysis
new hypotheses about possible exposure-disease relationships.
and the onset of clinical illness. In general, cohort studies are unsuited
rare diseases or conditions with a long latent period between exposure
and the onset of clinical illness. In general, cohort studies are unsuited
for investigating risk factors for new or rare diseases or for generating
new hypotheses about possible exposure-disease relationships.
Cohort studies provide data not only on whether an outcome occurs
but also, for those experiencing the outcome, on when it occurs. Analysis
of time-to-event data for outcomes such as death or illness is a powerful
approach to assess or compare the impacts of preventive or therapeutic
interventions. The probability of remaining event-free over time can be
expressed in a Kaplan-Meier survival curve where the event-free probability
is initially 1 and declines in a step-function as the outcomes of interest
occur (Fig. 1.1A). Time-to-event data also can be displayed as the cumula-
itive hazard of an event occurring among members of a cohort that
increases from 0 at enrollment (Fig. 1.1B). These 2 approaches are related
in that the hazard reflects the incident event rate, whereas survival reflects
the cumulative nonoccurrence of that outcome.6,7 With time-to-event
analysis, the association between exposure and disease often is expressed
as a hazard ratio. Like relative risk, hazard ratio is a comparative measure
of risk between exposed and unexposed groups. The primary difference
is that the hazard ratio compares event experience over the entire time
period, whereas the relative risk compares event occurrence only at the
study endpoint.6

Cross-Sectional Studies

In a cross-sectional study, or survey, a sample is selected, and at a single
point in time exposures and outcome are determined. Outcomes can
include disease status or behaviors and beliefs, and multiple exposures
and the onset of clinical illness. In general, cohort studies are unsuited
for investigating risk factors for new or rare diseases or for generating
new hypotheses about possible exposure-disease relationships.
Cohort studies provide data not only on whether an outcome occurs
but also, for those experiencing the outcome, on when it occurs. Analysis
of time-to-event data for outcomes such as death or illness is a powerful
approach to assess or compare the impacts of preventive or therapeutic
interventions. The probability of remaining event-free over time can be
expressed in a Kaplan-Meier survival curve where the event-free probability
is initially 1 and declines in a step-function as the outcomes of interest
occur (Fig. 1.1A). Time-to-event data also can be displayed as the cumula-
itive hazard of an event occurring among members of a cohort that
increases from 0 at enrollment (Fig. 1.1B). These 2 approaches are related
in that the hazard reflects the incident event rate, whereas survival reflects
the cumulative nonoccurrence of that outcome.6,7 With time-to-event
analysis, the association between exposure and disease often is expressed
as a hazard ratio. Like relative risk, hazard ratio is a comparative measure
of risk between exposed and unexposed groups. The primary difference
is that the hazard ratio compares event experience over the entire time
period, whereas the relative risk compares event occurrence only at the
study endpoint.6

Cross-Sectional Studies

In a cross-sectional study, or survey, a sample is selected, and at a single
point in time exposures and outcome are determined. Outcomes can
include disease status or behaviors and beliefs, and multiple exposures

TABLE 1.3 Types of Observational Studies and Their Advantages and Disadvantages

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Design and Characteristics</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>Prospective or retrospective</td>
<td>Ideal for outbreak investigations in defined populations</td>
<td>Unsuitable for rare diseases or those with long latency</td>
</tr>
<tr>
<td></td>
<td>Select study group</td>
<td>Prospective design ensures that exposure preceded disease</td>
<td>Expensive</td>
</tr>
<tr>
<td></td>
<td>Observe for exposures and disease</td>
<td>Selection of study group is unbiased by knowledge of disease status</td>
<td>Can require long follow-up periods</td>
</tr>
<tr>
<td></td>
<td>Outcome measures used: Relative risk (RR) or hazard ratio (HR) of disease given exposure</td>
<td>RR and HR accurately describe risk given an exposure</td>
<td>Difficult to investigate multiple exposures</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>Nondirectional</td>
<td>Rapid, easy to perform, and inexpensive</td>
<td>Timing of exposure and disease can be difficult to determine</td>
</tr>
<tr>
<td></td>
<td>Select study group</td>
<td>Ideal to determine knowledge, attitudes, and behaviors</td>
<td>Biases can affect recall of past exposures</td>
</tr>
<tr>
<td></td>
<td>Determine exposure and disease status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome measures used: Prevalence ratio for disease given exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td>Retrospective</td>
<td>Rapid, easy to perform, and inexpensive</td>
<td>Timing of exposure and disease can be difficult to determine</td>
</tr>
<tr>
<td></td>
<td>Identify cases with disease</td>
<td>Ideal for studying rare diseases, those with long latency, new diseases</td>
<td>Biases can occur in selecting cases and controls and determining exposures</td>
</tr>
<tr>
<td></td>
<td>Identify controls without disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Determine exposures in cases and controls</td>
<td>OR only provides an estimate of the RR if disease is rare</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outcome measures used: Odds ratio (OR) for an exposure given disease</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1.4 Measures of Association, Risk, and Impact

<table>
<thead>
<tr>
<th>Type of Measure</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute risk of association (ARR)</td>
<td>(A/(A + B)) – (C/(C + D))</td>
<td>Difference in the incidence of the outcome between exposed and unexposed</td>
</tr>
<tr>
<td>Number needed to treat (NNT): Number of individual subjects who must receive an intervention (or exposure) to prevent one negative outcome</td>
<td>1/ARR</td>
<td></td>
</tr>
<tr>
<td>Relative risk or risk ratio (RR): Risk (probability) of a health event in those with a given exposure divided by the risk in those without the exposure</td>
<td>A/(A + B) / C/(C + D)</td>
<td></td>
</tr>
<tr>
<td>Odds ratio (OR): Odds of a given exposure among those with a health event divided by odds of exposure among those without the health event</td>
<td>AD/BC</td>
<td></td>
</tr>
<tr>
<td>Measure of impact</td>
<td>Population attributable fraction: The proportion of disease in a population that results from the specific exposure</td>
<td>[P_e (RR – 1)]/[1 + P_e (RR – 1)]</td>
</tr>
<tr>
<td></td>
<td>Vaccine efficacy (VE): The percentage of reduction in incidence of a disease among persons who have received a vaccine compared with the incidence in persons who have not received the vaccine</td>
<td>(1 – RR) x 100 or (1 – OR) x 100</td>
</tr>
</tbody>
</table>
can be evaluated as explanations for the outcome. Associations are characterized by the \textit{prevalence ratio}, similar to the risk ratio in cohort studies. Because neither exposures nor outcomes are used in selection of the study group, prevalence is an estimate of that in the overall population. Because neither exposures nor outcomes are used in selection of the study group, prevalence is an estimate of that in the overall population.

Causal Inference and the Impact of Bias

The impact of potential bias is particularly important in observational studies. The \textit{validity} of a study is the degree to which inferences drawn from a study are warranted. \textit{Internal validity} refers to the correctness of study conclusions for the population from which the study sample was drawn, whereas \textit{external validity} refers to the extent to which the study results can be generalized beyond the population sampled. The validity of a study can be affected by bias, or \textit{systematic error}, in selecting the study participants (sampling), in ascertaining their exposures, or in analyzing and interpreting study data. For errors to result in bias, they must be systematic, or directional. \textit{Nonsystematic error} (random misclassification) decreases the ability of a study to identify a true association but does not result in detection of a spurious association.

Several sources of bias can occur in selection of study participants (Box 1.1). \textit{Diagnosis bias} results when persons with a given exposure are more likely to be diagnosed as having disease than are people without the exposure; this can occur because diagnostic testing is more likely to 4

**Case-Control Studies**

In a \textit{case-control study}, the investigator identifies a group of people with a disease or outcome of interest (cases) and compares their exposures with those in a selected group of people who do not have disease (controls). Differences between the groups are expressed by an \textit{odds ratio}, which compares the odds of an exposure in case and control groups (see Table 1.4). The odds ratio is not the same as a risk ratio; however, it provides an estimate of the risk ratio if the disease or outcome in question is rare. Case-control studies are retrospective in that disease status is known and serves as the basis for selecting the 2 comparison groups; exposures are then determined by reviewing available records or by interview.

A major advantage of case-control studies is their efficiency in studying uncommon diseases or diseases with a long latency. Case-control studies also can evaluate multiple exposures that may contribute to a single outcome; study subjects frequently can be identified from existing sources (e.g., hospital or laboratory records, disease registries, or surveillance reports) and, after identification of suitable control subjects, data on previous exposures can be collected rapidly. Case-control studies also have several drawbacks. Bias can be introduced during selection of cases and controls and in determining exposures retrospectively, and inferring causation from statistically significant associations can be complicated by difficulty in determining the temporal sequence of exposure and disease.

**Box 1.1 Potential Sources of Bias in Observational Studies**

<table>
<thead>
<tr>
<th>Bias</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Selection bias}</td>
<td>Differential sampling of cases based on an exposure or a variable associated with an exposure.</td>
</tr>
<tr>
<td>\textit{Referral bias}</td>
<td>Differential admission to hospital based on an exposure or a variable associated with an exposure.</td>
</tr>
<tr>
<td>\textit{Diagnosis bias}</td>
<td>Differential use of diagnostic tests in exposed and unexposed.</td>
</tr>
<tr>
<td>\textit{Nonresponse bias}</td>
<td>Differential outcome or exposures of responders and nonresponders.</td>
</tr>
<tr>
<td>\textit{Survival bias}</td>
<td>Differential exposures between those who survive to be included in a study and those who die following an illness.</td>
</tr>
<tr>
<td>\textit{Misclassification bias}</td>
<td>Systematic error in classification of disease status.</td>
</tr>
</tbody>
</table>

\textit{Survey bias:} differential surveillance or reporting for exposed and unexposed. \textit{Diagnosis bias:} differential use of diagnostic tests in exposed and unexposed. \textit{Referral bias:} differential admission to hospital based on an exposure or a variable associated with exposure. \textit{Selection bias:} differential sampling of cases based on an exposure or a variable associated with exposure. \textit{Nonresponse bias:} differential outcome or exposures of responders and nonresponders. \textit{Survival bias:} differential exposures between those who survive to be included in a study and those who die following an illness. \textit{Misclassification bias:} systematic error in classification of disease status.

**A more complete listing is provided by Sackett.**
be done or because the interpretation of a test may be affected by knowledge of exposure status. For hospital-based studies, differential referral also can bias selection of a study sample. This bias would occur if, for example, the frequency of an exposure varied with socioeconomic status and a hospital predominantly admitted persons from either a high-income group or a low-income group. Bias also can occur when eligible subjects refuse to participate in a study as cases or controls.

Determination of exposures can be affected by several types of bias. Recall of exposures can be different for persons who have had an illness compared with people who were well. This bias occurs in either direction: patients may be more likely to remember an exposure that they associate with their illness (e.g., what was eaten before an episode of diarrhea) or less likely to recall an exposure if a severe illness affected memory. Interviewers can introduce bias by questioning cases and controls differently about their exposures. Misclassification of exposures can result from errors in measurement such as can occur with the use of an inaccurate laboratory test. Although systematic misclassification can result in bias, misclassification of exposure often is random rather than systematic.

Even a carefully designed study that minimizes potential biases can lead to erroneous causal inferences. An exposure can falsely appear to be associated with disease because it is closely linked to the true, but undetermined, risk factor. Race often is found to be associated with disease because it is closely linked to the true, but undetermined, risk factor. The number needed to treat (NNT) is a measure of the number of individual subjects who must receive a treatment to prevent a single negative outcome and is calculated as the reciprocal of the absolute risk reduction. In addition to absolute measures, relative measures also are useful for describing the strength of an association. In a cohort study or survey, the relative risk or risk ratio compares the risk of disease for subjects with versus subjects without an exposure (see Table 1.4). In case-control studies, association is assessed by the odds ratio, which compares the odds of exposure among subjects with and without a disease or health outcome; when disease is uncommon (<10%) in both exposed and unexposed groups, the odds ratio approximates the relative risk. For time-to-event analyses, the comparative risk is expressed as the hazard ratio. Odds ratios, relative risks, and hazard ratios greater than 1 signify increased risk given exposure, and values less than 1 suggest that exposure decreases the risk of an outcome. Because observational studies generally do not include all members of a population, these measures of association represent an estimate of the true value within the entire population. Statistical analyses can help guide investigators in making causal inferences based on point estimates of these measures of association.

Statistical Significance

Statistical tests are applied to assess the likelihood that the study results were obtained by chance alone rather than representing a true difference within the population. Most investigators consider a P value <0.05 as being statistically significant, indicating a less than 5% risk that the observed association is the result of chance alone (designated a type I error, the probability of which is the alpha level). Although use of this cutoff for significance testing has become conventional, ignoring higher P values can lead to missing a real and important association, whereas blind faith in the significance of lower P values can lead to erroneous conclusions. Statistical testing should contribute to, but not replace, criteria for evaluating possible causation.

Statistical significance also can be defined based on 95% confidence intervals, which approximately correspond to a P value of 0.05. An odds ratio, relative risk, or hazard ratio is considered statistically significant if the 95% confidence interval does not include 1. An advantage of using confidence intervals to define statistical significance is that they provide information on whether a finding is statistically significant and on the possible range of values for the point estimate in the population, with 95% certainty.

One pitfall in interpreting statistical significance is ignoring the magnitude of an effect in favor of its “significance.” A very large study can identify as significant a small, perhaps trivial, difference among study groups. Some epidemiologists have proposed that, despite statistical significance, odds ratios less than 2 or 3 in an observational study should not be interpreted because unidentified bias or confounding could have accounted for a difference of this magnitude. Conversely, the relative risk or odds ratio associating an exposure and outcome can be large, but if the exposure is uncommon in both groups, it cannot explain most cases of illness. The public health importance of an exposure can be described by the population-attributable fraction, or the proportion of the disease in a population that is related to the exposure of interest.

Sample Size

A study can fail to identify a true risk factor as statistically significant because the sample size was too small (designated a type II error, the probability of which is the beta level). Statistical power is defined as 1 – β and is the complement of type II error; that is, power is the probability of correctly identifying a difference of specified size among groups, if such a difference truly exists. The problem of inadequate sample size in clinical studies was highlighted in an analysis of “negative” randomized controlled trials reported in 3 leading medical journals between 1975 and 1990. Of 70 reports, only 16% and 36% had sufficient statistical power.
In calculating sample sizes for testing hypotheses, investigators must select type I and type II errors and define the magnitude of the difference that is deemed clinically important. Often, the type II error is set at 0.2, indicating acceptance of a 20% likelihood that a true difference exists but would not be identified by the study. Sample size calculations can be performed using a range of computer software. The program Epi-Info can be used to perform sample size calculations as well as other statistical functions and is available at no charge from the Centers for Disease Control and Prevention (www.cdc.gov/epiinfo/).

Ensuring an adequate sample size is particularly important for studies attempting to prove equivalence or noninferiority of a new treatment compared with standard therapy. Food and Drug Administration guidance recommends that noninferiority trials adopt a null hypothesis that a difference exists among treatments; this hypothesis is rejected if the lower 95% confidence limit for the new treatment is within a specified margin of the point estimate for standard therapy. Because the null hypotheses can never be proven or accepted, the failure to reject a null hypothesis of no difference among treatments or exposure does not prove equivalence. The importance of this distinction is illustrated by an analysis of 25 studies claiming equivalence of therapies for pediatric bacterial meningitis. Twenty-three studies claimed equivalence based on a failure to detect a significant difference among treatment groups. However, only three of these trials would exclude a 10% difference in mortality, thus showing that many studies potentially missed a clinically significant difference.

In some situations, an investigator would want to detect a significant difference among study groups as soon as possible, for example, when a therapeutic or preventive intervention could be applied once a risk group is identified or when concerns exist about the safety of a drug or vaccine. One approach to this situation is to include in the study design an interim analysis after a specified number of subjects are evaluated. Because the likelihood of identifying chance differences as significant increases with the number of analyses, it is recommended that the threshold for defining statistical significance should become more stringent as the number of planned analyses increases. If each interim analysis can lead the investigators to stop the trial, this study design is considered a group sequential method. Another example of a group sequential design is when concordance or discordance in outcome is tabulated for each matched set exposed to alternate treatments. Results for each set are plotted on a graph, and data collection continues until a preset threshold for a significant difference among study groups is crossed or no significant difference is detected at a given power.

### Statistical Inference

Statistical testing is used to determine the significance of differences among study groups, and it provides guidance on whether to accept or reject the null hypothesis. Although providing details of specific statistical tests is outside the scope of this chapter, Table 1.5 gives examples of statistical tests that can be applied in analyzing different types of exposure and outcome variables.

Using appropriate analytic and statistical methods is important in identifying significant predictors of an outcome (i.e., risk factors) correctly. Confounding variables are associated with the disease of interest and with other exposure variables that are associated with the disease and are not part of the causal pathway (Fig. 1.2). For example, if recent antimicrobial use is associated with childcare attendance, failure to adjust for recent antimicrobial use as a confounding variable could result in overestimating the relationship between childcare and antibiotic-resistant pneumococcal carriage. Effect modifiers interact with risk factors to affect their impact on outcome but may or may not themselves affect outcome. Frequently, age is an effect modifier, with an exposure associated significantly with an outcome in one age group but not in another.

Several approaches are used to control for confounding variables and effect modifiers. In study design, an extraneous variable can be controlled for by randomization, restricting sampling to one category of the variable or by frequency matching to obtain similar proportions of cases and controls in each stratum of the variable. A more extreme form of matching is to select control subjects who are similar to individual cases for extraneous variables (e.g., age, sex, underlying disease) and to analyze whether exposures are concordant or discordant within matched sets. A newer approach to study design is the case-crossover or case series analysis. In this method, exposures occurring in a defined risk period before the outcome are compared with exposures occurring outside the risk window for the same individual subjects. This approach has been adapted to the study of adverse events after vaccination. If the vaccine causes the event, the rate of the event will be greater within a defined risk window than predicted by chance alone based on the expected distribution of the event. The strength of this approach is that each subject, or case, serves as his or her own control, thereby decreasing confounding.

At the analysis stage, the impact of confounding variables and effect modifiers can be limited by performing a stratified analysis or using a multivariable model. In a stratified analysis, the possible association between a risk factor and an outcome is determined separately within different categories, or strata, of the extraneous variable. Stratum-specific estimates can be combined into a single estimate using an appropriate statistical test (e.g., a Mantel-Haenszel odds ratio). If a stratification variable is an effect modifier, the relative risk or odds ratio will differ substantially among the strata; for example, an exposure can be a strong risk factor in one age group but not another. In this setting, a summary statistic should not be presented, and results for each stratum should be presented separately. When the extraneous variable is confounding, an unstratified analysis can identify an exposure as a significant risk factor, whereas stratifying analysis by the confounding variable abolishes the apparent association in each stratum and indicates that the exposure is not an independent risk factor for disease.

Because stratified analyses become confusing rapidly as the number of strata increases, techniques of mathematical modeling have been developed that permit simultaneous control of multiple variables. Significant risk factors determined in a multivariable model are interpreted as each contributing independently and significantly to the outcome, thus controlling for confounding. Effect modification can be taken into account by including terms expressing the interaction between a risk factor and another risk factor.
effect modifier in the model. Various multivariable models are appropriate for discrete, continuous, and time-dependent outcomes. A limitation of multivariable modeling is multicoollinearity, which occurs when 2 or more explanatory variables of interest are highly correlated, and it can result in inaccurate measures of association and decreased statistical power. The risk of multicoollinearity can be reduced by assessing correlations among potential risk factors and selecting which variables to include in the model. Various methods to identify and minimize multicoollinearity have been developed.17

VACCINE EFFICACY STUDIES

Most prelicensure efficacy studies are experimental, randomized, double-blind, controlled trials in which vaccine efficacy (VE) is calculated by comparing the attack rates (AR) for disease in the vaccinated and unvaccinated groups: VE (%) = (AR unvaccinated − AR vaccinated) / AR unvaccinated × 100; or (1 − RR) × 100.

After licensure, conducting controlled studies, which requires withholding vaccine from a control group, is no longer ethical. Therefore further studies of efficacy must be observational rather than experimental, by comparing persons who have chosen to be immunized with those who have not. In case-control efficacy studies, vaccination status of persons with disease is compared with vaccination status of healthy control subjects. The number of vaccinated and unvaccinated cases and controls is included in a 2x2 table, and vaccine efficacy is calculated as 1 minus the odds ratio: VE (%) = (1 − OR) × 100. When the proportion of cases vaccinated is less than the proportion of vaccinated controls, the odds ratio is <1, and the point estimate for efficacy indicates that immunization is protective. The precision of the estimate is expressed by the 95% confidence interval. A lower 95% confidence limit that is greater than 0% indicates statistically significant protection; often investigators set power of vaccine efficacy studies so that the expected lower confidence limit is much greater than 0 and consistent with meaningful levels of protection. The most important component of a case-control efficacy study is selecting control subjects who have the same opportunity for immunization as do cases. If cases have less opportunity to be immunized, results will be biased toward showing protection. Factors such as low socioeconomic status, which can increase the risk of disease and decrease the chance of being immunized, are potential confounding variables and can be controlled for by matching control subjects to cases for those factors.

Cohort studies also can be used to determine vaccine efficacy after licensure. A study design called the indirect-cohort method was developed by researchers at the Centers for Disease Control and Prevention to evaluate the efficacy of the pneumococcal polysaccharide vaccine by using data collected by disease surveillance.18 The study cohort included persons identified with invasive pneumococcal infections. The study hypothesis was that if vaccine was protective, the proportion of vaccinated persons infected with pneumococcal serotypes that are included in the vaccine formulation would be less than the proportion of unvaccinated persons infected with vaccine-type strains. Vaccine efficacy was calculated from the relative serotype distributions overall and for each individual serotype. In a study of pneumococcal polysaccharide vaccine efficacy that used this approach, the point estimate of efficacy for preventing invasive infection was 57% (95% confidence interval, 45% to 66%)19; this estimate is similar to that obtained in a case-control efficacy study.20

DISEASE CONTROL AND PUBLIC HEALTH POLICY

Outbreak Investigations

Outbreak investigations require knowledge of disease transmission and use of descriptive and analytic epidemiologic tools. Possible outbreaks can be identified from surveillance data showing an increased rate of an infection or an unusual clustering of infection in person, place, and time. Comparing the incidence rate of disease with a baseline rate from a previous period is helpful in validating the occurrence of an outbreak. Other explanations for changes in the apparent rate of disease occurrence, such as diagnostic error, seasonal variations, and changes in reporting, must be considered. Using sensitive molecular methods to assess similarities among isolates from cases can be helpful in documenting that an apparent cluster of cases represents an outbreak because most outbreaks are caused by a single strain.

After establishing the presence of an outbreak, the next steps of an investigation are to develop a case definition, identify cases, and characterize the descriptive epidemiology. An epidemic curve depicts the number of cases over time and can provide information on possible transmission. In an outbreak with a point source exposure, an index case may be identified, with other cases occurring after an incubation period or at multiples of an incubation period (Fig. 1.3). Plotting the location of cases on a spot map can be helpful in determining possible exposures. Describing patients’ characteristics can be important in identifying at-risk populations for further investigation or targeting control measures, as well as for developing hypotheses that can be investigated in an analytic study.

Not all outbreaks can be traced to a point source exposure. Outbreaks and epidemics also can result from increased transmission of an endemic disease (i.e., a disease or condition that normally occurs in a specific population or area). In this situation it can be challenging to determine when an increase in disease constitutes an outbreak rather than a normal fluctuation in disease incidence. To determine whether an outbreak is occurring, the current incidence of the disease must be compared with the baseline disease incidence in that area. Often,

![Figure 1.3](image-url)
no standard definition exists for when an increase in endemic disease incidence constitutes an outbreak or epidemic. For instance, in US pertussis epidemics the threshold for declaring an epidemic has varied by state. In California in 2010 and 2014, an epidemic was declared when the statewide case counts had reached 5 times the number of cases observed in a year with baseline pertussis incidence. By contrast, in Washington State in 2012, an epidemic was declared when the incidence of pertussis reached 2 standard deviations above the statewide 10-year average (Fig. 1.4).

Cohort studies are optimal for investigating outbreaks that occur in small, well-defined populations, including in schools, childcare settings, social gatherings, and hospitals. In populations that are not well defined, a case-control study is the most feasible approach. It is important to select control subjects who had an opportunity equal to that of cases for exposure to potential risk factors and development of disease. When the number of cases is relatively small, enrolling multiple control subjects per case increases the power of the study to find significant risk factors. After a standard questionnaire is administered, significant risk factors are determined by comparing exposures of cases and controls. The results of analyses can lead to inferences of causation and development of prevention and control strategies or to further hypotheses that can be evaluated later. The impact of intervention can be determined by ongoing surveillance and continued plotting of additional cases on the epidemic curve.

Impact and Economic Analysis of Disease Prevention

Assessing health and economic impacts of public health interventions is important in developing or supporting policy decisions. Health impacts can be expressed directly as cases of disease, deaths, and sequelae prevented. Vaccine efficacy is a specific example of the prevented fraction (PF), where PF = P (1 – RR), with P representing the proportion exposed to an intervention. Secondary measures of health impact include years of potential life lost (YPLLs) or quality-adjusted life-years (QALYs) lost, which quantify the impact of death, or death and disability, respectively, based on the age at which these events occur. A measure of the efficiency of an intervention is the number needed to treat, which describes how many persons must be exposed to an intervention to avoid a single case of adverse health outcome.

Cost-effectiveness analyses determine the cost per health outcome achieved, such as the cost per death or complication averted, and it permits comparison of an intervention with other potential users of resources. In a cost-effectiveness formula, costs appear in the numerator, and health benefits appear in the denominator. The numerator includes expenditures for the prevention program from which cost savings occurring with disease prevention are subtracted. In addition to direct costs averted (e.g., savings from decreased medical care), indirect cost savings occur from increased productivity of people who do not become ill or miss time from work while receiving care or caring for ill family members. Cost-utility calculations are similar to cost-effectiveness but assess cost per quality-adjusted life-years saved or years of potential life lost averted.

Cost-benefit analyses differ from cost-effectiveness analyses in that the calculation is made entirely in economic terms. Health benefits are assigned an economic value, and expenditures are compared with savings. One problem with this approach lies in the difficulty of assigning an economic value to a health effect. For example, the value of a life saved may be quantified as the estimated value of a person’s earnings over his or her lifetime, forgone earnings as a result of premature death, or by a standard amount; both economic and ethical issues can be raised by the choice of approach. Because the parameters used in economic analyses often are uncertain or based on limited data, and because choices made by the investigator (e.g., regarding the value of life) can be influential to the analysis, sensitivity analyses often are performed in which parameters are varied across a range of potential values. In addition to defining a range of possible economic outcomes, sensitivity analyses can identify the factors that most strongly influence the results, thus elucidating where further studies may be important.

EVALUATING THE MEDICAL LITERATURE

Steps in reviewing published medical research are shown in Box 1.2.22 The ability to assess published studies carefully often is limited by the information presented in the report. To improve reporting of randomized controlled trials, a group of investigators and editors developed the Consolidated Standards of Reporting Trials (CONSORT)23 (http://www.consort-statement.org/) and later extended these recommendations to reporting randomized trials of noninferiority and equivalences.24 Reporting often still does not adhere to the quality standards proposed.25,26 Although the guidelines refer to experimental rather than observational studies, most criteria apply to observational studies as well.

In assessing the medical literature, it is important to examine all published work on a particular topic rather than relying on a single report and to evaluate the strength of the combined evidence from these reports. One framework for evaluating the literature on a topic is the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, adopted by the US Advisory Committee on Immunization Practices (ACIP) in 2010.
In the United States, all clinical trials must be registered at clinicaltrials.gov in addition to reviewing the published literature can provide a more complete record shows that the study was conducted even if the results are never reported, even if that outcome is "lost to follow-up." Intent-to-treat analyses consider outcomes for all enrolled subjects, whether or not they completed the therapy (e.g., subjects who were nonadherent to therapy or who received only part of a vaccination series). By contrast, per protocol analyses include only those subjects who followed the study protocol throughout the full study duration. Usually, an intent-to-treat analysis is considered more conservative than a per protocol analysis, but this is not necessarily the case for noninferiority trials.

Bias can have an important impact on study results and must be carefully considered. Approaches to minimize bias should be described clearly. The direction and potential magnitude of remaining bias should be estimated and its impact on results considered. Potential confounding, the presence of important unmeasured variables, and possible effect modification can have a major impact on the results. Investigators should openly discuss the potential limitations of the investigation and describe the strategies applied to overcome those limitations.

Interpretation of study results includes assessing the magnitude of the associations, their relevance to practice, and the likelihood that the relationships observed are causal. The importance of an exposure in explaining an outcome can be expressed by the attributable proportion. The quality of randomized trial reporting in leading journals can have an important impact on one's own patient population still must be assessed.

**Understanding a Medical or Epidemiologic Study**

Assessing the research hypothesis allows readers to determine the relevance of the study to their practice and to judge whether the analyses were done to test the hypothesis or to identify other associations of interest. The ability to make causal inferences from a confirmatory study that tests a single hypothesis is greater than from an exploratory study in which multiple exposures are considered as potential explanations for an outcome.

Several components of study design are important to consider. Details should be presented regarding the criteria for selecting the cohort or cases and controls. Exposure and outcome variables should be clearly defined, and the potential for misclassification and its impact should be considered. Quantifying exposure can be important to establish a dose-response relationship. Sample size estimates should be presented, making clear the magnitude of difference among study groups considered clinically meaningful and the type I and type II error levels.

In the analysis, it is important that outcomes for all study subjects are reported, even if that outcome is "lost to follow-up." Intent-to-treat analyses consider outcomes for all enrolled subjects, whether or not they completed the therapy (e.g., subjects who were nonadherent to therapy or who received only part of a vaccination series). By contrast, per protocol analyses include only those subjects who followed the study protocol throughout the full study duration. Usually, an intent-to-treat analysis is considered more conservative than a per protocol analysis, but this is not necessarily the case for noninferiority trials.

The appropriateness of the statistical tests should be assessed; for example, if data are not normally distributed, they can be transformed to a scale that is more normally distributed (e.g., geometric mean titers), or nonparametric statistical tests should be used. In assessing a multivariable model, the reader should critically evaluate the type of model chosen, the variables included, and whether interaction terms were considered. Missing data pose a particular problem in modeling in that study subjects can be included only if data are available for each variable in the model; thus the power of a multivariate model can be much less than that predicted in a sample size calculation.

**KEY REFERENCES**

4. Greenberg RS. Medical Epidemiology. Norwalk, CT, Appleton and Lange, 1993, p 120.
REFERENCES

22. Greenberg RS. Medical Epidemiology. Norwalk, CT, Appleton and Lange, 1993, p 120.