Chapter 35

Introduction to Epidemiology

Donna K. Arnett and Steven A. Claas

INTRODUCTION

Epidemiology concerns itself with the incidence, distribution, cause, control and prevention of disease. In contrast to other branches of medical science that focus on the health status of *individuals*, epidemiology is the study of the health status of *populations*. Epidemiology employs the tools of basic science (e.g., observation, hypothesis testing, statistics, correlation and causal reasoning) to protect, promote and restore health and well-being in populations. This chapter discusses the fundamental principles of descriptive and analytic epidemiology. Our goal is not to provide definitive descriptions of all elements of population-based medical research, but rather to explain the rudiments of the science of epidemiology and create a context in which topics presented in subsequent chapters can be understood. We define the concepts of incidence and prevalence, outline the basic epidemiological principles of risk and association, discuss the various types of studies used to conduct epidemiological research, and caution against forces and factors that may threaten the validity and reliability of such studies. We also consider the theoretical and practical constraints on making claims of causation in epidemiology. We explore the differences between traditional, public health epidemiology and clinical epidemiology. Finally, we discuss the role of the concepts of sex, gender, race and ethnicity in epidemiology, and we encourage health professionals to define and deploy these socially charged categories with care.

DEFINITION AND ROLE OF EPIDEMIOLOGY

Conceived somewhat narrowly, epidemiology concerns itself with the incidence, distribution, cause, control and prevention of disease. Conceived more liberally, epidemiology's gaze extends beyond disease to include any marker of health status, such as disability, injury, mortality, hospitalization, or quality-of-life measures (see Chapter 9). In contrast to other branches of medical science that focus on the health status of *individuals*, epidemiology is the study of the health status of *populations*. Because epidemiology serves as the scientific basis of public health, it is – at its very core – an example of translational science (see Chapters 34 and 36). Epidemiology employs the tools of basic science (e.g., observation, hypothesis testing, statistics, correlation and causal reasoning) to protect, promote, and restore health and well-being in populations.

The practice of epidemiology can be conceptually organized into two approaches – descriptive and analytic. *Descriptive epidemiology* seeks to answer fundamental questions related to the 'where,' 'when,' and 'who' dimensions of the disease of interest. (Throughout the remainder of this chapter, we will often use the shorthand term 'disease' to represent the more verbose, albeit more accurate, term 'health-related state or event'.) The spatial distribution of a disease provides one descriptive measure of the extent of the problem. Characterizing the spatial distribution of a disease can also be an important first step in determining (or at least hypothesizing) how or if the disease might spread through a population. Depending on the disease under study, the scale of distribution can vary greatly, ranging from a hospital wing to the entire globe. The temporal distribution of a disease can also be important. As with spatial scale, the temporal scale relevant in an epidemiological inquiry varies widely depending on the disease and other circumstances of interest. For example, a recent study that followed patients admitted on weekends and those admitted on weekdays for a first myocardial infarction found that admission on weekends was associated with higher mortality (Kostis *et al.*, 2007). At the other end of the temporal spectrum, epidemiologists often track disease occurrence over the course of many years to determine long-term (also known as 'secular') trends of occurrence in populations. Not only can characterizing the incidence of

disease relative to time suggest possible causal mechanisms, understanding the timing of disease can help public health officials predict and possibly prevent future occurrences. Finally, describing who is affected by a disease is fundamentally important. Like spatial and temporal dimensions, the salience of any personal variable will depend on the disease of interest. Personal characteristics may be inherent (e.g., age, genotype) or acquired (e.g., immune status, educational level). Personal characteristics also include behavioral patterns (e.g., diet, exercise) and social circumstance (e.g., occupation, income). As discussed below, personal characteristics are often markers of risk, and collecting such data is the first step in determining which characteristics may be operative in the context of a particular disease.

It should be clear from the discussion above that gathering descriptive data on the epidemiologic variables of space, time and person can have intrinsic value for public health practitioners. However, the observations of descriptive epidemiology also frequently prompt questions or hypotheses that become the starting point of *analytic epidemiology*. For example, descriptive data may reveal that cases of a rare respiratory illness clustered around a particular industrial site. Although the descriptive data may raise flags for public health officials, these data say nothing about the causal role of the industrial site or any other possible relationship between the disease and the environment (other than the obvious spatial relationship). To investigate these questions, epidemiologists must draw upon the analytical tools available to them in the form of various study designs and statistical tests. These are described in greater detail in the Types of Epidemiological Studies section below.

Analytic epidemiology (with sound descriptive underpinnings) can be put to many uses. For instance, epidemiology can be used to search for factors that correlate with, modify the risk of and cause disease in populations. (For the moment, we suspend knotty epistemological issues associated with determining causation.) For example, in 1978 public health officials in a number of states reported to the US Centers for Disease Control (CDC) cases of a sometimes fatal condition dubbed toxic-shock syndrome (TSS). Although *Staphylococcus aureus* was (correctly) implicated as the causal pathogen early in the description of the syndrome (Todd *et al.*, 1978), the factors underlying an apparent increase in the occurrence of the syndrome remained a mystery. As expected, descriptive data emerged first: in the first 55 cases reported to the CDC, 95% were women; the average age of cases was about 25 years; for the 40 patients with menstrual history data, 95% reported symptoms within 5 days of the onset of menses; *S. aureus* was isolated from vaginal cultures in two-thirds of patients (Centers for Disease Control, 1980). These observations prompted a flurry of TSS research conducted by epidemiologists, clinicians and laboratory scientists. Two CDC-sponsored case-control studies found statistically significant associations between TSS and the use of tampons (Shands *et al.*, 1982). Although these early TSS studies were fraught with methodological complications (Todd, 1982), after ten years of epidemiologic surveillance and analysis, researchers reported that the incidence of menstrual TSS had decreased and the drop could be explained by the withdrawal of a particular brand of tampon from the market as well as changes in the composition of tampons on the market (Schuchat and Broome, 1991). Menstrual TSS and its association with tampon use remains incompletely understood, and it would be reckless to make simple claims of causation. However, it is clear that the work of epidemiologists (in collaboration with others) played a significant role in identifying factors that contributed to the increased occurrence of TSS in the US population in the late 1970s.

Although it is true that epidemiologists often rely on clinicians to provide reliable and consistent diagnoses, it is also true that epidemiologists generate knowledge useful for clinicians. Obviously, as members of multidisciplinary teams searching for the cause of newly identified diseases, epidemiologists make clinically valuable contributions. Beyond the search for causes, however, epidemiologists also help create an accurate and useful clinical picture of a disease. For example, during the late 1940s investigators began following a population of over 5000 individuals in Framingham, MA in an effort to identify factors that contribute to cardiovascular disease. This long-term epidemiological work identified age, hypertension, smoking, diabetes and hyperlipidemia as coronary risk factors. These risk factors have subsequently been incorporated into algorithms that return global risk scores used by clinicians to assess and characterize patients' risk of cardiovascular disease (Wilson *et al.*, 1998; Hemann *et al.*, 2007).

Epidemiology is also used to test hypotheses about the natural history and routes of exposure to disease. The tools of analytical epidemiology can also be used to plan and assess interventions, ameliorations and other public health initiatives. Ultimately, the fruits of epidemiology are put to use by individuals – whenever someone chooses to stop smoking, start exercising, lower their dietary fat, or wash their hands during cold and flu season, they are using knowledge gained through epidemiology to make decisions that protect, promote and restore health.

The remainder of this chapter will discuss the fundamental principles of descriptive and analytic epidemiology. Many of the topics discussed below (such as experimental design) are covered in more detail elsewhere in this volume. Therefore, our goal is not to provide definitive descriptions of all elements of population-based medical research, but rather to explain the rudiments of the science of epidemiology and create a context in which subsequent chapters can be understood.

MEASURING OCCURRENCE OF DISEASE

Defining diseases, outcomes, and other health-related states or events

Before any epidemiological data collection can begin, meaningful health-related states or events must be defined, characterized and accepted by all those involved in the research. This condition may seem so obvious as to be trivial, but it is assuredly not. A number of factors can complicate this fundamental prerequisite of epidemiological research. In some instances the health-related state or event is poorly understood and the clinical picture is incomplete, making definitions difficult to derive. For example, the Metabolic Syndrome is a group of conditions (including obesity, high blood sugar, high blood pressure and high triglycerides) associated with cardiovascular disease and diabetes. Although there is consensus that these conditions often occur together and associate with disease (and, therefore, may be a valuable health-related state from a public health perspective), there is no consensus on the exact definition of the syndrome. Not only do major public health organizations such as the National Cholesterol Education Program, the World Health Organization, the International Diabetes Federation have slightly different definitions of Metabolic Syndrome, but – significantly – using these different definitions can lead to different estimates of the occurrence of the syndrome in populations (Cheung *et al.*, 2006; Olijhoek *et al.*, 2007).

In some instances epidemiologists must decide between using a continuous variable or a categorical variable based on the continuous measure. For example, blood pressure is frequently measured as systolic and diastolic pressure in nominal units of millimeters of mercury. For clinical purposes, however, these two continuous measures, in combination with treatment status, are often translated into a single categorical diagnosis such as 'prehypertension' or 'stage 2 hypertension' (Chobanian *et al.*, 2003). Which type of variable to use in an epidemiologic study hinges on such considerations as which is likely to offer the greatest statistical power in the context of the hypothesis being tested and study design being used, which is most clinically relevant, and whether the thresholds that delimit the categories are empirically justified and widely accepted.

Epidemiologists must also sometimes choose between *primary* or *surrogate* outcomes. Primary outcomes, as the name suggests, are direct measures of the actual health-related event of interest. Surrogate outcomes are used in place of primary outcomes. For example, a decrease in blood pressure over time is a surrogate outcome often used as a predictor of decreased occurrence of cardiovascular disease, a primary outcome. The choice to use a surrogate measure often results from logistical constraints. For example, an epidemiologist may hypothesize that a particular diet maintained throughout life will be associated with fewer occurrences of fatal cardiovascular disease in a population. Ideally, a study would be designed to follow a population from birth to death, with detailed surveillance of diet, disease and cause of death throughout. Such a study would be both complicated and costly. A more feasible alternative might be a shorter-term study using surrogate measures, such as blood lipid concentrations, which have well-documented associations with both diet and cardiovascular disease.

In summary, the following questions should be considered when defining health-related states or events and selecting outcome measures for epidemiologic studies:

- Is the definition/outcome suitable for the hypothesis being tested?
- How relevant is the definition/outcome to clinical and/or public health practice?
- Is the definition/outcome compatible with earlier studies, and is it acceptable to others who are conducting or planning to conduct similar studies?
- How feasible (both logistically and financially) is it to collect the data prescribed by the definition/outcome?

Calculating incidence and prevalence

Up to this point we have used the term *occurrence* to loosely refer to the burden of disease in a population. Epidemiologists have adopted a number of common measures to quantify how often a disease occurs in a population; which measure is used depends on the question being asked or the hypothesis being tested.

The term *incidence rate* refers to the number of new disease cases reported in a population over a certain period of time among those without the disease at first measurement. Thus, the incidence rate is calculated as shown in Equation 1:

$$\text{Incidence rate} = \frac{\text{Number of new disease episodes}}{\text{Total time at risk for subjects followed}} \quad (1)$$

It is important to note that the denominator is neither an average time at risk for a population nor does it correspond to the entire follow-up period (i.e., time of surveillance) of the study – it is the sum of all time units at risk for all members of the population. For example, Chenoweth and colleagues (2007) sought to determine the incidence of pneumonia in patients receiving mechanical ventilation at home. They followed 57 patients from June 1995 through December 2001. During that period, the 57 patients accrued a combined total of 50 762 days on their ventilators, and there were 79 episodes of pneumonia (see Equation 2).

$$\text{Pneumonia incidence rate} = \frac{79 \text{ cases of pneumonia}}{50\,762 \text{ ventilator-days}} = 0.00155 \text{ cases/ventilator-day} \quad (2)$$

Incidence rates are often converted to more easily interpreted units. For example, Chenoweth *et al.* reported 1.55 episodes per 1000 ventilator-days.

Incidence rate is a useful measure because it does not require a uniform follow-up period for all participants in a study; because time-at-risk is part of the equation, meaningful data can be produced even if participants begin the study late or drop out early.

Unlike incidence rate, which concerns itself with new cases of disease, *prevalence* deals with existing cases of disease. Simply put, prevalence is the proportion of people in a population who have the disease of interest a particular time. Thus, prevalence is calculated as shown in Equation 3.

$$\text{Prevalence} = \frac{\text{Number of subjects with disease at a particular time}}{\text{Total number of subjects in the population}} \quad (3)$$

For example, Bisgaard and Szefler (2007) sought to determine the prevalence of asthma-like symptoms in preschool children in the US and Europe. These researchers acquired data from 7251 households and found that 3077 of 9490 children met their criteria (see Equation 4).

$$\text{Asthma symptom prevalence} = \frac{3077 \text{ children with symptoms}}{9490 \text{ children in population}} = 32\% \quad (4)$$

Although prevalence is typically calculated using the number of subjects with the disease at a single time point (sometimes referred to as 'point prevalence'), 'period prevalence' is calculated using the number of subjects with the disease at a starting time point plus the number of new cases that develop over the study period. Period prevalence is a useful measure for intermittent conditions (such as lower back pain) because point prevalence tends to underestimate the number of individuals in a population with episodic but chronic disease. Because prevalence represents the abstract 'burden' of disease in a population at a particular time, the measure can be used to make quality-of-life assessments or calculate the financial costs of disease in a population.

Deciding which measure of disease occurrence to use depends on both the natural history of the disease and the type of epidemiological question being asked. Prevalence is influenced by both the incidence and duration of the disease in individuals. This fact makes prevalence an especially meaningful measure of occurrence for chronic diseases. For example, herpes simplex virus type 2 infections are typically non-fatal but also incurable. The incidence rate may be low, but because the infections are lifelong, seroprevalence of the virus in a population can be considerably higher. In contrast, easily transmitted diseases of short duration, such as influenza, can have a low point prevalence but a high incidence and, therefore, are often quantified using the latter. An epidemiologist tracking the secular trends of atherosclerosis in a population would probably be best served by making a series of prevalence measures, whereas an epidemiologist tracking an outbreak of food borne *E. coli* O157:H7 would find it most useful to measure incidence.

MEASURING RISK AND ASSOCIATION

Incidence and prevalence are core epidemiological parameters used to answer the fundamental question, 'How much disease exists in a population?'. These measures of disease occurrence cannot, however, by themselves answer another fundamental epidemiological question, 'What factors increase the chances of being affected by a disease?' To answer this question, epidemiologists turn to measures of *risk* and *association* to identify and quantify risk factors (see Chapter 4). Identifying risk factors is a crucial epidemiological task because these factors help create a clinical picture of disease and can be used in public health practice. Some risk factors may even play an etiological role in the development of a disease. The term *association* is used to describe a statistically established relationship between a risk factor and a disease. It is important to remember that *association does not imply causation*. Below we describe the general concept of risk and two important measures of association, relative risk and odds ratio.

Calculating risk

Like incidence rate described above, *risk* is a measure of new disease episodes. Unlike incidence rate, however, the new episodes are normalized with respect to the number of subjects followed rather than the time the subjects are at risk. Thus, risk is calculated as shown in Equation 5.

$$\text{Risk} = \frac{\text{Number of new disease episodes during time period}}{\text{Number of subjects followed for the time period}} \quad (5)$$

It is important to note two differences between risk and incidence rate. First, the concept of risk assumes all subjects are followed and 'at risk' for the same time period. Second, when reporting risk, it is always necessary to specify a time period to which the risk applies (although it need not be the actual time of follow-up, i.e., adjustments can be made to ease interpretation). The work of Semrad and colleagues (2007) offers an example of the risk of venous thromboembolism (VTE) in patients with malignant

gliomas. During a 2-year period, 9489 subjects were monitored for VTE and 715 cases were reported (see Equation 6).

$$\text{VTE risk} = \frac{715 \text{ cases of VTE}}{9489 \text{ subjects followed for 2 years}} = 7.5\% \text{ 2-year risk} \quad (6)$$

Risk is a useful measure because it can be interpreted as the probability of developing a disease. Simple risk estimates can be difficult to make, however, because subject death or other losses to follow-up effectively eliminate them (and their data) from the sample population. For these and other reasons, risk is often considered less useful as a measure of disease occurrence but more useful as a predictive tool. As we will see below, in situations where study designs allow it, risk is also used to compare groups with different exposures to risk factors.

Quantifying associations

Given the risk parameter described above, it follows that a simple means to quantify how strongly a suspected risk factor is associated with a disease is to compare the risk of a group of individuals exposed to the risk factor of interest with the risk of an unexposed group. Thus, relative risk (RR, also known as the *risk ratio*) is defined as shown in Equation 7.

$$\text{Relative risk} = \frac{\text{Risk of group exposed to factor}}{\text{Risk of group unexposed to factor}} \quad (7)$$

At the most fundamental level, relative risk is simple to interpret:

- If RR = 1, then the risk factor shows no association with the disease.
- If RR > 1, then the risk factor is positively associated with the disease.
- If RR <1, then the risk factor is negatively associated with the disease.

There are a number of disadvantages to using relative risk as a measure of association. For example, it is analytically tricky to adjust relative risk estimates for potentially confounding covariates (see the Threats to Validity and Reliability section below). Certain types of study designs (such as case-control studies) preclude the calculation of relative risk. Finally, because they are expressed as ratios and because analysts decide how the outcome of interest is defined and how the ratio is set up, the same study data can be used to generate very different-looking (but perfectly legitimate) relative risk estimates. For example, imagine a sample population is monitored for one year for getting at least one new tooth cavity (the outcome of interest). During that time, 791 of 850 individuals exposed to a new toothpaste additive (the factor of interest) had no new cavities (probability or 'risk' of no new cavities = 0.93 per year), and 860 of 1000 individuals not using the additive had no new cavities (risk = 0.86 per year). Using these figures, the relative risk of no new cavities for users of the new toothpaste additive is 1.08. Consider, however, if the event of interest was expressed as getting new cavities (as opposed to not getting them). In this case, the figures would be 59/850 for additive users and 140/1000 for non-users. This results in a relative risk of new cavities for additive users of 0.49. If we express it as the inverse (the relative risk of new cavities for people who do not use the additive), the relative risk is 2.02. All of these relative risk estimates are correctly calculated and all point toward the same conclusion – that using a toothpaste with the new additive is associated with a lower risk of getting at least one new cavity during a year's time. However, the magnitudes of the relative risk estimates obviously appear quite different. This intrinsically tricky property of relative risk demands that, although the *direction* of an association is intuitively obvious from the measure, the *magnitude* of relative risk must always be interpreted with care.

Whereas relative risk compares the *probability* of being affected by disease between exposed and unexposed groups, as its name implies, an odds ratio compares the relative *odds* of being affected by disease between exposed and unexposed groups. Recall that probability expresses the ratio of the number of events of interest to the total number of events; odds expresses the ratio of the occurrences of one event to the occurrences of its corresponding non-event. Although people's intuitive reasoning about the likelihood of events tends towards probability rather than odds, at a fundamental level, odds ratios are interpreted like relative risk – the direction and magnitude of an association is represented by direction (i.e., either greater than or less than 1) and degree that an odds ratio varies from 1 (see Equation 8).

$$\text{Odds ratio} = \frac{(\text{Probability of event in group 1})/(1 - \text{Probability of event in group 1})}{(\text{Probability of event in group 2})/(1 - \text{Probability of event in group 2})} \quad (8)$$

We can use the toothpaste additive-dental cavity example above to calculate an odds ratio. For toothpaste additive users, the odds were (new cavity)59:791 (no new cavity), or 0.075:1. For additive non-users, the odds were (new cavity)140:860 (no new cavity), or 0.163:1. Thus, there is a 2.17-fold greater odds (0.163/0.075) of getting at least one new cavity for people who don't use the toothpaste additive compared to those who do.

Using multivariable logistic regression models, odds ratios can be adjusted for covariates. (See the Threats to Validity and Reliability section below.) Unlike relative risk,

odds ratios can be calculated from case-control studies. For relatively rare outcomes, odds ratios approximate relative risk because the denominator in the individual odds estimates approximates the total number of events. Readers who are interested in learning more about the calculation of relative risk and odds ratio and their interpretation should consult Davies and colleagues (1998) and Naylor and colleagues (1992) and the commentaries associated with these articles.

TYPES OF EPIDEMIOLOGICAL STUDIES

The careful design and implementation of public health surveillance programs underlie all descriptive epidemiological studies. Because this volume concerns itself with clinical and translational science, however, we will forbear further treatment of this topic and instead direct interested readers to the books edited by Teutsch and Churchill and Brookmeyer and Stroup (see Recommended Resources section below).

Analytical epidemiological studies are often classified as either *experimental* or *non-experimental,* where non-experimental studies are those in which researchers make observations without manipulating independent variables. In practice, the range of complex and compound study designs available to researchers makes this distinction somewhat theoretical. For purposes of discussion here, however, we will consider randomized and non-randomized clinical trials to be experimental and refer readers to the clinical trials section of Chapter 2 and the detailed discussion of experimental design in Chapter 5. After eliminating the design of surveillance programs and clinical trials, we are left to describe three classic epidemiological study designs: cross-sectional, cohort and case-control.

Cross-sectional studies

A cross-sectional study observes a sample population at a nominal single point in time. Although the cross-sectional design is often used for descriptive prevalence studies, this design is also used to investigate associations between risk factors and diseases. Cross-sectional studies allow investigators to collect data on multiple diseases and risk factors simultaneously, and they permit considerable analytical freedom to define and compare subgroups within the sample population. Cross-sectional studies often offer a relatively inexpensive way to collect a lot of epidemiological data in a short period of time.

The flexibility and efficiency of cross-sectional studies do, however, come at a cost. Of the various designs available to researchers, cross-sectional studies provide the least robust evidence that a risk factor plays a causal role in disease etiology – hence the use of the word 'association' to cautiously describe the relationship between a risk factor and a disease. Because cross-sectional studies provide a single snapshot in time of disease and risk factor status, it is impossible to determine if exposure to a risk factor occurred before, during, or after disease emergence. Cross-sectional studies are also susceptible to prevalence-incidence bias (see the Threats to Validity and Reliability section below), which can cause the association between potentially significant risk factors and a disease to be underestimated. Despite these epistemological and analytical shortcomings, cross-sectional studies provide an expedient means to generate hypotheses which can be subsequently tested in studies of other design.

The recent study by Ha and colleagues (2007) provides an excellent example of a cross-sectional investigation. Using data from the US National Health and Nutrition Examination Survey, these investigators found statistically significant associations between serum concentrations of persistent organic pollutants (POPs) and self-reported history of cardiovascular disease in women. This research illustrates characteristics of an appropriate, well designed and carefully reported cross-sectional study: the study investigated a recently reported risk factor-disease association about which little is known (Mastin, 2005); the study used a robust (n = 889), population-based sample; 21 different risk factors (POPs) were simultaneously investigated; self-reported affectation data (although suboptimal) was probably relatively quick, easy, and inexpensive to collect; and the authors organized and analyzed their risk data to allow comparisons between POPs and between participant demographic (i.e., sex, age) strata. Finally, these researchers concluded their report by calling for cautious interpretation of their cross-sectional data and urged more research on the reported association.

Cohort studies

Cohort studies are so named because they identify and follow a group of individuals through time (i.e., *longitudinally*). Individuals in the cohort are bound by a common characteristic, such as having been born in the same year, living in the same town, or having been exposed to the same risk factor. A defining characteristic of cohort studies is that the health-related event (or events) of interest is not present in the cohort at the start of follow-up. It would be somewhat misleading to say that a cohort is without disease at the beginning of a study, because sometimes the presence of a disease is, in fact, the characteristic that defines the cohort. However, in such cases, the disease itself is not the health-related state or event of interest. For example, a cohort of HIV-infected individuals may be followed to learn about specific events in the natural history of AIDS. Cohort studies can be open (i.e., the pool of participants changes during the study) or closed (i.e., the cohort remains static during the study); prospective (i.e., the cohort is assembled before outcome measures are made) or

retrospective (i.e., the cohort is defined after outcome data have been collected).

Although cohort studies are observational, they are also analytical because statistical comparisons can be made between the cohort and the general population from which the cohort was drawn or among subgroups with different risk exposures or other characteristics of interest within the cohort. Cohort studies are useful for describing the incidence and natural history of a disease and for finding associations between purported risk factors and health outcomes. However, it is important to remember that, like cross-sectional studies, cohort studies cannot demonstrate that a risk factor positively associated with a disease is a factor that causes the disease. It is arguable, however, that because the longitudinal nature of cohort studies allows investigators to determine whether exposure to a risk factor occurred before, during, or after disease emergence, the cohort design provides slightly stronger evidence for (or against) causation than the cross-sectional design. Like cross-sectional studies, cohort studies can simultaneously track multiple health-related events in the same sample. Finally, cohort studies are useful when experimental studies (such as randomized clinical trials) are practically or ethically impossible. For example, it would be unethical to intentionally expose a clinical trial group to a suspected carcinogen (e.g., tobacco smoke, asbestos), but it would not be unethical to compare cancer incidence between cohorts of smokers or asbestos minors and the general population.

Cohort studies are useful only if the group can be followed for a length of time sufficient to observe a statistically adequate number of outcomes. For this reason, using a cohort design to study rare events (e.g., certain kinds of cancer) is often not practical. Of course, the problem associated with studying a rare event can be offset by following a larger cohort and/or following the cohort for a longer period of time. These strategies, however, can make cohort studies prohibitively expensive. Cohort studies are susceptible to a number of biases, including differential confounding with risk, differential loss to follow-up, and the potential for zero-time bias. (These are discussed in the Threats to Validity and Reliability section below.)

Relative risk is often used to quantify hypothesized associations in cohort studies. The data used to calculate relative risk can be conveniently laid out in a 2 × 2 matrix. Fig. 35.1 contains the data from the toothpaste example above. Tests such as chi-square or Fisher's exact can be used to determine whether an association reaches the level of statistical significance.

The Framingham Study mentioned above is an example of a cohort study that has contributed significantly to our understanding of cardiovascular disease (Dawber *et al.*, 1951). The Nurses' Health Study was begun in 1976 with a cohort of about 120 000 women and has since provided data for more than 600 published articles concerning oral contraceptive and hormone use, smoking, menopausal

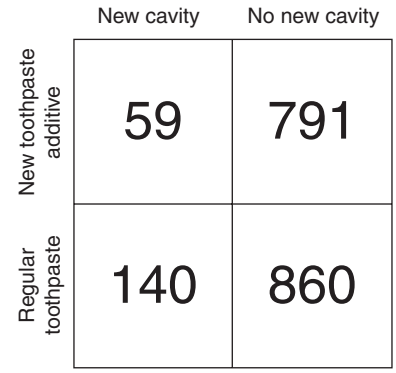

FIGURE 35.1 2 × 2 matrix showing results for a cohort study sorted by exposure and outcome. A color version of this figure is available on the *Clinical and Translational Science* companion website which can be accessed at www.elsevierdirect.com/companions/9780123736390

status and other health-related events (Belanger *et al.*, 1978). Examples of other notable cohort studies include the British Doctors Study (which followed 34 439 participants and, among other findings, provided convincing data which linked smoking to lung cancer) (Doll and Hill, 1956); the Bogalusa Heart Study (which followed 11 796 participants to investigate cardiovascular disease in children and young adults) (Berenson *et al.*, 2005); and the National Children's Study (which plans to follow 100 000 US children from birth to age 21, tracking a wide range of environmental factors and health markers) (Branum *et al.*, 2003).

Case-control studies

In contrast to cohort studies – which select participants on the basis of *not having experienced* the state or event of interest – case-control studies begin by selecting a group of participants *who have experienced* the state or event (*cases*) and a group of participants who have not (*controls*). By retrospectively comparing the proportion of cases with a potential risk factor to the proportion of controls with the same risk factor, these studies seek to identify variables that predict the outcome of interest. The predictive value of these factors is often quantified using odds ratios.

Because the case-control study begins with a sample of affected individuals, the design is particularly well suited for the study of rare diseases; it is not susceptible to the same problems related to statistical power that often afflict cohort studies (i.e., too few observed cases to make meaningful conclusions). In fact, for some diseases, a case-control study may offer the only feasible study design. Although each case-control study can concern itself with only one disease, many potential risk factors can be assessed with respect to that disease. The case-control design is often used for relatively quick, inexpensive studies intended to generate hypotheses about the significance of risk factors. Subsequent studies of different designs must be used to validate a risk factor or seek evidence to suggest the factor plays a causal role.

Ideally, cases should be randomly selected from the pool of all individuals in a population who have experienced the outcome. For many reasons, however, this is often impossible – including the frequently inescapable fact that not all affected individuals in a population may have been diagnosed at the time of case selection. Locating and selecting an appropriate control group can be even more difficult. For example, recruiting a 'convenience' sample may result in a control group that systematically differs – in ways unknown and unmeasured – from the case group or that is otherwise not representative of the general population to which the cases should be compared. In an effort to minimize unanticipated differences between groups, controls can be individually matched to cases on the basis of basic demographic variables and potential confounders such as sex and age. This, however, can result in 'overmatching' when cases and controls are matched on the basis of non-confounding variables (i.e., those associated with an exposure but not the disease), potentially resulting in an underestimation of the predictive value of an important risk factor. Other case-control design measures, such as randomly drawing controls from a population-based sample or using multiple control groups, can be used to minimize some of these potential biases.

Reports of case-control studies in the medical literature are ubiquitous. (A search of NCBI PubMed database returned over 100 000 instances.) Despite the fact that many of these case-control studies undoubtedly provided valuable (albeit less than definitive) data associating risk factors with disease, it is often the controversial – and confuted – studies that are remembered. For example, from the late 1970s to the middle 1990s, many studies, including case-control studies, reported cardioprotective effects of hormone replacement therapy in postmenopausal women (Stampfer and Colditz, 1991). However, by the late 1990s and early 2000s, evidence – especially from randomized control trials – began to emerge suggesting hormone replacement had no cardioprotective benefits and, in fact, may have deleterious cardiovascular associations (Hulley *et al.*, 1998; Herrington *et al.*, 2000; Waters *et al.*, 2002). How could this have happened? It is now generally accepted that a primary confounder of this association was socio-economic status (see the Threats to Validity and Reliability section below). That is, in the groups investigated, women undergoing hormone-replacement therapy were generally of higher socio-economic status – and a higher socio-economic status correlates with lower cardiovascular risk. Although some early studies statistically adjusted their analyses to account for socio-economic status, these adjustments were inadequate, and the confounding by socio-economic status long went undetected and unmitigated (Lawlor *et al.*, 2004). Rather than bemoan the fact that epidemiology has suffered such discomfitures, some have suggested these incidents be remembered as 'classics' in the history of the science (Smith, 2004). Certainly such controverted studies should serve to illustrate the limits of the case-control design and forewarn against overly ambitious interpretation.

Hybrid study designs

There are a number of study designs that incorporate elements of both cohort and case-control studies. These hybrid designs are used because they often offer both logistical and analytical advantages over any of the more basic designs.

Nested case-control studies are case-control studies built retrospectively upon cohort studies (Ernster, 1994). Nested case-control studies are typically used as a cost-efficient design to test a novel biomarker or other data collection instrument that is prohibitively expensive to administer in the cohort, but is feasible for testing in a smaller subset. After data are collected prospectively from the cohort, investigators retrospectively select samples of cases and controls and use the previously collected data and any newly acquired data collected from cases and controls as they would in the standard case-control design. As in a standard case-control study, cases and controls are usually matched on the basis of basic demographic parameters and potential confounders. The nested case-control design ameliorates the problem of recall bias which can be an issue in standard case-control studies. Recall bias is caused by systematic differences between cases and controls in the accuracy of their memory of affectation status, exposures, or other data collected by survey. Because cohort studies effectively detect cases as they occur, participant recall errors are eliminated. Including cases and controls from the same cohort can also reduce selection bias. However, the pool of cases selected retrospectively may not be fully representative of the original cohort due to death or losses to follow-up.

The *case-cohort* design is similar to a nested case-control study in that subsamples of a cohort are analyzed (Prentice, 1986). In a case-cohort study, the cases include all individuals in the cohort who meet the case criteria. This group is compared to a sub-cohort that is randomly drawn from the full cohort but does not include any cases. Unlike case-control studies, case-cohort studies allow direct estimation of prevalence and risk ratios, and different sub-cohorts can be drawn from the same cohort to be used as a control group for different disease outcomes. The case-cohort design can be more cost effective than cohort studies because only a subsample of the non-cases (i.e., the sub-cohort) needs to be assessed for risk exposure.

Choosing between the nested case-control and the case-cohort design depends on both logistical and analytical considerations. For example, a nested case-control study might be a better choice simply because sensitive biological specimens stored in batches need to be handled only

once for analysis after all cases and controls are selected. In contrast, if case specimens are analyzed as they emerge in a case-cohort study, specimen batches may need to be handled repeatedly (e.g., thawed and refrozen) which could cause deterioration. Conversely, a case-cohort study might be the better choice because it is not restricted to the same analytical constraints as case-control-type studies. Interested readers are referred to Wacholder (1991) and Langholz and Thomas (1990) for more detailed comparisons of these study designs.

Study design: summary

Even a cursory review of the literature will reveal that the study designs described above do not fully capture the complexity and sophistication of contemporary epidemiological research. For example, there are prospective case-control studies, retrospective case-cohort studies, *ecological studies* where group (rather than individual) data are used as the basis for analysis and comparison, and serial cross-sectional studies which comprise *pseudo-longitudinal* studies. Descriptions of all of these are beyond the scope of this chapter. Keeping in mind the fundamental characteristics of basic study designs, however, can provide a context in which designs of greater complexity can be understood. To this end, we provide the following summary:

- Cross-sectional studies compare – at a single point in time – disease prevalence rates between individuals with different exposures to risk factors. A higher prevalence of disease among individuals with greater risk exposure provides evidence of association between risk factor and disease.
- Cohort studies compare – over a period of time – disease incidence rates between individuals with different exposures to risk factors. A higher incidence of disease among individuals with greater risk exposure provides evidence of association between risk factor and disease.
- Case-control studies compare exposures to risk factors between diseased (cases) and non-diseased (controls) individuals. A higher level of risk exposure in cases provides evidence of association between risk factor and disease.

THREATS TO VALIDITY AND RELIABILITY

The many occasions within this chapter when we have pointed to this section serve to emphasize an important fact: epidemiologists must remain aware of, guard against and compensate for forces and factors that might make their findings invalid and unreliable. Although an exhaustive treatment of these forces and factors is outside the goals of this chapter (Delgado-Rodriguez and Llorca [2004] cataloged over 70 types of bias alone), a brief discussion of the major threats to validity and reliability are in order.

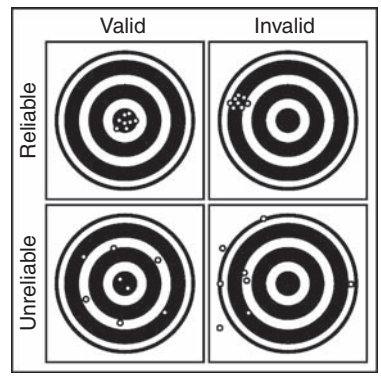

FIGURE 35.2 Relationship between validity and reliability. A color version of this figure is available on the *Clinical and Translational Science* companion website which can be accessed at www.elsevierdirect.com/companions/9780123736390

Defining and measuring threats to validity and reliability

A measurement (or study) is *valid if it measures what it sets out to measure*. A measurement (study) is *reliable if it produces the same result upon replication*. The relationship between these two concepts is best illustrated with the bull's-eye analogy. If we imagine that the center of a bull's-eye represents the correct answer to a research question and the shots taken at the target represent repeated attempts to learn the correct answer, Fig. 35.2 illustrates the four possible combinations of valid–invalid and reliable–unreliable findings. It is important to remember that a reliable measurement does not guarantee a valid measurement, nor does an unreliable measurement imply an invalid measurement.

A study is said to have *internal validity* if it accurately measures what it intends to measure *within the study's sample*. A study is said to have *external validity* if its findings are generalizable to a wider population. Validity is compromised by bias – any factor that systematically distorts our apprehension of the truth and, therefore, results in incorrect measurements or conclusions. As noted above, there are many types of bias and these have been classified by a number of different taxonomies. Below we outline and offer specific examples of a few of the most important types of bias.

Selection bias

Selection bias occurs when groups being compared in an analysis differ systematically in ways unknown or unintended. For example, in a cohort study, the exposed and unexposed groups may differ in ways other than their exposure to the risk factor under study (e.g., smokers might drink more sugared soft drinks per day than non-smokers). A common form of selection bias is *ascertainment bias*, in which, for example, a cohort sample is chosen non-randomly from a population (thus compromising external validity). A form of ascertainment bias known as *prevalence-incidence bias* (or Neyman bias) can occur when there is

a gap between exposure to the risk factor of interest and recruitment of participants. In a cross-sectional study, for example, prevalent cases include only those who did not immediately die from the disease and, therefore, may possess better-than-average survival characteristics. Incident cases, on the other hand, would represent the full spectrum of diseased individuals. Similarly, *spectrum bias* results when only clear or obvious instances of a diseased individuals are recruited as cases and, therefore, do not represent the full spectrum of actually diseased individuals in the population. In prospective cohort studies, *zero-time bias* results when participants are recruited in a manner resulting in unintended systematic differences between groups at the beginning of the study. *Differential loss to follow-up* bias occurs when participants are lost or removed at different rates from exposure or outcome groups.

Information bias

Information bias results from systematically incorrect observation, classification, or measurement of risk exposure and/or disease outcome. *Recall bias*, described above, can be a problem in case-control studies because cases tend to spend more time and effort searching their memory about possible causes of their disease (i.e., exposures), but controls have no such incentive. *Observer bias* occurs when data collectors have knowledge of a study's hypothesis or participants' exposure or affectation status. For example, a researcher making diagnoses for a study may inadvertently perform more thorough clinical exams on individuals she knows have had the exposure of interest. In studies where participants supply information via surveys or interviews, *reporting bias* may happen if participants alter their responses because they know (or believe they know) what the researchers want to hear.

Confounding

In short, confounding is a type of bias caused by a misattribution of cause. That is, confounding occurs when an association between a presumed risk factor and the outcome of interest is, in fact, accounted for by another factor. Confounding results not from erroneous measurement, but from incorrect interpretation of what may be an accurate measurement. A variable can be a confounder only if it is distributed differently between compared groups. For example, early studies reported associations between oral contraceptive use and increased risk of myocardial infarction. Subsequent studies, however, determined that a large proportion of oral contraceptive users were also cigarette smokers. Cigarette smoking was, in fact, the actual factor responsible for increasing the likelihood of heart attack (Burchard *et al.*, 2003). Note that for a factor to confound an association between a purported risk factor and a disease, the confounding factor must be associated both with the risk factor and the disease, and its relation to the disease must be independent of its association with the risk factor (i.e., it should not interact with the risk factor to have its effect on the disease; see below). Confounders may bias interpretation of an association either positively or negatively. Positive confounders result in an overestimated association while negative confounders lead to an underestimation.

Interaction

Unlike confounding, in which an association between a risk factor A and an outcome is merely a reflection of the association between risk factor B with risk factor A and the outcome, interaction results when the strength of the association between risk factor A and an outcome is influenced by risk factor B. In statistical terms, an interaction results when the effect of two (or more) variables on an outcome is not simply additive. Generally, when we speak of variables having an interactive effect on an outcome, we are not concerned with a threat to validity – interaction is 'real' within our dataset, not the result of mistaken interpretation of associations. However, if a real interaction between variables is not detected or is handled improperly during analysis, the true magnitude of certain associations may be obscured. Also, just as causal variables can interact, so can two or more threats to validity. For example, selection-maturation interaction occurs when comparison groups that entered a study with similar disease indicators are different in their rates of progressive worsening of the disease. This interaction of time $\times$ personal characteristic (i.e., disease maturation) may lead to a specious treatment effect.

Estimating and avoiding threats to validity

Some forms of bias can be averted with proper study design while other forms must be addressed during data analysis.

Selection bias can be avoided by careful recruitment of study participants. For example, in choosing a control group for a case-control study, the goal is to capture controls from the same population that gave rise to the cases, and, indeed, who would have been enrolled in the study as cases if they had developed the disease in question. For instance, a case-control study might be designed to test the association between coffee consumption and breast cancer. If cases are identified during routine mammography from a particular women's health center, an appropriate control group would be women seen for routine mammography from the same health center who do not have breast cancer. If a control group were recruited from the general population, selection bias might be introduced since women who get routine mammograms may be different in many ways related to the exposure and disease from those women who do not.

Careful study design is also necessary to reduce information bias. For example, if a case-control study seeks to determine whether exposure to a particular environmental factor is associated with lung cancer, a carefully chosen control group may be necessary to avoid recall bias. A control group that suffers from benign lung disease may be deemed appropriate since they would be expected to recall past exposures more likely than a healthy control group. A potential problem arises when choosing a control group with a disease other than the one being studied, however, if the exposure in question is associated with both diseases. To reduce observer bias, data collectors are often blinded to the disease/exposure status of the study participants. Thoughtful data collection instruments are necessary to avoid reporting bias. For example, if the information to be collected is deemed sensitive, such as information on sexual practices or illegal drug use, participants might be more honest if they are allowed to write their responses on paper rather than provide them in a face-to-face interview. Making the data collection process anonymous might also be necessary to get unbiased answers from participants.

Confounding is a major threat to validity, and potential confounders may be known and measured (either accurately or with error), known and not measured, or unknown. It is possible to minimize confounding during both the design phase and the statistical analysis phase of a study. For example, by design, randomized studies attempt to reduce confounding by distributing confounding factors (both known and unknown) equally between the groups to be compared. Matching is another way to reduce confounding during the design phase, and it also allows efficient management of confounding during the analysis phase. For example, if a researcher suspects that gender will confound the association of interest, comparison groups may be constructed so they are matched on gender, with an equal ratio of women to men in each group. The ratio need not be 50% to achieve a matched design, but the ratio must be the same in each comparison group. It is also possible to eliminate confounding by a particular factor by restricting the study population. For example, if smoking is known to be a confounder in the association of interest, the study population can be limited to non-smokers. During the analysis phase of a study, researchers can control for confounding factors by either computing standardized rates and risks of outcomes, conducting matched analysis in studies with a matched design, using stratified analyses (i.e., computing unbiased stratum-specific measures of association), or adjusting for confounding in a multivariable statistical model.

The nature of interactions must be uncovered in order to estimate the 'true' association for each level of the interacting variable. For example, if the true effect of a particular antihypertensive drug is beneficial to men but has no effect on women and the analysis is conducted without regard to gender, the analysis will return the 'average' effect, which is accurate for neither men nor women. If we statistically adjust for gender, we will get the 'weighted average' effect, which, again, is accurate for neither men nor women. To assess an interaction, one may conduct a stratified analysis to uncover the true effect for each group individually, and to determine whether the measure of association (drug effect on blood pressure, in our example) differs across the levels of the third variable (gender, in our example). Using a multivariable statistical model, the interaction effect may be tested for statistical significance.

Estimating and avoiding threats to reliability

Reliability is threatened whenever a system of measure is unable to produce consistent results when it should. For example, a large epidemiological study of blood pressure may enlist many practitioners in many clinics as data collectors. Even with standardized protocols, differences in equipment and personnel may lead to unreliable measurements.

Several methods are used in epidemiologic research in an attempt to achieve reliable results. A carefully prepared manual of study procedures is necessary for consistency. The manual should provide a detailed description of items such as participant eligibility requirements, variable definitions, clinic visit forms and precise laboratory methods. Training sessions for data collection staff are necessary to provide instruction on standardized protocols, making sure all staff are aware of and able to perform the techniques uniformly. There are ways to measure inter- and intra-rater reliability, which is the degree to which different personnel agree in their measurement of a particular variable, and the degree to which an individual will get the same value upon measuring a variable twice (test–retest reliability), respectively. If several different personnel are collecting data, verification that consistent procedures are used is necessary. Laboratories can be checked for reliability by submitting blind replicates of samples and evaluating the concordance of the lab findings.

MOVING FROM ASSOCIATION TO CAUSATION

Throughout this chapter we have avoided 'knotty epistemological issues' related to causation, warned against 'overly ambitious' interpretation of epidemiological studies, and have assiduously applied the word 'association' to describe the relationship between risk factor and disease. But what about *cause*? Although it should be clear from this chapter that the findings of epidemiologists are put to many uses other than discovering the causes of disease, discovering the causes of disease is a vitally important element of public health practice, clinical medicine and epidemiology. How, then, do we move from association to causation?

This is a difficult question to answer for philosophical, semantic and practical reasons. For our purposes, we can set aside the weightier philosophical issues (e.g., what, if any, evidentiary standards are sufficient to prove something is true and knowable?) – not because these are uninteresting questions, but because our goals are ultimately pragmatic, and we can at least come to some consensus on what constitutes a sufficient process to establish cause. The semantic issue is best illustrated by example: Do changes in the machinery that drives cell growth and death cause lung cancer? Do carcinogens in cigarette smoke cause lung cancer? Do cigarettes cause lung cancer? Does peer pressure and tobacco product advertising cause lung cancer? This simple web of component causes illustrates that the definition of *cause* in any particular case may not be easy to nail down. Ultimately, this issue can be addressed by not discussing causal relationships in simplistic or abbreviated terms.

We are left to discuss the practical demands of providing a sufficient process to establish a cause and effect relationship. Epidemiologists have typically approached this issue in two related ways. First, as we suggested in the study design section above, some designs provide stronger evidence for causation than others. The list below ranks various designs from strongest to weakest with respect to their ability to provide evidence of causation:

1. Experimental studies:
   A. Randomized controlled clinical trial
   B. Randomized cross-over clinical trial
   C. Randomized controlled laboratory study
2. Observational studies:
   A. Cohort study
   B. Case-control study
   C. Cross-sectional study
   D. Case series
   E. Case report

This list ranges from little more than anecdotal observation (a case report) to what is considered the 'gold standard' of research designs – the randomized controlled clinical trial. The differences between the end members of this continuum should be obvious. A case report might offer an observation of a single risk factor-disease association with potentially no information on the timing or extent of exposure or sources of confounding. Randomized controlled clinical trials generate data from a statistically sufficient sample population in which the degree of exposure to a risk factor is controlled by researchers, and confounding is avoided by randomly assigning participants to exposure groups. Differences in strength between other designs are less profound but are a function of similar considerations. This spectrum of study designs does not provide a definitive procedure for moving from association to causation, but it approximates the real-world progression from hypothesis-generating studies to studies that offer increasingly stronger evidence to support (or reject) a hypothesis of causation.

The other manner in which cause and effect is discussed in epidemiology is epitomized by the work of Sir Austin Bradford Hill. In his seminal 1965 article, Hill outlined nine different dimensions along which a purported causal relationship might be assessed. Some of Hill's dimensions will be familiar as they are the same criteria used to evaluate study designs and generate the spectrum of designs listed above.

- **Strength:** Is the magnitude of association sufficiently large that other risk factors can be ruled out?
- **Consistency:** Has the association been repeatedly observed by different researchers using different experimental designs and analytical methods?
- **Specificity:** Is the association valid only for a specific population group, specific risk factor(s), and specific disease?
- **Temporality:** Does exposure to the risk factor precede the disease?
- **Biological gradient:** Does the association between risk factor and disease exhibit a dose–response relationship?
- **Plausibility:** Is there a credible biological mechanism that might explain the association?
- **Coherence:** Is the association consistent with the known natural history biology of the disease?
- **Experimental evidence:** Does experimentally manipulating the risk factor change the nature of the association?
- **Analogy:** Is there an established cause–effect relationship similar to the association?

Hill is quick to point out that none of these nine 'viewpoints' (his own carefully chosen term) can provide indisputable evidence of a causal relationship. In fact, finding counter-examples in the literature to 'debunk' Hill's criteria is a trivial exercise – and, given philosophical precedents, ultimately a needless one. Although it is true that a 'checklist' approach to establishing causality in epidemiology is problematic (Lanes and Poole, 1984), it is also true that Hill's work (and similar efforts, e.g., Henle-Koch postulates [Evans, 1976]) ought not be considered a method but a heuristic.

Clearly, neither the study design spectrum nor Hill's viewpoints offer a peremptory path from association to causation. Given the complex, multicausal nature of disease and the apparent theoretical limits of science to 'prove' causation, moving from association to causation becomes as much a social as a scientific process. In addition to conducting more research and gathering more and better evidence, epidemiologists and their colleagues must discuss and debate purported causal relationships and their translational significance; Hill's viewpoints provide useful sites for such interrogation and deliberation. Scientists who feel uncomfortable with this seemingly unscientific

manner of reaching consensus would do well to remember Hill's words that 'all scientific work is incomplete' and 'all scientific work is liable to be upset or modified by advancing knowledge'. Those who seek to protect, promote and restore health and well-being, Hill wrote, cannot 'postpone the action[s]' dictated by current knowledge. Ultimately, any discussion of causation points to a quandary faced in one form or another by all those who seek to translate scientific knowledge into practice, including epidemiologists: the overly cautious epidemiologist who eschews public health action because causation has not been definitively proven commits as grave an error as the overly ambitious epidemiologist who initiates public health action predicated on a specious understanding of causation.

Readers interested in learning more about the theoretical and practical aspects of determining cause in epidemiology are referred to the work of Karhausen (2000) and Parascandola and Weed (2001).

CLINICAL EPIDEMIOLOGY

Clinical epidemiology, as its name implies, represents the marriage of clinical medicine and traditional epidemiology. Clinical epidemiology is 'clinical' in two senses – the research is usually conducted in a clinical setting using patients as study subjects, and the findings are typically used by clinicians to make better decisions about patient care. Clinical epidemiology is a major constituent of what has become known as 'evidence-based medicine,' that is, using the scientific method and its fruits to inform clinical practice. To younger clinicians, this may seem manifest. However, in addition to being a discipline driven by strong traditions, clinical medicine has historically been built upon extrapolations from the basic sciences of physics, chemistry and biology as well as anatomy and physiology. Clinical epidemiology is often used to test whether the extrapolations from relatively simple model systems to the spectacularly complex systems of humans in their environments are valid.

All of the fundamental epidemiological principles introduced in this chapter apply to clinical epidemiology. Differences between clinical and more traditional epidemiology are largely driven by purpose and context and are best illustrated by example. A traditional epidemiological study of heart disease may follow a cohort of individuals to collect outcome and potential risk factor data. A successful study will find associations between risk factors (e.g., smoking, diet, exercise) and the occurrence of a disease or event (e.g., coronary heart disease, myocardial infarction). Clinical epidemiology, in a sense, picks up where traditional, public health epidemiology leaves off by assessing the consequences of disease and the efficacy of possible interventions. A clinical study may follow a group of patients with coronary heart disease and assess whether a novel treatment provides more relief from the disease than a traditional treatment. It should be obvious from this description that the work of clinical epidemiologists usually falls into the realm of experimental epidemiology. As a result, clinical epidemiologists have at their disposal randomized controlled and cross-over study designs, tools exquisitely suited to advancing both clinical knowledge and practice. Much of the remainder of this volume is devoted to topics central to or strongly allied with clinical epidemiology (see Chapters 2, 20, 27, 34 and 36).

SEX, GENDER, RACE AND ETHNICITY IN EPIDEMIOLOGY

Epidemiologists often collect data on study participants' sex, race and ethnicity. The motives for doing this are pragmatic: the constructs of sex, race and ethnicity may be associated with (or be confounders of) outcomes of interest. Unfortunately, sex, race and ethnicity have also been (and still are) used as instruments of subordination, subjugation and discrimination. Many people are justifiably concerned that the very use of such categories in medical research potentially perpetuates harmful discrimination. Below we define these often misused terms, discuss the grounds for their contentiousness and offer a reasoned perspective on their role in epidemiological research.

Within the medical professions, 'sex' is defined by fairly straightforward biological criteria: females have two X chromosomes; males have one X and one Y chromosome. In humans, there are forms of intersexuality with well understood biological bases; however, these are relatively rare (Sax, 2002). 'Gender' refers to notions of what is feminine and what is masculine. Whereas the definition of sex is generally uniform across contemporary cultures, the traits and behaviors associated with gender vary from culture to culture. Although the purpose of some studies may warrant the collection of gender data, epidemiologists typically collect data on sex (albeit self-reported); care should be taken to use the correct term. A 'race' is defined as a category of people who share characteristics that are presumed to be biologically inherited. 'Ethnicity' describes the state of belonging to a social group defined by a shared cultural heritage. The US National Institutes for Health currently endorses the use of five racial categories (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White) and two ethnic categories (Hispanic or Latino, not Hispanic or Latino). It is important to note that both race and ethnicity are social constructs – people, after all, must decide which biological traits factor into definitions of race and which cultural criteria delimit ethnicity.

The use of these categories is contentious for a number of reasons. For example, some fear that assigning individuals to these categories places study participants at risk for

future discrimination (either inadvertent or intentional). Some believe that maps of the human genome and the advent of genetic epidemiology make categorical distinctions such as race obsolete. We now understand genetic variation to be a continuous phenomenon that cannot demarcate discrete population groups, and genetic variation between traditional racial groups is smaller than within groups (Burchard *et al.*, 2003). Ultimately the genetic characterization of differences or similarities between populations may provide significantly more useful epidemiologic data than crude phenotypic indicators such as skin color. However, others believe traditional racial categories ought to continue to play a role in medical research. Self-identified race is often associated with attitudes toward health and use of health services, and significant disparities in healthcare often occur among self-identified racial categories (Flores *et al.*, 2001). To ignore traditional racial categories is to ignore real-world determinants of public health.

Given the concerns associated with collecting sex, race and ethnicity data and the dubious veracity of some of these categorical distinctions in light of existing scientific evidence, what role do these concepts have in contemporary epidemiology? Ultimately, these categories have proven to be useful markers of risk in epidemiological studies, and their utility – even in the age of genomics – is unlikely to be surpassed in many circumstances (see Osborne and Feit, 1992 for the important distinction between 'risk marker' and 'risk factor' in this context). Ignoring these socially constructed categories may hamper discoveries that could reduce health disparities between groups. When making these categorical distinctions, however, epidemiologists must recognize they are simplifying complex and dynamic interactions among biological, social, geographic and cultural forces. We offer the following recommendations (based on the race/ethnicity work of Kaplan and Bennett [2003], who offer a more thorough discussion) for those using these socially charged categories in epidemiological research:

- When race or ethnicity data are presented, category definitions should be described and these definitions should be justified.
- When sex, race, or ethnicity data are presented, their use should be justified.
- When sex, race, or ethnicity data are presented, all potentially relevant confounding variables (such as socio-economic status) should be considered in the analyses.

SUMMARY

Epidemiology is a critically important translational science that serves both public health and clinical medicine. By employing the tools of basic science, epidemiology – whether deployed at the population level or in the clinic – seeks to answer fundamental questions related to the 'where,' 'when,' and 'who' dimensions of health-related states or events. Armed with such data, community health specialists can design and implement interventions to protect, promote and restore health and well-being in populations and clinicians can make better diagnoses and treatment decisions for their patients.

ACKNOWLEDGMENTS

The authors thank Amy I. Lynch, PhD for her valuable contributions to this chapter.

RECOMMENDED RESOURCES

Armitage, P., Berry, G. and Matthews, J.N.S. (2002) *Statistical Methods in Medical Research*. Malden, MA: Blackwell Science.

Aschengrau, A. and Seage, G.R. IIIrd (2008) *Essentials of Epidemiology in Public Health*. Sudbury, MA: Jones and Barlett.

Brookmeyer, R. and Stroup, D.F. (eds) (2004) *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*. New York: Oxford University Press.

Teutsch, S.M. and Churchill, R.E. (eds) (2000) *Principles and Practice of Public Health Surveillance*. New York: Oxford University Press.

REFERENCES

Belanger, C.F., Hennekens, C.H., Rosner, B. and Speizer, F.E. (1978) The nurses' health study. *Am. J. Nurs.* 78, 1039–1040.

Berenson, G.S. and Srnivasan, S.R. Bogalusa Heart Study Group (2005) Cardiovascular risk factors in youth with implications for aging: the Bogalusa Heart Study. *Neurobiol. Aging* 26, 303–307.

Bisgaard, H. and Szefler, S. (2007) Prevalence of asthma-like symptoms in young children. *Pediatr. Pulmonol.* 42, 723–728.

Branum, A.M., Collman, G.W., Correa, A., Keim, S.A., Kessel, W., Kimmel, C.A., Klebanoff, M.A., Longnecker, M.P., Mendola, P., Rigas, M. *et al.* (2003) The National Children's Study of environmental effects on child health and development. *Environ. Health Perspect.* 111, 642–646.

Burchard, E.G., Ziv, E., Coyle, N., Gomez, S.L., Tang, H., Karter, A.J., Mountain, J.L., Perez-Stable, E.J., Sheppard, D. and Risch, N. (2003) The importance of race and ethnic background in biomedical research and clinical practice. *N. Engl. J. Med.* 348, 1170–1175.

Centers for Disease Control (1980) Toxic-shock syndrome – United States. *MMWR.* 29, 229–230.

Chenoweth, C.E., Washer, L.L., Obeyesekera, K., Friedman, C., Brewer, K., Fugitt, G.E. and Lark, R. (2007) Ventilator-associated pneumonia in the home care setting. *Infect. Control. Hosp. Epidemiol.* 28, 910–915.

Cheung, B.M., Ong, K.L., Man, Y.B., Wong, L.Y., Lau, C.P. and Lam, K.S. (2006) Prevalence of the metabolic syndrome in the United States National Health and Nutrition Examination Survey 1999–2002 according to different defining criteria. *J. Clin. Hypertens. (Greenwich)* 8, 562–570.

Chobanian, A.V., Bakris, G.L., Black, H.R., Cushman, W.C., Green, L.A., Izzo, J.L. Jr., Jones, D.W., Materson, B.J., Oparil, S., Wright, J.T. Jr,

et al. (2003) Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* 42, 1206–1252.

Davies, H.T., Crombie, I.K. and Tavakoli, M. (1998) When can odds ratios mislead? *BMJ* 316, 989–991.

Dawber, T.R., Meadors, G.F. and Moore, F.E. Jr. (1951) Epidemiological approaches to heart disease: the Framingham Study. *Am. J. Public Health Nations Health* 41, 279–281.

Delgado-Rodriguez, M. and Llorca, J. (2004) Bias. *J. Epidemiol. Community Health* 58, 635–641.

Doll, R. and Hill, A.B. (1956) Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *BMJ* 2, 1071–1081.

Ernster, V.L. (1994) Nested case-control studies. *Prev. Med.* 23, 587–590.

Evans, A.S. (1976) Causation and disease: the Henle–Koch postulates revisited. *Yale J. Biol. Med.* 49, 175–195.

Flores, G., Fuentes-Afflick, E., Carter-Pokras, O., Claudio, L., Lamberty, G., Lara, M., Pachter, L., Ramos Gomez, F., Mendoza, F., Valdez, R.B. *et al.* (2001) Why ethnicity and race are so important in child health services research today. *Arch. Pediatr. Adolesc. Med.* 155, 1178–1179.

Ha, M.H., Lee, D.H. and Jacobs, D.R. (2007) Association between Serum Concentrations of Persistent Organic Pollutants and Self-Reported Cardiovascular Disease Prevalence: Results from the National Health and Nutrition Examination Survey, 1999–2002. *Environ. Health Perspect.* 115, 1204–1209.

Hemann, B.A., Bimson, W.F. and Taylor, A.J. (2007) The Framingham Risk Score: an appraisal of its benefits and limitations. *Am. Heart Hosp. J.* 5, 91–96.

Herrington, D.M., Reboussin, D.M., Brosnihan, K.B., Sharp, P.C., Shumaker, S.A., Snyder, T.E., Furberg, C.D., Kowalchuk, G.J., Stuckey, T.D., Rogers, W.J. *et al.* (2000) Effects of estrogen replacement on the progression of coronary-artery atherosclerosis. *N. Engl. J. Med.* 343, 522–529.

Hill, A.B. (1965) The Environment and Disease: Association or Causation?. *Proc. R. Soc. Med.* 58, 295–300.

Hulley, S., Grady, D., Bush, T., Furberg, C., Herrington, D., Riggs, B. and Vittinghoff, E. (1998) Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 280, 605–613.

Kaplan, J.B. and Bennett, T. (2003) Use of race and ethnicity in biomedical publication. *JAMA* 289, 2709–2716.

Karhausen, L.R. (2000) Causation: the elusive grail of epidemiology. *Med. Health Care Philos.* 3, 59–67.

Kostis, W.J., Demissie, K., Marcella, S.W., Shao, Y.H., Wilson, A.C. and Moreyra, A.E. and Myocardial Infarction Data Acquisition System (MIDAS 10) Study Group (2007) Weekend versus weekday admission and mortality from myocardial infarction. *N. Engl. J. Med.* 356, 1099–1109.

Lanes, S.F. and Poole, C. (1984) Truth in packaging? The unwrapping of epidemiologic research. *J. Occup. Med.* 26, 571–574.

Langholz, B. and Thomas, D.C. (1990) Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *Am. J. Epidemiol.* 131, 169–176.

Lawlor, D.A., Smith, G.D. and Ebrahim, S. (2004) Socioeconomic position and hormone replacement therapy use: explaining the discrepancy in evidence from observational and randomized controlled trials. *Am. J. Public Health* 94, 2149–2154.

Mastin, J.P. (2005) Environmental cardiovascular disease. *Cardiovasc. Toxicol.* 5, 91–94.

Naylor, C.D., Chen, E. and Strauss, B. (1992) Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness?. *Ann. Intern. Med.* 117, 916–921.

Olijhoek, J.K., van der Graaf, Y., Haffner, S.M. and Visseren, F.L. for the SMART Study Group (2007) Defining the metabolic syndrome: Resolving unresolved issues?. *Eur. J. Intern. Med.* 18, 309–313.

Osborne, N.G. and Feit, M.D. (1992) The use of race in medical research. *JAMA* 267, 275–279.

Parascandola, M. and Weed, D.L. (2001) Causation in epidemiology. *J. Epidemiol. Commun. Health* 55, 905–912.

Prentice, R.L. (1986) A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 73, 1–11.

Sax, L. (2002) How common is intersex? A response to Anne Fausto-Sterling. *J. Sex Res.* 39, 174–178.

Schuchat, A. and Broome, C.V. (1991) Toxic shock syndrome and tampons. *Epidemiol. Rev.* 13, 99–112.

Semrad, T.J., O'Donnell, R., Wun, T., Chew, H., Harvey, D., Zhou, H. and White, R.H. (2007) Epidemiology of venous thromboembolism in 9489 patients with malignant glioma. *J. Neurosurg.* 106, 601–608.

Shands, K.N., Schlech, W.F. IIIrd, Hargrett, N.T., Dan, B.B., Schmid, G.P. and Bennett, J.V. (1982) Toxic shock syndrome: case-control studies at the Centers for Disease Control. *Ann. Intern. Med.* 96, 895–898.

Smith, G.D. (2004) Classics in epidemiology: should they get it right? *Int. J. Epidemiol.* 33, 441–442.

Stampfer, M.J. and Colditz, G.A. (1991) Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev. Med.* 20, 47–63.

Todd, J., Fishaut, M., Kapral, F. and Welch, T. (1978) Toxic-shock syndrome associated with phage-group-I Staphylococci. *Lancet.* 2, 1116–1118.

Todd, J.K. (1982) Toxic shock syndrome: a perspective through the looking glass. *Ann. Intern. Med.* 96, 839–842.

Wacholder, S. (1991) Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 2, 155–158.

Waters, D.D., Alderman, E.L., Hsia, J., Howard, B.V., Cobb, F.R., Rogers, W.J., Ouyang, P., Thompson, P., Tardif, J.C., Higginson, L. *et al.* (2002) Effects of hormone replacement therapy and antioxidant vitamin supplements on coronary atherosclerosis in postmenopausal women: a randomized controlled trial. *JAMA* 288, 2432–2440.

Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H. and Kannel, W.B. (1998) Prediction of coronary heart disease using risk factor categories. *Circulation* 97, 1837–1847.