CHAPTER 1

Introduction

This chapter introduces health professionals in community medicine, health policy and management, nursing, and various health fields allied to epidemiology and biostatistics. We begin by defining *health* and presenting analytic competencies fundamental to public health. We then define *epidemiology*, along with study designs and the scientific method, which are integral parts of epidemiology. *Biostatistics* is also defined, and its supportive role in epidemiology is described. The role epidemiology and biostatistics play in providing health professionals with the skills necessary to meet the analytic competencies of public health is emphasized.

HEALTH

The World Health Organization (WHO) defines *health* as a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.[1] *Physical health* is the ability of the human body to function properly and includes physical fitness and activities of daily living; *mental health* is the ability to think clearly, reason objectively, and act properly; and *social health* is the ability to have satisfying relationships and interaction within social institutions and societal mores. In more recent years, three additional dimensions of health have been added: *emotional health* (one's ability to cope, adjust, and adapt), *spiritual health* (one's personal beliefs and choices), and *environmental health* (one's surroundings, such as habitat or occupation). The field of medicine concerned with safeguarding and improving the health of a community as a whole is *public health*.[2] Safeguarding and improving the public's health is accomplished by efforts in preventive medicine, health education, control of communicable diseases, application of sanitary measures, and monitoring environmental hazards.[3]

To meet the goals of public health, core functions have been set forth that involve assessment, policy development, and assurance.[4] Ten essential public health services have also been given.[5] The 3 core functions of public health and the 10 essential public health services categorized under the appropriate core functions are presented in **Table 1.1**. Analytic competencies for each of the 10 services are also shown in the table. Because the focus of this text is on analytic competencies in public health, other

TABLE 1.1 Competencies for the 10 Essential Public Health Services

I. Assessment	**Analytic competencies[6]**
• Monitor health status to identify and solve community health problems. • Diagnose and investigate health problems and health hazards in the community.	• Define the problem. • Determine appropriate use of data and statistical methods for problem identification and resolution, as well as program planning, implementation, and evaluation. • Select and define variables relevant to defined public health problems. • Evaluate the integrity and comparability of data and identify gaps in data sources. • Understand how the data illuminate ethical, political, scientific, economic, and overall public health issues. • Understand basic research designs used in public health. • Make relevant inferences from data.
II. Policy development	
• Inform, educate, and empower people about health issues.	
• Mobilize community partnerships and action to identify and solve health problems.	
• Develop policies and plans that support individual and community health efforts.	
III. Assurance	
• Enforce laws and regulations that protect health and ensure safety.	
• Link people to needed personal health services and assure the provision of health care when otherwise unavailable.	• Define a problem. • Make relevant inferences from data.
• Assure competent public and personal healthcare workforce.	• Determine appropriate use of data and statistical methods for problem identification and resolution, and program planning, implementation, and evaluation.

TABLE 1.1 Competencies for the 10 Essential Public Health Services (*continued*)

III. Assurance	Analytic competencies[6]
• Evaluate effectiveness, accessibility, and quality of personal and population-based health services.	• Evaluate the integrity and comparability of data and identify gaps in data sources. • Understand how the data illuminate ethical, political, scientific, economic, and overall public health issues. • Understand basic research methodologies used in public health and health services research. • Make relevant inferences from data.
• Research for new insights and innovative solutions to health problems.	• Define a problem. • Determine appropriate use of data and statistical methods for problem identification and resolution, and program planning, implementation, and evaluation. • Select and define variables relevant to defined public health problems. • Evaluate the integrity and comparability of data and identify gaps in data sources. • Understand how the data illuminate ethical, political, scientific, economic, and overall public health issues. • Understand basic research designs used in public health. • Make relevant inferences from data.

Data from: U.S. Department of Health and Human Services. The public health workforce: An agenda for the 21st century. A report of the Public Health Functions Project. http://www.health.gov/phfunctions/pubhlth.pdf. Accessed June 14, 2011.

competencies such as communication skills, policy and development/program planning skills, cultural skills, financial planning and management skills, and other basic public health skills are not listed.

Many disciplines are involved in public health research, but it rests upon a scientific core of epidemiology. Monitoring, diagnosing, and investigating disease and health-related events are primary functions of epidemiology that assist in decisions about appropriate action and identify whether progress is being made in prevention and control efforts. Biostatistical methods also play a critical

Definitions

Research is a scientific or scholarly investigation that aims to advance human knowledge.

Epidemiology is the study of the distribution and determinants of health-related states or events in specified human populations, and the application of this study to prevent and control health problems.[11]

Definitions

Surveillance is the process of observing or monitoring.

Monitoring is a process in which changes in health status over time or among populations are identified in order to recognize health problems and to assess progress toward health goals or objectives.

Diagnosing is the act or process of identifying or determining the nature of a case of a disease or distinguishing one disease from another.

Investigating is the act or process of inquiry or systematic study to discover and examine the facts so as to establish the truth.

Biostatistics is the science of statistics applied to biologic or medical data; it is a contraction of biology and statistics.

role in public health and, in particular, are fundamental to carrying out descriptive and analytic epidemiologic studies. Because epidemiology draws heavily on biostatistics, the distinction between the two is often unclear. However, there are some important differences that will be discussed in this chapter.

EPIDEMIOLOGY

The first known epidemiologist, also known as the father of medicine, is the Greek physician Hippocrates (460–377 BCE). Hippocrates addressed how selected diseases were related to time and seasons, place, environmental conditions, and especially in relation to water and the seasons. He also taught about causal factors, how diseases affect people, and how diseases spread. He classified illness as acute, chronic, endemic, and epidemic.[7–10]

The public health problem is established through observed or measured phenomena in the population of interest. Characterizing acute, chronic, endemic, and epidemic disease states and conditions according to person, place, and time factors remains an important way epidemiologists are able to identify unique features that may provide clues or evidence regarding the source of the agent and the nature of the exposure leading to given health problems.

The word *epidemiology* is based on three Greek words: *epi*, the prefix, meaning "on, upon, or befall"; *demos*, the root, meaning "the people"; and *logos*, the suffix, meaning "the study of." In accordance with medical terminology, the suffix is read first and then the prefix and the root, such that the word *epidemiology* literally means "the study of that which befalls the people." Today, *epidemiology* involves the study of the distribution and determinants of diseases and health-related events (e.g., body weight and diet). To better understand the word *epidemiology*, descriptions of the various parts of the definition are presented in **Table 1.2**.

Disease processes are complex and require that we understand biology, anatomy, physiology, histology, biochemistry, microbiology, and related medical sciences. A *disease* is any impairment of normal physiological function affecting all or part of an organism, especially a specific pathological change, which may be caused by pathogens (viruses, bacteria, fungi, or parasites capable of producing disease); chemicals

TABLE 1.2 Defining Epidemiology

Term	Description
Study	A careful examination of a phenomenon; implies sound methods of scientific investigation. Contributing factors to epidemiology include advances in biology, medicine, statistics, and social and behavioral sciences.
Distribution	Distribution refers to the frequency and pattern of a health-related state or event. Frequency is the number of occurrences of a health-related state or event; pattern involves presenting the distribution by person, place, and time characteristics.
Determinants	Determinants are factors that produce an effect, result, or consequence in another factor. A determinant is a cause. Determinants may be: • Physical stresses—excessive heat, cold, and noise; radiation (electromagnetic, ultrasound, microwave, or x-irradiation); climate change; ozone depletion; housing; and so on • Chemicals—drugs, acids, alkali, heavy metals (lead and mercury), poisons (arsenic), and some enzymes • Biological—disease-causing infectious agents or pathogens (viruses, bacteria, fungi, and parasites) • Psychosocial milieu—families and households, socioeconomic status, social networks and social support, neighborhoods and communities, access to health care, formal institutions, and public policy
Health-related states or events	*Health-related states or events* refers to the fact that epidemiology involves more than just the study of disease states; it also includes the study of events, behaviors, and conditions associated with health: • A *disease* is an interruption, cessation, or disorder of body functions, systems, or organs (e.g., cholera, angina, breast cancer, and influenza). • An *event* is something that takes place (e.g., vehicular collisions, workplace injuries, drug overdose, and suicide). • A *behavior* is a manner of conducting oneself (e.g., physical activity, diet, and safety precautions). • A *condition* is an existing circumstance (e.g., an unhealthy state, a state of fitness, something that is essential to the occurrence of something else).
Application	Application refers to the fact that the information obtained through epidemiology is applied to better prevent and control health problems.

Definitions

Epidemic refers to the occurrence of cases of an illness, specific health-related behavior, or other health-related events clearly in excess of normal expectancy in a community or region.

Endemic refers to the constant presence or usual frequency of a specific disease in a particular population; a disease is said to be endemic when it continually prevails in a region.

Pandemic is an epidemic that is extensive, involving large regions, countries, or continents.

Acute diseases or conditions are characterized by the rapid onset of symptoms that are usually severe with short duration.

Chronic diseases or conditions are long-lasting or recurrent, lasting three months or more.

(e.g., drugs, acids, heavy metals); inherent weaknesses (e.g., suppressed immune system, susceptibility gene); lifestyle (e.g., lack of physical activity, smoking, obesity); and physical stresses (e.g., radiation, extreme temperatures, workplace injuries), and is characterized by an identifiable group of signs or symptoms. The ability of a pathogen to get into a susceptible host and cause a disease is called *invasiveness*. The disease-evoking power of a pathogen is called *virulence*.

In general, diseases may be classified as congenital and hereditary diseases, allergies and inflammatory diseases, degenerative diseases, metabolic diseases, or cancer. *Congenital and hereditary diseases* are often caused by genetic and familial tendencies toward certain inborn abnormalities; injury to the embryo or fetus by environmental factors, chemicals, or agents such as drugs, alcohol, or smoking; or innate developmental problems possibly caused by chemicals or agents. They can also be a fluke of nature. Examples are Down syndrome, hemophilia, and heart disease present at birth.[12] *Allergies* and *inflammatory diseases* are caused by the body reacting to an invasion of or injury by a foreign object or substance. Allergies, viruses, bacteria, or other microscopic and microbiological agents can cause an inflammatory reaction in the body. Some inflammatory reactions may result in the body forming antibodies. Antibodies are formed as a first line of defense. They are protein substances or globulins derived from B and T lymphocytes that originate in the bone marrow.[12] *Degenerative diseases* cause a lower level of mental, physical, or moral state than is normal or acceptable. Degenerative diseases are often associated with the aging process but in some cases may not be age related. Arteriosclerosis, arthritis, and gout are examples of degenerative chronic diseases.[12] *Metabolic diseases* cause the dysfunction, poor function, or malfunction of certain organs or physiologic processes in the body, leading to disease states. Glands or organs that fail to secrete certain biochemicals to keep the metabolic process functioning in the body cause metabolic disorders. For example, adrenal glands may stop functioning properly, causing Addison's disease; the cells may no longer use glucose normally, causing diabetes; or the thyroid gland may fail, resulting in a goiter, hyperthyroidism, or cretinism (hypothyroidism).[12] *Cancer* is a collective name that refers to a group of many diseases with one common characteristic: uncontrolled cell growth

or the loss of the cell's ability to perform apoptosis (cell suicide). The gradual increase in the number of uncontrolled dividing cells creates a mass of tissue called a tumor (neoplasm). When a tumor is malignant, meaning it is capable of spreading to surrounding tissue or remote places in the body, it is called cancer.[13]

Infectious diseases are caused when pathogens enter, survive, and multiply in the host. An infectious disease may or may not be contagious. Infectious disease is contagious if it is capable of being communicated or transmitted. Infectious diseases may be either acute or chronic. Noninfectious diseases are noncommunicable, but may be either acute or chronic. Health-related events are noninfectious and may be acute or chronic (**Table 1.3**).

Cancer is a disease that may be infectious or noninfectious. Although cancer is generally thought of as being lifestyle related (e.g., tobacco, diet), 10–20% of all cancer has been related to parasites, bacteria, and viruses.[14,15] For example, Epstein-Barr virus is associated with Burkitt's lymphoma, human papillomavirus is associated with cervical

TABLE 1.3 Classification of Selected Diseases

	Communicable		Noncommunicable	
	Acute	**Chronic**	**Acute**	**Chronic**
Infectious	Influenza/pneumonia	Cancer	Anthrax	
	Cholera	Leprosy	Legionnaires' disease	
	Lyme disease	Polio	Tetanus	
	Mumps	Syphilis		
	Measles	Tuberculosis		
Noninfectious			Accidents	Alcoholism
			Drug abuse	Arthritis
			Epilepsy	Cancer
			Heart attack	Cardiovascular disease
			Homicide	
			Seizures	Diabetes mellitus
			Stroke	
			Suicide	Oral health problems
				Paralysis

cancer, hepatitis B virus is associated with liver cancer, human T-cell lymphotrophic virus is associated with adult T-cell leukemia, Kaposi's sarcoma–associated herpesvirus is associated with Kaposi's sarcoma, and bacterium *H. pylori* is associated with stomach cancer. Some known inherited conditions that increase the risk of cancer include hereditary retinoblastoma linked to retinoblastoma; xeroderma pigmentosum linked to skin cancer; Wilms' tumor linked to kidney cancer; Li-Fraumeni syndrome linked to sarcomas, leukemia, and cancers involving the brain and breast; familial adenomatous polyposis linked to colorectal cancer; Paget's disease of the bone linked to bone cancer; and Fanconi's aplastic anemia linked to leukemia, and liver and skin cancers.

Definitions

Primary prevention is preventing a disease or disorder before it happens through activities such as lifestyle changes; community health education; school health education; good prenatal care; good behavior choices; proper nutrition; and safe and healthful conditions at home, school, and the workplace.

Secondary prevention is aimed at the health screening and detection activities used to identify disease. It also aims to block the progression of disease or prevent an injury from developing into an impairment or disability.

Tertiary prevention consists of limiting any disability by providing rehabilitation (attempt to restore an afflicted person to a useful, productive, and satisfying quality of life) where a disease, injury, or disorder has already occurred and caused damage. Prompt diagnosis and treatment, followed by proper rehabilitation and posttreatment recovery, proper patient education, behavior changes, and lifestyle changes are all necessary to improve the patient's prognosis.[16]

Each disease has a natural history of progression if no medical intervention is taken and the disease runs its full course. There is a stage of susceptibility, a stage of presymptomatic disease, a stage of clinical disease, and a stage of outcome (recovery, disability, or death). The stage of susceptibility precedes the disease and reflects the likelihood a host has of becoming ill from a given source. The stage of presymptomatic disease involves the time from exposure and subsequent pathologic changes up to the time of clinical symptoms. This time is called the incubation period. In the context of chronic disease, it is called the latency period. The stage of clinical disease is when signs and symptoms are manifest. The final stage reflects the patient's prognosis. Factors that influence these stages include primary prevention, secondary prevention, and tertiary prevention.

Understanding and influencing health-related behaviors is also complex and involves psychology, sociology, and economics. Many behavior change models are based on the idea that health education itself is not sufficient to promote behavior change. The Health Belief Model is a widely used conceptual framework for understanding health behavior.[17] In this model, behavior change requires a rational decision-making process that considers perceived susceptibility to illness, perceived consequences or seriousness of the illness, belief that recommended action is appropriate or efficacious to reduce risk, and belief that the benefits of action outweigh the costs.[18–20] Two exten-

TABLE 1.4 Concepts, Definitions, and Applications of the Health Belief Model

Concept	Definition	Application
Perceived susceptibility	One's opinion of chances of getting a condition	Define population(s) at risk, risk levels; personalize risk based on a person's features or behavior; heighten perceived susceptibility if too low
Perceived severity	One's opinion of how serious a condition and its consequences are	Specify consequences of the risk and the condition
Perceived benefits	One's belief in the efficacy of the advised action to reduce risk or seriousness of impact	Define action to take; how, where, when; clarify the positive effects to be expected
Perceived barriers	One's opinion of the tangible and psychological costs of the advised action	Identify and reduce barriers through reassurance, incentives, assistance
Cues to action	Strategies to activate "readiness"	Provide how-to information, promote awareness, reminders
Self-efficacy	Confidence in one's ability to take action	Provide training, guidance in performing action

Data from: Glanz K, Marcus Lewis F, Rimer BK. *Theory at a Glance: A Guide for Health Promotion Practice.* Bethesda, MD: National Institutes of Health; 1997.

sions of these concepts in more recent years include cues to action and self-efficacy (**Table 1.4**).[21]

TRANSMISSION OF DISEASE

The ability of a disease to be transmitted from one person to another or to spread throughout the population is called *communicability*. The communicability of a disease is determined by how likely a pathogen or agent will be transmitted from a diseased or infected person to another susceptible person. Infectious communicable diseases may be transmitted vertically or horizontally. *Vertical transmission* refers to the transmission from an individual to its offspring through the placenta, milk, or vagina. *Horizontal transmission* refers to the transmission of an infectious agent from an infected individual to a susceptive contemporary.

Definitions

Fomites are objects such as clothing, towels, and utensils that can harbor a disease agent and are capable of transmitting it.

A **vector** is an invertebrate animal (e.g., tick, mite, mosquito, blood-sucking fly) that is capable of transmitting an infectious agent among vertebrates.

A **reservoir** is the habitat (living or nonliving) in or on which an infectious agent lives, grows, and multiplies and on which it depends for its survival in nature.

Zoonoses are those diseases and infections that are transmitted between vertebrate animals and humans.

A **vehicle** is a nonliving intermediary such as a fomite, food, or water that conveys the infectious agent from its reservoir to a susceptible host.

A **carrier** contains, spreads, or harbors an infectious organism.

Horizontal transmission may involve direct transmission, a common vehicle, an airborne pathogen, or a vector-borne pathogen. Examples of direct transmission include chlamydia, cytomegalovirus, genital warts, gonorrhea, herpes, intestinal parasites, scabies, and syphilis, which are caused by infectious agents passed from one person to another during sexual contact. Examples of a common vehicle include water-borne diseases caused by pathogenic microorganisms that are directly transmitted by consuming contaminated water, food-borne illness (commonly called food poisoning) caused by the consumption of a number of food-borne bacteria and viruses (e.g., *E. coli* O157:H7, salmonella), or blood-borne diseases caused mainly by parasites and viruses (e.g., malaria is one of the most common blood-borne diseases on earth, affecting nearly a half-billion people). Examples of airborne pathogens include pathogenic microorganism spread by droplet nuclei through the air by coughing, sneezing, singing, or talking (e.g., haemophilus influenza, measles, meningitis, mumps, pertussis, tuberculosis, diphtheria, plague, and varicella). Examples of vector-borne pathogens are pathogens transmitted by blood-feeding arthropods. Some vector-borne diseases include malaria, dengue, yellow fever, louse-borne typhus, plague, leishmaniasis, sleeping sickness, West Nile encephalitis, Lyme disease, Japanese encephalitis, Rift Valley fever, and Crimean-Congo hemorrhagic fever.

DESCRIPTIVE EPIDEMIOLOGY

The definition of epidemiology indicates both descriptive and analytic features. Various methods in descriptive epidemiology, which are related to "distribution" (see the definition of epidemiology), are used to monitor the health of a population, identify a public health problem and its extent, and identify unique characteristics associated with those experiencing the health problem. Thus, clues may be obtained as to the determinants of the health problem. However, it is analytic epidemiology that provides methods to test hypotheses about causal associations.

The application of epidemiologic methods to assess the distribution of health problems in human populations is the essence of descriptive epidemiology. Descrip-

History

An Irish cook, Mary Mallon (1869–1938), also called Typhoid Mary, was believed to be responsible for 53 cases of typhoid fever in a 15-year period. She was a cook who served in many homes that were stricken with typhoid. The disease followed, but never preceded, her employment. Bacteriologic examination of Mary Mallon's feces showed that she was a chronic carrier of typhoid. Mary seemed to sense that she was giving people sickness, because when typhoid appeared, she would leave with no forwarding address. Mary Mallon illustrated the importance of concern over the chronic typhoid carrier causing and spreading typhoid fever. Like 20% of all typhoid carriers, Mary suffered no illness from the disease. Epidemiologic investigations have shown that carriers might be overlooked if epidemiologic searches are limited to the water, food, and those with a history of the disease.[22,23]

tive epidemiology is used to describe the health of communities and to identify health problems and priorities according to person (who?), place (where?), and time (when?) factors (**Table 1.5**). Descriptive epidemiology also involves characterizing the nature of the health problem (what?).

Characteristics of disease cases themselves, in terms of a variety of person characteristics, may portray the uniqueness of the patient population. If a given attribute emerges as being common, at-risk groups may be identified and the idea of a specific exposure may be presented. Birthplace, workplace, community, and so on may identify differences in environmental factors that provide evidence of potential risk behaviors or factors for disease. Place characteristics may also play a role in disseminating environmental pathogens or chemical agents that may influence disease. Time plays various roles in descriptive epidemiology. Monitoring health status over time is useful for identifying and solving community health problems. Time trends can identify health problems that are endemic or epidemic. An epidemic may arise from a specific source or from infections transmitted from one infected person to another.[24] The time from the point of exposure until clinical symptoms arise is the incubation period. Knowing the incubation period of the disease can help narrow the likely cause and lead to a solution to the problem. Some examples of incubation periods are 12–36 hours for botulism, 2–3 weeks for chicken pox, 12–72 hours (usually about 24 hours) for the common cold, 1–3 days for influenza, 2 weeks or longer for rabies, 1–3 days for scarlet fever, 4–12 weeks for tuberculosis, and 7–21 days for whooping cough.[16] For chronic disease, the time from exposure to clinical symptoms is called the latency period. The latency period for several cancer types can be several years (e.g., about 20 to 25 years for smoking and lung cancer, and 40 or more years for a blistering sunburn during childhood or adolescence and melanoma of the skin). Comparing time trends of chronic disease

TABLE 1.5 Defining Descriptive Epidemiology

Characteristic	Question	Attributes	Examples
Person	Who is affected?	Inherent	Age, gender, race/ethnicity, blood type, genetics, family structure
		Acquired	Immunity, marital status, education, beliefs, traditions
		Activities	Occupation, leisure, medication use, religious practices
		Conditions	Air quality, water quality, food quality, housing conditions, living conditions, safety of the environment, areas for walking and exercising, health services
Place	Where are they affected?		Birthplace, workplace, classroom, school, cafeteria, jail, country, state, county, census tract; water supplies, sewage disposal outflows, ecological habitats, air-flow patterns in buildings, prevailing wind currents, milk distribution routes
Time	When are they affected?		Epidemic curve
			Incubation period
			Secular or seasonal trend
			Latency period
			Duration of the course of the health-related state or event
			Life expectancy of the host or pathogen
Health-related state or event	What observed features distinguish the health-related state or event?		An instance of something; an occurrence
			Clinical signs and symptoms

and conditions in relation to suspected risk factors may also provide clues as to the cause of the disease.

If the duration time of the epidemic is reflected on a graph, with the vertical axis representing frequency and the horizontal axis representing time, the shape of the graph is called the epidemic curve. The time begins at the point of exposure and

extends over the course of the epidemic, and may reflect hours, days, weeks, or longer. The shape of the epidemic curve reflects whether the exposure occurred at a point in time or was continuous.

A *case definition* consists of a set of criteria that identifies what is involved.[25] A standard case definition ensures that cases are consistently diagnosed, regardless of where or when they are diagnosed or by whom. If case definitions are inconsistently applied, classifying the health problem becomes problematic. A case definition may be limited according to person, place, and time criteria. Person criteria typically include clinical symptoms (e.g., diarrhea, stomach cramps, and sometimes vomiting or fever for *Salmonella* food poisoning) or clinical tests (e.g., pneumonia on chest X-ray). Place criteria may require the case be employed at a certain location, and time criteria may require that a case be diagnosed during a certain time period. For certain rare but lethal communicable diseases (e.g., plague) where a quick response is needed, a loose case definition may be appropriate. For certain other situations where a quick response is less critical and identifying causal associations is important (e.g., cancer), a stricter set of criteria for establishing the presence of the disease is warranted. The identifying features (e.g., signs, symptoms, disease progression, place and type of exposure, lab findings) will depend on the condition and disease under investigation.

Definition

A **case definition** in epidemiology is a set of standard criteria for diagnosing a particular disease or health-related condition, by specifying clinical criteria and limitations on person, place, and time factors.

ANALYTIC EPIDEMIOLOGY

Analytic epidemiology is the application of epidemiologic methods to study the determinants of health-related states or events in human populations. Analytic epidemiology answers questions about why and how a health-related state or event occurred. Analytic methods involve formulating and testing hypotheses about associations between exposure and outcome variables. Statistical measures of association are used, and causal theory is often employed. Analytic studies provide greater evidence than do descriptive studies to support a hypothesis of causal association.

Analytic epidemiologic study designs employ a comparison group in order to evaluate associations between exposure and outcome variables. For example, selected future health outcomes are compared between those exposed and those not exposed (cohort study), or previous exposure status is compared between cases and noncases (case-control study). Use of a comparison group is also important in evaluating the efficacy and effectiveness of prevention and control programs.[26] For

Definitions

Efficacy refers to the ability of a program to produce a desired effect among those who participate in the program compared with those who do not.

Effectiveness refers to the ability of a program to produce benefits among those who are offered the program.

A **case report** is an in-depth investigation of a problem or situation of a single individual; it includes qualitative descriptive research of the facts in chronological order.

A **case series** is a small group of patients with a similar diagnosis.

A **cross-sectional study** is a collection of data at a single point in time for each participant or system being studied. It is assumed that the phenomena of interest remain static through the period of study.

An **ecologic study** refers to aggregate data involved instead of individual-level data, such as comparing injury rates from one occupation to another.

In an **experimental study**, participants are deliberately manipulated for the purpose of studying an intervention effect. An intervention is assigned to selected participants to determine its effect on a given outcome.

example, researchers may be interested in whether the recovery time is quicker among heart attack patients who adhere to a given dietary program compared with those who do not. If those who comply with the program have much quicker recovery than those who do not comply, the dietary program is efficacious, but if the program is so strict that very few are able to adhere to it, it is not efficient.

STUDY DESIGNS

A study design is a detailed plan or approach for systematically collecting, analyzing, and interpreting data; it is a formal approach of scientific or scholarly investigation. Study designs in epidemiology are classified as descriptive or analytic. A descriptive study design is used to assess and monitor the health of communities and identify health problems and priorities according to who is affected, where they are affected, and when they were most likely affected. It also involves describing what the health problem is. These study designs include case studies, cross-sectional studies, and ecologic studies. Descriptive epidemiologic studies also lend support to more definitive evaluation using analytic methods.

Although these study designs are generally classified as descriptive, in some cases when a specific research hypothesis is formulated and cross-sectional or ecologic data is used to assess the hypothesis, we may refer to it as an analytic study.

Analytic study designs are used to test one or more predetermined hypotheses about associations between exposure and outcome variables. Analytic epidemiology provides information on how and why a health-related state or event occurred. There are both observational (i.e., the observed variables are not controlled by the investigator) and experimental (i.e., the exposure status is controlled by the investigator) study designs. For exposures that cannot be ethically assigned in an experimental study (e.g., requiring participants to smoke

or not smoke cigarettes), exposure–disease relationships are better assessed in observational study designs. Analytic observational study designs include case control, cohort, case crossover, and nested case control.

Experimental (or intervention) study designs may involve a between-group design, a within-group design, or a combination of both types of designs. The strongest methodological design is a between-group design in which the outcome of interest is compared between one group receiving the intervention of interest and another group receiving no active treatment (preferably a placebo) or a currently accepted treatment. A within-group design (also called a time-series design) may also be used, but the outcome of interest is compared before and after an intervention. The between-group design is more prone to individual type confounding factors (e.g., gender, race, genetic susceptibility). The within-group design is more susceptible to confounding from time-related factors (e.g., learning effects where participants do better on follow-up cognitive tests because they learned from the baseline test, influences from the media, or other external factors).

Definitions

A **case-control study** involves grouping people as cases (persons experiencing a health-related state or event) and controls, and investigating whether the cases are more or less likely than the controls to have had past experiences, lifestyle behaviors, or exposures.

A **cohort** is a group or body of people, often defined by experiencing a common event (e.g., birth, training, or enrollment) in a given time period.

A **cohort study** involves grouping people into exposure categories (e.g., smokers versus nonsmokers) and investigating whether exposure classifications are associated with one or more health outcomes in the future.

A **case-crossover study** compares the exposure status of a case immediately before its occurrence with that of the same case at a prior time.

A **nested case-control study** is a case-control study nested within a cohort study (also called a case-cohort study).

SCIENTIFIC METHOD

Epidemiology employs the scientific method to identify and solve public health problems. The scientific method involves using appropriate study designs and statistical methods for investigating an observable occurrence and acquiring new knowledge. Descriptive epidemiologic methods are used to identify the presence of a health problem. Once the research problem is established, hypotheses are formulated to evaluate possible explanations for the health problem. Hypothesis testing is central to the scientific method. After the researcher has formulated a research hypothesis, data are collected and assessed to either support or fail to support the hypothesis. If the hypothesis is rejected, alternatives are considered. If the data are consistent with the purported hypothesis, the hypothesis is retained. Information is

analyzed in epidemiology in much the same way as it is analyzed in prospective laboratory experiments, but because the investigator often does not have control over the exposure and outcome measures, extra care must be taken to find convincing evidence that the hypothesized chain of events is not due to bias.

BIOSTATISTICS

Biostatistics can be thought of as having two parts, *bio*, meaning biology (the science concerned with the phenomena of life and living organisms, and in public health specifically referring to human life), and *statistics*. The word *statistics* has multiple meanings, such as data or numbers, the process of collecting and analyzing data for patterns and relationships, and the description of a field of study. Working with numbers may involve using statistical methods that summarize and describe person, place, and time factors, and applying statistical methods to draw certain conclusions that can be applied to public health. The application of statistics is broad, but when applied to biology or medicine, it is termed *biostatistics*.[24]

Definition

Statistics is the science of data and involves collecting, classifying, summarizing, organizing, analyzing, and interpreting data.

Several important individuals helped develop the use of statistics in the context of biology, including Sir Ronald A. Fisher (1890–1962), Sewall G. Wright (1889–1988), and J. B. S. Haldane (1892–1964). In 1930, Fisher published *The Genetical Theory of Natural Selection*, wherein he developed a number of basic statistical methods; Wright and Haldane used statistics in developing modern population genetics. These

History

Florence Nightingale (1820–1910), a British nurse, possessed one of the greatest analytical minds of her time. In 1854, Nightingale was appointed head of a contingent of nurses sent to care for soldiers in the Crimean War, where she effected changes in hospital sanitation that significantly reduced soldiers' mortality rates from illness. In 1856, Nightingale returned to England and spent the rest of her life enacting reform of army sanitation, hospitals, and a number of other public health efforts. Much of her work was accomplished through gathering and sharing statistics and writing reports and letters. She used data as a tool for improving city and military hospitals; monitored disease mortality rates, which showed that with improved sanitary methods in hospitals, the rates of death decreased; and developed applied statistical methods to display her data, showing that statistics provided an organized way of learning and improving medical and surgical practices.[33–35]

and other biostatisticians played a critical role in adding quantitative discipline to the study of biology.[27–32]

Biostatistics may be divided into four general areas: descriptive, probability, inferential, and statistical techniques. Descriptive biostatistics involves methods of organizing, summarizing, and describing numerical data relating to living organisms. A second area of biostatistics study is probability as it relates to living organisms. Probability covers random variables and probability distributions; sampling methods, sampling distributions and the central limit theorem. A third area of biostatistics involves inference about a population's characteristics from information in the sample. Finally, there are several statistical methods that are used for investigating a range of problems involving biologic data, which we refer to as statistical techniques.

Biostatistics can involve the application of statistics to any living system, whether human, animal, plant, fungus, or microorganism. On the other hand, epidemiology focuses on human populations. Hence, biostatistics within epidemiology involves the study of the frequency, patterns, and relationships of health-related states or events in human populations. However, unlike epidemiology, biostatistics does not focus on the study of determinants of health-related states or events in human populations; that is, although understanding causal processes is a primary focus in epidemiology, it is not a focus of biostatistics.

EVALUATING THE LITERATURE

The ability to access the scientific literature is vital for health professionals in order to stay current in their knowledge about disease management and prevention. Medical practitioners should read the literature to stay abreast of the most effective screening and diagnostic procedures and treatments and the most effective approach for implementing the treatments. In addition, knowledge of likely health outcomes according to specific prognostic indicators is important for counseling patients. From a disease prevention point of view, the literature indicates risk factors associated with a given health-related state or event and describes health promotion and disease prevention programs. Public health officials are often required to draw upon the literature in their health planning and decision making for establishing health programs with appropriate priorities. The literature also may guide public health professionals in carrying out health programs. Publishers ask the health professionals themselves to critically assess the articles submitted for possible inclusion in their journals. In order for an article to properly inform our health efforts and not be misleading, it must

provide valid information about medicine and public health. Finally, participating in or directing research requires that we understand what is currently known about a topic before we can advance knowledge on that topic.

A review involving 585 articles in the medical literature indicated that physicians with prior training in epidemiology and biostatistics are better prepared for reading medical literature.[36] Several studies have assessed the extensive use of study designs and statistical methods in medical and health-related journals. These studies have involved biomedical journals,[37] general medical journals,[38–40] rheumatology and internal medicine journals,[41] psychiatric journals,[42] family practice journals,[43] ophthalmic journals,[44] rehabilitation literature,[45] radiology journals,[46] public health journals,[47] surgical journals,[48] and health education journals.[49] A high percentage of articles in these journals involve study designs and statistical methods. For example, Merrill and colleagues assessed the use and types of study designs and the statistical methods employed in three representative health education journals from 1994 through 2003.[49] Editorials, commentaries, program/practice notes, and perspectives represented 17.6% of the journals' content. More commonly, the articles used cross-sectional designs (27.5%), reviews (23.2%), and analytic designs (i.e., case-control, cohort, and experimental studies) (18.4%). In addition, over the study period, the use of cross-sectional study designs increased 3.3% annually and the use of analytic study designs increased 5.5% annually, whereas the use of review articles decreased by 9.3% annually. The use of statistics in the articles surveyed also showed an increasing trend (e.g., 2.4% for descriptive statistics, 4.4% for parametric test statistics, 3.5% for nonparametric test statistics, 6.8% for generalized linear models, and 6.7% for validation statistics).

Although peer-reviewed journals are often rigorously reviewed prior to acceptance and publication, almost all reported studies have limitations. In 1992, a comprehensive report involving 28 papers that examined the scientific adequacy of study design, data collection, and statistical methods of 4,235 published medical studies found that only 20% met the researcher's criteria for validity.[50] They also found that poorly designed studies were more likely to report positive findings (i.e., 80% versus 25%). In 2002, a review of 34 commonly reviewed journals found that many journals failed to report how or whether their studies dealt with confounding factors.[51] Many studies have also been shown to suffer from various statistical problems in medical research.[52–56] As health professionals, it is ultimately our responsibility to conclude whether the results of a published paper are valid.

A basic knowledge of epidemiology and biostatistics makes the medical and health-related literature accessible, such that current knowledge can be gained and

applied. The content of this text will provide health practitioners with the skills they need to meet the analytic competencies listed in the outset of this chapter.

EXERCISES

1. *Health* has been defined as consisting of six interactive dimensions. Are there any of these six dimensions that are not important to your own health?
2. The 3 core public health functions involve assessment, policy development, and assurance. Relate these functions to the 10 essential public health services.
3. Define the terms *monitoring* and *diagnosing* and indicate how they are distinct and why they are important in public health.
4. It has been said that public health intervention focuses on prevention rather than treatment. What are some ways in which disease prevention may be achieved?
5. What are two ways in which epidemiology is distinct from biostatistics?
6. What is meant by the phrase "application of this study" in the definition of epidemiology?
7. Epidemiology is commonly referred to as the foundation of public health. Why?
8. With respect to treating a health problem, how would an epidemiology approach differ from a clinical approach?
9. Epidemiology involves more than just the study of disease. Explain.
10. Classify each of the following as (A) primary prevention, (B) secondary prevention, or (C) tertiary prevention.

    ___ vitamin-fortified foods
    ___ fluoridation of public water supplies
    ___ wearing protective devices
    ___ health promotion
    ___ cancer screening
    ___ lifestyle changes
    ___ physical therapy for stroke victims
    ___ halfway houses for recovering alcoholics
    ___ shelter homes for the developmentally challenged
    ___ fitness programs for heart attack patients
    ___ community health education
    ___ ensuring healthful conditions at home, school, and workplace

11. Self-efficacy is a concept of the Health Belief Model. Define this term and state where it fits in the six dimensions of health.

12. Classify the following as involving (A) direct transmission or (B) indirect transmission.

___ mosquito conveys the infectious agent
___ mucous membrane to mucous membrane (STDs)
___ herpes type 1 acquired from contact with an infected animal

13. In what way can a pathogen be indirectly transmitted from an infected person or animal?

14. List three common types of indirect transmission.

15. Zoonoses are diseases and infections transmitted between vertebrate animals and humans. Provide examples of direct and indirect transmission of a pathogen involving a vertebrate animal.

16. Which question does a case definition answer in descriptive epidemiology?

17. Classify each of the following as (A) descriptive epidemiology or (B) analytic epidemiology.

___ Identify the extent of a public health problem.
___ Identify the efficacy of a new drug.
___ Monitor the change in obesity for a given community over time.
___ Identify primary risk factors for disease.

18. Which descriptive study designs are observational?

19. Which analytic study design is *not* observational?

20. List the four general areas of biostatistics introduced in this chapter.

21. A primary purpose for taking a class that covers principles of epidemiology and biostatistics is to provide you with skills for reading the health literature. What are some of the ways this may be so?

REFERENCES

1. World Health Organization. Definition of health. World Health Organization. https://apps.who.int/aboutwho/en/definition.html. Accessed June 16, 2011.
2. *Dorland's Medical Dictionary for Health Consumers*. Philadelphia, PA: Saunders; 2007.
3. *The American Heritage Medical Dictionary*. Boston, MA: Houghton Mifflin Company; 2007.
4. Institute of Medicine. *The Future of Public Health*. Washington, DC: National Academy Press; 1988.

5. Centers for Disease Control and Prevention. 10 essential public health services. National Public Health Performance Standards Program (NPHPSP). http://www.cdc.gov/nphpsp/essentialServices.html. Published December 9, 2010. Accessed June 16, 2011.
6. U.S. Department of Health and Human Services Public Health Service. The public health workforce: an agenda for the 21st century. A report of the Public Health Functions Project. http://www.health.gov/phfunctions/pubhlth.pdf. Published 1998. Accessed June 14, 2011.
7. Hippocrates. Airs, waters, places. In Buck C, Llopis A, Najera E, Terris M, eds. *The Challenge of Epidemiology: Issues and Selected Readings*. Washington, DC: World Health Organization; 1988:18–19.
8. *Dorland's Illustrated Medical Dictionary.* 25th ed. Philadelphia, PA: Saunders; 1974.
9. Cumston CG. *An Introduction to the History of Medicine.* New York, NY: Alfred A. Knopf; 1926.
10. Garrison FH. *History of Medicine.* Philadelphia, PA: Saunders; 1926.
11. Last JM, ed. *A Dictionary of Epidemiology.* 3rd ed. New York, NY: Oxford University Press; 1995.
12. Crowley LV. *Introduction to Human Disease.* 2nd ed. Boston, MA: Jones and Bartlett; 1988.
13. U.S. National Institutes of Health. Understanding cancer. National Cancer Institute. http://www.cancer.gov/cancertopics/understandingcancer/cancer. Published 2010. Accessed June 28, 2011.
14. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst.* 1981;66:1191–1308.
15. Doll R. Epidemiological evidence of the effects of behavior and the environment on the risk of human cancer. *Recent Results Cancer Res.* 1998;154:3–21.
16. Merrill RM. *Introduction to Epidemiology.* 5th ed. Sudbury, MA: Jones and Bartlett Publishers; 2010.
17. Prochaska JO, DiClemente CC. Stages of change in the modification of problem behaviors. *Prog Behav Modif.* 1992;28:184–218.
18. Rosenstock IM. Why people use health services. *Milbank Mem Fund Q.* 1966;44:94–127.
19. Rosenstock IM. Historical origins of the health belief model. *Health Educ Q.* 1974;2:328–335.
20. Janz NK, Becker MH. The health belief model: a decade later. *Health Educ Q.* 1984;11:1–47.
21. Glanz K, Marcus Lewis F, Rimer BK. *Theory at a Glance: A Guide for Health Promotion Practice.* Bethesda, MD: National Institutes of Health; 1997.
22. Nester EW, McCarthy BJ, Roberts CE, Pearsall NN. *Microbiology: Molecules, Microbes and Man.* New York, NY: Holt, Rinehart and Winston; 1973.
23. Health News. *Medical Milestone: Mary Mallon, Typhoid Mary.* New York, NY: New York Department of Health; November 1968.
24. *Stedman's Medical Dictionary for the Health Professions and Nursing: Illustrated 5th Edition.* New York, NY: Lippincott Williams & Wilkins; 2005.
25. *McGraw-Hill Concise Dictionary of Modern Medicine.* New York, NY: McGraw-Hill Companies, Inc.; 2002.
26. Oleckno WA. *Essential Epidemiology: Principles and Application.* Long Grove, IL: Waveland Press; 2002.
27. Hald A. *A History of Mathematical Statistics.* New York, NY: Wiley; 1998.
28. Wright S. *Evolution and the Genetics of Populations: Genetics and Biometric Foundations.* Vol. 1. New ed. Chicago, IL: University of Chicago Press; 1984.
29. Wright S. *Evolution and the Genetics of Populations: Genetics and Biometric Foundations.* Vol. 2. New ed. Chicago, IL: University of Chicago Press; 1984.

30. Wright S. *Evolution and the Genetics of Populations: Genetics and Biometric Foundations.* Vol. 3. New ed. Chicago, IL: University of Chicago Press; 1984.
31. Wright S. *Evolution and the Genetics of Populations: Genetics and Biometric Foundations.* Vol. 4. New ed. Chicago, IL: University of Chicago Press; 1984.
32. Majumder PP. Haldane's contributions to biological research in India. *Resonance.* 1998;3:32–35.
33. Cook ET. *The Life of Florence Nightingale.* Vols. 1–2. London: Macmillan; 1913.
34. Bostridge M. *Florence Nightingale: The Making of an Icon.* New York, NY: Farrar, Straus and Giroux; 2008.
35. McDonald L, ed. *Florence Nightingale: An Introduction to Her Life and Family.* Collected Works of Florence Nightingale, vol. 1. Waterloo, Ontario: Wilfrid Laurier University Press; 2001.
36. Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. *J Med Educ.* 1980;55(8):692–697.
37. Pilcik T. Statistics in three biomedical journals. *Physiol Res.* 2003;51:39–43.
38. Wang Q, Zhang B. Research design and statistical methods in Chinese medical journals. *JAMA.* 1998;280:283–285.
39. Colditz GA, Emerson JD. The statistical content of published medical research: some implications for biomedical education. *Med Educ.* 1985;19:248–255.
40. Emerson JD, Coldtiz GA. Use of statistical analysis in the *New England Journal of Medicine. N Engl J Med.* 1983;309:709–713.
41. Cardiel MH, Goldsmith CH. Type of statistical techniques in rheumatology and internal medicine journals. *Revista de Investigación Clínica.* 1995;47:197–201.
42. Miettunen J, Nieminen P, Isohanni M. Statistical methodology in general psychiatric journals. *Nordic J Psychiatry.* 2002;56:223–228.
43. Fromm BS, Snyder VL. Research design and statistical procedures used in the *Journal of Family Practice. J Fam Pract.* 1986;23:565–566.
44. Juzych MS, Shin DH, Seyedsadr M, Siengner SW, Juzych LA. Statistical techniques in ophthalmic journals. *Arch Opthalmol.* 1992;110:1225–1229.
45. Schwartz SJ, Sturr M, Goldberg G. Statistical methods in rehabilitation literature: a survey of recent publications. *Arch Phys Med Rehab.* 1996;77:497–500.
46. Elster AD. Use of statistical analysis in the *AJR* and *Radiology*: frequency, methods, and subspecialty differences. *AJR Am Journal Roentgenol.* 1994;163:711–715.
47. Levy PS, Stolte K. Statistical methods in public health and epidemiology: a look at the recent past and projections for the next decade. *Stat Meth Med Res.* 2000;9:41–55.
48. Reznick RK, Dawson-Saunders E, Folse JR. A rationale for the teaching of statistics to surgical residents. *Surgery.* 1987;101:611–617.
49. Merrill RM, Lindsay CA, Shields EC, Stoddard J. Have the focus and sophistication of research in health education changed? *Health Educ Behav.* 2007;34(1):10–25.
50. Williamson JW, Goldschmidt PG, Colton T. The quality of medical literature: an analysis of validation assessments. In Bailar JC, Mosteller F, eds. *Medical Uses of Statistics.* Boston, MA: Massachusetts Medical Society; 1992.
51. Mullner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med.* 2002;136:122–126.

52. Williams JL, Hathaway CA, Koster KL, Layne BH. Low power, type II errors, and other statistical problems in recent cardiovascular research. *Am J Physiol.* 1997;273:H487–H493.
53. Hyran M. Appropriate analysis and presentation of data is a must for good clinical practice. *Acta Neurochir Suppl.* 2002;83:121–125
54. Skovlund E. A critical review of papers from clinical cancer research. *Acta Oncol.* 1998;36:339–345.
55. Vrbos LA, Lorenz MA, Peabody EH, McGregor M. Clinical methodologies and incidence of appropriate statistical testing in orthopaedic spine literature. Are statistics misleading? *Spine*. 1993;18(8):1021–1029.
56. McCance I. Assessment of statistical procedures used in papers in the *Aust Vet J*. 1995; 72(9):322–328.