Introduction to epidemiology

Aug 17 2011 Introduction to Epidemiology Epidemiology is considered the basic science of public health, and with good reason. Epidemiology is: • • • A quantitative basic science built on a working knowledge of probability, statistics, and sound research methodology A method of causal reasoning based on developing and testing hypotheses pertaining to occurrence and prevention of morbidity and mortality A tool for public health action to promote and protect the public's health based on science, causal reasoning, and a dose of practical common sense (2).

As a public health discipline, epidemiology is instilled with the spirit that epidemiologic information should be used to promote and protect the public's health. Hence, epidemiology involves both science and public health practice. The term applied epidemiology is sometimes used to describe the application or practice of epidemiology to address public health issues.

Examples of applied epidemiology include the following: • • • • the monitoring of reports of communicable diseases in the community the study of whether a particular dietary component influences your risk of developing cancer evaluation of the effectiveness and impact of a cholesterol awareness program analysis of historical trends and current data to project future public health resource needs Objectives

After studying this document and answering the questions in the exercises, you should be able to do the following: • • • • • • • • • • • Define epidemiology Summarize the historical evolution of epidemiology Describe the elements of a case definition and state the effect of changing the value of any of the elements List the key features and uses of descriptive epidemiology List the key features and uses of analytic epidemiology List the three components of the epidemiologic triad List and describe Hill's criteria of causation Understand the natural history of disease and the three types of prevention Understand infectivity, pathogenicity, and virulence List and describe primary applications of epidemiology in public health practice List and describe the different modes of transmission of communicable disease in a population 1 Page 2 Applied Epidemiology I A number of exercises are provided. It is suggested you attempt to answer these questions and then compare your answers with those at the end of this document. Introduction The word epidemiology comes from the Greek words epi, meaning " on or upon," demos, meaning " people," and logos, meaning " the study of. Many definitions have been proposed, but the following definition captures the underlying principles and the public health spirit of epidemiology: " Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems. " (17) Key terms in this definition reflect some of the important principles of epidemiology. Study Epidemiology is a scientific discipline with sound methods of scientific inquiry at its foundation. Epidemiology is data-driven and relies on a systematic and unbiased approach to the collection, analysis, and interpretation of data.

Basic epidemiologic methods tend to rely on careful observation and use of valid comparison groups to assess whether what was observed, such as the number of cases of disease in a particular area during a particular time period or the frequency of an exposure among persons with disease, differs from what might be expected. However, epidemiology also draws on methods from other scientific fields, including biostatistics and informatics, with biologic, economic, social, and behavioral sciences. In fact, epidemiology is often described as the basic science of public health, and for good reason. First, epidemiology is a quantitative discipline that relies on a working knowledge of probability, statistics, and sound research methods.

Second, epidemiology is a method of causal reasoning based on developing and testing hypotheses grounded in such scientific fields as biology, behavioral sciences, physics, and ergonomics to explain health-related behaviors, states, and events. However, epidemiology is not just a research activity but an integral component of public health, providing the foundation for directing practical and appropriate public health action based on this science and causal reasoning. Determinants Epidemiology is also used to search for determinants, which are the causes and other factors that influence the occurrence of disease and other health-related events.

Epidemiologists assume that illness does not occur randomly in a population, but happens only when the right accumulation of risk factors or determinants exists in an individual. To search for these determinants, epidemiologists use analytic epidemiology or epidemiologic studies to provide the " Why" and " How" of such events. They assess whether groups with different rates of disease differ in their demographic characteristics, genetic or immunologic make-up, behaviors, environmental exposures, or other so-called potential risk factors. Ideally, the findings provide sufficient evidence to direct prompt and effective public health control and prevention measures. Health-related states or events

Epidemiology was originally focused exclusively on epidemics of communicable diseases3 but was subsequently expanded to address endemic communicable diseases and non-communicable infectious diseases. By the middle of the 20th Century, additional epidemiologic methods had been developed and applied to chronic diseases, injuries, birth defects, maternal-child health, occupational health, and environmental health. Then epidemiologists began to look at behaviors related to health and well-being, such as amount of exercise and seat belt use. Now, with the recent explosion in molecular methods, Introduction to Epidemiology – Epi 592J Page 3 epidemiologists can make important strides in examining genetic markers of disease risk.

Indeed, the term health related states or events may be seen as anything that affects the well-being of a population. Nonetheless, many epidemiologists still use the term “ disease” as shorthand for the wide range of healthrelated states and events that are studied. Specified populations Although epidemiologists and direct health-care providers (clinicians) are both concerned with occurrence and control of disease, they differ greatly in how they view “ the patient. ” The clinician is concerned about the health of an individual; the epidemiologist is concerned about the collective health of the people in a community or population. In other words, the clinician’s “ patient” is the individual; the epidemiologist’s “ patient” is the community.

Therefore, the clinician and the epidemiologist have different responsibilities when faced with a person with illness. For example, when a patient with diarrheal disease presents, both are interested in establishing the correct diagnosis. However, while the clinician usually focuses on treating and caring for the individual, the epidemiologist focuses on identifying the exposure or source that caused the illness; the number of other persons who may have been similarly exposed; the potential for further spread in the community; and interventions to prevent additional cases or recurrences. Application Epidemiology is not just “ the study of” health in a population; it also involves applying the knowledge gained by the studies to community-based practice.

Like the practice of medicine, the practice of epidemiology is both a science and an art. To make the proper diagnosis and prescribe appropriate treatment for a patient, the clinician combines medical (scientific) knowledge with experience, clinical judgment, and understanding of the patient. Similarly, the epidemiologist uses the scientific methods of descriptive and analytic epidemiology as well as experience, epidemiologic judgment, and understanding of local conditions in “ diagnosing” the health of a community and proposing appropriate, practical, and acceptable public health interventions to control and prevent disease in the community. Summary

Epidemiology is the study (scientific, systematic, data-driven) of the distribution (frequency, pattern) and determinants (causes, risk factors) of health-related states and events (not just diseases) in specified populations (patient is community, individuals viewed collectively), and the application of (since epidemiology is a discipline within public health) this study to the control of health problems. Evolution Although epidemiologic thinking has been traced from Hippocrates (circa 400 B. C.) through Graunt (1662), Farr, Snow (both mid-1800’s), and others, the discipline did not blossom until the end of the Second World War. The contributions of some of these early and more recent thinkers are described next. Hippocrates (circa 400 B. C.) attempted to explain disease occurrence from a rational instead of a supernatural viewpoint. In his essay entitled “ On Airs, Waters, and Places,” Hippocrates suggested that environmental and host factors such as behaviors might influence the development of disease.

Another early contributor to epidemiology was John Graunt, a London haberdasher who published his landmark analysis of mortality data in 1662. He was the first to quantify patterns of birth, death, and disease occurrence, noting male-female disparities, high infant mortality, urban-rural differences, and seasonal variations. No one built upon Graunt’s work until the mid-1800, when William Farr began to systematically collect and analyze Britain’s mortality statistics. Farr, considered the father of modern vital statistics and disease surveillance, developed many of the basic practices used today in vital statistics and disease classification. He extended the epidemiologic analysis of morbidity and mortality data, looking at Page 4 Applied Epidemiology I he effects of marital status, occupation, and altitude. He also developed many epidemiologic concepts and techniques still in use today. Meanwhile, an anesthesiologist named John Snow was conducting a series of investigations in London that later earned him the title “ the father of epidemiology. ” Twenty years before the development of the microscope, Snow conducted studies of cholera outbreaks both to discover the cause of the disease and to prevent its recurrence. Because his work classically illustrates the sequence from descriptive epidemiology to hypothesis generation to hypothesis testing (analytic epidemiology) to application, we will consider two of his efforts.

It is important to mention that at the time of John Snow's investigations the most widely accepted cause of diseases, including cholera, was due to miasma, or foul air. Therefore most believed that cholera was transmitted by air, especially foul-smelling air near water. The germ theory, that disease was transmitted by microbes, did not gain acceptance until later in the 1800s. Snow conducted his classic study in 1854 when an epidemic of cholera developed in the Golden Square of London. He began his investigation by determining where in this area persons with cholera lived and worked. He then used this information to map the distribution of cases on what epidemiologists call a spot map. His map is shown in Figure 1. 1.

Because Snow believed that water was a source of infection for cholera, he marked the location of water pumps on his spot map, and then looked for a relationship between the distribution of cholera case households and the location of pumps. He noticed that more case households clustered around certain pumps, especially the Broad Street pump, and he concluded that the Broad Street pump was the most likely source of infection. Questioning residents who lived near the other pumps, he found that they avoided certain pumps because the water they provided was grossly contaminated, and that other pumps were located too inconveniently for most residents of the Golden Square area.

From this information, it appeared to Snow that the Broad Street pump was probably the primary source of water for most persons with cholera in the Golden Square area. He realized, however, that it was too soon to draw that conclusion because the map showed no cholera cases in a two-block area to the east of the Broad Street pump. Perhaps no one lived in that area, or perhaps the residents were somehow protected. Upon investigating, Snow found that a brewery was located there and that it had a deep well on the premises where brewery workers, who also lived in the area, got their water. In addition, the brewery allotted workers a daily quota of malt liquor. Access to these uncontaminated rations could explain why none of the brewery's employees contracted cholera.

To provide further evidence that the Broad Street pump was the source of the epidemic, Snow gathered information on where persons with cholera had obtained their water. Consumption of water from the Broad Street pump was the one common factor among the cholera patients. According to legend, Snow removed the handle of the Broad Street pump and aborted the outbreak. Snow's second major contribution involved another investigation of the same outbreak of cholera that occurred in London in 1854. In a London epidemic in 1849, Snow had noted that districts with the highest mortalities had water supplied by two companies: the Lambeth Company and the Southwark and Vauxhall Company. At that time, both companies obtained water from the Thames River, at intake points downstream of London.

In 1852, the Lambeth Company moved their water works upstream from London, thus obtaining water free of London sewage. When cholera returned to London in 1853, Snow realized the Lambeth Company's relocation of its intake point would allow him to compare districts that were supplied with water upstream from London with districts that received water downstream from London. Table 1. 1 shows what Snow found when he made that comparison for cholera mortality over a 7-week period during the summer of

1854. Introduction to Epidemiology – Epi 592J Page 5 Figure 1. 1 Distribution of cholera cases in the Golden Square area of London, August-September 1854 Table 1. Mortality from cholera in the districts of London supplied by the Southwark and Vauxhall and the Lambeth Companies, July 9-August 26, 1854 Districts with Water Supplied by Population Deaths from Mortality Risk per (1851 Census) Cholera 1, 000 Population 167, 654 844 5. 0 Southwark and Vauxhall Co. only Lambeth Co. only Both companies Source: 27 19, 133 300, 149 18 652 0. 9 2. 2 Page 6 Applied Epidemiology I The data in Table 1. 1 show that the risk of death from cholera was more than 5 times higher in districts served only by the Southwark and Vauxhall Company than in those served only by the Lambeth Company. Interestingly, the mortality risks in districts supplied by both companies fell between the risks for districts served exclusively by either company.

These data were consistent with the hypothesis that water obtained from the Thames below London was a source of cholera. Alternatively, the populations supplied by the two companies may have differed on a number of other factors which affected their risk of cholera. To test his water supply hypothesis, Snow focused on the districts served by both companies, because the households within a district were generally comparable except for which company supplied water. In these districts, Snow identified the water supply company for every house in which a death from cholera had occurred during the 7-week period. Table 1. 2 shows his findings. Table 1. Mortality from cholera in London related to the water supply of individual houses in districts served by both the Southwark and Vauxhall Company and the Lambeth Company, July 9August 26, 1854 Water Supply of Individual

House Population Deaths from Mortality risk per (1851 Census) Cholera 1, 000 Population Southwark and Vauxhall Co. 98, 862 419 4. 2 Lambeth Co. Source: 27 154, 615 80 0. 5 This further study added support to Snow's hypothesis, and demonstrates the sequence of steps used today to investigate outbreaks of disease. Based on a characterization of the cases and population at risk by time, place, and person, Snow developed a testable hypothesis. He then tested this hypothesis with a more rigorously designed study, ensuring that the groups to be compared were comparable. After this study, efforts to control the epidemic were directed at changing the location of the water intake of the Southwark and Vauxhall Company to reduce sources of contamination.

Thus, with no knowledge of the existence of microorganisms, Snow demonstrated through epidemiologic studies that water could serve as a vehicle for transmitting cholera and that epidemiologic information could be used to direct prompt and appropriate public health action. More information on John Snow can be found at: www. ph. ucla. edu/epi/snow. html In the mid- and late-1800's, many others in Europe and the United States began to apply epidemiologic methods to investigate disease occurrence. At that time, most investigators focused on acute infectious diseases. In the 1900's, epidemiologists extended their methods to noninfectious diseases.

The period since the Second World War has seen an explosion in the development of research methods and the theoretical underpinnings of epidemiology, and in the application of epidemiology to the entire range of health-related outcomes, behaviors, and even knowledge and attitudes. The studies by Doll and Hill (13) linking smoking to lung cancer and the study of cardiovascular disease among residents of Framingham, Massachusetts (12), are two examples of how pioneering researchers have applied epidemiologic methods to chronic disease since World War II. Finally, during the 1960’s and early 1970’s health workers applied epidemiologic methods to eradicate smallpox worldwide.

This was an achievement in applied epidemiology of unprecedented proportions. Today, public health workers throughout the world accept and use epidemiology routinely. Epidemiology is often practiced or used by non-epidemiologists to characterize the health of their communities and to solve day-to-day problems. This landmark in the evolution of the discipline is less dramatic than the eradication of smallpox, but it is no less important in improving the health of people everywhere. Introduction to Epidemiology – Epi 592J Page 7 Uses Epidemiology and the information generated by epidemiologic methods have many uses. These uses are categorized and described below. Population or community health assessment.

To set policy and plan programs, public health officials must assess the health of the population or community they serve and determine whether health services are available, accessible, effective, and efficient. To do this, they must find answers to many questions: What are the actual and potential health problems in the community? Where are they? Who is at risk? Which problems are declining over time? Which ones are increasing or have the potential to increase? How do these patterns relate to the level and distribution of services available? The methods of descriptive and analytic epidemiology provide ways to answer these and other questions.

With answers provided through the application of epidemiology, the officials can make informed decisions that will lead to improved health for the population they serve. Individual decisions. People may not realize that they use epidemiologic information in their daily decisions. When they decide to stop smoking, take the stairs instead of the elevator, order a salad instead of a cheeseburger with French fries, or choose one method of contraception instead of another, they may be influenced, consciously or unconsciously, by epidemiologists' assessment of risk. Since World War II, epidemiologists have provided information related to all those decisions.

In the 1950's, epidemiologists documented the increased risk of lung cancer among smokers; in the 1960's and 1970's, epidemiologists noted a variety of benefits and risks associated with different methods of birth control; in the mid-1980's, epidemiologists identified the increased risk of human immunodeficiency virus (HIV) infection associated with certain sexual and drug-related behaviors; and, more positively, epidemiologists continue to document the role of exercise and proper diet in reducing the risk of heart disease. These and hundreds of other epidemiologic findings are directly relevant to the choices that people make every day, choices that affect their health over a lifetime. Completing the clinical picture. When studying a disease outbreak, epidemiologists depend on clinical physicians and laboratory scientists for the proper diagnosis of individual patients.

But epidemiologists also contribute to physicians' understanding of the clinical picture and natural history of disease. For example, in late 1989 three patients in New Mexico were diagnosed as having myalgias (severe muscle pains in chest or abdomen) and unexplained eosinophilia (an increase in the number of one type of white blood cell). Their physicians could not identify the cause of their symptoms, or put a name to the disorder. Epidemiologists began looking for other cases with similar symptoms, and within weeks had found enough additional cases of eosinophilia-myalgia syndrome (EMS) to describe the illness, its complications, and its risk of mortality.

Similarly, epidemiologists have documented the course of HIV infection, from the initial exposure to the development of a wide variety of clinical syndromes that include acquired immunodeficiency syndrome (AIDS). They have also documented the numerous conditions associated with cigarette smoking—from pulmonary and heart disease to lung and cervical cancer. Search for causes. Much of epidemiologic research is devoted to a search for causes, factors which influence one's risk of disease. Sometimes this is an academic pursuit, but more often the goal is to identify a cause so that appropriate public health action might be taken. It has been said that epidemiology can never prove a causal relationship between an exposure and a disease. Nevertheless, epidemiology often provides enough information to support effective action.

Examples include John Snow's removal of the pump handle and the withdrawal of a specific brand of tampon that was linked by epidemiologists to toxic shock syndrome. Another example is the recommendation that children not be given aspirin due to its association with Reye syndrome. Just as often, epidemiology and laboratory science converge to provide the evidence needed to establish causation. For example, a team of epidemiologists were able to identify a variety of risk factors during an outbreak of pneumonia among persons attending the American Page 8 Applied Epidemiology I Legion Convention in Philadelphia in 1976, called " Legionnaire's disease. However, the outbreak was not " solved" until the Legionnaires' bacillus was identified in the laboratory almost 6 months later. Disease control, elimination, and eradication. The ultimate goal of epidemiology is to improve the health of populations and through the reduction in disease. The definitions of disease control, elimination, and eradication as applied to infectious diseases are given below. (Dowdle WR. The principles of disease elimination and eradication. MMWR 48(SU01); 23-7, 1999.): Control: The reduction of disease incidence, prevalence, morbidity or mortality to a locally acceptable level as a result of deliberate efforts; continued intervention measures are required to maintain the reduction. Example: diarrheal diseases.

Elimination of disease: Reduction to zero of the incidence of a specified disease in a defined geographical area as a result of deliberate efforts; continued intervention measures are required. Examples: neonatal tetanus. Elimination of infections: Reduction to zero of the incidence of infection caused by a specific agent in a defined geographical area as a result of deliberate efforts; continued measures to prevent reestablishment of transmission are required. Example: measles, poliomyelitis. Eradication: Permanent reduction to zero of the worldwide incidence of infection caused by a specific agent as a result of deliberate efforts; intervention measures are no longer needed. Example: smallpox.

Extinction: The specific infectious agent no longer exists in nature or in the laboratory. Example: none. The above definitions are specific to infectious disease, but some of the concepts can carry over to other conditions, such as nutritional disorders, inborn errors of metabolism, and chronic diseases. Introduction to Epidemiology – Epi 592J Page 9 Exercise 1. 1 In the early 1980's, epidemiologists recognized that AIDS occurred most frequently in men who had sex with men and in intravenous drug users. Describe how this information might be used for each of the following: a. Population or community health assessment b. Individual decisions c. Search for causes Page 10 Applied Epidemiology I The Epidemiologic Approach

Like a newspaper reporter, an epidemiologist determines What, When, Where, Who, and Why. However, the epidemiologist is more likely to describe these concepts in slightly different terms: case definition, time, place, person, and causes. Case Definition (“ What? ”) The identification of disease can be based on symptoms, signs, and diagnostic tests. A symptom is a sensation or change in health experienced by an individual. Examples of symptoms reported by an individual are a cough, fatigue, anxiety, and back pain. Signs, or signs of disease, are an objective evidence of disease observed by someone other than the affected individual, such as a physician or nurse.

A case definition is a set of standard criteria for deciding whether a person has a particular disease or other health-related condition. By using a standard case definition we attempt to ensure that every case is diagnosed in the same way, regardless of when or where it occurred, or who identified it. We can then compare the number of cases of the disease that occurred in one time or place with the number that occurred at another time or another place. For example, with a standard case definition, we can compare the number of cases of hepatitis A that occurred in New York City in 1991 with the number that occurred there in 1990. Or we can compare the number of cases that occurred in New York in 1991 with the number that occurred in San Francisco in 1991. With a standard ase definition, when we find a difference in disease occurrence, we know it is likely to be due to a real difference or due to the quality of the disease reporting system rather than the result of differences in how cases were diagnosed. A case definition consists of clinical criteria and, sometimes, limitations on time, place, and person. The clinical criteria usually include confirmatory laboratory tests, if available, or combinations of symptoms (subjective complaints), signs (objective physical findings), and other findings. For example, see the case definition for rabies below; notice that it requires laboratory confirmation. Rabies, Human Clinical description Rabies is an acute encephalomyelitis that almost always progresses to coma or death within 10 days of the first symptom.

Laboratory criteria for diagnosis • Detection by direct fluorescent antibody of viral antigens in a clinical specimen (preferably the brain or the nerves surrounding hair follicles in the nape of the neck), or • Isolation (in cell culture or in a laboratory animal) of rabies virus from saliva, cerebrospinal fluid (CSF), or central nervous system tissue, or • Identification of a rabies-neutralizing antibody titer greater than or equal to 5 (complete neutralization) in the serum or CSF of an unvaccinated person Case classification Confirmed: a clinically compatible illness that is laboratory confirmed Comment Laboratory confirmation by all of the above methods is strongly recommended. Source: 3 Compare this with the case definition for

Kawasaki syndrome provided in Exercise 1. 3 on page 12. Kawasaki syndrome is a childhood illness with fever and rash that has no known cause and no specifically distinctive laboratory findings.

Notice that its case definition is based on the presence of fever, at least four of five specified clinical findings, and the lack of a more reasonable explanation. A case definition may have several sets of criteria, depending on the certainty of the diagnosis. For example, during an outbreak of measles, we might classify a person with a fever and rash as having a Introduction to Epidemiology – Epi 592J Page 11 suspect, probable, or confirmed case of measles, depending on what additional evidence of measles was present. In other situations, we may temporarily classify a case as suspect or probable until laboratory results are available. When we receive the laboratory report, we then reclassify the case as either confirmed or “ not a case,” depending on the lab results.

In the midst of a large outbreak of a disease caused by a known agent, we may permanently classify some cases as suspect or probable, because it is unnecessary and wasteful to run laboratory tests on every individual with a consistent clinical picture and a history of exposure (e. g. , chickenpox). Case definitions may also vary according to the purpose for classifying the occurrences of a disease. For example, health officials need to know as soon as possible if anyone has symptoms of plague or foodborne botulism so that they can begin planning what actions to take. For such rare but potentially severe diseases, where it is important to identify every possible case, health officials use a sensitive, or “ loose” case definition.

On the other hand, investigators of the causes of a disease outbreak want to be certain that any person included in the investigation really had the disease. The investigator will prefer a specific or " strict" case definition. For instance, in an outbreak of Salmonella agona, the investigators would be more likely to identify the source of the infection if they included only persons who were confirmed to have been infected with that organism, rather than including anyone with acute diarrhea, because some persons may have had diarrhea from a different cause. In this setting, a disadvantage of a strict case definition is an underestimate of the total number of cases. Exercise 1. 2

In the case definition for an apparent outbreak of trichinosis, investigators used the following classifications: Clinical criteria Confirmed case: signs and symptoms plus laboratory confirmation Probable case: acute onset of at least three of the following four features: myalgia, fever, facial edema, or eosinophil count greater than 500/mm3 Possible case: acute onset of two of the above four features plus a physician diagnosis of trichinosis Suspect case: unexplained eosinophilia Not a case: failure to fulfill the criteria for a confirmed, probable, possible, or suspect case Time Onset after October 26, 1991 Place Metropolitan Atlanta Person Any Assign the appropriate classification to each of the persons included in the line listing below. (All were residents of Atlanta with acute onset of symptoms in November.) Page 12 Applied Epidemiology I ID # 1 2 3 4 5 Last name Abels Baker Corey Dale Ring myalgia yes yes yes yes yes fever yes yes yes no no facial edema no yes no no no eosinophil count 495 pending 1, 100 2, 050 600 Physician diagnosis trichinosis trichinosis ? trichinosis EMS ? trichinosis Lab confirm yes pending pending pending not done Classification __________ __________ __________ __________ __________

Exercise 1. 3 The following is the official case definition for Kawasaki syndrome that is recommended by CDC: Kawasaki Syndrome Clinical case definition A febrile illness of greater than or equal to 5 days' duration, with at least four of the five following physical findings and no other more reasonable explanation for the observed clinical findings: • Bilateral conjunctival injection • Oral changes (erythema of lips or oropharynx, strawberry tongue, or fissuring of the lips) • Peripheral extremity changes (edema, erythema, or generalized or periungual desquamation) • Rash • Cervical lymphadenopathy (at least one lymph node greater than or equal to 1. cm in diameter) Laboratory criteria for diagnosis None Case classification Confirmed: a case that meets the clinical case definition Comment If fever disappears after intravenous gamma globulin therapy is started, fever may be of less than 5 days' duration, and the clinical case definition may still be met. Source: 3 Discuss the pros and cons of this case definition for the purposes listed below. (For a brief description of Kawasaki syndrome, see Benenson's Control of Communicable Diseases in Man). a. Diagnosing and treating individual patients b. Tracking the occurrence of the disease for public health records c. Doing research to identify the cause of the disease Introduction to Epidemiology – Epi 592J Page 13 Numbers and Risks

A basic task of a health department is counting cases in order to measure and describe morbidity. When physicians diagnose a case of a reportable disease they are suppose to report the case to their local health department. For most reportable conditions, these reports are legally required to contain information on time (when the case occurred), place (where the patient lived), and person (the age, race, and sex of the patient). The health department combines all reports and summarizes the information by time, place, and person. From these summaries, the health department determines the extent and patterns of disease occurrence in the area, and attempts to identify clusters or outbreaks of disease.

A simple count of cases, however, does not provide all the information a health department needs. To compare the occurrence of a disease at different locations, during different times, or in different subgroups, a health department converts the case counts into risks, which relates the number of cases to the size of the population. Risks are useful in many ways. With risks, the health department can identify groups in the community with an elevated risk of disease. These so-called high-risk groups can be further assessed and targeted for special intervention; the groups can be studied to identify risk factors that are related to the occurrence of disease.

Individuals can use knowledge of these risk factors to guide their decisions about behaviors that influence health. Descriptive Epidemiology In descriptive epidemiology, we organize and summarize data according to time, place, and person. These three characteristics are sometimes called the epidemiologic variables. Compiling and analyzing data by time, place, and person is desirable for several reasons. First, the investigator becomes intimately familiar with the data and with the extent of the public health problem being investigated. Second, this provides a detailed description of the health of a population that is easily communicated. Third, such analysis identifies the populations at greatest risk of acquiring a particular disease.

This information provides important clues to the causes of the disease, and these clues can be turned into testable hypotheses. Time (“ When? ”) Disease risks usually change over time. Some of these changes occur regularly and can be predicted. For example, the seasonal increase of influenza cases with the onset of cold weather is a pattern that is familiar to everyone. By knowing when flu outbreaks will occur, health departments can time their influenza vaccination campaigns effectively. Other diseases may make unpredictable changes in occurrence. By examining events that precede a disease increase or decrease, we may identify causes and appropriate actions to control or prevent further occurrence of the disease.

We usually show time data as a graph (Figure 1. 3). We put the number or risk of cases or deaths on the vertical, y-axis; we put the time periods along the horizontal, x-axis. We often indicate on a graph when events occurred that we believe are related to the particular health problem described in the graph. For example, we may indicate the period of exposure or the date control measures were implemented. Such a graph provides a simple visual depiction of the relative size of a problem, its past trend and potential future course, as well as how other events may have affected the problem. Studying such a graph often gives us insights into what may have caused the problem.

Depending on what event we are describing, we may be interested in a period of years or decades, or we may limit the period to hours, days, weeks, or months when the number of cases reported is greater than normal (an epidemic period). For some conditions—for many chronic diseases, for example—we are interested in long-term changes in the number of cases or risk of the condition. For other conditions, we may find it more revealing to look at the occurrence of the condition by season, month, day of the Page 14 Applied Epidemiology I week, or even time of day. For a newly recognized problem, we need to assess the occurrence of the problem over time in a variety of ways until we discover the most appropriate and revealing time period to use. Some of the common types of time-related graphs are further described below. Secular (long-term) trends.

Graphing the annual cases or risk of a disease over a period of years shows long-term or secular trends in the occurrence of the disease. We commonly use these trends to suggest or predict the future incidence of a disease. We also use them in some instances to evaluate programs or policy decisions, or to suggest what caused an increase or decrease in the occurrence of a disease, particularly if the graph indicates when related events took place, as depicted in Figure 1. 3 (note the scale of the y-axis). Figure 1. 3 Malaria by year, United States, 1930-1990 Works Progress Administration Malaria Control Drainage Program Relapses from Overseas Cases 1000 Reported Cases per 100, 000 Population 100

Relapses from Korean Veterans Returning Vietnam Veterans 10 Foreign Immigration 1 0. 1 0. 01 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 Source: 9 Year Seasonality. By graphing the occurrence of a disease by week or month over the course of a year or more we can show its seasonal pattern, if any. Some diseases are known to have characteristic seasonal distributions; for example, as mentioned earlier, the number of reported cases of influenza typically increases in winter. Seasonal patterns may suggest hypotheses about how the infection is transmitted, which behavioral factors increase risk, and other possible contributors to the disease or condition.

The seasonal pattern of an unknown disease is shown in Figure 1. 4. What factors might contribute to its seasonal pattern? From only the single year's data in Figure 1. 4, it is difficult to conclude whether the peak in June represents a characteristic seasonal pattern that would be repeated yearly, or whether it is simply an epidemic that occurred in the spring and summer of that particular year. You would need more than one year's data before you could conclude that the pattern shown there represents the seasonal variation in this disease. Introduction to Epidemiology – Epi 592J Page 15 Figure 1. 4 Cases of an unknown disease by month of onset 450 400 350 300 Cases 50 200 150 100 50 0 Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Source: 14 Month of Onset Day of week and time of day. Displaying data by days of the week or time of day may also be informative. Analysis at these shorter time periods is especially important for conditions that are potentially related to occupational or environmental exposures, which may occur at regularly scheduled intervals. In Figure 1. 5, farm tractor fatalities are displayed by days of the week. Does this analysis at shorter time periods suggest any hypothesis? In Figure 1. 5 the number of farm tractor fatalities on Sundays is about half the number on the other days. We can only speculate why this is.

One reasonable hypothesis is that farmers spend fewer hours on their tractors on Sundays than on the other days. Figure 1. 5 Fatalities associated with farm tractor injuries by day of death, Georgia, 1971-1981 Source: 15 Page 16 Applied Epidemiology I Examine the pattern of fatalities associated with farm tractor injuries by hour in Figure 1. 6. How might you explain the morning peak at 11: 00 AM, the dip at noon, and the afternoon peak at 4: 00 PM? Figure 1. 6 Fatalities associated with farm tractor injuries by time of day, Georgia, 1971-1981 Source: 15 Epidemic period. To show the time course of a disease outbreak or epidemic, we use a graph called an epidemic curve.

As with the other graphs you have seen in this section, we place the number of cases on the vertical axis and time on the horizontal axis. For time, we use either the time of onset of symptoms or the date of diagnosis. For very acute diseases with short incubation periods (i. e. , time period between exposure and onset of symptoms is short), we may show time as the hour of onset. For diseases with longer incubation periods, we might show time in 1-day, 2-day, 3-day, 1-week, or other appropriate intervals. Figure 1. 7 shows an epidemic curve that uses a 3-day interval for a foodborne disease outbreak. Notice how the cases are stacked in adjoining columns. By convention, we use this format, called a histogram, for epidemic curves.

The shape and other features of an epidemic curve can suggest hypotheses about the time and source of exposure, the mode of transmission, and the causative agent. Figure 1. 7 Date of onset of illness in patients with culture-confirmed Yersinia enterocolitica infections, Atlanta, November 1, 1988-January 10, 1989 8 7 6 Thanksgiving Christmas New Year’s Cases 5 4 3 2 1 0 1 4 7 10 13 16 19 22 25 28 1 4 7 10 13 16 19 22 25 28 1 4 7 10 November December January Source: 18 Date of Onset Introduction to Epidemiology – Epi 592J Page 17 Place (“ Where? ”) We describe a health event by place to gain insight into the geographical extent of the problem. For place, we may use place of residence, birthplace, employment, school district, hospital unit, etc. , epending on which may be related to the occurrence of the health event. Similarly, we may use large or small geographic units: country, state, county, census tract, street address, map coordinates, or some other geographical designation. Sometimes, we may find it useful to analyze data according to place categories such as urban or rural, domestic or foreign, and institutional or noninstitutional. Not all analyses by place will be equally informative. For example, examine the data shown in Table 1. 3. Where were the malaria cases diagnosed? What “ place” does the table break the data down by? Would it have been more or less useful to analyze the data according to the “ state of residence” of the cases?

We believe that it provides more useful information to show the data in Table 1. 3 by where the infection was acquired than it would have to show where the case-patients lived. By analyzing the malaria cases by place of acquisition, we can see where most of the malaria cases acquired their disease. Table 1. 3 Malaria cases by distribution of Plasmodium species and area of acquisition, United States, 1989 Species Area of Acquisition Vivax Falciparum Other Total Africa 52 382 64 498 Asia 207 44 29 280 Central America & Caribbean 107 14 9 130 North America 131 3 13 147 (United States) (5) (0) (0) (5) South America 10 1 2 13 Oceania 19 2 5 26 Unknown 6 2 0 8 Total 532 448 122 1, 102 Source: 6

By analyzing data by place, we can also get an idea of where the agent that causes a disease normally lives and multiplies, what may carry or transmit it, and how it spreads. When we find that the occurrence of a disease is associated with a place, we can infer that factors that increase the risk of the disease are present either in the persons living there (host factors) or in the environment, or both. For example, diseases that are passed from one person to another tend to spread more rapidly in urban areas than in rural ones, mainly because the greater crowding in urban areas provides more opportunities for susceptible people to come into contact with someone who is infected.

On the other hand, diseases that are passed from animals to humans often occur in greater numbers in rural and suburban areas because people in those areas are more likely to come into contact with disease-carrying animals, ticks, and the like. For example, perhaps Lyme disease has become more common because people have moved to wooded areas where they come into contact with infected deer ticks. Although we can show data by place in a table—as Table 1. 3 does—it is often better to show it pictorially in a map. On a map, we can use different shadings, color, or line patterns to indicate how a disease or health event has different numbers or risks of occurrence in different areas, as in Figure 1. 8. Page 18 Applied Epidemiology I Figure 1. 8 AIDS cases per 100, 000 population, United States, July 1991-June 1992 Source: 4

For a rare disease or outbreak, we often find it useful to prepare a spot map, like Snow's map of the Golden Square of London (Figure 1. 1), in which we mark with a dot or an X the relation of each case to a place that is potentially relevant to the health event being investigated—such as where each case lived or worked. We may also label other sites on a spot map, such as where we believe cases may have been exposed, to show the orientation of cases within the area mapped. Figure 1. 9 is a spot map for an outbreak of mumps that occurred among employees of the Chicago futures exchanges. Study the location of each case in relation to other cases and to the trading pits. The four numbered areas delineated with heavy lines are the trading pits.

Does the location of cases on the spot map lead you to any hypothesis about the source of infection? Figure 1. 9 Mumps cases in trading pits of exchange A, Chicago, Illinois, August 18-December 25, 1987 #1 #3 #2 #4 Key: Pit areas are numbered and delineated by heavy lines. Individual trading pits within pit areas are outlined by light lines. Affected person (N= 43) Desk areas Source: CDC, unpublished data, 1988 Introduction to Epidemiology – Epi 592J Page 19 You probably observed that the cases occurred primarily among those working in trading pits #3 and #4. This clustering of illness within trading pits provides indirect evidence that the mumps was transmitted person-to person. Person (“ Who?) In descriptive epidemiology, when we organize or analyze data by “ person” there are several person categories available to us. We may use inherent characteristics of people (for example, age, race, sex), their acquired characteristics (immune or marital status), their activities (occupation, leisure activities, use of medications/tobacco/drugs), or the conditions under which they live (socioeconomic status, access to medical care). These categories usually determine, to a large degree, who is at greatest risk of experiencing certain undesirable health conditions, such as becoming infected with a particular disease organism. We may show person-related characteristics in either tables or graphs.

In analyzing data by person, we often must try a number of different categories before we find which are the most useful and enlightening. Age and sex are most critical; we almost always analyze data according to these.

Depending on the health event we are studying, we may or may not break the data down by other attributes. Often we analyze data by more than one characteristic simultaneously; for example, we may look at age and sex simultaneously to see if the sexes differ in how they develop a condition that increases with age—such as with heart disease. Age. Age is probably the single most important “ person” attribute, because almost every health-related event or state varies with age.

A number of factors that also vary with age are behind this association: susceptibility, opportunity for exposure, latency or incubation period of the disease, and physiologic response (which affects, among other things, disease development). When we analyze data by age, we try to use age groups that are narrow enough to detect any agerelated patterns that may be present in the data. In an initial breakdown by age, we commonly use 5-year age intervals: 0 to 4 years, 5 to 9, 10 to 14, and so on. Larger intervals, such as 0 to 19 years, 20 to 39, etc. , may conceal variations related to age which we need to know to identify the true ages at greatest risk.

Sometimes, even 5-year age groups can hide important differences, especially in children less than five years of age. Take time to examine Figure 1. 10, for example, before you read ahead. What does the information in this figure suggest health authorities should do to reduce the number of cases of whooping cough? Where should health authorities focus their efforts? You probably said that health authorities should focus on immunizing infants against whooping cough during the first year of life. Now, examine Figure 1. 11. This figure shows the same data but they are presented in the usual 5-year intervals. Based on Figure 1. 11 where would you have suggested that health authorities focus their efforts?

Would this recommendation have been as effective and efficient in reducing cases of whooping cough? You probably said that health authorities should immunize infants and children before the age of 5. That recommendation would be effective, but it would not be efficient. You would be immunizing more children than actually necessary and wasting resources. Sex. In general, males have higher risks of illness and death than females do for a wide range of diseases. For some diseases, this sex-related difference is because of genetic, hormonal, anatomic, or other inherent differences between the sexes. These inherent differences affect their susceptibility or physiologic responses.

For example, premenopausal women have a lower risk of heart disease than men of the same age. This difference is attributed to higher estrogen levels in women. On the other hand, the sex-related differences in the occurrence of many diseases reflect differences in opportunity or levels of exposure. For example, Figure 1. 12 shows that hand/wrist disorders occur almost twice as often in females than in males. What are some sex-related differences that would cause a higher level of this disorder in females? Page 20 Applied Epidemiology I Figure 1. 10 Pertussis (whooping cough) incidence by age group, United States, 1989 Source: 9 Figure 1. 11 Pertussis (whooping cough) incidence by age group, United States, 1989 Source: 9 Figure 1. 2 Prevalence of hand/wrist cumulative trauma disorder by sex, Newspaper Company A, 1990 Source: NIOSH, unpublished data, 1991 Introduction to Epidemiology – Epi 592J Page 21 You may have attributed the higher level of disorders in females to their higher level of exposure to occupational activities that require repetitive hand/wrist motion such as typing or keyboard entry. With occupationally-related illness, we usually find that sex differences reflect the number of workers in those occupations. You may also have attributed the higher level of disorders in females to anatomical differences; perhaps women’s wrists are more susceptible to hand/wrist disorders. Ethnic and racial groups.

In examining epidemiologic data, we are interested in any group of people who have lived together long enough to acquire common characteristics, either biologically or socially. Several terms are commonly used to identify such groups: race, nationality, religion, or local reproductive or social groups, such as tribes and other geographically or socially isolated groups. Differences that we observe in racial, ethnic, or other groups may reflect differences in their susceptibility or in their exposure, or they may reflect differences in other factors that bear more directly on the risk of disease, such as socioeconomic status and access to health care. In Figure 1. 13, the risks of suicide for five groups of people are displayed. Figure 1. 3 Suicide death rates for persons 15 to 24 years of age according to race/ethnicity, United States, 1988 Source: 22 Clearly this graph displays a range of suicide death rates for the five groups of people. These data provide direction for prevention programs and for future studies to explain the differences. Socioeconomic status. Socioeconomic status is difficult to quantify. It is made up of many variables such as occupation, family income, educational achievement, living conditions, and social standing. The variables that are easiest to measure may not reflect the overall concept. Nevertheless, we commonly use occupation, family income, and educational achievement, while recognizing that these do not measure socioeconomic status precisely.

The frequency of many adverse health conditions increases with decreasing socioeconomic status. For example, tuberculosis is more common among persons in lower socioeconomic strata. Infant mortality and time lost from work due to disability are both associated with lower income. These patterns may reflect more harmful exposures, lower resistance, and less access to health care. Or they may in part Page 22 Applied Epidemiology I reflect an interdependent relationship which is impossible to untangle—does low socioeconomic status contribute to disability or does disability contribute to lower socioeconomic status? Some adverse health conditions are more frequent among persons of higher socioeconomic status.

These conditions include breast cancer, Kawasaki syndrome, and tennis elbow. Again, differences in exposure account for at least some of the differences in the frequency of these conditions. Exercise 1. 4 The following series of tables (Exercise 1. 4, Tables 1-4) show person information about cases of the unknown disease described in Figure 1. 4 on page 15. Look again at Figure 1. 4, study the information in the four exercise tables and then describe in words how the disease outbreak is distributed by time and person. Exercise 1. 4, Table 1 Incidence of the disease by age and sex in 24 villages surveyed for one year Males Females Age Group Population* # Cases Risk per Population* # Cases Risk per (years) 1, 000 1, 000