Analytic Epidemiology (1.7)

As noted earlier, descriptive epidemiology can identify patterns among cases and in populations by time, place and person. From these observations, epidemiologists develop hypotheses about the causes of these patterns and about the factors that increase risk of disease. In other words, epidemiologists can use descriptive epidemiology to generate hypotheses, but only rarely to test those hypotheses. For that, epidemiologists must turn to analytic epidemiology.

The key feature of analytic epidemiology is a comparison group. Consider a large outbreak of hepatitis A that occurred in Pennsylvania in 2003. Investigators found almost all of the case-patients had eaten at a particular restaurant during the 2–6 weeks (i.e., the typical incubation period for hepatitis A) before onset of illness. While the investigators were able to narrow down their hypotheses to the restaurant and were able to exclude the food preparers and servers as the source, they did not know which particular food may have been contaminated. The investigators asked the case-patients which restaurant foods they had eaten, but that only indicated which foods were popular. The investigators, therefore, also enrolled and interviewed a comparison or control group — a group of persons who had eaten at the restaurant during the same period but who did not get sick. Of 133 items on the restaurant's menu, the most striking difference between the case and control groups was in the proportion that ate salsa (94% of case-patients ate, compared with 39% of controls). Further investigation of the ingredients in the salsa implicated green onions as the source of infection. Shortly thereafter, the Food and Drug Administration issued an advisory to the public about green onions and risk of hepatitis A. This action was in direct response to the convincing results of the analytic epidemiology, which compared the exposure history of case-patients with that of an appropriate comparison group.

When investigators find that persons with a particular characteristic are more likely than those without the characteristic to contract a disease, the characteristic is said to be associated with the disease. The characteristic may be a:

- Demographic factor such as age, race, or sex;
- Constitutional factor such as blood group or immune status;
- Behavior or act such as smoking or having eaten salsa; or
- Circumstance such as living near a toxic waste site.

Identifying factors associated with disease help health officials appropriately target public health prevention and control activities. It also guides additional research into the causes of disease.

Thus, analytic epidemiology is concerned with the search for causes and effects, or the why and the how. Epidemiologists use analytic epidemiology to quantify the association between exposures and outcomes and to test hypotheses about causal relationships. It has been said that epidemiology by itself can never prove that a particular exposure caused a particular outcome. Often, however, epidemiology provides sufficient evidence to take appropriate control and prevention measures.

Epidemiologic studies fall into two categories: **experimental** and **observational**.

Experimental studies

In an experimental study, the investigator determines through a controlled process the exposure for each individual (clinical trial) or community (community trial), and then tracks the individuals or communities over time to detect the effects of the exposure. For example, in a clinical trial of a new vaccine, the investigator may randomly assign some of the participants to receive the new vaccine, while others receive a placebo shot. The investigator then tracks all participants, observes who gets the disease that the new vaccine is intended to prevent, and compares the two groups (new vaccine vs. placebo) to see whether the vaccine group has a lower rate of disease. Similarly, in a trial to prevent onset of diabetes among high-risk individuals, investigators randomly assigned enrollees to one of three groups — placebo, an anti-diabetes drug, or lifestyle intervention. At the end of the follow-up period, investigators found the lowest incidence of diabetes in the lifestyle intervention group, the next lowest in the anti-diabetic drug group, and the highest in the placebo group.

Observational studies

In an observational study, the epidemiologist simply observes the exposure and disease status of each study participant. John Snow's studies of cholera in London were observational studies. The two most common types of observational studies are cohort studies and case-control studies; a third type is cross-sectional studies.

Cohort study. A cohort study is similar in concept to the experimental study. In a cohort study the epidemiologist records whether each study participant is exposed or not, and then tracks the participants to see if they develop the disease of interest. Note that this differs from an experimental study because, in a cohort study, the investigator observes rather than determines the participants' exposure status. After a period of time, the investigator compares the disease rate in the exposed group with the disease rate in the unexposed group. The unexposed group serves as the comparison group, providing an estimate of the baseline or expected amount of disease occurrence in the community. If the disease rate is substantively different in the exposed group compared to the unexposed group, the exposure is said to be associated with illness.

The length of follow-up varies considerably. In an attempt to respond quickly to a public health concern such as an outbreak, public health departments tend to conduct relatively brief studies. On the other hand, research and academic organizations are more likely to conduct studies of cancer, cardiovascular disease, and other chronic diseases which may last for years and even decades. The Framingham study is a well-known cohort study that has followed over 5,000 residents of Framingham, Massachusetts, since the early 1950s to establish the rates and risk factors for heart disease. The Nurses Health Study and the Nurses Health Study II are cohort studies established in 1976 and 1989, respectively, that have followed over 100,000 nurses each and have provided useful information on oral contraceptives, diet, and lifestyle risk factors. These studies are sometimes called **follow-up** or **prospective** cohort studies, because participants are enrolled as the study begins and are then followed prospectively over time to identify occurrence of the outcomes of interest.

An alternative type of cohort study is a **retrospective** cohort study. In this type of study both the exposure and the outcomes have already occurred. Just as in a prospective cohort study, the investigator calculates and compares rates of disease in the exposed and unexposed groups. Retrospective cohort studies are commonly used in investigations of disease in groups of easily identified people such as workers at a particular factory or attendees at a wedding. For example, a retrospective cohort study was used to determine the source of infection of cyclosporiasis, a parasitic disease that caused an outbreak among members of a residential facility in Pennsylvania in 2004. The investigation indicated that consumption of snow peas was implicated as the vehicle of the cyclosporiasis outbreak.

Case-control study. In a case-control study, investigators start by enrolling a group of people with disease (at CDC such persons are called case-patients rather than cases, because case refers to occurrence of disease, not a person). As a comparison group, the investigator then enrolls a group of people without disease (controls). Investigators then compare previous exposures between the two groups. The control group provides an estimate of the baseline or expected amount of exposure in that population. If the amount of exposure among the case group is substantially higher than the amount you would expect based on the control group, then illness is said to be associated with that exposure. The study of hepatitis A traced to green onions, described above, is an example of a case-control study. The key in a case-control study is to identify an appropriate control group, comparable to the case group in most respects, in order to provide a reasonable estimate of the baseline or expected exposure.

Cross-sectional study. In this third type of observational study, a sample of persons from a population is enrolled and their exposures and health outcomes are measured simultaneously. The cross-sectional study tends to assess the presence (prevalence) of the health outcome at that point of time without regard to duration. For example, in a cross-sectional study of diabetes, some of the enrollees with diabetes may have lived with their diabetes for many years, while others may have been recently diagnosed.

From an analytic viewpoint the cross-sectional study is weaker than either a cohort or a case-control study because a cross-sectional study usually cannot disentangle risk factors for occurrence of disease (incidence) from risk factors for survival with the disease. (Incidence and prevalence are discussed in more detail in Lesson 3.) On the other hand, a cross-sectional study is a perfectly fine tool for descriptive epidemiology purposes. Cross-sectional studies are used

routinely to document the prevalence in a community of health behaviors (prevalence of smoking), health states (prevalence of vaccination against measles), and health outcomes, particularly chronic conditions (hypertension, diabetes).

In summary, the purpose of an analytic study in epidemiology is to identify and quantify the relationship between an exposure and a health outcome. The hallmark of such a study is the presence of at least two groups, one of which serves as a comparison group. In an experimental study, the investigator determines the exposure for the study subjects; in an observational study, the subjects are exposed under more natural conditions. In an observational cohort study, subjects are enrolled or grouped on the basis of their exposure, then are followed to document occurrence of disease. Differences in disease rates between the exposed and unexposed groups lead investigators to conclude that exposure is associated with disease. In an observational case-control study, subjects are enrolled according to whether they have the disease or not, then are questioned or tested to determine their prior exposure. Differences in exposure prevalence between the case and control groups allow investigators to conclude that the exposure is associated with the disease. Cross-sectional studies measure exposure and disease status at the same time, and are better suited to descriptive epidemiology than causation.