already succumbed to an infectious disease may no longer be susceptible to it for some period of time. If a substantial proportion of a population is affected by the outbreak, the number of susceptible people will decline gradually as the epidemic progresses and the attack rate increases. This change in the susceptible population leads to a more rapid decline over time in the number of new cases compared with the incidence rate in the susceptible population. The incidence rate declines more slowly than the number of new cases because in the incidence rate, the declining number of new cases is divided by a dwindling amount of susceptible person-time.

A *propagated epidemic* is one in which the causal agent is transmitted through a population. Influenza epidemics are propagated by person-to-person transmission of the virus. The epidemic of lung cancer during the 20th century was a propagated epidemic attributable to the spread of tobacco smoking through many cultures and societies. The curve for a propagated epidemic tends to show a more gradual initial rise and a more symmetric shape than for a point-source epidemic because the causes spread gradually through the population. Transmission of infectious disease within a population is discussed further in Chapter 6, which also presents the Reed-Frost model, a simple model that describes transmission of an infectious disease in a closed population.

Although we may think of point-source epidemics as occurring over a short time span, they are not always briefer than propagated epidemics. The epidemic of cancer attributable to exposure to the atomic bombs detonated in Hiroshima and Nagasaki was a point-source epidemic that began a few years after the explosions and continues into the present. Another possible point-source epidemic that occurred over decades was an apparent outbreak of multiple sclerosis in the Faroe Islands that followed the occupation of those islands by British troops during the Second World War[4] (although this interpretation of the data has been questioned[5]). Propagated epidemics can occur over extremely short time spans. An example is epidemic hysteria, a disease often propagated from person to person in minutes. An example of an epidemic curve for a hysteria outbreak is depicted in Figure 4–5. In this epidemic, 210 elementary school children developed symptoms of headache, abdominal pain, and nausea. These symptoms were attributed by the investigators to hysteric anxiety.[6]

Prevalence Proportion

Incidence proportion and incidence rate are measures that assess the frequency of disease onsets. The numerator of either measure is the frequency of events that are defined as the occurrence of disease. In contrast, *prevalence proportion*, often referred to simply as *prevalence*, does not measure disease onset. Instead, it is a measure of disease status.

The simplest way of considering disease status is to consider disease as being either present or absent. The prevalence proportion is the proportion of people in a population who have disease. Consider a population of size N, and suppose that P individuals in the population have disease at a given time. The prevalence proportion is P/N. For example, suppose that among 10,000 women residents of a town on July 1, 2001, 1200 have hypertension. The prevalence proportion of hypertension among women in that town on that date is $1200/10{,}000 = 0.12$,

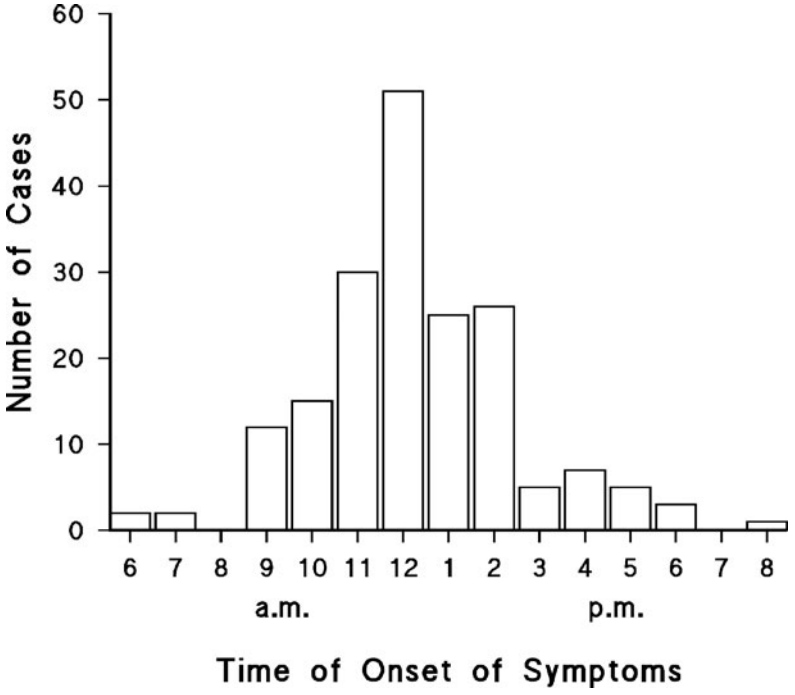

Figure 4–5 Epidemic curve for an outbreak of hysteria among elementary school children on November 6, 1985.

or 12%. This prevalence applies only to a single point in time, July 1, 2001. Prevalence can change with time as the factors that affect prevalence change.

What factors affect prevalence? Clearly, disease occurrence affects prevalence. The greater the incidence of disease, the more people there are who have it. Prevalence is also related to the length of time that a person has disease. The longer the duration of disease, the higher the prevalence. Diseases with short duration may have a low prevalence even if the incidence rate is high. One reason is that if the disease is benign, there may be a rapid recovery. For example, the prevalence of upper respiratory infection may be low despite a high incidence, because after a brief period, most people recover from the infection and are no longer in the disease state. Duration may also be short for a grave disease that leads to rapid death. The prevalence of aortic hemorrhage would be low even with a high incidence because it usually leads to death within minutes. The low prevalence means that, at any given moment, only an extremely small proportion of people are suffering from an aortic hemorrhage. Some diseases have a short duration because either recovery or death ensues promptly; appendicitis is an example. Other diseases have a long duration because, although a person cannot recover from them, they are compatible with a long survival time (although survival is often shorter than it would be without the disease). Diabetes, Crohn's disease, multiple sclerosis, parkinsonism, and glaucoma are examples.

Because prevalence reflects both incidence rate and disease duration, it is not as useful as incidence alone for studying the causes of disease. It is extremely

useful, however, for measuring the disease burden on a population, especially if those who have disease require specific medical attention. For example, the prevalent number of people in a population with end-stage renal disease predicts the need in that population for dialysis facilities.

In a *steady state*, which is the situation in which incidence rates and disease duration are stable over time, the prevalence proportion, P, has the following relation to the incidence rate:

$$\frac{P}{1-P} = I\bar{D} \qquad [4–2]$$

In Equation 4–2, I is the incidence rate and $\bar{D}$ is the average duration of disease. The quantity $P/(1-P)$ is known as the *prevalence odds*. In general, when a proportion, such as prevalence proportion, is divided by 1 minus the proportion, the resulting ratio is referred to as the *odds* for that proportion. If a horse is a 3-to-1 favorite at a racetrack, it means that the horse is thought to have a probability of winning of 0.75. The odds of the horse winning is $0.75/(1-0.75) = 3$, usually described as 3 to 1. Similarly, if a prevalence proportion is 0.75, the prevalence odds would be 3, and a prevalence of 0.20 would correspond to a prevalence odds of $0.20/(1-0.20) = 0.25$. For small prevalences, the value of the prevalence proportion and that of the prevalence odds are close because the denominator of the odds expression is close to 1. For small prevalences (eg, <0.1), we can rewrite Equation 4–2 as follows:

$$P \approx I\bar{D} \qquad [4–3]$$

Equation 4–3 indicates that, given a steady state and a low prevalence, prevalence is approximately equal to the product of the incidence rate and the mean duration of disease. Note that this relation does not hold for age-specific prevalences. In that case, $\bar{D}$ corresponds to the duration of time spent within that age category rather than the total duration of time with disease.

As we did earlier for risk and incidence rate, we should check this equation to make certain that the dimensionality and ranges of both sides of the equation are satisfied. For dimensionality, the right-hand sides of Equations 4–2 and 4–3 involve the product of a time measure, disease duration, and an incidence rate, which has units of reciprocal of time. The product is dimensionless, a pure number. Prevalence proportion, like risk or incidence proportion, is also dimensionless, which satisfies the dimensionality requirement for the two equations, 4–2 and 4–3. The range of incidence rate and that of mean duration of illness is $[0,\infty]$, because there is no upper limit to an incidence rate or the duration of disease. Therefore Equation 4–3 does not satisfy the range requirement, because the prevalence proportion on the left side of the equation, like any proportion, has a range of $[0,1]$. For this reason, Equation 4–3 is applicable only for small values of prevalence. The measure of prevalence odds in Equation 4–2, however, has a range of $[0,\infty]$, and it is applicable for all values, rather than just for small values of the prevalence proportion. We can rewrite Equation 4–2 to solve for the prevalence proportion as follows:

$$P = \frac{I\bar{D}}{1+I\bar{D}} \qquad [4–4]$$

Prevalence measures the disease burden in a population. This type of epidemiologic application relates more to administrative areas of public health than to causal research. Nevertheless, there are research areas in which prevalence measures are used more commonly than incidence measures, even to investigate causes. Examples are birth defects and birth-related phenomena such as birth weight or preterm birth. We use a prevalence measure when describing the occurrence of congenital malformations among liveborn infants in terms of the proportion of these infants who have a malformation. For example, the proportion of infants who are born alive with a defect of the ventricular septum of the heart is a prevalence. It measures the status of liveborn infants with respect to the presence or absence of a ventricular septal defect. Measuring the incidence rate or incidence proportion of ventricular septal defects would require ascertainment of a population of embryos who were at risk for developing the defect and measurement of the defect's occurrence among these embryos. Such data are usually not obtainable, because many pregnancies end before the pregnancy is detected, and the population of embryos is not readily identified. Even when a woman knows she is pregnant, if the pregnancy ends early, information about the pregnancy may never come to the attention of researchers. For these reasons, incidence measures for birth defects are uncommon. Prevalence at birth is easier to assess and often is used as a substitute for incidence measures. Although prevalence measures are easier to obtain, they have a drawback when used for causal research: Factors that increase prevalence may do so not by increasing the occurrence of the condition but by increasing the duration of the condition. For example, a factor associated with the prevalence of ventricular septal defect at birth could be a cause of ventricular septal defect, but it could also be a factor that does not cause the defect but instead enables embryos that develop the defect to survive until birth. On the other hand, there may be practical interest in understanding the factors that are related to being born alive with the defect.

Prevalence is sometimes used in research to measure diseases that have insidious onset, such as diabetes or multiple sclerosis. These are conditions for which it may be difficult to define onset, and it therefore may be necessary in some settings to describe the condition in terms of prevalence rather than incidence.

Prevalence of Characteristics

Because prevalence measures status, it is often used to describe the status of characteristics or conditions other than disease in a population. For example, the proportion of a population that engages in cigarette smoking often is described as the prevalence of smoking. The proportion of a population exposed to a given agent is often referred to as the exposure prevalence. Prevalence can be used to describe the proportion of people in a population who have brown eyes, type O blood, or an active driver's license. Because epidemiology relates many individual and population characteristics to disease occurrence, it often employs prevalence measures to describe the frequency of these characteristics.