chapter

3 Measuring the Occurrence of Disease: I. Morbidity

We owe all the great advances in knowledge to those who endeavor to find out how much there is of anything.
– James Maxwell, physicist (1831–1879)

If you can measure that of which you speak, and can express it by a number, you know something of your subject, but if you cannot measure it, your knowledge is meager and unsatisfactory.
– William Thomson, Lord Kelvin, engineer, mathematician and physicist (1824–1907)

In Chapter 2, we discussed how diseases are transmitted. It is clear from that discussion that in order to examine the transmission of disease in human populations, we need to be able to measure the frequency of both disease occurrence and deaths from the disease. In this chapter, we will therefore discuss how we use rates to express the extent of morbidity resulting from a disease, and in the next chapter (Chapter 4), we will turn to expressing the extent of mortality in quantitative terms.

Let us begin this discussion by considering the development and course of a disease in an individual over a period of time.

Figure 3–1 shows the progression of disease in a population as reflected by the levels of illness and medical care. The outside rectangle represents the total population, and the smaller rectangles represent progressively smaller subsets, from sick to hospitalized patients. Deaths straddle all of these rectangles, but the death rate is proportionately greater in groups with more severe illness.

Figure 3–2 shows the timeline for the development of a disease in an individual. An individual is healthy (i.e., without disease), and at some point, biologic onset of a disease occurs. The person is often unaware of the point in time when the disease begins. Later, symptoms develop and lead the patient to seek medical care. In certain situations, hospitalization may be required, either for diagnosis, for treatment, or for both. In any case, at some point a diagnosis is made and treatment is initiated. One of several outcomes can then result: cure, control of the disease, disability, or death.

What sources of data can be used to obtain information about the person's illness? For the period of the illness that necessitates hospitalization, medical and hospital records are useful. If hospitalization is not required, physicians' records may be the best source. If we want information about the illness even before medical care was sought, we may have to obtain this information from the patient, using a questionnaire or an interview. Not shown in this figure are the records of health insurers, which at times can provide very useful information.

The source of data from which cases are identified clearly influences the rates that we calculate for expressing the frequency of disease. For example, hospital records will not include data regarding patients who obtained care only in physicians' offices. Consequently, when we see rates for the frequency of occurrence of a certain disease, we must identify the sources of the cases and determine how the cases were identified before we interpret the rates and compare them to rates reported in other populations and at other times.

The occurrence of disease can be measured using rates or proportions. *Rates* tell us how fast the disease is occurring in a population, and *proportions* tell us what fraction of the population

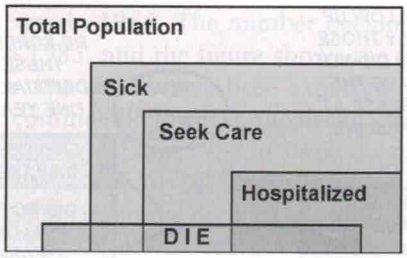

FIGURE 3–1▼ The population: progression from health to varying degrees of disease severity. (Adapted from White KL, Williams TF, Greenberg BG: The ecology of medical care. N Engl J Med 265:885–892, 1961.)

is affected. Let us turn to how we use rates and proportions for expressing the extent of disease in a community or other population. *Measures of illness* or *morbidity* are discussed in this chapter, followed by a discussion of *measures of mortality* in Chapter 4.

Measures of Morbidity

Incidence

The *incidence* of a disease is defined as the number of new cases of a disease that occur during a specified period of time in a population at risk for developing the disease.

$$\text{Incidence per 1,000} = \frac{\text{No. of new cases of a disease occurring in the population during a specified period of time}}{\text{No. of persons at risk of developing the disease during that period of time}} \times 1{,}000$$

In this rate, the result has been multiplied by 1,000 so that we can express the incidence per 1,000 persons. The choice of 1,000 is completely arbitrary—we could have used 10,000, 1 million, or any other figure.

The critical element in the definition of incidence is *new* cases of disease. Incidence is a measure of events—the disease is identified in a person who develops the disease and did not have the disease previously. Because incidence is a measure of events (i.e., transition from a nondiseased to a diseased state), incidence is a *measure of risk*. This risk can be looked at in any population group, such as a particular age group, males or females, an occupational group, or a group that has been exposed to a certain environmental agent, such as radiation or a chemical toxin.

The denominator of incidence represents the number of people who are at risk for developing the disease. For incidence to be meaningful, *any individual who is included in the denominator must have the potential to become part of the group that is counted in the numerator*. Thus, if we are calculating incidence of uterine cancer, the denominator must include only women, because men would not have the potential to become part of the group that is counted by the numerator—that is, men are not at risk for developing uterine cancer. Although this point seems obvious, it is not always so clear, and we shall return to this issue later in the discussion.

Another important issue with regard to the denominator is the issue of time. For incidence to be a measure of risk, we must specify a period of time, and we must know that all of the individuals in the group represented by the denominator have been followed up for that entire period. The choice of time period is arbitrary: We could calculate incidence in 1 week, incidence in 1 month, incidence in 1 year, incidence in 5 years, and so on. The important point is that whatever time period is used in the calculation must be clearly specified, and all individuals included in the calculation must have been observed (at risk) for the entire period. The incidence calculated using a period of time during which all of the individuals in the population are considered to be at risk

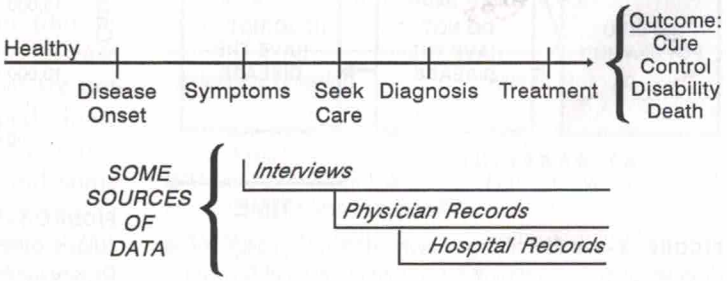

FIGURE 3–2▼ The natural history of disease and some sources of data relating to each interval.

for the outcome is called *cumulative incidence*, which is a measure of risk.

Practically speaking, how do we identify all new cases in a population during a specified time period? In certain situations it may be possible to monitor an entire population over time with tests which can detect newly developed cases of a disease. However, often this is not possible and instead a population is identified and screened for the disease at baseline (prevalent cases defined in the next section) (Fig. 3–3). Those who do not have the disease are followed for the specified time, e.g. one year, and they are then rescreened (Fig. 3–4). Any cases that are identified clearly developed during the one year period since those followed were free of disease at the beginning of the year. Thus these cases are new or incident cases and serve as the numerator for the incidence rate.

In Figure 3–4, the entire population is shown as being followed for the full 1-year period. Often, however, every individual in the denominator has not been followed for the full time specified. For a variety of reasons, including loss to follow-up, different individuals may be observed for different lengths of time. In such a case, we calculate an *incidence rate* (also called an *incidence density*), in which the denominator consists of the sum of the different times that each individual was at risk. This is often expressed in terms of person-years, which is further discussed in Chapter 6. This does not affect the number of cases counted in the numerator.

Occasionally, time may be specified implicitly rather than explicitly. For example, in investigating a foodborne disease outbreak, we know that most cases occur within a few hours or a few days after the exposure. Thus, cases that develop months later are not considered part of the same outbreak. However, in many situations, current knowledge of the biology and natural history of the disease does not clearly define a time frame, and so the time must be stated explicitly.

Although in most situations, it is necessary to express incidence by specifying a denominator, at times, the number of cases alone may be informative. For example, Figure 3–5 shows the number of expected and observed cases of tuberculosis reported in the United States from 1980 to 1992. (Note that the vertical axis is a logarithmic scale.) The smallest number of cases ever reported in a year in the United States (since reporting

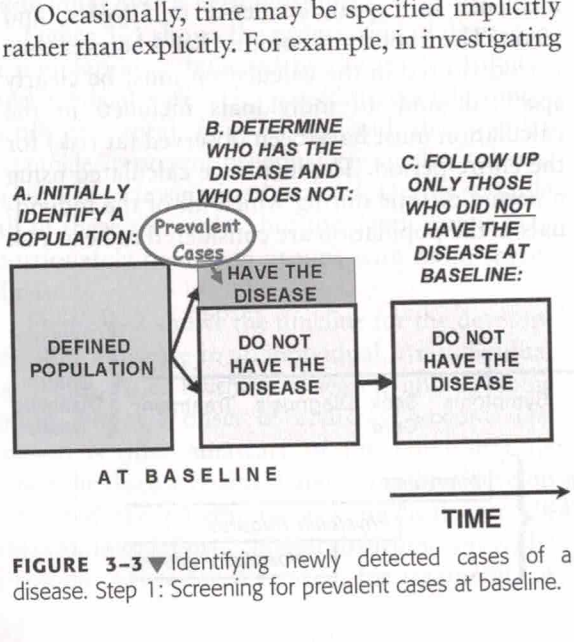

FIGURE 3–3 ▼ Identifying newly detected cases of a disease. Step 1: Screening for prevalent cases at baseline.

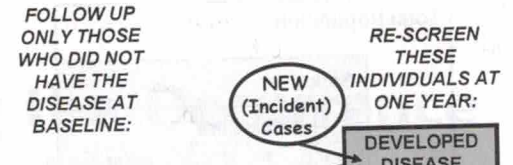

FIGURE 3–4 ▼ Identifying newly detected cases of a disease. Step 2: Follow-up and rescreening at 1 year to identify cases that developed during the year.

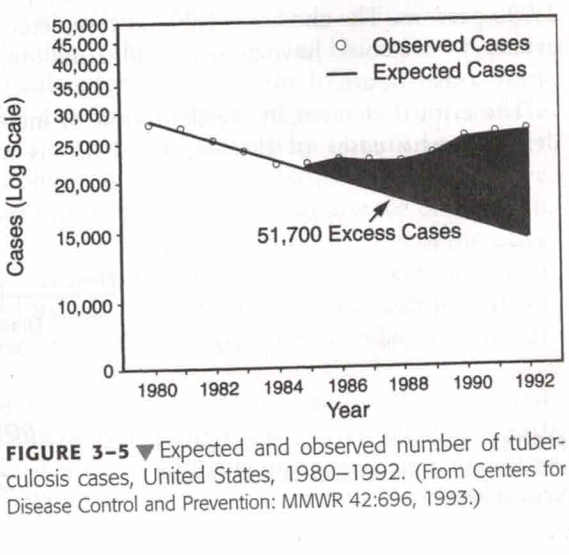

FIGURE 3–5 ▼ Expected and observed number of tuberculosis cases, United States, 1980–1992. (From Centers for Disease Control and Prevention: MMWR 42:696, 1993.)

began) was in 1985. The number declined from 1980 to 1985, and the figure shows the number of cases that would have been expected had the decline continued. However, the decline suddenly stopped in 1985. From 1985 to 1992, the reported number of cases of tuberculosis increased by 20%; had the projected decline continued, approximately 51,700 fewer cases would have been expected. Much of the increase in tuberculosis seen here was associated with simultaneous infection with human immunodeficiency virus (HIV). However, even before acquired immunodeficiency syndrome (AIDS) and HIV were recognized as major public health problems, tuberculosis had remained a serious, but often neglected, problem, particularly in certain urban areas of the United States. We see that even a graph that plots numbers of cases without a denominator can be very helpful when there is no reason to suspect a significant change in the denominator during a given time period.

Prevalence

Prevalence is defined as the number of affected persons present in the population at a specific time divided by the number of persons in the population at that time.

$$\text{Prevalence per 1,000} = \frac{\text{No. of cases of a disease present in the population at a specified time}}{\text{No. of persons in the population at that specified time}} \times 1{,}000$$

For example, if we are interested in knowing the prevalence of arthritis in a certain community on a certain date, we might visit every household in that community and, using interviews or physical examinations, determine how many people have arthritis on that day. This number becomes the numerator for prevalence. The denominator is the population in the community on that date.

What is the difference between *incidence* and *prevalence?* Prevalence can be viewed as a slice through the population at a point in time at which it is determined who has the disease and who does not. But in so doing, we are not determining *when* the disease developed. Some individuals may have developed arthritis yesterday, some last week, some last year, and some 10 or 20 years ago. *Thus, when we survey a community to estimate the prevalence of a disease, we generally do not take into account the duration* of the disease. Consequently, the numerator of prevalence includes a mix of people with different durations of disease, and as a result we do not have a measure of risk. If we wish to measure risk, we must use incidence, because in contrast to prevalence, it includes only new cases or events and a specified time period during which those events occurred.

In the medical and public health literature, the word *prevalence* is often used in two ways:

POINT PREVALENCE. Prevalence of the disease at a point in time—the usage we have just discussed.

PERIOD PREVALENCE. How many people have had the disease at any time during a certain period, such as during a single calendar year. Some people may have developed the disease during that period, and others may have had the disease before and died or been cured during that period. The important point is that every person represented by the numerator had the disease at some time during the period specified.

The two types of prevalence, as well as cumulative incidence, are illustrated in Table 3–1 using questions regarding asthma.

Returning to point prevalence, practically speaking, it is virtually impossible to survey an entire city on a single day. Therefore, although conceptually, we are thinking in terms of a single point in time, in reality, the survey would take much longer. When we see the word *prevalence* used without a modifier, it generally refers to point prevalence, and for the rest of this chapter, we will use *prevalence* to mean point prevalence.

Let us consider incidence and prevalence. Figure 3–6 shows five cases of a disease in a community in 2004. The first case of the disease occurred in 2003, and the patient died in 2004.

TABLE 3–1 Examples of Point and Period Prevalence and Cumulative Incidence in Interview Studies of Asthma

Interview Question	Type of Measure
"Do you currently have asthma?"	Point prevalence
"Have you had asthma during the last *(n)* years?"	Period prevalence
"Have you ever had asthma?"	Cumulative incidence

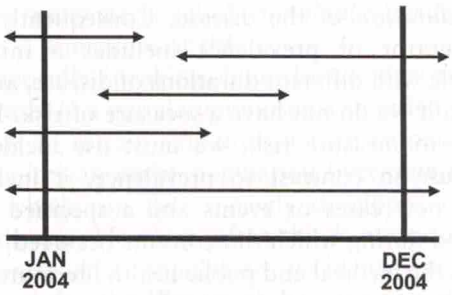

FIGURE 3–6 ▼ Example of incidence and prevalence: I.

The second case developed in 2004 and continued into 2005. The third case was a person who became ill in 2004 and was cured in 2004. The fourth case occurred in 2003, and the patient was cured in 2004. The fifth case occurred in 2003 and continued through 2004 and into 2005.

For this example, we will consider only the cases (numerators) and will ignore the denominators. In this example, what is the numerator for incidence in 2004? We know that incidence counts only *new* cases, and because two of the five cases developed in 2004, the numerator for incidence in 2004 is 2.

What about the numerator for point prevalence? This depends on when we do our prevalence survey (Fig. 3–7). If we do the survey in May, the numerator will be 4. If we do the survey in July, the numerator will also be 4. If we do the survey in September, however, the numerator will be 3, and if we do it in December, the numerator will be 2. Thus, the prevalence will depend on the point during the year at which the survey is performed.

Figure 3–8 shows the relationship between incidence and prevalence. A flask is shown that represents a community, and the beads in the flask represent the prevalence of a disease in the community. How can we add to or increase

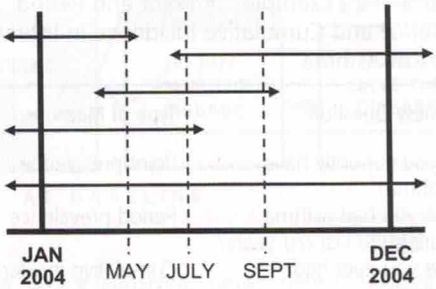

FIGURE 3–7 ▼ Example of incidence and prevalence: II.

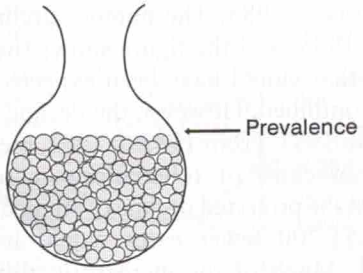

FIGURE 3–8 ▼ Relationship between incidence and prevalence: I.

the prevalence? As seen in Figure 3–9, we can do so through incidence—by the addition of new cases. What if we could drain beads from the flask and lower the prevalence? How might this be accomplished? As seen in Figure 3–10, it could occur through either death or cure. Clearly, these two outcomes represent a major difference to a patient, but with regard to prevalence, cure and death have the same effect: They reduce the number of diseased persons in the population and thus lower prevalence. Therefore, what exists is the dynamic situation shown in Figure 3–11. A continual addition of new cases (incidence) is increasing the prevalence, while death and/or cure is decreasing the prevalence.

This effect of lowering prevalence through either death or cure underlies an important issue in public health and clinical medicine. For example, when insulin first became available, what happened to the prevalence of diabetes? The prevalence increased because diabetes was not cured, but was only controlled. Many patients with diabetes who formerly would have died now survived; therefore, the prevalence

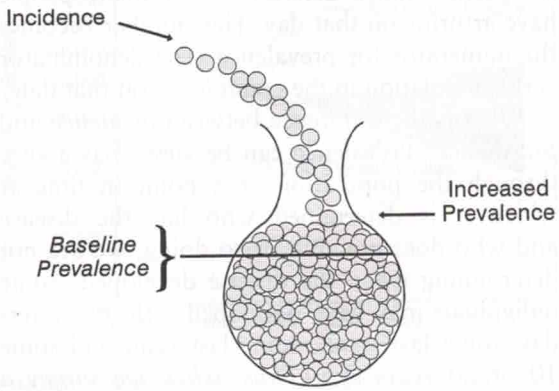

FIGURE 3–9 ▼ Relationship between incidence and prevalence: II.

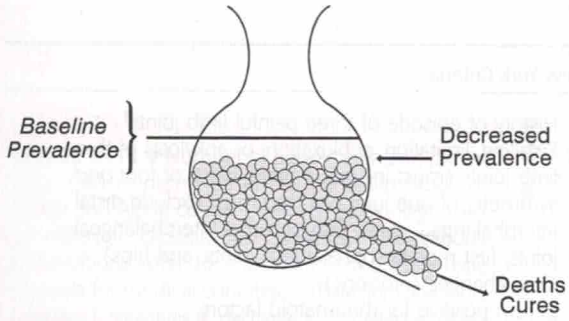

FIGURE 3–10 ▼ Relationship between incidence and prevalence: III.

increased. This seeming paradox is often the case with public health programs: A new measure is introduced that enhances survival or detects the disease in more people, and the net effect is an apparent increase in prevalence. It may be difficult to convince some people that a program is successful if the prevalence of the disease that is the target of the program actually increases. However, this clearly occurs when death is prevented and the disease is not cured.

We have said that prevalence is not a measure of risk. If so, why bother to estimate prevalence? Prevalence is an important and useful measure of the burden of disease in a community. For example, how many people in the community have arthritis? This information might help us to determine, for example, how many clinics are needed, what types of rehabilitation services are needed, and how many and what types of health professionals are needed. Prevalence is therefore valuable for planning health services. When we use prevalence, we also want to make future projections and anticipate the changes that are likely to take place in the disease burden. However, if we want to look at the cause or etiology of disease, we must explore the relationship between an exposure and the risk of disease, and to do this, we need incidence rates.

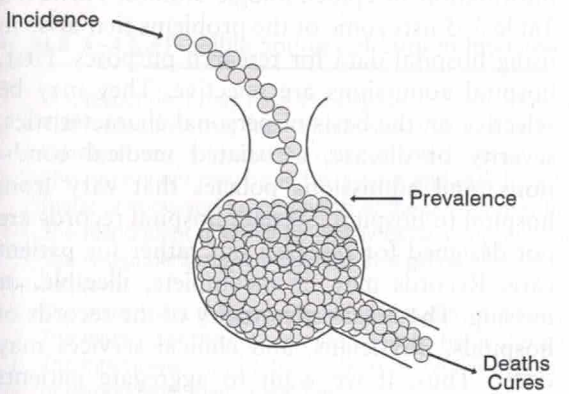

FIGURE 3–11 ▼ Relationship between incidence and prevalence: IV.

Table 3–2 lists some possible sources of morbidity statistics. Each has its limitations, primarily because most of these sources are not established for research purposes. Therefore, they may be characterized by incomplete or ambiguous data and, at times, by the use of a highly selected population that may not be representative of the population to which we would like to generalize the findings.

Problems with Incidence and Prevalence Measurements

PROBLEMS WITH NUMERATORS. The first problem is defining who has the disease. One example demonstrates this problem: Rheumatoid arthritis (RA) is often a difficult disease to diagnose, and when such a diagnostic difficulty arises, expert groups are often convened to develop sets of diagnostic criteria. Two sets of diagnostic criteria for RA are those of the New York Rheumatism Association and the American Rheumatism Association (Table 3–3). Figure 3–12 shows the

TABLE 3–2 ▽ Some Sources of Morbidity Statistics

1. Disease reporting—communicable diseases, cancer registries
2. Data accumulated as a by-product of insurance and prepaid medical care plans
   a. Group health and accident insurance
   b. Prepaid medical care plans
   c. State disability insurance plans
   d. Life insurance companies
   e. Hospital insurance plans—Blue Cross
   f. Railroad Retirement Board
3. Tax-financed public assistance and medical care plans
   a. Public assistance, aid to the blind, aid to the disabled
   b. State or federal medical care plans
   c. Armed Forces
   d. Veterans Administration
4. Hospitals and clinics
5. Absenteeism records—industry and schools
6. Pre-employment and periodic physical examinations in industry and schools
7. Case-finding programs
8. Records of military personnel
9. Morbidity surveys on population samples (e.g., National Health Survey, National Cancer Surveys)

TABLE 3–3 Criteria for Rheumatoid Arthritis*

American Rheumatism Association Criteria	New York Criteria
1. Morning stiffness	1. History of episode of three painful limb joints†
2. Joint tenderness or pain on motion	2. Swelling, limitation, subluxation, or ankylosis of three limb joints (must include a hand, wrist, or foot *and* symmetry of one joint pair *and* must exclude distal interphalangeal joints, fifth proximal interphalangeal joints, first metatarsophalangeal joints, and hips)
3. Soft-tissue swelling of one joint	3. X-ray changes (erosions)
4. Soft-tissue swelling of a second joint (within 3 mo)	4. Serum positive for rheumatoid factors
5. Soft-tissue swelling of symmetrical joints (excludes distal interphalangeal joint)	
6. Subcutaneous nodules	
7. X-ray changes	
8. Serum positive for rheumatoid factors	

*A score of three or four points indicates "probable" rheumatoid arthritis; five or more points indicates "definite" rheumatoid arthritis.
†Count each joint group (e.g., proximal interphalangeal joints) as one joint, scoring each side separately.
From O'Sullivan JB, Cathcart ES: The prevalence of rheumatoid arthritis. Ann Intern Med 76:573, 1972.

results of a survey conducted in Sudbury, Mass., using both sets of criteria. We see that the prevalence estimate is significantly affected by the set of criteria that is used.

More recently, a cohort of 1,879 men and women 65 years of age and older who were enrolled in the Canadian Study of Health and Aging (CSHA) were examined.[1] The proportion who were given a diagnosis of dementia using six commonly used classification systems was calculated. Depending on which diagnostic system was used, the proportion of subjects with dementia varied from 3.1% to 29.1% (Fig. 3–13). This marked variation in prevalence estimates has important potential implications both for research and for the provision of appropriate health services. When the results of any morbidity survey are reported, it is essential that the precise definition used for a case be clearly specified. The decision as to which definition to use is not always simple. Often it will largely depend on the specific purpose for which a given survey has been conducted.

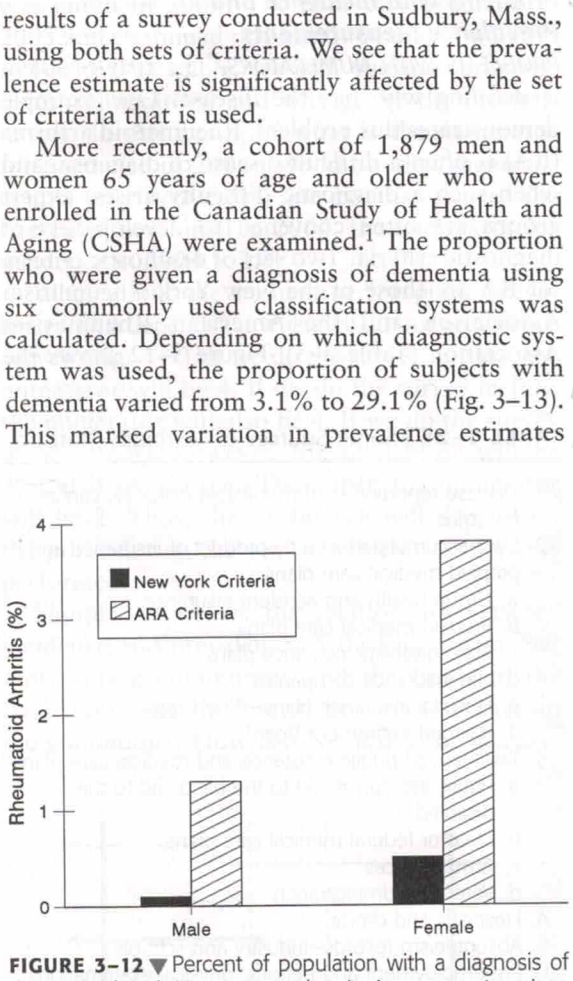

FIGURE 3–12 Percent of population with a diagnosis of rheumatoid arthritis. New York criteria versus American Rheumatism Association (ARA) criteria, Sudbury, Massachusetts, 1964. (Adapted from O'Sullivan JB, Cathcart ES: The prevalence of rheumatoid arthritis: Follow-up evaluation of the effect of criteria on rates in Sudbury, Massachusetts. Ann Intern Med 76:573–577, 1972.)

The next issue relating to numerators is that of ascertaining which persons should be included in the numerator. How do we find the cases? We can use regularly available data, or as discussed earlier in this chapter, we can conduct a study specifically designed to gather data for estimating incidence or prevalence. In many such studies the data are obtained from interviews, and some of the problems with interview data are listed in Table 3–4.

PROBLEMS WITH HOSPITAL DATA. Data from hospital records are one of the most important sources of information in epidemiologic studies. However, Table 3–5 lists some of the problems that arise in using hospital data for research purposes. First, hospital admissions are selective. They may be selective on the basis of personal characteristics, severity of disease, associated medical conditions, and admissions policies that vary from hospital to hospital. Second, hospital records are not designed for research but rather for patient care. Records may be incomplete, illegible, or missing. The diagnostic quality of the records of hospitals, physicians, and clinical services may differ. Thus, if we want to aggregate patients from different hospitals, we may have problems of comparability. Third, if we wish to calculate

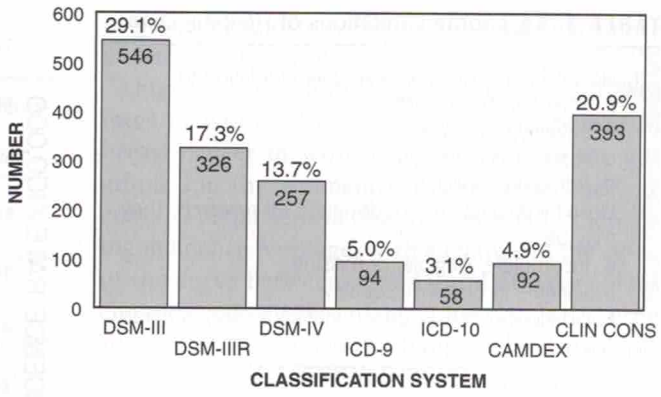

FIGURE 3–13 ▼ Number of people with and prevalence (%) of dementia in the Canadian Study of Health and Aging cohort (N = 1879) as diagnosed by different classification systems. The various abbreviations refer to commonly used diagnostic manuals for medical conditions. (Data from Erkinjuntti T, Østbye T, Steenhuis R, Hachinski V: The effect of different diagnostic criteria on the prevalence of dementia. N Engl J Med 337:1667–1774, 1997.)

rates, we have a problem defining denominators, because most hospitals do not have defined catchment areas that would require that all persons in a given area who are hospitalized be admitted to a particular hospital, and that none from outside the catchment area be admitted to that hospital.

In Chapter 5, we will turn to how we use screening and diagnostic tests to identify individuals who are ill (who are included in the numerator) and distinguish them from those in the population who are not ill.

PROBLEMS WITH DENOMINATORS. Many factors affect the denominators used. Selective undercounting of certain groups in the population may occur. For example, young males in ethnic minority groups have been missed in many counts of the population. Frequently, we wish to determine whether a certain group has a higher-than-expected risk of disease so that appropriate preventive measures can be directed to that group. We are therefore interested in the rates of disease for different ethnic groups rather than just for the population as a whole. However, there are different ways to classify people by ethnic group, such as by language, country of origin, heritage, or parental ethnic group. When different studies use different definitions, comparison of the results is difficult. What is most important in any study is that the working definition be clearly stated so that the reader can judge whether the results are truly comparable.

In an earlier section we stated that for a rate to make sense, everyone in the group represented by the denominator must have the potential to enter the group that is represented by the numerator. The issue is not a simple one. For example, hysterectomy is one of the most commonly performed surgical procedures in the United States. This raises a question about uterine cancer rates. For if we include women who have had hysterectomies in the denominator, clearly they are not at risk for developing uterine cancer. Figure 3–14 shows uterine cancer incidence rates from Alameda County, California; both uncorrected

TABLE 3–4 ▼ Possible Sources of Error in Interview Surveys

1. The respondent may have the disease, but may have no symptoms and may not be aware of the disease.
2. The respondent may have the disease and may have had symptoms, but may not have had medical attention and therefore may not know the name of the disease.
3. The respondent may have the disease and may have had medical attention, but the diagnosis may not have been made or conveyed to the person or the person may have misunderstood.
4. The respondent may not accurately recall an episode of illness or events and exposures related to the illness.
5. The respondent may be involved in litigation about the illness and may choose not to respond or may alter his or her response.
6. The respondent may provide the information, but the interviewer may not record it or may record it incorrectly.
7. The interviewer may not ask the question he or she is supposed to ask or may ask it incorrectly.
8. The interviewer may be biased by knowing the hypothesis being tested and may probe more intensively in one group of respondents than in another.
9. Problems of selection bias may occur, possibly including significant nonresponse rates.

TABLE 3–5 ▽ Some Limitations of Hospital Data

1. Hospital admissions are selective in relation to
   a. Personal characteristics
   b. Severity of disease
   c. Associated conditions
   d. Admission policies
2. Hospital records are not designed for research. They may be
   a. Incomplete, illegible, or missing
   b. Variable in diagnostic quality
3. Population(s) at risk (denominator) is (are) generally not defined

rates and rates corrected for hysterectomy are presented. We see that the corrected rates are higher. Why? Because in the corrected rates women who have had hysterectomies are removed from the denominator. Consequently, the denominator gets smaller and the rate increases. However, in this case the trend over time is not significantly changed whether we use corrected or uncorrected rates.

Relation Between Incidence and Prevalence

We have said that incidence is a measure of risk and that prevalence is not, because it does not take into account the duration of the disease. However, there is an important relationship between incidence and prevalence: In a steady-state situation, in which the rates are not changing and in-migration equals out-migration, the following equation applies:

Prevalence = Incidence × Duration of disease

This is demonstrated in the following hypothetical example. Using chest x-rays, 2,000 persons are screened for tuberculosis: 1,000 are upper-income individuals from Hitown and 1,000 are lower-income individuals from Lotown (Table 3–6). X-ray findings are positive in 100 of the Hitown people and in 60 of the Lotown people. Can we therefore conclude that the risk of tuberculosis is higher in Hitown people than in Lotown people? Clearly, we cannot, for what we are measuring with a chest x-ray is the point prevalence of disease—we do not know how long any of the people with positive x-rays have had their disease (Table 3–7). We could in fact consider a hypothetical scenario that might explain the higher prevalence in Hitown people that is not related

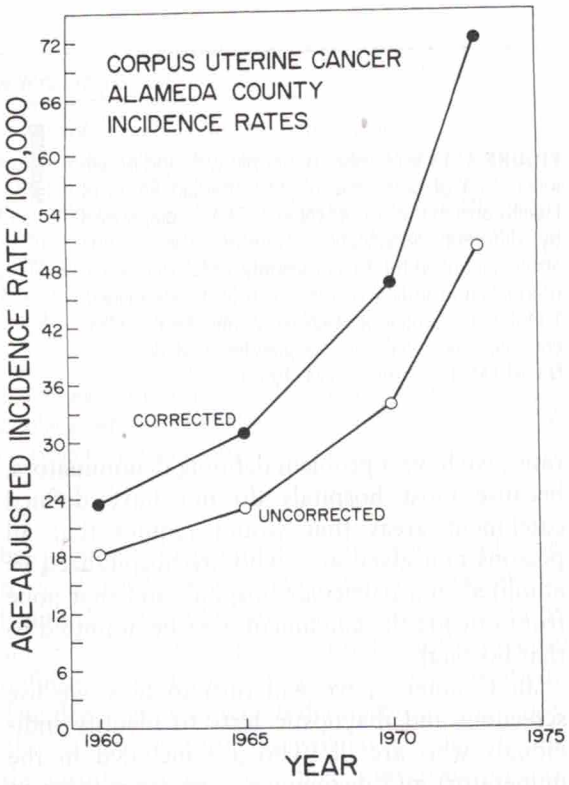

FIGURE 3–14 ▼ Age-adjusted uterine cancer incidence rates, corrected and uncorrected by hysterectomy status, Alameda County, California. (From Lyon JL, Gardner JW: The rising frequency of hysterectomy: Its effect on uterine cancer rates. Am J Epidemiol 105:439–443, 1977.)

to any higher risk in Hitown people (Table 3–8). We have said that prevalence = incidence × duration. Let us assume that Lotown people have a much higher risk (incidence) of tuberculosis than Hitown people—20 cases/year in Lotown people compared with 4 cases/year in Hitown people. But for a variety of reasons, such as poorer access to medical care and poor nutritional status, Lotown people survive with their disease, on average, for only 3 years, whereas Hitown people survive, on average, for 25 years. In this

TABLE 3–6 ▽ Hypothetical Example of Chest X-Ray Screening: I. Populations Screened and Numbers with Positive X-Rays

Screened Population	No. with Positive X-Ray
1,000 Hitown	100
1,000 Lotown	60

TABLE 3–7 Hypothetical Example of Chest X-Ray Screening: II. Point Prevalence

Screened Population	No. with Positive X-Ray	Point Prevalence per 1,000 Population
1,000 Hitown	100	100
1,000 Lotown	60	60

example, therefore, there is a higher prevalence in Hitown people than in Lotown people not because the risk of disease is higher in Hitown people, but because affected Hitown people survive longer; the prevalence of disease (incidence × duration) is therefore higher in Hitown people than in Lotown people.

Figure 3–15 shows the percent of all births in New Zealand that were extramarital from 1962 to 1979. Much concern was expressed because of the apparent steady rise in extramarital births. However, as seen in Figure 3–16, there had really been no increase in the *rate* of extramarital births; there had been a decline in total births that was largely accounted for by a decline in births to married women. The extramarital births, as a result, accounted for a greater percent of all births, even though the rate of extramarital births had not increased.

This example makes two points: First, a proportion is not a rate, and we shall return to this point in our discussion of mortality. Second, birth can be viewed as an event, just as the development of disease is an event, and appropriate rates can be computed. In discussing babies born with malformations, some people prefer to speak of the *prevalence* of malformations at birth rather than the *incidence* of malformations at birth, because the malformation was clearly present (but often unrecognized), even before birth. Furthermore, because some proportion of cases with malformations aborts before birth, any estimate of the frequency of malformations at birth is probably a significant underestimate of the true incidence. Hence, the term *prevalence at birth* is often used.

Figure 3–17 shows breast cancer incidence rates in women by age and the distribution of breast cancer in women by age. Ignore the bar graph for the moment, and consider the line curve. The pattern is one of continually increasing incidence with age, with a change in the slope of the curve between ages 45 and 50 years. This change is observed in many countries. It has been suggested that something happens near the time of menopause, and that premenopausal and postmenopausal breast cancer may be different diseases. Note that, even in old age, the incidence or risk of breast cancer continues to rise.

Now let us look at the histogram—the distribution of breast cancer cases by age. If the incidence is increasing so dramatically with age, why are only fewer than 5% of the cases occurring in the oldest age group of women? The answer is that there are very few women alive in that age group, so that even though they have the highest risk of breast cancer, the group is so small that they contribute only a small proportion of the total number of breast cancer cases seen at all ages. The fact that so few cases of breast cancer are seen in this age group has contributed to a false public impression that the risk of breast cancer is low in this group and that mammography is therefore not important in the elderly. This is a serious misperception. The need to change public thinking on this issue is a major public health challenge. We therefore see the importance of recognizing the distinction between the distribution of disease or the proportion of cases, and the incidence rate or risk of the disease.

SPOT MAPS. One approach to examining geographic or spatial differences in incidence is to plot the cases on a map, with each point representing a case. Figure 3–18 shows a spot map for rheumatic fever in Baltimore from 1960 to 1964. Rheumatic fever was frequently observed in this

TABLE 3–8 Hypothetical Example of Chest X-Ray Screening: III. Prevalence, Incidence, and Duration

Screened Population	Point Prevalence per 1,000	Incidence (Occurrences/yr)	Duration (yrs)
Hitown	100	4	25
Lotown	60	20	3
	Prevalence = Incidence × Duration		

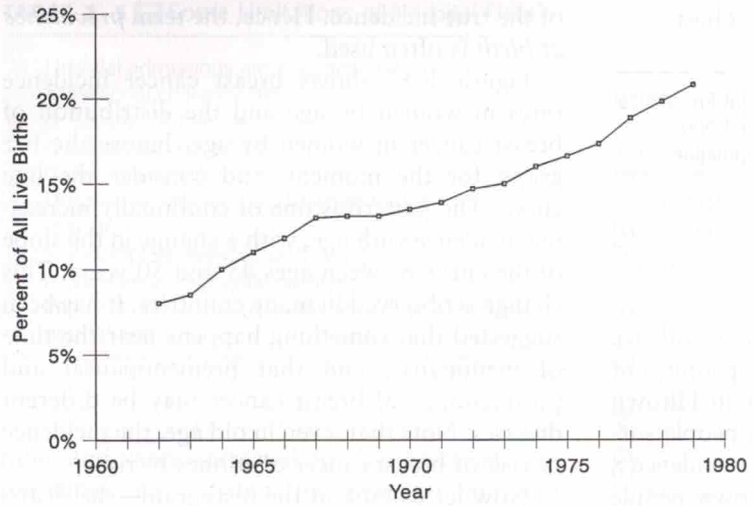

FIGURE 3–15 ▼ Percentage of births that were extramarital in New Zealand, 1962–1979, based on data from the Department of Statistics. (Adapted from Benfield J, Kjellstrom T: New Zealand ex-nuptial births and domestic purposes benefits in a different perspective. N Z Nurs J 74: 28–31, 1981.)

period, and as seen on the map, the cases clustered in the inner city, consistent with the often-made observation that rheumatic fever is strongly associated with low socioeconomic status. It should be pointed out that such a clustering seen on a spot map does not demonstrate a higher incidence in the area of the cluster. For if the population also clusters in this area, the rate in the area of the cluster may be no different than that elsewhere in the city. However, a spot map may offer important clues to disease etiology that can then be pursued with more rigorous studies.

Figure 3–19 shows such a spot map for 1977 to 1981. By this time, the disease had become almost nonexistent in Baltimore, despite the absence of any concerted program aimed at disease eradication.

Clustering, the phenomenon shown by spot maps, is often reported. Residents of a community may report apparent clusters of cancer deaths in children. For example, in Woburn, Massachusetts, a cluster of cases of childhood leukemia was reported and attributed to industrial contamination.[2] This cluster led to action in the courts.[3] However, many apparent clusters are due only to chance, and an important epidemiologic challenge is to investigate such groups of cases and rule out an environmental etiology for what appears to be a greater-than-expected proximity of cases of a disease in time and space.

Surveillance

Surveillance is a fundamental role of public health. Surveillance may be carried out to monitor

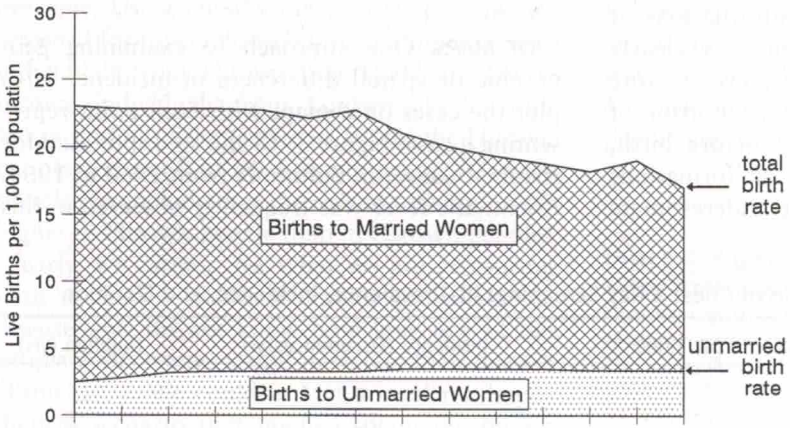

FIGURE 3–16 ▼ Births to married and unmarried women in New Zealand, 1965–1978, based on data from the Department of Statistics. (Adapted from Benfield J, Kjellstrom T: New Zealand ex-nuptial births and domestic purposes benefits in a different perspective. N Z Nurs J 74:28–31, 1981.)

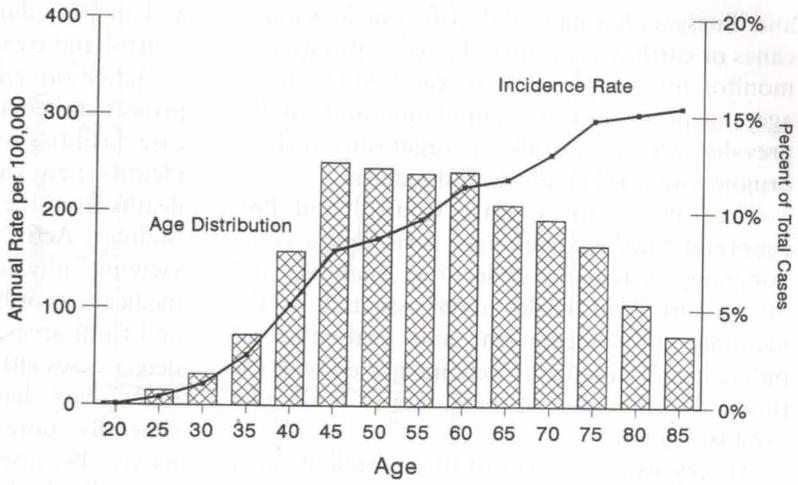

FIGURE 3–17 ▼ Breast cancer incidence rates in white women and distribution of cases by age. (Data from Cutler SJ, Young Jr JL: Third National Cancer Survey: Incidence data. Natl Cancer Inst Monogr 41, 1975.)

changes in disease frequency or to monitor changes in prevalence of risk factors. Much of our information about morbidity and mortality from disease comes from programs of systematic disease surveillance. Surveillance is most frequently conducted for infectious diseases, but in recent years it has become increasingly important in monitoring changes in other types of conditions such as congenital malformations, cancer, asthma, and chemical poisoning and for injuries

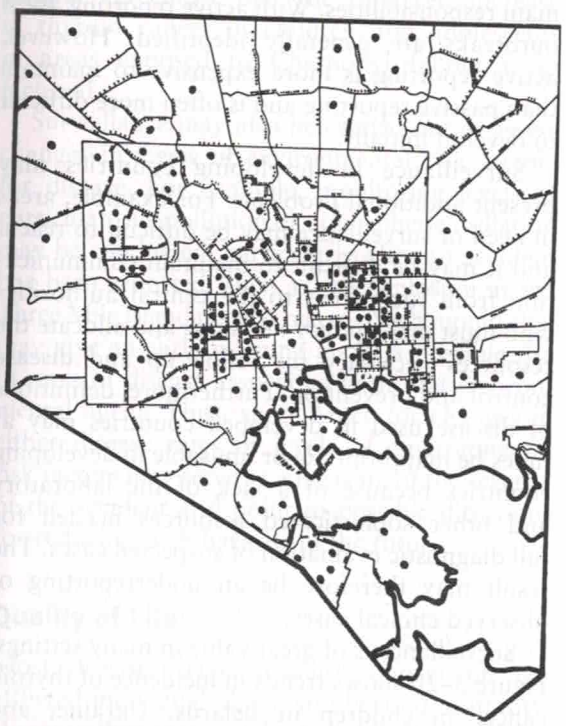

FIGURE 3–18 ▼ Spot map of residence distribution of patients with rheumatic fever, ages 5 to 19 years, hospitalized for first attacks, Baltimore, 1960–1964. (Reprinted from Gordis L, Lilienfeld A, Rodriguez R: Studies in the epidemiology and preventability of rheumatic fever: I. Demographic factors and the incidence of acute attacks. J Chronic Dis 21:645–654, 1969. Copyright 1969, with kind permission from Elsevier Science Ltd.)

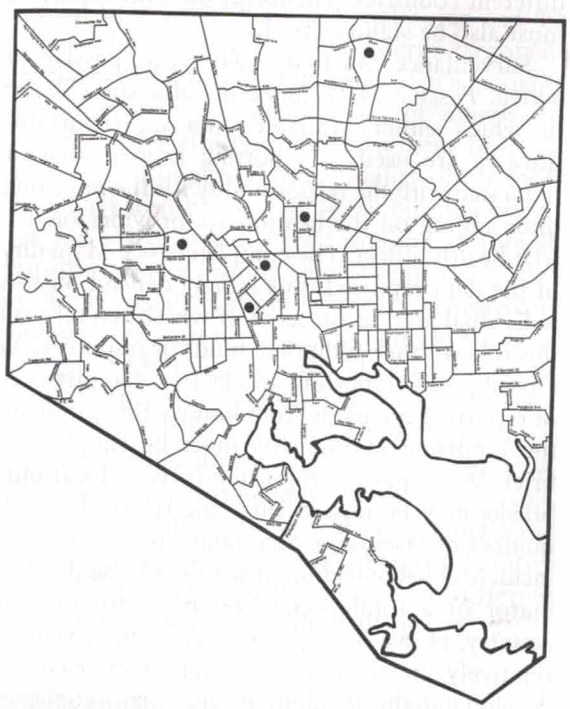

FIGURE 3–19 ▼ Spot map for patients with rheumatic fever, ages 5 to 19 years, hospitalized for first attacks in Baltimore, 1977–1981. (Reproduced with permission. From Gordis L: The virtual disappearance of rheumatic fever in the United States: Lessons in the rise and fall of disease. Circulation 72: 1155–1162, 1985. Copyright 1985, American Heart Association.)

and illnesses after natural disasters such as hurricanes or earthquakes. Surveillance is also used to monitor for completeness of vaccination coverage and protection of a population and for the prevalence of drug-resistant organisms such as drug-resistant tuberculosis and malaria.

The Centers for Disease Control and Prevention defined *epidemiologic surveillance* as the "ongoing systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice closely integrated with the timely dissemination of these data to those who need to know."[4]

An important element of this as well as other definitions of surveillance is providing decision-makers with guidance for developing and implementing the best strategies for programs for disease prevention and control. In order to enable countries or states to develop coordinated public health approaches, mechanisms for information exchange are essential. Consequently, standardized definitions of disease and diagnostic criteria are needed that can be applied in different countries. The forms used for reporting must also be standardized.

Surveillance can be of two types: passive or active. *Passive surveillance* denotes surveillance in which either available data on reportable diseases are used or reporting is mandated or requested with the responsibility for the reporting often falling on the health care provider or district health officer. The completeness and quality of the data reported thus largely depend on this individual and his or her staff who often take on this role without additional funds or resources. As a result, underreporting and lack of completeness of reporting are likely; to minimize this problem, the reporting instruments must be simple and brief. When passive reporting is used, local outbreaks may be missed since the relatively small number of cases often ascertained (numerator for incidence) become diluted within a large denominator of a total population of a province or country. However, a passive reporting system is relatively inexpensive and relatively easy to develop initially. In addition, since many countries have systems of passive reporting for a number of reportable diseases that are generally infectious, passive reporting allows for international comparisons that can identify areas which urgently need assistance in confirming new cases and in providing appropriate interventions for control and treatment.

Active surveillance denotes a system in which project staff make periodic field visits to health care facilities such as clinics and hospitals to identify new cases of a disease or diseases or deaths from the disease that have occurred (case finding). Active surveillance may involve interviewing physicians and patients, reviewing medical records, and in developing countries and rural areas, surveying villages and towns to detect cases either on a routine basis or after an index case has been reported. Reporting is generally more accurate when it is active than passive because active surveillance is conducted by individuals who have been specifically employed to carry out this responsibility. This is in contrast to passive surveillance in which those expected to report new cases are often overburdened by their primary responsibilities of providing health care and administering health services. For them, filing reports is an additional burden that they often view as peripheral to their main responsibilities. With active reporting, local outbreaks are generally identified. However, active reporting is more expensive to maintain than passive reporting and is often more difficult to develop initially.

Surveillance in developing countries may present additional problems. For example, areas in need of surveillance may be difficult to reach, and it may be difficult to maintain communication from such areas to the central authorities who must make policy decisions and allocate the resources necessary for follow-up and disease control and prevention. Furthermore, definitions of disease used in developed countries may at times be inappropriate or unusable in developing countries because of a lack of the laboratory and other sophisticated resources needed for full diagnostic evaluation of suspected cases. The result may therefore be an underreporting of observed clinical cases.

Surveillance is of great value in many settings. Figure 3–20 shows trends in incidence of thyroid cancer in children in Belarus, Ukraine, and Russia from 1986 to 1994 following the explosion in the Chernobyl reactor.[5] The highest incidence rates were found in the most contaminated areas—Gomel in southern Belarus and parts of northern Ukraine. However, a problem in interpreting such data is the possibility that

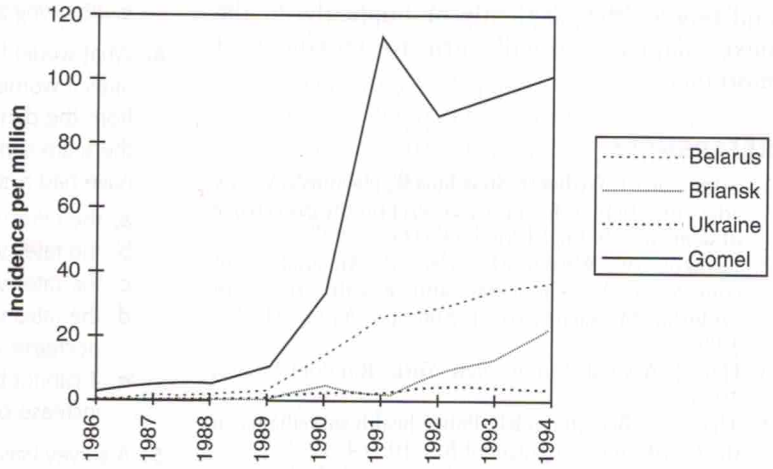

FIGURE 3–20 ▼ Trends of incidence of childhood thyroid cancer in Belarus, Ukraine, and Russia, 1986–1994. (From Bard D, Verger P, Hubert P: Chernobyl, 10 years after: Health consequences. Epidemiol Rev 19:187–204, 1997.)

the observed increase could be due to intensive screening following the accident, which could have identified tumors that would otherwise not have been detected. However, there is now general agreement that the observed increase in thyroid cancer in children and adolescents in areas exposed to Chernobyl fallout is in fact real.

Surveillance may also be carried out to assess changes in levels of environmental risk factors for disease. For example, monitoring levels of particulate air pollution or atmospheric radiation may be conducted, particularly after an accident has been reported, such as the explosion at the Three Mile Island nuclear reactor. Such monitoring may give an early warning about a possible rise in rates of disease associated with that environmental agent. Thus, surveillance for changes in either disease rates or levels of environmental risk factors may serve as a measure of the severity of the accident and point to possible directions for reducing such hazards in the future.

Quality of Life

Most diseases have a major impact on the afflicted individual above and beyond mortality. Diseases that may not be lethal may be associated with considerable suffering and disability. For this reason, it is also important to consider the impact of a disease as measured by its effect on a person's quality of life, even though such measures are not, in fact, measures of disease occurrence. For example, it is possible to examine the extent to which patients with arthritis are compromised by the illness in carrying out activities of daily living. Although considerable controversy exists about which quality-of-life measures are most appropriate and valid, there is general agreement that such measures can be reasonably used to plan short-term treatment programs for groups of patients. Such patients can be evaluated over a period of months to determine the effects of the treatment on their self-reported quality of life. Quality-of-life measures have also been used for establishing priorities for scarce health care resources. Although prioritization of health care resources is often primarily based on mortality data, because many diseases are chronic and non-life-threatening, quality of life must also be taken into account for this purpose. Patients may place different weights on different quality-of-life measures depending on cultural background, education, and, for example, religious values. As a result, measuring quality of life and developing valid indices that are useful for obtaining comparative data in different patients and in different populations remain a major challenge.

Conclusion

In this chapter, we have reviewed different approaches to measuring morbidity. We have seen that a rate involves specification of a numerator, a denominator of people at risk,

and time—either explicitly or implicitly. In the next chapter, we will turn to measures of mortality.

REFERENCES

1. Erkinjuntti T, Østbye T, Steenhuis R, Hachinski V: The effect of different diagnostic criteria on the prevalence of dementia. N Engl J Med 337:1667, 1997.
2. Lagakos SW, Wessen BJ, Zelen M: An analysis of contaminated well water and health effects in Woburn, Massachusetts. J Am Stat Assoc 81:583, 1986.
3. Harr J: A Civil Action. New York, Random House, 1995.
4. Thacker S, Berkelman RL: Public health surveillance in the United States. Epidemiol Rev 10:164, 1988.
5. Bard D, Verger P, Hubert P: Chernobyl, 10 years after: health consequences. Epidemiol Rev 19:187, 1997.

REVIEW QUESTIONS FOR CHAPTER 3

1. At an initial examination in Oxford, Mass., migraine headache was found in 5 of 1,000 men aged 30 to 35 years and in 10 of 1,000 women aged 30 to 35 years. The inference that women have a two times greater risk of developing migraine headache than do men in this age group is:

- a. correct
- b. incorrect, because a ratio has been used to compare male and female rates
- c. incorrect, because of failure to recognize the effect of age in the two groups
- d. incorrect, because no data for a comparison or control group are given
- e. incorrect, because of failure to distinguish between incidence and prevalence

2. A prevalence survey conducted from January 1st through December 31st, 2003 identified 1,000 cases of schizophrenia in a city of 2 million persons. The incidence rate of schizophrenia in this population is 5 per 100,000 persons each year. What percent of the 1,000 cases were newly diagnosed in 2003? _____

3. Which of the following is an advantage of active surveillance?

- a. requires less project staff
- b. is relatively inexpensive to employ
- c. more accurate due to reduced reporting burden for health care providers
- d. relies on different disease definitions to account for all cases
- e. reporting systems can be developed quickly

4. What would be the effect on age-specific incidence rates if women with hysterectomies were excluded from the denominator of calculations, assuming that there are some women in each age group who have had hysterectomies?

- a. the rates would remain the same
- b. the rates would tend to decrease
- c. the rates would tend to increase
- d. the rates would increase in older groups and decrease in younger groups
- e. it cannot be determined whether the rates would increase or decrease

5. A survey was conducted among the non-hospitalized adult population of the United States during 1988 through 1991. The results from this survey are shown below.

Age Group	Percent of Persons with Hypertension
18–29 years	4
30–39 years	10
40–49 years	22
50–59 years	43
60–69 years	54
70 and older	64

The researchers stated that there was an age-related increase in the risk of hypertension in this population. You conclude that the researchers' interpretation:

- a. is correct
- b. is incorrect because it was not based on rates
- c. is incorrect because incidence rates do not describe risk
- d. is incorrect because prevalence is used
- e. is incorrect because the calculations are not age-adjusted

Questions 6 and 7 use the information below:

Population of the city of Atlantis on March 30th, 2003 = 183,000

Number of new active cases of tuberculosis (TB) occurring between January 1st and June 30th, 2003 = 26

Number of active TB cases according to the city register on June 30th, 2003 = 264

6. The incidenc
6-month per
a. 7 per 100
b. 14 per 10
c. 26 per 10
d. 28 per 10
e. 130 per 1

6. The incidence rate of active cases of TB for the 6-month period was:

a. 7 per 1000,000 population
b. 14 per 100,000 population
c. 26 per 100,000 population
d. 28 per 100,000 population
e. 130 per 100,000 population

7. The prevalence rate of active TB as of June 30th, 2003 was:

a. 14 per 100,000 population
b. 130 per 100,000 population
c. 144 per 100,000 population
d. 264 per 100,000 population
e. none of the above