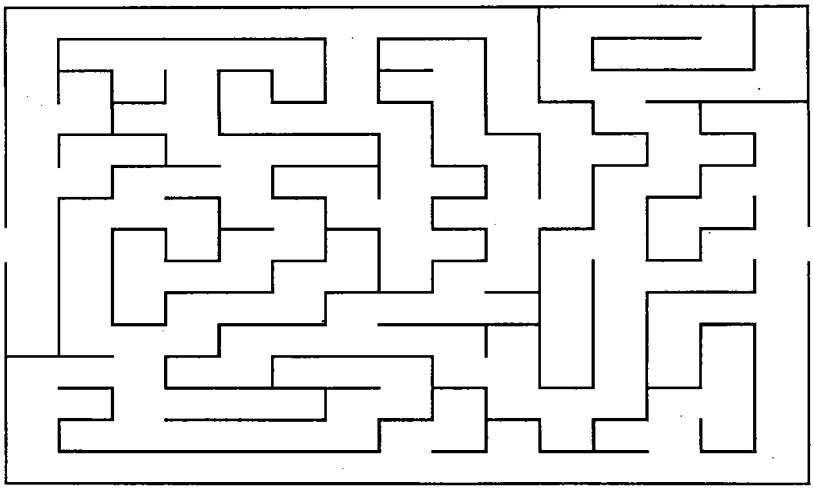
1

Disease Frequency



The building blocks of epidemiology are rooted in experience: we "know" what disease is, we know what a toxic exposure might be, and we can count noses. The language of epidemiology is full of terms from everyday speech, terms that have both formal meanings and all the connotations that the same words carry in general use. The associations with day-to-day ideas can be helpful in spirit, but they can be misleading in particulars. The purpose of this first chapter is therefore to define and illustrate, with an emphasis on the components of observation in a single population. Subsequent chapters will deal with comparisons, methods for structuring observation, and the barriers that stand between observation and inference.

Cross-Sectional Measures

The simplest population measure of disease burden is prevalence.

Prevalence. *The prevalence of a characteristic in a population is the fraction of individuals in the population who possess the characteristic.*

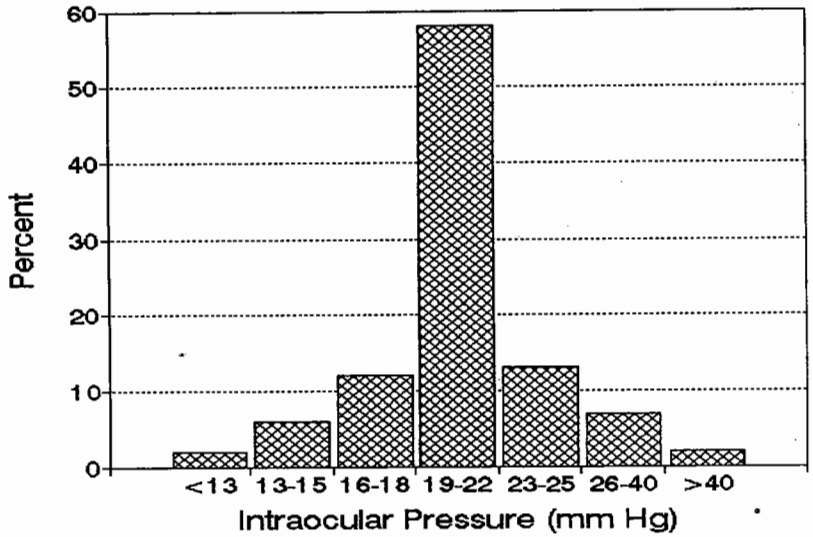


Figure 1.1 The prevalence of intraocular pressure readings in diabetics attending the Joslin Clinic in 1935

A prevalence is a status report. Since most characteristics of any interest vary with time as well as across populations, any mention of a prevalence ought to be accompanied by a specification of whom and of when. Figure 1.1 graphs the prevalence of various levels of intraocular pressure (IOP) among 2,002 diabetics over the age of 20 seen at the Joslin Clinic (in Boston) in the years 1925 through 1934.[1] Most of the readings summarized in Figure 1.1 were normal, and the data were interpreted as indicating that diabetics do not as a group suffer from intraocular hypertension. In the non-diabetic population, however, about five percent of adults have IOPs higher than 22 millimeters of mercury (mm Hg). On the basis of Figure 1.1, the

1. Waite JH, Beetham WP. N Engl J Med 1935;212:367-369

prevalence of IOPs greater than 22 mm Hg appears to have been about 20 percent. The authors' opinion notwithstanding, diabetics at the Joslin Clinic had an excess prevalence of intraocular hypertension.

Table 1.1 Use of NSAIDs, acetaminophen, and antacids two years prior to a first prescription for cimetidine

	Subsequent Users of Cimetidine N = 1327	Age-Matched Subsequent Non-Users N = 5308
Any NSAID	338 (25%)	907 (17%)
Acetaminophen	382 (29%)	906 (17%)
Antacids	521 (39%)	889 (17%)

Prevalence can be of special usefulness in assessing the disease burden of a community and in projecting demands for medical services. The annual cost, *per capita* of the general population, of caring for persons with AIDS can be estimated by multiplying the prevalence of AIDS by $25,000.

Table 1.1 gives the prevalence of use of several analgesics in persons aged 65 and older in two populations.[2] The first population consisted of people who would be diagnosed two years later as having significant peptic ulcer disease (indicated by the receipt of a first prescription for the anti-ulcer drug cimetidine[3]), the second population consisted of persons of similar age and sex who would have no such diagnosis. Are these prevalences compatible with the idea that there is no relation between the use of these drugs and peptic ulcer disease (PUD)? Even without formal tests of statistical significance, it seems scarcely credible that the differences in Table 1.1 could be ascribable to chance. One possibility is that the drugs considered *cause* PUD. Alternatively, it may be that the drugs are being used to treat early symptoms of PUD, even in advance of a

2. The study from which these data are drawn is described in: Hernandez Avila M, Walker AM, Romieu I, et al. Choice of nonsteroidal anti-inflammatory drug in persons treated for dyspepsia. Lancet 1988;ii:556-9

3. At present, use of cimetidine would be a very nonspecific indicator of peptic ulcer disease. In the late 1970s, when the data of Table 1.1 were collected, the connection was thought to be much closer.

specific diagnosis. The connection is evident in the case of antacids, but holds for the analgesics (nonsteroidal anti-inflammatory drugs -- "NSAIDs" -- and acetaminophen) as well. Persons with early, undiagnosed PUD may have nonspecific pains that are incorrectly diagnosed as arthritic, for which analgesics are prescribed. This supposition is strengthened particularly by the increased use of acetaminophen, for which (unlike the NSAIDs) there is little non-epidemiologic evidence of gastrointestinal toxicity.

Measures that Incorporate Fixed Intervals of Time

Prevalence does not capture the concept of elapsed time, and it offers no information about *transitions* between states of health and disease. For the resource planner, knowing the prevalent number of ill people may be enough, but for both the patient and the physician, the transition from one health state to another is a key event.

One way to examine transition probabilities is to string together a series of prevalences over time. Assuming that the entire population remains under observation, the change between prevalences at successive times is a measure of the net rate of transition into the morbid state. If more people get sick than get well, then the prevalence rises. If the state is irreversible, then the difference in successive prevalences is related to the number of new or *incident* cases.

Incident. *A case of disease is said to be "incident" at the moment at which the disease manifests signs or symptoms. Incident cases are newly occurring cases.*

The definition of an incident case depends on the current technical ability to recognize disease. The distinction of times of onset is important, but often approximate, particularly in the case of chronic diseases. Rather than wait for an unimpeachable definition of disease onset, most investigators proceed with the understanding that the operative definition may change, and that conclusions may have to be revised.

Even with reversible conditions, prevalences can be calculated at successive points in time as if the condition were irreversible. That is, people go from the state "has never suffered event X" to the state "has suffered event X at some time in the past." In the absence of loss to follow-up among the study subjects, prevalences obtained by this convention are measures of the *cumulative incidence*.[4]

Cumulative incidence. *The cumulative incidence from time t_0 to time t_1 for event X is the prevalence of "history of X" at time t_1 among all those persons who began observation at time t_0 and did not possess a "history of X" at time t_0.*

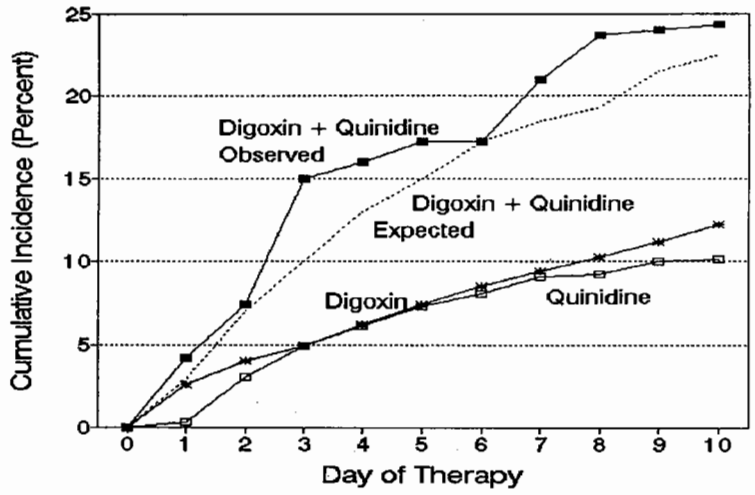


Figure 1.2 Cumulative incidence of symptoms of digitalis intoxication in patients receiving digoxin and/or quinidine

Figure 1.2[5] gives the cumulative incidence of signs and symptoms of digitalis intoxication in hospitalized patients receiving digoxin

4. Other terms for cumulative incidence include "incidence proportion" and "risk." The former provides a neat linguistic tie to "incidence rate," defined later; the latter emphasizes the connection to probability of disease; the principal drawback of the term "risk" is that the same word can be used to denote both observed events and the underlying forces of morbidity that generated the events.

5. Walker AM, Cody RJ, Greenblatt DJ, Jick H. Drug toxicity in patients receiving digoxin and quinidine. Am Heart J 1983;105:1025-8

alone and digoxin plus quinidine (instituted on Day 0). Also shown are the cumulative incidences of clinically similar events in patients receiving quinidine alone, and the sum of the curves for digoxin alone and quinidine alone. Administration of quinidine to patients receiving digoxin increases serum digoxin levels through displacement of digoxin from albumin and from tissue binding sites, and the expectation was that concurrent therapy would increase the risk for digoxin side effects. The observed digoxin-with-quinidine curve is higher than the curves for the drugs separately, but only slightly above the sum of the two.[6] Elevated blood levels notwithstanding, there is little evidence here for a clinical effect attributable specifically to the concurrent administration of digoxin and quinidine.

The cumulative incidence over each of the one-day intervals of Figure 1.2 might be referred to as a "daily cumulative incidence," and could be calculated as each day's number of incident cases divided by the number of persons who had not yet developed digitalis intoxication at the beginning of the day. The curve labeled "daily incidence" in Figure 1.3 presents the daily cumulative incidence of bleeding in patients receiving heparin therapy.[7]

Complementary to the cumulative incidence is the measure called *survival.*

Survival *is the complement of disease occurrence over a time interval. The observed survival is 1 minus the cumulative incidence of disease.*

Survivals in successive time periods can be multiplied together to obtain a cumulative survival. The cumulative incidence curves of

6. The method of comparison illustrated in Figure 1.2 is valid for the data shown, but should be generalized only with some caution because the addition of curves to give the digoxin-plus-quinidine "expected" curve overstates the expectation. First, from elementary probability theory, the combination of P_D, the probability of an adverse reaction as a result of digoxin, and P_Q, the corresponding probability for quinidine, to form P_{D+Q}, should not be $P_D + P_Q$ but rather $1-(1-P_D)(1-P_Q)$. Second, if there is any background probability of occurrence of something that might be misdiagnosed as digitalis toxicity, i.e. if signs or symptoms consistent with digitalis intoxication can be expected to occur in people receiving neither digitalis nor quinidine, then that background probability, P_0, has been counted twice when the curves are added. The true expected cumulative incidence curves would be given by $P_{D+Q+0} = 1-(1-P_D)(1-P_Q)(1-P_0)$, where the Ps refer to the probabilities uniquely associated with digoxin, quinidine, and background. In the present example, in which the component probabilities are not large, the true expected curve is close to the simple approximation graphed in Figure 1.2.

7. Walker AM, Jick H. Predictors of bleeding during heparin therapy. JAMA 1980;244:1209-1212

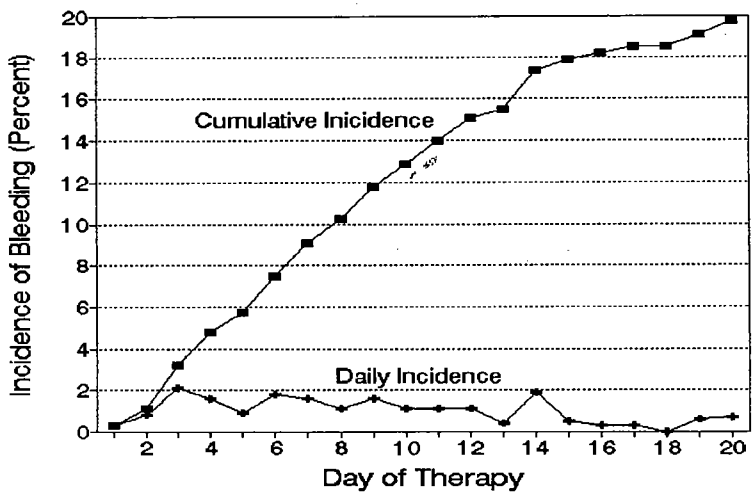


Figure 1.3 Cumulative incidence and daily incidence of bleeding in patients receiving heparin

Figures 1.2 and 1.3 were derived by calculating daily survivals, multiplying these to arrive at cumulative survivals, and subtracting the result from 1.

Measures for Variable Observation Times

The definition of cumulative incidence assumes an opportunity to observe a group of persons from beginning to end of a time interval. More commonly, individuals possess characteristics that define their class membership for variable periods of time. The occurrence of disease is then measured not over a fixed interval, but over whatever intervals are eligible for observation from each person under study. The intervals are collectively called the *person time* of observation of a population.

Person time *is the time during which a single individual meets all the definitions for inclusion in a study, and during which any disease event occurring in the individual would be known. The person time of observation in a population is the sum of the person times contributed by all the members of the population.*

There are three equivalent methods for calculating the person time of observation of a population under study.

(1) For each person, identify the amount of time contributed to the group's observation, then sum the times of the individual persons to get person time.

(2) Multiply the number of persons under observation by the average duration of observation per person.

(3) Multiply the length of the period of observation by the average number of persons under observation during the period.

Unlike "persons," who are discrete and easily imagined, "person time" is a continuous quantity that proves difficult for most people to intuit. To visualize person time, think of a very small period of a person's experience, such as a day or an hour, as a discrete unit of observation. Each study subject contributes some number of units to a grand pool of observation.

A measure of disease frequency can be obtained by dividing the number of events that are observed among eligible population members by the total person time of observation.

Incidence rate. *The incidence rate of an event in a pool of person time is the number of events observed divided by the amount of person time observed.*

The general method of calculating an incidence rate in a specified population group during a given period of time has three parts:

(1) sum the number of cases that occur among members of the population during the time period in question;

(2) calculate the amount of person time of observation in the population for the time period;

(3) divide the number of cases by the person time of observation.

An incidence rate can be thought of as the fraction becoming ill, adjusted for length of follow-up. To see this algebraically, look at the second method for obtaining the person time of observation. With a little rearrangement, the procedure for calculating the incidence rate could be laid out as follows: divide the number of cases by the number of people observed to get the fraction becoming ill; next divide the fraction becoming ill by the average duration of observation to obtain the incidence rate.

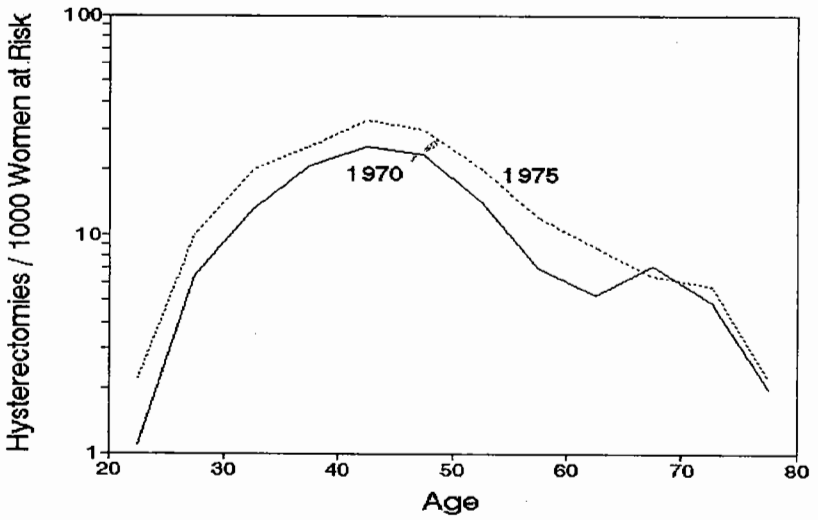


Figure 1.4 Incidence rate of hysterectomy among women with an intact uterus, by United States Census Region, 1970

To justify an incidence rate calculation as a sensible procedure the analyst has to assume that the incidence rate is nearly constant throughout the pool of person time observed. If the incidence rate appears to vary over calendar time or within subgroups of a population of interest, then the time period of observation and the population must be split into subcategories, within which the assumption of constancy nearly holds. A series of appropriately labeled rates for the categories of person time, presented in tabular or graphic form, then serves to characterize the disease process in the population.

Figure 1.4 shows the incidence rate of hysterectomy per 1000 woman years at risk for women of various ages in the United States in 1970 and 1975.[8] The person times of observation were obtained using the third person-time method above. The rate was calculated for each age group in each of the two calendar years by estimating the number of women who had a uterus (the population at risk),

8. Walker AM, Jick H. Temporal and regional variation in hysterectomy rates in the United States, 1970-1975. Am J Epidemiol 1979;110:41-6

multiplying that figure by one year (yielding woman years at risk in 1970 or 1975), and dividing the resulting number into an estimate of the number of hysterectomies performed.

In 1970, hysterectomy rates in the 65-69 year age group were higher than those in the immediately young or older age groups. One interpretation of the 1970 curve in Figure 1.4 is that there is a discontinuity at age 65 in what should properly be a picture more like that seen five years later, a smooth curve with a single peak at 40-44 years. In this light, the high value among women in their late 60s in 1970 is just the leading figure in an elevation of the whole post-65 year portion of the age-incidence curve. Medicare, the U.S. government's program of payment for medical care to the elderly, was introduced in 1967. According to cynics observing the U.S. medical scene in the early 1970s, the two principal indications for hysterectomy at the time were a uterus and the ability to pay. The secondary peak in 1970 may have represented "catch-up" procedures in patients who had newly acquired the second indication.

Some characteristics of the measures of disease presented so far are listed in Table 1.2 at the end of this chapter. The information below the solid line refers to material presented in later chapters, but is included here for reference.

Duration of Disease

The relation between the incidence rate of a disease and its prevalence involves the *duration* of disease.

Duration. *The duration of an illness is the length of the time interval that elapses from first manifestation of disease until complete resolution. For an irreversible disease process, duration is the length of the interval from first manifestation to death.*

A disease with a long duration may have a relatively high prevalence even if the disease has a low incidence. Multiple sclerosis (MS) has an overall incidence rate in the northern part of the United States of around 3 cases per 100,000 person years, less than one-tenth the incidence of cancer of lung, which is about 40 cases per 100,000 person years. Yet the prevalence of MS is much higher than that of cancer of the lung. The prevalence of MS is about 75 cases per 100,000 persons, versus about 40 cases per 100,000 persons for lung cancer. The MS prevalence is actually quite close to that of streptococcal pharyngitis, a commonly occurring disease (the incidence rate is on the order of 10,000 cases per 100,000 person years) that has a short duration. The discrepancy among the incidence rates and prevalences of these illnesses are accounted for by their different mean durations. The duration of MS is on the order of 25 years, that of lung cancer one year, and of symptomatic strep throat three days.

The interplay between the epidemiologic features of disease has a straightforward algebraic expression. When the prevalence is less than 10 percent (10,000 cases per 100,000 persons), the steady state relation between prevalence (Pr), incidence rate (IR), and mean duration $\overline{D}$ is

$$Pr = (IR)\overline{D}$$

The qualifier "steady state" imposes a number of strong restrictions on this relation. Newborns and immigrants (all healthy) must balance exactly the number without disease who die or emigrate, the number of diseased persons who immigrate must equal the number of diseased persons who emigrate, and the number of persons who become diseased per unit time must exactly equal the number of disease terminations, which may occur through a return to health or through death.

Although no free-living population is likely to meet the steady state criteria, the qualitative relation embodied in the preceding equation applies widely. A study of HLA types (a class of genetic markers) among children with acute lymphocytic leukemia (ALL) who attended an oncology clinic found that the prevalence of type A2 was higher than that in the general population.[9] The observation raised considerable interest, implying as it did that susceptibility to acute leukemia might be mediated by genetic factors. A follow-up study of a series of newly diagnosed leukemics found identical prevalences of the "high risk" type A2 in patients and in the general population.[10] The discordance between the two findings was due to an effect of HLA type on the mean duration of ALL. Far from being at high risk of ALL, children with HLA type A2 were at no

9. Rogentine GN, Yankee RA, Gart JJ, et al. HLA antigens and disease: acute lymphocytic leukemia. J Clin Invest 1972;51:2420-8. I am grateful to Philip Cole for pointing out this example in his chapter of introduction to Volume 1 of Statistical Methods in Cancer Research, The Analysis of Case-Control Studies, by N.E. Breslow and N.E. Day, International Agency for Research on Cancer, Lyon, 1980.

10. Rogentine GN, Trapani RJ, Yankee RA, Henderson ES. HLA antigens and acute lymphocytic leukemia: the nature of the association. Tissue Antigens 1973;3:470-6

increased risk, responded better to chemotherapy, had longer survivals, and were therefore overrepresented in the (prevalent) clinic population. The lesson is that if you want to study the determinants of incidence rate, you need incident rather than prevalent cases of disease.

A more general relation between prevalence, incidence rate, and duration, which holds for any prevalence, can be derived by observing that (a) if the total population size is N then the population at risk for becoming diseased is $(1-Pr)N$ and the population at risk for leaving the diseased state is PrN, and (b) the rate of leaving the diseased state equals the reciprocal of the mean duration of the diseased state. The equation to be solved becomes "number entering diseased state in time interval Δt = number leaving diseased state in Δt." That is

$$IR(1-Pr)N\Delta t = \frac{1}{D}Pr\,N\Delta t$$

From which

$$\frac{Pr}{1-Pr} = (IR)D$$

Table 1.2
Measures of Disease Occurrence

Measure	**Prevalence**	**Cumulative incidence**	**Survival**	**Incidence rate**
Incorporation of time	None	Specified follow-up	Specified follow-up	Incremental
Dimension	Cases per population	Cases per population after specified time	Nondiseased per population after specified time	Cases per population per unit time
Minimum Value	0	0	0	0
Maximum value	1	1	1	infinity
Parameter	Probability	Probability	Probability	Hazard
Probability Distribution	Binomial	Binomial	Binomial	Poisson
Variance	$\frac{Pr(1-Pr)}{N}$	$\frac{CI(1-CI)}{N}$	$\frac{S(1-S)}{N}$	$\frac{IR}{P}$