2 Frequency

Here, it is necessary to count.—**P.C.A. Louis**[†]
1787–1872

Key Words

Numerator
Denominator
Prevalence
Point prevalence
Period prevalence
Incidence
Duration of disease
Case fatality rate
Survival rate
Complication rate
Infant mortality rate
Perinatal mortality rate
Prevalence studies
Cross-sectional studies
Surveys
Cohort
Cohort studies
Cumulative incidence
Incidence density
Person-time
Dynamic population
Population at risk
Random sample
Probability sample
Sampling fraction
Oversample
Convenience samples
Grab samples
Epidemic
Pandemic
Epidemic curve
Endemic

Chapter 1 outlined the questions that clinicians need to answer as they care for patients.

Answers are usually in the form of probabilities and only rarely as certainties. Frequencies obtained from clinical research are the basis for probability estimates for the purposes of patient care. This chapter describes basic expressions of frequency, how they are obtained from clinical research, and how to recognize threats to their validity.

Example

A 72-year-old man presents with slowly progressive urinary frequency, hesitancy, and dribbling.

A digital rectal examination reveals a symmetrically enlarged prostate gland and no nodules. Urinary flow measurements show a reduction in flow rate, and his serum prostate-specific antigen (PSA) is not elevated. The clinician diagnoses benign prostatic hyperplasia (BPH). In deciding on treatment, the clinician and patient must weigh the benefits and hazards of various therapeutic options. To simplify, let us say the options are medical therapy with drugs or surgery. The patient might choose medical treatment but runs the risk of worsening symptoms or obstructive renal disease because the treatment is less immediately effective than surgery. Or he might choose surgery, gaining immediate relief of symptoms but at the risk of operative mortality and long-term urinary incontinence and sexual dysfunction.

Decisions such as the one this patient and clinician face have traditionally relied on clinical judgment based on experience at the bedside and in the clinics. In modern times, clinical research has become sufficiently strong and extensive that it is possible to ground clinical judgment in research-based probabilities—frequencies. Probabilities of disease, improvement, deterioration, cure, side effects, and death are the basis for answering most clinical questions. For this patient, sound clinical decision making requires accurate estimates of how his symptoms and complications of treatment will change over time according to which treatment is chosen.

ARE WORDS SUITABLE SUBSTITUTES FOR NUMBERS?

Clinicians often communicate probabilities as words (e.g., usually, sometimes, rarely) rather than as

[†]A 19th Century physician and proponent of the "numerical method" (relying on counts, not impressions) to understand the natural history of diseases such as typhoid fever.

numbers. Substituting words for numbers is convenient and avoids making a precise statement when one is uncertain about a probability. However, words are a poor substitute for numbers because there is little agreement about the meanings of commonly used adjectives describing probabilities.

Example

Physicians were asked to assign percentage values to 13 expressions of probability (1). These physicians generally agreed on probabilities corresponding to adjectives such as "always" or "never" describing very likely or very unlikely events but not on expressions associated with less extreme probabilities. For example, the range of probabilities (from the top to the bottom tenth of attending physicians) was 60% to 90% for "usually," 5% to 45% for sometimes, and 1% to 30% for "seldom." This suggests (as authors of an earlier study had asserted) that "difference of opinion among physicians regarding the management of a problem may reflect differences in the meaning ascribed to words used to define probability" (2).

Patients also assign widely varying probabilities to word descriptions. In another study, highly skilled and professional workers outside of medicine thought "usually" referred to probabilities of 35% to 100%; "rarely" meant to them a probability of 0% to 15% (3).

Thus, substituting words for numbers diminishes the information conveyed. Use numbers whenever possible.

PREVALENCE AND INCIDENCE

In general, clinically relevant measures of frequency are expressed as proportions, in which the **numerator** is the number of patients experiencing an event (cases) and the **denominator** is the number of people in whom the event could have occurred (population). The two basic measures of frequency are prevalence and incidence.

Prevalence

Prevalence is the fraction (proportion or percent) of a group of people possessing a clinical condition or outcome at a given point in time. Prevalence is measured by surveying a defined population and counting the number of people with and without the condition of interest. **Point prevalence** is measured at a single point in time for each person (although actual measurements need not necessarily be made at the same point in calendar time for all the people in the population). **Period prevalence** describes cases that were present at any time during a specified period of time.

Incidence

Incidence is the fraction or proportion of a group of people initially free of the outcome of interest that develops the condition over a given period of time. Incidence refers then to *new* cases of disease occurring in a population initially free of the disease or new outcomes such as symptoms or complications occurring in patients with a disease who are initially free of these problems.

Figure 2.1 illustrates the differences between incidence and prevalence. It shows the occurrence of lung cancer in a population of 10,000 people over the course of 3 years (2018–2020). As time passes, individuals in the population develop the disease. They remain in this state until they either recover or die—in the case of lung cancer, they usually die. Four people already had lung cancer before 2018, and 16 people developed it during the 3 years of observation. The rest of the original 10,000 people have not had lung cancer during these 3 years and do not appear in the figure.

To calculate prevalence of lung cancer at the beginning of 2018, four cases already existed, so the

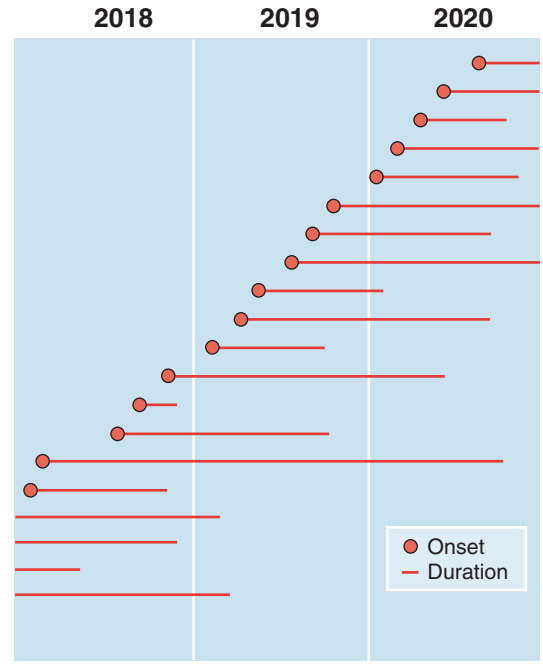

Figure 2.1. Incidence and prevalence. Occurrence of disease in 10,000 people at risk for lung cancer, 2018 to 2020.

TABLE 2.1

Characteristics of Incidence and Prevalence

Characteristic	Incidence	Prevalence
Numerator	New cases occurring during a period of time among a group initially free of disease	Existing cases at a point or period of time
Denominator	All susceptible people without disease at the beginning of the period	All people examined, including cases and noncases
Time	Duration of the period	Single point or period
How measured	Cohort study (see Chapter 6)	Prevalence (cross-sectional) study

prevalence at that point in time is 4/10,000. If all surviving people are examined at the beginning of each year, one can compute the prevalence at those points in time. At the beginning of 2019, the prevalence is 5/9,996 because two of the pre-2018 patients are still alive, as are three other people who developed lung cancer in 2018; the denominator is reduced by the four patients who died before 2019. Prevalence can be computed for each of the other two annual examinations and is 7/9,992 at the beginning of 2020 and 5/9,986 at the beginning of 2021.

To calculate the incidence of new cases developing in the population, we consider only the 9,996 people free of the disease at the beginning of 2018 and what happens to them over the next 3 years. Five new lung cancers developed in 2018, six developed in 2019, and five additional lung cancers developed in 2020. The 3-year incidence of the disease is all new cases developing in the 3 years (16) divided by the number of susceptible individuals at the beginning of the follow-up period (9,996), or 16/9,996 in 3 years. What are the annual incidences for 2018, 2019, and 2020? Remembering to remove the previous cases from the denominator (they are no longer at risk of developing lung cancer), we would calculate the annual incidences as 5/9,996 in 2018, 6/9,991 in 2019, and 5/9,985 in 2020.

Prevalence and Incidence in Relation to Time

Every measure of disease frequency necessarily contains some indication of time. With measures of prevalence, time is assumed to be instantaneous, as in a single frame from a motion picture film. Prevalence depicts the situation at that point in time for each patient, even though it may, in reality, have taken several months to collect observations on the various people in the population. However, for incidence, time is the interval during which susceptible people were observed for the emergence of the event of interest. Table 2.1 summarizes the characteristics of incidence and prevalence.

Why is it important to know the difference between prevalence and incidence? Because they answer two entirely different questions: on the one hand, "What proportion of a group of people has a condition?"; and on the other, "At what rate do new cases arise in a defined population as time passes?" The answer to one question cannot be obtained directly from the answer to the other.

RELATIONSHIPS AMONG PREVALENCE, INCIDENCE, AND DURATION OF DISEASE

Anything that increases the duration of disease increases the chances that the patient will be identified in a prevalence study. Another look at Figure 2.1 will confirm this. Prevalent cases are those that remain affected; to the extent that patients are cured, die of their disease, or leave the population under study, they are no longer a case in a prevalence survey. As a result, diseases of brief duration will be more likely to be missed by a prevalence study. For example, 15% of all deaths from coronary heart disease occur outside the hospital within an hour of onset and without prior symptoms of heart disease. A prevalence study would, therefore, miss nearly all these events and underestimate the true burden of coronary heart disease in the community. In contrast, diseases of long duration are well represented in prevalence surveys, even when their incidence is low. The incidence of ulcerative colitis in North America is only about 2 to 19 per 100,000/year, but its prevalence is much higher, 37 to 249/100,000, reflecting the chronic nature of the disease (4).

The relationship among incidence, prevalence, and **duration of disease** in a steady state, in which