Epidemiology
Chapter 1
Measures of Disease Frequency

Learning Objectives

1. Measures of disease frequency:
   a. Clarify the significance of a particular health problem
   b. Help guide resource allocation
   c. Provide basic insight into the pathogenesis of disease

2. Point prevalence describes the amount of disease at a particular point in time.
3. Incidence describes the number of new cases of disease that develop over time.
4. Incidence may be expressed as incidence proportion or incidence rate.
5. Incidence rate accounts for follow-up time
6. Measures of disease frequency can be stratified, or broken up, by person, place, or time characteristics to gain insight into a disease process.

In December 1998, a 55-year-old woman presented to her local emergency department complaining of profound weakness and difficulty walking. She first noticed pain and weakness in her shoulders about 7 days earlier. The weakness progressed to involve her thigh muscles; she then developed nausea and noticed that her urine appeared dark. Over the next 48 h, her weakness further intensified and she became unable to stand on her own power.

Her previous medical conditions included high blood pressure, asthma, and high serum cholesterol levels. Her father had died at an early age from heart disease. She did not smoke cigarettes and rarely drank alcohol. Her regular medications included aspirin, diltiazem, and cerivastatin. She first started taking cerivastatin 2 weeks earlier to treat high cholesterol levels.

She appeared ill. There was no fever, the blood pressure was 140/95 mmHg, and the pulse was 48 beats/min. She was unable to raise her hips or her shoulders against gravity and her quadriceps muscles were diffusely tender. The rest of her physical examination, including neurological function, was normal.

The urine was dark amber in color. Laboratory testing revealed a serum creatinine level of 8.9 mg/dl, indicating severe kidney failure, and a serum potassium level of 7.6 mEq/liter (normal level is 2.5–4.5 mEq/liter). She was admitted to the hospital for emergent dialysis.
Further diagnostic testing revealed a serum level of creatine kinase, an enzyme that normally exists inside of muscle tissue, of 178,000 units/liter (normal level is <200 units/liter). The patient was diagnosed with acute rhabdomyolysis, a condition characterized by severe, systemic muscle breakdown with release of muscle contents into the blood. One of these muscle components, myoglobin, is toxic to the kidney and causes kidney failure.

Cerivastatin (Baycol), a synthetic inhibitor of 3-hydroxy-3-methylglutaryl-coenzyme-A reductase, belongs to a class of cholesterol-lowering medications called “statins.” The drug was approved on the basis of lowering serum cholesterol levels in 1997. At the time of this patient’s presentation, no rhabdomyolysis cases associated with cerivastatin use had been reported in the literature. Could cerivastatin be causing this patient’s rare and potentially fatal condition?

Epidemiology is concerned with investigating the cause of disease. In this example, there are some reasons to suspect that cerivastatin might be causing rhabdomyolysis. The disease developed soon after initiation of cerivastatin. Similar cholesterol-lowering medications can also damage muscle tissue, though the severe rhabdomyolysis seen in this case would be rare.

During the first 100 days following approval, the cerivastatin manufacturer received seven case reports of rhabdomyolysis among people using cerivastatin. Should these seven cases be cause for concern? It is difficult to answer this question based on the case report data alone. The next step is to estimate the frequency of rhabdomyolysis in cerivastatin users. According to company sales data, there were 3,100 cerivastatin prescriptions dispensed during the first 100 days of drug approval. On the basis of these data, the estimated frequency of rhabdomyolysis in cerivastatin users is 7/3,100, or 0.2%.

This disease frequency seems relatively small, but rhabdomyolysis is a rare and potentially fatal condition. The next step is to compare the frequency of rhabdomyolysis among cerivastatin users to that of an appropriate control population. One possible control population might be people who were using similar cholesterol-lowering medications. In previous clinical trials, a total of 33,683 people had been assigned to a statin medication other than cerivastatin; 8 of these people developed rhabdomyolysis. On the basis of these trial data, the estimated frequency of rhabdomyolysis among people using other statin drugs is 8/33,683, or 0.02%.

While these findings may be partially distorted, due to differences in the compared populations, the observed tenfold greater frequency of rhabdomyolysis among cerivastatin users is concerning.

Two years after cerivastatin was approved, the manufacturer conducted an internal investigation of rhabdomyolysis rates. They found cerivastatin use to be associated with a 20-fold greater risk of rhabdomyolysis compared with other approved statin medications. Their findings were not reported or published. Case reports of rhabdomyolysis associated with cerivastatin use began to surface in the medical literature in 2000–2001.1 By August 2001, there were 31 fatal cases of rhabdomyolysis attributed to cerivastatin. At this time, the company voluntarily removed the drug from the market.2
1.2 Prevalence

The cerivastatin example demonstrates that measures of disease frequency represent key initial information needed to investigate the cause of disease. While it may be tempting to dive in and conduct novel discovery studies or high-profile clinical trials, an important first question about a disease process is, “how frequently does the disease occur?” Measures of disease frequency can help answer several important questions:

1. Measures of disease frequency can provide big picture information about a disease, framing public health questions and guiding resource allocation. For example, years after the invention of chronic dialysis for kidney failure, researchers observed that rates of cardiovascular death among dialysis patients were approximately 30-fold greater than those of the general population. These disease frequency data lead to a dramatic increase in funding for research of links between kidney and cardiovascular diseases.

2. Measures of disease frequency describe the absolute risk of a disease. For example, many studies have reported that smoking causes a more than tenfold increase in the relative risk of lung cancer. Rate data reveals that the cumulative lifetime risk of lung cancer for a person who smokes is approximately 18%. The rate data are important for counseling patients and understanding the impact of the disease on the population.

3. Measures of disease frequency can be categorized, or stratified, by person, place, and/or time characteristics to gain insight into the pathogenesis (mechanism) of disease. For example, rates of multiple sclerosis, an autoimmune disease that affects the central nervous system, vary considerably by geographic region within the USA. Areas with the lowest sunlight exposure have the highest incidence of multiple sclerosis. These rate data lead some researchers to investigate whether vitamin D deficiency might participate in the pathogenesis of multiple sclerosis. Vitamin D is obtained from sunlight exposure and can suppress inflammation and T-cell function.

The two most commonly used measures of disease frequency are prevalence and incidence.

1.2 Prevalence

Point prevalence measures the amount of a disease at one particular point in time. Prevalence is defined as the proportion of people who have the disease:

$$\text{Prevalence (\%) = \frac{\text{number of people with disease}}{\text{number of people in the population}} \times 100\%}$$
Because prevalence is always a ratio of some number of people and some number of people, prevalence estimates are often multiplied by 100% and expressed as %.

**Example 1.1.** What is the prevalence of anxiety disorder among second year medical students?

Solution: Administer a standardized test for anxiety disorder to 200 second year medical students; find that 12 meet the definition of anxiety disorder. $\text{Prevalence} = \frac{12}{200} \times 100\% = 6\%$.

In the medical literature, the term “prevalent” is also used to indicate a previous history of a chronic disease. For example, “prevalent diabetes” and “prevalent coronary disease” may be used in clinical research studies to indicate previous or current diagnoses of these conditions, because they are rarely cured and considered to be present indefinitely after diagnosis. In contrast, a previous history of a short-lived disease, such as influenza, would not be considered to represent prevalent disease, unless that condition was found to exist at the time of measurement.

Prevalence measures help to describe the current burden of a disease in a population in order to facilitate planning and resource allocation. For example, if the prevalence of anxiety disorder was truly 6% among second year medical students, the medical school might consider implementing specific counseling programs for students with this disorder. Analogously, if the prevalence of diabetes is found to be 40% among patients in a particular chronic kidney disease clinic, then that clinic might implement routine blood glucose monitoring.

### 1.3 Incidence

Incidence is a measure of the number of new cases of disease that develop over time. There are two definitions of incidence, differing only by the choice of the denominator:

\[
\text{Incidence proportion} = \frac{\text{number of new cases of disease}}{\text{population without disease at baseline}}
\]

\[
\text{Incidence rate} = \frac{\text{number of new cases of disease}}{\text{person - time at risk}}
\]

Another term for incidence rate is incidence density.

**Example 1.2.** What is the incidence of influenza infection among UW medical students during a 3-month period from January through March 2002?

Solution: Suppose that there are 500 UW medical students beginning in January 2000, and 5 new cases of influenza develop from January through March (3 months of follow-up).
Incidence proportion = \( \frac{5 \text{ cases}}{500 \text{ people}} \times 100\% = 1\%, \text{ or 1 per 100 people} \)

\[
\text{Incidence rate} = \frac{5 \text{ cases}}{(500 \text{ people} \times 3 \text{ months})} = \frac{5 \text{ cases}}{1500 \text{ person-months}} = 0.003 \text{ cases/ person-month} = 3 \text{ cases/ 1000 person-months}
\]

Incidence rates are typically reported as the number of cases of disease per some rounded measurement of time at risk, such as 1,000 or 100,000 person-years. The inclusion of time-at-risk in the denominator of incidence rate provides a more precise description of incidence than incidence proportion, particularly if study subjects contribute different amounts of time at risk to a study. For the influenza example, suppose that some of the medical students in the study are assigned to a distant clinical rotation for part of the study period, and cannot report influenza to a research study center during that time. Because the study could not detect the development of influenza for these away months, time at-risk should be adjusted to consider only months in which the disease could be captured. Table 1.1 presents data for the first six students in the study.

<table>
<thead>
<tr>
<th>Student</th>
<th>Unadjusted time at risk</th>
<th>Months away</th>
<th>Actual time at risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 months</td>
<td>0 month</td>
<td>3 months</td>
</tr>
<tr>
<td>2</td>
<td>3 months</td>
<td>0 month</td>
<td>3 months</td>
</tr>
<tr>
<td>3</td>
<td>3 months</td>
<td>0 month</td>
<td>3 months</td>
</tr>
<tr>
<td>4</td>
<td>3 months</td>
<td>1 month</td>
<td>2 months</td>
</tr>
<tr>
<td>5</td>
<td>3 months</td>
<td>2 month</td>
<td>1 month</td>
</tr>
<tr>
<td>6</td>
<td>3 months</td>
<td>3 month</td>
<td>0 month</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>12 months</td>
</tr>
</tbody>
</table>

The total time at risk contributed by these six students is 12 months. If two cases of influenza developed in these six students, then the incidence rate of influenza for these 6 students would be \( \frac{2 \text{ cases}}{12 \text{ person-months}} = 16.7 \text{ cases per 100 person-months} \).

The calculation of incidence rate from person-time data may be best appreciated using a diagram representing time at risk and disease status for each individual in a study, as shown in Fig. 1.1.

In this diagram, students one and three are followed for the full 3-month study period and do not develop influenza. Student two is also followed for the full 3-month period, but develops influenza at the end of the study. Student four develops influenza after only 2 months of follow-up. Since a study subject is longer at risk for developing incident (new) disease once the disease has occurred, we would consider the total time at risk for student four to be 2 months. Students five and six do not develop the disease and contribute approximately 1.25 and 0.25 months of time at-risk, respectively.
The importance of counting time at risk is highlighted by examples in which follow-up time differs between comparison groups. For example, a study compares rates of cellulitis, a common skin infection, between children seen in primary care clinics at a county and a university hospital. Investigators study 500 children from each site and follow them for up to 5 years. Results using incidence proportion data indicate that cellulitis is more common at the university hospital:

<table>
<thead>
<tr>
<th></th>
<th>Number of children</th>
<th>Cellulitis cases</th>
<th>Incidence proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>County hospital</td>
<td>500</td>
<td>10</td>
<td>2%</td>
</tr>
<tr>
<td>University hospital</td>
<td>500</td>
<td>15</td>
<td>3%</td>
</tr>
</tbody>
</table>

However, these raw rate data do not include time at risk. It is possible that children from the county hospital are lost to follow-up or dropout of the study more frequently than those from the university hospital. If this were the case, then the incidence proportion data would be misleading. Examining the same data using incidence rates reveal a different result:

<table>
<thead>
<tr>
<th></th>
<th>Number of children</th>
<th>Cellulitis cases</th>
<th>Time at risk (person-years)</th>
<th>Incidence rate (per 1,000 person-years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>County hospital</td>
<td>500</td>
<td>10</td>
<td>1,200</td>
<td>8.3</td>
</tr>
<tr>
<td>University hospital</td>
<td>500</td>
<td>15</td>
<td>2,000</td>
<td>7.5</td>
</tr>
</tbody>
</table>

The incidence rate of cellulitis is actually higher at the county hospital after accounting for person time at risk.

Incidence measures help to provide clues as to the cause or development of a disease. For the anxiety disorder example, suppose that the incidence proportion of anxiety disorder was 5% per year among medical students. This incidence measure would consider only new cases of anxiety disorder that developed during medical school; students with prevalent anxiety disorder at the beginning of medical school would not be counted. These incidence data suggest that certain aspects of medical school might contribute to anxiety disorder, prompting a more thorough search for possible causal factors. In contrast, the prevalence data for anxiety
disorder alerted to a high burden of disease in the student population, motivating implementation of treatment programs.

1.4 Relationship Between Prevalence and Incidence

The prevalence of a disease is a function of how often new cases develop and how long the disease state lasts. For example, the incidence of influenza may be relatively high during influenza season; however, the prevalence of influenza at any point in time is likely to be low because illness is short-lived; people either recover quickly, or in rare cases, die from the disease. In contrast, the prevalence of diabetes is likely to be high because there is a steady incidence of new cases, and the disease, though treatable, is rarely cured.

The mathematical relationship between prevalence and incidence is \( P = I \times D \), where \( P \) is the prevalence, \( I \) the incidence, and \( D \) the duration of disease. Figure 1.2 presents a graphical depiction of the relationship between incidence and prevalence.

Individuals in a population will acquire a disease at some rate (incidence). They will remain with the disease until they either get well, die, or leave the population (and cannot be counted).

1.5 Stratification of Disease Frequency by Person, Place, and Time

Once we calculate measures of disease frequency, we can examine whether these measures vary by personal characteristics, geography, and/or time periods. Stratification refers to the process of separating analysis by subgroups. For example, the prevalence of diabetes among all US adults is approximately 9.0%; the prevalence of diabetes stratified by race is 8.2% among whites and 14.9% among Native Americans.
Example 1.3

A study was conducted to describe the epidemiology of latex allergy in healthcare workers. At the beginning of the study (baseline), 500 females and 540 males underwent skin-prick testing; 40 females and 22 males tested positive for latex sensitivity. Follow-up skin tests were performed 2 years later, and 29 new cases of latex sensitivity were detected.

The overall prevalence of latex sensitivity among healthcare workers at baseline is $\frac{62}{1,040} \times 100\% = 6\%$.

The prevalence stratified by sex is $\frac{40}{500} \times 100\% = 8\%$ in females, and $\frac{22}{540} \times 100\% = 4\%$ in males. Therefore, latex sensitivity appears to be twice as common in women compared to men at baseline. To calculate incidence, the number of new cases of latex sensitivity that develop over time, we will exclude the 62 people who already had prevalent latex sensitivity at baseline.

\[
\text{Incidence proportion of latex allergy} = \frac{29}{1,040 - 62} \times 100\% = 3\%
\]

\[
\text{Incidence rate of latex allergy} = \frac{29}{(1,040 - 62) \times 2 \text{ years}} = 15 \text{ cases per 1,000 person-years}
\]

Information about sex is not provided for the new cases of latex sensitivity, so incidence data cannot be stratified by sex in this example.

1.5.1 Disease Frequency Measurements Stratified by Characteristics of Person

Examples of personal characteristics include age, race/ethnicity, and sex. For example, polycythemia vera is a myeloproliferative disorder characterized by an abnormal increase in red blood cell mass. The estimated prevalence of polycythemia vera among individuals aged 35–44 is 9 cases per 100,000, whereas the estimated prevalence in people aged 75–84 is 163 cases per 100,000. Polycythemia rates are also greater in men and in people of Jewish/Eastern European ancestry. These stratified disease frequency data begin to define risk factors for the disease.

1.5.2 Disease Frequency Measurements Stratified by Characteristics of Place

The incidence of multiple sclerosis varies considerably by geographic region within the USA. Areas with the lowest sunlight exposure, such as Seattle, have the
highest incidence of multiple sclerosis. Vitamin D is ascertained from sunlight exposure and may play an important role in suppressing autoimmunity. Circulating vitamin D levels are particularly low in regions with reduced sunlight exposure. These disease frequency data, stratified by place, suggest the hypothesis that vitamin D deficiency may play a role in the pathogenesis of multiple sclerosis.

### 1.5.3 Disease Frequency Measurements Stratified by Characteristics of Time

In 1970, approximately 5% of all births in the USA were by Cesarean section delivery. By the year 2000, nearly 25% of US babies were born by Cesarean section. These strong temporal changes in rates generate a number of hypotheses. One possibility is that maternal age has also increased during this time period, leading to more complicated pregnancies that may require Cesarean section. A second possibility is that improved fetal monitoring technology that can detect small changes in fetal status may prompt more surgical intervention. A third possibility is that the routine use of repeat Cesarean section has become standard practice in the USA because of data demonstrating an increased risk of uterine rupture in women who have a vaginal birth after a first Cesarean section. Disease frequency measurements stratified by time are often hypothesis generating, motivating further studies to uncover the true causes of a disease process.

### 1.5.4 Disease Frequency Measurements To Complement Experimental Data

Stratified measures of disease frequency can also be used to corroborate experimental data. For example, animal models have suggested that estrogen can slow the progression of chronic kidney disease by reducing expression of proinflammatory cytokines and decreasing the extent of fibrosis within the kidney. Can measures of disease frequency in humans be used to corroborate these provocative experimental data?

One possibility is to obtain estimates of the incidence of chronic kidney disease, stratified by sex and menopausal status, as depicted by the hypothetical data presented in Table 1.2.

<table>
<thead>
<tr>
<th>Table 1.2 Rates of chronic kidney disease according to sex and premenopausal status</th>
<th>Incident rate of chronic kidney disease (cases per 1,000 person years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prenopausal women</td>
<td>4.0</td>
</tr>
<tr>
<td>Postmenopausal women</td>
<td>6.3</td>
</tr>
<tr>
<td>Men</td>
<td>6.5</td>
</tr>
</tbody>
</table>
These data reveal a lower chronic kidney disease incidence rate among premenopausal women, compared with postmenopausal women and men. These disease frequency data support the hypothesis that estrogen protects against chronic kidney disease, and represent a first step toward investigation of this process in humans.
Epidemiology and Biostatistics
An Introduction to Clinical Research
Kestenbaum, B.
2009, XIV, 242 p. 46 illus., Softcover