3

MEASURES OF RISK

Lesson 2 described measures of central location and spread, which are useful for summarizing continuous variables. However, many variables used by field epidemiologists are categorical variables, some of which have only two categories — exposed yes/no, test positive/negative, case/control, and so on. These variables have to be summarized with frequency measures such as ratios, proportions, and rates. Incidence, prevalence, and mortality rates are three frequency measures that are used to characterize the occurrence of health events in a population.

Objectives

After studying this lesson and answering the questions in the exercises, you will be able to:

- *Calculate and interpret the following epidemiologic measures:*
  - *Ratio*
  - *Proportion*
  - *Incidence proportion (attack rate)*
  - *Incidence rate*
  - *Prevalence*
  - *Mortality rate*
- *Choose and apply the appropriate measures of association and measures of public health impact*

Major Sections

Frequency Measures 3-2
Morbidity Frequency Measures 3-10
Mortality Frequency Measures 3-20
Natality (Birth) Measures 3-38
Measures of Association 3-38
Measures of Public Health Impact 3-47
Summary 3-50

Numerator = upper portion of a fraction

Denominator = lower portion of a fraction

Frequency Measures

A measure of central location provides a single value that summarizes an entire distribution of data. In contrast, a frequency measure characterizes only part of the distribution. Frequency measures compare one part of the distribution to another part of the distribution, or to the entire distribution. Common frequency measures are **ratios**, **proportions**, and **rates**. All three frequency measures have the same basic form:

$$\frac{numerator}{denominator} \; x \; 10^n$$

Recall that:

$10^0 = 1$ (anything raised to the 0 power equals 1)
$10^1 = 10$ (anything raised to the ^{1st} power is the value itself)
$10^2 = 10 \text{ x } 10 = 100$
$10^3 = 10 \text{ x } 10 \text{ x } 10 = 1{,}000$

So the fraction of (numerator/denominator) can be multiplied by 1, 10, 100, 1000, and so on. This multiplier varies by measure and will be addressed in each section.

Ratio

Definition of ratio

A ratio is the relative magnitude of two quantities or a comparison of any two values. It is calculated by dividing one interval- or ratio-scale variable by the other. The numerator and denominator need not be related. Therefore, one could compare apples with oranges or apples with number of physician visits.

Method for calculating a ratio

$$\frac{\textit{Number or rate of events, items, persons, etc. in one group}}{\textit{Number or rate of events, items, persons, etc. in another group}}$$

After the numerator is divided by the denominator, the result is often expressed as the result "to one" or written as the result ":1."

Note that in certain ratios, the numerator and denominator are different categories of the same variable, such as males and females, or persons 20–29 years and 30–39 years of age. In other ratios, the numerator and denominator are completely different variables, such as the number of hospitals in a city and the size of the population living in that city.

EXAMPLE: Calculating a Ratio — Different Categories of Same Variable

Between 1971 and 1975, as part of the National Health and Nutrition Examination Survey (NHANES), 7,381 persons ages 40–77 years were enrolled in a follow-up study.[1] At the time of enrollment, each study participant was classified as having or not having diabetes. During 1982–1984, enrollees were documented either to have died or were still alive. The results are summarized as follows.

	Original Enrollment (1971–1975)	Dead at Follow-Up (1982–1984)
Diabetic men	189	100
Nondiabetic men	3,151	811
Diabetic women	218	72
Nondiabetic women	3,823	511

Of the men enrolled in the NHANES follow-up study, 3,151 were nondiabetic and 189 were diabetic. Calculate the ratio of non-diabetic to diabetic men.

Ratio = 3,151 / 189 x 1 = 16.7:1

Properties and uses of ratios

- Ratios are common descriptive measures, used in all fields. In epidemiology, ratios are used as both descriptive measures and as analytic tools. As a descriptive measure, ratios can describe the male-to-female ratio of participants in a study, or the ratio of controls to cases (e.g., two controls per case). As an analytic tool, ratios can be calculated for occurrence of illness, injury, or death between two groups. These ratio measures, including risk ratio (relative risk), rate ratio, and odds ratio, are described later in this lesson.

- As noted previously, the numerators and denominators of a ratio can be related or unrelated. In other words, you are free to use a ratio to compare the number of males in a population with the number of females, or to compare the number of residents in a population with the number of hospitals or dollars spent on over-the-counter medicines.

- Usually, the values of both the numerator and denominator of a ratio are divided by the value of one or the other so that either the numerator or the denominator equals 1.0. So the ratio of non-diabetics to diabetics cited in the previous example is more likely to be reported as 16.7:1 than 3,151:189.

EXAMPLES: Calculating Ratios for Different Variables

Example A: *A city of 4,000,000 persons has 500 clinics. Calculate the ratio of clinics per person.*

$$500 / 4{,}000{,}000 \times 10^n = 0.000125 \text{ clinics per person}$$

To get a more easily understood result, you could set $10^n = 10^4 = 10{,}000$. Then the ratio becomes:

$$0.000125 \times 10{,}000 = 1.25 \text{ clinics per } 10{,}000 \text{ persons}$$

You could also divide each value by 1.25, and express this ratio as 1 clinic for every 8,000 persons.

Example B: *Delaware's infant mortality rate in 2001 was 10.7 per 1,000 live births.[2] New Hampshire's infant mortality rate in 2001 was 3.8 per 1,000 live births. Calculate the ratio of the infant mortality rate in Delaware to that in New Hampshire.*

$$10.7 / 3.8 \times 1 = 2.8{:}1$$

Thus, Delaware's infant mortality rate was 2.8 times as high as New Hampshire's infant mortality rate in 2001.

A commonly used epidemiologic ratio: death-to-case ratio

Death-to-case ratio is the number of deaths attributed to a particular disease during a specified period divided by the number of new cases of that disease identified during the same period. It is used as a measure of the severity of illness: the death-to-case ratio for rabies is close to 1 (that is, almost everyone who develops rabies dies from it), whereas the death-to-case ratio for the common cold is close to 0.

For example, in the United States in 2002, a total of 15,075 new cases of tuberculosis were reported.[3] During the same year, 802 deaths were attributed to tuberculosis. The tuberculosis death-to-case ratio for 2002 can be calculated as 802 / 15,075. Dividing both numerator and denominator by the numerator yields 1 death per 18.8 new cases. Dividing both numerator and denominator by the denominator (and multiplying by $10^n = 100$) yields 5.3 deaths per 100 new cases. Both expressions are correct.

Note that, presumably, many of those who died had initially contracted tuberculosis years earlier. Thus many of the 802 in the numerator are not among the 15,075 in the denominator. Therefore, the death-to-case ratio is a ratio, but not a proportion.

Proportion

Definition of proportion

A proportion is the comparison of a part to the whole. It is a type of ratio in which the numerator is included in the denominator. You might use a proportion to describe what fraction of clinic patients tested positive for HIV, or what percentage of the population is younger than 25 years of age. A proportion may be expressed as a decimal, a fraction, or a percentage.

Method for calculating a proportion

$$\frac{\textit{Number of persons or events with a particular characteristic}}{\textit{Total number of persons or events, of which the numerator is a subset}} \times 10^n$$

For a proportion, 10^n is usually 100 (or n=2) and is often expressed as a percentage.

EXAMPLE: Calculating a Proportion

Example A: *Calculate the proportion of men in the NHANES follow-up study who were diabetics.*

Numerator = 189 diabetic men
Denominator = Total number of men = 189 + 3,151 = 3,340

Proportion = (189 / 3,340) x 100 = 5.66%

Example B: *Calculate the proportion of deaths among men.*

Numerator = deaths in men
= 100 deaths in diabetic men + 811 deaths in nondiabetic men
= 911 deaths in men

Notice that the numerator (911 deaths in men) is a subset of the denominator.

Denominator = all deaths
= 911 deaths in men + 72 deaths in diabetic women + 511 deaths in nondiabetic women
= 1,494 deaths

Proportion = 911 / 1,494 = 60.98% = 61%

Your Turn: *What proportion of all study participants were men? (Answer = 45.25%)*

Properties and uses of proportions

- Proportions are common descriptive measures used in all fields. In epidemiology, proportions are used most often as descriptive measures. For example, one could calculate the proportion of persons enrolled in a study among all those eligible ("participation rate"), the proportion of children in a village vaccinated against measles, or the proportion of persons who developed illness among all passengers of a cruise ship.

- Proportions are also used to describe the amount of disease that can be attributed to a particular exposure. For example, on the basis of studies of smoking and lung cancer, public health officials have estimated that greater than 90% of the lung cancer cases that occur are attributable to cigarette smoking.

- In a proportion, the numerator must be included in the denominator. Thus, the number of apples divided by the number of oranges is not a proportion, but the number of apples divided by the total number of fruits of all kinds is a proportion. Remember, the numerator is always a subset of the denominator.

- A proportion can be expressed as a fraction, a decimal, or a percentage. The statements "one fifth of the residents became ill" and "twenty percent of the residents became ill" are equivalent.

- Proportions can easily be converted to ratios. If the numerator is the number of women (179) who attended a clinic and the denominator is all the clinic attendees (341), the proportion of clinic attendees who are women is 179 / 341, or 52% (a little more than half). To convert to a ratio, subtract the numerator from the denominator to get the number of clinic patients who are not women, i.e., the number of men (341 – 179 = 162 men.)Thus, ratio of women to men could be calculated from the proportion as:

$$\begin{aligned} \text{Ratio} &= 179 / (341 - 179) \times 1 \\ &= 179 / 162 \\ &= 1.1 \text{ to } 1 \text{ female-to-male ratio} \end{aligned}$$

Conversely, if a ratio's numerator and denominator together make up a whole population, the ratio can be converted to a proportion. You would add the ratio's numerator and denominator to form the denominator of the proportion, as illustrated in the NHANES follow-up study examples (provided earlier in this lesson).

A specific type of epidemiologic proportion: proportionate mortality

Proportionate mortality is the proportion of deaths in a specified population during a period of time that are attributable to different causes. Each cause is expressed as a percentage of all deaths, and the sum of the causes adds up to 100%. These proportions are not rates because the denominator is all deaths, not the size of the population in which the deaths occurred. Table 3.1 lists the primary causes of death in the United States in 2003 for persons of all ages and for persons aged 25–44 years, by number of deaths, proportionate mortality, and rank.

Table 3.1 Number, Proportionate Mortality, and Ranking of Deaths for Leading Causes of Death, All Ages and 25–44 Year Age Group, United States, 2003

	All Ages			Ages 25–44 Years		
	Number	Percentage	Rank	Number	Percentage	Rank
All causes	2,443,930	100.0		128,924	100.0	
Diseases of heart	684,462	28.0	1	16,283	12.6	3
Malignant neoplasms	554,643	22.7	2	19,041	14.8	2
Cerebrovascular disease	157,803	6.5	3	3,004	2.3	8
Chronic lower respiratory diseases	126,128	5.2	4	401	0.3	*
Accidents (unintentional injuries)	105,695	4.3	5	27,844	21.6	1
Diabetes mellitus	73,965	3.0	6	2,662	2.1	9
Influenza & pneumonia	64,847	2.6	7	1,337	1.0	10
Alzheimer's disease	63,343	2.6	8	0	0.0	*
Nephritis, nephrotic syndrome, nephrosis	33,615	1.4	9	305	0.2	*
Septicemia	34,243	1.4	10	328	0.2	*
Intentional self-harm (suicide)	30,642	1.3	11	11,251	8.7	4
Chronic liver disease and cirrhosis	27,201	1.1	12	3,288	2.6	7
Assault (homicide)	17,096	0.7	13	7,367	5.7	5
HIV disease	13,544	0.5	*	6,879	5.3	6
All other	456,703	18.7		29,480	22.9	

* Not among top ranked causes

Data Sources: Centers for Disease Control and Prevention. Summary of notifiable diseases, United States, 2003. MMWR 2005;2(No. 54).
Hoyert DL, Kung HC, Smith BL. Deaths: Preliminary data for 2003. National Vital Statistics Reports; vol. 53 no 15. Hyattsville, MD: National Center for Health Statistics 2005: p. 15, 27.

As illustrated in Table 3.1, the proportionate mortality for HIV was 0.5% among all age groups, and 5.3% among those aged 25–44 years. In other words, HIV infection accounted for 0.5% of all deaths, and 5.3% of deaths among 25–44 year olds.

Rate

Definition of rate

In epidemiology, a rate is a measure of the frequency with which an event occurs in a defined population over a specified period of time. Because rates put disease frequency in the perspective of the size of the population, rates are particularly useful for comparing disease frequency in different locations, at different times, or among different groups of persons with potentially different sized populations; that is, a rate is a measure of risk.

To a non-epidemiologist, rate means how fast something is happening or going. The speedometer of a car indicates the car's speed or rate of travel in miles or kilometers per hour. This rate is always reported per some unit of time. Some epidemiologists restrict use of the term rate to similar measures that are expressed per unit of time. For these epidemiologists, a rate describes how quickly disease occurs in a population, for example, 70 new cases of breast cancer per 1,000 women per year. This measure conveys

Table 3.2 Epidemiologic Measures Categorized as Ratio, Proportion, or Rate

Condition	Ratio	Proportion	Rate
Morbidity (Disease)	Risk ratio (Relative risk) Rate ratio Odds ratio Period prevalence	Attack rate (Incidence proportion) Secondary attack rate Point prevalence Attributable proportion	Person-time incidence rate
Mortality (Death)	Death-to-case ratio	Proportionate mortality	Crude mortality rate Case-fatality rate Cause-specific mortality rate Age-specific mortality rate Maternal mortality rate Infant mortality rate
Natality (Birth)			Crude birth rate Crude fertility rate

Morbidity Frequency Measures

Morbidity has been defined as any departure, subjective or objective, from a state of physiological or psychological well-being. In practice, morbidity encompasses disease, injury, and disability. In addition, although for this lesson the term refers to the number of persons who are ill, it can also be used to describe the periods of illness that these persons experienced, or the duration of these illnesses.[4]

Measures of morbidity frequency characterize the number of persons in a population who become ill (incidence) or are ill at a given time (prevalence). Commonly used measures are listed in Table 3.3.

Table 3.3 Frequently Used Measures of Morbidity

Measure	Numerator	Denominator
Incidence proportion (or attack rate or risk)	Number of new cases of disease during specified time interval	Population at start of time interval
Secondary attack rate	Number of new cases among contacts	Total number of contacts
Incidence rate (or person-time rate)	Number of new cases of disease during specified time interval	Summed person-years of observation or average population during time interval
Point prevalence	Number of current cases (new and preexisting) at a specified point in time	Population at the same specified point in time
Period prevalence	Number of current cases (new and preexisting) over a specified period of time	Average or mid-interval population

Incidence refers to the occurrence of new cases of disease or injury in a population over a specified period of time. Although some epidemiologists use incidence to mean the number of new cases in a community, others use incidence to mean the number of new cases per unit of population.

Two types of incidence are commonly used — **incidence proportion** and **incidence rate**.

Incidence proportion or risk

Definition of incidence proportion

Synonyms for incidence proportion

- Attack rate
- Risk
- Probability of developing disease
- Cumulative incidence

Incidence proportion is the proportion of an initially disease-free population that develops disease, becomes injured, or dies during a specified (usually limited) period of time. Synonyms include attack rate, risk, probability of getting disease, and cumulative incidence. Incidence proportion is a proportion because the persons in the numerator, those who develop disease, are all included in the denominator (the entire population).

Method for calculating incidence proportion (risk)

$$\frac{\textit{Number of new cases of disease or injury during specified period}}{\textit{Size of population at start of period}}$$

EXAMPLES: Calculating Incidence Proportion (Risk)

Example A: *In the study of diabetics, 100 of the 189 diabetic men died during the 13-year follow-up period. Calculate the risk of death for these men.*

Numerator = 100 deaths among the diabetic men
Denominator = 189 diabetic men
$10^n = 10^2 = 100$

Risk = (100 / 189) x 100 = 52.9%

Example B: *In an outbreak of gastroenteritis among attendees of a corporate picnic, 99 persons ate potato salad, 30 of whom developed gastroenteritis. Calculate the risk of illness among persons who ate potato salad.*

Numerator = 30 persons who ate potato salad and developed gastroenteritis
Denominator = 99 persons who ate potato salad
$10^n = 10^2 = 100$

Risk = "Food-specific attack rate" = (30 / 99) x 100 = 0.303 x 100 = 30.3%

Properties and uses of incidence proportions

- Incidence proportion is a measure of the risk of disease or the probability of developing the disease during the specified period. As a measure of incidence, it includes only new cases of disease in the numerator. The denominator is the number of persons in the population at the start of the observation period. Because all of the persons with new cases of disease (numerator) are also represented in the denominator, a risk is also a proportion.

More About Denominators

The denominator of an incidence proportion is the number of persons at the start of the observation period. The denominator should be limited to the "population at risk" for developing disease, i.e., persons who have the potential to get the disease and be included in the numerator. For example, if the numerator represents new cases of cancer of the ovaries, the denominator should be restricted to women, because men do not have ovaries. This is easily accomplished because census data by sex are readily available. In fact, ideally the denominator should be restricted to women with ovaries, excluding women who have had their ovaries removed surgically (often done in conjunction with a hysterectomy), but this is not usually practical. This is an example of field epidemiologists doing the best they can with the data they have.

- In the outbreak setting, the term **attack rate** is often used as a synonym for risk. It is the risk of getting the disease during a specified period, such as the duration of an outbreak. A variety of attack rates can be calculated.

  Overall attack rate is the total number of new cases divided by the total population.

  A **food-specific attack rate** is the number of persons who ate a specified food and became ill divided by the total number of persons who ate that food, as illustrated in the previous potato salad example.

  A **secondary attack rate** is sometimes calculated to document the difference between community transmission of illness versus transmission of illness in a household, barracks, or other closed population. It is calculated as:

$$\frac{\textit{Number of cases among contacts of primary cases}}{\textit{Total number of contacts}} \times 10^n$$

Often, the total number of contacts in the denominator is calculated as the total population in the households of the primary cases, minus the number of primary cases. For a secondary attack rate, 10^n usually is 100%.

EXAMPLE: Calculating Secondary Attack Rates

Consider an outbreak of shigellosis in which 18 persons in 18 different households all became ill. If the population of the community was 1,000, then the overall attack rate was 18 / 1,000 x 100% = 1.8%. One incubation period later, 17 persons in the same households as these "primary" cases developed shigellosis. If the 18 households included 86 persons, calculate the secondary attack rate.

Secondary attack rate = (17 / (86 - 18)) x 100% = (17 / 68) x 100% = 25.0%

Incidence rate or person-time rate

Definition of incidence rate

Incidence rate or person-time rate is a measure of incidence that incorporates time directly into the denominator. A person-time rate is generally calculated from a long-term cohort follow-up study, wherein enrollees are followed over time and the occurrence of new cases of disease is documented. Typically, each person is observed from an established starting time until one of four "end points" is reached: onset of disease, death, migration out of the study ("lost to follow-up"), or the end of the study. Similar to the incidence proportion, the numerator of the incidence rate is the number of new cases identified during the period of observation. However, the denominator differs. The denominator is the sum of the time each person was observed, totaled for all persons. This denominator represents the total time the population was at risk of and being watched for disease. Thus, the incidence rate is the ratio of the number of cases to the total time the population is at risk of disease.

Method for calculating incidence rate

$$\frac{\textit{Number of new cases of disease or injury during specified period}}{\textit{Time each person was observed, totaled for all persons}}$$

In a long-term follow-up study of morbidity, each study participant may be followed or observed for several years. One person followed for 5 years without developing disease is said to contribute 5 person-years of follow-up.

What about a person followed for one year before being lost to follow-up at year 2? Many researchers assume that persons lost to follow-up were, on average, disease-free for half the year, and thus contribute ½ year to the denominator. Therefore, the person followed for one year before being lost to follow-up contributes 1.5 person-years. The same assumption is made for participants diagnosed with the disease at the year 2 examination — some may have developed illness in month 1, and others in months 2 through 12. So, on average, they developed illness halfway through the year. As a result, persons diagnosed with the disease contribute ½ year of follow-up during the year of diagnosis.

The denominator of the person-time rate is the sum of all of the person-years for each study participant. So, someone lost to follow-up in year 3, and someone diagnosed with the disease in year 3, each contributes 2.5 years of disease-free follow-up to the denominator.

Properties and uses of incidence rates

- An incidence rate describes how quickly disease occurs in a population. It is based on person-time, so it has some advantages over an incidence proportion. Because person-time is calculated for each subject, it can accommodate persons coming into and leaving the study. As noted in the previous example, the denominator accounts for study participants who are lost to follow-up or who die during the study period. In addition, it allows enrollees to enter the study at different times. In the NHANES follow-up study, some participants were enrolled in 1971, others in 1972, 1973, 1974, and 1975.

- Person-time has one important drawback. Person-time assumes that the probability of disease during the study period is constant, so that 10 persons followed for one year equals one person followed for 10 years. Because the risk of many chronic diseases increases with age, this assumption is often not valid.

- Long-term cohort studies of the type described here are not very common. However, epidemiologists far more commonly calculate incidence rates based on a numerator of cases observed or reported, and a denominator based on the mid-year population. This type of incident rate turns out to be comparable to a person-time rate.

- Finally, if you report the incidence rate of, say, the heart disease study as 2.5 per 1,000 person-years, epidemiologists might understand, but most others will not. Person-time is epidemiologic jargon. To convert this jargon to something understandable, simply replace “person-years” with “persons per year.” Reporting the results as 2.5 new cases of heart disease per 1,000 persons per year sounds like English rather than jargon. It also conveys the sense of the incidence rate as a dynamic process, the speed at which new cases of disease occur in the population.

EXAMPLES: Calculating Incidence Rates

Example A: *Investigators enrolled 2,100 women in a study and followed them annually for four years to determine the incidence rate of heart disease. After one year, none had a new diagnosis of heart disease, but 100 had been lost to follow-up. After two years, one had a new diagnosis of heart disease, and another 99 had been lost to follow-up. After three years, another seven had new diagnoses of heart disease, and 793 had been lost to follow-up. After four years, another 8 had new diagnoses with heart disease, and 392 more had been lost to follow-up.*

The study results could also be described as follows: No heart disease was diagnosed at the first year. Heart disease was diagnosed in one woman at the second year, in seven women at the third year, and in eight women at the fourth year of follow-up. One hundred women were lost to follow-up by the first year, another 99 were lost to follow-up after two years, another 793 were lost to follow-up after three years, and another 392 women were lost to follow-up after 4 years, leaving 700 women who were followed for four years and remained disease free.

Calculate the incidence rate of heart disease among this cohort. Assume that persons with new diagnoses of heart disease and those lost to follow-up were disease-free for half the year, and thus contribute ½ year to the denominator.

Numerator = number of new cases of heart disease
= 0 + 1 + 7 + 8 = 16

Denominator = person-years of observation
= (2,000 + ½ x 100) + (1,900 + ½ x 1 + ½ x 99) + (1,100 + ½ x 7 + ½ x 793) + (700 + ½ x 8 + ½ x 392)
= 6,400 person-years of follow-up

or

Denominator = person-years of observation
= (1 x 1.5) + (7 x 2.5) + (8 x 3.5) + (100 x 0.5) + (99 x 1.5) + (793 x 2.5) + (392 x 3.5) + (700 x 4)
= 6,400 person-years of follow-up

Person-time rate = $\dfrac{\text{Number of new cases of disease or injury during specified period}}{\text{Time each person was observed, totaled for all persons}}$
= 16 / 6,400
= .0025 cases per person-year
= 2.5 cases per 1,000 person-years

In contrast, the incidence proportion can be calculated as 16 / 2,100 = 7.6 cases per 1,000 population during the four-year period, or an average of 1.9 cases per 1,000 per year (7.6 divided by 4 years). The incidence proportion underestimates the true rate because it ignores persons lost to follow-up, and assumes that they remained disease-free for all four years.

Example B: *The diabetes follow-up study included 218 diabetic women and 3,823 nondiabetic women. By the end of the study, 72 of the diabetic women and 511 of the nondiabetic women had died. The diabetic women were observed for a total of 1,862 person-years; the nondiabetic women were observed for a total of 36,653 person-years. Calculate the incidence rates of death for the diabetic and non-diabetic women.*

For diabetic women, numerator = 72 and denominator = 1,862
Person-time rate = 72 / 1,862
= 0.0386 deaths per person-year
= 38.6 deaths per 1,000 person-years

For nondiabetic women, numerator = 511 and denominator = 36,653
Person-time rate = 511 / 36,653 = 0.0139 deaths per person-year
= 13.9 deaths per 1,000 person-years

Example C: *In 2003, 44,232 new cases of acquired immunodeficiency syndrome (AIDS) were reported in the United States.[5] The estimated mid-year population of the U.S. in 2003 was approximately 290,809,777.[6] Calculate the incidence rate of AIDS in 2003.*

Numerator = 44,232 new cases of AIDS
Denominator = 290,809,777 estimated mid-year population
10^n = 100,000

Incidence rate = (44,232 / 290,809,777) x 100,000
= 15.21 new cases of AIDS per 100,000 population

Prevalence

Definition of prevalence

Prevalence, sometimes referred to as **prevalence rate**, is the proportion of persons in a population who have a particular disease or attribute at a specified point in time or over a specified period of time. Prevalence differs from incidence in that prevalence includes all cases, both new and preexisting, in the population at the specified time, whereas incidence is limited to new cases only.

Point prevalence refers to the prevalence measured at a particular point in time. It is the proportion of persons with a particular disease or attribute on a particular date.

Period prevalence refers to prevalence measured over an interval of time. It is the proportion of persons with a particular disease or attribute at any time during the interval.

Method for calculating prevalence of disease

$$\frac{\textit{All new and pre-existing cases during a given time period}}{\textit{Population during the same time period}} \times 10^n$$

Method for calculating prevalence of an attribute

$$\frac{\textit{Persons having a particular attribute during a given time period}}{\textit{Population during the same time period}} \times 10^n$$

The value of 10^n is usually 1 or 100 for common attributes. The value of 10^n might be 1,000, 100,000, or even 1,000,000 for rare attributes and for most diseases.

EXAMPLE: Calculating Prevalence

In a survey of 1,150 women who gave birth in Maine in 2000, a total of 468 reported taking a multivitamin at least 4 times a week during the month before becoming pregnant.[7] Calculate the prevalence of frequent multivitamin use in this group.

Numerator = 468 multivitamin users
Denominator = 1,150 women

Prevalence = (468 / 1,150) x 100 = 0.407 x 100 = 40.7%

Properties and uses of prevalence

- Prevalence and incidence are frequently confused. Prevalence refers to proportion of persons who *have* a condition at or during a particular time period, whereas incidence refers to the proportion or rate of persons who *develop* a condition during a particular time period. So prevalence and incidence are similar, but prevalence includes new and pre-existing cases whereas incidence includes new cases only. The key difference is in their numerators.

  Numerator of incidence = new cases that occurred during a given time period

  Numerator of prevalence = all cases present during a given time period

- The numerator of an incidence proportion or rate consists only of persons whose illness began during the specified interval. The numerator for prevalence includes all persons ill from a specified cause during the specified interval **regardless of when the illness began**. It includes not only new cases, but also preexisting cases representing persons who remained ill during some portion of the specified interval.

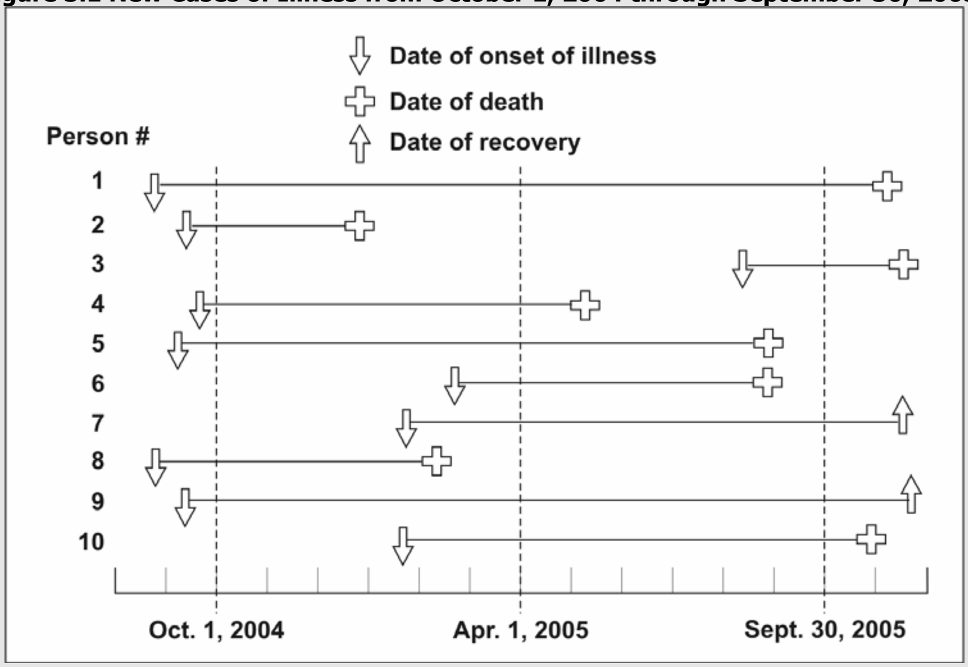
- Prevalence is based on both incidence and duration of illness. High prevalence of a disease within a population might reflect high incidence or prolonged survival without cure or both. Conversely, low prevalence might indicate low incidence, a rapidly fatal process, or rapid recovery.

- Prevalence rather than incidence is often measured for chronic diseases such as diabetes or osteoarthritis which have long duration and dates of onset that are difficult to pinpoint.

EXAMPLES: Incidence versus Prevalence

Figure 3.1 represents 10 new cases of illness over about 15 months in a population of 20 persons. Each horizontal line represents one person. The down arrow indicates the date of onset of illness. The solid line represents the duration of illness. The up arrow and the cross represent the date of recovery and date of death, respectively.

Figure 3.1 New Cases of Illness from October 1, 2004 through September 30, 2005



Example A: *Calculate the incidence rate from October 1, 2004, to September 30, 2005, using the midpoint population (population alive on April 1, 2005) as the denominator. Express the rate per 100 population.*

Incidence rate numerator	=	number of new cases between October 1 and September 30
	=	4 (the other 6 all had onsets before October 1, and are not included)
Incidence rate denominator	=	April 1 population
	=	18 (persons 2 and 8 died before April 1)
Incidence rate	=	(4 / 18) x 100
	=	22 new cases per 100 population

Example B: *Calculate the point prevalence on April 1, 2005. Point prevalence is the number of persons ill on the date divided by the population on that date. On April 1, seven persons (persons 1, 4, 5, 7, 9, and 10) were ill.*

Point prevalence	=	(7 / 18) x 100
	=	38.89%

Example C: *Calculate the period prevalence from October 1, 2004, to September 30, 2005. The numerator of period prevalence includes anyone who was ill any time during the period. In Figure 3.1, the first 10 persons were all ill at some time during the period.*

Period prevalence	=	(10 / 20) x 100
	=	50.0%

Mortality Frequency Measures

Mortality rate

A mortality rate is a measure of the frequency of occurrence of death in a defined population during a specified interval. Morbidity and mortality measures are often the same mathematically; it's just a matter of what you choose to measure, illness or death. The formula for the mortality of a defined population, over a specified period of time, is:

$$\frac{\textit{Deaths occurring during a given time period}}{\textit{Size of the population among which the deaths occurred}} \times 10^n$$

When mortality rates are based on vital statistics (e.g., counts of death certificates), the denominator most commonly used is the size of the population at the middle of the time period. In the United States, values of 1,000 and 100,000 are both used for 10^n for most types of mortality rates. Table 3.4 summarizes the formulas of frequently used mortality measures.

Table 3.4 Frequently Used Measures of Mortality

Measure	Numerator	Denominator	10^n
Crude death rate	Total number of deaths during a given time interval	Mid-interval population	1,000 or 100,000
Cause-specific death rate	Number of deaths assigned to a specific cause during a given time interval	Mid-interval population	100,000
Proportionate mortality	Number of deaths assigned to a specific cause during a given time interval	Total number of deaths from all causes during the same time interval	100 or 1,000
Death-to-case ratio	Number of deaths assigned to a specific cause during a given time interval	Number of new cases of same disease reported during the same time interval	100
Neonatal mortality rate	Number of deaths among children < 28 days of age during a given time interval	Number of live births during the same time interval	1,000
Postneonatal mortality rate	Number of deaths among children 28–364 days of age during a given time interval	Number of live births during the same time interval	1,000
Infant mortality rate	Number of deaths among children < 1 year of age during a given time interval	Number of live births during the same time interval	1,000
Maternal mortality rate	Number of deaths assigned to pregnancy-related causes during a given time interval	Number of live births during the same time interval	100,000

Crude mortality rate (crude death rate)

The crude mortality rate is the mortality rate from all causes of death for a population. In the United States in 2003, a total of 2,419,921 deaths occurred. The estimated population was 290,809,777. The crude mortality rate in 2003 was, therefore, (2,419,921 / 290,809,777) x 100,000, or 832.1 deaths per 100,000 population.[8]

Cause-specific mortality rate

The cause-specific mortality rate is the mortality rate from a specified cause for a population. The numerator is the number of deaths attributed to a specific cause. The denominator remains the size of the population at the midpoint of the time period. The fraction is usually expressed per 100,000 population. In the United States in 2003, a total of 108,256 deaths were attributed to accidents (unintentional injuries), yielding a cause-specific mortality rate of 37.2 per 100,000 population.[8]

Age-specific mortality rate

An age-specific mortality rate is a mortality rate limited to a particular age group. The numerator is the number of deaths in that age group; the denominator is the number of persons in that age group in the population. In the United States in 2003, a total of 130,761 deaths occurred among persons aged 25-44 years, or an age-specific mortality rate of 153.0 per 100,000 25–44 year olds.[8] Some specific types of age-specific mortality rates are neonatal, postneonatal, and infant mortality rates, as described in the following sections.

Infant mortality rate

The infant mortality rate is perhaps the most commonly used measure for comparing health status among nations. It is calculated as follows:

$$\frac{\textit{Number of deaths among children < 1 year of age reported during a given time period}}{\textit{Number of live births reported during the same time period}} \times 1{,}000$$

The infant mortality rate is generally calculated on an annual basis. It is a widely used measure of health status because it reflects the health of the mother and infant during pregnancy and the year thereafter. The health of the mother and infant, in turn, reflects a wide variety of factors, including access to prenatal care, prevalence of prenatal maternal health behaviors (such as alcohol or tobacco use and proper nutrition during pregnancy, etc.), postnatal care and behaviors (including childhood immunizations and proper nutrition), sanitation, and infection control.

Is the infant mortality rate a ratio? Yes. Is it a proportion? No, because some of the deaths in the numerator were among children born the previous year. Consider the infant mortality rate in 2003. That year, 28,025 infants died and 4,089,950 children were born, for an infant mortality rate of 6.951 per 1,000.[8] Undoubtedly, some of the deaths in 2003 occurred among children born in 2002, but the denominator includes only children born in 2003.

Is the infant mortality rate truly a rate? No, because the denominator is not the size of the mid-year population of children < 1 year of age in 2003. In fact, the age-specific death rate for children < 1 year of age for 2003 was 694.7 per 100,000.[8] Obviously the infant mortality rate and the age-specific death rate for infants are very similar (695.1 versus 694.7 per 100,000) and close enough for most purposes. They are not exactly the same, however, because the estimated number of infants residing in the United States on July 1, 2003 was slightly larger than the number of children born in the United States in 2002, presumably because of immigration.

Neonatal mortality rate

The neonatal period covers birth up to but not including 28 days. The numerator of the neonatal mortality rate therefore is the number of deaths among children under 28 days of age during a given time period. The denominator of the neonatal mortality rate, like that of the infant mortality rate, is the number of live births reported during the same time period. The neonatal mortality rate is usually expressed per 1,000 live births. In 2003, the neonatal mortality rate in the United States was 4.7 per 1,000 live births.[8]

Postneonatal mortality rate

The postneonatal period is defined as the period from 28 days of age up to but not including 1 year of age. The numerator of the postneonatal mortality rate therefore is the number of deaths among children from 28 days up to but not including 1 year of age during a given time period. The denominator is the number of live births reported during the same time period. The postneonatal mortality rate is usually expressed per 1,000 live births. In 2003, the postneonatal mortality rate in the United States was 2.3 per 1,000 live births.[8]

Maternal mortality rate

The maternal mortality rate is really a ratio used to measure mortality associated with pregnancy. The numerator is the number of deaths during a given time period among women while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes. The denominator is the number of live births reported during the same time period. Maternal mortality rate is usually expressed per 100,000 live births. In 2003, the U.S. maternal mortality rate was 8.9 per 100,000 live births.[8]

Sex-specific mortality rate

A sex-specific mortality rate is a mortality rate among either males or females. Both numerator and denominator are limited to the one sex.

Race-specific mortality rate

A race-specific mortality rate is a mortality rate related to a specified racial group. Both numerator and denominator are limited to the specified race.

Combinations of specific mortality rates

Mortality rates can be further stratified by combinations of cause, age, sex, and/or race. For example, in 2002, the death rate from diseases of the heart among women ages 45–54 years was 50.6 per 100,000.[9] The death rate from diseases of the heart among men in the same age group was 138.4 per 100,000, or more than 2.5 times as high as the comparable rate for women. These rates are a cause-, age-, and sex-specific rates, because they refer to one cause (diseases of the heart), one age group (45–54 years), and one sex (female or male).

EXAMPLE: Calculating Mortality Rates

Table 3.5 provides the number of deaths from all causes and from accidents (unintentional injuries) by age group in the United States in 2002. Review the following rates. Determine what to call each one, then calculate it using the data provided in Table 3.5.

a. Unintentional-injury-specific mortality rate for the entire population

This is a cause-specific mortality rate.

$$\text{Rate} = \frac{\text{number of unintentional injury deaths in the entire population}}{\text{estimated midyear population}} \times 100{,}000$$

= (106,742 / 288,357,000) x 100,000

= 37.0 unintentional-injury-related deaths per 100,000 population

b. All-cause mortality rate for 25–34 year olds

This is an age-specific mortality rate.

$$\text{Rate} = \frac{\text{number of deaths from all causes among 25–34 year olds}}{\text{estimated midyear population of 25–34 year olds}} \times 100{,}000$$

= (41,355 / 39,928,000) x 100,000

= 103.6 deaths per 100,000 25–34 year olds

c. All-cause mortality among males

This is a sex-specific mortality rate.

$$\text{Rate} = \frac{\text{number of deaths from all causes among males}}{\text{estimated midyear population of males}} \times 100{,}000$$

= (1,199,264 / 141,656,000) x 100,000

= 846.6 deaths per 100,000 males

d. Unintentional-injury-specific mortality among 25- to 34-year-old males

This is a cause-specific, age-specific, and sex-specific mortality rate

$$\text{Rate} = \frac{\text{number of unintentional injury deaths among 25–34 year old males}}{\text{estimated midyear population of 25–34 year old males}} \times 100{,}000$$

= (9,635 / 20,203,000) x 100,000

= 47.7 unintentional-injury-related deaths per 100,000 25–34 year olds

Table 3.5 All-Cause and Unintentional Injury Mortality and Estimated Population by Age Group, For Both Sexes and For Males Alone, United States, 2002

	All Races, Both Sexes			All Races, Males		
Age group (years)	**All Causes**	**Unintentional Injuries**	**Estimated Pop. (x 1000)**	**All Causes**	**Unintentional Injuries**	**Estimated Pop. (x 1000)**
0–4	32,892	2,587	19,597	18,523	1,577	10,020
5–14	7,150	2,718	41,037	4,198	1713	21,013
15–24	33,046	15,412	40,590	24,416	11,438	20,821
25–34	41,355	12,569	39,928	28,736	9,635	20,203
35–44	91,140	16,710	44,917	57,593	12,012	22,367
45–54	172,385	14,675	40,084	107,722	10,492	19,676
55–64	253,342	8,345	26,602	151,363	5,781	12,784
65+	1,811,720	33,641	35,602	806,431	16,535	14,772
Not stated	357	85	0	282	74	0
Total	2,443,387	106,742	288,357	1,199,264	69,257	141,656

Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. Available from: http://www.cdc.gov/ncipc/wisqars.

Age-adjusted mortality rate: a mortality rate statistically modified to eliminate the effect of different age distributions in the different populations.

Age-adjusted mortality rates

Mortality rates can be used to compare the rates in one area with the rates in another area, or to compare rates over time. However, because mortality rates obviously increase with age, a higher mortality rate among one population than among another might simply reflect the fact that the first population is older than the second.

Consider that the mortality rates in 2002 for the states of Alaska and Florida were 472.2 and 1,005.7 per 100,000, respectively (see Table 3.6). Should everyone from Florida move to Alaska to reduce their risk of death? No, the reason that Alaska's mortality rate is so much lower than Florida's is that Alaska's population is considerably younger. Indeed, for seven age groups, the age-specific mortality rates in Alaska are actually higher than Florida's.

To eliminate the distortion caused by different underlying age distributions in different populations, statistical techniques are used to adjust or standardize the rates among the populations to be compared. These techniques take a weighted average of the age-specific mortality rates, and eliminate the effect of different age distributions among the different populations. Mortality rates computed with these techniques are **age-adjusted** or **age-standardized mortality rates**. Alaska's 2002 age-adjusted mortality rate (794.1 per 100,000) was higher than Florida's (787.8 per 100,000), which is not surprising given that 7 of 13 age-specific mortality rates were higher in Alaska than Florida.

Death-to-case ratio

Definition of death-to-case ratio

The death-to-case ratio is the number of deaths attributed to a particular disease during a specified time period divided by the number of new cases of that disease identified during the same time period. The death-to-case ratio is a ratio but not necessarily a proportion, because some of the deaths that are counted in the numerator might have occurred among persons who developed disease in an earlier period, and are therefore not counted in the denominator.

Table 3.6 All-Cause Mortality by Age Group, Alaska and Florida, 2002

Age Group (years)	ALASKA Population	Deaths	Death Rate (per 100,000)	FLORIDA Population	Deaths	Death Rate (per 100,000)
<1	9,938	55	553.4	205,579	1,548	753.0
1–4	38,503	12	31.2	816,570	296	36.2
5–9	50,400	6	11.9	1,046,504	141	13.5
10–14	57,216	24	41.9	1,131,068	219	19.4
15–19	56,634	43	75.9	1,073,470	734	68.4
20–24	42,929	63	146.8	1,020,856	1,146	112.3
25–34	84,112	120	142.7	2,090,312	2,627	125.7
35–44	107,305	280	260.9	2,516,004	5,993	238.2
45–54	103,039	427	414.4	2,225,957	10,730	482.0
55–64	52,543	480	913.5	1,694,574	16,137	952.3
65–74	24,096	502	2,083.3	1,450,843	28,959	1,996.0
65–84	11,784	645	5,473.5	1,056,275	50,755	4,805.1
85+	3,117	373	11,966.6	359,056	48,486	13,503.7
Unknown	NA	0	NA	NA	43	NA
Total	3,030	3,030	472.2	16,687,068	167,814	1,005.7
Age-adjusted rate:			794.1			787.8

Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. Available from: http://www.cdc.gov/ncipc/wisqars.

Method for calculating death-to-case ratio

$$\frac{\text{Number of deaths attributed to a particular disease during specified period}}{\text{Number of new cases of the disease identified during the specified period}} \times 10^n$$

EXAMPLE: Calculating Death-to-Case Ratios

Between 1940 and 1949, a total of 143,497 incident cases of diphtheria were reported. During the same decade, 11,228 deaths were attributed to diphtheria. Calculate the death-to-case ratio.

Death-to-case ratio = 11,228 / 143,497 x 1 = 0.0783

or

= 11,228 / 143,497 x 100 = 7.83 per 100

Case-fatality rate

The case-fatality rate is the proportion of persons with a particular condition (cases) who die from that condition. It is a measure of the severity of the condition. The formula is:

$$\frac{\textit{Number of cause-specific deaths among the incident cases}}{\textit{Number of incident cases}} \times 10^n$$

The case-fatality rate is a proportion, so the numerator is restricted to deaths among people included in the denominator. The time periods for the numerator and the denominator do not need to be the same; the denominator could be cases of HIV/AIDS diagnosed during the calendar year 1990, and the numerator, deaths among those diagnosed with HIV in 1990, could be from 1990 to the present.

EXAMPLE: Calculating Case-Fatality Rates

In an epidemic of hepatitis A traced to green onions from a restaurant, 555 cases were identified. Three of the case-patients died as a result of their infections. Calculate the case-fatality rate.

Case-fatality rate = (3 / 555) x 100 = 0.5%

The case-fatality rate is a proportion, not a true rate. As a result, some epidemiologists prefer the term **case-fatality ratio**.

The concept behind the case-fatality rate and the death-to-case ratio is similar, but the formulations are different. The death-to-case ratio is simply the number of cause-specific deaths that occurred during a specified time divided by the number of new cases of that disease that occurred during the same time. The deaths included in the numerator of the death-to-case ratio are not restricted to the new cases in the denominator; in fact, for many diseases, the deaths are among persons whose onset of disease was years earlier. In contrast, in the case-fatality rate, the deaths included in the numerator are restricted to the cases in the denominator.

Proportionate mortality

Definition of proportionate mortality

Proportionate mortality describes the proportion of deaths in a specified population over a period of time attributable to different causes. Each cause is expressed as a percentage of all deaths, and the sum of the causes must add to 100%. These proportions are not mortality rates, because the denominator is all deaths rather than the population in which the deaths occurred.

Method for calculating proportionate mortality

For a specified population over a specified period,

$$\frac{\textit{Deaths caused by a particular cause}}{\textit{Deaths from all causes}} \quad x \;\; 100$$

The distribution of primary causes of death in the United States in 2003 for the entire population (all ages) and for persons ages 25–44 years are provided in Table 3.1. As illustrated in that table, accidents (unintentional injuries) accounted for 4.3% of all deaths, but 21.6% of deaths among 25–44 year olds.[8]

Sometimes, particularly in occupational epidemiology, proportionate mortality is used to compare deaths in a population of interest (say, a workplace) with the proportionate mortality in the broader population. This comparison of two proportionate mortalities is called a **proportionate mortality ratio**, or PMR for short. A PMR greater than 1.0 indicates that a particular cause accounts for a greater proportion of deaths in the population of interest than you might expect. For example, construction workers may be more likely to die of injuries than the general population.

However, PMRs can be misleading, because they are not based on mortality rates. A low cause-specific mortality rate in the population of interest can elevate the proportionate mortalities for all of the other causes, because they must add up to 100%. Those workers with a high injury-related proportionate mortality very likely have lower proportionate mortalities for chronic or disabling conditions that keep people out of the workforce. In other words, people who work are more likely to be healthier than the population as a whole — this is known as the healthy worker effect.

Years of potential life lost

Definition of years of potential life lost

Years of potential life lost (YPLL) is one measure of the impact of premature mortality on a population. Additional measures incorporate disability and other measures of quality of life. YPLL is calculated as the sum of the differences between a predetermined end point and the ages of death for those who died before that end point. The two most commonly used end points are age 65 years and average life expectancy.

The use of YPLL is affected by this calculation, which implies a value system in which more weight is given to a death when it occurs at an earlier age. Thus, deaths at older ages are "devalued." However, the YPLL before age 65 ($YPLL_{65}$) places much more emphasis on deaths at early ages than does YPLL based on remaining life expectancy ($YPLL_{LE}$). In 2000, the remaining life expectancy was 21.6 years for a 60-year-old, 11.3 years for a 70-year-old, and 8.6 for an 80-year-old. $YPLL_{65}$ is based on the fewer than 30% of deaths that occur among persons younger than 65. In contrast, YPLL for life expectancy ($YPLL_{LE}$) is based on deaths among persons of all ages, so it more closely resembles crude mortality rates.[10]

YPLL rates can be used to compare YPLL among populations of different sizes. Because different populations may also have different age distributions, YPLL rates are usually age-adjusted to eliminate the effect of differing age distributions.

Method for calculating YPLL from a line listing

Step 1. Decide on end point (65 years, average life expectancy, or other).

Step 2. Exclude records of all persons who died at or after the end point.

Step 3. For each person who died before the end point, calculate that person's YPLL by subtracting the age at death from the end point.

$$YPLL_{individual} = \text{end point} - \text{age at death}$$

Step 4. Sum the individual YPLLs.

$$YPLL = \sum YPLL_{individual}$$

Method for calculating YPLL from a frequency

Step 1. Ensure that age groups break at the identified end point (e.g., 65 years). Eliminate all age groups older than the endpoint.

Step 2. For each age group younger than the end point, identify the midpoint of the age group, where midpoint =

$$\frac{\text{age group's youngest age in years + oldest age + 1}}{2}$$

Step 3. For each age group younger than the end point, identify that age group's YPLL by subtracting the midpoint from the end point.

Step 4. Calculate age-specific YPLL by multiplying the age group's YPLL times the number of persons in that age group.

Step 5. Sum the age-specific YPLL's.

The **YPLL rate** represents years of potential life lost per 1,000 population below the end-point age, such as 65 years. YPLL rates should be used to compare premature mortality in different populations, because YPLL does not take into account differences in population sizes.

The formula for a YPLL rate is as follows:

$$\frac{\textit{Years of potential life lost}}{\textit{Population under age 65 years}} \times 10^n$$

EXAMPLE: Calculating YPLL and YPLL Rates

Use the data in Tables 3.9 and 3.10 to calculate the leukemia-related mortality rate for all ages, mortality rate for persons under age 65 years, YPLL, and YPLL rate.

1. Leukemia-related mortality rate, all ages

= (21,498 / 288,357,000) x 100,000 = 7.5 leukemia deaths per 100,000 population

2. Leukemia-related mortality rate for persons under age 65 years

$$= \frac{125 + 316 + 472 + 471 + 767 + 1{,}459 + 2{,}611}{(19{,}597 + 41{,}037 + 40{,}590 + 39{,}928 + 44{,}917 + 40{,}084 + 26{,}602)} = \times 100{,}000$$

= 6,221 / 252,755,000 = x 100,000

= 2.5 leukemia deaths per 100,000 persons under age 65 years

3. Leukemia-related YPLL

a. Calculate the midpoint of each age interval. Using the previously shown formula, the midpoint of the age group 0–4 years is (0 + 4 + 1) / 2, or 5 / 2, or 2.5 years. Using the same formula, midpoints must be determined for each age group up to and including the age group 55 to 64 years (see column 3 of Table 3.10).

b. Subtract the midpoint from the end point to determine the years of potential life lost for a particular age group. For the age group 0–4 years, each death represents 65 minus 2.5, or 62.5 years of potential life lost (see column 4 of Table 3.10).

c. Calculate age-specific years of potential life lost by multiplying the number of deaths in a given age group by its years of potential life lost. For the age group 0–4 years, 125 deaths x 62.5 = 7,812.5 YPLL (see column 5 of Table 3.10).

d. Total the age-specific YPLL. The total YPLL attributed to leukemia in the United States in 2002 was 117,033 years (see Total of column 5, Table 3.10).

4. Leukemia-related YPLL rate

= $YPLL_{65}$ rate
= YPLL divided by population to age 65
= (117,033 / 252,755,000) x 1,000
= 0.5 YPLL per 1,000 population under age 65

Table 3.9 Deaths Attributed to HIV or Leukemia by Age Group, United States, 2002

Age group (Years)	Population (X 1,000)	Number of HIV Deaths	Number of Leukemia Deaths
0–4	19,597	12	125
5–14	41,037	25	316
15–24	40,590	178	472
25–34	39,928	1,839	471
35–44	44,917	5,707	767
45–54	40,084	4,474	1,459
55–64	26,602	1,347	2,611
65+	35,602	509	15,277
Not stated		4	0
Total	288,357	14,095	21,498

Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. Available from: http://www.cdc.gov/ncipc/wisqars.

Table 3.10 Deaths and Years of Potential Life Lost Attributed to Leukemia by Age Group, United States, 2002

Column 1 Age Group (years)	Column 2 Deaths	Column 3 Age Midpoint	Column 4 Years to 65	Column 5 YPLL
0–4	125	2.5	62.5	7,813
5–14	316	10	55	17,380
15–24	472	20	45	21,240
25–34	471	30	35	16,485
35–44	767	40	25	19,175
45–54	1,459	50	15	21,885
55–64	2,611	60	5	13,055
65+	15,277	—	—	—
Not stated	0	—	—	—
Total	21,498			117,033

Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. Available from: http://www.cdc.gov/ncipc/wisqars.

Natality (Birth) Measures

Natality measures are population-based measures of birth. These measures are used primarily by persons working in the field of maternal and child health. Table 3.11 includes some of the commonly used measures of natality.

Table 3.11 Frequently Used Measures of Natality

Measure	Numerator	Denominator	10^n
Crude birth rate	Number of live births during a specified time interval	Mid-interval population	1,000
Crude fertility rate	Number of live births during a specified time interval	Number of women ages 15–44 years at mid-interval	1,000
Crude rate of natural increase	Number of live births minus number of deaths during a specified time interval	Mid-interval population	1,000
Low-birth weight ratio	Number of live births <2,500 grams during a specified time interval	Number of live births during the same time interval	100

Measures of Association

The key to epidemiologic analysis is comparison. Occasionally you might observe an incidence rate among a population that seems high and wonder whether it is actually higher than what should be expected based on, say, the incidence rates in other communities. Or, you might observe that, among a group of case-patients in an outbreak, several report having eaten at a particular restaurant. Is the restaurant just a popular one, or have more case-patients eaten there than would be expected? The way to address that concern is by comparing the observed group with another group that represents the expected level.

A measure of association quantifies the relationship between exposure and disease among the two groups. Exposure is used loosely to mean not only exposure to foods, mosquitoes, a partner with a sexually transmissible disease, or a toxic waste dump, but also inherent characteristics of persons (for example, age, race, sex), biologic characteristics (immune status), acquired characteristics (marital status), activities (occupation, leisure activities), or conditions under which they live (socioeconomic status or access to medical care).

The measures of association described in the following section compare disease occurrence among one group with disease occurrence in another group. Examples of measures of association include risk ratio (relative risk), rate ratio, odds ratio, and proportionate mortality ratio.

Risk ratio

Definition of risk ratio

A risk ratio (RR), also called relative risk, compares the risk of a health event (disease, injury, risk factor, or death) among one group with the risk among another group. It does so by dividing the risk (incidence proportion, attack rate) in group 1 by the risk (incidence proportion, attack rate) in group 2 . The two groups are typically differentiated by such demographic factors as sex (e.g., males versus females) or by exposure to a suspected risk factor (e.g., did or did not eat potato salad). Often, the group of primary interest is labeled the exposed group, and the comparison group is labeled the unexposed group.

Method for Calculating risk ratio

The formula for risk ratio (RR) is:

$$\frac{\textit{Risk of disease (incidence proportion, attack rate) in group of primary interest}}{\textit{Risk of disease (incidence proportion, attack rate) in comparison group}}$$

A risk ratio of 1.0 indicates identical risk among the two groups. A risk ratio greater than 1.0 indicates an increased risk for the group in the numerator, usually the exposed group. A risk ratio less than 1.0 indicates a decreased risk for the exposed group, indicating that perhaps exposure actually protects against disease occurrence.

EXAMPLES: Calculating Risk Ratios

Example A: *In an outbreak of tuberculosis among prison inmates in South Carolina in 1999, 28 of 157 inmates residing on the East wing of the dormitory developed tuberculosis, compared with 4 of 137 inmates residing on the West wing.[11] These data are summarized in the two-by-two table so called because it has two rows for the exposure and two columns for the outcome. Here is the general format and notation.*

Table 3.12A General Format and Notation for a Two-by-Two Table

	Ill	Well	Total
Exposed	a	b	$a + b = H_1$
Unexposed	c	d	$c + d = H_0$
Total	$a + c = V_1$	$b + d = V_0$	T

In this example, the exposure is the dormitory wing and the outcome is tuberculosis) illustrated in Table 3.12B. Calculate the risk ratio.

Table 3.12B Incidence of Mycobacterium Tuberculosis Infection Among Congregated, HIV-Infected Prison Inmates by Dormitory Wing, South Carolina, 1999

	Developed tuberculosis?		
	Yes	**No**	**Total**
East wing	a = 28	b = 129	H_1 =157
West wing	c = 4	d = 133	H_0=137
Total	32	262	T=294

Data source: McLaughlin SI, Spradling P, Drociuk D, Ridzon R, Pozsik CJ, Onorato I. Extensive transmission of Mycobacterium tuberculosis among congregated, HIV-infected prison inmates in South Carolina, United States. Int J Tuberc Lung Dis 2003;7:665–672.

To calculate the risk ratio, first calculate the risk or attack rate for each group. Here are the formulas:

Attack Rate (Risk)

Attack rate for exposed = a / a+b

Attack rate for unexposed = c / c+d

For this example:

Risk of tuberculosis among East wing residents = 28 / 157 = 0.178 = 17.8%

Risk of tuberculosis among West wing residents = 4 / 137 = 0.029 = 2.9%

The risk ratio is simply the ratio of these two risks:

Risk ratio = 17.8 / 2.9 = 6.1

Thus, inmates who resided in the East wing of the dormitory were 6.1 times as likely to develop tuberculosis as those who resided in the West wing.

EXAMPLES: Calculating Risk Ratios

Example B: *In an outbreak of varicella (chickenpox) in Oregon in 2002, varicella was diagnosed in 18 of 152 vaccinated children compared with 3 of 7 unvaccinated children. Calculate the risk ratio.*

Table 3.13 Incidence of Varicella Among Schoolchildren in 9 Affected Classrooms, Oregon, 2002

	Varicella	Non-case	Total
Vaccinated	a = 18	b = 134	152
Unvaccinated	c = 3	d = 4	7
Total	21	138	159

Data Source: Tugwell BD, Lee LE, Gillette H, Lorber EM, Hedberg K, Cieslak PR. Chickenpox outbreak in a highly vaccinated school population. Pediatrics 2004 Mar;113(3 Pt 1):455–459.

Risk of varicella among vaccinated children = 18 / 152 = 0.118 = 11.8%
Risk of varicella among unvaccinated children = 3 / 7 = 0.429 = 42.9%

Risk ratio = 0.118 / 0.429 = 0.28

The risk ratio is less than 1.0, indicating a decreased risk or protective effect for the exposed (vaccinated) children. The risk ratio of 0.28 indicates that vaccinated children were only approximately one-fourth as likely (28%, actually) to develop varicella as were unvaccinated children.

Rate ratio

A rate ratio compares the incidence rates, person-time rates, or mortality rates of two groups. As with the risk ratio, the two groups are typically differentiated by demographic factors or by exposure to a suspected causative agent. The rate for the group of primary interest is divided by the rate for the comparison group.

$$\text{Rate ratio} = \frac{\textit{Rate for group of primary interest}}{\textit{Rate for comparison group}}$$

The interpretation of the value of a rate ratio is similar to that of the risk ratio. That is, a rate ratio of 1.0 indicates equal rates in the two groups, a rate ratio greater than 1.0 indicates an increased risk for the group in the numerator, and a rate ratio less than 1.0 indicates a decreased risk for the group in the numerator.

EXAMPLE: Calculating Rate Ratios

Public health officials were called to investigate a perceived increase in visits to ships' infirmaries for acute respiratory illness (ARI) by passengers of cruise ships in Alaska in 1998.[13] The officials compared passenger visits to ship infirmaries for ARI during May–August 1998 with the same period in 1997. They recorded 11.6 visits for ARI per 1,000 tourists per week in 1998, compared with 5.3 visits per 1,000 tourists per week in 1997. Calculate the rate ratio.

$$\text{Rate ratio} = 11.6 / 5.3 = 2.2$$

Passengers on cruise ships in Alaska during May–August 1998 were more than twice as likely to visit their ships' infirmaries for ARI than were passengers in 1997. (Note: Of 58 viral isolates identified from nasal cultures from passengers, most were influenza A, making this the largest summertime influenza outbreak in North America.)

Odds ratio

An odds ratio (OR) is another measure of association that quantifies the relationship between an exposure with two categories and health outcome. Referring to the four cells in Table 3.15, the odds ratio is calculated as

$$Odds\ ratio\ =\ \left(\frac{a}{b}\right)\left(\frac{c}{d}\right)\ =\ ad\,/\,bc$$

where

a	=	number of persons exposed and with disease
b	=	number of persons exposed but without disease
c	=	number of persons unexposed but with disease
d	=	number of persons unexposed: and without disease
a+c	=	total number of persons with disease (case-patients)
b+d	=	total number of persons without disease (controls)

The odds ratio is sometimes called the **cross-product ratio** because the numerator is based on multiplying the value in cell "a" times the value in cell "d," whereas the denominator is the product of cell "b" and cell "c." A line from cell "a" to cell "d" (for the numerator) and another from cell "b" to cell "c" (for the denominator) creates an x or cross on the two-by-two table.

Table 3.15 Exposure and Disease in a Hypothetical Population of 10,000 Persons

	Disease	No Disease	Total	Risk
Exposed	a=100	b=1,900	2,000	5.0%
Not Exposed	c=80	d=7,920	8,000	1.0%
Total	180	9,820	10,000	

EXAMPLE: Calculating Odds Ratios

Use the data in Table 3.15 to calculate the risk and odds ratios.

1. Risk ratio

5.0 / 1.0 = 5.0

2. Odds ratio

(100 x 7,920) / (1,900 x 80) = 5.2

Notice that the odds ratio of 5.2 is close to the risk ratio of 5.0. That is one of the attractive features of the odds ratio — when the health outcome is uncommon, the odds ratio provides a reasonable approximation of the risk ratio. Another attractive feature is that the odds ratio can be calculated with data from a case-control study, whereas neither a risk ratio nor a rate ratio can be calculated.

The odds ratio is the measure of choice in a case-control study (see Lesson 1). A case-control study is based on enrolling a group of persons with disease ("case-patients") and a comparable group without disease ("controls"). The number of persons in the control group is usually decided by the investigator. Often, the size of the population from which the case-patients came is not known. As a result, risks, rates, risk ratios or rate ratios cannot be calculated from the typical case-control study. However, you can calculate an odds ratio and interpret it as an approximation of the risk ratio, particularly when the disease is uncommon in the population.

In a case-control study, investigators enroll a group of case-patients (distributed in cells a and c of the two-by-two table), and a group of non-cases or controls (distributed in cells b and d).

Measures of Public Health Impact

A measure of public health impact is used to place the association between an exposure and an outcome into a meaningful public health context. Whereas a measure of association quantifies the relationship between exposure and disease, and thus begins to provide insight into causal relationships, measures of public health impact reflect the burden that an exposure contributes to the frequency of disease in the population. Two measures of public health impact often used are the attributable proportion and efficacy or effectiveness.

Attributable proportion

Definition of attributable proportion

The attributable proportion, also known as the attributable risk percent, is a measure of the public health impact of a causative factor. The calculation of this measure assumes that the occurrence of disease in the unexposed group represents the baseline or expected risk for that disease. It further assumes that if the risk of disease in the exposed group is higher than the risk in the unexposed group, the difference can be attributed to the exposure. Thus, the attributable proportion is the amount of disease in the exposed group attributable to the exposure. It represents the expected reduction in disease if the exposure could be removed (or never existed).

Appropriate use of attributable proportion depends on a single risk factor being responsible for a condition. When multiple risk factors may interact (e.g., physical activity and age or health status), this measure may not be appropriate.

Method for calculating attributable proportion

Attributable proportion is calculated as follows:

$$\frac{\textit{Risk for exposed group} - \textit{risk for unexposed group}}{\textit{Risk for exposed group}} \; x \; 100\%$$

Attributable proportion can be calculated for rates in the same way.

EXAMPLE: Calculating Attributable Proportion

In another study of smoking and lung cancer, the lung cancer mortality rate among nonsmokers was 0.07 per 1,000 persons per year.[14] *The lung cancer mortality rate among persons who smoked 1–14 cigarettes per day was 0.57 lung cancer deaths per 1,000 persons per year. Calculate the attributable proportion.*

Attributable proportion = (0.57 – 0.07) / 0.57 x 100% = 87.7%

Given the proven causal relationship between cigarette smoking and lung cancer, and assuming that the groups are comparable in all other ways, one could say that about 88% of the lung cancer among smokers of 1-14 cigarettes per day might be attributable to their smoking. The remaining 12% of the lung cancer cases in this group would have occurred anyway.

Vaccine efficacy or vaccine effectiveness

Vaccine efficacy and vaccine effectiveness measure the proportionate reduction in cases among vaccinated persons. Vaccine efficacy is used when a study is carried out under ideal conditions, for example, during a clinical trial. Vaccine effectiveness is used when a study is carried out under typical field (that is, less than perfectly controlled) conditions.

Vaccine efficacy/effectiveness (VE) is measured by calculating the risk of disease among vaccinated and unvaccinated persons and determining the percentage reduction in risk of disease among vaccinated persons relative to unvaccinated persons. The greater the percentage reduction of illness in the vaccinated group, the greater the vaccine efficacy/effectiveness. The basic formula is written as:

$$\frac{\textit{Risk among unvaccinated group} - \textit{risk among vaccinated group}}{\textit{Risk among unvaccinated group}}$$

OR: $1 - \textit{risk ratio}$

In the first formula, the numerator (risk among unvaccinated – risk among vaccinated) is sometimes called the risk difference or excess risk.

Vaccine efficacy/effectiveness is interpreted as the proportionate reduction in disease among the vaccinated group. So a VE of 90% indicates a 90% reduction in disease occurrence among the vaccinated group, or a 90% reduction from the number of cases you would expect if they have not been vaccinated.

EXAMPLE: Calculating Vaccine Effectiveness

Calculate the vaccine effectiveness from the varicella data in Table 3.13.

VE = (42.9 – 11.8) / 42.9 = 31.1 / 42.9 = 72%

Alternatively, VE = 1 – RR = 1 – 0.28 = 72%

So, the vaccinated group experienced 72% fewer varicella cases than they would have if they had not been vaccinated.