In this 4-part series, we will discuss epidemiologic measures of disease burden and distribution.

In this third module, I present information related to epidemiologic measures that are commonly used to summarize the distribution of disease in a population.
Learning Objectives

• In epidemiologic studies,
  – Define ratio, proportion, rate and odds as measures of frequency
  – Calculate cumulative incidence, incidence rate and prevalence proportion and understand the difference among these measures
  – Identify and calculate commonly used measures of morbidity
EPIDEMIOLOGIC MEASURES
Let’s consider an example. What if 10 people were diagnosed with salmonella?

Does this indicate a public health concern? What additional information would you like to know?
We would want to know more information about the population that gave rise to the cases, in particular, the size of the population.

If 10 cases occurred in a dorm housing 100 students, then 10% became ill. This is different from a situation in which 10 cases occurred in state of OK (pop=3,579,212), where 2.8 cases per 1,000,000 became ill.

For a count to be descriptive, it must be considered relative to the population. A central concern in epidemiology is appropriate denominators and calculations of ratios to measure disease burden.
In summarizing disease distribution and burden, we need to define several measures.

Let’s first begin by focusing on the numerator of rates or ratios.

A count is the number of cases of a particular health event. Examples would include the number of high school students who smoke or the number of traffic fatalities in Oklahoma County during a 24-hour period.

Remember the definition of epidemiology: the study of the distribution and determinants of disease in human populations. Epidemiology is used to:

1) Describe patterns of disease (or the distribution of disease occurrence)
2) Identify disease determinants

To achieve either of these goals, it is necessary to measure the frequency of disease. We measure frequency of disease using counts, ratios, proportions, rates and odds.

Counts are essential for public health planners and health administrators who need
to determine the allocation of health care resources in a particular community; however, count data alone has limited utility for epidemiology.

Appropriate denominators are needed to compare and describe groups in a meaningful and useful way. For some diseases one case is sufficient to be of public health concern (e.g., Smallpox, Ebola virus!) – and would spark an immediate investigation. But in general, a simple count of cases is not very useful for comparisons without a meaningful reference.
As we think about denominators in epidemiology, let’s first consider a simple ratio, which is calculated by dividing one quantity by another. In general, there is no specified relationship between numerator and denominator.

An example would be the sex ratio for motorcycle fatalities 19:1 male to female, 2.1 male cases to 1 female case or 253 persons/square mile.

A proportion, in contrast, is a type of ratio in which the numerator is contained in the denominator. A ratio indicates what fraction of the population is affected and is expressed as a percentage, decimal proportion (0.0-1.0) or fraction. An example is 0.25 (25%) of Oklahoma adults are current smokers. Other types of proportions include the prevalence and cumulative incidence, which we will define later in this lecture.
A rate is a specific type of ratio that involves a measure of time. Time is key in interpreting a rate. As an example, we may estimate the velocity of a car and report the result as km/hour or estimate the incidence rate of cancer as the number of cases per unit of time at risk.

The odds is another ratio that is calculated as the division of a proportion by its complement, meaning, dividing the probability of an event by the probability of the event NOT occurring. As an example, in horse racing there may be a 3:1 odds of winning for a particular horse, which would arise if the probability of the horse is winning is 0.75.
Let’s look more closely at the calculation of the odds. The odds are calculated as the division of a proportion by its complement. If a horse has a 75% chance of winning, the probability of losing is 25%. Therefore, the odds of the horse winning would be 0.75/0.25 = 3:1.
Let’s consider another example of odds.

If there are 100 smokers and 60 develop a chronic cough, the probability of a smoker developing a cough is \( \frac{60}{100} = 0.60 \).

The probability of a smoker not developing a cough is \( \frac{40}{100} = 0.40 \).

Therefore, the odds in favor of developing a cough are 60:40 or 1.5:1.
### Terminology

- Strict use of term “rate” refers only to ratios representing changes over time
  - e.g., incidence density

- Note: medical literature may use terms “proportion”, “ratio” and “rate” interchangeably
  - e.g., prevalence “rate”, cumulative incidence “rate”

Keep in mind that a rate indicates how fast disease is occurring in a population. The essential characteristic of a rate is that time is an element in the denominator. It is important to be aware of how reported measures are actually defined and calculated. Terminology reported in the literature is not always accurate.
Now, let’s consider a general measure of disease frequency.

1) The count of events is the numerator
2) The defined population in which the events occur is the denominator. It refers to the size of the population from which the cases have been observed.
3) Person time combines the population at risk with the observation time
4) We then apply an arbitrary multiplier of some power of 10 (usually 1,000 or 100,000)

Different rates are expressed in a particular convention, e.g., cancer rates per 100,000 and infant mortality per 1,000 live births. The size of the multiplier is usually determined by whether the value/rate can be expressed as a whole number.

E.g., 4 per 100,000 is preferable to 0.04 per 1,000.
Morbidity refers to the presence of disease in a population.

Mortality refers to the occurrence of death in a population. Mortality will be the topic of the next module.

Most common measures of disease frequency fall into two broad categories: prevalence and incidence. The important distinction is that prevalent cases reflect existing cases in a population at a given point or period in time while incident cases reflect NEW cases identified over a specified period of time.

In the next slides, we will review specific measures of prevalence and incidence.
Prevalence quantifies the proportion of people in a population who have the disease at a specific point in time.

Prevalence is a proportion, not a rate. Prevalence data provide an indication of the extent of a health problem.

These data are useful for determining the health burden in a population and the scope of health services and resources needed in a community such as facilities and personnel.
Let’s consider an example. If we find that among 43,732 adults surveyed on June 13, 2009, 12,763 reported depression, the prevalence proportion would be 0.2918 or 291.8 per 1,000 adults surveyed on June 13, 2009.

We wonder if this is a reasonable estimate for other time points or period. Would point prevalence be the same in November?
There are two kinds of prevalence measures: point and period.

Point can refer to a specific point in calendar time or to a fixed point in the course of events that varies from person to person in real time, like onset of menopause or onset of puberty.

Period prevalence, on the other hand, reflects a period of time. Period prevalence is not frequently used. Period prevalence assumes a stable population for estimation. Period prevalence is sometimes used when it's difficult to determine when a disease is considered to be present, e.g., dx of mental illness.

Prevalence data is typically not helpful for studies of disease etiology. This is because of the possible influence of differential survival. Measures of disease etiology, or onset, will be discussed in more detail in upcoming slides.
There are several necessary criteria for calculating prevalence proportion.

First, we need a precise definition of who is diseased vs. not diseased. For example, hypertension may be defined as a self-report of hypertension; defined as a doctor’s diagnosis of hypertension (yes/no); defined as a pressure over 140/90 mm Hg; or defined as mild, moderate, severe or very severe hypertension (using various cut-points of diastolic and systolic pressures). We may utilize a variety of information sources including information from health records, registries, screening, and interviews.

It is important to remember that when comparing rates across different populations, the data sources and quality of data must be equivalent for the groups being compared. For example, we would not want to compare self-reported interview data to cancer registry data because the data accuracy and definitions are not comparable.

As a second piece of information, we need to know the number of existing cases of disease. Be clear whether measure represent number of events or number of individuals. As an example, we may calculate the frequency of myopia (nearsightedness) in a population of school children and this could represent the number of affected eyes in relation to total number of eyes or number of children affected in one or both eyes in relation to the total number of school
children. In general, epidemiologic rates refer to persons, not events, ex: # people who
develop a cold vs. # of colds.

Be sure that you understand the units being used in the numerator and denominator.

Next, we need to know the size of the total population and finally, we will focus on a period in time or a period of observation.

Prevalence is generally the preferred measures when it is difficult to define the onset of
disease (such as asthma) and is sometimes preferred for diseases of long duration (chronic conditions like arthritis or diabetes).
Now, let’s define cumulative incidence, which is a measure of risk. This is a measure of the probability that health people will develop the disease during a specified time.

The numerator = frequency of new cases.

The Denominator = population at risk, meaning individuals who have the potential to become a case. This is the number of individuals free of disease at the start of the study period and must be susceptible (at risk). We will exclude those who have already developed disease or are not capable of developing disease. For example, when calculating the incidence of ovarian cancer, the denominator is restricted to women w/ intact ovaries.

Incidence is a measure of risk because it’s a measure of transition from a non-diseased state to a diseased state.
In order to calculate the cumulative incidence, there are several criteria and pieces of information that must be available.

1) There must be adequate grounds on which to assess the health status of individuals in a population and to classify people as diseased or not diseased.

2) Unlike prevalence, for cumulative incidence you need to know when disease onset occurred.

For infectious diseases or other acute conditions like severe stroke or myocardial infarction, determining onset can be relatively simple and pinpointed to a specific hour. In contrast, chronic conditions like cancer have an indefinite time of onset so we may define onset by date of earliest definitive diagnosis (rather than the date symptoms appeared).

3) The numerator is then the number of new cases of disease during the
time period where existing cases are excluded from both the numerator and denominator.
4). We then determine the population at risk. Those who cannot develop disease for reasons such as age, immunization or prior removal of the involved organ should, in principle, be excluded from the denominator. It is often not possible to determine this info for each individual in a population, so correction not always made. The necessity for excluding these from the denominator depends on what proportion of the total population they make up. In general, including people who are not at risk in the denominator will result in an underestimate of the true incidence of disease. For most chronic diseases, the proportion of the general pop with a history of the condition is small, so failing to exclude these from the denominator will not have much affect on the incidence rate.

5). The period of observation is then specified, for example, one year.
Now, let’s consider an example. If we consider an estimate for the risk of West Nile virus infection in 2002. If we observe 22 cases in 2002 among the population of Oklahoma, the cumulative incidence is 0.63 new cases of WNV per 100,000 (for time period 2002 in OK). Important – assumes everybody is AT risk of WNV at the start of 2002.
When calculating numerator terms for the measures that we have discussed, keep in mind that the wording of the question is important. For example, asking “Do you currently have asthma?” could be used to estimate point prevalence. Asking “Have you had asthma during the last 2 months?” could be used to estimate period prevalence. Finally, asking “have you ever had asthma?” could be used to estimate the cumulative incidence of asthma over the lifetime of the respondent.
Recognizing that the follow-up time or time at risk may not be the same for all participants, we can calculate what is called the Incidence Rate or Incidence Density.

This measure is useful when individuals are observed for different lengths of time and provides a measure of how quickly people are acquiring the disease.

The numerator reflects the number of new cases during the time period while the denominator reflects the person-time at risk during the study period.
Let’s look more carefully at the measure of person time.

Person time is an estimate of the actual time at risk in months, months, or days that all persons contributed to the study. It is accumulated over the duration of the study and for all participants. Keep in mind that a subject is eligible to contribute person-time to the study as long as they remain susceptible to the disease of interest.
Let’s consider an example of the person-time calculation.

Consider a cohort of 11 patients who are followed over time. The patients will contribute differing amounts of time at risk.

In the first period, 10 patients complete the follow-up.

Then, one patient ends follow-up (e.g., moves out of the area) and one patient develops the event of interest. A third patient dies soon after the first follow-up visit. The remaining seven patients complete the next follow-up period. Then, one of these seven experiences the event soon after the follow-up time point and an additional patient is added to the cohort. Next two additional events are observed, resulting in five patients who complete the follow-up period.
Now, let's consider the time at risk.

We can calculate the total person time at risk by calculating the number of person-months accumulated at each study time point.

In the first period, all 10 patients complete 20 months of follow-up, resulting in 200 total person-months at risk.

In the second period, one patient dies within 1 month, one experiences the event of interest within 10 months and one is lost to follow-up after 10 months. Therefore, the person-months contributed during the second period is 231 months.

In the next period of 40 months of follow-up, an 11th patient has been added to the cohort. Among the 8 patients, events are experienced within 1, 20 and 30 months while the others complete 40 months of follow-up.

Finally, five remaining patients complete the 10-month final follow-up period.

The total person-months at risk is 732.

From this example, we see that the calculation of person-months at risk is much more complicated when the time at risk varies by patient.
Incidence Rate Units

- 732 person-months at risk
  - $732 \div 12 = 61.25$ person-months at risk
  - $732 \times 30.44 = 22,282$ person-days at risk
- Number of incident cases = 4
- Incidence Rate (Incidence Density)
  - $4 \div 732 = 0.0055$ case / person-month at risk
  - 5.5 cases / 1000 person-months at risk
  - 0.066 case / person-year at risk
  - 6.6 cases / 100 person-years at risk
  - 6,600 / 100,000 person-years at risk

We can express the incidence rate in a variety of units depending on the unit of time.

The numerator is the number of cases and the denominator is the number of person-months at risk.

We can multiply 0.0055 case/person month by 12 months (in a year) to get 0.0066 cases/person-year.

Then, we can multiply by factors of 10 to have a more easily interpretable estimates, such as 6.6 cases per 100 person-years at risk.
As a hospital-based example, we may want to estimate the incidence rate of Nosocomial infections. If we observe 53 infections in 1,290 patient-days at risk, we would estimate an incidence rate of 4.1 nosocomial infections per 100 patient days (for six-month time period).
It is important to distinguish incidence from prevalence. Both are measures of the extent of disease in a population. Incidence reflects probability of developing a disease (new onset) while prevalence reflects the probability of already having disease.

Prevalence is not a measure of risk because it does not reflect the onset of new cases. Prevalence is used to express “burden of disease” in a population and is used by health planners for determining workload for monitoring control programs for chronic conditions. Note that if incidence not available, can use prevalence to estimate importance of disease.
In contrast, incidence tells us about the change in status from non-diseased to diseased (limited to new cases). Incidence reflects etiology as a direct indicator of risk of disease (cumulative incidence). We can compare incidence rates in populations to better understand or to develop hypotheses related to disease etiology.

Note that prevalence is limited as a measure of disease etiology. Prevalence favors inclusion of chronic over acute cases. Also, with a cross-sectional assessment, we do not observe the onset of disease and therefore, temporality, such as the timing of disease onset relative to exposure, is uncertain.
Keep in mind that prevalence is a function of incidence and average disease duration.

If incidence is low, but those affected have the condition for a long period of time, the proportion of the population that has the disease at a particular point in time will be high relative to the incidence.

Similarly, if the incidence is high but the duration of disease is short, either through quick recovery or death, the prevalence will be low relative to the incidence.

In a steady state situation with low prevalence, where steady state means that the incidence rate is constant (no epidemic or marked reduction in disease AND no major changes have occurred in disease duration) and low prevalence means that the prevalence is < 0.1 (1 in 10).
In this case, prevalence is proportional to the product of the incidence rate and the average duration of the disease. Each new (incident) case enters a prevalence pool and remains there until either recovery or death. The change in prevalence from one time period to another may be the result of changes in incident rates, change in the duration of disease or both.
It is helpful to visualize a BATHTUB analogy where incidence = flow of water into tub. Prevalence = measures proportion of tub filled w/water (proportion of total population made up of existing cases). Death/Recovery represents the prevalent cases that leave the prevalence pool by recovery or death.
Now, let’s consider some review problems.
When would a low incidence produce a high prevalence?

- When recovery and death rates are low

When recovery and death rates are low (few prevalent cases are leaving the pool)
When is prevalence roughly equal to incidence?

- Acute disease (e.g., common cold) with a high recovery rate

In an acute disease (e.g., common cold) setting with a high recovery rate, meaning a high rate of cases leaving the pool.
When is prevalence greater than incidence?

- Chronic disease (e.g., diabetes) with low death and low recovery rate

In a chronic disease (e.g., diabetes) setting with low death and low recovery rate, meaning few cases are leaving the pool or are leaving slowly.
In conclusion, when summarizing disease distribution and burden, we will use measures of frequency, including the ratio, proportion, rate and odds. In considering specific types of ratios, we discussed differences among cumulative incidence, incidence rate and prevalence proportion and understand the difference among these measures. Finally, you are now familiar with identifying and calculating commonly used measures of morbidity. In the next module, we will focus on measures of mortality.