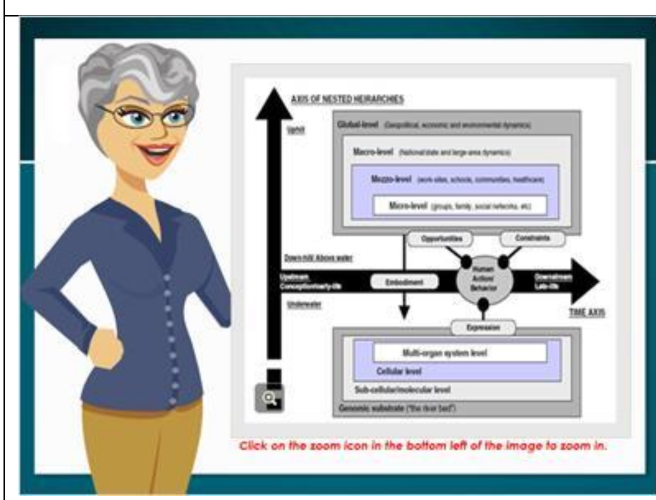
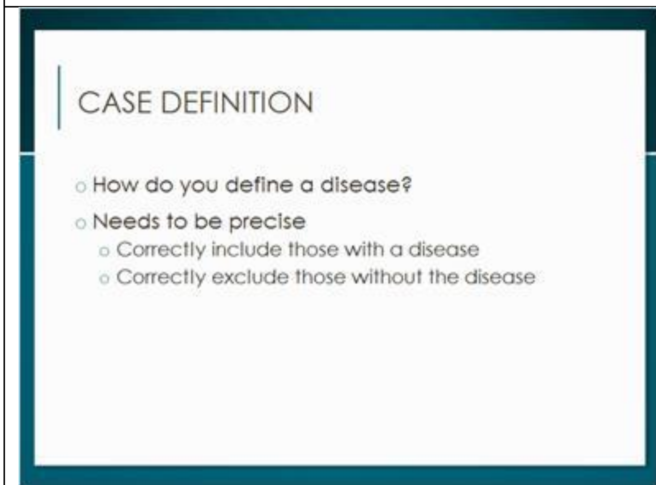
Incidence & Prevalence Handout

<table>
<tr>
<td>
</td>
<td>This module will review common epidemiological measures and also explain how to calculate important measures.</td>
</tr>
<tr>
<td>
</td>
<td>Before we begin looking at epidemiological measures, I wanted to remind you of why this is important. The overall goal of epidemiology is to improve the health of populations. This quote by the World Health Organization made in 1948 remains true today. We want to look at health as a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. In this regard, we care about physical health, mental health, and social health. Because the vision of health is so broad, epidemiology is likewise broad. Epidemiologists not only study cancer, infectious diseases, cardiac disease, but also domestic violence, unintentional injury, handgun safety, depression, etc. The field is as broad as you can imagine.</td>
</tr>
<tr>
<td>
</td>
<td>This somewhat complex graphic illustrates current understanding of the multi-level view of health using a systems approach. It also considers the effect of time. We know that health changes over the course of an individual's life. In addition, health is impacted by many factors. We are effected by our genetics, maternal behaviors during pregnancy, and even behaviors of our grandmothers during their pregnancies. The field of epigenetics has demonstrated that experiences of people can turn on and off certain genes and that these changes in genes are inherited much like our genome is. Traditionally epidemiologists have looked at individual behavior as a risk factor for disease. Smoking causes cancer, or a high fat diet leads to heart disease.   But now we are appreciating that many other factors, our family and friends, work and</td>
</tr>
</table>

school sites, national laws and policies as well as our overall political, economic and physical environment. As we move forward in this class, we need to consider more than just the traditional behavioral factors in studying disease and improving health.



The numerator therefore is obtained by counting people with a disease or other condition. To accurately count people with a disease, we need to have a case definition, i.e., a definition of the disease that allows us to accurately identify people with the disease and exclude people without the disease. Case definitions can be based upon physical symptoms, a diagnostic exam, laboratory tests, or a combination of all three.

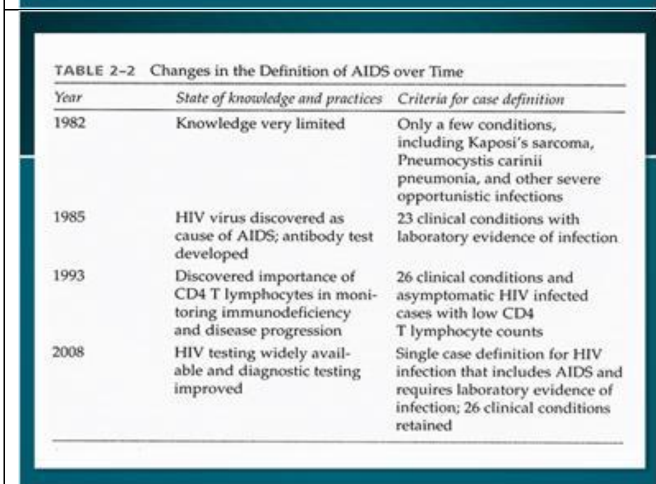
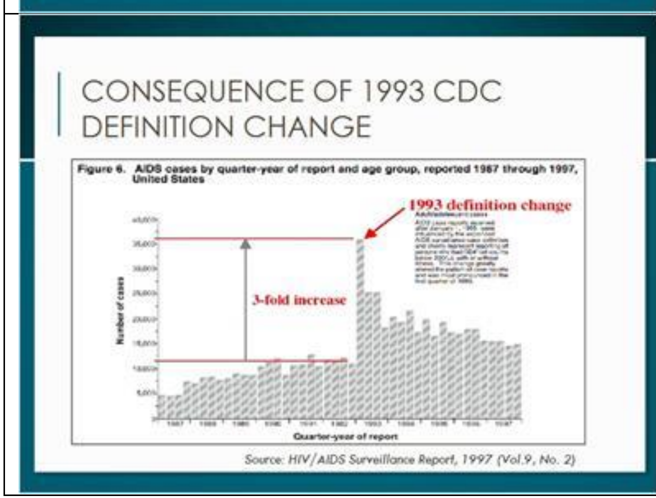


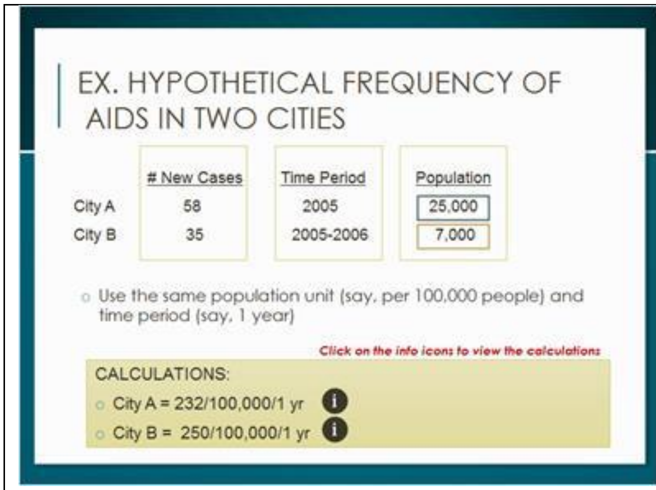
TABLE 2–2 Changes in the Definition of AIDS over Time

Year	*State of knowledge and practices*	*Criteria for case definition*
1982	Knowledge very limited	Only a few conditions, including Kaposi's sarcoma, Pneumocystis carinii pneumonia, and other severe opportunistic infections
1985	HIV virus discovered as cause of AIDS; antibody test developed	23 clinical conditions with laboratory evidence of infection
1993	Discovered importance of CD4 T lymphocytes in monitoring immunodeficiency and disease progression	26 clinical conditions and asymptomatic HIV infected cases with low CD4 T lymphocyte counts
2008	HIV testing widely available and diagnostic testing improved	Single case definition for HIV infection that includes AIDS and requires laboratory evidence of infection; 26 clinical conditions retained

To compound the confusion, case definitions change over time. This table shows the changes in the case definition of AIDS over a number of years. Each time the definition was changed, it included more cases, as more conditions were found to be symptoms of AIDS. Thus, the numbers of people with AIDS increased dramatically over time, more than would have been expected just due to increases in people who became ill.



This graph shows the apparent three fold increase in the number of AIDS cases in 1993 that in reality was due to the change in the case definition which added many conditions to the case definitions.



This slide shows the hypothetical frequency of AIDS in two cities, city A and city B. As you can see the underlying population of the two communities differ. A has 25,000 people and B has 7,000 people. Note that there are three components: cases of AIDS, population at risk, and time period. In order to compare the two populations you need to determine the number of cases per the same denominator, in this case they used 100,000.

The way you do the math is you divide the numerator by the denominator 58/25,000 and then multiply it by the population size you want (100,000).

Try it to be sure you can get these numbers.

Also note that city B covers 2 years and city A covers one year.

So you need to make these equivalent as well by dividing the result by 2. 35/7000=0.005

Then divide by 2 which equals 0.0025. When you multiply this by 100,000, you get 250. So you can see that City B has more AIDS deaths than city A. I know some people will find this math confusing but if you stop and look at it, it actually is pretty straightforward.

I want to take some time to introduce you to very important epidemiological measures: incidence and prevalence

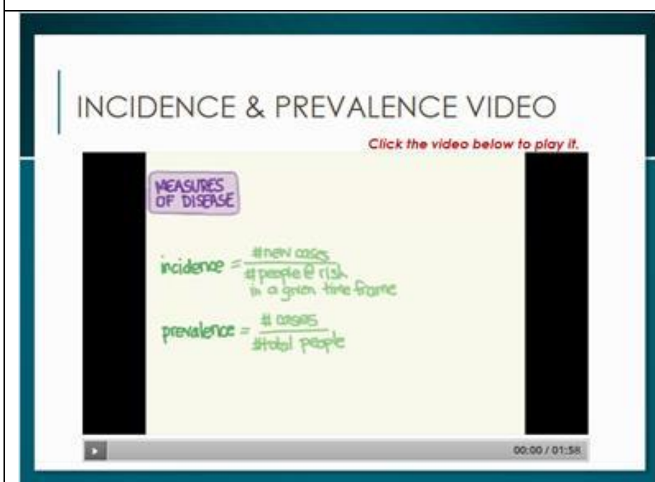
INCIDENCE

Incidence refers to the number of new cases of a disease during a specified time period. When we consider incidence, we don’t count pre-existing cases. One also needs to allow time for cases of a disease to develop.

This slide shows the formula for incidence and how it is calculated. In the denominator, one can either count the number in the population at risk or the amount of time people were followed during the study. I will get to that in a few minutes. In the example shown people in City A were followed for one year to see who developed arthritis. There were 70,000 people in that city and 70 developed arthritis, which is 0.001 or 1 case per thousand. When we talk about incidence, we usually express it as a number per 1,000 people because that is more understandable than 0.001. You can talk about the number of cases per 100, 1,000, 10,000 or whatever. Just use the number that best helps you explain your finding to the general public. If you used the number per 100 in this case, it would be 0.1 per 100 and that would be hard for people to understand. Remember, you want to present your data in a way that is easy to understand. So 1 per 1,000 works really well.

PREVALENCE

Prevalence is an important epidemiological term. It is a proportion. Prevalence refers to the number of cases of a disease divided by everyone at risk for disease. Prevalence is also usually expressed per 1,000 people. You can measure prevalence at a certain time period since you do not have to wait for people to develop the disease or you can measure prevalence during a given time period. In the example below prevalence was measured on January 1st and on that day 7,000 of 70,000 people living in City A had arthritis. This is 10% of the population or 100 cases per each thousand people. In this case you may choose to say 10 cases per every hundred people or 1 in 10, as these are easily understandable. They all equal the same thing.



So, for prevalence, you can just count existing cases but for incidence you need to get them time to get the disease. I am providing you with a clip from a video on prevalence from Dr. Patwari. The link to the full video, entitled Incidence and Prevalence, is included in the course materials. I want to add a comment that when he refers to incidence or incidence proportion, he is actually talking about cumulative incidence.

There are two ways in which we measure incidence, depending upon what we put in the denominator. For Cumulative Incidence, abbreviated CI, we include the number of people in the denominator within a certain time period. For Incidence Rate, abbreviated IR, we use person time in the denominator.

Let me spend a few minutes explaining the difference between the two measures, and why we would choose one over the other.

CUMULATIVE INCIDENCE (CI)

Cumulative incidence is the number of new cases of a disease in a population over a period of time. We first eliminate all people who already have the disease and then count all new cases of disease during the time period. The population which is in the denominator as the people at risk of getting the disease. People who already have the disease are not included in the denominator. This measure is usually used for relatively stable populations where it is assumed that everyone is followed for the same time period.

This slide shows the cumulative e incidence of cases of Sudden Infant Death Syndrome during the first year of life. The denominator consists off all live births during that period, and there were 10 cases of SIDS out of 1,000 with a 1% cumulative incidence rate over 1 year.

CI EXAMPLE

In that example, you assume that you followed the entire population for the entire follow-up period. Often this does not really reflect the reality. Children move in and out of the area and children may die from other causes. In a dynamic population which we discussed previously individuals enter populations over time and some become lost to follow up. So we know in reality, most times the length of follow-up is not uniform for everyone. The other measure of incidence, incidence rates does not assume equal follow up and takes into account the actual follow up time.

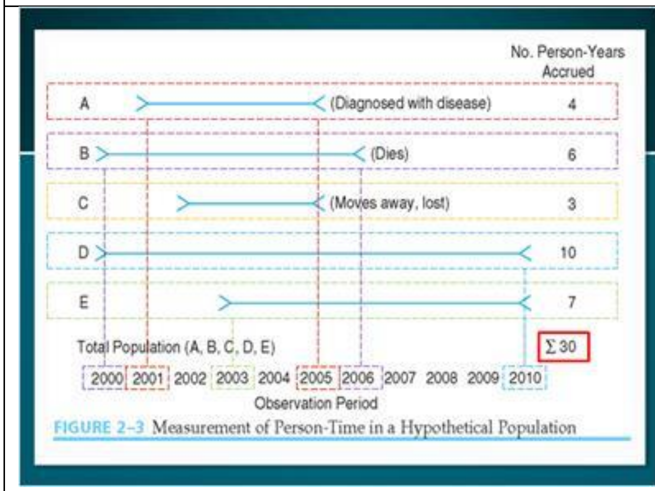
INCIDENCE RATE (IR)

When you use incidence rates, the denominator is the person time of observation. You count the number of new cases of disease divided by the person time of follow up for everyone in a study. This measures is a true rate because it directly considers time in the equation.

This clip, again from Dr. Patwari shows a comparison of what he calls incidence proportion and incidence rate. We use the term cumulative incidence, but it is a nice short review. The larger video, entitled incidence rates reviews what I covered in these slides and it is below for your viewing.

COMPARISON

This clip, again from Dr. Patwari shows a comparison of what he calls incidence proportion and incidence rate. We use the term cumulative incidence, but it is a nice short review. The larger video, entitled incidence rates reviews what I covered in these slides and it is below for your viewing.



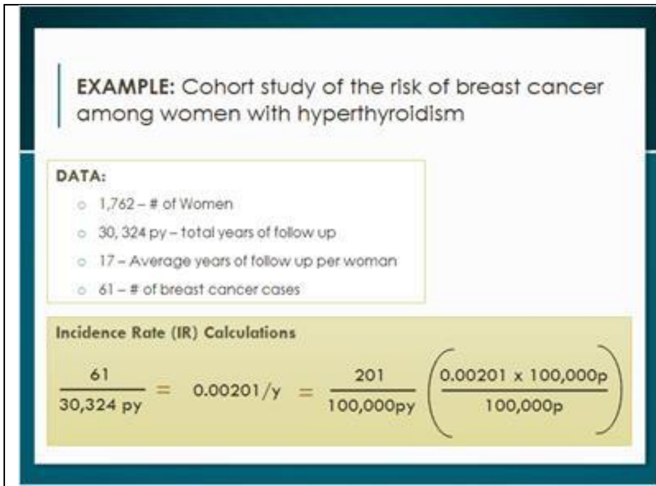
This slide shows a line drawing of people in a research study. First let's look at in terms of cumulative incidence. Assuming the person who died C died from other causes 1 person out of 5 was diagnosed with the disease over the ten years of the study which equals 1 person per 5 divided by 10 years or 1 person per 50 for each year of the study, probably better stated 2individuals per 100 or 2% acquired the disease annually. You could also say 20 per 1,000 per year. The first calculation assumes everyone in the study was followed for the same amount of time. This clearly is not true.

If you look at it in terms of incidence rate, your results are different. First identify the time period for all participants. You can see that person A entered in 2001 and was diagnosed in 2005, contributing 4 person years to the study, person B provided 6 years to the study (from the start until 2006), Person C provided 3 years to the study 2001 until 2005), person D was followed the entire 10 years, and person E was followed for 7 years for a total of 30 years of person time. You had 1 case of disease for every 30 years of person follow-up, which equals 0.03 cases per year or 3 cases per 100 years of follow-up or 30 per 1,000 years of follow-up. Clearly using person time provides a much more accurate measure. It is not as intuitively easy to understand, though.

So how do we get person years? It can come in a variety of was as shown in this slide.

- 100 people followed 1 year each = 100 py
- 10 people followed 10 years each= 100 py
- 50 people followed 1 year plus 25 people followed 2 years = 100 py

The time unit does not always need to be person years. For a rapidly progressing disease, such as Ebola, we may measure person days or even person hours.



This slide shows a data from a calculation of the incidence of breast cancer among women with hypothyroidism. They divided the number of women who had breast cancer by the total years of follow-up. As you can see women were followed on an average of 17 years but that is not included in the calculation. The rate was presented per 100,000 person years of follow-up which can be used when you are evaluating cancer, which is a relatively rare disease.

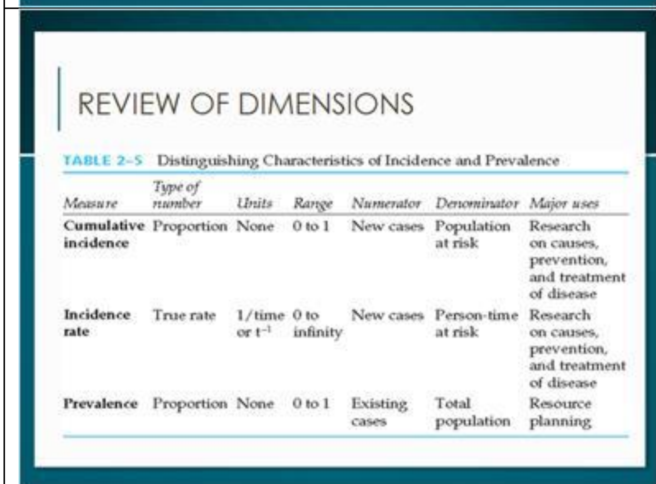
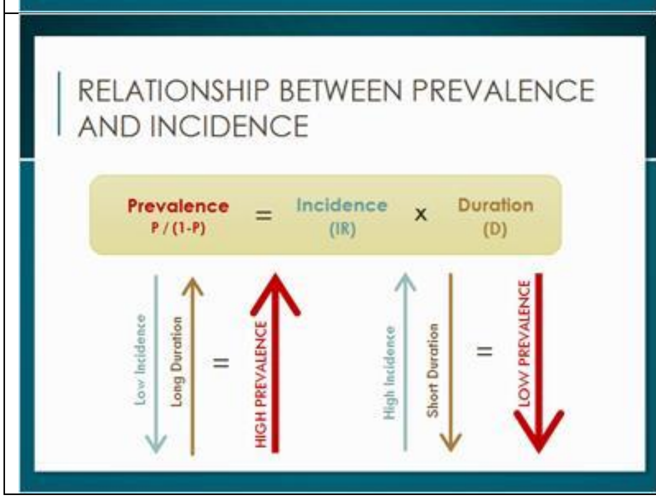


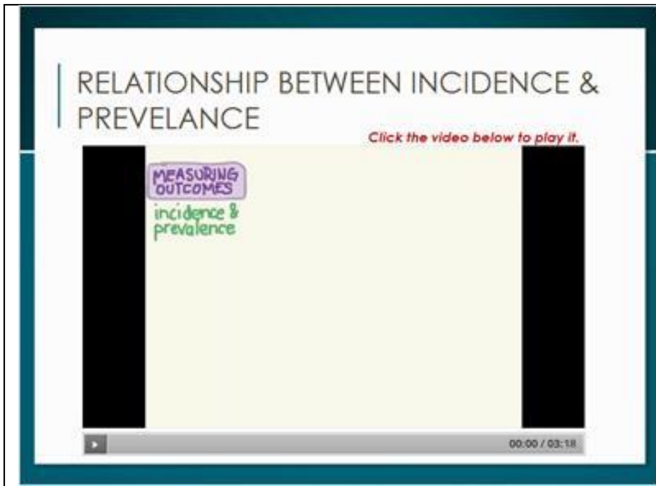
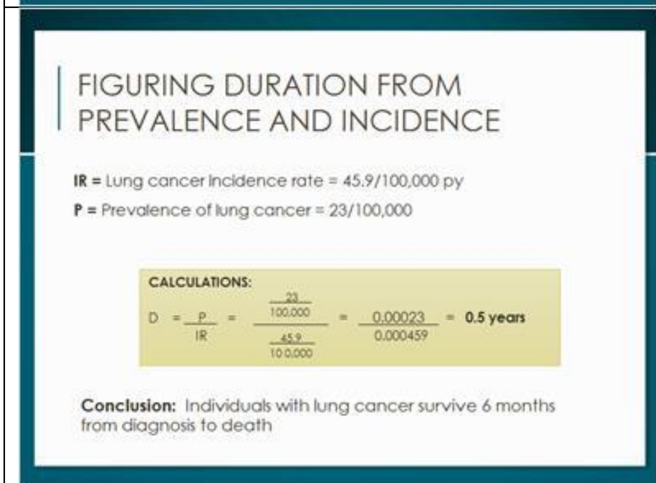
TABLE 2–5 Distinguishing Characteristics of Incidence and Prevalence

Measure	Type of number	Units	Range	Numerator	Denominator	Major uses
Cumulative incidence	Proportion	None	0 to 1	New cases	Population at risk	Research on causes, prevention, and treatment of disease
Incidence rate	True rate	1/time or t^{-1}	0 to infinity	New cases	Person-time at risk	Research on causes, prevention, and treatment of disease
Prevalence	Proportion	None	0 to 1	Existing cases	Total population	Resource planning

This slide shows a comparison between the three measures. As you can see cumulative incidence and prevalence are proportions using people in the denominator although cumulative incidence is more likely to follow people over time than prevalence which can be a one-time measure. In contrast incidence rate is a true rate and the denominator is time of follow-up.



There is a relationship between incidence and prevalence. Quite simple, prevalence equals incidence times duration. Duration refers to the length of a disease from onset to its termination by either cure or death. If incidence is low and duration is long the prevalence will be relatively high. However, if incidence is high but duration is short, the prevalence is relatively low. This relationship only really holds up if the duration of the disease remains constant.

	The analogy of a bathtub is also used and I have a good video which describes this relationship quite well. Again, we can turn to Dr. Patwari as he provides a nice video of the relationship between incidence and prevalence.
	You can perform a number of calculations using this relationship. This slide shows how to identify the average duration of lung cancer if you have the incidence and prevalence. You can divide the prevalence by the incidence rate to obtain the duration.
	In practicality, incidence and prevalence are used for different reasons. If we need to know how many hospital beds we need to treat people with a disease, it is generally much more helpful to know the overall number of people with the disease and not just new cases, assuming people need care throughout the entire course of their disease. As you will learn when we conduct research studies incident cases are more useful in determining the causes of disease. This is because time is an important factor to take into consideration. Once people become ill they change their behavior so it is better to measure the exposures before they became ill. Again, we will cover this later in the semester.