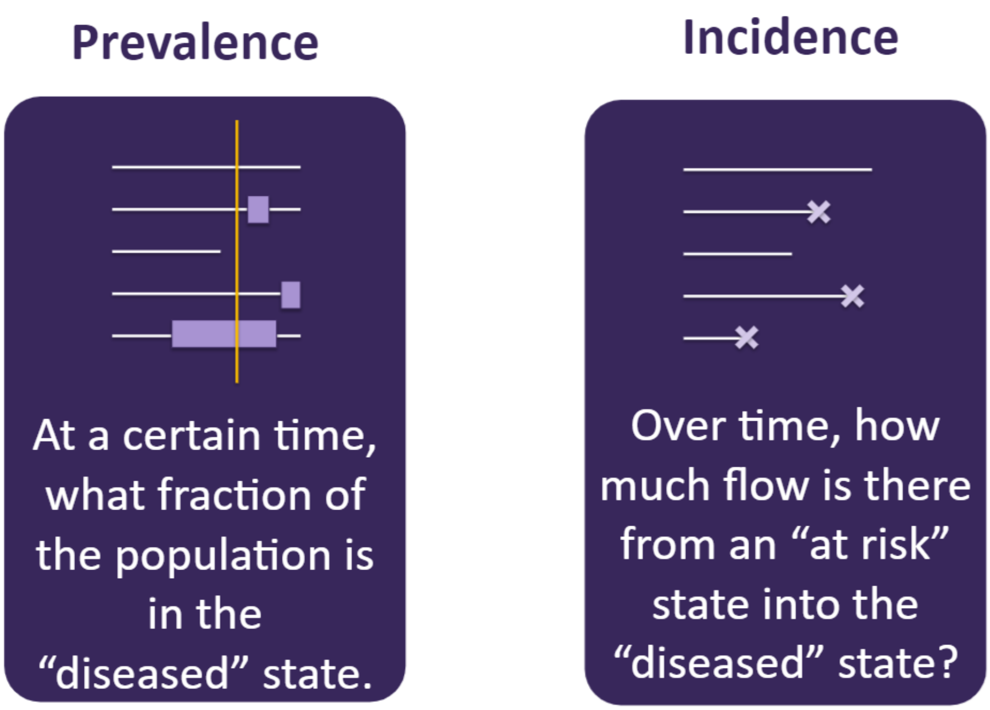
Incidence and Cumulative Incidence

In this lecture, you will learn about incidence and cumulative incidence.

The two most common measures of disease frequency within the population are prevalence and incidence. A separate video reviews prevalence.

Prevalence and Incidence

Prevalence calculates the fraction of the population that is in the “diseased” state, or more generally has a certain condition at a certain time. Incidence on the other hand quantifies the transition between being at risk into actually being in the “diseased” state over time. Consider a hypothetical population where everyone starts off at risk, and some people develop the “disease” over time. Prevalence counts the “diseased” folks at a certain time whereas incidence measures the “flow”, if you will, from “non diseased” to “diseased” state. To be an “incident case”, you must have started as “non diseased”.

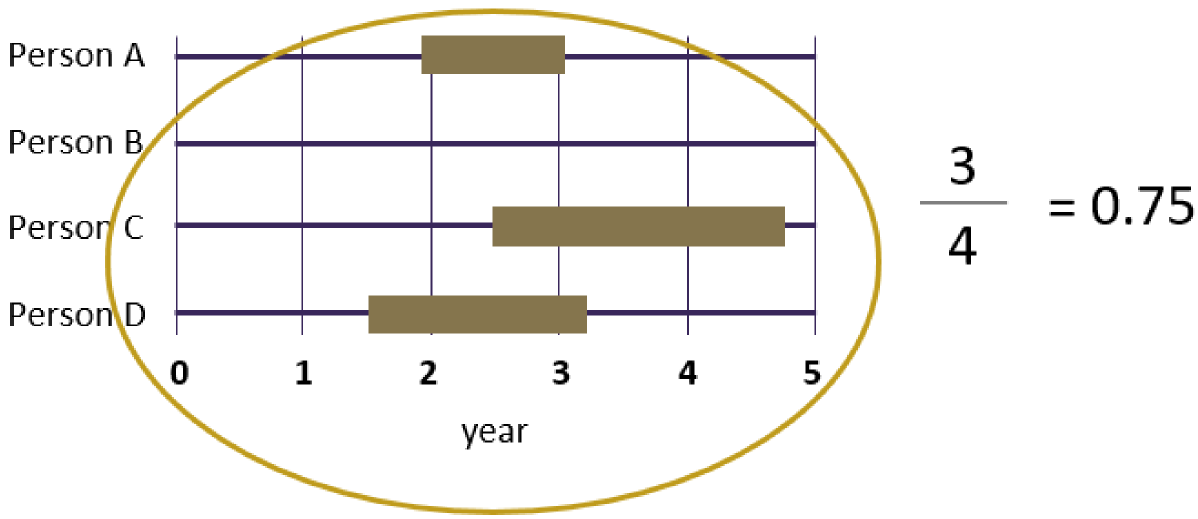


Incidence quantifies the frequency of transition from the “at risk” state to the “diseased” state. In incidence, we count incident cases which is the number of events where an “at risk” individual develops disease. In other words, people have to be at risk of developing the disease in order to be included in the calculation of incidence. For incidence, time is not frozen.
We can rephrase our definition to say incidence is “the frequency of disease events or new cases over a specific period of time.”

We discuss incidence rate and cumulative incidence in this video. Incidence rate is more flexible and applies to both open populations, where individuals can go in and out, and to closed populations, where there is no flow in and out of the population. In contrast, cumulative incidence is more restrictive and only applies to closed populations.

For a closed population, all initially at risk and all people are followed for a specified period of time, so cumulative incidence is defined as the number of people who develop the disease among the number of people who were initially at risk. It is also called the incidence proportion. Cumulative incidence doesn't consider when the disease occurred within the time period of interest. We need to have a closed population of individuals who are all at risk initially and follow all individuals for the same period of time.

Let's look at an example. In this graph, all the individuals were initially at risk and followed for 5 years. The purple horizontal lines represent the time followed and the gold bars represent the time when they had the disease. The cumulative incidence over 5 years in this population is 3 divided by 4 because persons A, C, and D developed the disease but person B did not develop the disease.



Features of Cumulative Incidence

Let's look at three features of cumulative incidence. Like prevalence, cumulative incidence is also a proportion between 0 to 1. It could also be presented as a percentage from 0 to 100%. Also, it is unitless, just like prevalence. And, most importantly, the timing of disease occurrence within the time period of interest does not affect results.

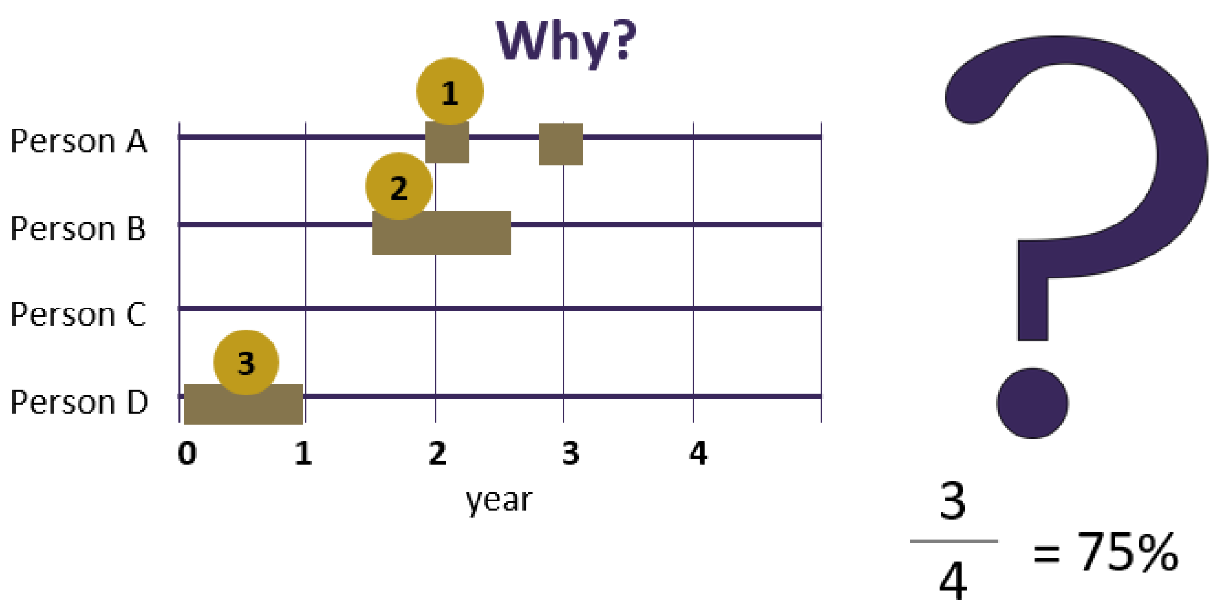
Estimating Cumulative Incidence

We can calculate cumulative incidence directly for closed populations. Let us first define a closed population. This is when individuals do not enter or leave the population during follow-up. So we have full information regarding all time periods when the individual is either at risk, not at risk, or diseased. In a closed population, each individual is either at risk, not at risk, or diseased.

In contrast, in open populations, individuals can enter or leave the population during follow-up. So we may not have full information about when the individual is either at risk, not at risk, or diseased. Some people could leave the risk category before an outcome occurs. For example, some people can be lost to follow up. During this time we wouldn't observe if an outcome had happened, or not. Or, if a participant enters the study as already diseased, we would not observe their prior "at risk" or "not at risk" periods.

Let's look at this example, where person A got the disease at year 2, got better, and then got it again at year 3. Suppose we're interested in calculating cumulative incidence over the 5-year period. In calculating cumulative incidence, each person may count as a case only once. Why? Because cumulative incidence is a proportion, and therefore has to be between 0 and 1.

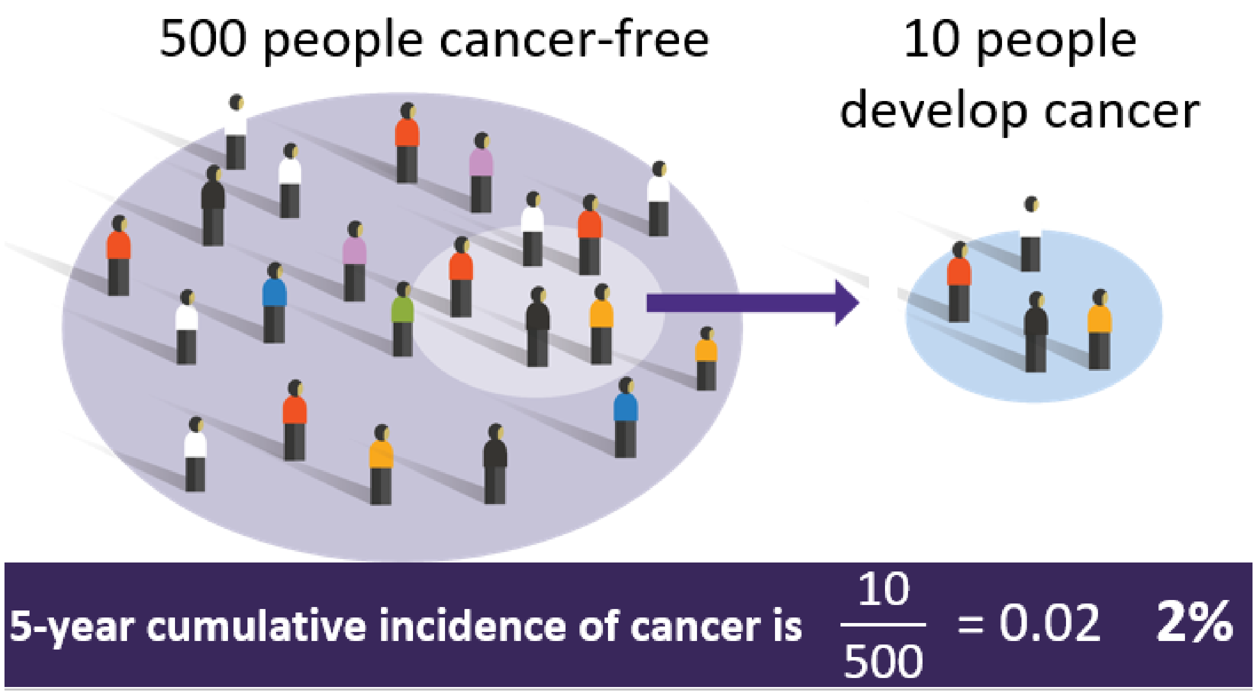
Let's calculate the cumulative incidence first. The numerator has information on whether the person got the disease or not, not how many times they got the disease. Here, person A, B and D all got the disease. Person C did not get disease. So the total number of people who got diseased is 3 and there are 4 people followed up. Therefore the 5-year cumulative incidence is 3/4 or 75%.



If we were allowed to include both events for person A, then the cumulative incidence would be 4/4, which is 1. But this does not make sense since Person C remained disease-free, while a cumulative incidence of 1 means everyone in the population got the disease.

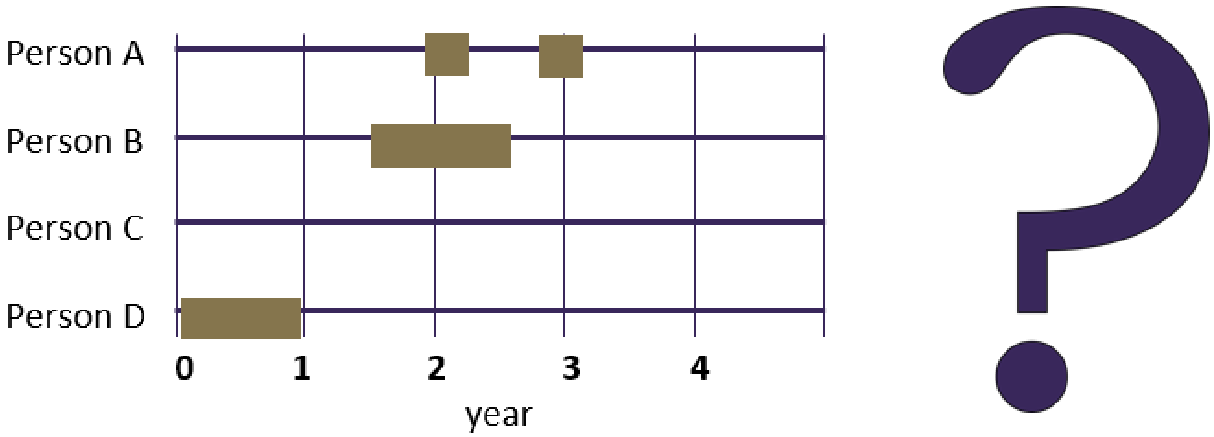
A takeaway here is that for cumulative incidence, we only count the first occurrence of the event. However, we will see that incidence rates allow us to count something multiple times, if it's scientifically relevant of course.

Now let's look at an example of cumulative incidence. Suppose we have a total of 500 people who we follow for a 5-year period. Suppose all 500 individuals were at risk of cancer at the beginning of the five-year period. Now suppose that 10 individuals developed cancer over this 5-year period. This means that the remaining 490 individuals did not develop cancer. To make our lives easy, we also make the simplifying assumptions that there were no deaths or loss to follow-up. The 5-year cumulative incidence of cancer in this study population is 10/500 = 0.02 or 2%.

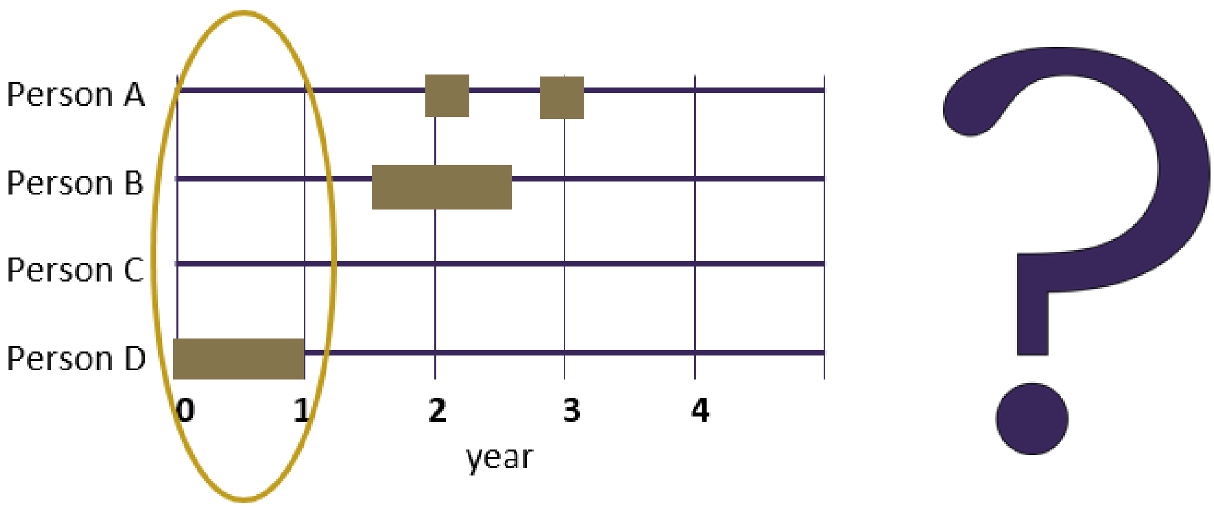


The next example that we will go through demonstrates why it's important to be explicit about the definition of time in cumulative incidence. Why is it crucial to report the element of time in the description of cumulative incidence? Because the answers could differ depending on the time interval.

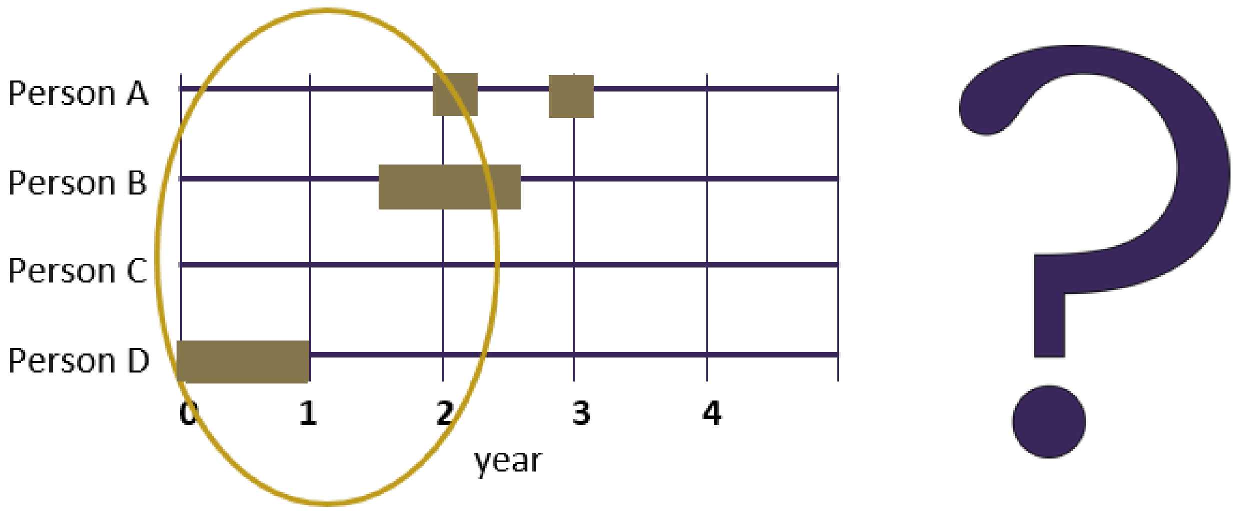
Let us walk through an example. In this example we assume a disease is recurrent and that this is a closed population, with all individuals followed for 5 years.
Here, the gold bars represent time spent in a diseased state and the purple lines represent time spent in a non-diseased state, or at risk.



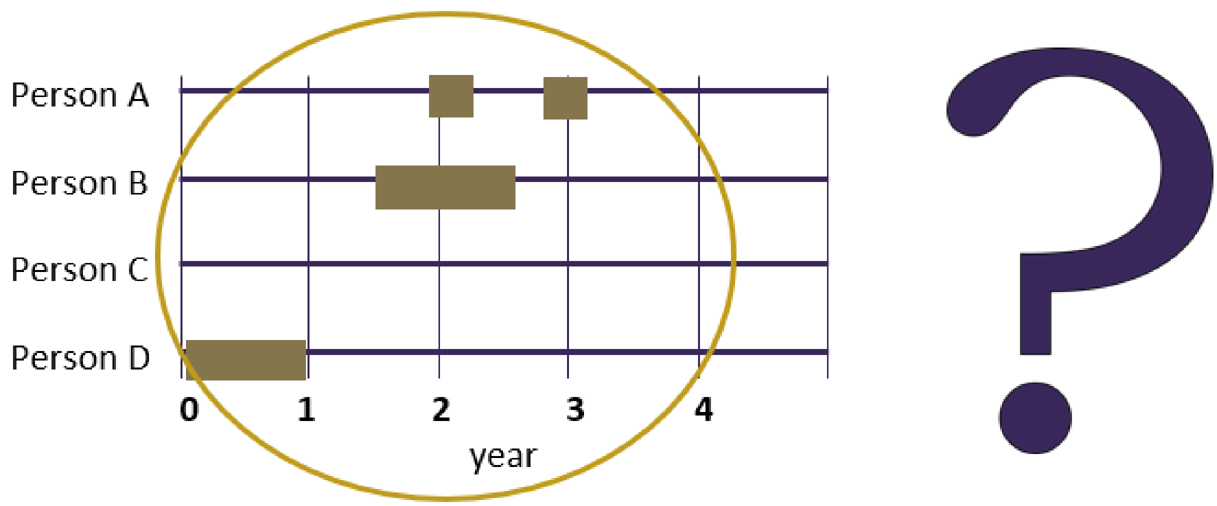
Here person D already had a disease at the beginning of the study, so was not at risk. No one else developed disease during the first year. So the one year cumulative incidence is 0, since no new disease occurred in that interval.



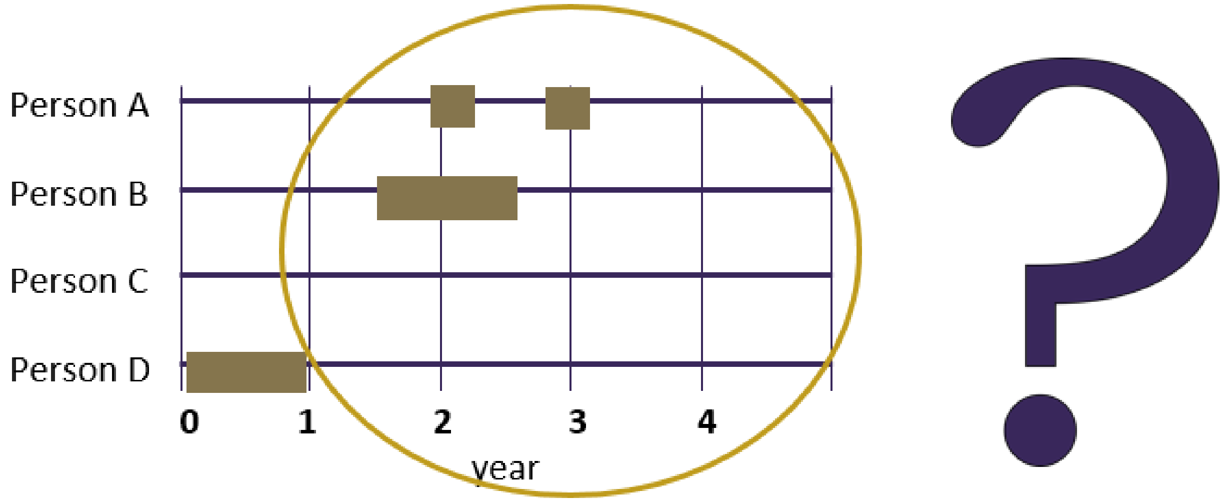
The two year cumulative incidence spanning the first two years of follow-up is 2/3, since persons A and B both developed disease, and persons A, B and C were all at risk. Person D is not included because they already had the disease.



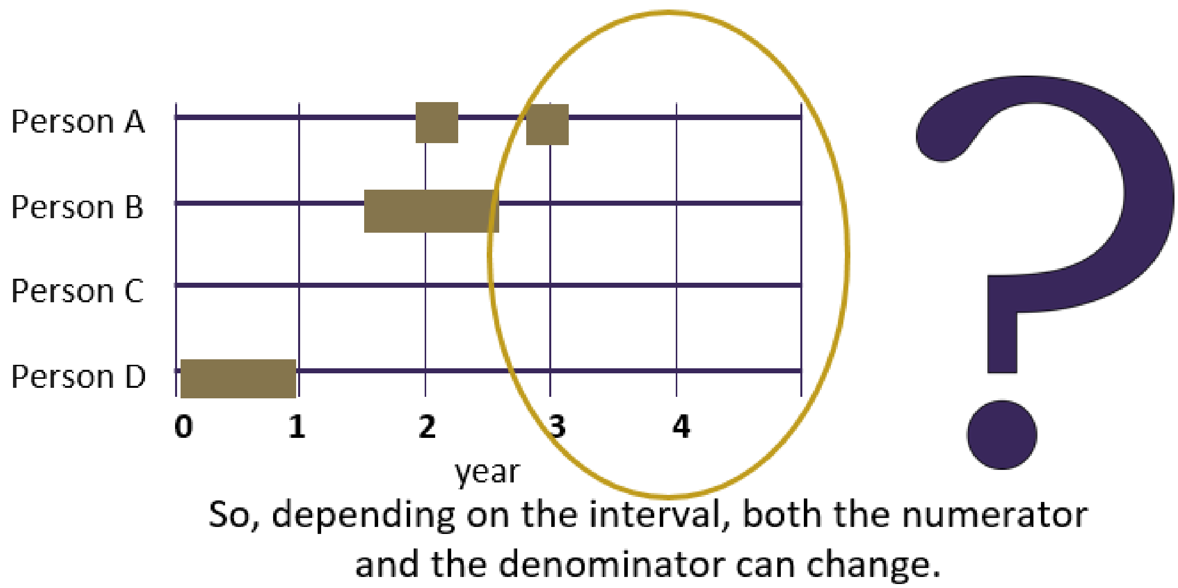
This stays the same for the four year cumulative incidence, which is still 2/3 since it's still only persons A and B who developed the disease, and the same 3 people were at risk. Remember the second, recurrent event in person A does not count for cumulative incidence.



Consider what happens if we started the follow-up time at year 1, instead of year zero. This means we assume we ignore what happened earlier. If we look at year 1 and later, person D did not have an event and is at risk. Persons A and B still each have an event and person C is at risk. So the cumulative incidence would be 2/4 or ½ or 50%!



If we only look at years 3 and later, person A already had disease, so was not at risk.
And persons B, C, and D were at risk. So the cumulative incidence would be 0/3 or 0.
In summary, depending on the time interval, both the numerator and the denominator can change!



Our next example pertains to COVID-19 infections on the Diamond Princess cruise ship at the beginning of the COVID-19 pandemic. This example comes from a journal article published in 2020 in the journal, European Surveillance.

Of 3,711 passengers and crew on board the cruise ship, between February 5 and February 13 2020 prior to any disembarkations, 218 tested positive for SARS-CoV-2. During this period, the cumulative incidence of SARS-CoV-2 infections was 218/3,711 = 0.06 or 6%.

Now, let's discuss the incidence rate. For a specified population of people at risk of disease who may be followed for different amounts of time since the onset of disease, we can calculate the incidence rate. This is also known as incidence density or person-time incidence rate. To calculate the incidence rate, we use the number of incident events as the numerator and divide this by the amount of person-time at-risk for the disease or condition as the denominator.

Notice that we no longer know the denominator and we must estimate it. If we don't know the exact amount of time at risk for each person in the population, we can estimate person-time at risk as the average, or mean number of individuals at risk multiplied by the duration of the study period. It can be person days, person years, or so forth.

At risk generally means susceptible. So, we would include only individuals who are susceptible in that at-risk population. Incidence rate is more commonly used as it is applicable in both open or closed populations and applies to a broader range of questions. What makes incidence rate different from cumulative incidence?

Features of Incidence Rate

Let's review the main features. It is a rate, meaning the number of events per unit of opportunity at risk (usually time). Unlike prevalence and cumulative incidence which are unitless, incidence rate is in the unit of inverse time. This allows it to range anywhere between 0 and infinity.

$$\text{Unit: time}^{-1} \text{ (i.e., } \frac{1}{time}\text{)}$$

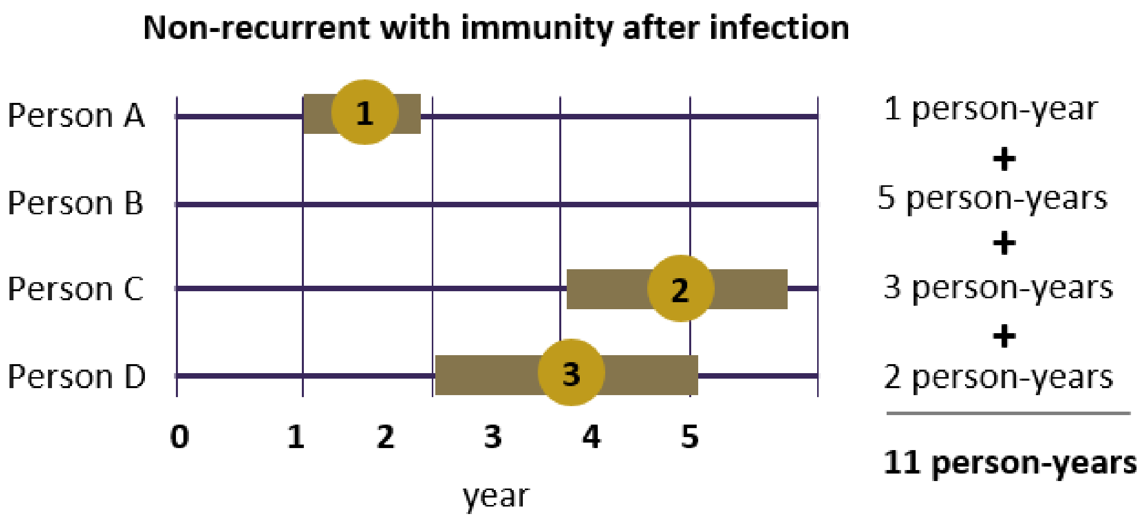
It is in general very flexible. It can be used in closed or open populations as long as person-time at risk can be measured (or at least estimated). It can be calculated when persons are at risk for different amounts of time. It can be used for non-recurrent and recurrent events. For example, measles can only occur once; a urinary tract infection (UTI) could occur many times in one's life. To calculate incidence rate, we need to add up all qualifying events and divide through by sum of the total time at risk. The denominator is pretty complex. How do we add up all the "time at risk"?

Person-time

Person-time is used to achieve that. It measures the amount of "at risk" experience. It is a special kind of time, it measures units of time that "persons" were "at risk". The total person-time is the sum of all persons' contribution to time at risk by the end of the observation period.

Let's look at an example to better understand person-time. In this closed population, we can see that the number of incident events is 3. Let's say this event is non-recurrent and people develop immunity after the infection. Remember, we calculate events and not the individuals. Person A developed the disease after one year of enrollment. Person A contributed one person-year at risk. After year 1, person A can no longer contribute person-time at risk. Person B has not developed the disease at the end of year 5 and so contributed 5 person-years at risk. Person C contributed 3 person-years at risk and Person D contributed 2 person-years at risk. We can add up the at-risk time for each individual before they develop the disease.

In this case, it is 11 person-years. The numerator is the number of incident events and the denominator is the total person time at risk. So, the incidence rate is 3/11 or 0.27 person years as shown here, which translates into 270 per 1000 person-years. And that's how the incidence would be reported. Incidence rate is a rate, the units include time, and the range is 0 to infinity.



$$\frac{3}{11} = 0.27 \text{ person-years} \times 1000 = 270 \text{ per } 1000 \text{ person-years}$$

Let’s look at another example, calculating incidence rate here. There are 110 new injuries from car accidents detected in 2020 in community X. In June, the population of the community was 2,200 people. So, what is the incidence rate of car accident injuries in 100 person-years? Pause the video while you calculate the answer. Once you are ready to continue with the lecture, press play.

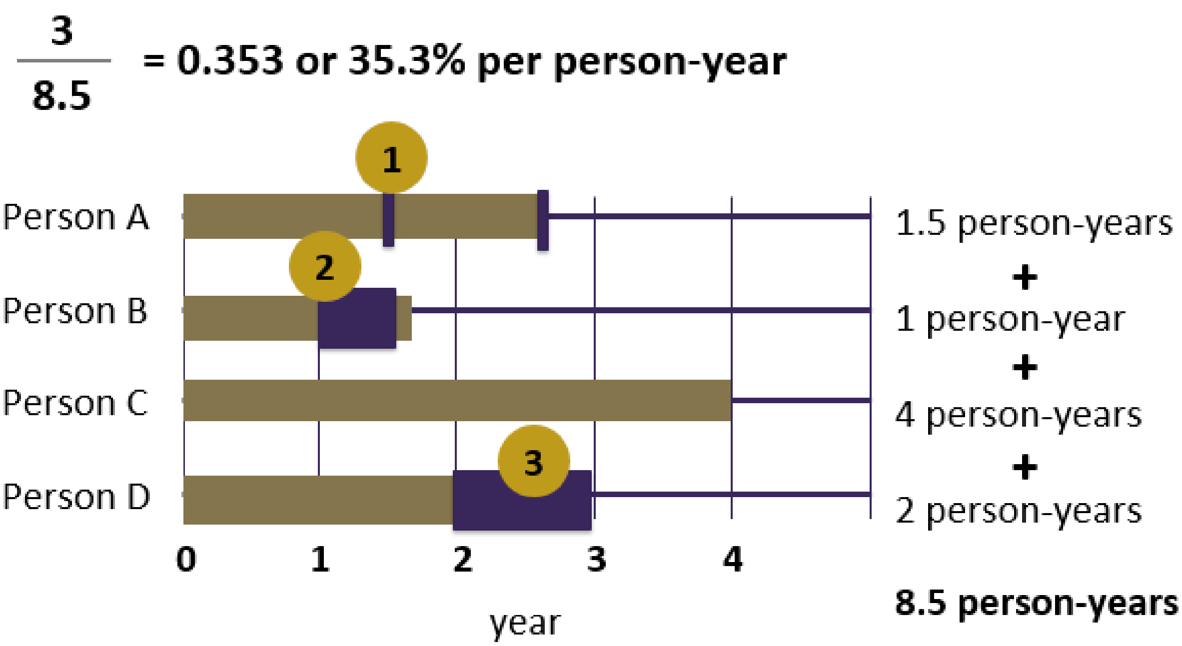
The incidence rate of car accident injuries is 110/2200 persons * 1 year = 1/20 person-years or 5 per 100 person-years. Person-time can be calculated in the following ways:

- Worker-days
- Child-years
- Woman-months
- Athlete-hours
- Driver-miles (Note that this isn’t a unit of time)

Let’s look at this example, and consider the question: If recurrent events do not count, what is the incidence rate of first occurrence of disease over the observation period? The gold bars represent time at risk and purple bars are time when some has the disease.

Persons A, B and D all had disease, so the numerator is 3. Person A had their first event in the middle of the second year, contributing 1.5 person years. Person B had their first event at the end of the first year, contributing to 1 person-year. Person C never had an event but was followed up for 4 years. And person D had their first event at the end of the second year, contributing to 2 person years.

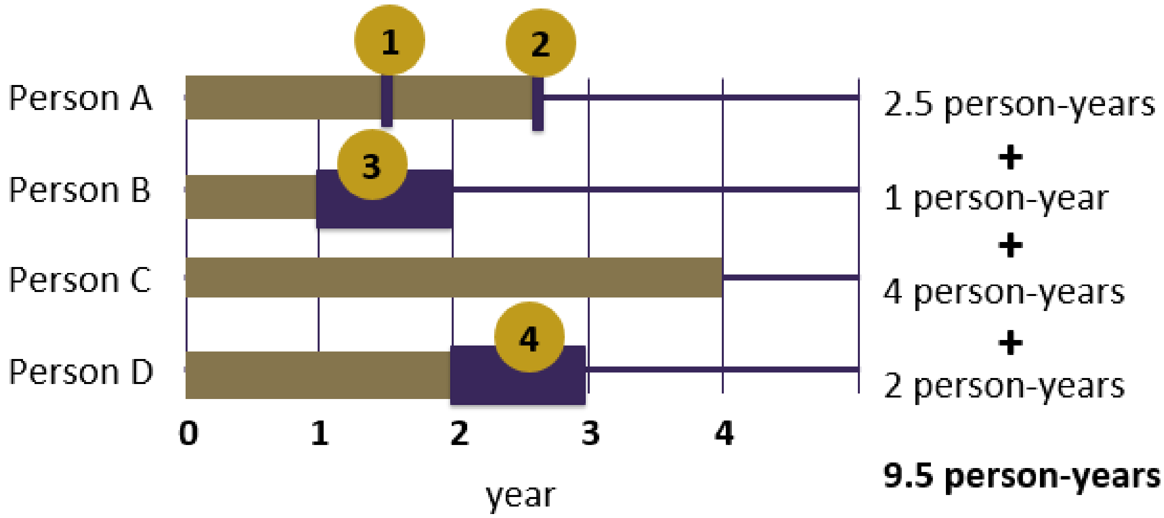
Notice that we didn’t count any time at risk after the first event, since it already happened. The total person years in the denominator is thus 1.5 plus 1 plus 4 plus 2, which gives 8.5 total person years. The incidence rate is therefore 3/8.5 = 0.353 first cases of disease per person-year, or 35.3%.



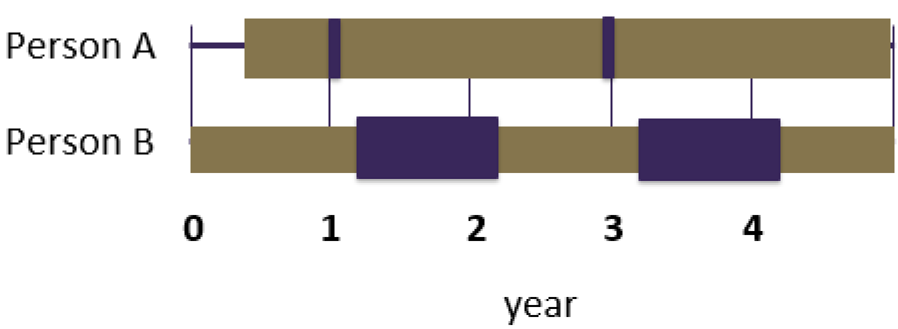
If recurrent events do count: What is the incidence rate of disease over the observation period?

The numerator is now 4 events, since person A had two. The time at risk for person A also increases by one year, giving a total person time of 9.5 years. So the incidence rate is 4/9.5 which is 0.421 or 42.1%.

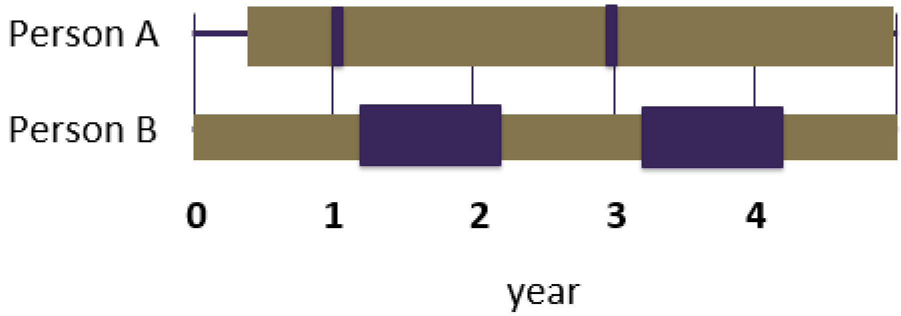
$$\frac{4}{9.5} = 0.421 \text{ per person-year or } 42.1\%$$



Let's now look at this example of two people. Does the <u>duration</u> of disease affect the incidence rate of first disease occurrence and why? No, because it occurs after the occurrence of disease and here we "stop the clock" the moment disease is diagnosed.



Does the <u>duration</u> of disease affect the incidence rate of disease over the observation period and why? Yes, it does, since it takes away from the time at risk. A person can't be at risk when they are in the diseased state.



This slide compares cumulative incidence and incidence rate in terms of key features. In terms of the Numerator, cumulative incidence is the number of persons who become new cases whereas the incidence rate is the number of events. The Denominator for cumulative incidence is the initial size of population at risk and person-time for incidence rate. The mathematical nature for cumulative incidence is proportion, and for incidence rate is rate. The mathematical nature for cumulative incidence is a proportion, and for incidence rate is a rate. Cumulative incidence rate is unitless, and incidence rate is one over time. Range for cumulative incidence is 0 to 1, and for incidence rate is 0 to infinity. Cumulative incidence is calculated directly in a closed population, and incidence rate is calculated directly in a closed or open population.

This last example compares cumulative incidence and incidence rates. Consider two populations A and B shown in this slide, with two people in each population. Assume both populations are closed for the disease of interest. What is the cumulative incidence of disease in each population?

Among persons 1 and 2 in population A, person 1 got diseases in Year 2. Among persons 3 and 4 in population B, person 3 got diseases in Year 5. The cumulative incidence for both populations is then 1 person who ever had disease divided by two people in the population, which is 0.5.

Now, let's consider: What is the incidence rate of disease in each population? Assume we are interested in recurrent disease. Person 1 was disease free for 4 years, and person 2 disease free for 5 years. Total person time in population A is 9 person-years. Incidence rate is therefore 1 divided by 9 person-years which is 1/9 or 0.11.

Same holds for population B. What if we only consider the incidence rate of first occurrence? Then person 1 was followed for 1-person year before they developed disease, so total person-time in population A is 6 person-years. However, person 3 in population B developed disease in the fifth year after they were followed up for 4 years, so total person time in population B is 9 person years. Therefore, the incidence rate of first disease occurrence is 0.17 for population A vs 0.11 (same as previous slide) for population B. We see in this example that the timing of when someone develops disease can make a difference in calculation of the incidence of first disease occurrence!

Key Points

In summary:

- Incidence measures new disease over time.
- Like prevalence, cumulative incidence is a proportion and is also unitless.
- Incidence rate is a rate. Its unit is inverse of time. These are two very important concepts that we will revisit again.