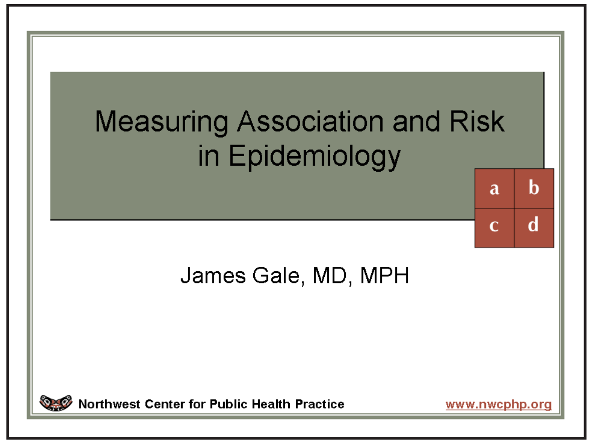
Measuring Risk in Epidemiology
Transcript

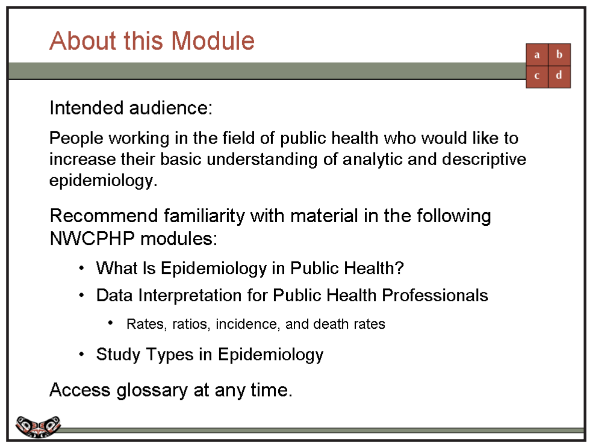
a b
c d



Measuring Risk in Epidemiology

Welcome to "Measuring Association and Risk in Epidemiology."

My name is Jim Gale. I am professor emeritus in the Department of Epidemiology at the University of Washington School of Public Health and Community Medicine. I spent 12 years as a local health officer in Kittitas County, Washington, and have also been a US Public Health Service Epidemic Intelligence Service officer.



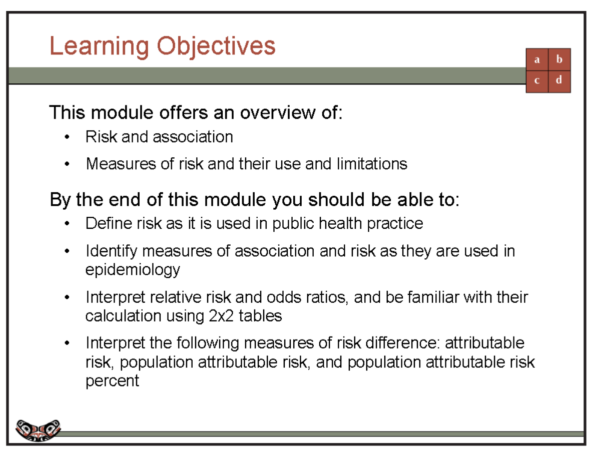
About this Module

This module and others in the epidemiology series from the Northwest Center for Public Health Practice are intended for people working in the field of public health who are not epidemiologists but would like to increase their familiarity with and understanding of the basic terms and concepts used in epidemiology. Before you go on with this module we recommend that you become familiar, if you haven't already, with the material presented in the following modules, which you can find on the Center's Web site:

- What Is Epidemiology in Public Health?
- Data Interpretation for Public Health Professionals
- Study Types in Epidemiology

It is particularly important to understand the concepts of rates, ratios, and measures of disease frequency, such as incidence rates and death rates. These concepts are covered in the modules "What is Epidemiology in Public Health Practice?" and "Data Interpretation for Public Health Professionals."

If you want to review definitions of epidemiological terms, you may access the glossary at any time from the attachments link at the top of the screen. You can also find the transcript in the attachments. You may want to print a copy now and make notes on it as you go through the module.



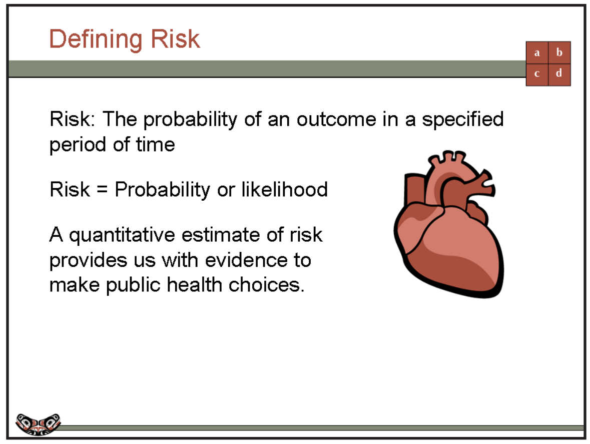
Learning Objectives

In this module we will explore the meaning of risk and association and their relationship to incidence rates in epidemiology. We will also look at measures of risk and their use and limitations in considering causal relationships. By the end of this 45-minute module you should be able:

- To define risk as it is used in public health practice

- To identify measures of association and risk as they are used in epidemiology
- To interpret relative risk and odds ratios and be familiar with their calculation using 2x2 tables, and
- To interpret the following measures of risk differences: attributable risk, population attributable risk, and population attributable risk percent.

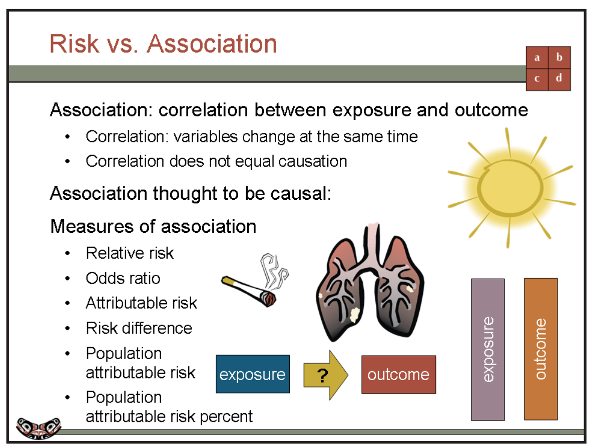
Defining Risk



Epidemiologists use the term risk to mean the probability of an outcome (often a negative outcome) in a specified period of time. In epidemiology, risk usually implies a quantifiable concept, such as the risk of dying or the risk of a heart attack, rather than a more general concept such as the risk of offending someone by speaking frankly. In this module, I will use risk, probability, and likelihood interchangeably, since they're measured the same way.

Why is it important to know about risk? In the practice of public health, we are faced with many choices. A quantitative estimate of risk is useful in making decisions about a course of action or intervention as well as how to allocate finite resources of time and money.

Risk vs. Association



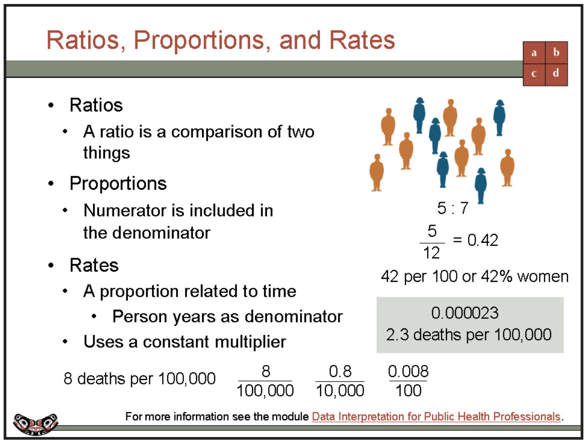
In epidemiology, an association means a correlation, often between an exposure and an outcome.

I'd like to pause for a moment to talk about correlation and causation. Correlation is the situation in which two or more variables, in this case exposure and outcome, change at the same time. For example, as exposure to the sun increases, the incidence of some types of cancer increases. It's important to remember that just because both variables change, it does not mean that one causes the other to change. In other words, correlation does not equal causation. Some other unknown factor may be causing both the exposure and the outcome to change.

Deciding the degree to which an association might be causal, occupies a great deal of an epidemiologist's thinking and effort. Where an association is thought to be causal, as in the association between smoking and lung cancer, we often use the term risk factor in referring to the exposure we are comparing.

Measures of association refer to specific mathematical expressions that measure the degree to which an exposure, such as exercise or smoking, is associated with an outcome of interest, such as health status or disease. Some measures of association we'll explore later in this module are relative risk, odds ratio, attributable risk, risk difference, population attributable risk, and population attributable risk percent. Now we'll pause for the first of several interactive exercises that will allow you to answer questions about the material we have just covered. Please note that the exercises can take several seconds to load.

Ratios, Proportions, and Rates



Before we go any further, let's review three terms we'll be using in this module: ratios, proportions, and rates.

A ratio compares two dissimilar things by dividing one quantity by another. For example, say a group of people has 5 women and 7 men. The ratio of women to men is 5 to 7.

We can also express the relationship of women to the entire group as a proportion, or a percentage, of the total number of people.

A proportion or a percentage is a special kind of ratio, because the group affected, or the numerator, must come out of the population at risk, which is the denominator. For example, our group contains 12 people. The proportion of women in the group is 5/12. We can also express this fraction as 42 per 100, or 42%. The value of a proportion never exceeds 1, or 100 percent.

A rate is also written as a numerator and a denominator, but a rate has an element of time. You will often see person-years used as a denominator in calculating rates.

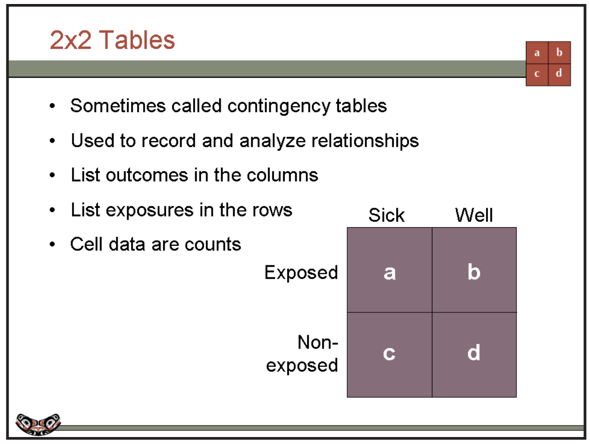
A multiple of 10 is often used to create a numerator with at least one number to the left of the decimal point. For example, the risk of dying last year from stomach cancer may be 0.000023. However, for most people, it's easier to think about this rate when it is expressed as 2.3 deaths per 100,000 per year.

Rates can be written in several ways. For example, the rate of 8 deaths per 100,000 from suicide in 20 to 24 year olds in Oregon in 1995 could also be expressed in these ways.

If you want to review the arithmetic of counts, numerators, denominators, ratios, proportions, and rates, they are well covered in slides 11 through 14 in the module "Data Interpretation for Public Health Professionals."

Before we go on to look at measures of risk, or measures of association as we usually call them, we should also look at a tool often used in calculating these measures, the 2 by 2 table.

2 x 2 Tables



Epidemiologists use 2 by 2 tables, which are also sometimes called contingency tables, to record and analyze the relationship between variables. Since we're often comparing two exposures and two states of illness, the 2x2 table has four cells, or options. The usual format of the table is to list the outcomes or disease categories in the vertical columns, and the exposures, or attributes, in the table rows.

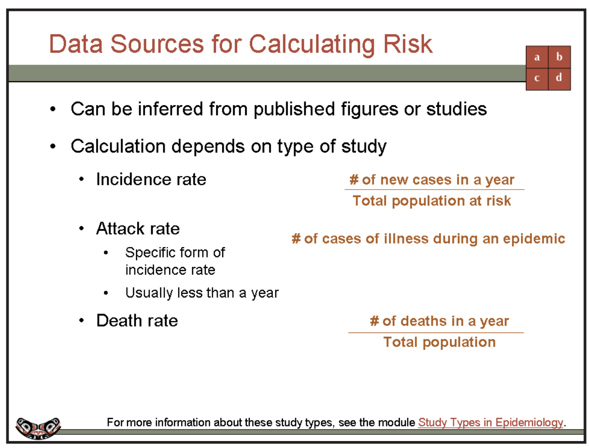
In this example, we'll label the exposure rows exposed and non-exposed, and we'll label the outcome columns, sick and well.

Just for this example, we'll represent the counts in the four cells as a, b, c, and d.

A key point to remember about 2x2 tables is that the data in the cells are actual counts, not rates.

As you go through the module, you'll see how to use a 2x2 table to calculate various measures of association and risk.

Data Sources for Calculating Risk



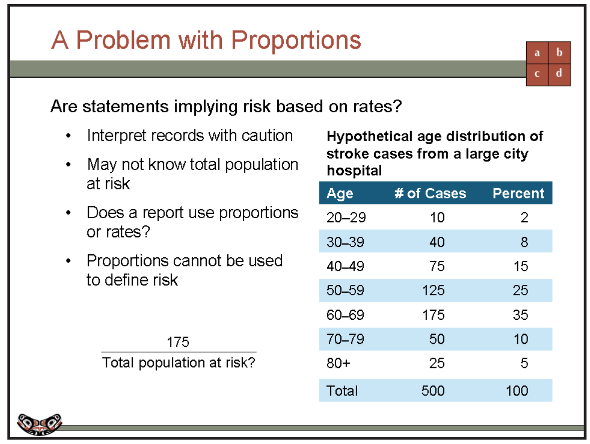
As I mentioned before, risk is the same as the probability, chance, or likelihood that something will happen. Risks can be inferred from published death rates, from incidence rates from an existing disease registry, such as a cancer registry, or from attack rates from an outbreak investigation. But more frequently special studies, such as cohort studies, must be carried out to determine these rates in populations that are or are not subject to specific exposures. The way in which we calculate risk depends on the type of study we conduct. (For more information about these study types, see the module Study Types in Epidemiology.)

An incidence rate is the number of new cases of illness occurring in a population over a specific period of time, usually a year, divided by the total population at risk.

An attack rate is a specific form of an incidence rate, indicating the incidence of disease in a population during an epidemic or outbreak, usually less than a year.

A death rate is the number of deaths (in general, or due to a specific cause) in a population during a specific time (usually a year) divided by the total population.

A Problem with Proportions



I'd like to emphasize an important point.

As you interpret reports or news stories, always ask yourself if statements implying risk are based on rates. In particular, data from such sources as in-patient records of hospitals must be interpreted with caution, since you usually do not know the population from which each group comes. The report is probably based on only a proportion of cases within one facility, not on all persons at risk in a defined population. It would be very difficult to define the population from which these different age groups are drawn, except under special circumstances.

To make this point clearer, let's look at this list of hypothetical cases of stroke. It might be tempting to say that the risk of dying is highest in the 60 to 69 year age group. But we don't know the number of persons at risk in each age category; that is we can't define the denominators to calculate risks for any of the age categories. Although we may know in general, that strokes increase with age, we couldn't use the data in this list to reinforce that assumption, because there could be alternative reasons for the number of strokes. All the younger persons may have left the community for work elsewhere, for example, which would mean that the denominators would be smaller, but the rates might be higher.

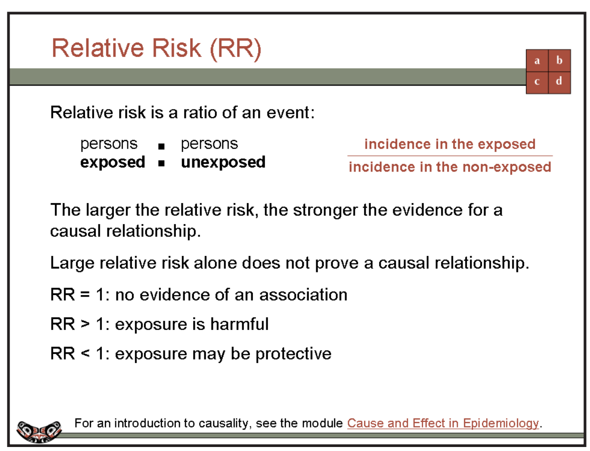
The point I'm making is always to notice whether a report uses proportions or rates when talking about risk. Keep in mind that proportions cannot be used to make statements regarding risk.

Let's pause now so you can answer some more questions about what you have just learned.

Practice: Ratios, Proportions, and Rates

Exercise 1

Relative Risk (RR)



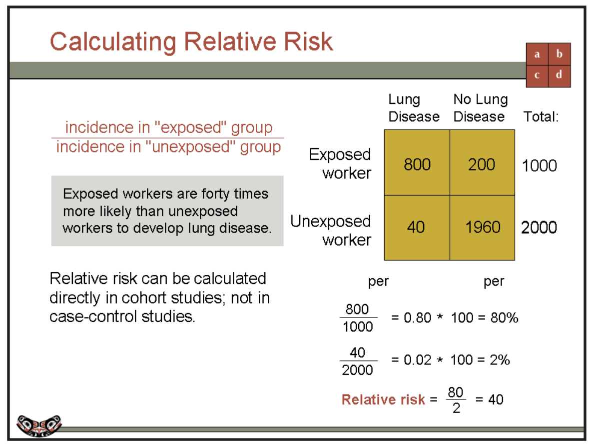
Relative risk is the first measure of association we will consider. It is a ratio of the risk of an event (or of developing a disease) in two groups: persons who are and who are not exposed to some factor. It can be expressed, or written, as the incidence in the exposed divided by the incidence in the non-exposed.

The size of the relative risk is one of the criteria we use in estimating the strength of a causal relationship between an exposure and an outcome. The larger the relative risk, the stronger the evidence is for a causal relationship, all other things being equal. For example, a relative risk of 10 provides better evidence for a causal relationship than would a relative risk of 3. But remember, a large relative risk alone does not establish a causal relationship.

If the relative risk is equal to 1, we say there is no evidence of an association. If the relative risk is greater than 1, than the exposure is harmful, and if the relative risk is significantly less that one, we can say there is evidence that the exposure may be protective.

Now let's look in detail at how you calculate relative risk.

Calculating Relative Risk



As I said, we calculate relative risk by dividing the incidence in the exposed group by the incidence in the non-exposed groups. In this fictitious example of a cohort study, we found that in a factory of 3000 workers, 1000 workers were exposed to toxic substances, and 2000 were not.

We'll use this 2x2 table to calculate the relative risk of developing lung disease among exposed workers compared to unexposed workers.

During the study period, 800 of the exposed workers developed some type of lung disease, and only 40 of the non-exposed workers developed a lung disease .

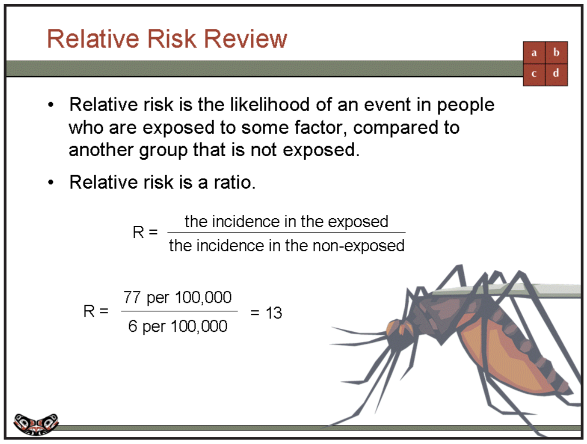
The incidence of lung disease during the study period among exposed workers was 800 per 1000 person per year. The incidence of lung disease among non-exposed workers during this time was 40 per 2000. So that we have comparable rates, we'll convert this to 20 per 1000. By subtracting both 800 and 40 from their respective row totals, we can get the number of persons with no lung disease in each exposure row.

Note that the incidence in exposed workers was 800/1000, or 80%. The incidence in those not exposed was 40/2000, or 2%. Comparing the incidence in the exposed to the incidence in the non-exposed, we get 80/2, for a relative risk of 40. We could also say that workers exposed to toxic substances would be forty times more likely than non-exposed workers to develop lung disease.

This elevated relative risk strongly suggests, but does not prove, a causal relationship between exposure to this toxic substance and developing lung disease.

Notice that this is an example of a cohort study. In case-control studies, relative risk cannot be calculated directly. It must be estimated using an odds ratio. We'll discuss odds ratios later on in this module.

Relative Risk Review



Let's review a few key points about relative risk. Relative risk is the likelihood of an event in people who are exposed to some factor compared to the likelihood in another group that is not exposed. Relative risk is a ratio of two incidence rates. To calculate it, you divide the incidence in the exposed by the incidence in the non-exposed.

Early in my work as an epidemiologist, I was part of a team investigating a large outbreak of St. Louis encephalitis, known as SLE, in Houston, Texas. This serious illness is transmitted through the bite of a mosquito.

We found that the incidence of SLE in people living in central Houston who were more likely to be exposed to standing water, where mosquitoes could breed, was 13 times the incidence in people living in neighborhoods on the outskirts of the city who had little or no exposure to standing water. After calculating the relative risk of SLE relative to exposure to standing water, we concluded that because the relative risk was higher in the central part of the city, those neighborhoods might benefit from spraying with chemicals to kill mosquito larvae.

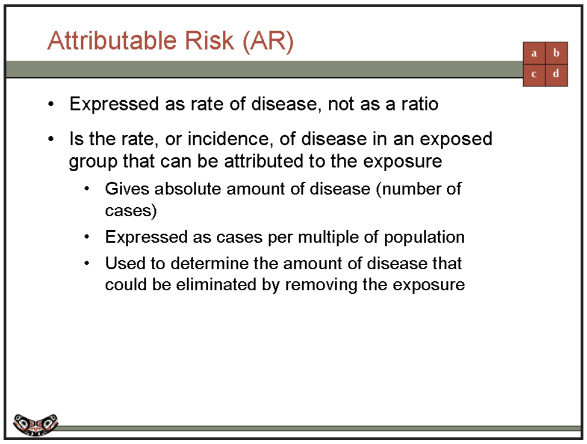
Let's pause now so you can answer some more questions on what you have just learned.

Practice: Calculating Relative Risk (part 1)

Practice: Calculating Relative Risk (part 2)

Exercise 2

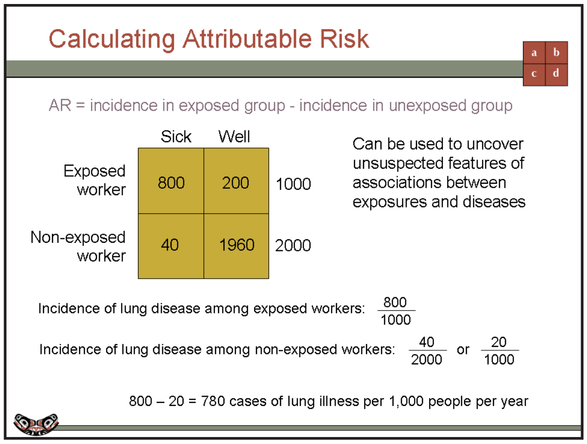
Attributable Risk (AR)



Remember that relative risk is the likelihood of an event (or of developing a disease) in persons who are exposed to some factor, as compared to another group which is not.

Attributable risk is more concrete. It is expressed as an absolute rate of disease, rather than as a ratio. It's the rate, or incidence, of disease in an exposed group that is attributable to the exposure. Attributable risk in an exposed group gives us the absolute amount of disease (or the number of cases) resulting from the exposure. We express it as cases per multiple of the population. We can use attributable risk to determine the amount of disease that could theoretically be eliminated or prevented by removing the exposure in the exposed population.

Calculating Attributable Risk



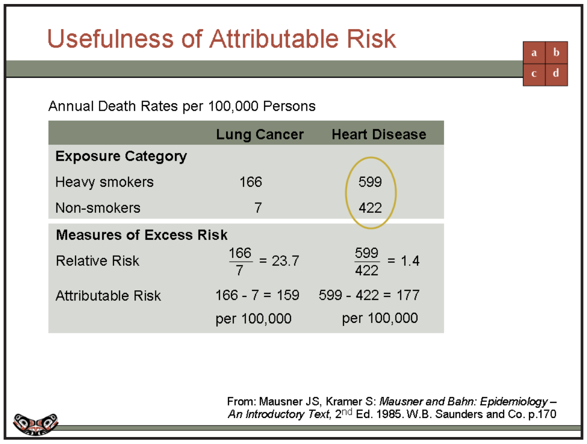
We calculate attributable risk by subtracting the incidence in the unexposed from the incidence in the exposed.

Let's look at the example of factory workers exposed to a toxic substance and who develop lung disease to see how this works.

Remember, in our fictitious cohort study, 800 of the exposed workers developed some type of lung disease during the study period, and 40 of the non-exposed workers developed a lung disease. So the incidence of lung disease among exposed workers was 800/1000 and the incidence in non-exposed workers was 40 per 2000. Since we have to use incidence rates with the same denominators, we'll convert this to 20 over 1000. Plugging these numbers into the attributable risk equation, we get an attributable risk of 780 cases of lung disease per 1000 factory workers exposed to the toxic substance attributable to their exposure.

Attributable risk can be used to uncover unsuspected features of associations between exposures and diseases. Let's look at an example to clarify this important idea.

Usefulness of Attributable Risk



In 1956, a study of lung cancer and smoking in British physicians reported these results. The study showed that although heavy smokers face a relative risk of dying from lung cancer 24 times higher than non-smokers, they face a relative risk of only 1.4 times of dying of heart disease, as compared to non-smokers.

Remember, to calculate the relative risk of death from lung cancer, you divide the death rate from lung cancer in the heavy smokers by the death rate from lung cancer in the non-smokers; and to calculate the relative risk of dying from heart disease, you divide the death rate from heart disease in the heavy smokers by that in the non-smokers.

By the way, we use death rates here in the same way as we'd use incidence rates because they are a form of incidence rate.

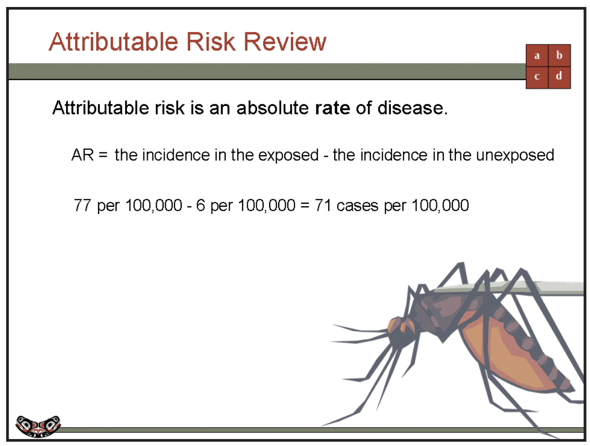
The point I want to make here is that if we examine only relative risk, we'll miss a crucial point.

By calculating the attributable risk for lung cancer deaths in heavy smokers as compared to non-smokers (that is, subtracting the lung cancer death rate in the non-smokers from that in the heavy smokers), and the attributable risk for deaths from heart disease, we can see that over all, heavy smokers are more likely to die of heart disease than of lung cancer, because the death rate from heart disease is much higher than the death rate from lung cancer in both groups.

Heart disease presumably has many other causes besides heavy smoking. However, if one were to eliminate smoking, the attributable risk tells us that more people would be saved from death by heart disease than from death by lung cancer.

This information would be important in arguing for funding a big anti-smoking campaign, since it would provide evidence that many additional lives could be saved by removing smoking as a risk factor for heart disease death as well as for death from lung cancer.

Attributable Risk Review



Let's review attributable risk before we go on. Attributable risk is an absolute rate of disease, rather than a ratio. It's the incidence of disease in an exposed group that is attributable to, or thought to be caused by, the exposure. The calculation is simple. We subtract the incidence rate in the unexposed from the incidence rate in the exposed. In the St. Louis encephalitis outbreak in Houston that I mentioned earlier, we would have calculated the incidence rate of encephalitis that theoretically could have been reduced by spraying mosquito larvicide in the high-incidence neighborhoods and discovered that 71 cases per 100,000 people living in the high risk areas would theoretically have been prevented by the spraying if standing water were the only cause of the difference in incidence.

Let's pause now so you can answer some questions on what you have just learned.

Exercise 3

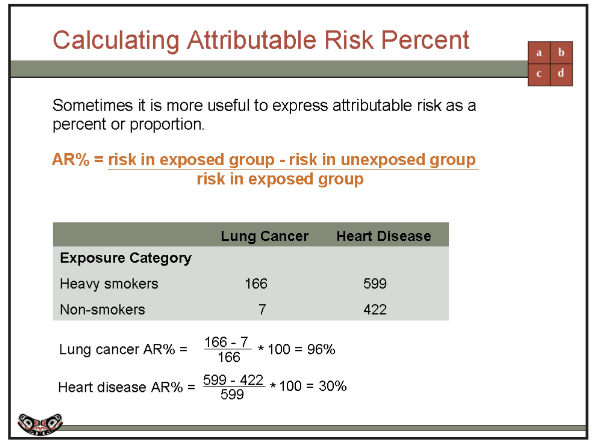
Calculating Attributable Risk Percent

Sometimes it is more useful to express attributable risk in terms of a percent, or proportion. Attributable risk can be calculated as a percentage by subtracting the risk in the unexposed from the risk in the exposed and dividing the result by the risk in the exposed.

Here's an example of that calculation using the attributable risks from the British study of deaths due to lung cancer and smoking that we just looked at. For lung cancer deaths, the attributable risk percent is 96%. This figure is known as the attributable risk percent in the exposed.

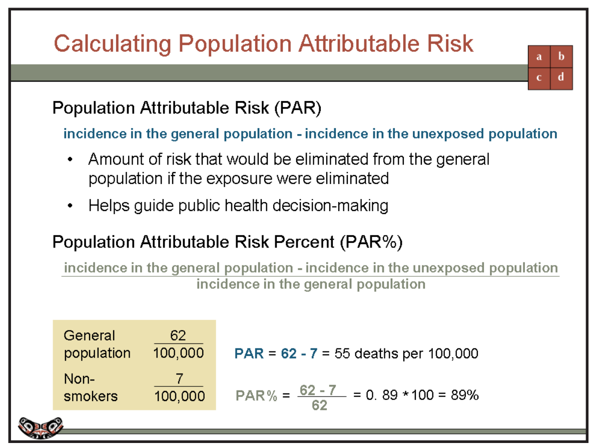
For heart disease deaths, the attributable risk percent is 30%.

When we compare the attributable risk percents for deaths due to lung cancer and to heart disease, it appears that eliminating smoking would reduce a larger percentage of deaths due to lung cancer than of deaths due to heart disease in smokers, but as I said before, since the frequency of heart disease deaths in the general population is so much greater than the frequency of lung cancer deaths, more heart disease deaths would be prevented, even though the relative risk of dying of heart disease in smokers, as compared to non-smokers, is smaller.



Calculating Population Attributable Risk

If we know the death rate due to lung cancer in the general population and in non-smokers, we can also calculate the risk in the general population of lung cancer deaths attributable to smoking. To do this we take the lung cancer death rate in the general population and subtract the lung cancer death rate in the non-smokers. The result is the population attributable risk (or PAR). This rate tells us the number of lung cancer deaths that would be eliminated from the general population if the exposure, in this case, smoking, were eliminated.



Suppose we know that the lung cancer death rate in the general population is 62 per 100,000 persons and in non-smokers is 7 per 100,000 persons. The rate of lung cancer deaths in the general population attributable to smoking is 62 minus 7 or 55 per 100,000 persons per year.

Population attributable risk is helpful for guiding public health decisions about where to focus resources most effectively to protect the public's health.

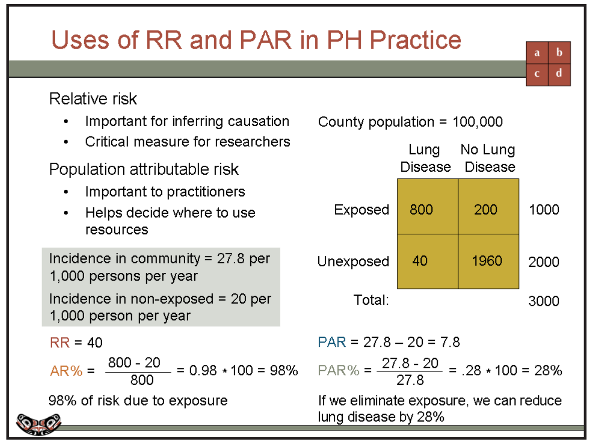
As with attributable risk, population attributable risk can be expressed as a percent.

We calculate it in a way similar to calculating attributable risk percent: the incidence in the general population minus the incidence in the unexposed population divided by the incidence in the general population.

Turning our smoking example into a population attributable risk percent, we can say that of the 62 per 100,000 deaths per year due to lung cancer in the general population, 62 minus 7 divided by 62, or 89%, can be attributed to smoking.

In this example, I have assumed that all persons are either smokers or non-smokers, and I have not taken secondary smoke exposure into consideration.

Uses of RR and PAR in PH Practice



As I mentioned before, relative risk is important for inferring causation. For example, public health researchers looking at the cause or source of a disease may use it to evaluate the role of an exposure in causing the disease.

Population attributable risk, on the other hand, is more important to public health practitioners and policy makers since they want to know the numbers of deaths or illnesses prevented in order to use their resources to greatest advantage to control or prevent disease in populations in their jurisdictions. The size of the problem in actual numbers counts for much when allocating limited public health resources.

Let's look at our factory workers again to clarify this point.

Suppose the factory is located in a county of 100,000 people. We know that the relative risk of developing lung disease for factory workers exposed to this toxic substance compared to other factory workers is 40, as we showed earlier. This high relative risk is very suggestive that the exposure is causal. But how big is this problem in the rest of the county?

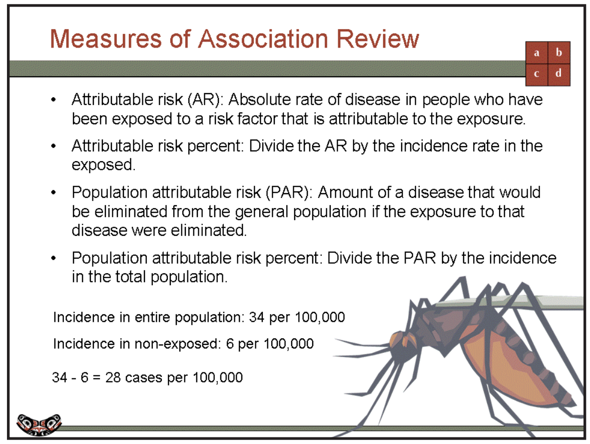
When we calculate the attributable risk percent of the exposed factory workers, we get 98 percent. So we can say that 98 percent of the lung disease in the exposed group is attributable to their exposure.

If we assume that the incidence of lung disease in the rest of the county is the same as that in the non-exposed factory workers, we find that the overall incidence in the community is 27.8 per thousand persons per year. Keep in mind that this figure includes the factory workers because they're part of the county population.

The population attributable risk is the incidence in the entire population (27.8 per thousand per year) minus the incidence in the non-exposed (20 per thousand per year), or 7.8 cases per thousand persons per year. The population attributable risk percent is the PAR divided by the incidence in the total population (2.8/27.8), or 28%. We can say that in this community, 28% of the cases of lung disease are attributable to the exposure in the factory. Expressed another way, if we eliminate the exposure to toxic substances in the factory, we can reduce lung disease in the county by 28%.

Clearly, exposed workers in the plant have a high risk for developing lung disease, but if we look at the county as a whole, we can see that this risk does not constitute as large a public health problem as the exposed factory workers' attributable risk percent of 98 would suggest.

Measures of Association Review



Let's take a moment to review the measures of association we've covered.

Attributable risk, or AR, is the absolute rate of a health event in people who have been exposed to a risk factor that is attributable to the exposure.

We can express attributable risk as a percentage of the overall risk by subtracting the incidence rate in the unexposed from the incidence rate in the exposed, which gives us the AR, and then dividing the resulting AR by the incidence rate in the exposed.

Population attributable risk, or PAR, on the other hand, tells us the amount of a disease that would be eliminated from the general population if the exposure to that disease were eliminated.

We calculate the PAR by subtracting the incidence in the non-exposed from the incidence in the entire population.

Why would we want to know the population attributable risk? Because it can help guide public health decisions about where to focus resources most effectively. In Houston, for example, the incidence of St. Louis encephalitis, or SLE, in the entire city was 34 per 100,000. We can calculate the risk of getting SLE for the general population by subtracting the incidence rate in the people who weren't exposed to standing water, which was 6 per 100,000, from the incidence rate for the entire city. And we get a PAR of 28 cases per 100,000 people that could theoretically be eliminated if standing water were eliminated, and if exposure to standing water was responsible for the entire excess in SLE incidence.

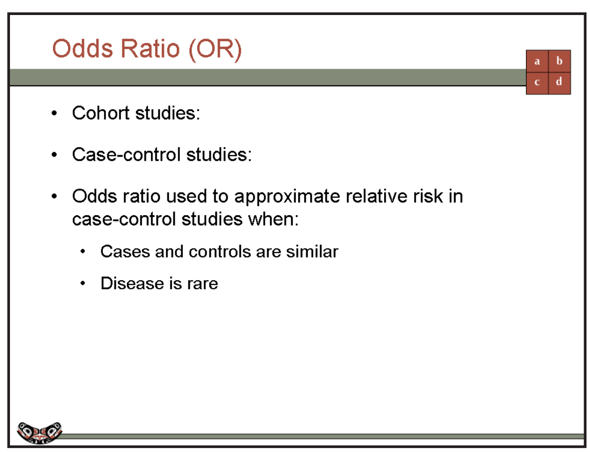
As with the AR percent, to turn the PAR into a population attributable risk percent, we divide the PAR by the incidence rate in the exposed population.

Let's pause now so you can answer some questions on what you have just learned.

Practice: Measures of Association

Exercise 4

Odds Ratio (OR)



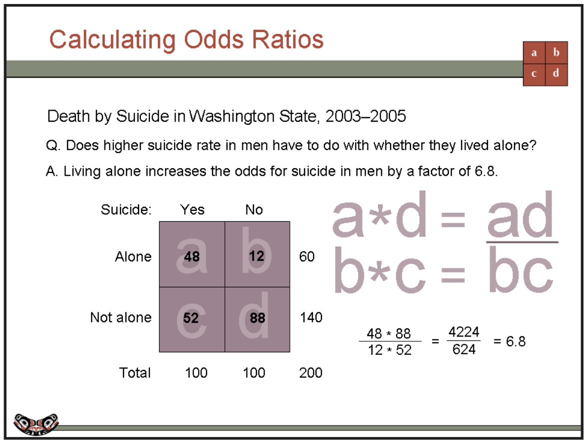
So far, the calculations we've discussed are useful when incidence rates (perhaps from cohort studies) are available, since you can calculate relative risk directly.

However, since cohort studies are time-consuming and expensive, epidemiologists often investigate association and risk using data from case-control studies. Case-control studies, though, do not produce incidence rates, so we cannot calculate a true relative risk.

Fortunately, case-control studies can produce odds ratios, which approximate relative risks in certain situations: when the cases are from the same populations as the controls, and when the disease is relatively rare in the population (say, less than 5%).

Let's use an example to see how to calculate odds ratios.

Calculating Odds Ratio

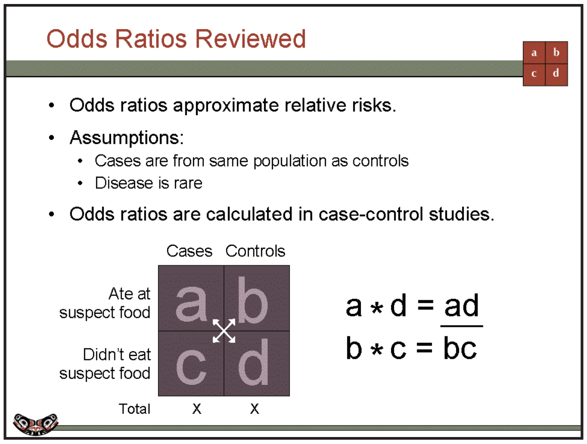


When we examine the rates of death by suicide among men and women living in Washington State between 2003 and 2005, we find that men had a higher suicide rate. Suppose we have an idea that the higher rate in men might have something to do with whether they were living alone. We decide to do a case-control study and interview 100 family members of persons who died by suicide in 2003 and later, and 100 family members of persons of the same age who did not die from suicide.

To calculate the odds ratio, we multiply the number in cell a (the suicides who lived alone) times the number in cell d (the non-suicides who didn't live alone) and the numbers in cells b (non-suicides living alone) and c (suicides not living alone). We then divide the product ad by the product bc. Here is the calculation.

This ratio is our estimation of the relative risk for suicide for men living alone as compared to men not living alone. We could say that living alone increases the risk for suicide in men by a factor of 6.8.

Odds Ratio Reviewed



Odds ratios approximate, or allow you to estimate, relative risk when incidence data in exposed and non-exposed populations are not available. We also need to make two assumptions: the cases are from the same populations as the controls and the disease is relatively rare in the population.

Odds ratios are calculated from case-control studies. We usually set up the data in a 2x2 table and do the calculations by cross-multiplying the data in the corner cells, and dividing the products. The number of controls usually equals the number of cases and are expressed in the column totals.

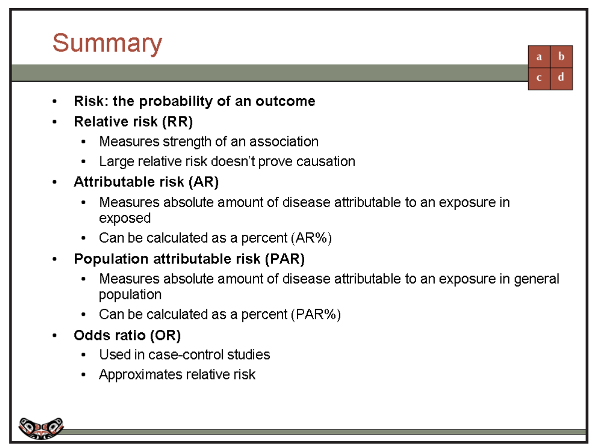
In public health practice, one of the most common uses of an odds ratio is in determining a possibly responsible food item in a restaurant-based food-borne outbreak. A health department investigator questions both ill and well people who ate at the restaurant before the beginning of the outbreak. The investigator then compares the food consumption history for each food in people who are and who are not ill to try to identify the food that might be responsible for the illness.

Let's pause now so you can answer some questions on what you have just learned.

Practice: Calculating Odds Ratio (part 1)

Practice: Calculating Odds Ratio (part 2)

Summary



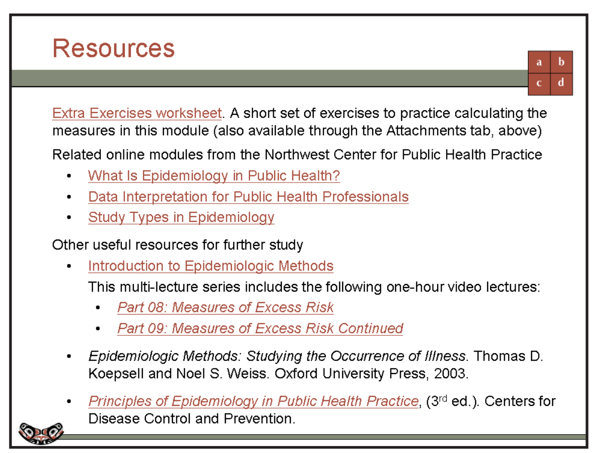
To summarize, in this module we've talked about risk and association as used by epidemiologists. Risk is the probability of an outcome (often a negative outcome) in a specified period of time.

Relative risk is the ratio of the incidence in the exposed to the incidence in the non-exposed groups. Relative risk measures the strength of an association and is useful when looking at possible causes of disease. A larger relative risk provides stronger evidence for a causal role for the exposure but does not prove causation.

Attributable risk and population attributable risk measure the absolute amount of disease attributable to an exposure in either the exposed or total population. Both can be calculated as percents. Public health policy makers use these measures to help decide where to allocate scarce resources.

An odds ratio is an estimate of relative risk, using data from case-control studies. Odds ratios are used when we don't have actual incidence rates. As with relative risk, the larger the odds ratio, the stronger the evidence for a causal relationship between an exposure and the outcome of interest.

Resources



If you would like to learn more about the concepts in this module, you might want to explore some of the resources listed here.

Now, if you're ready, please go on to the final assessment.