Epidemiologic Measures of Association and Impact

Part III

Welcome to this third module in a three-part series focused on epidemiologic measures of association and impact.
This three-part series focuses on the estimation of the association between exposures and outcomes.

In the first two modules, we learned about epidemiologic measures that can be used to determine if there is an association between exposure and disease. Furthermore, we learned about measures that can be used to summarize the strength of the association between exposure and disease outcome.

In this module, we will learn about epidemiologic measures that can be used to quantify the potential for prevention. Specifically, we will learn about measures that can be used to quantify the amount of disease incidence that can be attributed to a particular exposure. Throughout this series, we will focus on the interpretation of study results, and in particular, measures of the association between exposure and disease outcome.
In this module, we will move beyond measures of association and will discuss measures for estimating the potential for prevention, which relates to measures of impact.
The four measures of impact that we will discuss are:

1. Attributable risk
2. Attributable risk %
3. Population attributable risk
4. Population attributable risk %
When estimating the attributable risk, or the impact of a particular exposure, we will utilize an attributable risk model in which we assume that the risk of disease in the unexposed group is a function of the risk not due to the particular exposure under investigation. This is the background incidence of disease.

Then, among participants in the exposed group, we assume that the risk of disease is a function of the risk not due to the particular exposure under investigation (the background incidence) as well as the risk due to the particular exposure under investigation. Under this model, we can subtract the risk in the unexposed group from the risk in the exposed group to derive an estimate of the risk of the disease outcome that can be attributed to the exposure of interest.
Based on the model summarized on the previous slide, we define the attributable risk, or the risk difference, as the amount of disease risk, among exposed, that can be attributed to a specific exposure.

The attributable risk is of importance for public health and clinical practice because it helps us to determine how much of the disease risk, among the exposed, can be prevented if we eliminate the exposure?

The attributable risk is calculated as the risk in the exposed group minus the risk in the unexposed group where the risk in the unexposed group is assumed to be the background risk.
Let’s now consider an example where we focus on the risk of lung cancer that can be attributed to smoking.

In practice, we would like to know how much of the risk of lung cancer, among smokers, is due to smoking? Also, we would like to know how much of the risk of lung cancer, among smokers, can be prevented if they did not smoke?

Attributable risk can be calculated to answer these questions.
Let's consider a data example where we followed 800 smokers and 1200 non-smokers for the development of lung cancer over a 20 year period.

We first calculate the risk of lung cancer among the smokers. Among the 800 smokers, 90 developed lung cancer. This results in a risk estimate of 0.113 or 113 per 1,000 people.

\[
\text{Risk}_e = \frac{90}{800} = 0.113 = 113 \text{ per 1,000}
\]
Next, we will calculate the risk of lung cancer among the non-smokers. Among the 1200 non-smokers, 10 developed lung cancer. This results in a risk estimate of 0.008 or 8 per 1,000 people.
The attributable risk is then calculated as the risk of lung cancer among the smokers minus the risk of lung cancer among the non-smokers. The attributable risk is $0.113 - 0.008 = 0.105$.

105 of the 113 incident cases of lung cancer among 1,000 smokers are attributable to smoking.

The attributable risk is then calculated as the risk of lung cancer among the smokers minus the risk of lung cancer among the non-smokers. The attributable risk is $0.113 - 0.008 = 0.105$.

We would interpret this as 105 of the 113 incident cases of lung cancer among 1,000 smokers are attributable to smoking.
We can then calculate an attributable risk proportion or percentage to determine what proportion of risk in exposed persons is due to the exposure.

The attributable risk proportion is of importance for public health and clinical practice because it helps us to determine what proportion of the disease, among exposed, can be prevented if we eliminate the exposure?

The attributable risk proportion is calculated as the risk in the exposed group minus the risk in the unexposed group and then this difference is divided by the risk in the exposed group. The percentage is found by multiplying the proportion by 100%.
Let’s now consider an example where we focus on the risk of lung cancer that can be attributed to smoking.

In practice, we would like to know what proportion of the risk of lung cancer, among smokers, is due to smoking? Also, we would like to know what proportion of the risk of lung cancer, among smokers, can be prevented if they did not smoke?

The attributable risk proportion can be calculated to answer these questions.
We will consider the same data example where we followed 800 smokers and 1200 non-smokers for the development of lung cancer over a 20 year period.

We first calculate the risk of lung cancer among the smokers. Among the 800 smokers, 90 developed lung cancer. This results in a risk estimate of 0.113 or 113 per 1,000 people.

\[ R_e = \frac{90}{800} = 0.113 = 113 \text{ per 1,000} \]
Next, we will calculate the risk of lung cancer among the non-smokers. Among the 1200 non-smokers, 10 developed lung cancer. This results in a risk estimate of 0.008 or 8 per 1,000 people.

\[
R_0 = \frac{10}{1200} = 0.008 = 8 \text{ per 1,000}
\]
The attributable risk proportion is then calculated as the risk of lung cancer among the smokers minus the risk of lung cancer among the non-smokers, where this difference is divided by the risk among the smokers. The attributable risk proportion is \((0.113 - 0.008)/0.113 = 0.93\).

We would interpret this as 93% of the risk of lung cancer is attributable to smoking.
Now, let’s return to our listing of measures of impact. We have discussed two measures of impact, namely, the attributable risk and the attributable risk percentage. Both of these measures provide a measure of the impact of the exposure among the exposed.

In practice, we may want to derive measures of the impact that are relevant for the entire population, noting that the population is a mixture of exposed and unexposed.

We will now discuss the population attributable risk and the population attributable risk percentage; these measures of impact are relevant for the entire population.
We define the population attributable risk, as the amount of disease risk, in the population, that can be attributed to a specific exposure.

The population attributable risk is of importance for public health and clinical practice because it helps us to determine how much of the disease risk, in the population, can be prevented if we eliminate the exposure?

The population attributable risk is calculated as the risk in the population minus the risk in the unexposed group where the risk in the unexposed group is assumed to be the background risk.
Let’s now continue our discussion of the example where we focus on the risk of lung cancer that can be attributed to smoking.

In practice, we would like to know how much of the total risk of lung cancer, in the population, is attributable to smoking? Also, we would like to know how much of the total risk of lung cancer, in the population, can be prevented if we eliminated smoking?

Population attributable risk can be calculated to answer these questions.
Let's consider the same data example where we followed 800 smokers and 1200 non-smokers for the development of lung cancer over a 20 year period.

We first calculate the risk of lung cancer in the total population. Among the 2000 participants, 100 developed lung cancer. This results in a total risk estimate of 0.050 or 50 per 1,000 people.

\[ R_t = \frac{100}{2000} = 0.05 = 50 \text{ per 1,000} \]
Next, we will calculate the risk of lung cancer among the non-smokers. Among the 1200 non-smokers, 10 developed lung cancer. This results in a risk estimate of 0.008 or 8 per 1,000 people.
The population attributable risk is then calculated as the risk of lung cancer in the total population minus the risk of lung cancer among the non-smokers. The population attributable risk is $0.050 - 0.008 = 0.042$.

Based on these results, we can state that if smoking were eliminated, the risk of lung cancer in the population would be reduced by 42 cases per 1,000 population.
We can then calculate a population attributable risk proportion or percentage to determine what proportion of risk in population is due to the exposure.

The population attributable risk proportion is of importance for public health and clinical practice because it helps us to determine what proportion of the disease, in the total population, can be prevented if we eliminate the exposure?

The population attributable risk proportion is calculated as the risk in the total population minus the risk in the unexposed group and then this difference is divided by the risk in the total population. The percentage is found by multiplying the proportion by 100%.
Let’s now continue the discussion of our data example where we focus on the risk of lung cancer that can be attributed to smoking.

In practice, we would like to know what proportion of the total risk of lung cancer, in the population, is due to smoking? Also, we would like to know what proportion of the total risk of lung cancer, in the population, can be prevented if they did not smoke?

The population attributable risk proportion can be calculated to answer these questions.
We first calculate the risk of lung cancer in the total population. Among the 2000 participants, 100 developed lung cancer. This results in a total risk estimate of 0.050 or 50 per 1,000 people.

Next, we will calculate the risk of lung cancer among the non-smokers. Among the 1200 non-smokers, 10 developed lung cancer. This results in a risk estimate of 0.008 or 8 per 1,000 people.

The population attributable risk proportion is then calculated as the risk of lung cancer in the total population minus the risk of lung cancer among the non-smokers and then dividing this result by the risk of lung cancer in the population.

The population attributable risk proportion is \((0.050 - 0.008)/0.050 = 0.042/0.050 = 0.84\).

We would interpret this result as 84% of lung cancer incidence in total pop is due to smoking.
Assumptions of AR

- The association is causal
- All other variables equally distributed between exposed and unexposed
- i.e. the only difference is the presence of the exposure

Note that the attributable risk model, upon which our calculations are based, includes several key assumptions. First, the model assumes that the relation between the exposure and outcome is causal (the exposure causes the disease outcome). Next, we assume that all other variables, or patient characteristics and factors, are equally distributed, or balanced, between the exposed and unexposed groups. Therefore, under these assumptions, we can focus on the causal impact of the exposure of interest alone and avoid the confounding impact of other related factors.

These assumptions may not hold in practice, and a more complicated attributable risk model would need to be used.
Now, let’s relate a few of the concepts that we have learned in this module.

Let’s discuss the relation between relative risk and population attributable risk.
We will consider two different diseases, lung cancer and cardiovascular disease (CVD).

In this table, we have the mortality rates per 1,000 person years for each disease type according to smoking status.
We first note that among smokers, the risk of CVD death is higher than the risk of death from lung cancer.
We can use the incidence rates on the previous slide to calculate the relative risk of lung cancer and CVD associated with smoking. We can also calculate the attributable risk percentage for the risk of death due to each cause that is attributable to smoking.

<table>
<thead>
<tr>
<th></th>
<th>Lung Cancer</th>
<th>CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>18.5</td>
<td>1.3</td>
</tr>
<tr>
<td>AR%</td>
<td>95%</td>
<td>23%</td>
</tr>
</tbody>
</table>

Based on these summaries, which cause of death has a stronger association with smoking?
Lung cancer-related deaths have a stronger association with smoking. The risk of lung cancer-related death for smokers is 18.5 times the risk for non-smokers. Furthermore, 95% of the risk of lung cancer-related death among smokers can be attributed to smoking. These values are much higher than the values of 1.3 and 23% seen when we focus on CVD-related death.
Now, let’s consider the impact of smoking at a population level for deaths due to lung cancer and deaths due to CVD.

When we calculate the population attributable risk associated with smoking we see that the PAR value for CVD is much higher than the PAR for lung cancer-related deaths. Why is this the case?
Recall that the population attributable risk is calculated as the risk in the population minus the risk in the unexposed. If the risk in the total population is high, the population attributable risk will be high, assuming a fixed risk in the unexposed.

Reminder

- PAR = Risk in population – Risk in unexposed
- When risk in population is high -> PAR
In our example, deaths due to CVD were high (more people die from CVD than from lung cancer, in the population).

<table>
<thead>
<tr>
<th></th>
<th>Smokers</th>
<th>Non-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Cancer</td>
<td>1.30</td>
<td>0.07</td>
</tr>
<tr>
<td>CVD</td>
<td>9.51</td>
<td>7.32</td>
</tr>
</tbody>
</table>

In our example, the mortality rates for CVD are much higher than the mortality rates for lung cancer. Meaning, more people die from CVD than from lung cancer in the population. As a result, the population attributable risk will be higher for CVD than for lung cancer.

In summary, although smoking has a stronger association with lung cancer-related deaths, at a population level, because death due to CVD is more common, the PAR associated with smoking is higher for CVD than for lung cancer-related deaths.
Up to this point, we have focused on point estimates for the measures of association. In practice, we will couple the point estimates, say, the estimate for the relative risk, with a confidence interval. Recall from a previous module that the confidence interval is useful for quantifying the degree of uncertainty and variability in our estimate.
The confidence interval provides a range of reasonable estimates for the true parameter value, for example, the relative risk or the odds ratio.

The confidence interval width reflects the precision of our estimation. A wide interval corresponds to poor precision while a narrow interval corresponds to better precision.

The level of confidence reflects our certainty in the estimates upon repeated sampling from the population. The confidence interval provides an indication of how much the estimates would vary if we drew a sample from the population and calculated the parameter estimate multiple times with each repeated sample.
The level of confidence reflects the probability, across repeated samples from the population, that our calculated interval includes the true population parameter value.

If we sample the same population in the same way 100 times, for 95 of the times, the true value of the parameter of interest (OR or RR or IRR or PPR) would be included in our 95% CI. The figure represents this repeated sampling process.

This means that across repeated samples, 5% of the time, our interval will not include the true parameter value. A given confidence interval either will or will not include the true parameter value; we don’t know for certain if our interval includes the true value or not, we only know that across repeated samples, 95% of the time, the interval will include the true value.
Let’s consider an example. Let’s say that we calculated the odds of smoking for patients with lung cancer and compared that to the odds of smoking among participants without lung cancer. If the resulting odds ratio is 2.5, we would state that the odds of smoking for participants with lung cancer are 2.5 times the odds of smoking for participants without lung cancer.

We will couple this point estimate with an estimated confidence interval. The confidence interval ranges from 1.4 to 4.6.

We interpret this interval as, 95 times out of 100, the true OR will lie between the calculated interval (1.4 and 4.6 in this example). We are 95% confident that the true population OR lies between 1.4 and 4.6.

We note that this interval does not include the “null hypothesis” value of 1, which would indicate that the odds of smoking are equal between those with lung cancer and those without lung cancer. Therefore, we can state, that because the confidence interval lies entirely above 1, that the odds of smoking as significantly higher for participants with lung cancer compared to those without lung cancer based on a 2-sided 0.05 alpha level (corresponding to the 95% level of confidence).
Recall from an earlier module that we can use the estimated confidence interval to conduct a hypothesis test.

When considering differences in parameter values, say the difference in the mean total cholesterol between treated and control patients, we compared the confidence interval to a “null hypothesis” value of 0 (meaning, a value of no difference).

When analyzing ratios, such as the risk ratio or the odds ratio, the “null hypothesis” value is 1 instead of 0. A ratio of 1 indicates that the numerator and denominator are equal. For example, the risk of lung cancer among smokers (numerator) is equal to the risk of lung cancer among non-smokers (denominator).

To test the null hypothesis, of equal risk or equal odds between the groups, we can compare the confidence interval to a “null hypothesis” value of 1. If the 95% confidence interval includes the value of 1 then we know that the p-value will be greater than 0.05. On the other hand, if the interval does not include 1, then we know that the p-value is less than or equal to 0.05 and the results are statistically significant.
Consider the confidence intervals in each of these three examples. Let’s assume that these intervals are for the odds of smoking comparing cases with CVD to controls without CVD.

In the first example, the 95% confidence interval for the OR is wide and includes the “null hypothesis” value of 1. We would conclude that there is no significant association between the odds of smoking and CVD.

In the second example, the 95% confidence interval for the OR lies entirely above the “null hypothesis” value of 1. We would conclude that there is a significantly increased odds of smoking for patients with CVD compared to those without CVD.

In the third example, the 95% confidence interval for the OR lies entirely below the “null hypothesis” value of 1. We would conclude that there is a significantly lower odds of smoking for patients with CVD compared to those without CVD.
We noted that the width of the confidence interval conveys information regarding the precision of our estimation.

The width of the interval is a function of the degree of uncertainty in our estimate (i.e., the level of confidence) and also the standard error of our estimate, which is a function of the sample size and variance.

A larger sample size results in a narrower interval and as a result, our estimate is more precise.

A smaller sample size results in a wider interval and as a result, our estimate is less precise.
Let’s consider an example, which confidence interval would arise from a larger sample size when considering sampling from a study setting and population with a fixed OR value?

Answer: the interval 1.4 to 1.9 is narrower and therefore, we expect that this interval would arise from a larger sample size. As the sample size increases, the width of the confidence interval decreases.
Let’s now consider a few review questions.

Choose the measure that would best address the following question.

How many lung cancer cases could be prevented among smokers if smoking were eliminated?

- Relative Risk
- Attributable Risk
- Attributable Risk Proportion
- Population Attributable Risk

The key phrases are “how many” and “among smokers”. Therefore, we recognize that the attributable risk would answer this question.
Now, choose the measure that would best address the following question.

What proportion of lung cancer risk in the total population is attributable to smoking?

- Relative Risk
- Attributable Risk
- Attributable Risk Proportion
- Population Attributable Risk Proportion

The key phrases are “proportion” and “total population”. Therefore, we recognize that the population attributable risk proportion would answer this question.
In summary, we have discussed four new measures of the impact of a given exposure on health outcomes among the exposed and among the total population. We recognize the importance of coupling our point estimates with confidence intervals when presenting estimates. The confidence interval conveys information regarding the variability and uncertainty in our estimates. Wider intervals correspond to less precise estimates. As the sample size increases, the width of the confidence interval will decrease. Finally, we can use confidence intervals to implement hypothesis tests. For tests of differences, we compare the interval to 0, for tests of ratios, we compare the interval to a "null hypothesis" value of 1.

This concludes the series focused on epidemiologic measures of association and impact.