In this video, we are going to introduce the relative risk, a statistical measure of association between two categorical variables. We will see what it is used for, how to calculate it along with its confidence interval, and how to interpret it.

Definition

Exposure group	Outcome present		
	Yes	No	
Yes	a	b	**a+b (fixed)**
No	c	d	**c+d (fixed)**
	a+c	**b+d**	**Total**

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

The relative risk (RR) is a descriptive statistic that measures the association between the exposure to a risk factor and an outcome. It is the ratio of the probability of an outcome in an exposed group to the probability of an outcome in an unexposed group. Relative risk is particularly appropriate for cohort studies, a prospective study design that samples individuals based on their exposure status. The exposed and unexposed groups are followed over time, counting how many participants from each group go on to have the outcome of interest.

This information can be presented in a 2x2 contingency table. The “fixed” term in the table refers to the numbers under the investigator’s control and essentially fixed by the study design. However, the totals from each column are not fixed, as we cannot control what happens to each participant (if anything happens at all).

In a cohort study the Relative Risk (RR) is essentially calculating the ratio of the incidence in each exposure group.

So using the table above, we can calculate the risk of the outcome in each exposure group:

The risk of the outcome in the exposed group = a/(a+b) = incidence in the exposed group.

The risk of the outcome in the unexposed group = c/(c+d) = incidence in the unexposed group.

Therefore, the relative risk is:
a/(a+b) / c/(c+d)
The relative risk does not provide any information about the absolute risk of the outcome occurring, but rather the higher or lower likelihood of the outcome in the exposure versus the non-exposure group.

Confidence interval

$$ln\ 95\%\ CI = ln\ RR \pm 1.96 \times SE(ln\ RR)$$

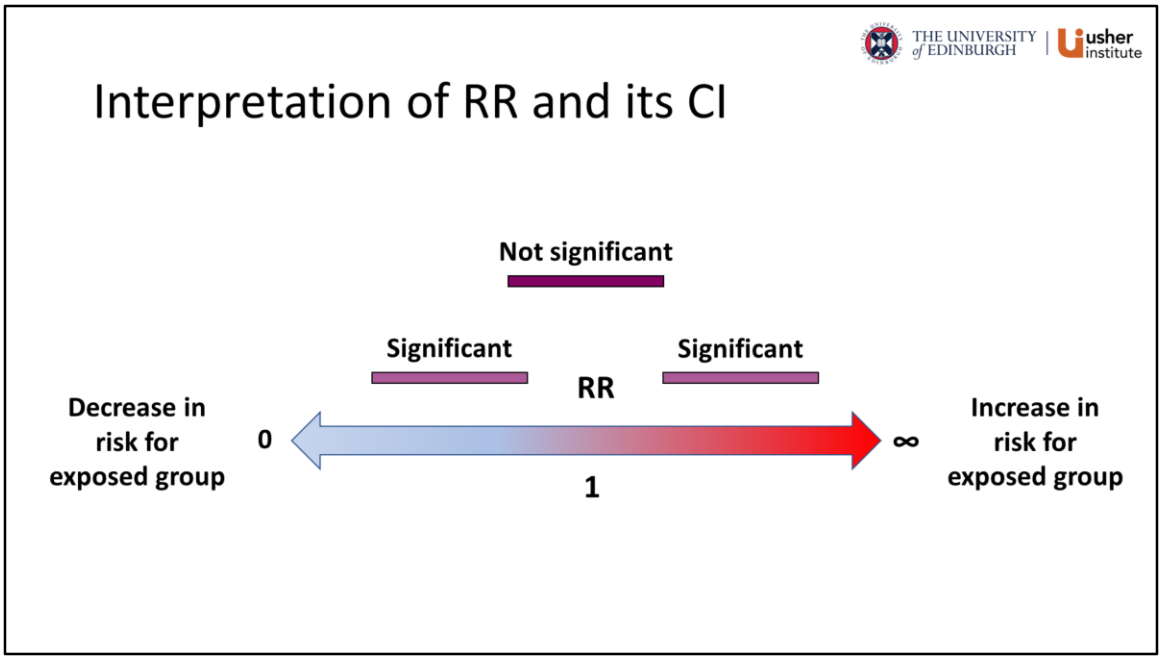
$$ln\ 95\%\ CI = ln\ RR \pm 1.96 \times \sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}$$

Take the exponential of the logarithm of the CI boundaries

$$95\%\ CI = e^{ln\ 95\%\ CI}$$

It is important to provide both the relative risk estimate itself and the accompanying 95% confidence interval. This shows both the significance and the precision of the estimate.

The 95% CI is calculated by using a log transformation of the RR and adding or subtracting the standard error of the log of the relative risk multiplied by 1.96. As with other confidence interval calculations, 1.96 comes from the fact that 95% of the area of a normal distribution is within 1.96 standard deviations of the mean. The boundaries of this log of the confidence interval are then back transformed to be useful. Most statistical software packages will generate the CI, but be careful when interpreting to make sure you are looking at the CI and not the log transformation of it.



When interpreting the confidence interval (CI) for a relative risk estimate, it is important to look at where the boundaries lie relative to 1.

It is worth noting that, although the relative risk is greater than zero, it can be any positive number – the figure above is not symmetrical around 1.

A relative risk of 1 indicates that both groups have the same risk – so that there is no difference between the exposed and unexposed groups. If the 95% CI contains 1, then you cannot say, with 95% certainty, that the groups are different.

If both boundaries of the CI are greater than 1, there is a significant increase in risk for the exposed group, relative to the unexposed group.

If both boundaries of the CI are smaller than 1, there is a significant decrease in risk for the exposed group, relative to the unexposed group.

The further away from 1 the confidence interval is located (in either direction), the larger the association we are seeing. The smaller the interval, the more precise the estimate.

Example

Image by Marijana from Pixabay

Exposure group	Breast cancer		
	Yes	No	
Smoker	65	435	**500**
Non-smoker	50	450	**500**
	115	**885**	**Total**

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{65/500}{50/500} = 1.3$$

$$ln\ 95\%\ CI = ln\ 1.3 \pm 1.96 \times \sqrt{\frac{435/65}{500} + \frac{450/50}{500}}$$

$$95\%\ CI = 0.91; 1.84$$

Let's say that we have a sample of 1000 women, 500 of whom smoke and 500 of whom do not smoke, and we count how many of them have had breast cancer in their lifetime. This is the cross-tabulation presenting the results. The relative risk estimate of getting breast cancer for the smoking group is slightly elevated (1.3) but because the confidence interval contains 1, that difference in risk is not significant. The interval seems very narrow and therefore is very precise: we can conclude that the risk of breast cancer is equivalent in the smoking and non-smoking groups in this population.

Example Doll & Hill

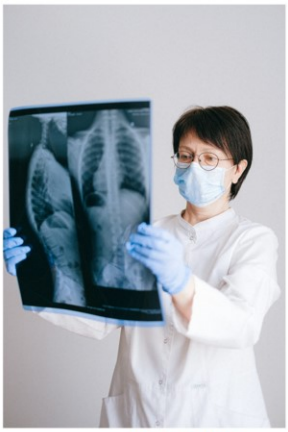


Photo by Anna Shvets

Exposure group	Death		Person years at risk
	Yes	No	
Smoker	133	102467	**102600**
Non-smoker	3	42797	**42800**
	136	**145264**	**145400**

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{133/102600}{3/42800} = 18.5$$

$$ln\ 95\%\ CI = ln\ 18.5 \pm 1.96 \times \sqrt{\frac{102467/133}{102600} + \frac{42797/3}{42800}}$$

$$95\%\ CI = 5.9; 58.1$$

We will now look at a real example, from a classic epidemiology paper by Doll and Hill, which studied lung cancer mortality and smoking in physicians.
The RR for smokers compared to non-smokers is 18.5, which clearly indicates an increase in death from smoking compared to non-smoking. It is worth noting that the protective effect of not smoking would be 1/18.5, so 0.054, which might sound less dramatic but conveys an identical message.
However, the 95% CI is important to gauge whether this is a significant result. The CI is 5.9 to 58.1, which does not contain 1. Therefore, the observed risk increase can be considered significant. The population relative risk might be as small as 5.9, which is still a sizable effect, or as large as 58.1, which is a huge effect.
The CI is wide, which is due to the relatively small number of deaths compared to the person years at risk, particularly in the non-smoking group.

Summary - TBC

- **Relative risk**
  - Risk in exposed / risk in unexposed
- **Confidence interval**
  - Use logarithm and 1.96
- **Interpretation**
  - Effect size
  - Overlap with 1
  - Width

In this video, we have seen that the relative risk is the ratio of the probability of an outcome in an exposed group over the probability of the outcome in an unexposed group. The confidence interval is calculated for the log-transformation of the RR, and then back-transformed. Finally, to interpret the RR and its CI, you should comment on the effect size, the overlap of the CI with 1 and the width of the CI.