THE UNIVERSITY of EDINBURGH | usher institute

Odds Ratio

Confidence interval

In this video, we are going to introduce the Odds ratio, 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**
No	c	d	**c+d**
	a+c (fixed)	**b+d (fixed)**	**Total**

$$OR = \frac{a/b}{c/d} = \frac{a \times d}{b \times c}$$

For a case control design, it isn't appropriate to use a Relative Risk. Indeed, although the results can be presented similarly in a 2x2 table, the underlying structure is different.
For a case-control study, the number of participants is fixed by the investigator for the outcome groups – not for the exposure groups.
This structure prevents relative risk from being a meaningful statistic in this study design.
As a/(a+b) or c/(c+d) would now involve two quantities that are not fixed by the design, the RR will change depending on the proportion of cases chosen relative to controls, even if the mechanism is the same.
We therefore need to use a method which is based on the calculations within each exposure group – we can calculate the odds of the outcome in each exposure group and look at the ratio of these odds. This is called the Odds ratio (OR).
- The odds of an individual in the exposed group having the outcome is a/b
- The odds of an individual in the unexposed group having the outcome is c/d
The Odds ratio is symmetrical, it is worth noting that you can just flip the equation to see the OR for the alternative comparison (A compared to B vs B compared to A). So an OR of 3:2 (1.5) is exactly the same as an OR of 2:3 (0.66).
Odds ratios are also commonly used in multivariate analyses such as logistic

regression, when effect estimates need to be adjusted for factors such as age which may differ between the two groups.

Confidence interval

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

$$\ln 95\%\ CI = \ln OR \pm 1.96 \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

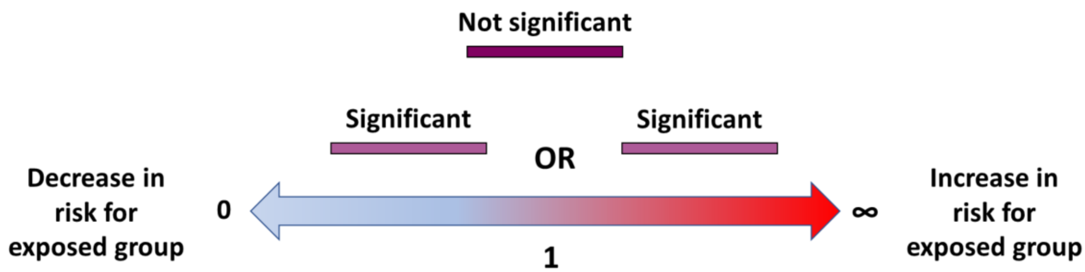
Take the exponential of the logarithm of the CI boundaries

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

It is important to provide both the odds ratio 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 OR and adding or subtracting the standard error of the log of the Odds ratio 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.

Interpretation of OR and its CI



When interpreting the confidence interval (CI) for an odds ratio estimate, it is important to look at where the boundaries lie relative to 1.
It is worth noting that, although the Odds ratio 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 odds – 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 odds 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 odds 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.

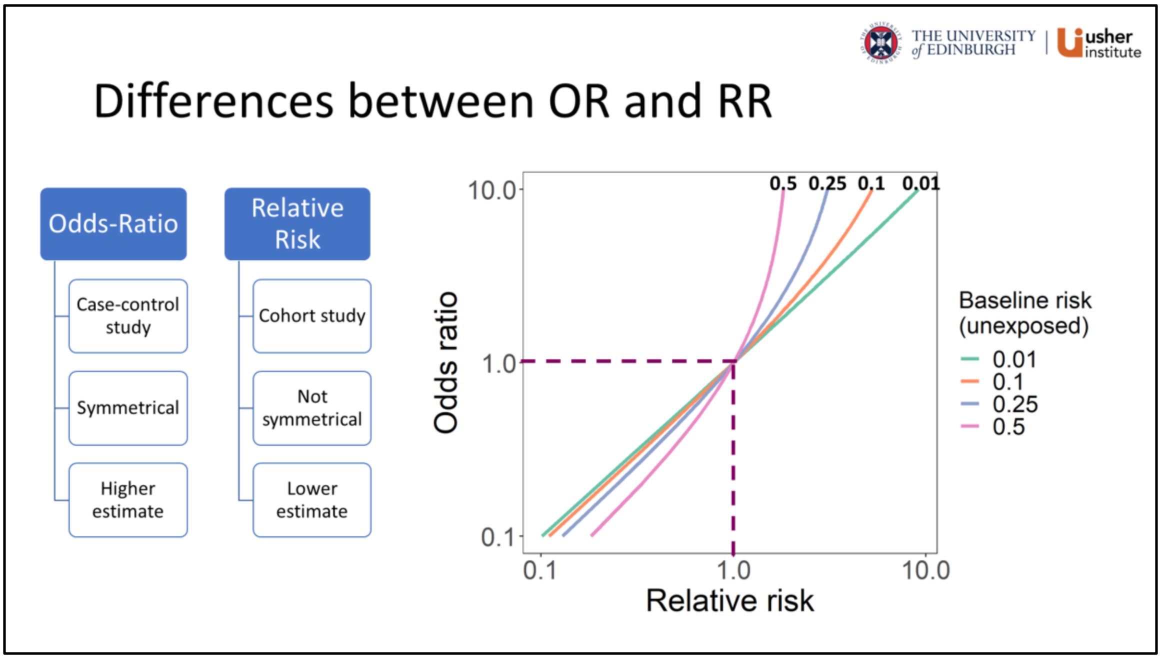
Similarities with Relative Risk

- Measure of association between exposure and outcome
- Relative measures
- Similar value when
  - there is no association
  - the outcome is rare

RR and OR

Exposure → **Outcome**

We can compare the OR and the RR to see their similarities. The OR and RR both indicate the nature of association between exposure and outcome. Both the RR and the OR are relative measures of effect, they are independent of the baseline risk or odds and therefore should give the same result in different regions with different disease levels for example. But this means that neither measure on its own gives an indication of how many outcomes there are. The OR and RR have similar values when there is no association or when the outcome is rare.



However, the RR and the OR also differ in several ways.
First, they are used in different contexts. For a cohort study, the RR is the more easily interpreted measure of association, but it is worth noting that an OR would not be incorrect. However for a case-control study, it is important to use the OR.
The calculations of both the estimate and the standard error are different between the RR and the OR. The OR is symmetrical whereas the RR isn't.
The relationship between the OR and the RR is complex and depends on outcome rates, as illustrated by the line plot here. When there is no association between exposure and outcome, the OR and RR estimates are identical and close to 1. When there is an association between an exposure and an outcome, the OR exaggerates the estimate of their relationship compared with the RR. Thus, when the RR <1, the OR is lower than RR, and when the RR is more than 1, the OR is higher than the RR. When the outcome is rare (typically <10%), the value of the OR is not too different from that of RR, and the two can be used interchangeably irrespective of whether the risk is lower or higher in the exposed group as compared to the unexposed. As outcome rates increase, the two ratios diverge and can no longer be used interchangeably, as you can see from the lines getting more and more curved away from the green straight line as the baseline risk increases.

Example

Photo by Shane Kong on Unsplash

Exposure group	Gastric cancer		
	Yes	No	
Chilli	476	275	**751**
No chilli	24	225	**249**
	500	**500**	**1000**

$$OR = \frac{a \times d}{b \times c} = \frac{476 \times 225}{275 \times 24} = 16.2$$

$$ln\ 95\%\ CI = ln\ 16.2 \pm 1.96 \times \sqrt{\frac{1}{476} + \frac{1}{275} + \frac{1}{24} + \frac{1}{225}}$$

$$95\%\ CI = 10.4; 25.4$$

Let's say that we have surveyed a sample of 1000 people, 500 of whom have had gastric cancer and 500 of whom have not had gastric cancer, and we have asked them whether they eat chilli peppers or not. This is the cross-tabulation presenting the results. An odds ratio of 16.2 means that the odds of having been exposed to chilli are 16 times higher among gastric cancer patients than for individuals who didn’t have gastric cancer. Because the odds ratio is greater than 1, exposure to chilli might be a risk factor for gastric cancer. The magnitude of the odds ratio suggests a strong association. The data suggest that the OR is at least 10 and may be as large as 25 (which would be a huge effect!).

THE UNIVERSITY of EDINBURGH | usher institute

Example Cordina-Duverger et al.

Exposure group	Breast cancer		
	Yes	No	
Night shift work	250	306	**556**
No night shift work	949	1473	**2422**
	1199	**1779**	**2978**

$$OR = \frac{a \times d}{b \times c} = \frac{250 \times 1473}{306 \times 949} = 1.27$$

$$ln\ 95\%\ CI = ln\ 1.27 \pm 1.96 \times \sqrt{\frac{1}{250} + \frac{1}{306} + \frac{1}{949} + \frac{1}{1473}}$$

$$95\%\ CI = 1.05; 1.53$$

Now let's look at a published pooled analysis of case-control studies on breast cancer and night shift work (Cordina-Duverger et al. 2018, European Journal of Epidemiology). The researchers wanted to know whether night shift work might increase the risk of breast cancer. Here night shift work is defined as working for at least 3 hours between midnight and 5 a.m. The researchers pooled case-control studies from different countries and compared the numbers of women who had breast cancer between the group who have never worked a night shift and the group who have. The data for women in Australia is presented in the crosstabulation here. Out of a sample of 2978 women, 1199 had breast cancer and 1779 have not had breast cancer. An odds ratio of 1.27 means that the odds of having done night work are 1.27 times higher among breast cancer patients than for women who didn't have breast cancer. Because the odds ratio is greater than 1, night work might be a risk factor for breast cancer. The magnitude of the odds ratio suggests a weak association. The data suggest that the OR is at least 1.05 and may be as large as 1.53.

THE UNIVERSITY of EDINBURGH | usher institute

Summary

- **Odds ratio**
  - $\frac{Odds\ of\ event\ in\ one\ group}{Odds\ of\ event\ in\ another\ group}$
- **Similar to RR**
  - Association between exposure and outcome
  - Relative measure
- **Different from RR**
  - OR for case-control study
  - OR symmetrical
  - OR larger estimates

Nick Youngson, CC BY-SA 3.0 via Alpha Stock Images

In this video, we have seen that the Odds ratio is the ratio of odds of an event in one group and the odds of that event in another group.
We have seen the similarities between the OR and the RR. They both indicate the nature of association between exposure and outcome, and they are both relative measures of effect.
We have also seen how the OR and the RR are different. While the RR is used in cohort studies, the OR is the appropriate measure of association in a case-control study. The calculations of both the estimate and the standard error are different between the RR and the OR, with the OR being symmetrical, but not the RR.
Finally, when there is an association between an exposure and an outcome that is not rare, the OR exaggerates the estimate of their relationship compared with the RR.