Effect estimates for binary `data:Risk` ratios (RRs) and Odds ratios (ORs)

Alice Kamau

Feb 18, 2016

Learning objective

At the end of this session, you should be able to at least;

- Know how to calculate the measures of disease (risk and odds).
- Know how to calculate the different measures of effect (risk ratios and odds ratios) and understand their value in scientific research.
- Know how to calculate the 95% confidence intervals of these ratios.
- Interpreting the relative measures.

Introduction: defining risk

- The risk (cumulative incidence) is the proportion of persons in a population, intially free of disease, who develop the disease within a specified time interval.
- It is the probaility or risk that an individual will develop a disease during a specified period of time.

$$Risk\,of\,disease = \frac{No.\,of\;\,new\,cases\,in\,a\,given\,time\,period}{No.\,of\,persons\,at\,the\,beginning\,of\,that\,time\,period} \tag{1}$$

Introduction: defining odds

- The odds of disease is a ratio of the probability of getting the disease to the probability of not getting the disease.

- The denominator for an odds measure is all people who are NOT cases.

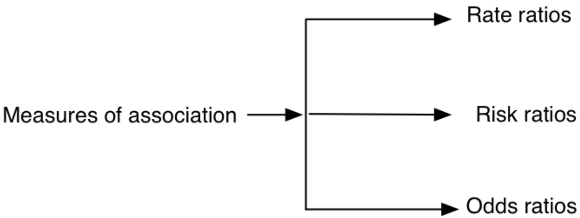
$$Odds\,of\,disease = \frac{No.\,of\ \,new\,cases\,in\,a\,given\,time\,period}{No.\,of\,persons\,that\,weren't\,a\,case\,during\,that\,period} \tag{2}$$

Comparing disease frequency in two groups

- The aim of epidemiological research is to investigate the association between exposure to risk factors (smoking) and occurence of disease (lung cancer).
- It requires incidence/odds of disease in a group of person exposed to a risk factor to be compared with the incidence/odds of disease in a group of persons unexposed.

- Association is a statistical relationship between two or more variables.

- The main measures of association are;



- Question; is there excess risk associated with a given exposure?

Measures of association - risk ratio

	Exposure	No exposure	Total
Disease	A	C	A+C
No Disease	B	D	B+D
Total	A+B	C+D	N

$$Risk\,ratio = \frac{Risk\,in\,the\,exposed\,group}{Risk\,in\,the\,unexposed\,group} \tag{3}$$

$$Risk\,ratio = \frac{A/(A+B)}{C/(C+D)} \tag{4}$$

Measures of association - odds ratio

	Exposure	No exposure	Total
Disease	A	C	A+C
No Disease	B	D	B+D
Total	A+B	C+D	N

$$Odds\,ratio = \frac{Odds\ of\ disease\ in\ the\ exposed\ group}{Odds\ of\ disease\ in\ the\ unexposed\ group} \tag{5}$$

$$Odds\,ratio = \frac{A/B}{C/D} \tag{6}$$

Calculating the 95% confidence intervals (CI) of the ratios.

- Sample estimate gives an imprecise estimate of the population value as it is subject to sampling variation.
- The impression is summarized using the standard error (s.e).
- The s.e is then used to calculate the CI.
- A CI is a way of moving from identifying a single value (for example odds ratio, risk ratio etc.) to a range of likely values that cover the true unknown population value.
- CI gives information on the size of the effect and the impression around this effect

Risk ratio Confidence Interval

$$s.e(log\,RR) = \sqrt{(\frac{1}{A} - \frac{1}{A+B} + \frac{1}{C} - \frac{1}{C+D})} \tag{7}$$

The standard error is derived using the delta method

$$95\%\ CI\ for\ log\ RR = log\ RR \pm 1.96 \times s.e(log\,RR) \tag{8}$$

$$95\%\ CI\ for\ RR = \frac{RR}{EF}\quad to\quad RR \times EF \tag{9}$$

$$Where\ EF = Error\ Factor = exp(1.96 \times s.e(log\,RR)) \tag{10}$$

Odds ratio Confidence Interval

$$s.e(log\,OR) = \sqrt{(\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D})} \tag{11}$$

The formula for s.e.(log OR) is known as Woolf 's formula.

$$95\%\ CI\ for\ log\ OR = log\ OR \pm 1.96 \times s.e(log\,OR) \tag{12}$$

$$95\%\ CI\ for\ OR = \frac{OR}{EF} \quad to \quad OR \times EF \tag{13}$$

$$Where\ EF = Error\ Factor = exp(1.96 \times s.e(log\,OR)) \tag{14}$$

Hypothesis testing

-If the null hypothesis of no difference between the risks/odds in the two groups is true, then;

```
Ho: RR = 1 and hence logRR = 0 or
Ho: OR = 1 and hence logOR = 0
```

-We use the logRR and logOR and their standard error to derive a z statistic and test the null hypothesis in the usual way:

$$z = \frac{log\ RR}{s.e(log\ RR)} \quad (15)$$

$$z = \frac{log\ OR}{s.e(log\ OR)} \quad (16)$$

Interpreting relative measures

The two relative measures are measures of how many times more/less likely people in the exposed group develop the disease than those in the unexposed group.

- Value of 1 indicates that the incidence of disease in the exposed and unexposured groups are identical thus no association between the exposure and disease.
- Value greater than 1 indicate positive/higher association/increased risk among those exposed than among those unexposured.
- Value less than 1 indicates decreased/lower association among those exposed than among those unexposed. Suggesting that the risk factor may be protective.

For example;

	Lung cancer	No lung cancer	Total	Odds	Risk
Smokers	709 (A)	142 (B)	851	709/142 = 4.99	709/851=0.83
Non-smokers	154 (C)	308 (D)	462	154/ 308 = 0.5	154/462=0.33
Total	863	450	1,313	OR ~ 10	RR ~ 2.5

- A risk ratio of 2.5 indicates that smokers are 2.5 times more likely to be diagnosed with lung cancer than non-smokers.
- An odds ratio of 10 indicates that the odds of smoking among lung cancer patients is 10 times higher than the odds of smoking among non-lung cancer patients.

- If the risk ratio was 0.5 it would indicates that smokers are half as likely to be diagnosed with lung cancer than non-smokers.
- If the odds ratio was 0.5 it would indicates that the odds of smoking among lung cancer patients is half less likely than the odds of smoking among non-lung cancer patients.
- A risk/odds ratio of 1 indicates no difference between the groups.

Conclusion

- In case-control study, only OR can be calculated as a measure of association.
- In cohort study, either RR or OR is a valid measure of association.
- Note that risk/odds ratio are always positive number.
- A common mistake in the literature is to interpret an odds ratio as if it were a risk ratio.