stcp-rothwell-risk

## Measures of risk

Medical statistics is a branch of statistics which focuses on medical applications. It introduces new methods for analysing proportions of events, which can be defined as $$proportion = \frac{number\ of\ events\ occuring}{number\ in\ group}$$ or $p = \frac{a}{n}$ where $a$ is a subset of $n$, in terms of risk. For example, the proportion of people who had toast for breakfast would have $a$ as the number of people who ate toast from the sample questioned and $n$ as the total number of people in the sample. This sheet will briefly explain various terms which arise in medical statistics regularly.

**Risk/ Prevalence (P):** The prevalence (or risk) of an event occurring is calculated as:

$$P_{event} = \frac{number\ of\ people\ experiencing\ event}{total\ number\ of\ people\ within\ group}$$

Where the event could be a disease, condition, risk factor, etc.

Risks are not always negative e.g. the risk of death following treatment and the probability of winning the lottery are calculated in the same way.

We are often interested in comparing risks of an event occurring, e.g. death, between different groups and there are several ways of doing this. The following table will be used to demonstrate the formulae involved.

|  | Event occurs | Event does not occur | Total | Risk of Event by group |
|---|---|---|---|---|
| **Exposed/ Treated** | a | b | a+b | $p_{exp} = \frac{a}{(a+b)}$. |
| **Not Exposed/ Not treated** | c | d | c+d | $p_{unexp} = \frac{c}{(c+d)}$ |

**Relative risk**: This measures how much more likely the event is to occur in one group compared to another.

The risk of the event occurring for the exposed population is $P_{exp} = \frac{a}{(a+b)}$.

The risk of the event occurring for the unexposed population is $P_{unexp} = \frac{c}{(c+d)}$.

---

**Relative risk = $RR = \dfrac{risk\ of\ event\ in\ exposed\ group}{risk\ of\ event\ in\ unexposed\ group} = \dfrac{P_{exp}}{P_{unexp}} = \dfrac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}} = \dfrac{a(c+d)}{c(a+b)}$**

This is sometimes called a Risk Ratio. If $RR > 1$ then the risk of disease for the exposed group is larger than the risk of disease for the unexposed group.

**Example:** A randomised controlled trial investigated mortality rates over a year for 300 patients with lung cancer. The first group received a new chemotherapy treatment for lung cancer (New treatment) and the other the standard chemotherapy treatment (Control treatment).

|  | Died | Survived | Total | Risk of dying |
|---|---|---|---|---|
| Control treatment | 50 | 150 | 200 | 50/200 = 0.25 |
| New treatment | 10 | 90 | 100 | 10/100 = 0.1 |

The relative risk is $\dfrac{risk\ of\ dying\ in\ control\ group}{risk\ of\ dying\ in\ new\ treatment\ group} = \dfrac{0.25}{0.1} = 2.5$. This means that those in the control group were 2.5 times more likely to die than those in the treatment group. The relative risk is interpreted in terms of the risk of the group in the numerator. This means that when calculating relative risks, it is easier to use the group with the highest risk in the numerator.

Sometimes a confidence interval is reported with the Relative Risk calculated from a sample. A confidence interval gives a range of values for the population relative risk.

**95% Confidence Interval (CI)** for a RR: For large samples this can be calculated using the natural logarithm (ln) because the confidence interval is not symmetrical.

First the variance of $\ln(RR)$ needs to be calculated:

$$Var(lnRR) = \frac{1}{a} - \frac{1}{a+c} + \frac{1}{b} - \frac{1}{b+d} = \frac{1}{50} + \frac{1}{50+10} + \frac{1}{150} + \frac{1}{150+90} = 0.0475$$

95% Confidence interval for the ln(RR):

$$\ln(RR) \pm 1.96 \times \sqrt{Var(lnRR)} = \ln(2.5) \pm 1.96 \times \sqrt{0.0475} = 0.916 \pm 0.427 =$$
$$(0.489, 1.343)$$

To get the confidence interval for the actual relative risk, take the exponential of the upper and lower value, the **95% confidence interval for a RR** is calculated as follows:

$$\left( e^{\ln(RR) - 1.96 \times \sqrt{Var(lnRR)}}, \; e^{\ln(OR) + 1.96 \times \sqrt{Var(lnRR)}} \right) = (e^{0.489}, e^{1.343}) = (1.63, 3.83).$$

**Reporting Relative Risk**

The relative risk for the whole population is likely to be between 1.63 and 3.83. If the confidence interval does not include 1, the risk in one group is significantly higher than the risk in the second group. In this example, both the upper and lower limit of the 95% CI are greater than 1 so the risk in the control group is significantly higher and the result is reported as: RR 2.5; 95% CI (1.63, 3.83).

The **Relative Risk Difference (RRD)** is given by $RRD = 1 - RR$ when $(RR < 1)$.

Put the smallest risk (treatment group) on top to get a RR under 1: $\frac{0.1}{0.25} = 0.4$. Therefore the RRD is 1 – 0.4 = 0.6. The risk of dying is reduced by 60% in the treatment group.

The **Absolute Risk Difference (ARD)** is given by

$ARR = |P_{exp} - P_{unexp}| = |0.25 - 0.1| = 0.15.$ The absolute risk has decreased by 15%.

The **Number Needed to Treat (NNT)** is defined as the additional number of people you would need to give a new treatment to in order to cure one extra person compared to the old treatment.

The **Number Needed to Treat (NNT)** is calculated using the **Absolute Risk Difference (ARD)** and is given by $NNT = \frac{1}{ARD} = \frac{1}{0.15} = 6.67$. This means that 7 people would need to receive the new treatment in order for one extra person to survive compared to the old treatment.

Another common measure used in medical statistics is the **Odds Ratio (OR).** First, odds are calculated using $odds = \frac{p}{1-p}$, where $p$ is the probability of an event occurring.

Therefore, the odds of the event in the exposed group would be $odds_{exp} = \frac{P_{exp}}{1-P_{exp}}$ and similarly the odds of event in the unexposed group would be $odds_{unexp} = \frac{P_{unexp}}{1-P_{unexp}}$.

Then the odds ratio is $OR = \frac{odds_{exp}}{odds_{unexp}}$, in terms of the 2x2 table defined earlier this can be calculated as $OR = \frac{ad}{bc}$, which would be the odds of disease in the exposed group compared to the unexposed group. If the $OR > 1$ then the odds of disease occurring in the exposed group are larger than the odds of disease in the unexposed group, so exposure to the factor has increased the risk of contracting the disease. For our example the odds ratio is $OR = \frac{ad}{bc} = \frac{50 \times 90}{10 \times 150} = 3$. The odds ratio looks at the odds of being in a particular treatment group given that you had the disease. So, in this example, those who improved were 3 times more likely to have received the new treatment than the control treatment.

**Note:** The odds ratio and relative risk are similar if the total sample is large and the event is rare.

**95% Confidence Interval (CI)** for an OR: This can be calculated for large samples and must be carried out using the natural logarithm (ln) because the confidence interval is not symmetrical.

First the variance of $\ln(OR)$ needs to be calculated. This is done using t hollowing formula based on a 2x2 table.

$$Var(lnOR) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} = \frac{1}{50} + \frac{1}{150} + \frac{1}{10} + \frac{1}{90} = 0.138$$

95% Confidence interval for the ln(OR):

$$\ln(OR) \pm 1.96 \times \sqrt{Var(lnOR)} = \ln(3) \pm 1.96 \times \sqrt{0.138} = (0.362, 1.818)$$

To get the confidence interval for the actual odds ratio, take the exponential of the upper and lower value, the **95% confidence interval for a OR** is calculated as follows:

$$\left(e^{\ln(OR)-1.96\times\sqrt{Var(lnOR)}}, \ e^{\ln(OR)+1.96\times\sqrt{Var(lnOR)}}\right) = (e^{0.362}, e^{1.818}) = (1.45, 6.21)$$

## Reporting Odds Ratio

The odds ratio comparing death rates after the standard treatment to the new treatment was OR 3 suggesting that those on the control treatment are more likely to die. As with the relative risk, the odds ratio is said to be significant if the confidence interval does not include 1. In this example the upper and lower limit of the 95% CI are both greater than 1 so the odds of death are said to be significantly higher in the control group and the result is reported as: OR 3; 95% CI (1.43, 6.16).

## Steps in SPSS

Use *Analyze → Descriptive statistics → Crosstabs* to produce a cross-tabulation of data in 2x2 format and within *Crosstabs → Statistics* select Cochran's and Mantel-Haenszel statistics to produce an odds ratio.

|  |  | Event | | Total |
|---|---|---|---|---|
|  |  | Died | Survived |  |
| Treatment | Control | 50 | 150 | 200 |
|  | New | 10 | 90 | 100 |
| Total |  | 60 | 240 | 300 |

**Mantel-Haenszel Common Odds Ratio Estimate**

| Estimate | | | 3.000 |
|---|---|---|---|
| ln(Estimate) | | | 1.099 |
| Std. Error of ln(Estimate) | | | .371 |
| Asymp. Sig. (2-sided) | | | .003 |
| Asymp. 95% Confidence Interval | Common Odds Ratio | Lower Bound | 1.449 |
| | | Upper Bound | 6.210 |
| | ln(Common Odds Ratio) | Lower Bound | .371 |
| | | Upper Bound | 1.826 |