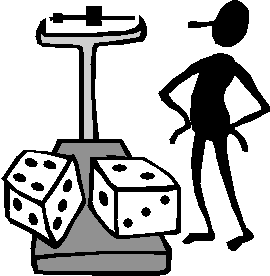
**The Difference Between "Probability" and "Odds"**

The probability that an event will occur is the fraction of times you expect to see that event in many trials.  Probabilities always range between 0 and 1. The odds are defined as the probability that the event will occur divided by the probability that the event will not occur.

If the probability of an event occurring is Y, then the probability of the event not occurring is 1-Y. (Example: If the probability of an event is 0.80 (80%), then the probability that the event will not occur is 1-0.80 = 0.20, or 20%.

The odds of an event represent the ratio of the (probability that the event will occur) / (probability that the event will not occur). This could be expressed as follows:

Odds of event = Y / (1-Y)

So, in this example, if the probability of the event occurring = 0.80, then the odds are 0.80 / (1-0.80) = 0.80/0.20 = 4 (i.e., 4 to 1).

* If a race horse runs 100 races and wins 25 times and loses the other 75 times, the probability of winning is 25/100 = 0.25 or 25%, but the odds of the horse winning are 25/75 = 0.333 or 1 win to 3 loses.
* If the horse runs 100 races and wins 5 and loses the other 95 times, the probability of winning is 0.05 or 5%, and the odds of the horse winning are 5/95 = 0.0526.
* If the horse runs 100 races and wins 50, the probability of winning is 50/100 = 0.50 or 50%, and the odds of winning are 50/50 = 1 (even odds).
* If the horse runs 100 races and wins 80, the probability of winning is 80/100 = 0.80 or 80%, and the odds of winning are 80/20 = 4 to 1.

NOTE that when the probability is low, the odds and the probability are very similar.

With the case-control design we cannot compute the probability of disease in each of the exposure groups; therefore, we cannot compute the relative risk. However, we can compute the odds of disease in each of the exposure groups, and we can compare these by computing the odds ratio. In the hypothetical pesticide study the odds ratio is

OR= (7/10) / (5/57) = 6.65

Notice that this odds ratio is very close to the RR that would have been obtained if the entire source population had been analyzed. The explanation for this is that if the outcome being studied is fairly uncommon, then the odds of disease in an exposure group will be similar to the probability of disease in the exposure group. Consequently, the odds ratio provides a relative measure of effect for case-control studies, and it provides an estimate of the risk ratio in the source population, provided that the outcome of interest is uncommon.

We emphasized that in case-control studies the only measure of association that can be calculated is the odds ratio. However, in cohort-type studies, which are defined by following exposure groups to compare the incidence of an outcome, one can calculate both a risk ratio and an odds ratio.

If we arbitrarily label the cells in a contingency table as follows:

|  |  |  |
| --- | --- | --- |
|  | **Diseased** | **Non-diseased** |
| **Exposed** | a | b |
| **Non-exposed** | c | d |

then the odds ratio is computed by taking the ratio of odds, where the odds in each group is computed as follows:

OR = (a/b) / (c/d)

As with a risk ratio, the convention is to place the odds in the unexposed group in the denominator. In addition, like a risk ratio, odds ratios do not follow a normal distribution, so we use the lo g transformation to promote normality. As a result, the procedure for computing a confidence interval for an odds ratio is a two step procedure in which we first generate a confidence interval for Ln(OR) and then take the antilog of the upper and lower limits of the confidence interval for Ln(OR) to determine the upper and lower limits of the confidence interval for the OR. The two steps are detailed below.

**Computing the Confidence Interval for an Odds Ratio**

To compute the confidence interval for an odds ratio use the formula

/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/ada-reference.gif/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/equation_image163.gif

1. Compute the confidence interval for Ln(OR) using the equation above.
2. Compute the confidence interval for OR by finding the antilog of the result in step 1, i.e., exp(Lower Limit), exp (Upper Limit).

The null, or no difference, value of the confidence interval for the odds ratio is one. If a 95% CI for the odds ratio does not include one, then the odds are said to be statistically significantly different. We again reconsider the previous examples and produce estimates of odds ratios and compare these to our estimates of risk differences and relative risks.

Example:

Consider again the hypothetical pilot study on pesticide exposure and breast cancer:

|  |  |  |
| --- | --- | --- |
|  | **Diseased** | **Non-diseased** |
| **Pesticide Exposure** | 7 | 10 |
| **Non-exposed** | 6 | 57 |

We noted above that

OR= (7/10) / (5/57) = 6.6

We can compute a 95% confidence interval for this odds ratio as follows:

/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/ada-reference.gif/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/equation_image163.gif

Substituting we get the following:

/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/ada-reference.gif/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/equation_image164.gif

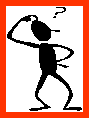
This gives the following interval (0.61, 3.18), but this still need to be transformed by finding their antilog (1.85-23.94) to obtain the 95% confidence interval.

Interpretation: The odds of breast cancer in women with high DDT exposure are 6.65 times greater than the odds of breast cancer in women without high DDT exposure. We are 95% confident that the true odds ratio is between 1.85 and 23.94. The null value is 1, and because this confidence interval does not include 1, the result indicates a statistically significant difference in the odds of breast cancer women with versus low DDT exposure.

Note that an odds ratio is a good estimate of the risk ratio when the outcome occurs relatively infrequently (<10%). Therefore, odds ratios are generally interpreted as if they were risk ratios.

Note also that, while this result is considered statistically significant, the confidence interval is very broad, because the sample size is small. As a result, the point estimate is imprecise. Notice also that the confidence interval is asymmetric, i.e., the point estimate of OR=6.65 does not lie in the exact center of the confidence interval. Remember that we used a log transformation to compute the confidence interval, because the odds ratio is not normally distributed. Therefore, the confidence interval is asymmetric, because we used the log transformation to compute Ln(OR) and then took the antilog to compute the lower and upper limits of the confidence interval for the odds ratio.

Remember that in a true case-control study one can calculate an odds ratio, but not a risk ratio. However, one can calculate a risk difference (RD), a risk ratio (RR), or an odds ratio (OR) in cohort studies and randomized clinical trials. Consider again the data in the table below from the randomized trial assessing the effectiveness of a newly developed pain reliever as compared to the standard of care. Remember that a previous quiz question in this module asked you to calculate a point estimate for the difference in proportions of patients reporting a clinically meaningful reduction in pain between pain relievers as (0.46-0.22) = 0.24, or 24%, and the 95% confidence interval for the risk difference was (6%, 42%). Because the 95% confidence interval for the risk difference did not contain zero (the null value), we concluded that there was a statistically significant difference between pain relievers. Using the same data, we then generated a point estimate for the risk ratio and found RR= 0.46/0.22 = 2.09 and a 95% confidence interval of (1.14, 3.82). Because this confidence interval did not include 1, we concluded once again that this difference was statistically significant. We will now use these data to generate a point estimate and 95% confidence interval estimate for the odds ratio.



We now ask you to use these data to compute the odds of pain relief in each group, the odds ratio for patients receiving new pain reliever as compared to patients receiving standard pain reliever, and the 95% confidence interval for the odds ratio.

|  |  |  |  |
| --- | --- | --- | --- |
| **Treatment Group** | **n** | **# with Reduction**  **of 3+ Points** | **Proportion with Reduction**  **of 3+ Points** |
| **New Pain Reliever** | 50 | 23 | 0.46 |
| **Standard Pain Reliever** | 50 | 11 | 0.22 |

## Answer to Odds Ratio Problem on Page 10

We now ask you to use these data to compute the odds of pain relief in each group, the odds ratio for patients receiving new pain reliever as compared to patients receiving standard pain reliever, and the 95% confidence interval for the odds ratio.

|  |  |  |  |
| --- | --- | --- | --- |
| **Treatment Group** | **n** | **# with Reduction**  **of 3+ Points** | **Proportion with Reduction**  **of 3+ Points** |
| **New Pain Reliever** | 50 | 23 | 0.46 |
| **Standard Pain Reliever** | 50 | 11 | 0.22 |

It is easier to solve this problem if the information is organized in a contingency table in this way:

|  |  |  |
| --- | --- | --- |
|  | **Pain Relief 3+** | **Less Relief** |
| **New Drug** | 23 | 27 |
| **Standard Drug** | 11 | 39 |

Odds of pain relief 3+ with new drug = 23/27 0.8519

Odds of pain relief 3+ with standard drug = 11/39 = 0.2821

Odds Ratio = 0.8519 / 0.2821 = 3.02

To compute the 95% confidence interval for the odds ratio we use

/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/ada-reference.gif/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/equation_image163.gif

Substituting we get

/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/ada-reference.gif/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/equation_image166.gif

Since we used the log (Ln), we now need to take the antilog to get the limits of the confidente interval.

/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/ada-reference.gif/var/folders/4q/_168gj4x40l9t_x1x1pk0f700000gn/T/com.microsoft.Word/WebArchiveCopyPasteTempFiles/equation_image167.gif

The point estimate of the odds ratio is OR=3.2, and we are 95% confident that the true odds ratio lies between 1.27 and 7.21. This is statistically significant because the 95% confidence interval does not include the null value (OR=1.0).



When the study design allows for the calculation of a relative risk, it is the preferred measure as it is far more interpretable than an odds ratio. The odds ratio is extremely important, however, as it is the only measure of effect that can be computed in a case-control study design. When the outcome of interest is relatively rare (<10%), then the odds ratio and relative risk will be very close in magnitude. In such a case, investigators often interpret the odds ratio as if it were a relative risk (i.e., as a comparison of risks rather than a comparison of odds which is less intuitive).