Statistics Notes

The odds ratio

J Martin Bland, Douglas G Altman

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE

J Martin Bland *professor of medical statistics*

ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF

Douglas G Altman *professor of statistics in medicine*

Correspondence to: Professor Bland

BMJ 2000;320:1468

In recent years odds ratios have become widely used in medical reports—almost certainly some will appear in today's *BMJ*. There are three reasons for this. Firstly, they provide an estimate (with confidence interval) for the relationship between two binary ("yes or no") variables. Secondly, they enable us to examine the effects of other variables on that relationship, using logistic regression. Thirdly, they have a special and very convenient interpretation in case-control studies (dealt with in a future note).

The odds are a way of representing probability, especially familiar for betting. For example, the odds that a single throw of a die will produce a six are 1 to 5, or 1/5. The odds is the ratio of the probability that the event of interest occurs to the probability that it does not. This is often estimated by the ratio of the number of times that the event of interest occurs to the number of times that it does not. The table shows data from a cross sectional study showing the prevalence of hay fever and eczema in 11 year old children.[1] The probability that a child with eczema will also have hay fever is estimated by the proportion 141/561 (25.1%). The odds is estimated by 141/420. Similarly, for children without eczema the probability of having hay fever is estimated by 928/14 453 (6.4%) and the odds is 928/13 525. We can compare the groups in several ways: by the difference between the proportions, 141/561 – 928/14 453 = 0.187 (or 18.7 percentage points); the ratio of the proportions, (141/561)/(928/14 453) = 3.91 (also called the relative risk); or the odds ratio, (141/420)/(928/13 525) = 4.89.

Association between hay fever and eczema in 11 year old children[1]

	Hay fever		
Eczema	Yes	No	Total
Yes	141	420	561
No	928	13 525	14 453
Total	1069	13 945	15 522

Now, suppose we look at the table the other way round, and ask what is the probability that a child with hay fever will also have eczema? The proportion is 141/1069 (13.2%) and the odds is 141/928. For a child without hay fever, the proportion with eczema is 420/13 945 (3.0%) and the odds is 420/13 525. Comparing the proportions this way, the difference is 141/1069 – 420/13 945 = 0.102 (or 10.2 percentage points); the ratio (relative risk) is (141/1069)/(420/13 945) = 4.38; and the odds ratio is (141/928)/(420/13 525) = 4.89. The odds ratio is the same whichever way round we look at the table, but the difference and ratio of proportions are not. It is easy to see why this is. The two odds ratios are

$$\frac{141/420}{928/13\,525} \text{ and } \frac{141/928}{420/13\,525}$$

which can both be rearranged to give

$$\frac{141 \times 13\,525}{928 \times 420}$$

If we switch the order of the categories in the rows and the columns, we get the same odds ratio. If we switch the order for the rows only or for the columns only, we get the reciprocal of the odds ratio, 1/4.89 = 0.204. These properties make the odds ratio a useful indicator of the strength of the relationship.

The sample odds ratio is limited at the lower end, since it cannot be negative, but not at the upper end, and so has a skew distribution. The log odds ratio,[2] however, can take any value and has an approximately Normal distribution. It also has the useful property that if we reverse the order of the categories for one of the variables, we simply reverse the sign of the log odds ratio: log(4.89) = 1.59, log(0.204) = – 1.59.

We can calculate a standard error for the log odds ratio and hence a confidence interval. The standard error of the log odds ratio is estimated simply by the square root of the sum of the reciprocals of the four frequencies. For the example,

$$\mathrm{SE}(\log OR) = \sqrt{\frac{1}{141} + \frac{1}{420} + \frac{1}{928} + \frac{1}{13\,525}} = 0.103.$$

A 95% confidence interval for the log odds ratio is obtained as 1.96 standard errors on either side of the estimate. For the example, the log odds ratio is $\log_e(4.89) = 1.588$ and the confidence interval is $1.588 \pm 1.96 \times 0.103$, which gives 1.386 to 1.790. We can antilog these limits to give a 95% confidence interval for the odds ratio itself,[2] as exp(1.386) = 4.00 to exp(1.790) = 5.99. The observed odds ratio, 4.89, is not in the centre of the confidence interval because of the asymmetrical nature of the odds ratio scale. For this reason, in graphs odds ratios are often plotted using a logarithmic scale. The odds ratio is 1 when there is no relationship. We can test the null hypothesis that the odds ratio is 1 by the usual χ^2 test for a two by two table.

Despite their usefulness, odds ratios can cause difficulties in interpretation.[3] We shall review this debate and also discuss odds ratios in logistic regression and case-control studies in future Statistics Notes.

We thank Barbara Butland for providing the data.

1 Strachan DP, Butland BK, Anderson HR. Incidence and prognosis of asthma and wheezing illness from early childhood to age 33 in a national British cohort. *BMJ*. 1996;312:1195-9.
2 Bland JM, Altman DG. Transforming data. *BMJ* 1996;312:770.
3 Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *Evidence-Based Med* 1996;1:164-6.

BMJ: first published as 10.1136/bmj.320.7247.1468 on 27 May 2000. Downloaded from http://www.bmj.com/ on 17 April 2024 by guest. Protected by copyright.