Hypothesis: The Research Page

Odds ratios and relative risks

Marshall Godwin, MD, CCFP, FCFP

This first article in a series on basic statistics deals with odds ratios (OR) and relative risks (RR), how to calculate them from a 2×2 table, how prevalence affects OR and RR, and why OR can be used to estimate RR in case-control studies.[1-3]

Suppose you want to know whether elevated cholesterol levels are associated with peripheral vascular disease (PVD). You do a cohort study: you choose 3000 people with elevated cholesterol and 3000 people without elevated cholesterol and follow them for 10 years. **Table 1** shows the results of your study.

Table 1. Fictitious cohort study of elevated cholesterol and peripheral vascular disease: *Numbers are for demonstration only.*

CHOLESTEROL LEVEL	PERIPHERAL VASCULAR DISEASE	NO PERIPHERAL VASCULAR DISEASE	TOTAL
Elevated	375	2625	3000
Normal	75	2925	3000
TOTAL	450	5550	6000

Relative risk = (a/a+b)/(c/c+d)
Odds ratio = ([a/a+b]/[b/a+b])/([c/c+d]/[d/c+d]) = (a/b)/(c/d) = ad/bc

What is the risk of developing PVD if you have elevated cholesterol compared with the risk of developing PVD if you do not have elevated cholesterol levels? Comparing risk with and without exposure to a potentially causative agent is called RR. Risk of developing PVD if you have elevated cholesterol is 375/3000 =.125, or 12.5% (all numbers are fictitious and used for demonstration only). Risk of developing PVD if you do not have elevated cholesterol is 75/3000 =.025, or 2.5%. The ratio of these two numbers, .125/.025 = 5, is the RR. So, if you have elevated cholesterol, you are five times more likely to develop PVD than someone who does not have elevated cholesterol.

Another comparison, called the OR, is also used. The odds of developing PVD if you have elevated cholesterol are determined by comparing the proportion of people with elevated cholesterol who get PVD with the proportion of people with elevated cholesterol who do not get PVD (375/3000 =.125 compared with 2625/3000 =.875). Then .125/.875 =.143 are the odds of getting PVD if you have elevated cholesterol. If you similarly calculate the odds of getting PVD if you do not have elevated cholesterol, you get .026 (75/3000 =.025; 2925/3000 =.975; .025/.975 =.026). The OR (ratio of these two odds) is then .143/.026 = 5.5. Notice that the RR and the OR are very similar.

Look again at **Table 1** and the formulas. In 2×2 tables, disease or outcome are always displayed in the columns with presence of disease or outcome listed first and absence of disease or outcome listed second. Similarly, exposure or treatment is listed in the rows with presence of exposure listed first. This allows us to label the cells a, b, c, and d and to always be certain which data are in each cell. The formula for RR is straightforward and cannot be reduced. The formula for OR can be reduced from very cumbersome algebra to something much simpler. The long version is used in the calculations above.

A cohort study is better for determining an association between an exposure (eg, elevated cholesterol) and an outcome (eg, PVD) than a case-control study. Cohort studies are, however, more costly and take longer to complete than case-control studies. Suppose you had limited funds and you wanted to find out whether an association existed between these two things. You design a case-control study: you find 100 people with PVD (cases) and 300 similar people without PVD (controls). You then check their hospital or laboratory records to determine their cholesterol levels about 10 years ago (**Table 2**).

There is a problem: you cannot calculate a true RR because the results would be meaningless. You cannot calculate across the table because you have manipulated the actual number of people who did and did not have PVD. If you had chosen different numbers of cases and controls, the results would be different. The OR was nearly the same as the RR in the cohort study described above. Maybe you could use it to estimate the RR. The problem is we also calculated the OR across the table and we know we cannot do that in this situation. This is where a mathematical curiosity of the OR comes in to play. It turns out that if you calculate the OR up and down the table you get the same result as if you calculate it across the table. The OR formula in **Table 2** can be reduced to exactly the same as the OR formula in **Table 1** (OR = ad/bc however you look at it).

In **Table 2**, OR = (83×158)/(142×17) = 5.4, which is very close to the RR of 5 obtained in the cohort study. So we can do an inexpensive, fast, retrospective case-control study and, by using the OR, we can estimate what the RR would have been in an expensive, time-consuming, prospective cohort study.

Table 2. Fictitious case-control study of elevated cholesterol and peripheral vascular disease: *Numbers are for demonstration only.*

CHOLESTEROL LEVEL	PERIPHERAL VASCULAR DISEASE	NO PERIPHERAL VASCULAR DISEASE	TOTAL
Elevated	85	142	224
Normal	17	158	176
TOTAL	100	300	400

Odds ratio = ([a/a+c]/[c/a+c])/([b/b+d]/[d/b+d]) = (a/c)/(b/d) = ad/bc

There is a limitation, however. This neat mathematical feature of OR predicting RR is true only for diseases or outcomes that are relatively rare. You will notice that the prevalence of PVD in the cohort study, which is more likely to approximate true prevalence in the population, was 450 in 6000, or 7.5%. It seems the OR approximates the RR only for conditions with a prevalence of 10% or less. **Table 3** shows how the difference between RR and OR increases with increasing prevalence. At high prevalence rates, the OR overestimates the RR.

Table 3. Effect of changing prevalence on odds ratios and relative risks

PREVALENCE (%)	ODDS RATIO	RELATIVE RISK
1	5.6	5.6
7.5	5.5	5.0
10	5.6	4.9
25	5.5	3.8
50	5.5	2.6

Finally, in randomized controlled trials (RCTs), where the goal is usually to decrease occurrence of an event by treatment (eg, decrease mortality using cholesterol-lowering agents), the frequency of an event in the treatment group is lower than in the control group if the treatment is effective. This will give an RR of less than 1 with effective treatment. So an RR of 0.6 means that people taking treatment were only 60% as likely to die as those not taking treatment. The relationship between RR and OR is the same for RCTs as for cohort studies. Because RR can be measured directly in both these study designs, OR are not used. Only in case-control studies are OR used to estimate RR. (Odds ratios are also used in systematic reviews and when doing logistic regression, but that will be discussed in another article.)

Dr Godwin *is an Associate Professor and Director of Research in the Department of Family Medicine at Queen's University in Kingston, Ont, and is currently on sabbatical at the Centre for Evidence-Based Medicine in Oxford, England.*

Correspondence to: *Dr Marshall Godwin, Family Medicine Centre, 220 Bagot St, Kingston, ON K7L 5E9; e-mail* godwinm@post.queensu.ca

References

1. Norman GR, Streiner N. *PDQ statistics*. Philadelphia, Pa: B.C. Decker Inc; 1986.
2. Abramson JH. *Making sense of data: a self instruction manual on the interpretation of epidemiologic data*. 2nd ed. New York, NY: Oxford University Press; 1994.
3. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. 2nd ed. Toronto, Ont: Churchill Livingstone; 2000.