

Understanding the Odds Ratio and the Relative Risk

STEPHEN D. SIMON

From the Office of Medical Research, Children's Mercy Hospital, Kansas City, Missouri.

The simplest of all possible statistical problems ought to be exploring the relationship between binary variables. But binary variables are tricky. Binary, of course, means two possible levels. You might be interested in how a binary outcome variable such as live/dead, pregnant/not pregnant, diseased/healthy, etc, is related to another binary variable such as treatment/control or exposed/unexposed.

Two common measures you might see in such a situation are the odds ratio and the relative risk. For example, Bracken et al (1990) showed a strong relationship between cocaine usage in the last 2 years (yes/no) and sperm counts (above/below 20×10^6 mL) by reporting an odds ratio of 2.1. Lin et al (1996) showed a strong relationship between lead exposure (workers with at least 5 years of lead exposure/professional bus drivers) and fathering a child during the years 1981–1992 (yes/no) by reporting a relative risk of 0.38. What do these numbers mean, and why would you use one instead of the other?

Consider the following data on survival of passengers on the Titanic. There were 462 women: 308 survived and 154 died. There were 851 men: 142 survived and 709 died.

Clearly, a man on the Titanic was more likely to die than a woman. But how much more likely? You can compute either the odds ratio or the relative risk to answer this question.

The odds ratio compares the relative odds of death in each group. For women, the odds were exactly 2 to 1 against dying ($154/308 = 0.5$). For men, the odds were almost 5 to 1 in favor of death ($709/142 = 4.993$). The odds ratio is 9.986 ($4.993/0.5$). There is a 10-fold greater odds of death for men than for women.

The relative risk (sometimes called the risk ratio) compares the probability of death in each group rather than the odds. For women, the probability of death is 33% ($154/462 = 0.3333$). For men the probability is 83% ($709/851 = 0.8331$).

Andrology Lab Corner

851 = 0.8331). The relative risk of death is 2.5 ($0.8331/0.3333$). There was 2.5 times as much probability for death among the men than among the women.

Both measurements show that men were more likely to die. But the odds ratio implies that men were a lot worse off than the relative risk would imply. Which number is a fairer comparison?

There are three issues here. The relative risk measures events in a way that is interpretable and consistent with the way people really think. The relative risk, though, cannot always be computed in a research design. Also, the relative risk can sometimes lead to ambiguous and confusing situations. But first, it helps to remember the general properties of ratios.

Understanding Ratios

Ratios represent comparison by division rather than subtraction. So a ratio of 1.0 for either the odds ratio or the relative risk implies perfect similarity between the two groups. Deviations from 1.0 imply dissimilarity between the two groups. The more the ratio deviates from 1.0, the greater the dissimilarity.

It is important, however, to remember that ratios are skewed measures; deviations less than 1.0 are compressed, compared with deviations greater than 1.0. For example, a ratio of 0.80 on the low end and a ratio of 1.25 on the high end both represent a comparable amount of dissimilarity. This seems strange, but it helps to represent these ratios as fractions ($4/5$ and $5/4$, respectively). Similarly, a ratio of 0.67 ($2/3$) on the low end corresponds to a ratio of 1.5 ($3/2$) on the high end. Researchers will often use log scales in graphical displays of odds ratios and relative risks to correct for the inherent skewness in these measures.

When examining a confidence interval for an odds ratio or a relative risk, examine whether the interval includes the values of 1.0 and 2.0. This is a clear indication of an inadequate sample size. There is so little precision that both a null effect and a doubling of risk are consistent with the data. We can draw a similar conclusion about a confidence interval that includes the values of 0.5 and 1.0.

Interpretability

The relative risk comes closer to what most people think of when they think of the relative likelihood of events. That is partly because people are more used to thinking about probabilities such as 10% and 50% and are less comfortable with the corresponding odds (9 to 1 against

Correspondence to: Stephen D Simon, PhD, Research Biostatistician, Office of Medical Research, Children's Mercy Hospital, Room HHC-600, 2401 Gillham Road, Kansas City, MO 64108 (e-mail:ssimon@cmh.edu).

Received for publication February 15, 2001; accepted for publication February 15, 2001.

and 1 to 1 odds, respectively). But it goes further than that.

Suppose there are two groups. The first group has a 10% chance of mortality, equivalent to an odds of 0.1111 (or 9 to 1 odds against). Assume that the second group has a 50% chance of mortality, equivalent to an odds of 1.0 (even odds). Most people would say that the latter group has it 5 times as bad ($= 0.50/0.10$). But the odds ratio has a surprisingly larger value of 9 ($= 1.0/0.1111$). The odds ratio and the relative risk will not always disagree by this much. Large effects on groups with high initial risk seem to cause the most problems. See Davies et al (1998) for some useful guidelines for when the odds ratio and relative risk are likely to differ. When they do differ, the relative risk represents the typical interpretation that most people make.

There are some additional issues about interpretability that are beyond the scope of this paper. In particular, both the odds ratio and the relative risk are relative measures, computed by division. In contrast, absolute measures, computed by subtraction rather than division, produce estimates with quite different interpretations (Fahey et al, 1995; Naylor et al, 1992).

Designs That Rule Out the Use of the Relative Risk

Some research designs make it difficult to compute a relative risk. A case-control design is the most notable design of this type; it prevents us from directly computing a relative risk. Also, some designs require covariate adjustment. These adjustments are much easier for an odds ratio than a relative risk.

A case-control design involves the selection of research subjects on the basis of the outcome measurement rather than on the basis of the exposure. Consider a case-control study of prostate cancer risk and male-pattern balding (Lotufo et al, 2000). The goal of this research was to examine whether men with certain hair patterns are at greater risk of prostate cancer. In that study, roughly equal numbers of patients with prostate cancer and controls were selected. Among the cancer patients, 72 out of 129 had either vertex or frontal baldness compared with 82 out of 139 among the controls. No information on hair patterning was available for 47 patients.

In this type of study we can estimate the probability of balding for cancer patients, but we cannot calculate the probability of cancer for bald patients. The prevalence of prostate cancer was artificially inflated to almost 50% by the nature of the case-control design. So we would need additional information or a different type of research design to estimate the relative risk of prostate cancer for patients with different types of male-pattern balding.

Contrast this with data from a cohort study of 19 112 US male physicians (Lotufo et al, 2000). In this study of the association between male-pattern baldness and coro-

nary heart disease, the researchers could estimate relative risks, because 1446 physicians had coronary heart disease events during the 11-year follow-up period.

For example, among the 8159 doctors with no baldness, 548 (6.7%) developed coronary heart disease. Among the 1351 doctors with severe vertex balding, 127 (9.4%) developed coronary heart disease. The relative risk would be 1.4 ($= 0.094/0.067$).

We can always calculate and interpret the odds ratio in a case control study. It has a reasonable interpretation as long as the outcome event is rare (Breslow and Day, 1980). The interpretation of the odds ratio in a case-control design is also dependent on how the controls were recruited (Pearce, 1993).

Another situation that calls for the use of odds ratios is covariate adjustment. It is easy to adjust an odds ratio for confounding variables; the adjustments for a relative risk are much trickier.

A study on the likelihood of pregnancy among people with epilepsy (Schupf and Ottman, 1994) provides a good example of covariate adjustment. In this study, 232 out of 586 men with idiopathic/cryptogenic epilepsy had fathered 1 or more pregnancies. In the control group, the respective counts were 79 out of 109. The simple relative risk is 0.546, and the simple odds ratio is 0.249. Clearly, though, the probability of fathering a child is strongly dependent on a variety of demographic variables, especially age (the issue of marital status was addressed through a separate analysis). The control group was 8.4 years older on average (43.5 years vs 35.1). Without a covariate adjustment, we don't know whether part or all of the dissimilarity can be accounted for by the large age discrepancy. With a multivariate logistic regression model that included age (and also education, ethnicity, and sibling number), the adjusted odds ratio for epilepsy status was 0.36. Although this ratio was closer to 1.0 than the simple odds ratio, it was still highly significant. A comparable adjusted relative risk would be more difficult to compute.

Ambiguous and Confusing Situations

Relative risk can sometimes produce ambiguous and confusing situations. Consider an interesting case of relative comparison that comes from a puzzle on the "Car Talk" radio show (Magliozzi et al, 1999). We have a hundred pound sack of potatoes. Let's assume that these potatoes are 99% water. That means 99 parts water and 1 part potato. These are soggy potatoes than I am used to seeing, but it makes the problem more interesting.

If we dried out the potatoes completely, they would only weigh 1 pound. But let's suppose we wanted to dry out the potatoes only partially, until they were 98% water. How much would they weigh then?

The answer is surprising to most people. Ninety-eight

percent water means 49 parts water and 1 part potato. Each part weighs a pound, so the sack would weigh 50 pounds. An alternative way of thinking about the problem is that in order to double the concentration of potato (from 1% to 2%), we have to remove about half the water.

Relative risks have the same sort of counterintuitive behavior. A small relative change in the probability of a common event's occurrence can be associated with a large relative change in the opposite probability (the probability of the event not occurring). Consider a recent study on physician recommendations for patients with chest pain (Schulman et al, 1999). This study found that when doctors viewed a videotape of hypothetical patients, race and sex influenced their recommendations. One of the findings was that doctors were more likely to recommend cardiac catheterization for men than for women. Three hundred twenty-six out of 360 (90.6%) doctors viewing the videotape of hypothetical male patients recommended cardiac catheterization, whereas only 305 out of 360 (84.7%) of the doctors who viewed videotapes of hypothetical female patients made this recommendation. The authors reported an odds ratio of 0.6, which indicated that doctors viewing the hypothetical female patients were less likely to recommend catheterization.

A critique of this study (Schwartz et al, 1999) noted among other things that the odds ratio overstated the effect, and that the relative risk was a much more modest 0.93. In this study, however, it is not entirely clear that 0.93 is the appropriate risk ratio. Look at this from a different perspective. The rates for recommending a less-aggressive intervention than catheterization was 15.3% for doctors viewing the female patients and 9.4% for doctors viewing the male patients, a relative risk of 1.7. This is much more extreme than the relative risk of 0.93, and is about the same as the odds ratio for the less-aggressive intervention (1.7, which is simply the inverse of the previous odds ratio).

This is the same thing that we just saw in the "Car Talk" puzzler: a small relative change in the water content implies a large relative change in the potato content.

For every problem, there are 2 possible ways to compute relative risk. Sometimes it is obvious which relative risk is appropriate. For the Titanic passenger, the appropriate risk is for death rather than survival. But what about a study of breastfeeding? Are we trying to measure how much an intervention increases the probability of breastfeeding success or are we trying to find how much the intervention decreases the probability of breastfeeding failure? For example, Deeks (1998) expresses concern about an odds ratio calculation in a study aimed at increasing the duration of breastfeeding. At 3 months, 32 out of 51 (63%) of the mothers in the treatment group had stopped breastfeeding compared with 52 out of 57

Table. Comparison of the odds ratio and the relative risk

Odds Ratio	Relative Risk
Null value is 1.0	Null value is 1.0
Difficult to interpret	Easy to interpret
Available in all research designs	Not available in case-control designs
Covariate adjustment is easy	Covariate adjustment is hard
Invariant to labeling of events and nonevents	Labeling of events and non-events creates ambiguity.

(91%) in the control group. Although the relative risk for these data (0.69) is much less extreme than the odds ratio (0.16), one has to wonder why Deeks chose to compare probabilities of breastfeeding failures rather than successes. The rate of successful breastfeeding at 3 months was 4.2 times higher (= 37%/9%) in the treatment group than the control group. This is still not as extreme as the odds ratio; the odds ratio for successful breastfeeding is 6.25, the inverse of the odds ratio for breastfeeding failure. Still, it illustrates quite well that a relative risk close to 1.0 can become much more extreme when we redefine what we call an "event."

One advantage of the odds ratio is that it is invariant (not dependent on) whether we focus on the event's occurrence or its failure to occur. If the odds ratio for an event deviates substantially from 1.0, the odds ratio for the event's failure to occur will also deviate substantially from 1.0, although in the opposite direction.

Summary

Both the odds ratio and the relative risk compare the relative likelihood of an event occurring between two groups. The relative risk is easier to interpret and is consistent with general intuition. Some designs, however, allow only for the calculation of the odds ratio. Covariate adjustment is easier for an odds ratio. Finally, the odds ratio avoids ambiguity by being invariant to the labeling of the outcome measure.

The Table summarizes the advantages and disadvantages of the odds ratio and relative risk. When reading research that summarizes data using odds ratios, or relative risks, be aware of the limitations of both of these measures.

References

- Bracken MB, Eskenazi B, Sachse K, McSharry JE, Hellenbrand K, Leo-Summers L. Association of cocaine use with sperm concentration, motility, and morphology. *Fertil Steril.* 1990;53:315-322.
- Breslow NE, Day NE. *Statistical Methods in Cancer Research.* Vol 1. Lyon, France: International Agency for Research on Cancer; 1980.
- Davies HTO, Crombie IK, Tavakoli M. When can odds ratios mislead? *BMJ.* 1998;316:989-991.
- Deeks J. When can odds ratios mislead? Odds ratios should be used only

- in case-control studies and logistic regression analyses [letter]. *BMJ*. 1998;317:1155–1156.
- Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *BMJ*. 1995;311:1056–1059.
- Lin S, Hwang SA, Marshall EG, Stone R, Chen J. Fertility rates among lead workers and professional bus drivers: a comparative study. *Ann Epidemiol*. 1996;6:201–208.
- Lotufo PA, Chae CU, Ajani UA, Hennekens CH, Manson JE. Male pattern baldness and coronary heart disease. *Arch Intern Med*. 2000;160:165–171.
- Magliozzi T, Magliozzi R, Berman D. *A Haircut in Horse Town: And Other Great Car Talk Puzzlers*. New York, NY: Perigee; 1999.
- Naylor C, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trials results alter perceptions of therapeutic effectiveness? *Am Coll Phys*. 1992;117:916–921.
- Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol*. 1993;22:1189–1192.
- Schulman KA, Berlin JA, Harless W, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. *N Engl J Med*. 1999;340:618–626.
- Schupf N, Ottman R. Likelihood of pregnancy in individuals with idiopathic/cryptogenic epilepsy: social and biologic influences. *Epilepsia*. 1994;35:750–756.
- Schwartz L, Woloshin S, Welch H. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N Engl J Med*. 1999;341:279–283.