Practical Assessment, Research, and Evaluation

2006

Bringing balance and technical accuracy to reporting odds ratios and the results of logistic regression analyses

Jason W. Osborne

Follow this and additional works at: https://scholarworks.umass.edu/pare

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Bringing balance and technical accuracy to reporting odds ratios and the results of logistic regression analyses

Jason W. Osborne
North Carolina State University

Logistic regression and odds ratios (ORs) are powerful tools recently becoming more common in the social sciences. Yet few understand the technical challenges of correctly interpreting an odds ratio, and often it is done incorrectly in a variety of different ways. The goal of this brief note is to review the correct interpretation of the odds ratio, how to transform it into the more easily understood and intuitive relative risk (RRs) estimate, and a suggestion for dealing with odds ratios or relative risk estimates that are below 1.0 so that perceptually their magnitude is equivalent of Ors or RRs greater than 1.0.

Logistic regression is becoming more widely used in the social sciences as more texts (e.g., Pedhazur 1997) include chapters on the technique and more articles aimed at the social science researcher introduce the concept (e.g., Davis and Offord 1997; Peng, Lee et al. 2002). However, with more widespread adoption of the technique comes more opportunity for researchers to incorrectly interpret the results of this analysis. As Pedhazur (1997) and others (e.g., Davies, Crombie et al. 1998; Holcomb, Chaiworapongsa et al. 2001); have pointed out, correctly interpreting odds ratios for either a scientific or practitioner audience is particularly challenging, and often done incorrectly. For example, Holcomb et al (2001) reported that in a survey of high-quality medical journals over one-quarter of the articles explicitly mis-interpreted odds ratios. As the technique is newer to the social sciences, it is more likely that misinterpretation is happening in these literatures.

The goal of this methodological note is to briefly review the challenges to successfully and (more importantly) correctly interpreting the odds ratio (as compared to the more intuitive probability ratio or relative risk estimate), to highlight a simple way for transforming odds ratios to the more easily interpreted relative risk estimate, and to highlight a method of dealing with ORs and RRs that are less than 1.0 to bring them into perceptual balance with those mathematically identical (but perceptually different) ratios over 1.0.

What is an Odds Ratio (OR)?

The odds ratio has a long tradition in epidemiological and medical research where one is examining whether different factors contribute to disease (morbidity) or mortality. There are several ways to produce odds ratios, from hand calculations based on 2 x 2 contingency tables to (more commonly) logistic regression analyses. Logistic regression brings the general processes of ordinary

least squares regression (including multiple regression) to bear on dependent variables that are either categorical (yes/no outcomes, such as whether students have dropped out, individuals have become pregnant, voted, purchased a specific product, etc.) or discrete and categorical (e.g., choice of major, purchase of one of several products, and many behavioral outcomes such as educational attainment).

Yet with the advantages of logistic regression comes a challenge: interpreting the standardized coefficients, which are not betas but rather odds ratios (Exp (b)). To more concretely understand these odds ratio (and also relative risk, as well as the computational and conceptual differences between the two), refer to Table 1, which utilizes some sample data.

Table 1: Sample Data for Student Sex and Remedial Reading Classification			
	Not recommended (coded as 0)	Recommended (coded as 1)	Total
Boys	65	35	100
Girls	90	10	100
Total	155	45	200

In order to understand the difference between relative risk (probability ratios) and odds ratios, one can simply examine the data above. The *probability* that a student would be recommended to remedial reading is computed as the number recommended divided by the total possible, which equals 45/200 or 0.225. However the probability (or risk) of being recommended to remedial reading varies as a function of student sex. Specifically, the probability that a boy would be recommended is 35/100 (0.35) while the probability that a girl would be recommended is 10/100 (0.10). These are straightforward to interpret.

While probability uses group total as the denominator, the odds of a student being recommended is the number recommended divided by *the number not recommended*. Note that this is a very different denominator, especially when researching an outcome that is relatively common such as in this data. The odds of a student being recommended is 45/155 (0.29), but again that varies by student sex. The odds of a boy being recommended is 35/65 (0.54) while the odds of a girl being recommended is 10/90 (0.11).

Note that the difference between probability (risk) and odds is the denominator, which then influences their interpretation. Probability (risk) is interpreted in a straightforward manner: boys will be recommended to remedial reading about 35% of the time, on average, while girls are recommended about 10% of the time. Odds are less well understood by researchers, practitioners, and the lay public (Davies, Crombie et al. 1998; Holcomb, Chaiworapongsa et al. 2001). The odds of a boy being recommended to remedial reading is 0.54:1, while for girls it is 0.11:1. In other words, for each boy *not* recommended to remedial reading, 0.54 boys will be recommended. A similar interpretation would be offered for girls.

This brings up two important points: first, as authors such as Davies, Crombie, and Tavakoli (1998) point out and as this example demonstrated empirically, odds tend to inflate the effect size of an analysis (this is particularly true when events being studied are relatively common; see Davies et al. (1998) for more explication). Second, because of the different denominators between probabilities (risk) and odds, probabilities are relatively

straightforward to interpret, yet odds can be tricky. This will be discussed more below.

Ratios. Analyses rarely end with the calculations of odds or probabilities. Generally researchers want to calculate probability ratios (also called relative risk or RR) and /or odds ratios (OR). Using the data above, if one wants to know if boys are at greater risk of being recommended to remedial reading then girls, we can calculate a relative risk by dividing the probability for boys by the probability for girls (.35/.10) which yields a relative risk of 3.50. In other words, boys are 3.50 times more likely to be recommended for remedial reading than girls. This is intuitive, yet this statistic is *rarely the one reported in research*. Odds ratios are much more common, partly because many popular software packages readily report ORs. The odds ratio for these data is the odds for boys divided by the odds for girls (.54/.11) which yields an odds ratio of 4.91. In this case, the odds for boys are 4.91 that of girls.

However, that does not mean one can say that boys are 4.91 times as likely, or 4.91 times more likely to be recommended to remedial reading than girls. Technically, the *odds* of being assigned are 4.91 times greater for boys relative to girls. But since odds are tricky to understand, the meaning of this is less clear. Technically, it means that for every boy *not recommended* to remedial reading, 4.91 times as many boys will be recommended for remediation (0.54) than the number of girls recommended for every girl not recommomended.

Confused? You should be. Unless you work with odds and probabilities for a living, you should find relative risk (probability ratios) much easier to understand than odds ratios. Odds are not intuitive like probabilities are, and the language needed to technically describe an odds ratio is (as you see) can be quite convoluted.

The situation is not helped by authors' tendency to whitewash this important distinction and use probabilistic language when discussing odds ratios. Even highly sophisticated researchers will summarize odds ratios using language similar to: "boys are 4.91 times more likely to be recommended to remedial reading than girls" or "boys are 4.91 times as likely to be recommended…" when technically the odds ratio should be summarized as "the odds of boys being recommended are 4.91 times greater than the odds of girls being recommended" which does not address exactly what it means for odds to be greater in one group than another. Pedhazur (1997, pp. 760-761) takes great pains to highlight this common error, as do other authors (e.g., Davies, Crombie et al. 1998; Cohen 2000; Holcomb, Chaiworapongsa et al. 2001). Holcomb et al. (2001) report at 26% of authors in top-tier medical journals explicitly mis-interpreted ORs as RRs.

Why is this an issue? First, it is incorrect. While the OR and RR will be in the same direction (both will be either above 1.0 or below 1.0 if they are significant), ORs can illegitimately inflate the effect size substantially, as Davies et al. (1998) demonstrates. This effect is particularly egregious when the outcome being examined is not rare (e.g., occurs in more than 5% of the population) and becomes magnified as the RR moves away from 1.0-- commonly exceeding 80-90% inflation.

Two suggestions for fixing this issue. First, one can report odds ratios as long as an accurate interpretation of the OR is provided. However, as noted above, odds are non-intuitive. It is probably more effective to calculate relative risk directly, or, if that is not possible, calculate it from the following formula (presented in Davies et al., 1998; Holcomb et al. 2001; original work presented by Zhang & Yu (1998)).

$$RR = OR\ /\ [(1-P0)+ (P0\ x\ OR)] \qquad (1)$$

where RR=relative risk, OR= calculated odds ratio, and P0 = the proportion of non-exposed individuals that experience the outcome in question. In the case of our example, P0 would be .10, the probability that girls would be referred to remedial reading, and the OR is 4.91. Completing the calculations we end up with an estimated RR of 3.53, a very close approximation to the actual RR of 3.50.

How should we interpret odds ratios or relative risks less than 1.0?

One significant problem with RR and ORs is that they are asymmetrical. A value of 1.0 means there is no difference in risk or odds (i.e., there is no effect of the independent variable). Ratios less than 1.0 indicate that being in the exposed /selected group *decreases* the odds/risk of experiencing the outcome, whereas ratios greater than 1.0 indicate that being in the exposed/selected group *increases* the odds/risk of experiencing the outcome. The imbalance comes with the fact that increasing ratios are unbounded. They can vary from 1.0 to infinity, yet decreasing odds ratios are bounded by 0. They can range from 1.0 to 0 only, yet they encompass, technically, the same infinite range.

Two issues arise here. First, use of directional language such as "individuals in group 1 are X times *more likely* to experience a specific outcome than in group 2" or "individuals in group 2 are X times *less likely*…." Leaving for a moment the difficulty with cogently describing an odds ratio, the difficulty here comes in the common mistake people in describing decreasing ratios. If you have a RR of 3.50, as we did, it is straightforward to say "boys are 3.50 times more likely to be referred than girls." But what if we had coded the variables differently, so we were comparing girls to boys? With the same numbers we would have gotten a RR of 0.29 (rounded), meaning exactly the same thing-- that girls are much less likely than boys to be referred. Yet the careless author might be tempted to say "girls are 0.29 times less likely than boys" when in fact that is not the case. And furthermore, to say that girls are .71 times less likely fails to convey the same magnitude as a RR=3.50 although mathematically they are identical. My advice has always been to use "as

likely" rather than less or more likely. Saying "girls are 0.29 times as likely as boys" is more accurate and foolproof (providing you are discussing a RR rather than OR), and carries some more of the psychological gravity as saying "3.50 times as likely"… but not quite. Which brings us to the second issue: the psychological impact of ratios, and accurately conveying effect sizes when the effect sizes themselves vary depending on whether they are increasing or decreasing odds.

Taking a more extreme example, imagine a drug that made the risk of experiencing a cancer relapse RR= 0.001 compared to people who do not take the drug. Mathematically that is identical to saying that taking the drug makes you 1000 times less likely to experience relapse, or not taking the drug makes you 1000 times more likely to have a relapse. But are they perceptually identical? No.

Thus the final recommendation for those of you engaged in logistic regression and similar analyses using odds or probability ratios-- when possible, refrain from reporting RRs or ORs less than 1.0. It would make sense to standardize the reporting of this effect size so that all ratios be reported as >1.0. Analyses that result in ratios less than 1.0 would take the inverse of the RR or OR, and reverse the categories or the description of the results to keep the conclusion consistent.

Not only do ratios less than 1.0 have different psychological impact (despite being mathematically identical) but as Figure 1 shows, the relationship between ratios > 1.0 and their mathematically equivalent < 1.0 counterpoints is nonlinear. This is sub-optimal for an effect size, with RR and OR are, de facto, and a situation easily remedied in most cases.

Figure 1
The nonlinear relationship between increasing and decreasing ratios
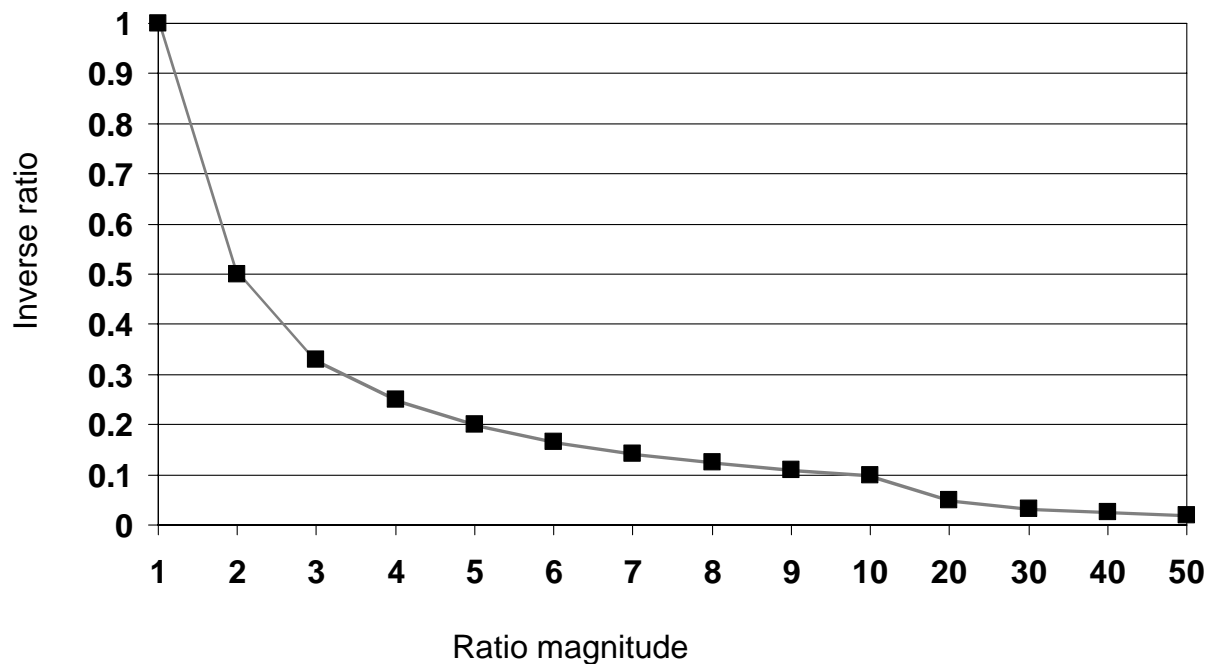
Summary

In sum, procedures such as logistic regression are powerful and useful tools to scientists. However, the commonly-reported odds ratio is difficult to understand conceptually, quite often mis-interpreted, and particularly difficult to disseminate to a lay/practitioner audience. Relative risk (probability ratios) are more intuitive and much easier to disseminate, so when possible researchers should report and interpret RRs rather than ORs. Secondly, ORs/RRs are relatively unique in the effect size world in that they are asymmetrical. Ratios below 1.0 behave very differently than ratios above 1.0 because they asymptote toward 0.0 whereas ratios above 1.0 are unbounded. More importantly, this creates asymmetry in perception of effect size, which is also undesirable. The second recommendation therefore is to convert all ratios < 1.0 to their corresponding ratio counterpart above 1.0 by taking the inverse of the RR or OR and adjusting the narrative accordingly.

These simple steps should increase the technical quality of reporting these analyses and standardize the metric of the effect size being used.

References

Burton, R. F. (2001) Quantifying the effects of chance in multiple choice and true/false tests: item selection and guessing of answers. *Assessment & Evaluation in Higher Education,* vol 26(1) 41-50.

Cohen, M. P. (2000). "Note on the Odds Ratio and the Probability Ratio." *Journal of Educational and Behavioral Statistics* 25(2): 249-252.

Davies, H. T. O., I. K. Crombie, et al. (1998). "When can odds ratios mislead?" *British Medical Journal* 316: 989-991.

Davis, L. J. and K. P. Offord (1997). "Logistic Regression." *Journal of Personality Assessment* 68(3): 497-507.

Holcomb, W. L., Jr., T. Chaiworapongsa, et al. (2001). "An odd measure of risk: use and misuse of the odds ratio." *Obstetrics and Gynecology* 84(4): 685-688.

Pedhazur, E. (1997). *Multiple Regression in behavioral research.* New York, Harcourt Brace College Publishers: 714-765.

Peng, C. J., K. Lee, L., et al. (2002). "An introduction to logistic regression analysis and reporting." *The Journal of Educational Research* 96(1): 3-14.

Zhang, J. and K. Yu (1998). "What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes." *Journal of the American Medical Association* 280: 1690-1.

Citation

Osborne, J.W. (2006). Bringing balance and technical accuracy to reporting odds ratios and the results of logistic regression analyses. *Practical Assessment Research & Evaluation*, 11(7). Available online: http://pareonline.net/getvn.asp?v=11&n=7

Author

Jason W. Osborne,
Educational Psychology,
Curriculum and Instruction
North Carolina State University
Poe Hall 602, Campus Box 7801
Raleigh NC 27695-7801
919-244-3538

Comments and correspondence should be directed via email to Jason W. Osborne at: jason_osborne@ncsu.edu.