What's the Relative Risk?: A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes

Jun Zhang; Kai F. Yu

JAMA. 1998;280(19):1690-1691 (doi:10.1001/jama.280.19.1690)

http://jama.ama-assn.org/cgi/content/full/280/19/1690

What's the Relative Risk?

A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes

Jun Zhang, MB, PhD; Kai F. Yu, PhD

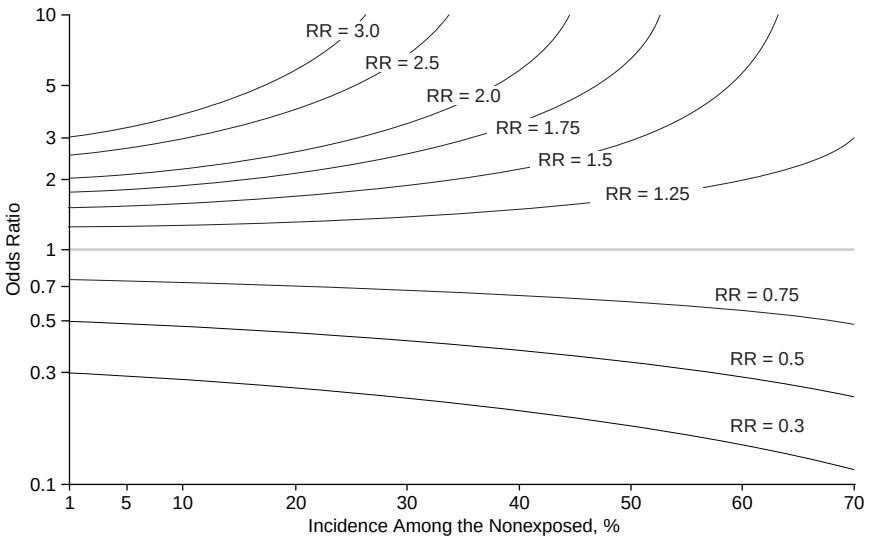
Logistic regression is used frequently in cohort studies and clinical trials. When the incidence of an outcome of interest is common in the study population (>10%), the adjusted odds ratio derived from the logistic regression can no longer approximate the risk ratio. The more frequent the outcome, the more the odds ratio overestimates the risk ratio when it is more than 1 or underestimates it when it is less than 1. We propose a simple method to approximate a risk ratio from the adjusted odds ratio and derive an estimate of an association or treatment effect that better represents the true relative risk.

JAMA. 1998;280:1690-1691

RELATIVE RISK has become one of the standard measures in biomedical research. It usually means the multiple of risk of the outcome in one group compared with another group and is expressed as the risk ratio in cohort studies and clinical trials. When the risk ratio cannot be obtained directly (such as in a case-control study), the odds ratio is calculated and often interpreted as if it were the risk ratio. Subsequently, the term *relative risk* commonly refers to either the risk ratio or the odds ratio. However, only under certain conditions does the odds ratio approximate the risk ratio. The Figure shows that when the incidence of an outcome of interest in the study population is low (<10%), the odds ratio is close to the risk ratio. However, the more frequent the outcome becomes, the more the odds ratio will overestimate the risk ratio when it is more than 1 or underestimate the risk ratio when it is less than 1.

Logistic regression is a widely used technique to adjust for confounders, not only in case-control studies but also in cohort studies.[1] However, logistic regression yields an odds ratio rather than a risk ratio, even in a cohort study. Under the same rule, when the outcome of interest is common in the study population (though it could be rare in the general population), the adjusted odds ratio from the logistic regression may exaggerate a risk association or a treatment effect. For instance, a previous study assessed the performance of neonatal units in Hospital A and Hospital B by comparing neonatal mortality in very low birthweight neonates between these 2 hospitals.[2] At first glance, Hospital A had a lower mortality rate than Hospital B (18% vs 24%, risk ratio, 18%:24% [0.75]). However, after adjusting for clinical variables and initial disease severity using logistic regression, the adjusted odds ratio of Hospital A vs Hospital B was 3.27 (95% confidence interval, 1.35-7.92). Can one therefore conclude that neonates with very low birthweight in Hospital A had 3 times the risk of death than those in Hospital B? Probably not, because the outcome (neonatal death) was common in this study population. To provide a measure that more accurately represents the concept of relative risk, correction of the odds ratio may be desirable.

A modified logistic regression with special macro functions has been devel-



The relationship between risk ratio (RR) and odds ratio by incidence of the outcome.

From the Division of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Md.

Corresponding author: Jun Zhang, MB, PhD, Division of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, Bldg 6100, Room 7B03, Bethesda, MD 20892-7510 (e-mail: Jun_Zhang@NIH.GOV).

Reprints not available from the authors.

Correcting the Odds Ratio in Cohort Studies of Common Outcomes in a Series of Simulated Cohorts*

True RR	Crude RR (95% CI)	Crude OR (95% CI)	Logistic OR (95% CI)	M-H RR (95% CI)	Corrected RR (95% CI)	P_0, %
7.4	6.5 (4.4-9.7)	9.4 (5.9-15.2)	14.1 (7.8-27.5)	8.0 (5.2-12.2)†	8.3 (5.4-11.4)	5
4.2	3.6 (2.8-4.8)	5.6 (3.9-8.1)	8.7 (5.5-14.3)	4.3 (3.2-5.8)†	4.6 (3.6-5.6)	12
3.0	2.9 (2.5-3.2)	25.0 (17.2-35.7)	27.4 (17.2-45.8)	3.0 (2.6-3.4)	2.9 (2.8-3.0)	32
2.0	0.93 (0.73-1.2)	0.90 (0.66-1.2)	4.5 (2.7-7.8)	2.0 (1.6-2.5)	2.3 (1.8-2.7)	27
0.37	0.34 (0.27-0.42)	0.23 (0.17-0.32)	0.25 (0.17-0.37)	0.37 (0.28-0.48)	0.36 (0.25-0.49)	40
0.14	0.13 (0.09-0.17)	0.08 (0.06-0.11)	0.09 (0.05-0.14)	0.14 (0.10-0.20)	0.14 (0.08-0.21)	40

*RR indicates risk ratio; CI, confidence interval; OR, odds ratio; logistic OR, odds ratio from logistic regression; M-H RR, risk ratio from Mantel-Haenszel estimate; P_0, incidence of outcome of interest in the nonexposed group; and corrected RR, risk ratio corrected by the above formula using logistic OR.

†Due to the sample sizes used in the simulation and the need to round numbers to integers, the M-H RR differs from the true RR.

oped to address this issue.[3] However, it is mathematically complex and uses a General Linear Interactive Modeling System (Numerical Algorithms Group, Oxford, England). Consequently, this method is rarely used. Another alternative is to use the Mantel-Haenszel method,[4] which can adjust for 1 or 2 confounders and still provide a risk ratio in a cohort study. However, this method becomes inefficient when several factors, especially continuous variables, are being adjusted for simultaneously. We herein propose an easy approximation with a simple formula that can be applied not only in binary analysis[5] but also in multivariate analysis.

In a cohort study, P_0 indicates the incidence of the outcome of interest in the nonexposed group and P_1 in the exposed group; OR, odds ratio; and RR, risk ratio: $OR = (P_1/1 - P_1)/(P_0/1 - P_0)$; thus, $(P_1/P_0) = OR/[(1 - P_0) + (P_0 \times OR)]$. Since $RR = P_1/P_0$, the corrected

$$RR = \frac{OR}{(1 - P_0) + (P_o \times OR)}.$$

We can use this formula to correct the adjusted odds ratio obtained from logistic regression and derive an estimate of an association or treatment effect that better represents the true relative risk. It can also be used to correct the lower and upper limits of the confidence interval by applying this formula to the lower and upper confidence limits of the adjusted odds ratio. In the above example, after the odds ratio is corrected (where $OR = 3.27$ and $P_0 = 0.24$), the risk ratio becomes 2.12 (95% confidence interval, 1.25-2.98), ie, very low birthweight neonates in Hospital A had twice the risk of neonatal death than those in Hospital B.

To examine the validity of this correction method in various scenarios, we simulated a series of hypothetical cohorts based on predetermined risk ratios (called true RR). Each cohort consists of 1000 subjects with 1 binary outcome (0,1), 1 exposure variable (0,1), and 2 confounders. Both confounders have 3 levels (1,2,3). The true risk ratio is kept constant across strata of the confounders. As expected, with an increase in incidence of outcome and risk ratio, the discrepancy between risk ratio and odds ratio increases (Table). The corrected risk ratio, which is calculated based on the odds ratio from logistic regression after having adjusted for the confounders, is very close to the true risk ratio. This procedure can be applied to both unmatched and matched cohort studies. It can further be used in cross-sectional studies, in which the prevalence ratio rather than the risk ratio will be generated. It enables us to obtain a corrected prevalence ratio very close to the one obtained from a complex statistical model[6] (data not shown).

Due to the differences in underlying assumptions between Mantel-Haenszel risk ratio and logistic regression odds ratio, some discrepancy between the Mantel-Haenszel risk ratio and the corrected risk ratio is expected (detailed discussion of which is beyond the scope of this work). More importantly, the validity of the corrected risk ratio relies entirely on the appropriateness of logistic regression model, ie, only when logistic regression yields an appropriate odds ratio will the correction procedure provide a better estimate. Therefore, in a cohort study, whenever feasible, the Mantel-Haenszel estimate should be used.

In summary, in a cohort study, if the incidence of outcome is more than 10% and the odds ratio is more than 2.5 or less than 0.5, correction of the odds ratio may be desirable to more appropriately interpret the magnitude of an association.

References

1. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: John Wiley & Sons Inc; 1989.
2. Tarnow-Mordi W, Ogston S, Wilkinson AR, et al. Predicting death from initial disease severity in very low birthweight infants: a method for comparing the performance of neonatal units. *BMJ*. 1990;300:1611-1614.
3. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol*. 1986;123:174-184.
4. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22:719-748.
5. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol*. 1994;47:881-889.
6. Lee J. Odds ratio or relative risk for cross-sectional data? *Int J Epidemiol*. 1994;23:201-202.