The Korean Communications
in Statistics Vol. 11 No. 1, 2004
pp. 181-187

Comparison Of Interval Estimation For Relative Risk Ratio With Rare Events[1)]

Yong Dai Kim[2)], Jin-Kyung Park

Abstract

One of objectives in epidemiologic studies is to detect the amount of change caused by a specific risk factor. Risk ratio is one of the most useful measurements in epidemiology. When we perform the inference for this measurement with rare events, the standard approach based on the normal approximation may fail, in particular when there are no disease cases observed. In this paper, we discuss and evaluate several existing methods for constructing a confidence interval of risk ratio through simulation when the disease of interest is a rare event. The results in this paper provide guidance with how to construct interval estimates for risk difference and risk ratio when there are no disease cases observed.[3)]

Keywords: Risk ratio, Bayesian probability interval, Rare events

1. Introduction

In epidemiologic research the calculation of appropriate measures of disease frequency is the basis for the comparison of populations and, therefore, the identification of disease determinants. To do this most efficiently and informatively, the two frequencies being compared can be combined into a single summary parameter that estimates the association between the exposure and the risk of developing a disease. This can be accomplished by calculating risk ratio as well as risk difference. This measure is a useful measurement to detect the amount of change caused by a specific risk factor. Risk ratio or relative risk is the ratio of the incidence of disease in the exposed group divided by the corresponding incidence of disease in the nonexposed group. Risk ratio is the measure used most commonly by those evaluating possible determinant of disease because it represents the magnitude of the association and provides information that can be used in making a judgement of casuality.

---

1) This research was done while the first author was in Ewha Womans Unversity. The first author is supported in part by 2001 Ewha Womans University research fund, and the second author is supported in part by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University

2) Professor, Department of Statistics, Seoul National Universityl, Seoul, Korea
E-mail : ydkim@stats.snu.ac.kr

Thus, the risk ratio is valuable in etiologic studies of diseases.

Suppose x_1 and x_2 are disease frequencies of two independent populations with sizes n_1 and n_2 respectively. The risk ratio is defined by p_1/p_2 where p_1 and p_2 are the probabilities of disease in the two populations. The standard method of constructing confidence interval of the risk ratio is based on the normal approximation, which provides the interval estimate:

$$\exp[(\log(\widehat{p}_1) - \log(\widehat{p_{2)})}\pm z_{\alpha/2}\sqrt{\frac{1-\widehat{p}_1}{n_1\widehat{p}_1}+\frac{1-\widehat{p}_2}{n_2\widehat{p}_2}}] \tag{1.1}$$

where $\widehat{p}_1 = x_1/n_1$ and $\widehat{p}_2 = x_2/n_2$. This interval estimate is a standard one in most statistical software packages so they are used by non-statisticians widely. Along with the computational simplicity, this estimate has the apparent advantage of producing intercal centered on the point estimate, thus resembling those for the mean of a continuous Normal variate.

However, when $x_1 = x_2 = 0$, we have a problem with using the interval estimate (1.1). The confidence interval (1.1) is not defined. The situation in which no cases occur in a binomial experiment arises quite frequently when p_1 and p_2 are small. Examples are an epidemiologic study where disease of interest is a rare event and a diagnostic test in which it is common to deal with a small false negative rate (the probability of a disease individual testing negative).

This paper is concerned with the interval estimates of the risk ratio when the probabilities of cases are small. We consider four interval estimates including (1.1) and compare their performances such as coverage probabilities and interval lengths by simulation. The results in this paper provide guidance with how to construct interval estimates for risk ratio, in particular when there are no disease cases observed.

Various interval estimates of risk ratio have been proposed by Neother(1957), Walter(1975), Katz et al.(1978), Aitchison and Bacon-Shone(1981), Koopman(1984), Mee(1984), Miettinen and Nurminrn(1985), Gart and Nam(1998) and Ewell(1996). All of these studies, however, have focused on improving the small and moderate sample performace of the interval estimates. To our knowledge, our simulation experiment is the first comparative studies for interval estimates of the risk ratio with small probabilities of cases.

This paper is organized as follows. In section 2, we describe several interval estimates for the risk ratio. In section 3, simulation results are provided and discussions follow in section 4.

2. Various interval estimates of Risk Ratio

Gart and Nam(1988) grouped several interval estimates inte three subsets based on their mode of derivation. The interval estimate (1.1) is the one and the other two are the Fieller-like method and likelihood method. In this subsection, we describe briefly these two methods as well as the Baysian method.

1.Fieller-like method(BF)

Let $\phi=p_1/p_2$. The Fieller-like interval uses the statistic $T=\hat{\phi}-\phi$ where $\hat{\phi}=\hat{p}_1/\hat{p}_2$. It can be shown that T is asymptotically normal with variance

$$var(T)=\frac{\phi^2 q_2}{n_2 p_2}+\frac{\phi^2 q_1}{n_1 p_1}.$$

p_1 in this variance formula is substituted by ϕp_2 and p_2 is estimated by $\hat{p}_2$. Finally, the 100(1- α)% confidence interval using this estimated variance V(T) is given as the solution of following quardratic equation.

$$(\hat{\phi}-\phi)^2=z_\alpha^2 V(T) \tag{2.1}$$

The above Fieller-like interval estimate is proposed by Neother (1975). Different Fieller-like interval estimates for risk ratio based on the statistics $T'=\hat{p}_1-\phi\hat{p}_2$ have been proposed by Katz et al. (1978) and Bailey (1987).

2. Methods Based on Likelihood Methods(BL)

Miettinen and Nurminen(1985) have proposed a method that use a maximum likelihood estimator(MLE) of the nuisance parameter, p_2, for given values of $\phi=p_1/p_2$. Let $\tilde{p}_2$ be the MLE of p_2 given ϕ and let $\tilde{p}_1=\phi\tilde{p}_2$. They started with the statistic $T=\hat{p}_1-\phi\hat{p}_2$and estimated its variance by $\tilde{p}_1\tilde{q}_1/n_1+\phi^2\tilde{p}_2\tilde{q}_2/n_2$. Then, they obtained the limits of 100(1- α)% confidence interval as the roots to the equation

$$\chi^2_{MN}(\phi)=\frac{(\hat{p}_1-\phi\hat{p}_2)^2}{(\tilde{p}_1\tilde{q}_1)/n_1+(\phi^2\tilde{p}_2\tilde{q}_2)/n_2}=z^2_{\alpha/2} \tag{2.2}$$

Koopman(1984) derived the same confidence limits with different derivations.

3. Bayesian Method with Binomial likelihood (BB)

The Bayesian probability interval of the risk ratio is constructed as follows. Let $\pi(p_1, p_2)$ be the prior distribution of (p_1, p_2). Then the posterior distribution is given by

$$\pi(p_1, p_2 \mid x_1, x_2) \propto p_1^{x_1}(1-p1)^{n_2-x_2} p_2^{x_2}(1-p_2)^{n_2-x_2}\pi(p_1, p_2)$$

Let $\theta=p_1/p_2$ and $\psi=p_2$, and let $\pi(\theta,\psi \mid x_1,x_2)$ be the corresponding posterior distribution of θ and ψ, which can be obtained by use of the variable transfomation technique. Now, the equal tail 100(1- α)% probability interval has the form of (L,U) which satisfies

$$\int_U^\infty \int_0^1 \pi(\theta,\psi \mid X_1, X_2)d\psi d\theta=\alpha/2$$

and

$$\int_0^L \int_0^1 \pi(\theta,\psi \mid X_1, X_2)d\psi d\theta=\alpha/2$$

See Gelman et al(1995). In practice, we can obtain L and U by using a simple Monte-Carlo method as follows. First, we calculate the posterior distribution of (p_1, p_2) and obtain the posterior distribution of p_1/p_2 by use of the variable transformation technique. Finally, we get the limits of the equal tail $100(1-\alpha)\%$ probability interval as the upper and low $100(\alpha/2)$ percentiles of the posterior distribution of p_1/p_2. These percentiles can be easily calculated using the simple Monte-Carlo method

3. Simulations

In this section, we present the simulation results for comparing the performance of the aforementioned interval estimates for the risk ratio when the probabilities of the disease case are small. In the Bayesian methods, independent uniform distributions are used as prior distributions for p_1 and p_2 to represent prior ignorance. Then, a posterior p_1 and p_2 are independent beta distributions with parameters (x_1+1, n_1-x_1+1) and (x_2+1, n_2-x_2+1) respectively.

We assume that the sample sizes of the two populations are the same (i.e. $n_1=n_2$). We set the true risk difference p_1/p_2 at 0.5 and 1.0, and investigate the coverage probabilities and interval lengths as we change the value of p_2 from 0.001 to 0.1. We add 0.5 when a generated datum shows zero.

Figure 3.1 draws that coverage probabilities of the four 95% interval estimates - the interval estimate based on the normal approximation (BZ), Fieller-like method (BF), likelihood method (BL) and Bayesian method (BB) when the sample sizes are 50 and 200.

Apparently, the all four methods perform competitively well. The interval estimate based in the normal approximation performs reasonably well. This is partly because we add 0.5 when a generated datum shows zero. Table I and Table II present the values of the coverage probabilities and interval lengths. The Fieller-like method gives the shortest interval lengths with proper coverage probabilities. Thus, we recommend using the Fieller-like method when the probability of the disease case is expected to be small.

4. Discussion

We performed simulation for evaluating various interval estimates for risk ratio when the probabilities of the disease case is small. Our simulation results showed that all four methods method performs competitively well. In particular, the performance of the Fieller-like interval estimates is superior comapred to the other three interval estimates. From the computational point of view, the Bayesian method is most appealing since calculation of the Bayesian probability intervals only requires random number generation which can be done easily by use of standard softwares such as SAS or Splus and the interval estimates of the risk ratio can be constructed simultaneously using the same generated random numbers.

REFERENCES

[1] Aitchison, J. and Bason-Shone, J.(1981). Bayesian Risk Ratio Analysis. *The Amerian Statistician* 35 254-257

[2] Ewell, M.(1998). Comparison Methods for Calculating Confidence intervals for Vaccine Efficiency. *Statistics in Medicine* 15 2379-2392

[3] Gart, J.J., and Nam, J.M.(1998). Approximate interval estimation of the ratio of binomial parameters : A review and corrections for skewness. *Biometrics* 44, 323-338

[4] Katz, D., Baptista, J., Azen, S.P., and Pike,M.C.(1978). Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies. *Biometrics* 34 469-474

[5] Koopman, P.A.R.(1984). Confidence limits for the ratio of two binomial proportions. *Biometics* 40, 513-517

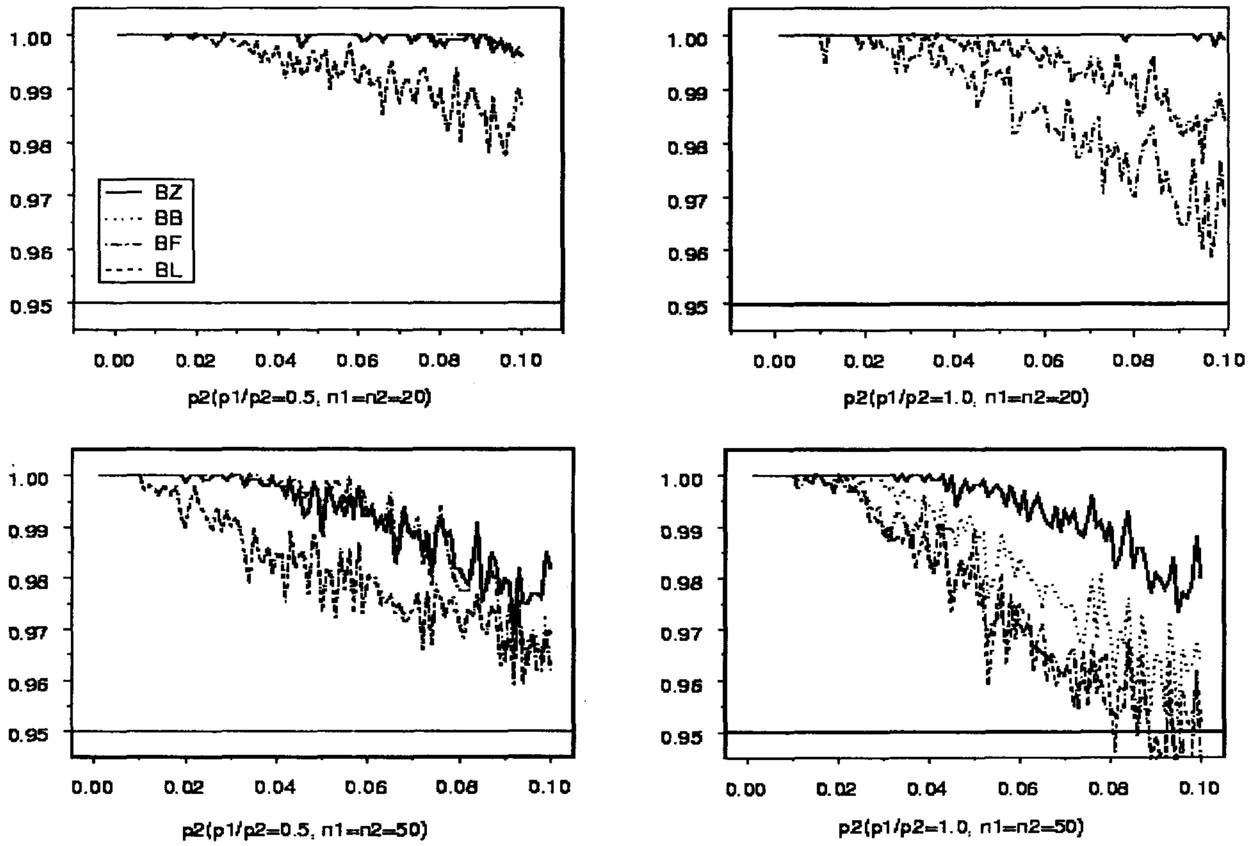
[6] Mee, R. W.(1984). Confidence bounds for the difference between two probabilities, *Biometrics*, 40, 1175-1176

[7] Miettinen, O.S., and M.Nurminen.(1985). Comparative analysis of two rates. *Statist. Med* 4, 213-226

[8] Neother, G.E.(1957). Two confidence intervals for the ratio of two probabilities and some measures of effectiveness. *Journal of the American Statistics Association* 52,36-42

[9] Walter, S.D.(1975). The distribution of Levin's measure of attributable risk. *Biometrika* 62, 371-375

[Received February 2004, Accepted April 2004]



3.1: Results of risk ratio

Comparison of two proportions

Table Ⅰ. Coverage Probabilities of Risk Ratio Interval Estimates

n		$p_1/p_2 = 0.5$				$p_1/p_2 = 0.5$			
$(n_1 = n_2)$	p_2	BZ	BB	BF	BL	BZ	BB	BF	BL
20	0.001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.010	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.040	1.000	0.998	1.000	0.998	1.000	0.999	0.994	0.999
	0.060	1.000	0.991	1.000	0.991	1.000	0.998	0.986	0.998
	0.080	0.999	0.990	1.000	0.990	1.000	0.993	0.969	0.993
	0.100	0.996	0.988	0.997	0.987	0.999	0.984	0.967	0.984
50	0.001	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.010	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.020	0.999	0.990	0.998	0.990	1.000	1.000	1.000	0.999
	0.040	0.998	0.985	0.997	0.984	1.000	0.994	0.983	0.983
	0.060	0.992	0.981	0.992	0.980	0.993	0.984	0.970	0.975
	0.080	0.982	0.972	0.977	0.971	0.991	0.969	0.954	0.956
	0.100	0.982	0.973	0.961	0.969	0.980	0.963	0.940	0.952

Table Ⅱ. Average Lengths of Risk Ratio Interval Estimates

n		$p_1/p_2 = 0.5$				$p_1/p_2 = 0.5$			
$(n_1 = n_2)$	p_2	BZ	BB	BF	BL	BZ	BB	BF	BL
20	0.001	47.595	14.424	3.973	16.820	47.682	14.432	3.973	16.910
	0.010	42.247	13.358	3.855	15.729	43.280	13.866	3.993	16.734
	0.020	38.609	12.678	3.652	15.045	40.791	13.679	3.852	17.031
	0.040	29.578	10.718	3.653	12.750	33.072	12.293	4.047	15.721
	0.060	22.394	9.031	3.428	10.691	26.867	11.165	4.007	14.516
	0.080	17.433	7.499	2.973	8.935	22.008	9.717	3.602	12.787
	0.100	13.468	6.390	2.849	7.377	17.381	8.505	3.552	10.794
50	0.001	47.981	14.763	4.473	16.826	48.236	14.903	4.493	17.070
	0.010	36.639	12.466	4.135	14.508	39.358	13.685	4.418	16.825
	0.020	28.277	10.672	3.898	12.497	32.592	12.616	4.453	16.068
	0.040	15.624	7.223	3.251	8.296	20.340	9.604	4.092	12.219
	0.060	8.570	4.926	2.651	5.467	12.512	7.237	3.565	8.783
	0.080	5.014	3.412	2.105	3.606	7.283	5.116	2.922	5.746
	0.100	3.759	2.691	1.699	2.833	5.719	4.132	2.369	4.624