KATZ 78
SI STATS

BIOMETRICS 34, 469–474
September, 1978

# *Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies*

D. KATZ, J. BAPTISTA, S. P. AZEN[1] and M. C. PIKE

Departments of Community and Family Medicine and Biomedical Engineering,
University of Southern California, 2025 Zonal Avenue,
Los Angeles, California 90033, U.S.A.

### *Summary*

*Three methods of obtaining confidence intervals for the risk ratio of failure in two independent binomial samples are compared. The three are (A) the method of Thomas and Gart (1977), (B) an adaptation of the method of Fieller using the normal distribution, and (C) a proposed method using a logarithmic transformation. On the basis of extensive simulations we have concluded that Method A is reasonable but conservative, Method B is erratic and should not be used, and Method C is reasonable and less conservative than Method A. Method C, computationally the simplest, is recommended.*

## *1. Introduction*

In a two-group cohort study a parameter of major interest is the ratio $R$, sometimes termed the *risk ratio*, of the probabilities of failure in the two groups (see, for example, Fleiss 1973). We assume that we have data from two independent binomial populations: $x \sim Bnl(N_t, p_t)$ and $y \sim Bnl(N_c, p_c)$ where $p_t$ and $p_c$ are the respective failure probabilities, and $R = p_t/p_c$. The data may be arranged in $2 \times 2$ table form as:

$$\begin{array}{ccc} x & n_t - x & N_t \\ y & N_c - y & N_c \\ m & N - m & N \end{array}$$

The most commonly used estimate of $R$ is $\hat{R} = \hat{p}_t/\hat{p}_c$, where $\hat{p}_t = x/N_t$ and $\hat{p}_c = y/N_c$.

In this paper we compare by simulation three methods of obtaining approximate confidence limits for $R$. Since many examples of cohort studies are characterized by relatively small $p$'s and reasonably large sample sizes, this situation is emphasized in the simulation studies.

## *2. Approximate Methods*

We consider different methods of calculating a lower $(1 - \alpha)$ confidence limit $R_L$ for $R$. Upper $(1 - \alpha)$ confidence limits for $R$ may be derived in a similar fashion. In all methods $N_t$ and $N_c$ are assumed to be fixed.

No method is known for calculating $R_L$ such that

[1] All correspondence and reprint requests should be directed to Dr. S. P. Azen, Department of Community and Family Medicine, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, U.S.A.

*Key Words:* Risk ratio; Relative risk; Cohort study; Confidence intervals.

$$Pr(R_L \leq R) \geq 1 - \alpha \tag{1}$$

for all appropriate $p_t$ and $p_c$ and with equality for some $p_t$ and $p_c$. Nor is a method known with this property when the sample results are conditioned on fixed $m$. However, at least three essentially different approaches to calculating $R_L$ to achieve approximate equality in (1) for all appropriate $p_t$ and $p_c$ are possible. The first (Method A) considers the odds ratio, the second (Method B) the variable $(\hat{p}_t - R\hat{p}_c)$, and the third (Method C) the variable $\log(\hat{p}_t/\hat{p}_c)$. Details are as follows:

*Method A:* The probability distribution of $x$ conditioned on fixed $m$ depends on $p_t$ and $p_c$ only through $\psi = [p_t(1 - p_c)]/[p_c(1 - p_t)]$, the so-called odds ratio, and may be written as

$$f(x|m) = [C(N_t, x)C(N_c, m - x)\,\psi^x] \Big/ \left[\sum_{i=0}^{m} C(N_t, i)C(N_c, m - i)\,\psi^i\right].$$

A lower $(1 - \alpha)$ confidence limit, $\psi_L$, for $\psi$ is the solution of

$$\sum_{j=x}^{m} f(j|m) = \alpha.$$

Given this $\psi_L$, Thomas and Gart (1977) suggested that one first solve

$$\psi_L = \{x_L(x_L + N_c - m)\}/\{(m - x_L)(N_t - x_L)\} \tag{2}$$

for $x_L$ $(0 \leq x_L \leq \min(N_t, m))$, and then set

$$R_L = (x_L/N_t)/\{(m - x_L)/N_c\}. \tag{3}$$

They claimed that this procedure for calculating $R_L$ was exact in the sense that $Pr(R_L \leq R|m) \geq 1 - \alpha$ for all $R$ compatible with $\psi$.

Table 1 shows, however, that this method only produces approximate confidence intervals. The table gives for $N_t = N_c = 20$ (the smallest sample size considered by Thomas and Gart), $m = 24$, $1 - \alpha = 0.8$, and $\psi = 0.125$ the exact probability that $R_L \leq R$ for different values of $R$. The probability that $R_L \leq R$ increases steadily with $E(m) = N_tp_t + N_cp_c$, the expected value of $m$, and is equal to the nominal probability (*viz.* 0.808) only when $m \simeq E(m)$.

*Method B:* The variate $\hat{p}_t - R\hat{p}_c$ is approximately normally distributed with zero mean and approximate variance $\hat{p}_t\hat{q}_t/N_t + R^2\hat{p}_c\hat{q}_c/N_c$, where $\hat{q}_t = 1 - \hat{p}_t$ and $\hat{q}_c = 1 - \hat{p}_c$. An approximate $R_L$ may be obtained by solving

$$(\hat{p}_t - R_L\hat{p}_c)/[\hat{p}_t\hat{q}_t/N_t + R_L{}^2\hat{p}_c\hat{q}_c/N_c]^{1/2} = z_{1-\alpha}\,,$$

where $z_{1-\alpha}$ is the $100(1 - \alpha)$ percentage point of the $N(0, 1)$ distribution (Fieller 1944). This method can lead to negative or complex values for $R_L$.

*Method C:* The variate $\log(\hat{p}_t/\hat{p}_c) = \log(\hat{p}_t) - \log(\hat{p}_c)$ is approximately normally distributed with approximate mean $\log(R)$ and estimated variance $(1 - \hat{p}_t)/x + (1 - \hat{p}_c)/y$. Thus, a third approximate $R_L$ may be obtained by solving

$$[\log(\hat{p}_t/\hat{p}_c) - \log(R_L)]/[(1 - \hat{p}_t)/x + (1 - \hat{p}_c)/y]^{1/2} = z_{1-\alpha}\,.$$

### 3. Simulation Study

For each of the three methods the same 5000 independent pairs of samples were generated using the pseudo-random number generator of Pike and Hill (1966) with a refinement suggested by Hill in Atkinson and Pearce (1976).

In all three methods $R_L$ was taken to be zero when $m = 0$ or $x = 0$. In addition, in

A caso $x = A$

C risk $1-p = \dfrac{C-A}{C}$ $\qquad \dfrac{C-A}{AC} = \dfrac{1}{A} - \dfrac{1}{C}$

Me
by .
sub
esti
whi
me:
in f
$N_t$
anc
0.0
obt
Me
typ
$R$ ;
ser
coı
illu

TABLE 1

Probabilities of R Being Greater Than or Equal to $R_L$ as Calculated by Method A: $\alpha = 0.2, \psi = 0.125, P\cdot(\psi \geq \psi_L) = 0.808, m = 24, N_t = N_c = 20$

| $p_t$ | $p_c$ | R | $Pr(R_L \leq R)$ | $E(m) = N_t p_t + N_c p_c$ |
|---|---|---|---|---|
| 0.029 | 0.191 | 0.15 | 0.008 | 4.4 |
| 0.086 | 0.429 | 0.20 | 0.008 | 10.3 |
| 0.143 | 0.571 | 0.25 | 0.073 | 14.3 |
| 0.200 | 0.667 | 0.30 | 0.267 | 17.3 |
| 0.257 | 0.735 | 0.35 | 0.557 | 19.8 |
| 0.314 | 0.785 | 0.40 | 0.808 | 22.0 |
| 0.371 | 0.825 | 0.45 | 0.808 | 23.9 |
| 0.429 | 0.857 | 0.50 | 0.942 | 25.7 |
| 0.486 | 0.883 | 0.55 | 0.988 | 27.4 |
| 0.543 | 0.905 | 0.60 | 0.988 | 29.0 |
| 0.600 | 0.921 | 0.65 | 0.998 | 30.1 |
| 0.657 | 0.937 | 0.70 | 0.998 | 31.6 |
| 0.714 | 0.952 | 0.75 | 1.000 | 33.3 |
| 0.771 | 0.964 | 0.80 | 1.000 | 34.7 |
| 0.829 | 0.975 | 0.85 | 1.000 | 36.1 |
| 0.886 | 0.984 | 0.90 | 1.000 | 37.4 |
| 0.943 | 0.992 | 0.95 | 1.000 | 38.7 |

Methods B and C, $y = 0$ was replaced by $y = 0.5$, and in Method B negative $R_L$ was replaced by $R_L = 0$. If Method B gave complex roots or a disjoint interval, then Method C was substituted.

The methods were compared in terms of (i) the proportion of times that $R \geq R_L$, i.e., an estimate of the one-sided confidence coefficient, (ii) the mean distance below $R$ of all $R_L$ which were less than $R$, in order to assess the conservativeness of the method, and (iii) the mean distance above $R$ of all $R_L$ which exceeded $R$, in order to assess the magnitude of error in failing to cover a true $R$.

Cases which were considered were $\alpha = 0.025, 0.05$ and $0.1$; with selected combinations of $N_t = N_c = 25, 50, 100, 200$ and $400$; $R = 0.25, 0.5, 0.667, 1, 1.5, 2$ and $4$; and $N_c p_c = 5, 10, 20$ and $40$. Some combinations such as $N_c = 100$, $N_c p_c = 40$ and $R = 4$ are impossible.

## 4. *Results*

The estimated one-sided confidence coefficients for the three methods for the cases $\alpha = 0.025$, $N_t = N_c = 100$, $N_c p_c = 10, 20$ and $40$ are shown in Table 2. Similar results were obtained for Methods A and C but Method A tended to be slightly more conservative. Method B was erratic and failed to produce a meaningful $R_L$ 0.7% of the time. These were typically cases where $x$ was small. The other two measures of the mean distances from $R_L$ to $R$ provided no additional information for choosing among the methods and are not presented.

Examination of the remaining cases revealed that when $N_c p_c$ is small, Method A is considerably more conservative than Method C and Method B is even more erratic than illustrated by the results in Table 2.

TABLE 2
Simulation Results for Lower 97½% Confidence Limits for All Methods: $N_t = N_c = 100$.

| R | $N_c p_c$ | Method A | Method B | Method C |
|---|---|---|---|---|
| 0.25 | 10 | 0.988 | 0.994 | 0.969 |
| | 20 | 0.982 | 0.992 | 0.968 |
| | 40 | 0.982 | 0.987 | 0.966 |
| 0.5 | 10 | 0.986 | 0.985 | 0.976 |
| | 20 | 0.983 | 0.982 | 0.973 |
| | 40 | 0.982 | 0.981 | 0.970 |
| 0.667 | 10 | 0.988 | 0.987 | 0.978 |
| | 20 | 0.982 | 0.978 | 0.974 |
| | 40 | 0.981 | 0.978 | 0.973 |
| 1.0 | 10 | 0.985 | 0.973 | 0.979 |
| | 20 | 0.982 | 0.974 | 0.978 |
| | 40 | 0.981 | 0.972 | 0.972 |
| 1.5 | 10 | 0.984 | 0.968 | 0.984 |
| | 20 | 0.982 | 0.968 | 0.978 |
| | 40 | 0.980 | 0.970 | 0.979 |
| 2 | 10 | 0.985 | 0.960 | 0.985 |
| | 20 | 0.981 | 0.965 | 0.980 |
| | 40 | 0.969 | 0.966 | 0.983 |
| 4 | 10 | 0.983 | 0.949 | 0.987 |
| | 20 | 0.970 | 0.955 | 0.985 |
| | 40 | — | — | — |

## 5. *An Application*

To determine the risk ratio of urethral strictures in the male offspring of diethylstilbestrol exposed pregnant women, a cohort study was performed (Henderson, Benton, Cosgrove, Baptista, Aldrich, Townsend, Hart and Mack 1976). From $N_c = 111$ unexposed pregnancies, 17 offspring with symptoms of urogenital tract abnormalities and two offspring with problems passing urine were found. From $N_t = 225$ exposed pregnancies, the numbers of offspring with the above two symptoms were 55 and 29, respectively. Thus, the estimated failure probabilities and risk ratio were $\hat{p}_c = 0.153$, $\hat{p}_t = 0.244$ and $\hat{R} = 1.6$ for urogenital tract abnormalities, and $\hat{p}_c = 0.081$, $\hat{p}_t = 0.129$ and $\hat{R} = 7.2$ for problems passing urine. The three lower 97.5% confidence limits obtained from these data are presented in Table 3. Method B failed in one case.

## 6. *Discussion*

Methods A and C produce similar and appropriate results. Method B, which is erratic and may fail, is not recommended.

Although computation by Method A requires the use of a digital computer, an approximate value for $x_L$ in (2) may be obtained using the method of Cornfield (1956). The

TABLE 3
Three Lower 97.5% Confidence Limits for the Risk Ratio of Urethral Strictures

| Method | Symptoms of Urogenital Tract Abnormalities ($\hat{R}$ = 1.6) | Problems Passing Urine ($\hat{R}$ = 7.2) |
|---|---|---|
| A | 0.965 | 1.862 |
| B | 1.019 | —* |
| C | 0.974 | 1.738 |

* Method B led to a disjoint interval for this situation.

approximate value of $x_L$ is the smallest root of the equation

$$[(x - x_L) - 0.5] \times [x_L^{-1} + (m - x_L)^{-1} + (N_t - x_L)^{-1} + (x_L + N_c - m)^{-1}]^{1/2} = z_{1-\alpha} ,$$

which can be calculated rapidly in an iterative fashion (Gart 1971). $R_L$ is then calculated from (3). Simulation studies with this approximation yielded results similar to those of Method A.

Another method based on the Poisson distribution for the situation when $p_t$ and $p_c$ are small was suggested by Bross (1954). Simulation studies for this method yielded reasonable, although conservative, results similar to those of Method A.

Finally, another method based on the variate $\hat{p}_t - R\hat{p}_c$ was suggested by Noether (1957). When applied to the current problem, this method is asymmetric with respect to which sample appears in the numerator of $\hat{R}$, and for that reason is not recommended.

All methods can be used to provide upper limits. Since we studied $R < 1$ as well as $R \geq 1$, similar conclusions would apply to upper limits. For all methods, two-sided limits can be obtained as the combined upper and lower limits.


## *Acknowledgments*

The authors wish to thank Lorene Deckert and Gail Williams for their technical assistance. Appreciation is given to the referees for their helpful suggestions.

The first and third authors of this paper were supported in part by Grant NIH GM23732-01 from the National Institutes of Health, United States Health Service. The second and fourth authors were supported by Contract NOI-CP-53500 and Grant POICA-17054 from the National Cancer Institute, National Institutes of Health, United States Public Health Service.



## *Résumé*

*On compare trois méthodes pour obtenir des intervalles de confiance pour le rapport de risque de défaillance de deux échantillons binomiaux indépendants Ce sont :*

*(A) la méthode de Thomas et Gart,*

*(B) une adaptation de la méthode de Fieller utilisant la distribution gaussienne, et*

*(C) une méthode proposée utilisant la transformation logarithmique. Sur la base de simulations extensives nous avons conclu que la méthode A est raisonnable mais conservative, la méthode B irrégulière et ne doit pas être utilisée, et la méthode C est raisonnable et moins conservative que la méthode A. La méthode C, la plus simple pour le calcul est recommandée.*


## *References*


Atkinson, A. C. and Pearce, M. C. (1976). The computer generation of beta, gamma and normal random variables. *Journal of the Royal Statistical Society, Series A 139*, 431–451.

Bross, I. (1954). A confidence interval for a percentage increase. *Biometrics 10*, 245–250.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. *Proceedings of the Third Berkeley Symposium IV*, J. Neyman (ed), University of California Press, Berkeley.

Fieller, E. C. (1944). A fundamental formula in the statistics of biological assay and some applications. *Quarterly Journal of Pharmacy and Pharmacology 17*, 117–123.

Fleiss, J. L. (1973). *Statistical Methods for Rates and Proportions*, Wiley and Sons, New York.

Gart, J. J. (1971). The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification. *Review of the International Statistical Institute 39*, 148–169.

Henderson, B. E., Benton, B., Cosgrove, M., Baptista, J., Aldrich, J., Townsend, D., Hart, W. and Mack, T. M. (1976). Urogenital tract abnormalities in sons of women treated with diethylstilbestrol. *Pediatrics 58*, 505–507.

Noether, G. E. (1957). Two confidence intervals for the ratio of two probabilities and some measures of effectiveness. *Journal of the American Statistical Association 52*, 36–45.

Pike, M. C. and Hill, J. D. (1966). Remark on algorithm 266 (G5) pseudo-random numbers. *Communications of the ACM 9*, 687.

Thomas, D. G. and Gart, J. J. (1977). A table of exact confidence limits for differences and ratios of two proportions and their odds ratio. *Journal of the American Statistical Association, 72*, 73–76.

*Received July* 1977; *Revised February* 1978