STATISTICS IN MEDICINE, VOL. 12, 691-698 (1993)

SAMPLE SIZE AND POWER FOR PROSPECTIVE ANALYSIS OF RELATIVE RISK

WILLIAM C. BLACKWELDER

National Institute of Allergy and Infectious Diseases, Solar Building, Room 3A24, Bethesda, Maryland 20892, U.S.A.

SUMMARY

In a placebo-controlled vaccine efficacy trial or a trial of equivalence of vaccines, one may wish to show that relative risk of disease is less than a specified value R_0, not equal to one. This paper compares three methods for estimating relative risk in the binomial setting, based on a logarithmic transformation, likelihood scores, and a Poisson approximation. Exact power and size of test are calculated by enumeration of possible binomial outcomes, and power is approximated from asymptotic formulations. Although the score method is generally preferable, for most studies of practical interest the log and score methods are comparable, and the Poisson method is also appropriate for small risks, up to about 0·05. When true and null relative risks are less than one, unequal allocation of study individuals can increase power, and the asymptotic formula for the log method may substantially underestimate power; in such a study the power approximation for the score method is more reliable, even if the log method is used in analysis. Exact power calculations are helpful in planning studies. The log and Poisson methods, but not the score method, apply readily in the case of unequal follow-up.

1. INTRODUCTION

Medical research is frequently concerned with estimation or hypothesis testing of the ratio of binomial probabilities (relative risk, risk ratio). Standard methods are available for testing the null hypothesis that relative risk is one.[1,2] There may be interest, however, in a null value of relative risk which is not one. For example, when a new therapy is thought to be equivalent to a standard therapy, one might design a study to show that the new therapy is not worse than the standard by as much as some specified quantity. This situation has been considered when the *difference* of risks is of interest.[3,4] Other methods are required, however, when interest is in the *ratio* of risks. An example is a randomized trial designed to show that efficacy of a vaccine (defined as one minus relative risk in vaccinated, compared with unvaccinated, individuals), is greater than a prespecified value. Alternatively, one might design a trial to show that risk of disease or a possible side effect with a new vaccine, relative to a standard vaccine, is less than a specified value. Even in a trial designed to test the conventional null hypothesis of no difference, a criterion that a confidence limit for relative risk exceeds or is less than a specified value may have use for interim monitoring.[5]

Gart and Nam[6] compared various methods of estimating the ratio of binomial probabilities for small and moderate sample sizes. They concluded that a likelihood score method[7-9] was generally preferable to one based on the logarithm of a binomial proportion.[10] In this report I compare these two methods, as well as one based on a Poisson approximation, for a prospective study designed to show that relative risk is greater than, or less than, a specified value other than

0277-6715/93/070691-08$09.00

Received October 1990
Revised July 1992

one. I approximate power from asymptotic sample size and power formulations, derived herein for the log and Poisson methods and developed previously for the likelihood score method.[11] I calculate actual power and size of test using the relevant binomial distributions, and discuss extension to the setting of unequal follow-up.

2. METHODS OF ESTIMATION AND TESTING

Suppose we observe x_1 and x_2 cases in binomial random samples of size n_1 and n_2 from two populations subject to disease risks p_1 and p_2. Let $N = n_1 + n_2$, and let $k = n_1/N$ denote the proportion of the total sample from population 1. We wish to show that $R = p_1/p_2$, the ratio of risk in population 1 ('treatment' or 'experimental therapy') to risk in population 2 ('control' or 'standard therapy') is less than some value R_0. That is, we wish to reject the hypothesis H_0: $R \geqslant R_0$, by a test at significance level α, in favour of the one-sided alternative H_1: $R < R_0$; equivalently, we wish to obtain a 100 $(1 - 2\alpha)$ per cent confidence interval for R with upper limit less than R_0. (The sample size and power calculations described herein are also valid, with slight modification, for testing H_0: $R \leqslant R_0$.)

2.1. Logarithmic transformation

From the asymptotic normality of the natural logarithm of an observed proportion,[10] for large samples $\log \hat{p}_i$ is approximately distributed $N(\log p_i, q_i/n_i p_i)$, $i = 1, 2$, where $\hat{p}_i = x_i/n_i$ and $q_i = 1 - p_i$. We reject H_0 if the statistic

$$z_L = (\log \hat{R} - \log R_0)/\hat{\sigma} \tag{1}$$

is less than $-z_\alpha$, where $\hat{R} = \hat{p}_1/\hat{p}_2$, $\hat{\sigma} = (\hat{q}_1/n_1\hat{p}_1 + \hat{q}_2/n_2\hat{p}_2)^{1/2}$, and z_α is the upper 100α per cent point of the standard normal distribution. The corresponding 100 $(1 - 2\alpha)$ per cent confidence interval for R is $\hat{R}$ exp $(\pm z_\alpha\hat{\sigma})$. (A modification which avoids the problem of z_L undefined for $\hat{p}_i = 0$ or $\hat{p}_i = 1$ is to add 1/2 to both x_i and n_i in $\hat{p}_i$ and $\hat{\sigma}$.[6,12,13])

For true risks p_1 and p_2, the probability $1 - \beta$ that the upper confidence limit is less than R_0, or power of the test, is just $P(z_L < -z_\alpha | p_1, p_2)$. Straightforward algebra and substitution of population parameters for estimates lead to the approximation

$$1 - \beta = \Phi[-z_\alpha + (\log R_0 - \log R)/(q_1/n_1p_1 + q_2/n_2p_2)^{1/2}], \tag{2}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Solving (2) for the total sample size N yields

$$N = (z_\alpha + z_\beta)^2[q_1/kp_1 + q_2/(1 - k)p_2]/(\log R_0 - \log R)^2, \tag{3}$$

if $R < R_0$.

O'Neill[14] has given a formula for the sample size necessary for a confidence interval for $\log R$ to have half width d. Setting $k = 0{\cdot}5$ and $z_\beta = 0$ in (3) and $d = \log R_0 - \log R$, we see that O'Neill's formula is a special case of (3) for equal-sized groups and power = 50 per cent. (The normal approximation implies a probability of 50 per cent that the observed $\hat{R}$ < the true R, and hence that a confidence interval for $\log R$ with half width $(\log R_0 - \log R)$ has upper limit $< \log R_0$). Similarly, we can show that a formula given by Lemeshow *et al.*[15] for estimating relative risk with specified precision is a special case of (3) for 50 per cent power.

2.2. Likelihood scores

We can base hypothesis testing and interval estimation of R on likelihood scores.[7,8] Setting the partial derivative of the log-likelihood with respect to p_2 equal to 0, we obtain $\tilde{p}_2$ as a solution of the quadratic equation

$$NR_0\tilde{p}_2^2 - [(x_2 + n_1)R_0 + x_1 + n_2]\tilde{p}_2 + x_1 + x_2 = 0.$$

The test statistic is

$$z_S = [(x_1 - n_1\tilde{p}_1)/\tilde{q}_1]/\tilde{\sigma}, \tag{4}$$

where $\tilde{p}_1 = R_0\tilde{p}_2$, $\tilde{\sigma} = (\tilde{q}_1/n_1\tilde{p}_1 + \tilde{q}_2/n\tilde{p}_2)^{-1/2}$, and $\tilde{q}_i = 1 - \tilde{p}_i$. A version of this statistic with a correction for skewness has also been derived.[6,8] We obtain a confidence interval for R by iterative methods.[7,8]

Miettinen and Nurminen[9] derived a statistic different in form but equivalent to the statistic (4) multiplied by $[(N-1)/N]^{1/2}$. Using their formulation, Farrington and Manning[11] obtained a sample size expression equivalent to

$$N = \{z_\alpha[\bar{p}_1\bar{q}_1/k + R_0^2\bar{p}_2\bar{q}_2/(1-k)]^{1/2} + z_\beta[p_1q_1/k + R_0^2p_2q_2/(1-k)]^{1/2}\}^2/(R_0p_2 - p_1)^2, \tag{5}$$

for $R < R_0$, where $\bar{p}_i$ is the asymptotic value of $\tilde{p}_i$ and $\bar{q}_i = 1 - \bar{p}_i$. The corresponding power formulation is

$$1 - \beta = \Phi(\{-z_\alpha[\bar{p}_1\bar{q}_1/k + R_0^2\bar{p}_2\bar{q}_2/(1-k)]^{1/2} + N^{1/2}(R_0p_2 - p_1)\}/[p_1q_1/k + R_0^2p_2q_2/(1-k)]^{1/2}). \tag{6}$$

For $R_0 = 1$, the square of the statistic (4) is the Pearson chi-square statistic,[6,8] and for $R_0 = 1$ and $k = 1/2$, (5) reduces to a standard formula for comparing two proportions.[2]

2.3. Poisson approximation

For sufficiently large n_i and small p_i, the number of cases in population i is distributed approximately as a Poisson variate with parameter n_ip_i.[16] We let x_i be the number of cases observed from population i and let $X = x_1 + x_2$. Then, conditional on X, x_1 is binomially distributed with parameters X and P, where

$$P = p_1/(p_1 + hp_2) = R/(h + R) \tag{7}$$

and $h = (1-k)/k$.[16] Given x_1 and x_2, we test H_0: $R \geqslant R_0$ by testing the equivalent hypothesis H_0: $P \geqslant P_0$, where $P_0 = R_0/(h + R_0)$. We obtain a 100 $(1 - 2\alpha)$ per cent confidence interval for R by setting $R = hP/(1-P)$ from (7) and substituting the corresponding confidence limits for P.

We can use exact methods to test H_0 and estimate R, but these methods are not explored in this paper. For sufficiently large numbers of cases, the asymptotic normality of the estimate $\hat{P} = x_1/X$ leads to the statistic

$$z_P = (\hat{P} - P_0)/[P_0(1-P_0)/X]^{1/2}, \tag{8}$$

which is equivalent to a likelihood score statistic derived under a Poisson assumption. We approximate power by

$$1 - \beta = \Phi\{[-z_\alpha(P_0(1-P_0))^{1/2} + X^{1/2}(P_0 - P)]/[P(1-P)]^{1/2}\} \tag{9}$$

and the total number of cases required by

$$X = \{z_\alpha[P_0(1 - P_0)]^{1/2} + z_\beta[P(1 - P)]^{1/2}\}^2/(P_0 - P)^2, \quad (10)$$

if $P < P_0$. The formula (10) is a generalization of a sample size formula given by Breslow and Day[17] for the case $P_0 = 1/2$. A formula for the case $P_0 = 1/2$ based on the arc sine transformation has also been suggested.[18]

The total number of subjects N is the number which we expect would result in X cases, that is,

$$N = X/[kp_1 + (1 - k)p_2]. \quad (11)$$

For testing H_0: $R \leqslant R_0$, we obtain sample size from (3), (5), (10), and (11), now valid for $R > R_0$; we obtain power by changing the signs of $(\log R_0 - \log R)$ in (2), $(R_0p_2 - p_1)$ in (6), and $(P_0 - P)$ in (9).

3. EXAMPLES AND COMPARISONS

In this section I compare the log, score, and Poisson methods with respect to power and size of test for two vaccine trials as well as a series of more general examples. The comparisons use exact power calculations for each method, obtained as follows. Given p_1, p_2, α, and either k and N or n_1 and n_2, we can generate the probabilities of all possible outcomes (2×2 tables), as Gart and Nam did in their evaluation of confidence coefficients.[6] The power is then the sum of the probabilities of outcomes which lead to rejection of H_0. We obtain the size of test (equivalently, the probability that the upper confidence limit is less than the true value) by setting $p_1 = R_0p_2$, and the sample size required for a given power from iterations of the power calculation. The extensive repetitive computations required for large sample sizes are readily programmed for a computer.

3.1. Placebo-controlled vaccine trial

For a randomized placebo-controlled trial of pertussis (whooping cough) vaccines in Sweden,[19] the sample size determination required 0·8 probability that the lower limit of a 90 per cent confidence interval for vaccine efficacy (VE) be greater than 70 per cent. Since $VE = 1 - R$, where $R = p_1/p_2$ is the ratio of pertussis risk in children who received vaccine to risk in placebo recipients, an equivalent requirement is that the upper confidence limit for R be less than 0·3, or that we reject H_0: $R \geqslant 0{\cdot}3$ at the one-sided 5 per cent significance level. The investigators anticipated a pertussis risk of 4 per cent in placebo recipients, during approximately 15 months of follow-up, and true vaccine efficacy of 90 per cent. For $\alpha = 0{\cdot}05$, $1 - \beta = 0{\cdot}8$, $k = 0{\cdot}5$, $p_1 = 0{\cdot}004$, $p_2 = 0{\cdot}04$, and $R_0 = 0{\cdot}3$, the sample size from (3) is 2797, which is 34 per cent larger than the 2088 from exact calculation (Table I). Sample sizes from (5) and (11) are 2119 and 2032, respectively, and the corresponding exact sample sizes are 2029 and 2032. Hence, in contrast to the approximate calculations, the actual sample sizes required are similar for the three methods.

For this study, however, both the null and (assumed) true relative risks are far from 1, and equal-sized groups are not optimum. Direct search for the optimum k for the log method, using exact power calculations, gives $k = 0{\cdot}61$ and a sample size of 1856; for this sample size, all three methods have 80 per cent power (Table I). The sample size from (3) is 2406, an overestimate by 30 per cent; in contrast, the approximations for both the score and Poisson methods (1925 and 1819, respectively) are close to the sample size needed for the log method.

In this example the size of test is lower than the nominal 0·05 for all three methods, but closest to 0·05 for the score method (Table I). The score method with skewness correction improves the size of test (to 0·050 or 0·051 for the above sample sizes) and also gives slightly higher power than the uncorrected score method.

Table I. Power and size of three tests for relative risk, by asymptotic formula and exact calculation

						Power (asymptotic formula)			Exact power and size of test		
p_1	p_2	R_0	N	k	α,one-sided	Log	Score	Poisson	Log	Score	Poisson
(a) $R_0 < 1$											
0·004	0·04	0·3	2,797	0·5	0·05	0·800	0·902	0·920	0·903 0·043	0·910 0·045	0·909 0·044
0·004	0·04	0·3	2,088	0·5	0·05	0·693	0·794	0·812	0·800 0·041	0·812 0·044	0·812 0·044
0·004	0·04	0·3	1,856	0·61	0·05	0·705	0·785	0·809	0·801 0·043	0·802 0·046	0·801 0·044
0·01	0·05	0·3	10,400	0·5	0·025	0·768	0·800	0·795	0·801 0·023	0·803 0·023	0·797 0·022
0·01	0·1	0·3	1,000	0·5	0·025	0·657	0·765	0·769	0·760 0·018	0·785 0·020	0·775 0·018
0·1	0·3	0·5	1,000	0·5	0·025	0·768	0·801	0·722	0·798 0·022	0·803 0·024	0·741 0·014
0·005	0·05	0·3	500	0·5	0·05	0·320	0·296	0·272	0·052 0·009	0·323 0·039	0·300 0·036
0·005	0·05	0·5	500	0·5	0·05	0·531	0·639	0·666	0·513 0·035	0·698 0·050	0·667 0·038
0·005	0·05	0·5	500	0·6	0·05	0·581	0·679	0·721	0·640 0·039	0·728 0·044	0·728 0·042
(b) $R_0 = 1$											
0·025	0·05	1·0	2,000	0·5	0·025	0·821	0·838	0·837	0·839 0·024	0·846 0·025	0·838 0·022
0·05	0·1	1·0	500	0·5	0·025	0·544	0·565	0·534	0·552 0·022	0·572 0·024	0·540 0·019
0·15	0·3	1·0	200	0·5	0·025	0·688	0·722	0·615	0·713 0·024	0·729 0·025	0·632 0·010
(c) $R_0 > 1$											
0·01	0·01	1·5	18,910	0·5	0·025	0·800	0·796	0·797	0·797 0·026	0·799 0·026	0·795 0·025
0·05	0·05	1·5	3,628	0·5	0·025	0·800	0·796	0·780	0·796 0·026	0·799 0·026	0·784 0·022
0·05	0·05	1·5	1,000	0·5	0·025	0·312	0·323	0·306	0·314 0·025	0·317 0·026	0·303 0·023
0·1	0·1	1·5	1,000	0·5	0·025	0·570	0·573	0·532	0·570 0·026	0·573 0·026	0·532 0·019
0·15	0·15	1·5	1,000	0·5	0·025	0·768	0·765	0·702	0·763 0·026	0·767 0·026	0·714 0·016
0·5	0·5	1·5	200	0·5	0·025	0·818	0·804	0·532	0·807 0·029	0·805 0·025	0·544 0·001
0·05	0·025	4·0	2,000	0·5	0·025	0·821	0·786	0·784	0·784 0·028	0·793 0·029	0·778 0·026
0·1	0·05	4·0	1,000	0·5	0·025	0·834	0·796	0·784	0·798 0·029	0·800 0·029	0·782 0·023
0·15	0·075	4·0	650	0·5	0·025	0·838	0·798	0·775	0·799 0·029	0·799 0·028	0·775 0·021

3.2. Comparison of two vaccines

Another type of trial compares an acellular pertussis vaccine to the standard whole-cell type of vaccine. Acellular vaccines, which contain specific antigens rather than the whole *Bordetella pertussis* bacterium, are considered safer than the whole-cell vaccine and probably equally effective in preventing typical pertussis disease. Suppose one wishes to show that the risk of disease with an acellular vaccine is less than 1·5 times the risk with whole-cell vaccine. If we assume the true risk is 0·01 for each vaccine during the planned follow-up period, test at the one-sided 2·5 per cent level, and desire 80 per cent power for two equal-sized groups, then we obtain similar total sample sizes (rounded to four significant digits): 18,910 for the log method, 19,110 for the score method, and 19,070 under the Poisson assumption. Exact calculations show that the actual power for $n = 18{,}910$ is 0·80 for all three methods, and size of test is close to the nominal 0·025 (Table I).

3.3. Other comparisons

Numerous further comparisons of the methods, including some with $R_0 = 1$, have been made. Some representative comparisons, chosen to illustrate certain characteristics of the methods, appear in Table I. The examples shown also correspond to tests of H_0: $R \leqslant R_0$ for R redefined as p_2/p_1 and R_0 redefined as the reciprocal of the tabulated value; k is still the proportion from population 1. The log statistic with the 1/2 correction and the skewness-corrected score statistic were calculated, but results for these are not included. In the table, the 1/2 correction was used only for $x_i = 0$ or n_i, but not otherwise; for $x_1 = n_1$ and $x_2 = n_2$, the log statistic was not calculated.

The major findings from the comparisons are as follows:

1. For $R \leqslant 1$, of the three tests the score test usually has highest power and size of test nearest the nominal value.
2. However, for large studies of practical interest (such as vaccine efficacy trials and equivalence studies), power and size for the log and score tests are comparable.
3. For $R < 1$ in large studies involving small risks, size of test is generally smaller than the nominal value for all three tests. In some cases size may be larger than nominal, especially for $R \geqslant 1$.
4. The skewness correction almost always brings the size of the score test closer to the nominal value (as in previous comparisons[6]) and increases or decreases power along with size (data not shown).
5. The Poisson approximation is good for very small risks, and adequate for risks up to about 0·05 (0·1 in some cases when R and $R_0 > 1$), except when expected numbers of cases are small. As risk increases, size of test tends to be farther from the nominal value and power to be lower than for the other methods; the asymptotic formulation, however, still gives a good approximation to the actual power.
6. Asymptotic formulations for the score and (if applicable) Poisson methods may estimate power for the log test better than the formula derived directly from the log statistic, especially for reasonably high power. This is the case when $R < R_0 \leqslant 0{\cdot}5$, as in a placebo-controlled vaccine trial, for which the approximating formula for the log method may substantially underestimate power.
7. In a placebo-controlled vaccine trial, both power and size of test can be improved if $k > 0{\cdot}5$. Direct search suggests an optimum k generally in the range of 0·6 to 0·7. However, power is not sensitive to the choice of k near the optimum.

8. Use of the 1/2 modification in the log test generally produces about the same or less satisfactory results than those tabulated (data not shown).
9. The log test is more apt than the score test and, when applicable, the Poisson test to have low power and incorrect size when expected numbers of events are small. There is no simple rule of thumb, however, for determining expected numbers necessary for adequate size or adequate agreement between asymptotic and exact power calculations.

4. COMMENTS

Gart and Nam[6] compared various approximate methods for interval estimation of R when sample sizes are small or moderate. They found the score method generally preferable to the log method and, with a skewness correction, the preferred method overall. The log method, however, has been suggested for clinical trials of vaccine efficacy[14] and used in the design of one such trial;[19] hence evaluation of its performance in such settings is relevant.

The log statistic (1) is a Wald statistic;[20,21] that is, we substitute unrestricted maximum likelihood estimates for the parameters throughout, including the expression for the variance of $\log(\hat{p}_1/\hat{p}_2)$. In some situations Wald statistics show aberrant behaviour for alternative hypotheses or maximum likelihood estimates far from the null value.[13,22,23] Indeed, for the types of studies considered in this paper, the log statistic may not increase monotonically as $\hat{p}_1$ tends to 0 for fixed $\hat{p}_2$. The comparisons in Section 3, however, indicate that concern about its use in such studies is more theoretical than practical. The log method gives good results for large prospective studies, except in some cases where the expected number of events in a group is very small. Use of the 1/2 correction in the statistic does not improve its power or size; similarly, we would not expect a skewness correction to be helpful.[6]

An important consideration is that in most large prospective studies follow-up among subjects is not equal. Both the log and Poisson methods apply readily in the setting of unequal follow-up, but the score method does not. To apply the log method, we let p^* and $V^*(p^*)$ be estimates of risk and its variance (for example, from Kaplan–Meier calculations[24]); then $\hat{\sigma}$ in (1) becomes $[V^*(p_1^*)/(p_1^*)^2 + V^*(p_2^*)/(p_2^*)^2]^{1/2}$. We apply the Poisson approximation by defining k as the proportion of follow-up (or the proportion of exposure to risk) in population 1. We can approximate sample size for the log and Poisson methods by letting p_1 and p_2 be risks associated with the average follow-up time; we can also express sample size for the Poisson method in terms of disease rates and follow-up time needed, rather than risks and the number of individuals.

Another possible statistic for the unequal follow-up situation is based on the difference $p_1^{1/3} - R^{1/3}p_2^{1/3}$.[25] We might also base estimation on incidence density calculations,[26] with a suitable variance expression. Under a proportional hazards assumption (as is warranted in the Poisson case, for example), we may estimate relative risk, with adjustment for covariates if necessary, from a regression model for time to disease.[24]

In summary, the likelihood score method is generally preferable to the logarithmic transformation for estimating and testing the ratio of binomial proportions. Both methods, however, are appropriate in large prospective studies designed to show that relative risk is less than, or greater than, some R_0 not equal to one. When true and null relative risks are less than one, as in a placebo-controlled vaccine efficacy trial, (1) we can increase power by allocating more individuals to the group with smaller risk, and (2) the actual power and size of test tend to be similar for the two methods, although asymptotic results for the log method may be misleading. Hence the asymptotic formulation for the score test[11] is more generally applicable, even if analysis using the log statistic is planned. Exact power calculations are also helpful when planning studies. The

Poisson approximation is appropriate in general for risks up to about 0·05. The log and Poisson methods, but not the score method, apply readily to the case of unequal follow-up.

ACKNOWLEDGEMENTS

The author thanks Patrick Olin and Roger Bernier for suggesting the problems which led to this work, John Gart for valuable discussions and comments, and reviewers for helpful comments.

REFERENCES

1. Armitage, P. *Statistical Methods in Medical Research*, Wiley, New York, 1971.
2. Halperin, M., Rogot, E., Gurian, J. and Ederer, F. 'Sample sizes for medical trials with special reference to long-term therapy', *Journal of Chronic Diseases*, **21**, 13–24 (1968).
3. Makuch, R. and Simon, R. 'Sample size requirements for evaluating a conservative therapy', *Cancer Treatment Reports*, **62**, 1037–1040 (1978).
4. Blackwelder, W. C. '"Proving the null hypothesis" in clinical trials', *Controlled Clinical Trials*, **3**, 345–353 (1982).
5. Reed, G. F. and Blackwelder, W. C. 'Interim monitoring for vaccine efficacy', *Controlled Clinical Trials*, **12**, 669 (1991).
6. Gart, J. J. and Nam, J. 'Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness', *Biometrics*, **44**, 323–338 (1988).
7. Koopman, P. A. R. 'Confidence intervals for the ratio of two binomial proportions', *Biometrics*, **40**, 513–517 (1984).
8. Gart, J. J. 'Approximate tests and interval estimation of the common relative risk in the combination of 2×2 tables', *Biometrika*, **72**, 673–677 (1985).
9. Miettinen, O. and Nurminen, M. 'Comparative analysis of two rates', *Statistics in Medicine*, **4**, 213–226 (1985).
10. Katz, D., Baptista, J., Azen, S. P. and Pike, M. C. 'Obtaining confidence intervals for the risk ratio in cohort studies', *Biometrics*, **34**, 469–474 (1978).
11. Farrington, C. P. and Manning, G. 'Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk', *Statistics in Medicine*, **9**, 1447–1454 (1990).
12. Walter, S. D. 'The distribution of Levin's measure of attributable risk', *Biometrika*, **62**, 371–374 (1975).
13. Pettigrew, H. M., Gart, J. J. and Thomas, D. G. 'The bias and higher cumulants of the logarithm of a binomial variate', *Biometrika*, **73**, 425–435 (1986).
14. O'Neill, R. T. 'On sample sizes to estimate the protective efficacy of a vaccine', *Statistics in Medicine*, **7**, 1279–1288 (1988).
15. Lemeshow, S., Hosmer, D. W. and Klar, J. 'Sample size requirements for studies estimating odds ratios or relative risks', *Statistics in Medicine*, **7**, 759–764 (1988).
16. Johnson, N. L. and Kotz, S. *Discrete Distributions*, Houghton Mifflin, Boston, 1969.
17. Breslow, N. E. and Day, N. E. *Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies*, International Agency for Research on Cancer, Lyon, 1987.
18. Gail, M. 'Power computations for designing comparative Poisson trials', *Biometrics*, **30**, 231–237 (1974).
19. Ad Hoc Group for the Study of Pertussis Vaccines. 'Placebo-controlled trial of two acellular pertussis vaccines in Sweden – protective efficacy and adverse events', *Lancet*, **i**, 955–960 (1988).
20. Wald, A. 'Tests of statistical hypotheses concerning several parameters when the number of observations is large', *Transactions of the American Mathematical Society*, **54**, 426–482 (1943).
21. Kotz, S. and Johnson, N. L. (Eds.) *Encyclopedia of Statistical Sciences, Vol. 9*, Wiley, New York, 1988.
22. Hauck, W. W. and Donner, A. 'Wald's test as applied to hypotheses in logit analysis', *Journal of the American Statistical Association*, **72**, 851–853 (1977). Corrigendum, *Journal of the American Statistical Association*, **75**, 482 (1980).
23. Vaeth, M. 'On the use of Wald's test in exponential families', *International Statistical Review*, **53**, 199–214 (1985).
24. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
25. Bailey, B. J. R. 'Confidence limits to the risk ratio', *Biometrics*, **43**, 201–205 (1987).
26. Kleinbaum, D. G., Kupper, L. L. and Morgenstern, H. *Epidemiologic Research*, Lifetime Learning Publications, Belmont, California, 1982.