BIOMETRICS 44, 323–338
June 1988

Approximate Interval Estimation of the Ratio of Binomial Parameters: A Review and Corrections for Skewness

John J. Gart and Jun-mo Nam

Mathematical Statistics and Applied Mathematics Section,
Biostatistics Branch, National Cancer Institute,
Bethesda, Maryland 20892, U.S.A.

SUMMARY

Various methods for finding confidence intervals for the ratio of binomial parameters are reviewed and evaluated numerically. It is found that the method based on likelihood scores (Koopman, 1984, *Biometrics* **40**, 513–517; Miettinen and Nurminen, 1985, *Statistics in Medicine* **4**, 213–226) performs best in achieving the nominal confidence coefficient, but it may distribute the tail probabilities quite disparately. Using general theory of Bartlett (1953, *Biometrika* **40**, 306–317; 1955, *Biometrika* **42**, 201–203), we correct this method for asymptotic skewness. Following Gart (1985, *Biometrika* **72**, 673–677), we extend this correction to the case of estimating the common ratio in a series of two-by-two tables. Computing algorithms are given and applied to numerical examples. Parallel methods for the odds ratio and the ratio of Poisson parameters are noted.

1. Introduction

The estimation of the ratio of the parameters from independent binomial variates is pervasive in biometry. In laboratory studies in animals and in cohort studies in humans, this ratio is called the *relative risk* (Gart, 1985a) or *risk ratio* (Katz et al., 1978). In case–control studies in epidemiology, it also arises in the estimation of the *attributable risk* (Walter, 1975, 1976). Many approximate methods have been proposed for finding confidence intervals for this parameter. Noether (1957) suggests two easily computed methods. Thomas and Gart (1977) suggest an "exact" method based on fixed marginals in the two-by-two tables. Santner and Snell (1980) consider related exact methods for finding such intervals. After considering previously proposed approximations as well as a method based on Fieller's theorem, Katz et al. (1978) recommend a method based on the logarithmic transformation. Recently Bailey (1987) suggests a modification of this Fieller's method. Koopman (1984) and Miettinen and Nurminen (1985) propose methods related to the likelihood. We show that these two methods are identical except for the $n - 1$ correction used by Miettinen and Nurminen in the variance estimation. These methods, unlike those previously proposed, are consistent with the usual Pearson chi-square test of a two-by-two table. Using the generalized divergence statistic of Cressie and Read (1984), Bedrick (1987) suggests an interval which is consistent with this statistic for the adjustable parameter equal to $\frac{1}{2}$ rather than to unity for which value this statistic yields the Pearson chi-square under the null hypothesis.

Gart (1985a) derives Koopman's method as a special case of Bartlett's (1953, 1955) general likelihood method based on score statistics. Gart also extends these results to the stratified case and shows that the resulting intervals are consistent with Radhakrishna's (1965) optimal extension of Cochran's (1954) test to the stratified ratio of binomial

Key words: Attributable risk; Binomial variate; Likelihood; Relative risk; Risk ratio; Score statistic.

parameters. Miettinen and Nurminen also give a method for the stratified case, but their interval is not a score method and is not consistent with the optimal test.

In this paper, we investigate several of these methods by means of exact computations. We find the likelihood-based method, which we call the score method, together with the logarithmic method, come closest to achieving the nominal confidence coefficients in moderate sample sizes. We go on to compare the upper and lower tail areas and find that they are less disparate in the score method. Using Bartlett's general results, we derive a skewness correction which improves the score method further in this respect. The skewness correction is extended to the stratified case. Examples are given for the single-stratum and multiple-strata cases. Parallels to the odds ratio, to the limiting Poisson case, and to differences in proportions are noted.

2. Model and Notation

Consider a pair of independent binomial variates x_0 and x_1 with corresponding parameters p_0 and p_1 and sample sizes n_0 and n_1. We wish to find approximate confidence limits for $\phi = p_1/p_0$ that are adequate in moderate sample sizes. Also define $p_j = 1 - q_j$, $\hat{p}_j = x_j/n_j = 1 - \hat{q}_j$, $j = 0, 1$; and $\hat{\phi} = \hat{p}_1/\hat{p}_0$. Summation is denoted by dots, e.g., $x_. = x_0 + x_1$.

When considering the multiple-strata case we add a second subscript, $i = 1, \ldots, I$, everywhere except that we assume $\phi = p_{1i}/p_{0i}$ is independent of i.

3. Review of the Proposed Methods

The various approximate methods may be grouped into three subsets based on their mode of derivation.

3.1 *Crude Methods Using First-Order Variance Estimation*

These limits are simply point estimates plus and minus a multiple of its standard error estimated from a first-order Taylor's series expansion.

Noether's (1957) method I_2 is based on the estimator $\hat{\phi}$ whose asymptotic variance is $\phi^2 u(p_0, p_1)$ where $u(p_0, p_1) = q_0/(n_0 p_0) + q_1/(n_1 p_1)$. The $1 - \alpha$ confidence limits are taken to be

$$\hat{\phi}(1 \pm z_{\alpha/2}\sqrt{\hat{u}}),$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the normal distribution and $\hat{u} = u(\hat{p}_0, \hat{p}_1)$. The limits are not defined when x_0 and/or x_1 is zero.

Log-limits (Katz et al., 1978) may be based on the asymptotic normality of $\log(\hat{\phi})$, whose variance is $u(p_0, p_1)$. Unfortunately, both $\log(\hat{\phi})$ and $\hat{u}$ are undefined when x_1 and/or $x_0 = 0$. Walter (1975) suggests that a nearly unbiased point estimator of $\log(\phi)$ is

$$\log(\hat{\phi}_{1/2}) = \log\left(\frac{x_1 + \frac{1}{2}}{n_1 + \frac{1}{2}}\right) - \log\left(\frac{x_0 + \frac{1}{2}}{n_0 + \frac{1}{2}}\right).$$

Pettigrew, Gart, and Thomas (1986) show that this estimator's variance is estimated nearly unbiasedly by

$$\hat{u}_{1/2} = \frac{1}{x_1 + \frac{1}{2}} - \frac{1}{n_1 + \frac{1}{2}} + \frac{1}{x_0 + \frac{1}{2}} - \frac{1}{n_0 + \frac{1}{2}}.$$

The resulting limits for ϕ are

$$\hat{\phi}_{1/2}\exp(\pm z_{\alpha/2}\sqrt{\hat{u}_{1/2}}).$$

These limits always exist but yield a degenerate interval of $(1, 1)$ when $x_1 = n_1$ and $x_0 = n_0$.

3.2 *Fieller-like Intervals*

Like Fieller's method for the ratio of normal means, these limits are the solutions to a quadratic equation.

The Fieller interval of Katz et al. (1978) is based on the statistic $T = \hat{p}_1 - \phi\hat{p}_0$. They estimate its variance by $V(T) = (\hat{p}_1\hat{q}_1)/n_1 + (\phi^2\hat{p}_0\hat{q}_0)/n_0$. The limits are the two roots to the quadratic equation in ϕ, $T^2 = z^2_{\alpha/2}V(T)$. These roots may be complex, in which case the interval is taken to be $(0, \infty)$, or the limits may be exclusive. In the latter case, the confidence interval is disjoint, i.e., $(0, a)$ and (b, ∞) where $b > a$.

Bailey's (1987) modified Fieller interval uses the statistic $T^* = \hat{p}_1^{1/3} - (\phi\hat{p}_0)^{1/3}$, where the cube roots are chosen to minimize the skewness. The variance of T^* is estimated by the usual first-order variance formula with ϕ left unspecified but $p_0 = \hat{p}_0$ and $p_1 = \hat{p}_1$. Setting $(T^*)^2 = z^2_{\alpha/2}V(T^*)$ yields a quadratic equation in ϕ that yields Bailey's limits. Bailey suggests certain modifications of the general formula when x_1 or $x_0 = 0$, which appears to avoid complex or exclusive limits whenever $z^2_{\alpha/2} \leqslant 4.5$.

Noether's I_1 limits are based on a general theorem of Geary and also lead to limits that are the roots of a quadratic equation. The statistic is $T' = \hat{\phi} - \phi$, which is asymptotically normal with variance

$$\operatorname{var}(T') = \frac{\phi^2 q_0}{n_0 p_0} + \frac{\phi^2 q_1}{n_1 p_1}.$$

Noether substitutes ϕp_0 for p_1 in this variance formula and then lets $p_0 = \hat{p}_0$ as well. Denoting this estimated variance by $V(T')$, Noether solves the quadratic equation

$$(\hat{\phi} - \phi)^2 = z^2_{\alpha/2}V(T')$$

to obtain his interval I_1. These limits do not exist for x_1 and/or $x_0 = 0$, but otherwise the equation always yields real roots.

3.3 *Methods Based on Likelihood Methods*

Two recent papers (Koopman, 1984; Miettinen and Nurminen, 1985) have proposed methods that use the maximum likelihood estimator (MLE) of the nuisance parameter, p_0, for given values of ϕ. These methods are in fact almost identical and may be most easily derived, as well as generalized to multiple strata, by using the general theory of score tests (Gart, 1985a). Consider the log-likelihood to be the sum of the logarithms of two binomials with parameters p_0 and p_1. Reparameterize by letting $p_1 = \phi p_0$, so that the log-likelihood is a function of ϕ and p_0. The score for ϕ is then

$$S_\phi(\phi, p_0) = \frac{\partial L(\phi, p_0)}{\partial \phi} = \frac{x_1 - n_1 p_1}{q_1 \phi}. \tag{3.1}$$

Also, we have that

$$\frac{\partial L(\phi, p_0)}{\partial p_0} = \frac{x_0 - n_0 p_0}{q_0} + \frac{x_1 - n_1 p_1}{q_1}. \tag{3.2}$$

In both (3.1) and (3.2), recall that $p_1 = \phi p_0$. The MLE of p_0, $\tilde{p}_0$, is the solution to the equation, $\partial L(\phi, \tilde{p}_0)/\partial p_0 = 0$, which is the appropriate solution to the quadratic equation

$$a\tilde{p}_0^2 + b\tilde{p}_0 + c = 0, \tag{3.3}$$

where $a = n_. \phi$, $b = -[(x_0 + n_1)\phi + x_1 + n_0]$, and $c = x_. = x_0 + x_1$.

The score method is based on the statistic $S_\phi(\phi, \tilde{p}_0)$. It follows from general results of Bartlett (1953) that its variance is estimated by

$$\mathrm{var}[S_\phi(\phi, \tilde{p}_0)] = 1/[\phi^2 u(\tilde{p}_0, \tilde{p}_1)] = v(\phi, \tilde{p}_0)/\phi^2,$$

where, in the notation of Gart (1985a), $v(\phi, \tilde{p}_0) = 1/[u(\tilde{p}_0, \tilde{p}_1)]$. Thus, the approximate $1 - \alpha$ confidence limits are the two solutions to the equation

$$\chi^2_{\mathrm{K}}(\phi) = \frac{(x_1 - n_1\tilde{p}_1)^2}{\tilde{q}_1^2 v(\phi, \tilde{p}_0)} = z^2_{\alpha/2}.$$

This equation, which is identical with that proposed by Koopman, may be combined with (3.3) to yield a cubic equation in ϕ or may be solved iteratively (see §8). When $\phi = 1$, $\chi^2_{\mathrm{K}}(1)$ reduces to the usual Pearson chi-square test statistic.

Miettinen and Nurminen's derivation looks quite different. Like the Fieller method, it also starts with the statistic, $T = \hat{p}_1 - \phi\hat{p}_0$, but they estimate its variance by $(\tilde{p}_1\tilde{q}_1)/n_1 + (\phi^2\tilde{p}_0\tilde{q}_0)/n_0$. The limits are the roots to the equation

$$\chi^2_{\mathrm{MN}}(\phi) = \frac{(\hat{p}_1 - \phi\hat{p}_0)^2}{(\tilde{p}_1\tilde{q}_1)/n_1 + (\phi^2\tilde{p}_0\tilde{q}_0)/n_0} = z^2_{\alpha/2}.$$

We show in Appendix 1 that $\chi^2_{\mathrm{MN}}(\phi) = \chi^2_{\mathrm{K}}(\phi)$ and thus that these limits are identical. In actual application, Miettinen and Nurminen use a variance correction, that is, they multiply $\chi^2_{\mathrm{MN}}(\phi)$ by $(n_. - 1)/n_.$ before solving. The resulting limits are, in moderate sample sizes, slightly wider than those given by Koopman's formulation.

Bedrick (1987) uses the generalized power divergence statistic of Cressie and Read (1984), $I^\lambda(\phi)$, which includes both the Pearson statistic ($\lambda = 1$) and the likelihood-ratio statistic ($\lambda = 0$) as specific cases. His optimal interval is based on a statistic midway between these values, $\lambda = \frac{1}{2}$. Like the score method, it involves iterative calculation.

4. Properties of the Methods

Consider a variety of properties that are desirable in a confidence interval:

(1) *Consistency with Pearson's χ^2 test* When testing $\phi = 1$, if the one-tailed $P < \alpha/2$ the $1 - \alpha$ interval should exclude 1, and if $P > \alpha/2$ the interval should include $\phi = 1$.
(2) *Invariance* The interval for $1/\phi$ found by taking reciprocals of the limits for ϕ should be identical with those found by reversing the 0 and 1 subscripts in the original method.
(3) *Absence of incomputable or aberrant cases* The method should be computable in virtually all possible outcomes without the need to "fix up" the formulas. All the methods, except the log method, fail when $x_1 = x_0 = 0$. As this may be considered an informationless outcome, this failure is not necessarily a liability. Some of the methods sometimes yield negative limits and degenerate, infinite, or disjoint intervals.
(4) *Ease of computability* Noniterative methods are usually preferred to those requiring iterative calculations. With the wide availability of programmable desk calculators, as well as high-speed computers, this is no longer such an important issue.

Table 1 summarizes these properties for the various methods. Only the likelihood or score method is consistent with the usual chi-square test. Both the log method and the score method enjoy the invariance property. These two methods also have no incomputable cases of any consequence. The other three methods all involve division by zero whenever x_1 or $x_0 = 0$ and can, for small numbers, yield negative limits. Fieller's method, as noted above, may also yield disjoint or infinite limits. Three of the methods, including the log method as noted by Miettinen and Nurminen, yield a degenerate interval of (1, 1) whenever

Table 1
Properties of the methods

Method	Consistency with χ^2 test	Invariance	Incomputable[a] or aberrant cases	Non-iterative
Noether's I_2	No	No	Fails for x_0 or $x_1 = 0$; can yield negative limits; when $\hat{p}_0 = \hat{p}_1 = 1$, degenerate.[b]	Yes
Log($\hat{\phi}_{1/2}$)	No	Yes	When $\hat{p}_0 = \hat{p}_1 = 1$, degenerate.[b]	Yes
Fieller and Bailey's modification	No	No	Fails for x_0 or $x_1 = 0$; when $\hat{p}_0 = \hat{p}_1 = 1$, degenerate[b]; can yield negative limits or disjoint and infinite intervals.[c]	Yes
Noether's I_1	No	No	Fails for x_0 or $x_1 = 0$; can yield negative limits.	Yes
Score method and Bedrick's method	Yes	Yes	None	No

[a] All methods except log($\hat{\phi}_{1/2}$) fail when $x_0 = x_1 = 0$.
[b] When $\hat{p}_0 = \hat{p}_1 = 1$, this method yields the degenerate interval, (1, 1).
[c] When Bailey's suggestions for x_0 or $x_1 = 0$ are used, his method is always computable for $z^2_{\alpha/2} \leq 4.5$.

$\hat{p}_1 = \hat{p}_0 = 1$. For the most part Bailey's and Bedrick's methods share the properties of Fieller's and the score method, respectively.

5. Exact Evaluation of the Methods

We evaluate the exact confidence coefficients by numerical computation of the probabilities of all $(n_1 + 1)(n_0 + 1)$ possible outcomes. Thus, all methods must be defined at each possible point. When $x_0 = x_1 = 0$, we follow Koopman and define the interval to be $(0, \infty)$ for all methods except the logarithmic. If either $x_1 = 0$ or $x_0 = 0$, we let $\hat{p}_0 = 1/(2n_0)$ or $\hat{p}_1 = 1/(2n_1)$ in both Noether methods and the Fieller methods. When a negative lower limit is found in any method, it is set to zero. When the Fieller limits yield disjoint intervals, they are evaluated directly. When complex roots result, the interval is taken to be $(0, \infty)$. Katz et al., in their numerical evaluations, substitute the log method in such cases.

The exact probability of each outcome was computed from the product of binomials for a variety of values of p_0, p_1, n_0, and n_1. The exact probability of the interval covering the two values was computed for $\alpha = .05$ and .01.

For five of the methods, Table 2 gives typical values for $n_1 = n_0 = 15$ and $n_1 = 10$ and $n_0 = 20$ for 95% limits. In these moderate sample sizes, both of Noether's methods yield actual confidence coefficients that deviate appreciably, both above and below the nominal value. The Fieller limits typically have too large a coefficient, probably to some extent due to the frequency of disjoint and infinite limits. However, for unequal sample sizes, it may also yield too small a coefficient.

The log method and the score method yield actual coefficients reasonably close to the nominal value. Typically, the score method is somewhat better. This similarity in overall confidence coefficient does not necessarily imply that the methods yield similar limits in particular applications. Consider two specific examples.

Example 1 Let $x_1/n_1 = \frac{8}{15}$ and $x_0/n_0 = \frac{4}{15}$. The two methods yield nominal 95% intervals:

Log method: (.768, 4.65);
Score method: (.815, 5.34).

The interval for the score method is clearly shifted to the right of that for the log method.

Table 2
Actual coverage percentage for nominal 95% confidence intervals for ϕ

$n_. = 30$:		$n_0 = n_1 = 15$					$n_0 = 20, \quad n_1 = 10$				
	$\phi =$	$\frac{1}{2}$	1	2	4	7	$\frac{1}{2}$	1	2	4	7
$p_0 = .125$	Noether I_2	95.6	92.1	87.6	88.4	88.7	99.9	92.7	90.1	89.4	90.5
	$\text{Log}(\hat{\phi}_{1/2})$	99.4	99.9	97.8	94.6	92.5	97.6	99.0	99.7	96.5	94.5
	Fieller	99.7	99.3	98.4	99.1	99.3	100.0	99.1	98.6	98.7	99.3
	Noether I_1	100.0	97.4	92.3	88.8	87.3	100.0	99.8	95.8	89.4	85.4
	Score	96.3	96.6	96.3	96.1	96.1	93.1	96.5	95.8	95.1	96.1
	$\phi =$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$
$p_0 = .250$	Noether I_2	99.3	88.4	88.8	91.4	91.6	100.0	97.3	88.1	90.8	92.3
	$\text{Log}(\hat{\phi}_{1/2})$	96.3	97.8	98.6	96.6	93.3	93.4	95.6	97.4	97.8	95.1
	Fieller	99.8	97.4	97.3	98.1	98.3	99.8	98.7	95.3	97.5	98.5
	Noether I_1	100.0	99.7	94.4	91.4	85.9	99.9	100.0	97.9	91.7	85.3
	Score	94.7	96.3	94.6	94.7	95.8	95.6	95.3	95.7	94.6	95.2
	$\phi =$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$
$p_0 = .500$	Noether I_2	100.0	92.7	90.5	93.9	91.5	100.0	99.4	90.0	94.3	93.9
	$\text{Log}(\hat{\phi}_{1/2})$	93.4	94.6	96.6	96.9	93.2	88.7	91.7	93.8	96.3	95.6
	Fieller	99.8	95.0	93.0	95.3	95.0	99.8	99.1	91.4	92.9	93.3
	Noether I_1	98.9	99.5	98.0	93.9	87.0	97.9	98.3	99.3	95.2	83.9
	Score	95.9	96.1	94.7	95.4	95.0	95.3	96.3	95.2	94.8	93.9

Example 2 Let $x_1/n_1 = \frac{6}{10}$ and $x_0/n_0 = \frac{6}{20}$. The nominal 95% intervals are

Log method: (.883, 4.32);
Score method: (.844, 4.59).

In this case with unequal sample sizes, the log method yields an interval entirely within that of the score method.

In both examples, $\hat{\phi} = 2$. An inspection of Table 2 at $\phi = 2$ shows that, if anything, the average confidence coefficient for the log method exceeds that of the score method. On these superficial grounds, we might expect the interval for the log method to be wider.

A more reasonable approach to understanding the discrepancies between the two methods is to examine the actual probabilities in each of the tails beyond the upper and lower limits. These are given in Table 3. The actual lower-tail probabilities for the log method are typically much less than the nominal value of 2.5% when ϕ exceeds unity. Conversely, the upper-tail probabilities are typically greater than nominal when ϕ exceeds 1. Thus, it is apparent, even when the actual coefficient is nearly nominal, that the interval is shifted to the left. More importantly, the upper limit of the interval will be less than the true ϕ more than 2.5% of the time. On the other hand, the score method comes closer to achieving the nominal tail probabilities, although for small ϕ it too can have widely disparate tails.

Returning to our examples, we see how the exact computation helps to explain the discrepancies between these methods. In Example 1, the upper limit is beyond 4 with $p_0 \approx .25$. Table 3 shows that, for $\phi = 3.5$ and $p_0 = .25$, the log method leaves 6.7% in the upper tail while the score method leaves only 2.9%. Thus, we should expect the log method to yield the smaller upper limit. Conversely, the lower limit is somewhat less than one. At both $\phi = 1$ and $\phi = \frac{1}{2}$, the log method has a tail less than nominal while the score method's results are somewhat above nominal. Here we expect the log method to yield a smaller

Table 3
Lower and upper tail probabilities for nominal 95% *confidence intervals for the log and score methods*

$n_. = 30$:		$n_0 = n_1 = 15$					$n_0 = 20, \quad n_1 = 10$				
	$\phi =$	$\frac{1}{2}$	1	2	4	7	$\frac{1}{2}$	1	2	4	7
$p_0 = .125$	Lower log	.6	.0	.0	.0	.0	2.4	1.0	.0	.0	.0
	Lower score	3.7	1.7	.8	.0	.0	6.9	3.3	2.0	1.2	.0
	Upper log	.0	.0	2.2	5.4	7.5	.0	.0	.2	3.5	5.5
	Upper score	.0	1.7	2.9	3.9	3.9	.0	.2	2.2	3.7	3.9
	$\phi =$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$
$p_0 = .250$	Lower log	3.7	2.2	.7	.0	.0	6.6	4.4	2.6	.7	.0
	Lower score	5.3	2.9	2.7	2.0	1.3	4.4	4.6	2.8	2.4	1.7
	Upper log	.0	.0	.7	3.4	6.7	.0	.0	.0	1.6	4.9
	Upper score	.0	.8	2.7	3.3	2.9	.0	.0	1.5	3.0	3.1
	$\phi =$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$
$p_0 = .500$	Lower log	6.6	5.4	3.4	1.6	.4	11.3	8.3	6.2	3.5	.8
	Lower score	4.1	3.9	3.3	2.3	1.9	4.7	3.7	3.4	2.8	3.7
	Upper log	.0	.0	.0	1.6	6.5	.0	.0	.0	.2	3.7
	Upper score	.0	.0	2.0	2.3	3.1	.0	.0	1.4	2.6	2.4

lower limit, as indeed it does in Example 1. By examining the right panel of Table 3, one can similarly explain the discrepancies between the two methods found in Example. 2.

Although the performance of the score method clearly surpasses that of the log method in assigning tail probabilities, it can still at times yield quite disparate results, particularly for values of ϕ far from unity. We shall consider how this may be corrected and compare this correction to the recently proposed methods of Bailey (1987) and Bedrick (1987).

6. Correcting the Score Method for Skewness

The general theory of the score method (Bartlett, 1953, 1955) permits the correcting of the resulting confidence coefficients for both bias and skewness including terms of order $O(n_0^{-1/2}, n_1^{-1/2})$. Consider the normal deviate which is the square root of $\chi^2_K(\phi)$, namely,

$$z(\phi) = \frac{x_1 - n_1 \tilde{p}_1}{\tilde{q}_1 [v(\phi, \tilde{p}_0)]^{1/2}},$$

where $\tilde{p}_0$ is a solution to (3.3) for a given ϕ and $\tilde{p}_1 = \phi \tilde{p}_0$. As shown in Appendix 3,

$$\mathrm{E}[z(\phi)] = O(n_0^{-3/2}, n_1^{-3/2});$$

that is, the bias to order $O(n_0^{-1/2}, n_1^{-1/2})$ is zero. The skewness to this order is

$$\gamma_1[z(\phi)] = [v(\phi, p_0)]^{3/2} \left[\frac{q_1(q_1 - p_1)}{(n_1 p_1)^2} - \frac{q_0(q_0 - p_0)}{(n_0 p_0)^2} \right]. \tag{6.1}$$

The skewness-corrected interval is based on the Cornish–Fisher corrected statistic,

$$z_s(\phi) = z(\phi) - \frac{\tilde{\gamma}_1(\phi)(z^2_{\alpha/2} - 1)}{6},$$

where $\tilde{\gamma}_1(\phi)$ is (6.1) with $\tilde{p}_0$ and $\tilde{p}_1$ substituted for p_0 and p_1, respectively. The skewness-corrected limits are the two solutions to the equation

$$z_s(\phi) = \pm z_{\alpha/2}$$

where the plus and minus signs indicate the lower and upper limits, respectively. As noted by Gart (1985a), the secant method may be used to find an iterative solution. This is considered in detail in Section 8. This interval will always be consistent with the skewness-corrected normal deviate test of no difference [see Gart (1985b) and (7.6) below].

Table 4
Skewness-corrected score method: Upper and lower tail probabilities and actual confidence coefficients for nominal 95% intervals

$n_. = 30$:		$n_0 = n_1 = 15$					$n_0 = 20,\ n_1 = 10$				
	$\phi =$	1	2	4	8	14	1	2	4	8	14
$p_0 = .0625$	Lower tail	.5	.3	.0	.0	.0	.7	.8	.5	.0	.0
	Upper tail	.5	1.3	2.3	2.0	1.7	.0	.9	1.7	2.2	2.7
	Conf. coeff.	99.1	98.5	97.7	98.0	98.3	99.3	98.3	97.8	97.8	97.3
	$\phi =$	$\frac{1}{2}$	1	2	4	7	$\frac{1}{2}$	1	2	4	7
$p_0 = .125$	Lower tail	1.3	1.7	2.0	.8	.0	1.9	1.6	2.0	2.8	1.8
	Upper tail	.3	1.7	1.9	2.4	2.8	.0	.8	2.2	2.8	2.7
	Conf. coeff.	98.5	96.6	96.1	96.8	97.2	98.1	97.6	95.8	94.3	95.5
	$\phi =$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$
$p_0 = .250$	Lower tail	2.3	1.9	2.7	2.4	2.2	2.7	2.3	2.7	3.2	1.7
	Upper tail	.0	2.0	2.7	2.2	2.2	.0	.1	3.0	2.3	2.6
	Conf. coeff.	97.7	96.1	94.6	95.3	95.6	97.3	97.6	94.3	94.6	95.9
	$\phi =$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$
$p_0 = .500$	Lower tail	2.0	2.4	2.2	2.3	1.9	2.2	2.7	2.4	2.8	2.3
	Upper tail	.0	.8	2.4	2.3	3.1	.0	.0	2.3	2.8	2.4
	Conf. coeff.	98.0	96.8	95.3	95.4	95.0	97.8	97.3	95.2	94.5	95.3
$n_. = 120$:		$n_0 = n_1 = 60$					$n_0 = 80,\ n_1 = 40$				
	$\phi =$	1	2	4	8	14	1	2	4	8	14
$p_0 = .0625$	Lower tail	2.4	2.5	2.6	2.2	2.1	2.6	2.5	2.3	2.4	2.9
	Upper tail	2.4	2.4	2.5	2.4	2.5	2.0	2.4	2.4	2.3	2.5
	Conf. coeff.	95.3	95.1	94.9	95.3	95.4	95.4	95.1	95.3	95.3	94.6
	$\phi =$	$\frac{1}{2}$	1	2	4	7	$\frac{1}{2}$	1	2	4	7
$p_0 = .125$	Lower tail	2.4	2.4	2.2	2.3	2.3	2.5	2.4	2.5	2.5	2.4
	Upper tail	2.5	2.4	2.7	2.4	2.4	2.3	2.6	2.5	2.6	2.7
	Conf. coeff.	95.1	95.2	95.0	95.3	95.3	95.2	95.0	95.0	95.0	94.9
	$\phi =$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	$3\frac{1}{2}$
$p_0 = .250$	Lower tail	2.5	2.7	2.6	2.6	2.4	2.2	2.3	2.6	2.7	2.3
	Upper tail	2.6	2.2	2.6	2.5	2.5	1.9	2.4	2.5	2.5	2.4
	Conf. coeff.	94.9	95.0	94.8	94.9	95.1	95.8	95.3	95.0	94.8	95.3
	$\phi =$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	$1\frac{3}{4}$
$p_0 = .500$	Lower tail	2.4	2.4	2.5	2.7	2.6	2.3	2.5	2.6	2.3	2.5
	Upper tail	2.2	2.3	2.6	2.7	2.5	1.6	2.6	2.3	2.3	2.8
	Conf. coeff.	95.3	95.3	94.9	94.5	94.9	96.0	95.0	95.1	95.3	94.7

Consider the effect of this correction in our two examples:

	Example 1	**Example 2**
Log method	(.768, 4.65)	(.883, 4.32)
Score method	(.815, 5.34)	(.844, 4.59)
Bailey	(.810, 5.85)	(.880, 4.86)
Bedrick	(.813, 5.74)	(.835, 4.79)
Skewness-corrected	(.806, 6.15)	(.822, 4.95)

Of particular note is the substantial increase in the upper limits of the skewness-corrected method as well as Bailey's and Bedrick's methods over both the log method and the uncorrected score method. In both instances, the correction widens the interval found by the score method.

The actual performance of the correction has been investigated and typical results are given in Table 4. In the top panel wherein $n_1 + n_0 = 30$, it is seen that the tails that are too large in the uncorrected method are usually corrected to the vicinity of 2.5%. However, those that had too small a tail are corrected in the right direction but not enough. Consequently, although the corrected tails are typically closer to the nominal value of 2.5%, the overall confidence coefficient may be larger than that for the uncorrected method. This is particularly true wherever the minimum of $(n_0 p_0, n_0 q_0, n_1 p_1, n_1 q_1)$ is less than 1. For instance, at $\phi = \frac{1}{2}$ where $p_0 = \frac{1}{8}$ and $p_1 = \frac{1}{16}$, the actual confidence coefficient for $n_0 = n_1 = 15$ is 98.5% for the corrected score method while it is 96.3% for the uncorrected version. However, it is useful to note in these sample sizes that the corrected method usually ensures an actual coefficient of at least the nominal value. In only four of the entries wherein $n_. = n_0 + n_1 = 30$ does the actual coefficient fall below the nominal value, and then by no more than .7%.

Table 5

Comparison of Bailey's (1987) *and Bedrick's* (1987) *methods to score and skewness-corrected score method for nominal* 95% *intervals*

p_0	ϕ	n_0	n_1		Score	Bailey[a]	Bedrick[a]	Skewness-corrected score
.2	1	10	25	Lower tail	1.2	.2	2.5	2.5
				Upper tail	3.2	3.0	3.2	3.2
				Conf. coeff.	95.6	96.7	94.3	94.3
.2	4	10	25	Lower tail	.0	.0	.3	.0
				Upper tail	3.6	3.7	3.2	2.7
				Conf. coeff.	96.4	96.3	96.5	97.2
.2	4	25	25	Lower tail	1.9	2.2	2.4	2.4
				Upper tail	3.1	3.1	2.9	2.4
				Conf. coeff.	95.0	94.6	94.7	95.2
.2	4	50	50	Lower tail	1.8	2.3	2.2	2.6
				Upper tail	3.0	2.9	2.7	2.4
				Conf. coeff.	95.3	94.8	95.1	94.9
.041[b]	16	50	50	Lower tail	.0	.0	.0	.0
				Upper tail	4.2	2.2	2.9	2.2
				Conf. coeff.	95.8	97.8	97.1	97.8
.041[b]	16	80	50	Lower tail	.4	.7	3.2	3.2
				Upper tail	3.6	2.3	2.8	2.4
				Conf. coeff.	96.0	97.0	94.0	94.4

[a] *Source*: Bedrick (1987).

[b] Exactly $p_0 = (.65)/16$.

The bottom panel of Table 4 shows results for samples wherein $n_{.} = n_0 + n_1 = 120$. In all these cases, $\min(n_0 p_0, n_0 q_0, n_1 p_1, n_1 q_1) \geqslant 2.5$. Here the maximum discrepancy in the tails is .6% and in the confidence coefficient it is 1.0%. When $\min(n_0 p_0, n_0 q_0, n_1 p_1, n_1 q_1) \geqslant 5$, these maxima are .3 and .4%, respectively.

The log method might be similarly corrected for skewness by using the results of Pettigrew et al. (1986). However, one must first correct the limits for the dependence of the variance on ϕ. Gart and Thomas (1982) explore this correction for the analogous problem of the odds ratio and the logit transformation. They conclude that such a corrected logit method is no better than that based directly on the underlying binomial variates, while being more complicated to compute. A similar conclusion follows in this problem.

Starting with some of the results of Bedrick, Table 5 compares the corrected score method to those of Bailey and Bedrick. Its performance is comparable to that of Bedrick's method, perhaps slightly better. Both of these methods seem to perform better than Bailey's method in assigning roughly equal probabilities to the two tails.

7. Extension of Results to the Stratified Case

Consider pairs of mutually independent binomial variates, x_{0i} and x_{1i}, with associated parameters p_{0i} and p_{1i}, and sample sizes n_{0i} and n_{1i}, respectively, and $\phi = p_{1i}/p_{0i}$, for $i = 1, \ldots, I$. From Gart (1985a), we have

$$S_{\phi \cdot}(\phi, \tilde{p}_{0i}) = \sum_i \frac{x_{1i} - n_{1i}\tilde{p}_{1i}}{\tilde{q}_{1i}\phi},$$

where $\tilde{p}_{1i} = \phi\tilde{p}_{0i}$ and the $\tilde{p}_{0i}$ are the appropriate roots to the quadratic equation [cf. (3.3)],

$$a_i \tilde{p}_{0i}^2 + b_i \tilde{p}_{0i} + c_i = 0, \tag{7.1}$$

where $a_i = n_{.i}\phi$, $b_i = -[(x_{0i} + n_{1i})\phi + x_{1i} + n_{0i}]$, and $c_i = x_{.i}$. From the results of Appendix 2, we have

$$\operatorname{var}[S_{\phi \cdot}(\phi, \tilde{p}_{0i})] = \sum_i \frac{1}{\phi^2 u_i(p_{0i}, p_{1i})} = \sum_i \frac{v_i(\phi, p_{0i})}{\phi^2}, \tag{7.2}$$

where $i = 1, \ldots, I$. The uncorrected limits are based on the asymptotically normal statistic,

$$z_I(\phi) = \frac{\phi S_{\phi \cdot}(\phi, \tilde{p}_{0i})}{[v_{.}(\phi, \tilde{p}_{0i})]^{1/2}} = \sum_i \frac{(x_i - n_{1i}\tilde{p}_{1i})/\tilde{q}_{1i}}{[\sum_i v_i(\phi, \tilde{p}_{0i})]^{1/2}}. \tag{7.3}$$

When $\phi = 1$, $\tilde{p}_{1i} = \tilde{p}_{0i} = \bar{p}_i = x_{.i}/n_{.i}$ and $z_I(1)$ yields the optimal test statistic derived by Radhakrishna (1965) in extending Cochran's (1954) test for multiple two-by-two tables for a constant odds ratio to the assumption of a constant ϕ. The $1 - \alpha$ uncorrected limits are the appropriate roots of the equation

$$z_I(\phi) = \pm z_{\alpha/2},$$

where the plus and minus signs indicate the lower and upper limits, respectively. These roots will always be consistent with the test of Radhakrishna.

The skewness of $z_I(\phi)$ is derived in Appendix 4 to be

$$\gamma_{1I}(\phi) = \frac{\sum_i [q_{1i}(q_{1i} - p_{1i})/(n_{1i}p_{1i})^2 - q_{0i}(q_{0i} - p_{0i})/(n_{0i}p_{0i})^2][v_i(\phi, p_{0i})]^3}{[\sum_i v_i(\phi, p_{0i})]^{3/2}}. \tag{7.4}$$

The skewness-corrected limits are the appropriate roots to the equation

$$z_{Is}(\phi) = z_I(\phi) - \frac{\tilde{\gamma}_{1I}(\phi)(z_{\alpha/2}^2 - 1)}{6} = \pm z_{\alpha/2}, \tag{7.5}$$

where $\tilde{\gamma}_{1I}(\phi)$ is (7.4) evaluated at ϕ by employing the solution to (7.1). The associated skewness-corrected one-tailed version of Radhakrishna's test is found from

$$z_I(1) - \frac{\gamma_{1I}(1)(z_P^2 - 1)}{6} = z_P. \tag{7.6}$$

The appropriate solution z_P is associated with the approximate P-value read from the normal table.

Both Bailey's and Bedrick's methods may also be extended to the stratified case. Each contains an arbitrary constant, $\frac{1}{3}$ and $\frac{1}{2}$ respectively, chosen on empirical grounds. It is not clear whether these choices also lead to reasonable methods in the stratified case.

8. Iterative Algorithm for Computing Point and Interval Estimates

Although the uncorrected confidence limits for the single-stratum case may be found by solving a cubic equation, the skewness-corrected limits and multiple-strata case require an iterative solution. Easily computed initial estimates are based on the inefficient, but nearly efficient point estimator of Tarone (1981) [see also Tarone, Gart, and Hauck (1983)],

$$\hat{\phi}_{\mathrm{B}} = \frac{\sum_i [(n_{0i}x_{1i})/(n_{.i} - x_{.i})]}{\sum_i [(n_{1i}x_{0i})/(n_{.i} - x_{.i})]}.$$

As $\mathrm{var}(\hat{\phi}_{\mathrm{B}})$ is approximately $\phi^2 \sum_i [1/u_i(p_{0i}, p_{1i})]$, initial estimates of the upper and lower limits, $\hat{\phi}_{\mathrm{U}}$ and $\hat{\phi}_{\mathrm{L}}$, may be found from

$$\log(\phi_{\mathrm{U,L}}) = \left[1 \pm \frac{z_{\alpha/2}}{\{\sum_i [1/u_i(p_{0i}, p_{1i})]\}^{1/2}}\right]\log(\phi_{\mathrm{B}}),$$

where the plus and minus signs indicate the upper and lower limits, respectively. The iteration proceeds by taking a second value of ϕ in the vicinity of the limit in question. Say, for the lower, let the two estimates be $\hat{\phi}_{\mathrm{L}}$ and $.9\hat{\phi}_{\mathrm{L}}$, compute $z_I(\hat{\phi}_{\mathrm{L}})$ and $z_I(.9\hat{\phi}_{\mathrm{L}})$ from (7.3) [or $z_{Is}(\phi)$ from the skewness-corrected limits]. Whichever value is closest to $z_{\alpha/2}$ designate as ϕ_1, and call the other ϕ_0. A new value of ϕ, ϕ_2, is found by applying the secant method in the log-scale of ϕ,

$$\log(\phi_2) = \log(\phi_0) + \frac{z_{\alpha/2} - z_I(\phi_0)}{z_I(\phi_1) - z_I(\phi_0)} \log\left(\frac{\phi_1}{\phi_0}\right), \tag{8.1}$$

calculate $z(\phi_2)$ and use it together with ϕ_1 to repeat the process until convergence is obtained. For the upper limit $-z_{\alpha/2}$ is substituted in (8.1).

The MLE of ϕ, $\tilde{\phi}_{\mathrm{ML}}$, may be found from (8.1) by letting $\alpha/2 = .5$ or $z_{\alpha/2} = 0$ and solving iteratively starting with $\phi = \hat{\phi}_{\mathrm{B}}$. A test for homogeneity of ϕ over the tables may be based on the approximate chi-square with $I - 1$ degrees of freedom,

$$\chi^2(I - 1) = \sum_i z_i^2(\tilde{\phi}_{\mathrm{ML}}),$$

where

$$z_i(\tilde{\phi}_{\mathrm{ML}}) = \frac{x_{1i} - n_{1i}\tilde{p}_{1i}}{q_i[v_i(\tilde{\phi}_{\mathrm{ML}}, \tilde{p}_{0i})]^{1/2}}.$$

See Tarone (1988) for the theoretical basis of this test.

Example 3 Table 6 shows the results of these methods to the example analyzed in Gart (1985a). The skewness-corrected 95% limits are again seen to be wider than the uncorrected limits. The correctness to the 99% limits are even greater. Note that the consistency of the

Table 6

Analyses of data on the numbers of pulmonary tumors among four sex-strain combinations of mice

i	=	1	2	3	4
x_{1i}/n_{1i}		4/16	2/16	4/18	1/15
x_{0i}/n_{0i}		5/79	3/87	10/90	3/82
$\bar{p}_i = x_{.i}/n_{.i}$		9/95	5/103	14/108	4/97

Analysis	Initial estimates	ML-score	Skewness-corrected
Test of $\phi = 1$		$z_I(1) = 2.88$, $P = .002$	$z_P = 2.59$, $P = .004$
Point estimates	$\hat{\phi}_B = 2.66$	$\tilde{\phi}_{ML} = 2.65$	
95% intervals	(1.37, 5.18)	(1.35, 5.03)	(1.31, 5.08)
99% intervals	(1.11, 6.38)	(1.10, 6.07)	(1.02, 6.24)
Test of homogeneity of ϕ: $\chi^2(3) = .95$, $P = .81$			

skewness-corrected limits and Radhakrishna's test is emphasized by the fact that the lower limit is 1.02 and the P-value is .004. For a 99.2% interval, the corrected method would yield a lower limit of unity.

9. Final Comments

The numerical results indicate that the skewness-corrected score method is quite adequate for 95% limits when $\min(n_0 p_0, n_0 q_0, n_1 p_1, n_1 q_1) \geqslant 2.5$. Of course, in an actual set of data we do not know the p's and q's. We suggest that their maximum likelihood estimators be used to evaluate this criterion at each limit. If the criterion is not met, Santner and Snell's exact method might be considered at that limit. It appears that Bedrick's limit performs almost as well in a single two-by-two table, but it has not been shown that its good properties carry over to the stratified case. Bailey's method, while perhaps not performing quite so well, is easily computable and is preferred to the log method as an initial estimator for either of these iterative methods.

As shown in Gart (1985b), the score method for the common odds ratio, $\psi = (p_{1i}q_{0i})/(p_{0i}q_{1i})$, $i = 1, \ldots, I$, closely parallels the method described here. That method requires simultaneous solution of quadratic equations having different coefficients from (7.1). It differs from the present result in that the first-order bias is not identically zero. However, as the $q_{ji} \to 1$ for $j = 0, 1$ and $i = 1, \ldots, I$, $\psi \to \phi$ and the two methods yield results identical with those for Poisson variates and ratios [see, e.g., Gart (1978)].

The score method may be used to find similar results for a constant difference in proportions, that is, $\Delta = p_{1i} - p_{0i}$, $i = 1, \ldots, I$. For this parameter, the method depends on the simultaneous solution to a series of cubic equations; see Mee (1984) and Miettinen and Nurminen (1985) for the single two-by-two table case.

ACKNOWLEDGEMENTS

The authors appreciate the helpful comments of the referees and also thank Professors Bedrick and Bailey for providing us with prepublication copies of their manuscripts.

RÉSUMÉ

Différentes méthodes pour trouver des intervalles de confiance pour le rapport des paramètres de lois binomiales sont rappelées et évaluées numériquement. On montre que la méthode basée sur les "scores de vraisemblance" (Koopman, 1984, *Biometrics* **40,** 513–517; Miettinen et Nurminen, 1985, *Statistics in Medicine* **4,** 213–226) est la plus adaptée pour obtenir le degré de confiance nominal; mais elle peut conduire à une forte dissymétrie. Utilisant une théorie générale due à Bartlett (1953, *Biometrika* **40,** 306–317; 1955, *Biometrika* **42,** 210–203), nous corrigeons cette méthode de l'asymétrie. Suivant Gart (1985, *Biometrika* **72,** 673–677), nous étendons cette correction au cas de l'estimation d'un rapport commun dans une série de tables 2 × 2. Des algorithmes sont donnés et appliqués à des exemples numériques. On indique des méthodes semblables pour le "odds-ratio" et le rapport de paramètres de lois de Poisson.

References

Bailey, B. J. R. (1987). Confidence limits to the risk ratio. *Biometrics* **43,** 201–205.

Bartlett, M. S. (1953). Approximate confidence intervals, II. More than one unknown parameter. *Biometrika* **40,** 306–317.

Bartlett, M. S. (1955). Approximate confidence intervals, III. A bias correction. *Biometrika* **42,** 201–203.

Bedrick, E. J. (1987). A family of confidence intervals for the ratio of two binomial proportions. *Biometrics* **43,** 993–998.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10,** 417–451.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B* **46,** 440–464.

Gart, J. J. (1978). The analysis of ratios and cross-product ratios of Poisson variates with application to incidence rates. *Communications in Statistics* **A7,** 917–937.

Gart, J. J. (1985a). Approximate tests and interval estimation of the common relative risk in the combination of 2 × 2 tables. *Biometrika* **72,** 673–677.

Gart, J. J. (1985b). Analysis of the common odds ratio: Corrections for bias and skewness. *Bulletin of the International Statistical Institute*, 45th Session, Book 1, 175–176.

Gart, J. J. and Thomas, D. G. (1982). The performance of three approximate confidence limit methods for the odds ratio. *American Journal of Epidemiology* **115,** 453–470.

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.

Koopman, P. A. R. (1984). Confidence limits for the ratio of two binomial proportions. *Biometrics* **40,** 513–517.

Mee, R. W. (1984). Confidence bounds for the difference between two probabilities (letter). *Biometrics* **40,** 1175–1176.

Miettinen, O. and Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine* **4,** 213–226.

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.

Pettigrew, H. M., Gart, J. J., and Thomas, D. G. (1986). The bias and higher cumulants of the logarithm of a binomial variate. *Biometrika* **73,** 425–435.

Radhakrishna, S. (1965). Combination of results from several 2 × 2 contingency tables. *Biometrics* **21,** 86–98.

Santner, T. J. and Snell, M. K. (1980). Small-sample confidence intervals for $p_1 - p_2$ and p_1/p_2 in 2 × 2 contingency tables. *Journal of the American Statistical Association* **73,** 386–394.

Tarone, R. E. (1981). On summary estimators of relative risk. *Journal of Chronic Diseases* **34,** 463–468.

Tarone, R. E. (1988). Homogeneity score tests with nuisance parameters. *Communications in Statistics—Theory and Methods* **17,** in press.

Tarone, R. E., Gart, J. J., and Hauck, W. W. (1983). On the asymptotic inefficiency of certain noniterative estimators of a common relative risk or odds ratio. *Biometrika* **70,** 519–522.

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

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

Walter, S. D. (1976). The estimation and interpretation of attributable risk in health research. *Biometrics* **32,** 829–849.

Received May 1986; *revised March and October* 1987.

Appendix 1

Proof That $\chi^2_{\mathrm{MN}}(\phi) = \chi^2_{\mathrm{K}}(\phi)$

Multiplying the numerator and denominator of $\chi^2_{\mathrm{MN}}(\phi)$ by $(n_1/\tilde{p}_1)^2$ yields

$$\chi^2_{\mathrm{MN}}(\phi) = \frac{(x_1 - n_1\phi\hat{p}_0)^2}{(n_1\tilde{p}_1)^2 u(\tilde{p}_0, \tilde{p}_1)}. \tag{A.1}$$

Setting $\partial L(\phi, p_0)/\partial p_0 = 0$ yields [see (3.2)] the relation

$$\frac{x_0 - n_0\tilde{p}_0}{\tilde{q}_0} = -\frac{(x_1 - n_1\tilde{p}_1)}{\tilde{q}_1},$$

from which we have

$$-n_1\phi\hat{p}_0 = -n_1\tilde{p}_1 + \frac{n_1\tilde{q}_0\phi(x_1 - n_1\tilde{p}_1)}{n_0\tilde{q}_1}.$$

Substituting this into (A.1) and multiplying the resulting numerator and denominator by $\tilde{q}_1^2/(n_1\tilde{p}_1)^2$ yields $\chi^2_{\mathrm{K}}(\phi)$.

Appendix 2

General Results for the Score Method

The log-likelihood is (see Gart, 1985a),

$$L_.(\phi, p_{0i}) = \sum_i L_i(\phi, p_{0i}),$$

where

$$L_i(\phi, p_{0i}) = x_{1i}\log(\phi) + (n_{1i} - x_{1i})\log(1 - \phi p_{0i}) + x_{.i}\log(p_{0i}) + (n_{0i} - x_{0i})\log(1 - p_{0i}),$$

for $i = 1, \ldots, I$. From this we find

$$S_{\phi i}(\phi, p_{0i}) = \frac{\partial L_i(\phi, p_{0i})}{\partial\phi} = \frac{x_{1i} - n_{1i}p_{1i}}{\phi q_{1i}}$$

and

$$S_{0i}(\phi, p_{0i}) = \frac{\partial L_i(\phi_i, p_{0i})}{\partial p_{0i}} = \frac{\phi(x_{1i} - n_{1i}p_{1i})}{p_{1i}q_{1i}} + \frac{x_{0i} - n_{0i}p_{0i}}{p_{0i}q_{0i}}.$$

The corresponding elements of the information matrix are

$$I_{\phi\phi}(\phi, p_{0i}) = \sum_i \left(\frac{n_{1i}p_{1i}}{\phi^2 q_{1i}}\right),$$

$$I_{\phi i}(\phi, p_{0i}) = n_{1i}/q_{1i},$$

$$I_{ii}(\phi, p_{0i}) = \frac{n_{1i}\phi^2}{p_{1i}q_{1i}} + \frac{n_{0i}}{p_{0i}q_{0i}},$$

and

$$I_{ij}(\phi, p_{0i}) = 0 \quad \text{for } i \neq j = 1, \ldots, I.$$

The score method is based on the asymptotically normal statistic

$$S_{\phi.}(\phi, \tilde{p}_{0i}) = \sum_i S_{\phi i}(\phi, \tilde{p}_{0i}),$$

where $\tilde{p}_{0i}$, $i = 1, \ldots, I$, are the simultaneous solutions for a given value of ϕ, of the I quadratic equations [see (7.1)] derived from

$$S_{0i}(\phi, \tilde{p}_{0i}) = 0 \quad \text{for } i = 1, \ldots, I.$$

The asymptotic mean of this statistic is zero and its variance is estimated by

$$I_{\phi\phi}(\phi, \tilde{p}_{0i}) - \sum_i \frac{[I_{\phi i}(\phi, \tilde{p}_{0i})]^2}{I_{ii}(\phi, \tilde{p}_{0i})},$$

which is explicitly $[\sum_i v_i(\phi, \tilde{p}_{0i})]/\phi^2$.

Appendix 3

Bias of the Test Statistic

Although the asymptotic mean of $z_I(\phi)$ is zero, it is possible for the score statistic derived by this theory (Bartlett, 1955) to have means of order $O(n_{ij}^{-1/2})$. In this case, it is sufficient, in order to prove that the terms of this order vanish, to show that

$$\mathrm{E}[\partial^3 L_i(\phi, p_{0i})/\partial\phi\partial p_{0i}^2] + 2\partial I_{\phi i}(\phi, p_{0i})/\partial p_{0i} = 0$$

and

$$\mathrm{E}[\partial^3 L_i(\phi, p_{0i})/\partial p_{0i}^3] + 2\partial I_{ii}(\phi, p_{0i})/\partial p_{0i} = 0$$

for $i = 1, \ldots, I$. Straightforward mathematics show that

$$\mathrm{E}[\partial^3 L_i(\phi, p_{0i})/(\partial\phi\partial p_{0i}^2)] = -2n_{1i}\phi/q_{1i}^2 = -2\partial I_{\phi i}(\phi, p_{0i})/\partial p_{0i}$$

and

$$\mathrm{E}[\partial^3 L_i(\phi, p_{0i})/\partial p_{0i}^3] = \frac{2n_{0i}(q_{0i}-p_{0i})}{(p_{0i}q_{0i})^2} + \frac{2n_{1i}\phi^3(q_{1i}-p_{1i})}{(p_{1i}q_{1i})^2} = -2\partial I_{ii}(\phi, p_{0i})/\partial p_{0i}$$

for $i = 1, \ldots, I$. Thus, the bias of order $O(n_{ji}^{-1/2})$ is zero.

Appendix 4

Skewness of $Z_I(\phi)$

In finding the third moment of $z_I(\phi)$, Bartlett (1953) considers the statistic

$$T_I(\phi) = S_{\phi\cdot}(\phi, p_{0i}) - \sum_i \frac{I_{\phi i}(\phi, p_{0i})S_{0i}(\phi, p_{0i})}{I_{ii}(\phi, p_{0i})} = \sum_i \left[v_i(\phi, p_{0i})\left(\frac{x_{1i} - n_{1i}p_{1i}}{n_{1i}\phi p_{1i}} - \frac{x_{0i} - n_{0i}p_{0i}}{n_{0i}p_{1i}}\right)\right].$$

To this order of approximation, $T_I(\phi)$ has the same second and third moments as $S_{\phi\cdot}(\phi, \tilde{p}_{0i})$. Thus,

$$\mathrm{var}[S_{\phi\cdot}(\phi, \tilde{p}_{0i})] = \mathrm{var}[T_I(\phi)] = \sum_i \left\{[v_i(\phi, p_{0i})]^2\left[\frac{\mu_2(x_{1i})}{(n_{1i}\phi p_{1i})^2} + \frac{\mu_2(x_{0i})}{(n_{0i}p_{1i})^2}\right]\right\},$$

and

$$\mu_3[S_{\phi\cdot}(\phi, \tilde{p}_{0i})] = \mu_3[T_I(\phi)] = \sum_i \left\{[v_i(\phi, p_{0i})]^3\left[\frac{\mu_3(x_{1i})}{(n_{1i}\phi p_{1i})^3} - \frac{\mu_3(x_{0i})}{(n_{0i}p_{1i})^3}\right]\right\},$$

where the second and third central moments of the binomial variates are

$$\mu_2(x_{ji}) = n_{ji}p_{ji}q_{ji}$$

and

$$\mu_3(x_{ji}) = n_{ji}^2(p_{ji}q_{ji})(q_{ji} - p_{ji}), \quad \text{for } j = 0, 1; \quad i = 1, \ldots, I.$$

Thus, we find

$$\mathrm{var}[S_{\phi\cdot}(\phi,\ \tilde{p}_{0i})] = \frac{[\sum_i v_i(\phi,\ p_{0i})]}{\phi^2},$$

as given in (7.2) and

$$\gamma_1[S_{\phi\cdot}(\phi,\ \tilde{p}_{0i})] = \frac{\mu_3[T_I(\phi)]}{\{\mathrm{var}[T_I(\phi)]\}^{3/2}} = \gamma_1([z(\phi)],$$

which yields (7.4).