BIOMETRICS 55, 597–602
June 1999

On Logit Confidence Intervals for the Odds Ratio with Small Samples

Alan Agresti

Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, U.S.A.
email: aa@stat.ufl.edu

SUMMARY. Unless the true association is very strong, simple large-sample confidence intervals for the odds ratio based on the delta method perform well even for small samples. Such intervals include the Woolf logit interval and the related Gart interval based on adding .5 before computing the log odds ratio estimate and its standard error. The Gart interval smooths the observed counts toward the model of equiprobability, but one obtains better coverage probabilities by smoothing toward the independence model and by extending the interval in the appropriate direction when a cell count is zero.

KEY WORDS: Bayes estimate; Chi-squared test; Contingency table; Delta method; Exact inference; Two-by-two table.

1. Introduction

One of the most important parameters in the study of contingency tables is the odds ratio. For a 2×2 contingency table with expected frequencies $\{\mu_{ij}\}$, the true odds ratio is $\theta = (\mu_{11}\mu_{22})/(\mu_{12}\mu_{21})$. The usual estimators of the odds ratio have the form $\hat{\theta} = (\hat{\mu}_{11}\hat{\mu}_{22})/(\hat{\mu}_{12}\hat{\mu}_{21})$ for estimates $\{\hat{\mu}_{ij}\}$ of the expected frequencies. For cell counts $\{n_{ij}\}$ obtained with Poisson, multinomial, or independent binomial sampling, many estimators of $\hat{\theta}$ use $\hat{\mu}_{ij} = n_{ij} + c$ for some nonnegative constant c. Denote this estimator by $\hat{\theta}_c$. The unconditional maximum likelihood (ML) estimator has $c = 0$, whereas using $c > 0$ ensures that undefined estimates or estimates on the boundary of the parameter space cannot occur. The most popular estimator of this form, with $c = .5$ (Haldane, 1956; Anscombe, 1956), achieves bias of order $O(n^{-2})$ for large samples in the estimation of $\log(\theta)$ by $\log(\hat{\theta}_c)$.

For large samples, by the delta method (Bishop, Fienberg, and Holland, 1975, Section 14.6), the estimator $\log(\hat{\theta}_c)$ of $\log(\theta)$ is approximately normal with asymptotic standard error that can be estimated by

$$\hat{\sigma}\left(\log \hat{\theta}_c\right) = \sqrt{\frac{1}{n_{11}+c} + \frac{1}{n_{12}+c} + \frac{1}{n_{21}+c} + \frac{1}{n_{22}+c}}.$$

Thus, an approximate $100(1-\alpha)\%$ confidence interval for θ results from exponentiating the end-points of

$$\log\left(\hat{\theta}_c\right) \pm z_{\alpha/2}\hat{\sigma}\left(\log \hat{\theta}_c\right), \qquad (1)$$

where $z_{\alpha/2}$ is the $\alpha/2$ standard normal quantile. This interval, often referred to as the logit interval because of the expression of $\log(\hat{\theta}_c)$ as the difference between two empirical logits, was proposed by Woolf (1955) for $c = 0$ and by Gart (1966) for $c = .5$. Several related intervals have been proposed (cf., Gart, 1971). I refer to the logit interval with $c = 0$ as the Woolf interval and the case $c = .5$ as the Gart interval (though perhaps Haldane–Anscombe–Gart would be more appropriate).

Many textbooks state that logit intervals behave poorly for small samples in the direction of having actual coverage probabilities too low (e.g., Breslow and Day, 1980, p. 134; Fleiss, 1981, p. 74; Agresti, 1990, p. 69). This apparently dates back to evaluations by Gart and Thomas (1972), who claimed that the logit interval was much too narrow, and to Woolf's statement in proposing (1) for $c = 0$ that "(t)his is a 'large-sample' treatment and the formulae cease to be applicable if any of the observed frequencies is small," (1995, p. 252). However, the Gart and Thomas (1972) remark referred to comparing the logit interval with an interval based on the exact conditional approach, which is necessarily conservative because of discreteness (Neyman, 1935), rather than comparing attained coverage probabilities to the nominal confidence level. This exact confidence interval for θ, proposed by Cornfield (1956), consists of the collection of θ_0 values for which the P-value exceeds $\alpha/2$ in conducting each exact one-sided test of H_0: $\theta = \theta_0$ using the nonnull hypergeometric conditional distribution derived from conditioning on row and column totals. The exactness refers to the conditional distribution being free of nuisance parameters. With Cornfield's method, the actual confidence coefficient, defined as the infimum of the coverage probabilities for all possible parameter values, has the nominal confidence level as a lower bound.

This paper reports the behavior of the coverage probabilities of the logit confidence interval. Unless the true odds ratio is very large, this method seems to be acceptable even for small sample sizes. An improved logit interval results from using cell probability estimates based on a smoothing different than the crude one of adding c to each cell.

2. Comparisons of Coverage Probabilities and Lengths

The coverage probabilities and expected lengths of the Woolf and Gart versions of the logit confidence interval were evaluated for a variety of sampling methods, true probabilities, small sample sizes, and nominal confidence coefficients. Tables 1 and 2 show typical results. Table 1 refers to nominal 95% confidence intervals when the data result from two independent binomial samples of sizes $n_1 = n_2 = 10$ with various choices for the binomial parameters p_1 and p_2. Here, $\theta = [p_1/(1-p_1)]/[p_2/(1-p_2)]$. When any cell count equals zero, we take the Woolf interval to be the entire real line. All expected lengths are reported conditional on the set of tables for which all counts are positive; the coverage probabilities, however, are based on all tables. The tables report expected log lengths, which are the same when values of p_1 and p_2 are interchanged. However, somewhat modest differences in expected log lengths can translate to substantively important differences on the scale actually used for interpretation, especially when the true log odds ratio is large. For example, when $p_1 = .4$ and $p_2 = .1$, the expected lengths of the confidence intervals for the odds ratio itself are 59.9 for the Woolf interval, 31.8 for the Gart interval, and 250.8 for the exact interval.

Table 1 reveals that, considering the small sample size, the logit methods behave surprisingly well for these parameter settings. Their coverage probabilities exceed the nominal confidence level. The expected lengths are smaller for the logit intervals than for Cornfield's exact interval, reflecting the conservativeness of exact conditional methods for small samples. (The exact intervals were obtained using StatXact [Cytel, 1995]). Similar results hold in other highly discrete problems, even when conditioning is not involved (e.g., Agresti and Coull, 1998). Since our main focus in this note is on the performance of the delta method-based logit formula (1), we will not further consider exact approaches.

Table 2 evaluates the coverage probabilities of the nominal 95% logit intervals for the somewhat larger binomial sample size cases $(n_1, n_2) = (20, 20)$ and $(30, 10)$. Except in one case with the Gart interval, coverage probabilities again exceed the nominal confidence level. With both of these logit intervals, underestimation of $|\log\theta|$ (the true log odds ratio falling farther from zero than the interval bounds) is much more common than overestimation, the discrepancy being somewhat larger for the Gart interval. For the eight cases summarized in Table 1 for which $p_1 \neq p_2$, e.g., the average probability of underestimation was .013 for the Woolf interval and .018 for the Gart interval and the average probability of overestimation was .002 for the Woolf interval and .001 for the Gart interval.

To investigate the logit intervals' performance over a much wider range of possible parameter values, coverage probabilities were calculated (using Splus) using each of a random sample of 10,000 pairs of independent binomial parameters from the uniform distribution over the unit square. Table 3 summarizes the results when $n_1 = n_2 = 10$. It evaluates the nominal 95% versions of the Woolf and Gart logit methods by reporting mean coverage probabilities, mean expected log lengths, the mean absolute distance of the coverage probability from .95, and the proportion of cases for which the coverage probability falls at least .03 below that nominal value. Expected lengths were calculated conditional on all cell counts being positive, but coverage probability measures again used the entire distribution. (Table 3 also summarizes performance of the exact confidence interval, and some methods discussed in Sections 3 and 4.)

Table 3 also reports minimum coverage probabilities for each method over the entire set of 10,000 odds ratios considered. To illustrate the behavior for the portion of the parameter space usually dealt with in practice, Table 3 also reports the minimum coverage probabilities for the (4909, 6519, 7739) of the 10,000 probability pairs in which the true odds ratio and its inverse have magnitude less than (5, 10, 20). The Woolf logit interval behaved well, and logit intervals with $c > 0$ be-

Table 1
Coverage probabilities and expected lengths of nominal 95% confidence intervals for the log odds ratio for binomial samples with $n_1 = n_2 = 10$ and parameters p_1 and p_2

		Coverage probability			Expected length		
p_1	p_2	Woolf	Gart	Exact	Woolf	Gart	Exact
.5	.5	.961	.960	.986	3.7	3.5	4.5
.5	.3	.973	.978	.982	3.9	3.6	4.9
.5	.1	.971	.959	.991	4.5	4.0	6.0
.4	.4	.967	.965	.984	3.8	3.6	4.7
.4	.2	.984	.979	.981	4.2	3.8	5.5
.4	.1	.975	.973	.990	4.6	4.1	6.1
.3	.3	.983	.980	.977	4.1	3.8	5.3
.3	.1	.988	.980	.990	4.7	4.1	6.4
.2	.2	.996	.995	.974	4.6	4.1	6.3
.2	.1	.996	.995	.990	4.9	4.3	6.9
.2	.05	.993	.984	.987	5.1	4.4	7.3
.1	.1	.9999	.9998	.991	5.2	4.5	7.5
.1	.05	.9996	.9992	.996	5.4	4.6	7.9

Note: Gart interval is the logit interval based on adding .5 to all cells for estimate and standard error. Expected lengths are conditional on all counts being positive.

Table 2
Coverage probabilities of nominal 95% logit confidence intervals for the odds ratio for binomial samples with $(n_1, n_2) = (30, 10)$ and $(20, 20)$ and parameters p_1 and p_2.

		(30, 10)		(20, 20)	
p_1	p_2	Woolf	Gart	Woolf	Gart
.5	.5	.967	.968	.961	.961
.5	.3	.975	.973	.959	.957
.5	.1	.971	.954	.972	.963
.4	.4	.969	.971	.960	.964
.4	.2	.974	.964	.967	.967
.4	.1	.969	.955	.970	.965
.3	.3	.971	.972	.959	.968
.3	.1	.971	.958	.976	.974
.2	.2	.971	.971	.977	.978
.2	.1	.974	.957	.983	.982
.1	.1	.984	.981	.998	.998
.2	.05	.971	.941	.978	.975
.1	.05	.980	.960	.995	.993

haved well except when the true log odds ratio was very large. Similar results occurred for other sample sizes considered in this study. To illustrate, Table 4 shows results for uniform random generation of binomial probabilities for some other sample size combinations. In the random sample size case, for each of the 10,000 binomial probability combinations, the two sample sizes were chosen uniformly and independently between 5 and 50. The Woolf interval performed very well when the true odds ratio is less than 10. The Gart interval performed well when the true odds ratio is less than five and the samples sizes are not highly unbalanced; e.g., in those 2932 of the random sample size cases in which the odds ratio is less than 5 and $.5 \le n_1/n_2 \le 2$, the minimum coverage probability was .947. Consistently in these analyses, the probability of underestimation of $|\log\theta|$ exceeded the probability of overestimation, the discrepancy between the two being greater for the Gart interval. For instance, for the random sample size case, the average probability of underestimation was .023 for the Woolf interval and .039 for the Gart interval and the average probability of overestimation was .008 for the Woolf interval and .007 for the Gart interval. Similar results as those displayed in Tables 1–4 occurred in evaluations conducted with 99% confidence coefficients.

So, what disadvantage is there in using the logit intervals? Not apparent from Tables 1 and 2 is a deficiency that the logit intervals using $\{\hat{\mu}_{ij} = n_{ij} + c\}$ with fixed $c > 0$ necesssarily have. For any such interval with given n_1 and n_2, there exists $\theta_{L0} < \theta_{U0}$ such that, for all $\theta < \theta_{L0}$ and for all $\theta > \theta_{U0}$, the actual coverage probability equals zero. For instance, one can simply take θ_{U0} to be the maximum of the method's upper confidence limits for all tables with those sample sizes. This is the upper confidence limit for the table with entries $(n_1, 0)$ in row 1 and $(0, n_2)$ in row 2 since that table has the largest estimated odds ratio and since one can easily show that that table and the one with entries $(0, n_1)$ in row 1 and $(n_2, 0)$ in row 2 have the greatest width. In this sense, all logit intervals with $c > 0$ have actual confidence coefficient (infimum of coverage probabilities) of 0. In addition, this property contributes toward the imbalance between probabilities of underestimation and overestimation.

For the case $n_1 = n_2 = 10$, Figure 1 shows the coverage probability for the Gart logit interval plotted against the absolute value of the logarithm of the odds ratio for the 10,000 pairs of binomial probabilities. The coverages behave very well until $|\log\theta| > 2.4$ (odds ratio exceeds about 11), with serious cases of deterioration occurring when $|\log\theta| > 4$ (odds ratio exceeds about 55). A similar figure occurs for any such interval with $c > 0$, with the deterioration starting at smaller log odds ratio values for estimates employing greater shrinkage. Figure 1 also shows the coverage probability of the Woolf interval. It behaves better, of course, since not replacing 0 counts by .5 implies that the interval can contain indefinitely large true log odds ratios.

Table 3
Summary of results for nominal 95% confidence intervals for odds ratio, applied with 10,000 randomly generated pairs of binomial distributions, when $n_1 = n_2 = 10$

	Method						
Criterion	Woolf	Gart	Indep.	Exact	Mid-P	Pearson	Like.
Mean coverage probability	.977	.956	.969	.986	.970	.955	.944
Mean expected log length	4.54	4.03	4.17	5.92	5.11	4.14	4.83
Mean |coverage probability − .95|	.027	.030	.024	.036	.020	.012	.023
Proportion coverage probability < .92	.000	.077	.025	.000	.000	.002	.173
Minimum coverage probability	.941	.000	.866	.970	.938	.910	.836
Mininimum coverage probability, odds ratio < 5	.949	.956	.958	.970	.945	.910	.910
Minimum coverage probability, odds ratio < 10	.949	.952	.956	.970	.940	.910	.900
Minimum coverage probability, odds ratio < 20	.949	.908	.945	.970	.938	.910	.888

Note: Indep., independence-smoothed logit interval; Pearson, inversion of Pearson chi-squared test; Like., profile likelihood interval based on inverting likelihood-ratio chi-squared test. Expected lengths are conditional on positive cell counts.

Table 4
Summary of results for nominal 95% confidence intervals for odds ratio θ, applied with 10,000 randomly generated pairs of binomial distributions

	$n_1 = n_2 = 20$			$n_1 = 30,\ n_2 = 10$			Random $5 \le n_i \le 50$		
Criterion	Woolf	Gart	Indep.	Woolf	Gart	Indep.	Woolf	Gart	Indep.
Mean coverage probability	.969	.954	.963	.973	.954	.968	.969	.954	.964
Mean expected log length	3.49	3.18	3.28	3.83	3.45	3.64	3.39	3.08	3.22
Mean |coverage probability − .95|	.019	.021	.016	.023	.026	.021	.019	.021	.017
P(coverage probability < 92)	0	.057	.024	0	.067	.016	0	.058	.019
Minimum coverage probability	.930	0	.863	.927	0	.889	.925	0	.867
Minimum coverage probability, $\theta < 5$	.943	.944	.944	.958	.909	.958	.943	.896	.947
Minimum coverage probability, $\theta < 10$	.943	.944	.944	.951	.880	.946	.943	.886	.946
Minimum coverage probability, $\theta < 20$	.943	.891	.944	.936	.871	.925	.932	.863	.929

Note: Expected lengths are conditional on positive cell counts.

3. Confidence Intervals Based on Smoothing Toward Independence

So far, this paper has merely reported the performance of logit intervals rather than suggested new interval estimators. There seems to be scope for the development of simple confidence intervals that are much shorter than the exact interval but have better coverage performance than the logit intervals, and this is a useful topic for future research. For instance, intervals that use a variance estimate for the logit estimator of the log odds ratio that reduces the contribution of small cell counts (e.g., Bedrick, 1984) may provide improved results. In addition, just because the Gart estimator $\log(\hat{\theta}_{.5})$ has small bias in estimating $\log(\theta)$ does not mean that that estimator should fall in the middle of a confidence interval for $\log(\theta)$.

The logit interval (1) adding c to each cell count corresponds to using an effective sample size of $n + 4c$ and cell proportion estimates

$$\hat{p}_{ij} = \frac{n}{n+4c}\left(\frac{n_{ij}}{n}\right) + \frac{4c}{n+4c}(.25)$$

that smooth toward the model of equiprobability. As pointed out by Bishop et al. (1975, p. 421), it is usually more sensible to smooth toward the model of independence, with cell proportion estimates such as

$$\tilde{p}_{ij} = \frac{n}{n+4c}\left(\frac{n_{ij}}{n}\right) + \frac{4c}{n+4c}\left(\frac{n_{i+}n_{+j}}{n^2}\right)$$

in which relatively more smoothing (in absolute terms) occurs for cells with larger row and column totals. If one uses $c =$ .5 and an effective sample size of $n + 2$, as in the Gart interval, then one replaces the count n_{ij} in the log odds ratio and standard error formulas by $n_{ij} + 2(n_{i+}n_{+j}/n^2)$ rather than n_{ij}+.5. For $n_1 = n_2 = 10$, the first panel in Figure 2 shows the coverage probability for this independence-smoothed logit interval. Serious deterioration does not occur until the log odds ratio exceeds about 7.5, compared to 4.0 for the Gart interval (compare to the panel in Figure 1 for the Gart interval).

A second potential improvement relates to the comment at the end of the last section about a deficiency of the logit intervals. If any $n_{ij} = 0$, the Woolf interval is completely uninformative, yet the Gart interval and others with $c > 0$ have the disadvantage of ruling out sufficiently large log odds ratio values. It seems more sensible to modify logit intervals for the log odds ratio to have a lower endpoint of $-\infty$ if and only if $\min(n_{11}, n_{22}) = 0$ and an upper endpoint of ∞ if and only if $\min(n_{12}, n_{21}) = 0$ since indefinitely large log odds ratio values are completely consistent with such data.

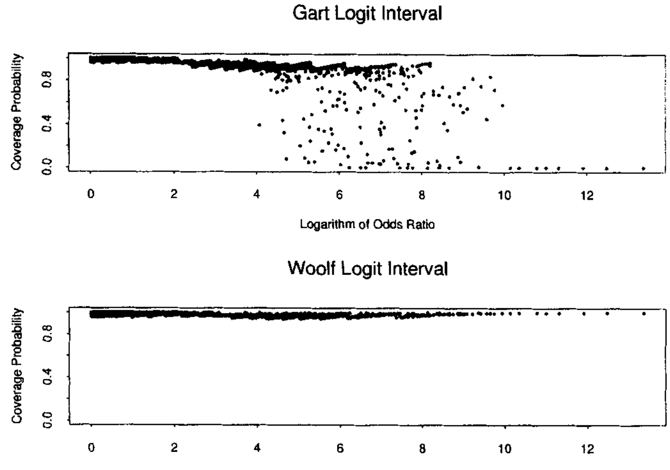

Figure 1. Coverage probabilities for Gart and Woolf logit intervals for 10,000 pairs of binomials with $n_1 = n_2 = 10$.

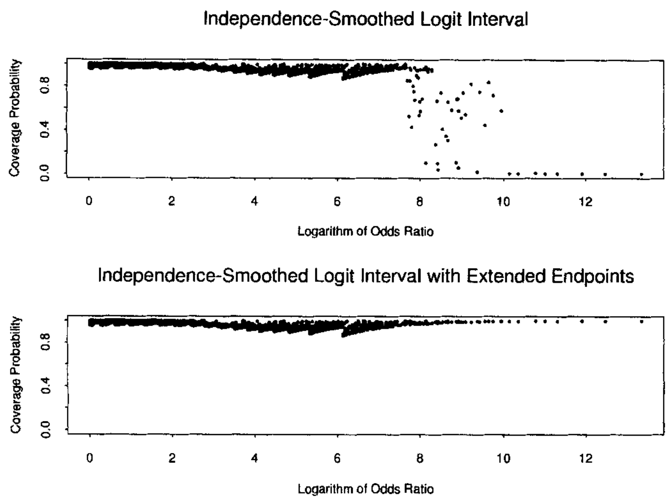

Figure 2. Coverage probabilities for independence-smoothed logit intervals for 10,000 pairs of binomials with $n_1 = n_2 = 10$.

Tables 3 and 4 show results (labeled Indep.) of using the logit confidence interval with the independence-smoothed counts with $c = .5$ but with the modified endpoints when any cell counts are zero. These intervals share the Gart property of being relatively short, but they exhibit much better behavior for the minimum coverage probability. The coverage probabilities behave well unless the absolute log odds ratio is very large. To illustrate, the second panel of Figure 2 shows coverage probabilities for this interval, again for the case $n_1 = n_2 = 10$. The coverage probability exceeds .94 for all odds ratios less than 24.5, it exceeds .92 for all odds ratios less than 33.8, and it exceeds .90 for all odds ratios less than 86.5. In the random sample size case, these odds ratio values are (10.2, 35.8, 81.6). Since these are based on a sample of cases, they are not absolute minima, but similar patterns occurred when 100,000 binomial pairs with particular sample size combinations were considered. The use of the modified endpoints with zero counts affects coverage probabilities almost entirely when the odds ratio is extremely large. The effect is that the coverage probabilities do not diminish sharply as they otherwise would, and the overall appearance of the coverage probabilities is similar to that for the Woolf interval, which is completely uninformative when any counts equal zero. The results on minimum coverage are better yet if one enters $\{\tilde{p}_{ij}\}$ in the delta-method formula with effective sample size n instead of $n+2$, as in the Woolf approach (i.e., $\hat{\mu}_{ij} = n\tilde{p}_{ij}$ instead of $\hat{\mu}_{ij} = (n+2)\tilde{p}_{ij}$), but those results are not reported here for lack of space.

The proportion estimates for this approach and the ones for the Gart interval also result from a Bayesian construction (Bishop et al., 1975). Bayesian approaches to interval estimation of θ can use a variety of potential estimators, depending on such factors as whether one formulates the problem in terms of multinomial parameters in the entire table or binomial parameters in each row, the form of the prior (e.g., Dirichlet for probabilities or normal for logits of probabilities), and how one selects parameters for the prior. One might also (as in Bishop et al., 1975, Section 12.5) use an empirical Bayes approach in which the amount of smoothing depends on the roughness of the data, although some limited work done with this estimator shows poorer coverage performance.

4. Other Methods of Interval Estimation for Odds Ratios

This note has focused on the logit interval because it is the simplest method and the one commonly discussed in textbooks. Of course, this is not the only existing large-sample method for interval estimation of the odds ratio. To approximate the Cornfield $100(1-\alpha)\%$ exact conditional interval, Cornfield (1956) and Fisher (1962) proposed the interval consisting of odds ratios resulting from expected frequencies having the same margins as the observed counts and for which the Pearson chi-squared statistic with continuity correction has a P-value exceeding α, i.e., an endpoint of the interval has the form $[(n_{11}-d)(n_{22}-d)]/[(n_{12}+d)(n_{21}+d)]$, where

$$(|d|-.5)^2\left(\frac{1}{n_{11}-d}+\frac{1}{n_{22}-d}+\frac{1}{n_{12}+d}+\frac{1}{n_{21}+d}\right) = z^2_{\alpha/2}.$$

This interval also tends to be conservative since it mimics the exact one. Much shorter intervals result while usually maintaining coverage probability near the desired level by inverting this test (which is the score test) without the continuity correction.

Alternatively, one could invert the likelihood-ratio chi-squared test, which gives the profile likelihood interval. Or one could form intervals using the conditional log likelihood, taking the set of odds ratio values for which twice the conditional log likelihood falls within $z^2_{\alpha/2}$ of its maximum (Aitkin et al., 1989, p. 198). The inversion of the Pearson or likelihood-ratio test also does not have the Woolf-interval deficiency of being uninformative in both directions when a cell count is zero. Another interval, which is shorter than the Cornfield exact interval, is the adaptation of that interval based on inverting the test using the mid-P value. Mehta and Walsh (1992) showed that this method, although not guaranteed to achieve at least the nominal confidence, usually does so and without the strong conservativeness that the exact interval can exhibit for small samples.

In the cases studied, the interval based on inversion of the Pearson test (without the continuity correction) performed well, with coverages probabilities fluctuating around the nominal confidence level and with much shorter intervals than the exact or mid-P method. The coverage probabilities of the profile likelihood intervals behaved somewhat more erratically. The interval based on the test with mid-P value behaved well in terms of rarely having actual coverage probability much below the nominal confidence level, but on the average it was somewhat conservative. Though its intervals were considerably shorter than the exact conditional intervals, they tended to be longer than the logit intervals or the intervals based on inverting the Pearson test. Table 3 illustrates the performance of these methods for the case $n_1 = n_2 = 10$.

5. Conclusions

Although Tables 1–4 describe behavior of the logit confidence intervals in a limited variety of cases, they do suggest some tentative recommendations: If one can tolerate the minimum coverage probability dipping below the nominal confidence level as long as it never falls substantially below that level, then the Woolf interval seems adequate. If, in addition, the effective parameter space for the given application consists of small odds ratios (say, less than about five) and the sample sizes are not highly unbalanced, the Gart interval also seems adequate. The advantage of changing the choice of logit interval in cases where it is reasonable to assume a restricted parameter space is that the intervals are somewhat shorter, although they also have smaller minimum coverage probability. The independence-smoothed logit interval is recommended, however, over the Gart interval, i.e., add $2(n_{i+}n_{+j}/n^2)$ rather than .5 to each cell and adjust endpoints appropriately for empty cells. In all cases considered here, the coverage probability for this approach rarely fell substantively below the nominal level (cf., the second panel of Figure 2). Its interval width tended to be only slightly larger than the Gart interval, but its minimum coverages performed much better. There is scope for further improvement, but this logit interval seems adequate for most purposes.

In some applications, the Woolf or independence-smoothed logit interval may even be preferable to the exact conditional interval because of the marked reduction in interval width at relatively little risk of the coverage probability falling much below the nominal confidence level. Of course, if one has no tolerance for the coverage probability possibly falling below the nominal confidence level, none of the logit intervals is adequate and one should use an exact interval.

Acknowledgements

This research was partially supported by a grant from NIH. The author is very appreciative of several helpful comments from Dr Robert G. Newcombe and from a referee.

Résumé

A moins que la véritable association soit très forte, de simples intervalles de confiance pour l'odds ratio, établis à partir de la méthode delta et pour de grands échantillons, ont de bonnes performances, même pour des petits échantillons. Ces intervalles comprennent l'intervalle logit de Woolf et l'intervalle associé de Gart, où l'on ajoute .5 avant le calcul de l'estimateur du logarithme de l'odds ratio et de son erreur standard. L'intervalle de Gart équilibre les valeurs observées vers le modèle uniforme, mais on obtient de meilleures probabilités de recouvrement en équilibrant vers le modèle d'indépendance, et en étendant l'intervalle dans la direction appropriée lorsqu'une cellule correspond à une valeur nulle.

References

Agresti, A. (1990). *Categorical Data Analysis.* New York: Wiley.

Agresti, A. and Coull, B. C. (1998). Approximate is better than 'exact' for interval estimation of binomial parameters. *The American Statistician* **52,** 119–126.

Aitkin, M., Anderson, D., Francis, B., and Hinde, J. (1989). *Statistical Modelling in GLIM.* Oxford: Clarendon Press.

Anscombe, F. J. (1956). On estimating binomial response relations. *Biometrika* **43,** 461–464.

Bedrick, E. (1984). Estimating the variance of empirical logits and contrasts in empirical log probabilities. *Biometrics* **40,** 805–809.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis.* Cambridge, Massachusetts: MIT Press.

Breslow, N. and Day, N. E. (1980). *Statistical Methods in Cancer Research,* Volume 1. Lyon: IARC.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **4,** 135–148.

Cytel. (1995). *StatXact 3 for Windows.* Cambridge, Massachusetts: Cytel Software.

Fisher, R. A. (1962). Confidence limits for a cross-product ratio. *Australian Journal of Statistics* **4,** 41.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions,* 2nd ed. New York: Wiley.

Gart, J. J. (1966). Alternative analyses of contingency tables. *Journal of the Royal Statistical Society, Series B* **28,** 164–179.

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.

Gart, J. J. and Thomas, D. G. (1972). Numerical results on approximate confidence limits for the odds ratio. *Journal of the Royal Statistical Society, Series B* **34,** 441–447.

Haldane, J. B. S. (1956). The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics* **20,** 309–311.

Mehta, C. R. and Walsh, S. J. (1992). Comparison of exact, mid-p, and Mantel–Haenszel confidence intervals for the common odds ratio across several 2×2 contingency tables. *The American Statistician* **46,** 146–150.

Neyman, J. (1935). On the problem of confidence limits. *Annals of Mathematical Statistics* **6,** 111–116.

Woolf, B. (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics* **19,** 251–253.

Received January 1998. *Accepted April* 1998.