Journal of Epidemiology and Community Health, 1983, **37**, 82–84

Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure

F D K LIDDELL

From the Department of Epidemiology and Health, McGill University, Montreal, PQ, Canada

SUMMARY In a case-referent study of matched pairs with a single dichotomous exposure variable, the relative risk R is estimated from the "discordant pairs," r and s, by $\hat{R} = r/s$. In this paper the exact $(1 - \alpha)$ confidence limits, R_L and R_U, have been simplified, and a new exact test of the null hypothesis that $R = 1$ has been derived. The exact methods depend on the F distribution; they are simple to calculate and all have been fully validated. Even after simplification, methods for obtaining approximations to R_L and R_U remain more unwieldy than the exact methods, and an evaluation shows that these approximations may occasionally be far from the truth. It is argued that there is no excuse for not using the exact methods.

This note aims to clarify, simplify, and illustrate the methods summarised by Breslow and Day[1] (hereinafter designated *B&D*) for analysing a case-control study in which, for each case, one referent has been chosen matched to the case for some characteristic(s) and where a single dichotomous factor of exposure has been evaluated. Armitage's notation[2] is used for the classification of the $\frac{1}{2}n$ pairs (n subjects) as follows:

Case	Referent: Factor +	Referent: Factor −
Factor +	t	r
Factor −	s	u

Clearly, $t + r + s + u = \frac{1}{2}n$. The "concordant pairs" (t and u) provide no information about the relative risk; only the "discordant pairs," r and s, contribute. All procedures are based on the condition that $(r + s)$ is fixed; departures from the null hypothesis that $E(r) = E(s)$ are examined on the basis of s being distributed binomially with parameters $\frac{1}{2}$ and $(r + s)$. Exact procedures arise from the mathematical properties linking the binomial and F distributions.[3]

Estimates of relative risk (R)

The maximum likelihood estimate of relative risk R was shown by Mantel and Haenszel[4] to be

$$\hat{R} = r/s, \tag{1}$$

where $r \geqslant s$. (If $\hat{R} < 1$, r and s can be interchanged as will be shown later.)

The exact $(1 - \alpha)$ interval estimate of R requires two critical values of the F distribution, each for $\frac{1}{2}\alpha$. The lower $(1 - \alpha)$ confidence limit is

$$R_L = r/(s + 1)F_L \tag{2}$$

where the degrees of freedom for F_L are $2(s + 1)$ and $2r$. The expression (2) comes from *B&D*'s (5.7) and (5.9), remembering that their n_{10} and n_{01} are Armitage's r and s, so that their $F_{\alpha/2}(2n_{01} + 2, 2n_{10})$ is my F_L. The upper $(1 - \alpha)$ confidence limit is

$$R_U = (r + 1)F_U/s \tag{3}$$

with $2(r + 1)$ and $2s$ degrees of freedom for F_U at the $\frac{1}{2}\alpha$ probability level. Derivation of (3) is again straightforward.

Approximate confidence limits can be found by the usual approximation to the binomial tail probabilities from expressions (5.10) in *B&D*. These can be reduced to two quadratic equations but remain more unwieldy than (2) and (3) above. Other, more approximate, limits can be based on the asymptotic variance of $\ln(\hat{R})$, but as their simplest versions are more complex still, they are not discussed further.

Test of the hypothesis that $R = 1$

For a test of significance, the tail of the binomial distribution can always be calculated.[1] The link between the binomial and F distributions, however, can be exploited by solving (5.7) of *B&D* for $\pi_L = \frac{1}{2}$.

This yields

$$\tfrac{1}{2} = r/\{r + (s + 1)F\}$$

whence $\quad r + (s + 1)F = 2r$

and $\quad F = r/(s + 1), \qquad (4)$

where the degrees of freedom are $2(s + 1)$ and $2r$, as in (5.7); the one-sided probability of F being at least as large as that calculated at (4) can be readily obtained. The justification for the earlier assertion that, if $r < s$, they can be interchanged is seen in the solution of (5.7) in *B&D* for $\pi_U = \tfrac{1}{2}$, leading to $F = s/(r + 1)$ with $2(r + 1)$ and $2s$ degrees of freedom.

The "standard" test of the hypothesis that $R = 1$ is the McNemar χ^2 test which refers

$$(|r - s| - 1)^2/(r + s) \qquad (5)$$

to the χ^2 distribution with 1 degree of freedom, to yield a two-sided p-value. The derivation of (5) is given neatly by *B&D* (p 165); its computation is also simple, but this test remains approximate.

Example, and evaluation of procedures

In the example of *B&D* (pp 167–8) $t = 27$, $u = 4$, $r = 29$, and $s = 3$. The point estimate of relative risk is

$$\hat{R} = 29/3 = 9{\cdot}67.$$

The exact confidence limits of R depend on F_L with 8 & 58 df and F_U with 60 & 6 df. These can be found from tables either directly or by double-linear interpolation with harmonic arguments[5] (see discussion). The 95% confidence limits are:

$$R_L = r/(s + 1)F_L = 29/(3 + 1)2{\cdot}4196 = 2{\cdot}9964$$

and $R_U = (r + 1)F_U/s = (29 + 1)4{\cdot}9589/3 = 49{\cdot}589.$

These values, which are as found by *B&D*, can be verified by calculating the precise probabilities of the binomial distributions at the upper and lower limits, using (5.2) in *B&D* to obtain π_L and π_U, the binomial parameters at these limits, the other binomial parameter being $(r + s) = 32$.

$$\pi_L = R_L/(R_L + 1) = 0{\cdot}749\,772,$$

and

$$\pi_U = R_U/(R_U + 1) = 0{\cdot}980\,233.$$

The lower tail area is found with π_L from terms up to and including $s = 3$; the upper area with π_U by subtracting from unity terms up to but excluding s. The probabilities are as follows:

	π_L		π_U
0	0·000 099	0	0·527 881
1	0·001 062	1	0·340 644
2	0·005 496	2	0·106 475
3	0·018 342		
		0 to 2	0·975 000
0 to 3	0·024 999	3 to 32	0·025 000

The approximations to R_L and R_U were also calculated using simplifications of (5.10) of *B&D*, together with the corresponding binomial tail areas. The R_L approximation was 2·8216 (tail area 0·017 968), which is liberal but not outrageously so; that for R_U was 39·7677 (0·043 148), which is substantially conservative.

Lower and upper 95% confidence limits on R

Values of s		Values of r/s: 1·5	2	3	5	9
1	Exact*	—	0·10;118·0	0·24;157·5	0·56;236·5	1·25;394·5
	Approx†	—	0·14;55·65	0·28;74·81	0·57;113·1	1·18;189·8
2	Exact	0·17;17·96	0·29;22·11	0·54;30·39	1·07;46·93	2·15;79·98
	Approx	0·21;12·78	0·32;15·67	0·55;21·45	1·04;33·00	2·02;56·10
4	Exact	0·36;7·23	0·54;9·08	0·91;12·76	1·67;20·12	3·23;34·81
	Approx	0·38;6·30	0·55;7·87	0·90;11·00	1·62;17·25	3·07;29·74
8	Exact	0·56;4·23	0·81;5·40	1·30;7·72	2·31;12·37	4·33;21·64
	Approx	0·57;4·01	0·81;5·09	1·29;7·25	2·25;11·55	4·19;20·16
16	Exact	0·76;3·02	1·07;3·90	1·67;5·66	2·90;9·16	5·36;16·16
	Approx	0·77;2·96	1·06;3·81	1·66;5·50	2·85;8·89	5·25;15·65
32	Exact	0·94;2·42	1·29;3·16	2·00;4·62	3·41;7·54	6·25;13·38
	Approx	0·94;2·40	1·28;3·13	1·98;4·57	3·37;7·45	6·17;13·20

*From expressions (2) and (3) in text.
†From equations (5.10) in *B&D*, after simplification.

The test of the null hypothesis is, from (4), $F = 29/4 = 7{\cdot}25$ with 8 & 58 df. The associated (one-tailed) probability is $1{\cdot}278\,007 \times 10^{-6}$. This is identical to that obtained from the tail of the binomial distribution with parameters $\frac{1}{2}$ and $(r+s)$ — that is, $\Pr(i \leqslant 3 \mid \mathrm{Bi}[i \mid \frac{1}{2}, 32])$ is also $1{\cdot}278\,007 \times 10^{-6}$. The two-sided p-value is, therefore, $2{\cdot}556 \times 10^{-6}$.

The McNemar test is to refer $(29-3-1)^2/(29+3) = 19{\cdot}5313$ to the χ^2 distribution with 1 df, giving (two-sided) $p = 9{\cdot}898 \times 10^{-6}$, which is 3·9 times as great as the exact value found above. (*B&D* appear to have quoted the one-sided value.)

The table gives the exact and approximate lower and upper confidence limits for $\frac{1}{2}\alpha = 0{\cdot}025$ over ranges of values for $s = 1, 2, 4, 8, 16$, and 32; and for $(r/s - 1) = \frac{1}{2}, 1, 2, 4$, and 8. The approximate lower limits were often close to those obtained by the exact method, and there seems little chance of this process being *seriously* misleading. The approximate upper limits, however, were often far from the truth, particularly for small s.

For every one of the exact values of the probability associated with F for specified degrees of freedom, the algorithms explored (with H-P 67 and Monroe 1930) were in complete agreement; even more importantly, they also agreed with the calculated tail of the binomial distribution. In other words, the p-values based on the F distribution from (4) were undoubtedly exact. When the p-values were large, so that the null hypothesis could hardly be rejected, the true p was close to, and in its highest values a little greater than, the p-value from the McNemar approximation. In the range of usual interest $(0{\cdot}2 > p > 0{\cdot}001)$ the approximation remained fairly close to the truth, but was slightly conservative. Where the p-values were very low, even the conservative McNemar test would hardly affect the decision to reject the null hypothesis. A further dozen examples, taken from work in progress, conformed to the same principles.

Discussion

Throughout this paper, the condition that $(r+s)$ is fixed must be remembered. Thus the word "exact" has to be placed in that context, but the implication is the same for every procedure.

The usual reason for using approximate methods, when exact methods are available, is that the former are "quick" if sometimes "dirty." There is no such justification in the present circumstances. The simplified calculations, (2) and (3), of confidence limits are very "quick," when interpolation for critical values of F is not required, and completely "clean." Despite warnings,[5 6] double-linear interpolation with harmonic arguments has been found adequate: all binomial tail areas for the exact limits in the table are 0·025 to 3 dp except where $s = 32$ and $r/s = 3$, 5, and 9, when the areas are 0·026 to 3 dp, the largest being 0·0258. A short interpolation program renders the procedures of (2) and (3) simpler as well as more reliable than any approximations. Although the hypothesis test (4) appears new, it is based on well-known theory and has been fully validated; it also is "quick" and "clean." The McNemar test is a fraction "slower" and although usually fairly reliable can in certain circumstances be quite "dirty."

Thus there appears no excuse for not using the exact methods. Programs have been written for the H-P 67 to carry out the exact procedures, and double interpolation for F; copies can be obtained from the author. Details of all evaluations are also available on request.

References

[1] Breslow NE, Day NE. *Analysis of case-control studies.* (Scientific publication 32.) Lyon: International Agency for Research on Cancer, 1980: 164–9.
[2] Armitage P. *Statistical methods in medical research.* Oxford: Blackwell, 1971: 430.
[3] Fisher RA. The mathematical distributions used in the common tests of significance. *Econometrica* 1935; **3:** 353–65.
[4] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22:** 719–48.
[5] Fisher RA, Yates F. *Statistical tables for biological, agricultural and medical research workers.* 5th ed. Edinburgh: Oliver and Boyd, 1957: 2.
[6] Pearson ES, Hartley HO. *Biometrika tables for statisticians.* Vol II. Cambridge: University Press, 1976: 12–3.