STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:689–699 (DOI: 10.1002/sim.1017)

Analysis of the ratio of marginal probabilities in a matched-pair setting§

Jun-mo Nam[1,*,†] and William C. Blackwelder[2,‡]

[1] *Biostatistics Branch, National Cancer Institute, Executive Plaza South, Room 8028, 6120 Executive Boulevard, MSC 7244, Rockville, Maryland 20852-4909, U.S.A.*
[2] *Division of Microbiology and Infectious Diseases, National Institute of Allergy and Infectious Diseases, 6700 B Rockledge Drive, MSC 7630, Bethesda, Maryland 20892-7630, U.S.A.*

SUMMARY

Statistical methods for testing and interval estimation of the ratio of marginal probabilities in the matched-pair setting are considered in this paper. We are especially interested in the situation where the null value is not one, as in one-sided equivalence trials. We propose a Fieller-type statistic based on constrained maximum likelihood (CML) estimation of nuisance parameters. For a series of examples, the significance level of the CML test is satisfactorily close to the nominal level, while a Wald-type test is anticonservative for reasonable sample sizes. We present formulae for approximate power and sample size for the CML and Wald tests. The matched design is seen to have a clear advantage over the unmatched design in terms of asymptotic efficiency when the two responses of the pair are highly positively correlated. We recommend the CML method over the Wald method, especially for small or moderate sample sizes. Published in 2002 by John Wiley & Sons, Ltd.

KEY WORDS: constrained maximum likelihood estimator; efficiency of matching; equivalence test in matched–pair studies; power; ratio of marginal probabilites; sample size determination

1. INTRODUCTION

There has been considerable interest recently in analysis of data arising from a matched-pairs design, when the aim is not necessarily to test a hypothesis of no difference. An example is the comparison of two screening or diagnostic tests. Such comparisons include, but are not limited to, one- and two-sided tests of equivalence (Blackwelder [1]). For example, we might assume a new screening or diagnostic test is equivalent to an established procedure, or nearly so, and set out to show that the performance of the new test is in some respect within a prespecified quantity of that of the standard test, on either the arithmetic or multiplicative scale.

* Correspondence to: Jun-mo Nam, Biostatistics Branch, National Cancer Institute, Executive Plaza South, Room 8028, 6120 Executive Boulevard, MSC, 7244, Rockville, Maryland 20852-4909, U.S.A.
† E-mail: namj@mail.nih.gov
‡ Current address: 8613 Hempstead Avenue, Bethesda, Maryland 20817-6711, U.S.A.

Received September 2000
Accepted April 2001

Table I. Observations and probabilities in a matched-pair study

Experimental treatment	Control treatment		Total	Experimental treatment	Control treatment		Total
	Positive	Negative			Positive	Negative	
Positive	x_{11}	x_{10}	$x_{1.}$	Positive	p_{11}	p_{10}	P_1
Negative	x_{01}	x_{00}	$x_{0.}$	Negative	p_{01}	p_{00}	$1-P_1$
Total	$x_{.1}$	$x_{.0}$	n	Total	P_0	$1-P_0$	1

Testing a hypothesis of a non-zero difference between marginal probabilities in a matched pairs design has been considered by Lu and Bean [2], Nam [3] and Tango [4, 5]. In some situations it is more natural, however, to consider the ratio of the marginal probabilities; for example, in comparing the sensitivities of two tests when the null value is not one. Lachenbruch and Lynch [6] approached a two-sided version of this problem by means of a Wald statistic – that is, statistics in which the variance is estimated by maximum likelihood with no constraints. VanRaden *et al.* [7] suggested a likelihood scores approach as described by Bartlett [8], but they did not present a solution in closed form.

In this paper we derive a Fieller-type statistic, similar to that of Lachenbruch and Lynch [6], but with the variance estimated by maximum likelihood constrained by a specified value of the ratio of marginal probabilities. We show that the type I error rate of the resulting test procedure is quite close to nominal values for reasonable sample sizes, whereas that of the Wald statistic may be considerably higher than the nominal value. We also suggest a confidence interval procedure that corresponds to the test statistic and derive asymptotic expressions for power and sample size. We provide examples for numerical illustration and discuss them in some detail.

2. MODEL AND NOTATION

Suppose we have n pairs of matched observations on a dichotomous variable, with one observation of the pair resulting from an experimental treatment (for example, experimental drug, new test) and the other from a control treatment (for example, standard therapy or test, or placebo). Let x_{ij} be the number of observations with response i for the experimental treatment and response j for the control treatment, where i and $j=1$ or 0 for a positive or negative response, respectively. We assume the x_{ij} follow a multinomial distribution with probabilities p_{ij} $(i,j=0$ or $1)$, where the p_{ij} sum to 1, and define $P_1=p_{11}+p_{10}$, $P_0=p_{11}+p_{01}$. Then P_1 and P_0 are the probabilities of a positive response with the experimental and control treatments, respectively. We may display the observations and probabilities as shown in Table I.

We are interested in inferences on the ratio of marginal probabilities $\phi=P_1/P_0$. Note that, for $\phi\neq 1$, $p_{11}=(p_{10}-\phi p_{01})/(\phi-1)$ and $p_{00}=\{\phi(1-p_{10})+p_{01}-1\}/(\phi-1)$. Thus for $\phi\neq 1$ the multinomial probabilities p_{ij} can be expressed as functions of ϕ, p_{10} and p_{01}.

3. TEST STATISTIC AND INTERVAL ESTIMATION

From the above, we can write the log-likelihood function as

$$\ln L = x_{11}\ln(p_{10} - \phi p_{01}) - (x_{11} + x_{00})\ln(\phi - 1) + x_{10}\ln p_{10} + x_{01}\ln p_{01} + x_{00}\ln\{\phi(1 - p_{10}) + p_{01} - 1\}$$

where ϕ is the parameter of interest and p_{10} and p_{01} are considered nuisance parameters. The (constrained) maximum likelihood (CML) estimators of p_{10} and p_{01} for a given value of ϕ are obtained by solving the two equations $\partial \ln L/\partial p_{10} = 0$ and $\partial \ln L/\partial p_{01} = 0$. Denoting the solutions by $\tilde{p}_{10}$ and $\tilde{p}_{01}$, we have

$$\tilde{p}_{10} = [-\hat{P}_1 + \phi^2(\hat{P}_0 + 2\hat{p}_{10}) + \{(\hat{P}_1 - \phi^2\hat{P}_0)^2 + 4\phi^2\hat{p}_{10}\hat{p}_{01}\}^{1/2}]/\{2\phi(\phi + 1)\} \tag{1}$$

and

$$\tilde{p}_{01} = \phi\tilde{p}_{10} - (\phi - 1)(1 - \hat{p}_{00})$$

where $\hat{P}_1 = x_{1.}/n$, $\hat{P}_0 = x_{.1}/n$, $\hat{p}_{10} = x_{10}/n$, and $\hat{p}_{01} = x_{01}/n$ (Appendix A). Note that values of $\tilde{p}_{10}$ and $\tilde{p}_{01}$ are located in $[0, 1)$ (Appendix B).

Suppose, for example, that we wish to demonstrate that the proportion of positive results with a new test is at least a factor ϕ_0 of the proportion of positives with an established test. Then we test the null hypothesis $H_0: \phi \leqslant \phi_0$ against the one-sided alternative $H_1: \phi > \phi_0$. For $\phi_0 < 1$, H_1 has the form of a one-sided equivalence, or non-inferiority, hypothesis. For $\phi_0 = 1$, H_0 is the conventional null hypothesis that $P_1 \leqslant P_0$.

In this paper we are primarily interested in the one-sided setting. However, we may be interested in inference about ϕ in both directions. In a two-sided equivalence setting, we would test $H_0: \phi \leqslant 1/\phi_0$ or $\phi \geqslant \phi_0$, for $\phi_0 > 1$, against $H_1: 1/\phi_0 < \phi < \phi_0$.

We can express H_0 as two separate one-sided hypotheses, both of which must be rejected in order to reject H_0. If each of the separate hypotheses is tested at significance level α, then the overall significance level is $\leqslant \alpha$, but the type II error rate for the alternative that $\phi = 1$ is approximately 2β, if β is the type II error rate for each of the one-sided tests [1]. From the relation $P_1 - \phi P_0 = 0$, we consider a Fieller-type statistic $T(\phi) = \hat{P}_1 - \phi\hat{P}_0$. For $\phi = \phi_0$, $T(\phi_0)$ is asymptotically normally distributed with expectation 0 and variance $\phi_0(p_{10} + p_{01})/n$ (Appendix C, (C1)). An appropriate test statistic for large n is then

$$z(\phi_0) = n^{1/2}(\hat{P}_1 - \phi_0\hat{P}_0)/\{\phi_0(\tilde{p}_{10} + \tilde{p}_{01})\}^{1/2} \tag{2}$$

where $\tilde{p}_{10}$ and $\tilde{p}_{01}$ are evaluated at $\phi = \phi_0$. We refer to $z(\phi_0)$ as a constrained maximum likelihood (CML) statistic. We reject $H_0: \phi \leqslant \phi_0$ in favour of $H_1: \phi > \phi_0$ if $z(\phi_0) > z_{1-\alpha}$, where $z_{1-\alpha}$ is the upper $100(1 - \alpha)$ percentage point of the standard normal distribution.

The lower and upper limits, ϕ_L and ϕ_U, of an estimated $100(1 - 2\alpha)$ per cent confidence interval for ϕ are the solutions of the equations

$$z(\phi) = \pm z_{1-\alpha} \tag{3}$$

(Appendix D). Then an equivalent test of H_0 is to reject if $\phi_L > \phi_0$. The limits ϕ_L and ϕ_U can be found by an iterative procedure, for example, the Newton–Raphson algorithm or the bisection method [9].

A Wald-type statistic for testing H_0 can be written as

$$z_w(\phi_0) = n^{1/2}(\hat{P}_1 - \phi_0 \hat{P}_0)/\{\phi_0(\hat{p}_{10} + \hat{p}_{01})\}^{1/2} \quad (4)$$

Note that z and z_w differ only in the standard error estimate in the denominator. The statistic z employs CML estimates of the nuisance parameters p_{10} and p_{01}, given $\phi = \phi_0$, whereas z_w uses unconstrained maximum likelihood estimates. For $\phi_0 = 1$, both statistics reduce to McNemar's statistic [10], just as the one-sided null hypothesis reduces to the conventional hypothesis that $\phi \leqslant 1$. (Note, however, that the two-sided null hypothesis in an equivalence setting, $H_0: \phi \leqslant 1/\phi_0$ or $\phi \geqslant \phi_0$, does not reduce to the usual hypothesis that $\phi = 1$. The statistics (2) and (4) are in general form and are applied more broadly than in equivalence trials.) For simplicity, we shall refer to z and z_w as CML and Wald statistics, respectively.

We could define hypotheses and tests based on $\psi = (1 - P_1)/(1 - P_0)$. However, since $\psi = 1 + \{(1 - \phi)P_0/(1 - P_0)\}$ is a function of both ϕ and P_0, a hypothesis about ϕ cannot in general be expressed in terms of ψ. Unless there is interest in both ϕ and ψ, we can always define the parameters so that the ratio of interest is ϕ.

4. POWER AND SAMPLE SIZE

We denote the asymptotic limits of the CML estimates $\tilde{p}_{10}$ and $\tilde{p}_{01}$ by $\bar{p}_{10}$ and $\bar{p}_{01}$, respectively. Then $\bar{p}_{10} = [-P_1 + \phi_0^2(P_0 + 2p_{10}) + \{(P_1 - \phi_0^2 P_0)^2 + 4\phi_0^2 p_{10} p_{01}\}^{1/2}]/\{2\phi_0(\phi_0 + 1)\}$ and $\bar{p}_{01} = \phi_0 \bar{p}_{10} - (\phi_0 - 1)q_{00}$ where $q_{00} = 1 - p_{00}$.

For convenience we denote $T(\phi_0) = \hat{P}_1 - \phi_0 \hat{P}_0$ by T_0. The asymptotic form of the estimated variance of T_0 under the null hypothesis is

$$\bar{V}_0(T_0) = \phi_0(\bar{p}_{10} + \bar{p}_{01})/n$$

The expectation and variance of T_0 under the alternative hypothesis $H_1 : \phi = \phi_1$, where $\phi_1 > \phi_0$, are

$$E_1(T_0) = (\phi_1 - \phi_0)P_0 \quad (5)$$

and

$$V_1(T_0) = \{(\phi_1 + \phi_0^2)P_0 - 2\phi_0 p_{11} - (\phi_1 - \phi_0)^2 P_0^2\}/n$$

(Appendix C, (C2)). For large samples the approximate power $1 - \beta$ of the CML statistic (2) for testing $H_0 : \phi \leqslant \phi_0$, given the alternative that $\phi = \phi_1 > \phi_0$, is obtained by

$$\text{power } 1 - \beta = 1 - \Phi(u) \quad (6)$$

where $u = [z_{1-a}\{\bar{V}_0(T_0)\}^{1/2} - E_1(T_0)]/\{V_1(T_0)\}^{1/2}$ and $z_{1-\alpha}$ and Φ are the upper $(1 - \alpha)$ quantile and cumulative distribution function, respectively, of the standard normal distribution. The approximate sample size required for power $1 - \beta$ can be found from the equation (see, for example, reference [11])

$$E_1(T_0) = z_{1-a}\{\bar{V}_0(T_0)\}^{1/2} + z_{1-\beta}\{V_1(T_0)\}^{1/2}$$

Then the approximate sample size (pairs of matched observations) for testing the one-sided hypothesis H_0 against alternative H_1 based on the CML statistic, using (5), is

$$n = \{z_{1-\alpha}\bar{v}_0^{1/2} + z_{1-\beta}v_1^{1/2}\}^2/\{(\phi_1 - \phi_0)P_0\}^2 \tag{7}$$

where $\bar{v}_0 = \phi_0(\bar{p}_{10} + \bar{p}_{01}), v_0 = \phi_0(p_{10} + p_{01})$ and $v_1 = (\phi_1 + \phi_0^2)P_0 - 2\phi_0 p_{11} - (\phi_1 - \phi_0)^2 P_0^2$. For testing the two-sided hypothesis that $\phi \leqslant \phi_0$ or $\phi \geqslant 1/\phi_0$, where $\phi_0 \leqslant 1$, if the true values P_1 and P_0 are assumed to be equal, then the approximate sample size for power $1-\beta$ can be calculated from (7) by replacing β with $\beta/2$ [12].

A formula for approximate sample size based on the Wald statistic is

$$n_{\mathrm{w}} = \{z_{1-\alpha}v_0^{1/2} + z_{1-\beta}v_1^{1/2}\}^2/\{(\phi_1 - \phi_0)P_0\}^2 \tag{8}$$

5. EVALUATION OF CML AND WALD METHODS

In practice, sample sizes for a matched-pair design may not be large. Thus it is important to assess the properties of any asymptotic test for small and medium sample sizes. There are $(n+3)(n+2)(n+1)/6$ possible outcomes for a given sample size n. For the case $x_{00} = n$, both the CML and Wald statistics are indeterminate. For either statistic the exact significance level for testing $H_0{:}\phi \leqslant \phi_0$ can then be expressed in the form

$$\sum_{\mathbf{x} \in R} \Pr(\mathbf{x} \mid \phi = \phi_0)/\{1 - \Pr(x_{00} = n)\}$$

where $\mathbf{x}' = (x_{11}, x_{10}, x_{01}, x_{00}), R$ is the critical region in which H_0 is rejected, and $\Pr(\mathbf{x} \mid \phi = \phi_0)$ is the multinomial probability of observing $\mathbf{x}$ for $\phi = \phi_0$. Table II summarizes computations of significance levels corresponding to a nominal one-sided significance level α of 0.05 for $\phi_0 = 0.8$, 0.9; $P_0 = 0.8$, 0.65, 0.5; $p_{10} = 0.05$, 0.10, 0.15; and $n = 25$, 50, 100. The significance levels shown are exact. In all these cases, as in many practical applications, the correction factor $\{1 - \Pr(x_{00} = n)\}$ is negligible. The table indicates that, for the cases studied, the CML statistic has significance level reasonably close to the nominal 0.05 and is generally slightly conservative. On the other hand, the Wald statistic is generally anticonservative – that is, significance levels associated with it are generally higher than the nominal level.

Approximate sample sizes from (7) for the CML statistic are shown in Table III for 80 per cent power and various combinations of the other parameter values. Calculations of the exact power for these $n \leqslant 100$ are also shown. The calculations indicate that, even for fairly small sample sizes, the power is close to, and often slightly larger than, the nominal value. A large sample size is needed when a difference between null and alternative is small. The required sample size increases as P_{10} increases and/or P_0 decreases. Sample sizes from (8) for the Wald statistic are somewhat smaller than those shown in Table III, but the anticonservative nature of the Wald statistic falsely inflates its power and makes it undesirable for general use.

6. EXAMPLES

We provide two illustrative examples, corresponding to relatively small and large sample sizes.

Table II. Actual levels of significance, from exact calculation, for Wald and CML tests at nominal 5 per cent level.

ϕ_0	P_0	p_{10}	$n=25$		$n=50$		$n=100$	
			Wald	CML	Wald	CML	Wald	CML
0.8	0.80	0.05	0.078	0.045	0.065	0.047	0.059	0.048
		0.10	0.065	0.050	0.058	0.049	0.054	0.050
		0.15	0.059	0.052	0.056	0.050	0.052	0.050
	0.65	0.05	0.075	0.047	0.065	0.048	0.059	0.049
		0.10	0.064	0.048	0.057	0.049	0.054	0.050
		0.15	0.059	0.049	0.054	0.050	0.052	0.050
	0.50	0.05	0.074	0.047	0.064	0.048	0.057	0.049
		0.10	0.061	0.049	0.055	0.049	0.052	0.050
		0.15	0.056	0.050	0.053	0.051	0.051	0.050
0.9	0.80	0.05	0.083	0.043	0.071	0.046	0.062	0.048
		0.10	0.072	0.046	0.060	0.048	0.056	0.050
		0.15	0.066	0.048	0.057	0.050	0.053	0.050
	0.65	0.05	0.082	0.047	0.068	0.047	0.061	0.048
		0.10	0.067	0.050	0.058	0.049	0.055	0.050
		0.15	0.061	0.050	0.056	0.050	0.053	0.050
	0.50	0.05	0.093	0.041	0.066	0.048	0.059	0.048
		0.10	0.066	0.048	0.057	0.051	0.054	0.050
		0.15	0.060	0.049	0.054	0.051	0.052	0.050

6.1. *Example 1*

Consider data of $x_{11}=17$, $x_{10}=2$, $x_{01}=1$ and $x_{00}=10$; then $n=30$ pairs. The rate of positives on the new treatment and the standard are $\hat{P}_1=19/30$ and $\hat{P}_0=18/30$ and the rate ratio is $\hat{\phi}=1.0556$. We want to show that the positive response rate with the new treatment is at least a factor $\phi=0.9$ of the rate with standard treatment, and we test $\phi\leqslant 0.9$ against $\phi>0.9$. The CML estimates are $\tilde{p}_{10}=0.0382$ and $\tilde{p}_{01}=0.1011$ from (1), and the CML and Wald statistics are $z=1.444(p=0.074)$ and $z_{\mathrm{w}}=1.703(p=0.044)$ from (2) and (4), respectively. Considering the anticonservative nature of the Wald test, particularly for small sample sizes, that test may falsely claim significance at the 5 per cent level. Conventional 90 per cent confidence intervals for ϕ by the CML and Wald methods are (0.872, 1.303) and (0.905, 1.231), respectively.

6.2. *Example 2*

Blake *et al.* [13] investigated whether certain vaginal infections could be satisfactorily diagnosed without using a speculum. Vaginal specimens were collected from participants, both with and without the use of a speculum, and evaluated by microscopic examination for the presence of trichomonas organisms. Ninety-nine patients among 686 study participants were positive for trichomonas by one of the two study methods or by culture, which served as

Table III. Approximate sample size required for 80 per cent power for CML tests at $\alpha = 0.05$ and actual power from exact calculation for those $n \leqslant 100$.

P_0	p_{10}	$\phi_0 = 0.8$ versus $\phi_1 = 1.0$		$\phi_0 = 0.9$ versus $\phi_1 = 1.0$
		n	(exact power)	n
0.8	0.05	34	(0.83)	112
	0.10	50	(0.82)	189
	0.15	67	(0.81)	272
0.65	0.05	47	(0.83)	160
	0.10	71	(0.81)	280
	0.15	97	(0.81)	406
0.50	0.05	71	(0.83)	254
	0.10	113		462
	0.15	159		679
0.40	0.05	102		381
	0.10	170		713
	0.15	243		1055
0.20	0.05	343		1429
	0.10	636		2801
	0.15	939		4185

the reference standard. Of these 99, 67 were positive for trichomonas by both study methods, 9 were positive only when the speculum was not used, 7 were positive only when the speculum was used, and 16 were negative by both methods. Then for this example $x_{11} = 67$, $x_{10} = 9$, $x_{01} = 7$, $x_{00} = 16$, $\hat{P}_1 = 76/99 = 0.77$ and $\hat{P}_0 = 74/99 = 0.75$. Since all 99 of the females were considered positive by at least one method, the estimates $\hat{P}_1$ and $\hat{P}_0$ can be considered estimates of sensitivity, assuming there were no false positives. The estimated ratio of sensitivities is 76/74, or 1.03. It was desired to demonstrate that the non-speculum collection method is at least 90 per cent as sensitive as the speculum method. For $\phi_0 = 0.9$, from (1) we have $\tilde{p}_{10} = 0.0591$ and $\tilde{p}_{01} = 0.1370$; from (2) the CML statistic is $z(0.9) = 2.248$, and the p-value for testing H_0 is 0.012. We can then reject the null hypothesis and conclude that the non-speculum method is more than 90 per cent as sensitive as the speculum method. The Wald statistic (4) is greater than the CML statistic ($z_{\mathrm{w}}(0.9) = 2.447$), which is consistent with the anti-conservativeness of the Wald statistic. From (3), the usual two-sided 90 per cent confidence interval for ϕ is (0.937, 1.130).

Now suppose we are planning a trial with $\phi_0 = 0.9$, $\phi_1 = 1$, $\alpha = 0.05$, $P_0 = 0.75$ and $p_{01} = 0.07$. The approximate sample sizes from (7) for 80, 90, and 95 per cent power are $n = 162$, 221 and 276, respectively. From (6), the asymptotic power of the CML test for the above assumptions and $n = 99$ is 61 per cent, which indicates that the study by Blake *et al.* had fair power if the two diagnostic tests were equally sensitive.

For this example we might also evaluate the ratio of false negative rates, the observed values of which are 25/99 and 23/99. If data were available for the two diagnostic tests on specimens known to be negative for trichomonas, we could perform hypothesis tests and interval estimation on the ratio of specificities.

7. EFFICIENCY OF MATCHING

Consider applying the two treatments to two independent samples of equal size n. The observed proportions $\hat{P}_1^*$ and $\hat{P}_0^*$ will be independent, and the variance of $T_0^* = \hat{P}_1^* - \phi_0\hat{P}_0^*$ can be written $V(T_0^*) = \{P_1(1-P_1) + \phi_0^2 P_0(1-P_0)\}/n$. A statistic for testing $H_0{:}\phi \leqslant \phi_0$ is

$$z(\phi_0) = n^{1/2}(\hat{P}_1^* - \phi_0\hat{P}_0^*)/\{\tilde{P}_1(1-\tilde{P}_1) + \phi_0^2\tilde{P}_0(1-\tilde{P}_0)\}$$

where $\tilde{P}_1$ and $\tilde{P}_0$ are maximum likelihood estimates of P_1 and P_0 for $\phi = \phi_0$ (see, for example, references [12] and [14]). Rearranging terms from (5) and substituting $P_1 = \phi P_0$ in the above expression for $V(T_0^*)$, we see that $V(T_0^*) - V(T_0) = 2\phi_0(p_{11} - P_1P_0)/n = 2\phi_0\rho\{V(\hat{P}_1)V(\hat{P}_0)\}^{1/2}$, where ρ is the correlation coefficient between $\hat{P}_1$ and $\hat{P}_0$. Then if $\rho > 0, V(T_0^*) > V(T_0)$ and the matching increases efficiency. The asymptotic relative efficiency (ARE) of the unmatched test relative to the matched test is

$$\text{ARE} = V(T_0)/V(T_0^*) = 1 - 2\rho\phi_0\{V(\hat{P}_1)V(\hat{P}_0)\}^{1/2}/V(T_0^*)$$

$\text{ARE} < 1$ when $\rho > 0$, and ARE decreases as ρ increases. Usually ρ will be considerably greater than 0, and the gain in efficiency will be appreciable. This result is consistent with the increased efficiency of the matched design for testing a non-zero difference between two treatments [3]. In some applications, such as the example in the previous section of testing for trichomonas infection, there are also the added advantages, including lower cost, of having only n individuals in the study, rather than $2n$ as in the unmatched design.

8. DISCUSSION

We have derived a Fieller-type statistic for testing and estimation of the ratio of marginal proportions when the data are in the form of matched pairs. The variance for this statistic was estimated by maximum likelihood constrained by a specified value of the ratio. The test statistic appears to have significance level satisfactorily close to the nominal value for sample sizes as small as 25 pairs. Power is also close to that calculated from an asymptotic approximation. A Wald statistic does not have appropriate significance levels for sample sizes likely to be attained in many applications. Therefore, we recommend the CML statistic for general use in the matched-pair setting for analysis of the ratio of marginal proportions.

APPENDIX A: CONSTRAINED MAXIMUM LIKELIHOOD ESTIMATORS OF NUISANCE PARAMETERS

From the likelihood in Section 3, the partial derivatives with respect to p_{10} and p_{01} are

$$\frac{\partial \ln L}{\partial p_{10}} = \frac{x_{11}}{p_{10} - \phi p_{01}} + \frac{x_{10}}{p_{10}} - \frac{x_{00}\phi}{\phi(1-p_{10}) + p_{01} - 1}$$

and

$$\frac{\partial \ln L}{\partial p_{01}} = -\frac{x_{11}\phi}{p_{10} - \phi p_{01}} + \frac{x_{01}}{p_{01}} + \frac{x_{00}}{\phi(1-p_{10}) + p_{01} - 1}$$

Define $\hat{p}_{ij} = x_{ij}/n$ for $i, j = 0, 1$. The MLEs of p_{10} and p_{01} for a given value of ϕ, $\tilde{p}_{10}$ and $\tilde{p}_{01}$, are a solution to $\partial \ln L/\partial p_{10} = 0$ and $\partial \ln L/\partial p_{01} = 0$ evaluated at a given value of ϕ:

$$\begin{aligned}&\tilde{p}_{11}\tilde{p}_{10}(\phi - 1 - \phi\tilde{p}_{10} + \tilde{p}_{01}) + \hat{p}_{10}(\tilde{p}_{10} - \phi\tilde{p}_{01})(\phi - 1 - \phi\tilde{p}_{10} + \tilde{p}_{01})\\ &\quad - \hat{p}_{00}\phi(\tilde{p}_{10} - \phi\tilde{p}_{01}) = 0 \end{aligned} \tag{A1}$$

and

$$-\hat{p}_{11}\phi\tilde{p}_{01}(\phi - 1 - \phi\tilde{p}_{01}) + \hat{p}_{01}(\tilde{p}_{10} - \phi\tilde{p}_{01})(\phi - 1 - \phi\tilde{p}_{10} + \tilde{p}_{01}) + \hat{p}_{00}\tilde{p}_{01}(\tilde{p}_{10} - \phi\tilde{p}_{01}) = 0 \tag{A2}$$

Dividing by $(\tilde{p}_{10} - \phi\tilde{p}_{01})$ after adding (A1) and (A2), we have a relation

$$\tilde{p}_{01} = \phi\tilde{p}_{10} - (\phi - 1)(1 - \hat{p}_{00}) \tag{A3}$$

We can simplify the (A1) after substituting (A3) in (A1) as

$$\phi(\phi + 1)\tilde{p}_{10}^2 + \{\hat{P}_1 - \phi^2(1 + \hat{p}_{10} - \hat{p}_{00})\}\tilde{p}_{10} + \phi(\phi - 1)\hat{p}_{10}(1 - \hat{p}_{00}) = 0$$

where $\hat{P}_1 = \hat{p}_{11} + \hat{p}_{10}$. A solution of the above quadratic equation is the MLE of p_{10} for a given value of ϕ in (1), Section 3 and the constrained MLE of p_{01} follows from (A3).

APPENDIX B: RANGE OF CML ESTIMATORS OF p_{10} and p_{01}

1. From (1), we can rewrite the CML estimator of p_{10} for a given value of ϕ as

$$\tilde{p}_{10} = \{-\hat{P}_1 + \phi^2\hat{P}_0 + 2\phi^2\hat{p}_{10} + \Delta\}/\{2\phi(\phi + 1)\} \tag{B1}$$

where $\Delta = \{(\hat{P}_1 - \phi^2\hat{P}_0)^2 + 4\phi^2\hat{p}_{10}\hat{p}_{01}\}^{1/2}$.

Since $|\hat{P}_1 - \phi^2\hat{P}_0| \leqslant \Delta \leqslant |\hat{P}_1 - \phi^2\hat{P}_0| + 2\phi(\hat{p}_{10}\hat{p}_{01})^{1/2}$, we can show the following inequalities from (B1):

$$0 \leqslant \frac{\phi\hat{p}_{10}}{\phi + 1} \leqslant \tilde{p}_{10} \leqslant \frac{\phi\hat{p}_{10} + (\hat{p}_{10}\hat{p}_{01})^{1/2}}{\phi + 1} < 1$$

when $\hat{P}_1 - \phi^2\hat{P}_0 > 0$ and

$$0 \leqslant \frac{\phi^2\hat{P}_0 - \hat{P}_1 + \phi^2\hat{p}_{10}}{\phi(\phi + 1)} \leqslant \tilde{p}_{10} \leqslant \frac{\phi^2\hat{P}_0 - \hat{P}_1 + \phi^2\hat{p}_{10} + \phi(\hat{p}_{10}\hat{p}_{01})^{1/2}}{\phi(\phi + 1)} < 1$$

when $\hat{P}_1 - \phi^2\hat{P}_0 \leqslant 0$. Therefore, the $\tilde{p}_{10}$ is in the interval $[0, 1)$.

2. Recall $\tilde{p}_{01} = \phi\tilde{p}_{10} + (1 - \phi)(1 - \hat{p}_{00})$. Then, we can rewrite $\tilde{p}_{01}$ as

$$\tilde{p}_{01} = (\hat{P}_1 - \phi^2\hat{P}_0 + 2\hat{p}_{01} + \Delta)/\{2(\phi + 1)\}.$$

From the above inequality for Δ, we have the following:

$$0 \leqslant \frac{\hat{P}_1 - \phi^2\hat{P}_0 + \hat{p}_{01}}{\phi + 1} \leqslant \tilde{p}_{01} \leqslant \frac{\hat{P}_1 - \phi^2\hat{P}_0 + \hat{p}_{01} + \phi(\hat{p}_{10}\hat{p}_{01})^{1/2}}{\phi + 1} < 1$$

when $\hat{P}_1 - \phi^2\hat{P}_0 \geqslant 0$ and

$$0 \leqslant \frac{\hat{p}_{01}}{\phi + 1} \leqslant \tilde{p}_{01} \leqslant \frac{\hat{p}_{01} + \phi(\hat{p}_{10}\hat{p}_{01})^{1/2}}{\phi + 1} < 1$$

when $\hat{P}_1 - \phi^2\hat{P}_0 < 0$.

Therefore, $\tilde{p}_{01}$ is also in the interval $[0, 1)$.

APPENDIX C: VARIANCE OF T

The variance of $T = \hat{P}_1 - \phi\hat{P}_0$ is $V(T) = \mathrm{var}(\hat{P}_1) + \phi^2\,\mathrm{var}(\hat{P}_0) - 2\phi\,\mathrm{cov}(\hat{P}_1, \hat{P}_0)$.

Since $\mathrm{var}(\hat{P}_i) = P_iQ_i/n$ where $Q_i = 1 - P_i$ for $i = 0, 1$ and $\mathrm{cov}(\hat{P}_1, \hat{P}_0) = (p_{11} - P_1P_0)/n$, the variance can be expressed as

$$V(T) = \phi(p_{10} + p_{01})/n \tag{C1}$$

using a relation of $P_1 = \phi P_0$. Similarly, we obtain the variance of $T_0 = \hat{P}_1 - \phi_0\hat{P}_0$ under $H_0{:}\phi = \phi_0$ and that under $H_1{:}\phi = \phi_1$ as $V_0(T_0) = \phi_0(p_{10} + p_{01})/n$ and

$$V_1(T_0) = \{(\phi_1 + \phi_0^2)P_0 - 2\phi_0 p_{11} - (\phi_1 - \phi_0)^2 P_0^2\}/n \tag{C2}$$

respectively.

APPENDIX D: CONFIDENCE INTERVAL BY CML METHOD

Denote $z'(\phi_0) = (\hat{P}_1 - \phi_0\hat{P}_0)/\{\phi_0(p_{10} + p_{01})/n\}^{1/2}$ for any positive real number ϕ_0. The $z'(\phi)$ is continuous and monotone decreasing with respect to ϕ where $0 < \phi < \infty$, and it is distributed asymptotically normal with mean 0 and variance 1 for a given value of ϕ. We may write the form (2) as

$$z(\phi_0) = \{(p_{10} + p_{01})/(\tilde{p}_{10} + \tilde{p}_{01})\}^{1/2} z'(\phi_0) \tag{D1}$$

where ϕ_0 is a positive real number. Since CML estimators of p_{10} and p_{01} for $\phi = \phi_0$, $\tilde{p}_{10}$ and $\tilde{p}_{01}$, are consistent estimators of the parameters, the first factor of the right-hand side (D1) converges in probability to one. Thus, $z(\phi_0)$ becomes equivalent in probability to $z'(\phi_0)$.

Recall ϕ_{L} and ϕ_{U} as the solutions of the equations, $z(\phi_{\mathrm{L}}) = z_{(1-\alpha)}$ and $z(\phi_{\mathrm{U}}) = -z_{(1-\alpha)}$, respectively. Using the inverse relation between ϕ and $z(\phi)$, we have $z(\phi) \leqslant z(\phi_{\mathrm{L}})$ if and only if $\phi \geqslant \phi_{\mathrm{L}}$ and $z(\phi) \geqslant z(\phi_{\mathrm{U}})$ if and only if $\phi < \phi_{\mathrm{U}}$. By combining them, we have

$$-z_{(1-\alpha)} \leqslant z(\phi) \leqslant z_{(1-\alpha)} \quad \text{if and only if } \phi_{\mathrm{L}} \leqslant \phi \leqslant \phi_{\mathrm{U}} \tag{D2}$$

Thus, from (D2), we have $\Pr(\phi_{\mathrm{L}} \leqslant \phi \leqslant \phi_{\mathrm{U}}) = \Pr\{-z_{(1-\alpha)} \leqslant z(\phi) \leqslant z_{(1-\alpha)}\} = 1 - 2\alpha$ and the interval defined by $[\phi_{\mathrm{L}}, \phi_{\mathrm{U}}]$ is approximate $(1 - 2\alpha)$ confidence interval for ϕ.

ACKNOWLEDGEMENTS

The authors thank the referees for helpful comments to improve the presentation of this paper.

REFERENCES

1. Blackwelder WC. Equivalence trials. In *Encyclopedia of Biostatistics* Vol. 2, Armitage P, Colton T (eds). Wiley: New York, 1998; 1367–1372.
2. Lu Y, Bean JA. On the sample size for one-sided equivalence of sensitivities based upon McNemar's test. *Statistics in Medicine* 1995; **14**:1831–1839.
3. Nam J. Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics* 1997; **53**:1422–1430.
4. Tango T. Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine* 1998; **17**:891–908.
5. Tango T. Letter to the Editor. *Statistics in Medicine* 1999; **18**:3511–3513.
6. Lachenbruch PA, Lynch CJ. Assessing screening tests: extensions of McNemar's test. *Statistics in Medicine* 1998; **17**:2207–2217.
7. VanRaden MJ, Blackwelder WC, Van Sant FD, Shrager RI, Nam J. Paired dichotomous data: test and confidence interval for ratio of marginal proportions. In *Proceedings of the Biometrics Section*, *American Statistical Association. American Statistical Association*: Alexandria, VA, 1997.
8. Bartlett MS. Approximate confidence intervals. II: More than one unknown parameter. *Biometrika* 1953; **40**: 306–317.
9. Cheney W, Kincaid D. *Numerical Mathematics and Computing*. Brooks/Cole: Belmont, California, 1985.
10. McNemar Q. Note on the sampling error of the differences between correlated proportions or percentages. *Psychometrika* 1947; **12**:153–157.
11. Nam J. A simple approximation for calculating sample sizes for detecting linear trend in proportions. *Biometrics* 1987; **43**:701–705.
12. Farrington CP, 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* 1990; **9**:1447–1454.
13. Blake DR, Duggan A, Quinn T, Zenilman J, Joffe A. Evaluation of vaginal infections in adolescent women: can it be done without a speculum? *Pediatrics* 1998; **102**:939–944.
14. Nam J. Sample size requirements for stratified prospective studies with null hypothesis of non-unity relative risk using the score test. *Statistics in Medicine* 1994; **13**:79–86.