STATISTICS IN MEDICINE
Statist. Med. **18**, 3511–3513 (1999)

LETTER TO THE EDITOR

IMPROVED CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN BINOMIAL PROPORTIONS BASED ON PAIRED DATA

by Robert G. Newcombe, *Statistics in Medicine*, **17**, 2635–2650 (1998)

From: T. Tango
Division of Theoretical Epidemiology
Department of Epidemiology
National Institute of Public Health
6-1 Shirokanedai 4 Chome
Minato-ku Tokyo 108, Japan

I read this article with great interest. The purpose of this letter is to point out that there is another score method[1] for constructing confidence intervals for the difference in correlated proportions, which I think could be included in the list of *better methods* the author has recommended.

Let us consider a random paired-sample (a, b, c, d) from a multinomial distribution with underlying probabilities $(q_{11}, q_{12}, q_{21}, q_{22})$. An example of the data structure and the probability model is given in Table I. The score based $100(1-\alpha)$ per cent confidence limits for the difference $\lambda = \pi_N - \pi_S = q_{12} - q_{21}$ are the two solutions to the equation

$$\frac{b - c - n\lambda}{\sqrt{\{n(2\hat{q}_{21} + \lambda(1-\lambda))\}}} = \pm Z_{\alpha/2} \tag{1}$$

where $Z_{\alpha/2}$ denotes the upper $\alpha/2$ percentile of the standard Normal distribution and $\hat{q}_{21}$ is the maximum likelihood estimator for q_{21}:

$$\hat{q}_{21} = \frac{\sqrt{(B^2 - 4AC)} - B}{2A}$$

where $A = 2n$, $B = -b - c + (2n - b + c)\lambda$ and $C = -c\lambda(1-\lambda)$.

The plus and minus signs of equation (1) indicate the lower limit L and the upper limit U, respectively. Table II shows 95 per cent confidence intervals for chosen combinations of $a + d$, b and c which are the same as those in Table II of this article. The characteristics of this method are summarized as follows:

无异常 1. No anomalies. Estimated limits cannot violate the $[-1, 1]$ bounds on the difference between proportions, or produce zero-width intervals. For the case of zero off-diagonal cells ($b = c = 0$) to which most existing methods cannot apply, we have

$$[L, U] = \left[-\frac{Z^2_{\alpha/2}}{n + Z^2_{\alpha/2}}, \frac{Z^2_{\alpha/2}}{n + Z^2_{\alpha/2}} \right].$$

It should be noted that when $n = 0$, that is, we have no information, this interval becomes a quite reasonable interval $[-1, 1]$.

计算简单 2. Computationally simple. These two limits can be easily found by the secant method (see, for example, Gart and Nam[2]).

良好的经验覆盖概率 3. Good empirical coverage probabilities. The empirical coverage probabilities are generally quite close to the nominal confidence level $(1 - \alpha)$, however, the method is *conservative* when one of the off-diagonal cell probabilities is near zero.

Table I. An example of 2×2 frequency table for the paired-sample design. Parameters in parentheses indicate the expected proportions

		Standard treatment		
		Responded	Non-responded	Total
New treatment	Responded	a (q_{11})	b (q_{12})	$a+b$ (π_N)
	Non-responded	c (q_{21})	d (q_{22})	$c+d(1-\pi_N)$
	Total	$a+c(\pi_S)$	$b+d(1-\pi_S)$	n (1)

Table II. The score-based 95 per cent confidence intervals for selected combinations of $a+d$, b and c which are the same as those in Table II of Newcombe's paper

Cell frequencies			95 per cent	
$(a+d)$	b	c	confidence interval	
36	12	2	0·0611,	0·3448
36	14	0	0·1748,	0·4167
2	97	1	0·8698,	0·9866
0	29	1	0·6666,	0·9882
2	98	0	0·9068,	0·9945
0	30	0	0·7730,	1·0000
54	0	0	− 0·0664,	0·0664

Furthermore, I would like to comment that this method provides us with a unified set of procedures for the paired-sample study design: the test statistic for *any difference* (for example, clinical equivalence test) and *the sample size formula* in addition to the confidence interval estimator. In our original paper,[1] the calculation of approximate sample size was not discussed, although its derivation is straightforward. For example, the sample size n required for the clinical equivalence hypothesis (equivalence is defined as *not more than* $100\Delta_0$ *per cent inferior*, significance level $= \alpha$ and power $= 1-\beta$) of the difference in proportions:

$$H_0 : \pi_N - \pi_S = q_{12} - q_{21} = -\Delta_0 \ (<0)$$

$$H_1 : \pi_N - \pi_S = q_{12} - q_{21} = -\Delta_1 \ (> -\Delta_0)$$

is given by

$$n = \left\{ \frac{2q_{21}^* - \Delta_0 - \Delta_0^2}{\Delta_0 - \Delta_1} \left(\frac{Z_\alpha}{\sqrt{(2q_{21}^* - \Delta_0 - \Delta_0^2)}} + \frac{Z_\beta}{\sqrt{(2q_{21} - \Delta_1 - \Delta_1^2)}} \right) \right\}^2 \tag{2}$$

where

$$q_{21}^* = \frac{\sqrt{(B'^2 - 4A'C')} - B'}{2A'}$$

and $A' = 2$, $B' = \Delta_1(1-\Delta_0) - 2(q_{21} + \Delta_0)$ and $C' = q_{21}\Delta_0(\Delta_0 + 1)$.

This formula is derived by considering the asymptotic properties of the partial derivative of the log-likelihood U evaluated by the maximum likelihood estimator under the null hypothesis

$$U = \left[\left. \frac{\partial L}{\partial \theta} \right|_{H_0 : \theta = -\Delta_0} \right] = \frac{(b-c)/n + \Delta_0}{(2\hat{q}_{21} - \Delta - \Delta^2)/n}$$

where $\theta = q_{12} - q_{21}$. The asymptotic properties are summarized as follows:

$$E_{H_0}(U) = 0$$

$$\lim_{n \to \text{large}} E_{H_1}(U) = \frac{n(\Delta_0 - \Delta_1)}{2q^*_{21} - \Delta_0 - \Delta_0^2}$$

$$\lim_{n \to \text{large}} \text{var}_{H_0}(U) = \frac{n}{2q^*_{21} - \Delta_0 - \Delta_0^2}$$

$$\lim_{n \to \text{large}} \text{var}_{H_1}(U) = \frac{n}{2q_{21} - \Delta_1 - \Delta_1^2}.$$

Therefore, the score test statistic is given by $U/\sqrt{\{\text{var}_{H_0}(U)\}}$ which has already been discussed in detail[1] and the sample size formula (2) is derived.

REFERENCES

1. Tango, T. 'Equivalence test and confidence interval for the difference in proportions for the paired-sample design', *Statistics in Medicine*, **17**, 891–908 (1998).
2. Gart, J. J. and Nam, J. 'Approximate interval estimation of the ratio of binomial parameters: a review and corrections for skewness', *Biometrics*, **44**, 323–338 (1988).

AUTHOR'S REPLY

In his 1998 paper Dr. Tango has set out a method for calculating a confidence interval for a difference between paired proportions, which appears to have excellent properties. Unlike methods 8 to 10 of my article, which are merely based on score intervals for single proportions, it is appropriate to describe this method directly as a score method, with the theoretical advantages that result. When applied to my seven illustrative examples, Dr. Tango's method gives generally fairly similar intervals to mine, in some cases identical. The method is evidently free from aberrations such as overshoot and inappropriate tethering as described in my paper. The mean empirical coverage probabilities for the parameter combinations chosen for inclusion in Table IV of Dr. Tango's article are excellent, from the standpoint of aligning mean coverage with the specified nominal $1 - \alpha$. Dr. Tango notes that his method compares very favourably for coverage with existing unconditional and conditional methods – which is exactly what I claim for my methods 5, 6 and 10. While it does not have the computational simplicity of my closed-form method 10, it is clearly more practicable than my more complex methods 5 and 6 when cell frequencies are large.

Accordingly, I thank Dr. Tango for the opportunity to state that, from the evidence presented in his article, I regard his method as an excellent one, and as such, I commend it to users and software producers.

ROBERT G. NEWCOMBE
Department of Medical Statistics
University of Wales College of Medicine
Heath Park
Cardiff CF4 4XN, U.K.

