云南大学学报（自然科学版） 2003, 25 (3): 202~206 CN 53-1045/N ISSN 0258-7971
Journal of Yunnan University

Sample size determination for testing nonzero difference of two proportions in matched-pair design*

TANG Niang-sheng
(Research Centre of Applied Statistics, Yunnan University, Kunming 650091, China)

Abstract: Two approximate sample size formulae are proposed for testing the null hypothesis of nonzero rate difference of two proportions in matched-pair design based on Tango's score test statistic. The formulae can be used to produce sample size estimates that guarantees a prespecified power of a hypothesis test at a certain significance level and controls the width of a confidence interval with a certain confidence level. Our empirical studies confirm that the proposed sample size formulae perform satisfactorily. A real example is used to illustrate our methods.

Key words: asymptotic inference; matched-pair design; power; sample size; score test
CLC number: O 212.2 **Document code**: A **Article ID**: 0258-7971(2003)03-0202-05
MSC(2000): 62F03

In a matched-pair study, we usually wish to know whether a new treatment is significantly better than or at least as effective as the standard one. The conventional significance testing of a null hypothesis of zero rate difference between the response rates for the two treatments is inappropriate when the intention of the trial is to establish either close equivalence or materially important difference. A null hypothesis appropriate for this situation is a prespecified nonzero rate difference. Statistical inference for testing a null hypothesis of nonzero difference in binomial trials has received much attention in recent years (see e. g. Farrington and Manning[1]; Yanagawa et al.[2]; Nam[3]; Lu and Bean[4]). However, when there are zero frequencies in the off-digonal cells under a matched-pair design, the statistics derived by the above cited authors become invalid. To solve this difficulty, Tango[5] derived a one-sided test statistic for testing the equivalence via nonzero rate difference of two proportions in the matched-pair study based on the efficient score method, and showed that the test had empirical significance levels closer to the nominal α-level than the other tests as given by Lu and Bean[4] via MonteCarlo simulation study. In addition, Tango[6] considered the score-based confidence intervals for the rate difference and sample size formulae, and pointed out that his confidence interval had better empirical coverage probability than those of the published methods including both unconditional and conditional ones. We note that one must specify the value of q_{21} to apply Tango's[6] sample size formula. However, in practice study, it is difficult to exactly know the value of q_{21}. Here, an alternative method is considered for calculating power and sample size.

The purpose of this article is to propose reliable method for calculating sample sizes for matched-pair study for unknown q_{21} based on Tango's[5] score statistic. Section 1 presents two different approaches for sample size calculations i. e. the significance test approach and the confidence interval approach (see Tang et al[7]). In Section 2, we investigate the accur-

* **Received date**: 2003-03-12
Foundation item: The National Natural Science Foundation(10226005).

racy of the proposed sample size formulae under different settings. The proposed approach is illustrated by a real example in Section 3.

1 Power calculation and sample size formula

Following Tango[5], we assume that each individual subject in the study is administrated both the new and standard tests. This results in paired data, and there are four possible outcomes for each pair. These outcomes can be represented in a 2 × 2 table (see Tab. 1).

Tab. 1 Data structure of a matched pair 2× 2 table

New test	Standard test		Total
	Response (+)	Nonresponse (−)	
Response (+)	$a(q_{11})$	$b(q_{12})$	$a+b(\pi_N)$
Nonresponse (−)	$c(q_{21})$	$d(q_{22})$	$c+d(1-\pi_N)$
Total	$a+c(\pi_S)$	$b+d(1-\pi_S)$	$n(1.0)$

Here q_{11} is the probability that a positive response is observed for both treatments, q_{12} is the probability that a positive response is observed for the new treatment and anegative response for the standard treatment, etc. Then $q_{11}+q_{12}+q_{21}+q_{22}=1.0$. Let $\pi_N=q_{11}+q_{12}$ and $\pi_S=q_{11}+q_{21}$ be the respective sensitivities of the new and standard treatments. The numbers of subjects falling into the four cells are denoted by a, b, c and d as in Table 1. Following Tango[5], the equivalence of both treatments is inferred by testing the following hypothesis

H_0: $\pi_N=\pi_S-\Delta_0$ vs.

H_1: $\pi_N>\pi_S-\Delta_0$,

where $\Delta_0(>0)$ is a pre-specified acceptable difference in two proportions. The new treatment is concluded to be effective/noninferior when the null hypothesis is rejected. Some practical choices for Δ_0 include 0.05 or 0.1 (Tango[5]; Lu and Bean[4]).

To test hypothesis H_0, Tango[5] proposed the following score statistic

$$T=T(\Delta_0)=\frac{b-c+n\Delta_0}{\sqrt{n(2\hat{q}_{21}-\Delta_0(\Delta_0+1))}},\qquad(1)$$

which has asymptotically a standard normal distribution under H_0, where $\hat{q}_{21}$ is the maximum likelihood estimator of q_{21} under H_0 and satisfies

$$\hat{q}_{21}=\hat{q}_{21}(\Delta_0)=(\sqrt{B^2-4AC}-B)/(2A),$$

with $A=2n$, $B=-b-c-(2n-b+c)\Delta_0$ and $C=c\Delta_0(\Delta_0+1)$. Then, H_0 is rejected at the nominal level α if the statistic T is greater than or equal to $z_{(1-\alpha)}$, where $z_{(1-\alpha)}$ is the $100\times(1-\alpha)$ percentile point of the standard normal distribution.

Let $\Delta=\pi_N-\pi_S$. The expectation and variance of $b-c$ is respectively given by $E(b-c|H_1:\Delta=\Delta_1)=n\Delta_1$, $\mathrm{Var}(b-c|H_1:\Delta=\Delta_1)=n\{2q_{21}+\Delta_1(1-\Delta_1)\}$. Let $\tilde{q}_{21}$ be the maximum likelihood estimator of q_{21} under H_1. Similarly, it is easily shown that $\tilde{q}_{21}$ is the larger root of the quadratic equation $2nx^2-(b+c-(2n-b+c)\Delta_1)x-c\Delta_1(1-\Delta_1)=0$, $\tilde{q}_{21}$ is $\sqrt{n}$-consistent, and test statistic $(b-c-n\Delta_1)/\{n[2\tilde{q}_{21}+\Delta_1(1-\Delta_1)]\}^{1/2}$ has asymptotically the standard normal distribution under H_1. Therefore, for a true rate difference of the sensitivities $\pi_N-\pi_S=\Delta_1(>-\Delta_0)$, the asymptotic power function for T is given by $Pr\{T\geqslant z_{(1-\alpha)}|H_1:\Delta_1=\Delta_1\}=1-\Phi(u)$, where $u=[z_{(1-\alpha)}\{n(2\bar{q}_{21}-\Delta_0(\Delta_0+1))\}^{1/2}-n(\Delta_1+\Delta_0)]/\{n(2\bar{q}^*_{21}+\Delta_1(1-\Delta_1))\}^{1/2}$, where $\bar{q}_{21}$ and $\bar{q}^*_{21}$ are respectively the asymptotic limits of $\hat{q}_{21}$ and $\tilde{q}_{21}$ for sufficiently large n given a true difference $\Delta_1=\pi_N-\pi_S$, i.e. $\bar{q}_{21}=(B_0+\sqrt{B_0^2-8C_0})/4$ with $B_0=(2q_{21}+\Delta_1)+(2-\Delta_1)\Delta_0$ and $C_0=q_{21}\Delta_0(1+\Delta_0)$, and $\bar{q}^*_{21}=(E_0+\sqrt{E_0^2+8F_0})/4$ with $E_0=2q_{21}-\Delta_1(1-\Delta_1)$ and

$F_0 = q_{21} - \Delta_1(1 - \Delta_1)$, and $\Phi(\cdot)$ is the standard normal distribution function. Similar to Tango[6], the approximate sample size required for a power of $1-\gamma$ based on the score test can be shown to be

$$n_{TS} = \{(z_{1-\alpha}v_0^{1/2} + z_{1-\gamma}v_1^{1/2})/(\Delta_1 + \Delta_0)\}^2,$$

where $v_0 = 2\bar{q}_{21} - \Delta_0(1+\Delta_0)$, $v_1 = 2\bar{q}^*_{21} + \Delta_1(1-\Delta_1)$. To apply the sampl esize formula, we require the exact specification of the value of q_{21} under H_1. In practice, an investigator can usually specify the desirable sensitivities, π_N and π_S, but may not have complete knowledge of q_{21}. In this case, the sample size formula without the specification of q_{21} is desirable. Note that n_{TS} is an increasing function of q_{21} that satisfies the following inequality: $\max[0, -\Delta_1] \leqslant q_{21} \leqslant \min[(1-\Delta_1)/2, \pi_S]$. Hence, we could adopt the midpoint level of q_{21}, given as $\min[(1-\Delta_1)/4, \pi_S/2]$ for $\Delta_1 \geqslant 0$ and $\min[(1-3\Delta_1)/4, (\pi_S - \Delta_1)/2]$ for $\Delta_1 < 0$ to obtain the midpoint sample size (denoted as n_{TM}) (cf. Lu and Bean[4], Tang et al.[7]).

Next, we consider the sample size determination based on the method controlling the width of a confidence interval with acertain confidence level. Following Tango[6], the $(1-\alpha)\times 100\%$ confidence interval for the risk difference $\Delta = \pi_N - \pi_S$ based on the score statistic T is given by $T^2(\Delta) = \chi^2_{1,\alpha}$, where $\chi^2_{1,\alpha}$ is the upper α-percentile of the central chi-square distribution with 1 d. f. It is easily shown from (1) that the lower and upper limits of the confidence interval are the two roots of the following quadratic equation: $A_1\Delta^2 + B_1\Delta + C_1 = 0$, with $A_1 = n(n+\chi^2_{1,\alpha})$, $B_1 = n[2(b-c) + \chi^2_{1,\alpha}]$, $C_1 = (b-c)^2 - 2n\chi^2_{1,\alpha}\hat{q}^*_{21}$, and $\hat{q}^*_{21} = \hat{q}_{21}(\Delta)$, as defined in (1). Thus, the half width of the confidence interval is given by

$$w = \frac{\sqrt{4[n(b-c) + 2n^2\hat{q}^*_{21} - (b-c)^2]\chi^2_{1,\alpha} + [1+8\hat{q}^*_{21}]n(\chi^2_{1,\alpha})^2}}{2\sqrt{n}(n+\chi^2_{1,\alpha})}.$$

Let $\tilde{q}_0$ be theasymptotic limit of $\hat{q}^*_{21}$ for a large n and given values of q_{21} and Δ, then the asymptotic limit of the right-hand side of the above equation can be expressed as

$$w = \frac{\{4n[2\tilde{q}_0 + \Delta(1-\Delta)]\chi^2_{1,\alpha} + (1+8\tilde{q}_0)(\chi^2_{1,\alpha})^2\}^{1/2}}{2(n+\chi^2_{1,\alpha})},$$

and $\tilde{q}_0 = [(B_2^2 - 8C_2)^{1/2} + B_2]/4$,

with $B_2 = 2q_{21} + (3-\Delta)\Delta$,

$C_2 = q_{21}\Delta(1+\Delta)$.

Therefore, the desired sample size n_{CS} based on the score statistic T is given by

$$n_{CS} = [B_3 + \sqrt{B_3^2 + A_3C_3}]\chi^2_{1,\alpha}/(2A_3),$$

where $A_3 = w^2$, $B_3 = 2\tilde{q}_0 + \Delta(1-\Delta) - 2w^2$, and $C_3 = 1 + 8\tilde{q}_0 - 4w^2$. Similarly, without the knowledgevalue of q_{21}, we can adopt the midpoint level of q_{21} to obtain the midpoint sample size (n_{CM}) which is regarded as a compromise between the maximum or conservative (n_{CC}) and the minimum sample sizes.

2 Evaluation of Performance

To examine the accuracy of the above approximate power formula controlled sample size formula, we compute their respective exact powers under different settings of Δ_0, Δ_1 and q_{21} with $\alpha = 0.05$, $\pi_S = 0.8$ based on the sample sizes obtained from n_{TS}, n_{TC} and n_{TM}. The exact power for any particular sample size n at Δ_1 is computed by

$$\sum_{x\in R} Pr(x:p) = \sum_{x\in R} \frac{n!}{b!\,c!\,(n-b-c)!} \cdot q_{12}^b q_{21}^c (1-q_{12}-q_{21})^{n-b-c},$$

where $q_{12} = q_{21} + \Delta_1$, and $x' = (b, c)$, $p' = (q_{12}, q_{21})$, $R = \{x: 0 \leqslant b, c, b+c \leqslant n$ such that $T \geqslant z_{(1-\alpha)}\}$ are the sampling point, alternative hypothesis and critical region, respectively. For calculations of the actual size, we simply replace Δ_1 by Δ_0. Table 2 reports the results for various settings of Δ_0, Δ_1 and q_{21} with nominal power being 90%, of one-sided test at 5% significance level. In general, the power controlled sample size formula could provide fairly accurate sample size estimates in the sense that the exact power based on the estimated sample size is usually pretty close to the nominal power. Generally, the sample size n_{TS} is sufficient to guarantee the desired power. In all cases, the midpoint sample size

seems to provide a reasonable sample size estimation without prior information of q_{21}. Table 3 reports the desired sample size based on n_{CS}, n_{CM} and n_{CC} to control the half width of a 90% confidence interval at $w = 0.01, 0.05$ and 0.08 for various true values of Δ and q_{21} with $\pi_S = 0.8$.

3 Numerical examples

Consider a numerical example adapted from an investigation of whether a particular body fluid gives results equivalent to the testing of plasma and analysed by Lachenbruch & Lynch[8]. The data are reported in Table 4. In this trial, we are interested in the equivalence of two test. We may consider the testing of alternative body fluid as effective as the testing of plasma of a decrease of the result of testing by alternative body fluid is no more than 5 per cent. Under the null hypothesis H_0: $\Delta = 0.05$, we obtain the MLE of q_{21} is $\hat{q}_{21} = 0.052$, and the one-sided score statistic for testing H_0: $\pi_N = \pi_S - 0.05$ against $\pi_N > \pi_S - 0.05$ is $z = 6.03$ (p-value < 0.01). Therefore, we reject the null hypothesis and conclude that the alternative body fluid produces results equivalent to plasma samples for the investigation. This is the same as Lachenbruch & Lynch's[8] result. Here, we want to know whether the present study has sufficiently large sample size for the exist-

Tab. 2 Controlling power sample sizes calculated by score (1) for nominal power being 80 percent of a one-tailed test for H_0: $\Delta = \Delta_0$ against H_1: $\Delta = \Delta_1$ with $\pi_S = 0.8$ at $\alpha = 0.05$ level and corresponding exact powers(%) and α-levels (%)

Δ_0	Δ_1	q_{21}	n_{TS}	Exact		n_{TM}	Exact		n_{TC}	Exact	
				power	size		power	size		power	size
0.0	0.05	0.10	852	90.08	4.98	1 795	90.02	5.04	3 423	90.01	5.01
0.0	0.05	0.30	2 223	90.34	5.01	—	—	—	—	—	—
0.0	0.20	0.10	81	90.62	4.95	124	90.54	5.12	210	90.42	5.07
0.0	0.20	0.30	167	89.98	4.99	—	—	—	—	—	—
0.05	0.00	0.10	698	90.17	4.98	1 713	90.12	5.07	3 422	90.10	5.01
0.05	0.00	0.30	2 054	89.92	5.01	—	—	—	—	—	—
0.05	0.10	0.10	115	91.03	5.10	208	90.13	5.14	378	90.48	4.95
0.05	0.10	0.30	265	90.10	5.16	—	—	—	—	—	—

Tab. 3 Sample size for 90% confidence intervals of half width $w = 0.01, 0.05$ and 0.08 with $\pi_S = 0.8$

Δ	q_{21}	$w = 0.01$			$w = 0.05$			$w = 0.08$		
		n_{CS}	n_{CM}	n_{CC}	n_{CS}	n_{CM}	n_{CC}	n_{CS}	n_{CM}	n_{CC}
0.00	0.10	5 412	13 527	27 053	218	540	1081	86	211	421
0.00	0.30	16 232	—	—	648	—	—	253	—	—
0.10	0.10	14 338	20 442	32 194	572	816	1 286	223	318	501
0.10	0.30	24 303	—	—	970	—	—	378	—	—

Tab. 4 Plasma compared to alternative body fluid

		Plasma sample		
		+	−	Total
Alternative body	+	446	5	451
fluid sample	−	16	690	706
	Total	462	695	1 157

ing test procedures to detect a nonzero rate difference at 0.05 nominal level with power 0.90. To answer this, we set q_{21} = 0.1, Δ_0 = 0.01, Δ_1 = 0.0, γ = 0.1 and α = 0.05, the desired sample size is n_{TS} = 17 150. Without the knowledge of value of q_{21}, the corresponding conservative sample size is given by 85 668, while the respective midpoint sample size is given by 42 836. Suppose an investigator would like to adopt the confidence interval approach and would like to guarantee the half width of the resultant 90% test-based confidence intervals being controlled at w = 0.05 with Δ = 0.0 and q_{21} = 0.1. In this case, the desired sample size is n_{CS} = 217. Whilst n_{CC} = 1 081, n_{CM} = 540.

References:

[1] FORRINGTON C P, 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[J]. Statistics in Medicine, 1990, 9(12): 1 447—1 454.

[2] YANAGAWA T, TANGO T, HEIJIMA Y. Mantel-Haenszel-Type test for testing equivalence or more than equivalence in comparative clinical trials[J]. Biometrics, 1994, 50(3): 859—864.

[3] NAM J. Sample size determination in stratified trials to establish the equivalence of two treatments[J]. Statistic-sin Medicine, 1995, 14(18): 2 037—2 049.

[4] LU Y, BEAN J A. On the sample size for one-sided equivalence of sensitivities based upon McNemar's test [J]. Statistics in Medicine, 1995, 14(17): 1 831—1 839.

[5] TANGO T. Equivalence test and confidence interval for the difference in proportions for the paired-sample design [J]. Statistics in Medicine, 1998, 17(8): 891—908.

[6] TAMGP T. Improved confidence intervals for the difference between binomial proportions based on paired data [J]. Statistics in Medicine, 1999, 18(24): 3 511—3 513.

[7] TANG M L, TANG N S, CHAN I S F, et al. Sample size determination for establishing equivalence/noninferiority via ratio of two proportions in matched-pair design[J]. Biometrics, 2002, 58(4): 957—963.

[8] LACHENBRUCH P A, LYNCH C J. Assessing screening tests: extensions of McNemar's test[J]. Statistics inMedicine, 1998, 17(19): 2 207—2 217.

配对设计试验中检验两个比值非零差的样本量确定*

唐年胜

(云南大学 应用统计研究中心, 云南 昆明 650091)

摘要: 对配对设计试验, 基于 Tango(1998) 得分统计量导出了检验 2 个比值非零差的 2 种近似样本量公式. 由这些公式得到的样本量能达到预先指定的功效和控制置信区间的宽度. 一个实例和一些经验结果验证了方法的有效性.

关键词: 渐近推断; 配对设计; 功效; 样本量; 得分检验

* 作者简介: 唐年胜(1968—), 男, 博士后, 副教授, 主要从事数据统计、生物医学统计研究.