STATISTICS IN MEDICINE, VOL. 11, 1521-1525 (1992)

ON THE SAMPLE SIZE FOR STUDIES BASED UPON McNEMAR'S TEST

PETER A. LACHENBRUCH

UCLA Department of Biostatistics, Los Angeles, CA 90024-1772, U.S.A.

SUMMARY

When computing the sample size for studies using McNemar's test, one needs to know the probability of discordance and the odds ratio to be detected. In many studies, the investigator is unable to specify the probability of discordance, but can state, at least approximately, the marginal probabilities of each variable. This information leads to restrictions on the possible values of the cell probabilities and provides a range of admissible values for the off-diagonal cells. We compute the sample size needed in these circumstances and compare them to the results cited by Schlesselman and Connett *et al.* These sample sizes for the method are quite close to those found in the Monte Carlo study of Connett *et al.*

INTRODUCTION

A new and inexpensive assay to detect human monoclonal antibody is being compared with a standard assay. The standard assay detects the antibody with probability p_{1+} equal to 0·9. The new assay detects the antibody with probability p_{+1} between 0·7 and 0·85. A detection probability of 0·70 is not acceptable. What sample size is needed to detect a difference between the assays at the 0·05 significance level with a power of 0·9 when p_{+1} equals 0·7?

McNemar's test is designed for this classical comparison of two correlated binomial proportions. Examples of its use include 'before and after' studies, or comparisons of new and standard assays performed on the same biological samples. The test can be derived in several different ways. Consider the table of counts with cell probabilities in parentheses (Table I).

We may treat this table as a set of multinomial counts with probabilities p_{ij} for the ith row and jth column. The marginal probability in the ith row is p_{i+} and that in the jth column is p_{+j}. The null hypothesis is usually written as H_0: $p_{0+} = p_{+0}$, which, of course, implies the same relationship between p_{1+} and p_{+1}. Either of these implies that $p_{01} = p_{10}$. The alternative is usually that of inequality. Using this model, all of the information is contained in the discordant pairs. Conditional on being discordant, the test is equivalent to testing if $p_{10}/(p_{01} + p_{10}) = 0{\cdot}5$, and under the null hypothesis the test statistic, $X^2 = (b - c)^2/(b + c)$, is chi-squared with 1 degree of freedom. The number of discordant pairs needed is readily computed for any alternative of interest, and the chance of being discordant is then used to compute the sample size needed for the experiment. Unfortunately, this probability is rarely known and must be guessed. A second way of studying this problem is to encode each subject's responses as 1 if yes', 0 if no' and to treat the mean difference between pairs of responses as a Normal random variable. A paired t test statistic is computed and compared to the t distribution. If one uses the expected value of the difference (which is 0) in the computation of the variance, the square of the t statistic is the same chi-squared test mentioned above.

0277-6715/92/111521-05$07.50

Received February 1991
Revised March 1992

Table I. Schema of paired binomial array

	New assay		
Standard assay	No ($j = 1$)	Yes ($j = 0$)	Total
No ($i = 1$)	a (p_{11})	b (p_{10})	$a + b$ (p_{1+})
Yes ($i = 0$)	c (p_{01})	d (p_{00})	$c + d$ (p_{0+})
Total	$a + c$ (p_{+1})	$b + d$ (p_{+0})	N (1)

Many procedures have been proposed to compute the sample size for this test. They require an assumption about the value of the odds ratio $\Phi = p_{10}/p_{01}$, and an estimate of the proportion of discordant pairs. For example, the Schlesselman[1] formula is written by Connett *et al.*[2] as

$$N_S = [Z_\alpha(\Phi + 1) + 2Z_\beta\sqrt{\Phi}]^2/[(\Phi - 1)^2(\Phi + 1)p_{01}],$$

where the Z_α and Z_β are percentiles of the normal distribution.

Connett *et al.*[2] propose an alternative formulation based on the t-test approach and derive the expression

$$N_C = [Z_\alpha\sqrt{(\Phi + 1)} + Z_\beta\sqrt{\{(\Phi + 1) - (\Phi - 1)^2 P_{01}\}}]^2/[(\Phi - 1)^2 p_{01}].$$

This derivation uses the sample variance of the differences rather than the null hypothesis variance, so the test statistic is slightly different from the usual McNemar test statistic.

Suissa and Shuster[3] provide an exact unconditional analysis of this problem, and give exact critical values with the corresponding power. Their table 4 corresponds to Table II in this paper. The samples sizes given by the unconditional analysis are slightly smaller than those given by Connett *et al.*[2]

Connor[4] presents results, points out bounds on the cell probabilities when the margins are fixed, and suggests a value for the probability of discordance, based on an independence argument, equal $p_{1+} + p_{+1} - 2p_{1+}p_{+1}$. We use these bounds on cell probabilities in our approach.

Parker and Bregman[5] generalize the problem to allow for a varying probability of discordance among the matched pairs. The formula they derive incorporates a factor that increases the sample size slightly.

AN ALTERNATIVE APPROACH

In many studies, the investigator is unable to specify the probability of discordance, but can state, approximately, the marginal probabilities, p_{1+} and p_{+1}. Under the null hypothesis, $p_{i+} = p_{+i}$, but under the alternative, the different values of p_{1+} and p_{+1} impose constraints on the values of p_{01} and p_{10}. From this, we may determine a range of possible alternative values: the concordant (1, 1) cell can have probabilities ranging from $\min(p_{1+}, p_{+1})$ to $(p_{1+} + p_{+1} - 1)$. Corresponding to each value of p_{11}, we have values of $p_{10} = p_{1+} - p_{11}$ and $p_{01} = p_{+1} - p_{11}$. These values of p_{01} and p_{10} give a way of computing the number of discordant pairs needed, and also the total number required. Note that this method does not specify an exact alternative, but only a range of possible alternatives are specified. The difference $p_{10} - p_{01}$ is given exactly by specifying the marginal probabilities, but the fraction discordant must lie in the range $1 - p_{1+} - p_{+1}$ to $p_{1+} + p_{+1} - 2\min(p_{1+}, p_{+1})$. Furthermore, the odds ratio p_{10}/p_{01} is constrained to lie within a

Table II. Sample sizes for conditional, unconditional and unadjusted approaches

p_{+1}	p_{1+}	min p_{11}	max p_{11}	s min	N_L min	N_L max	N_L mid	N_S min	N_S max	N_C min	N_C max
0·8	0·9	0·7	0·8	0·33	316	106	210	302	39	312	101
0·7	0·9	0·6	0·7	0·25	106	53	79	95	20	101	49
0·6	0·9	0·5	0·6	0·20	59	35	47	50	13	55	31
0·5	0·9	0·4	0·5	0·17	40	27	33	32	10	36	22
0·7	0·8	0·5	0·7	0·40	526	106	316	518	39	522	101
0·6	0·8	0·4	0·6	0·33	158	53	106	151	20	154	49
0·5	0·8	0·3	0·5	0·29	82	36	59	76	13	78	31
0·4	0·8	0·2	0·4	0·25	53	27	40	48	10	49	22

range of values. Assume that $p_{+1} < p_{1+}$. The conditional probability of $p_{01}/(p_{01} + p_{10})$ may be written as $s = (p_{+1} - p_{11})/(p_{+1} + p_{1+} - 2p_{11})$ and we can compute sample sizes based on the smallest and largest possible values of s. We can also include a value of s corresponding to p_{11} midway between these two values. We call this the midpoint method. The sample size in terms of s may be written:

$$N_L = 0{\cdot}25(Z_\alpha + Z_\beta)^2/[0{\cdot}5 - s)^2|p_{+1} + p_{1+} - 2p_{11}|].$$

We refer to N_L as the unadjusted approach. This uses the null variance to compute the probability of falsely accepting the null hypothesis. If one uses the variance under the alternative ($s(1 - s)$), the procedure is identical to that of Schlesselman (N_S). The probability of discordance is estimated as $p_{1+} + p_{+1} - 2p_{1+}p_{+1}$ by Connor, and as $p_{+1} + (p_{1+} - 1)/2$ for the midpoint method. Connor's estimate is smaller than the midpoint estimate if p_{+1} is greater than 0·25. This is the case in Table II and leads to larger required sample sizes using Connor's approach than using the midpoint method. Table II gives some examples for a two-sided test with significance level of 0·05 and a power of 0·9. Sample sizes are calculated corresponding to the minimum and maximum values of p_{11}. The values for s corresponding to the minimum p_{11} are given; for the maximum p_{11}, $s = 0$. The values given by the minimum, maximum and midpoint methods suggest to the user level of uncertainty resulting from the specification of the alternative hypotheses.

Returning to our example of the human monoclonal antibody assay, recall that we wish to detect a change in rate from 0·9 to 0·7 at the 0·05 level with a power of 0·9. From the second row of Table II, using the N_L columns, we see that at least 53 samples are needed, and at most 106 samples would be needed. The midpoint value of 79 samples is a reasonable compromise.

Parker and Bregman's[5] approach incorporates a factor that depends on the distribution of variability of exposure among the pairs. Assuming a beta distribution with parameters (a, b), the necessary sample size is increased by a factor $(a + b + 1)/(a + b)$. If the variation of exposure is great (small a and b), the sample size can be substantially increased, but only a small increase would be found for small variation of exposure (large a and b).

The uncertainty associated with the minimum and maximum values of p_{11} might be modelled by a prior distribution, and this would lead to a single answer for sample size. The proposed method gives slightly larger sample sizes than the Connett *et al.* method which in turn gives slightly larger values than the Schlesselman method. The Schlesselman approach yields smaller values in the cases in Table II, sometimes much smaller for very small odds ratios. Connett *et al.* noted the difference between the conditional and unconditional methods and compared the

Table III. Comparison of sample size requirements (Adapted from Connett *et al.*[2])

Odds ratio	p_{01}	Schlesselman	Connett	Monte Carlo	Unadjusted
1·5	0·10	1035	1047	1055	1052
1·5	0·15	690	697	701	701
2·0	0·15	201	207	204	211
3·0	0·15	63	66	65	71
4·0	0·15	34	35	34	39
1·5	0·20	518	522	517	526
2·0	0·20	151	154	152	158
3·0	0·20	48	49	48	53
4·0	0·20	25	25	24	30
1·5	0·30	345	347	345	351
2·0	0·30	101	100	101	106

results in a Monte Carlo study. This study simulated 2000 trials with sample size given by the formulae, and adjusted the sample size to a larger or smaller one to get the desired power. The studies indicated that the unconditional approach was generally closer to the value found in their Monte Carlo study. I compared the values with the unadjusted method for 90 per cent power. These are given in Table III. There are some minor differences in the numbers in my table compared to Connett *et al.* because I rounded sample size to the next higher integer. Apparently, Connett *et al.* rounded down if the fractional part was less than 0·5. For these cases, the results of the three methods are very close, and are also close to the Monte Carlo study results. The conditional method is the most conservative.

COMMENTS

When planning a study, a decision must be made about the sample size to obtain. As statisticians, we can determine an admissible range of possible alternative probabilities, which we can explain to the investigator. The plausible values, which must be a subset of the admissible ones, will determine the range of possible sample sizes. Deciding the sample size is the responsibility of the investigator, but needs the guidance of the statistician.

When there is a range plausible alternative hypotheses, using a prior distribution allows one to determine a specific recommended sample size. However, using the posterior mean or median of the estimated sample size leaves one with the question of how the prior was chosen. I have not used prior distributions for this reason.

There is little difference among the three methods and the methods of Parker and Bregman or Suissa and Schuster. In my experience, investigators usually arrive able to specify the marginal probabilities (if anything!), rather than the probability of being discordant. Indeed, estimating p_{01} and p_{10} is often an important by-product of such a study.

Finally, it is interesting to note that the formulae are asymmetric about an odds ratio of 1·0. For example, Table IV shows sample size computations for $\alpha = 0{\cdot}05$ and 90 per cent power. This occurs because the inverse odds ratio causes a differences in standard deviation units.

Table IV. Sample sizes for alternatives on opposite sides of 1

Odds ratio	p_{01}	N_S (Schlesselman)	N_C (Connett)	N_L (Lachenbruch)
0·5	0·2	302	312	316
2·0	0·2	151	154	158
0·33	0·2	142	154	158
3·0	0·2	48	49	53
0·25	0·2	100	113	117
4·00	0·2	25	25	29

ACKNOWLEDGEMENTS

This work was supported by CA-16042-15 from the National Cancer Institute. I thank Cindy Chang who brought this problem to my attention. I also thank Rob Weiss for several useful discussions, and the referees for their important comments.

REFERENCES

1. Schlesselman, J. J. *Case-Control Studies*, Oxford University Press, New York, 1982.
2. Connett, J. E., Smith, J. A. and McHugh, R. B. 'Sample size and power for pair-matched case-control studies', *Statistics in Medicine*, **6**, 53–59 (1987).
3. Suissa, S. and Shuster, J. J. 'The 2 × 2 matched-pairs trial: exact unconditional design and analysis', *Biometrics*, **47**, 361–372 (1991).
4. Connor, R. J. 'Sample size for testing differences in proportions for the paired-sample design', *Biometrics*, **43**, 207–211 (1987).
5. Parker, R. A. and Bregman, D. J. 'Sample size for individually matched case-control studies', *Biometrics*, **42**, 919–926 (1986).