Fisher´s Mid-P-Value Arrangement in 2×2 Comparative Trials

Martín Andrés, A. *, Sánchez Quevedo, M. J. **, Silva Mato, A. ***

* Bioestadística. Facultad de Medicina, Universidad de Granada,. 18071 Granada, Spain. E-mail: amartina@goliat.ugr.es

** Estadística. Universidad de Cádiz, 11003 Cádiz, Spain

*** Bioestadística. Facultad de Medicina. Universidad de Alcalá de Henares, 28871 Alcalá de Henares, (Madrid), Spain

Abstract: Martín and Silva (1994) studied nine existing unconditional methods for comparing two proportions (independent samples), selecting Z methods (based on the classic chi-square statistics) and F1 (based on Fisher's exact test) as the optimals, because their power is a higher than that of their competitors in computation time, and not too much lower than Barnard's optimal method (the original B or its approximation B′). However Z or F1 are optimals depending on the value of K (which depends on the imbalance of the sample sizes), which means that a program is need which deals with both methods. In this paper the authors study thirteen new methods and show that the new method based on Fisher's mid-p value is a solution halfway between methods Z and F1 (for every values of K), which it frequently surpasses in power, and approaching the B and B′ methods, especially in large samples (where B and B′ can not be applied). The authors conclude that the arrangement based on Fisher's exact test mid-p value (for one- and two-tailed versions) is the optimal, because it maintains n adequate balance between its power and the computation time required.

(SSNinCSDA 29, 107-115 (November 1998))

Keywords: Barnard's test. Chi-square statistics. Comparison of two proportions. Fisher's mid-p-value. Unconditional test. 2×2 tables.

Received: July 1997 *Revised:* September 1998

1. Introduction

If x_i (i=1, 2) are observed values of two independent binomial random variables $x_i \rightarrow B(n_i;p_i)$, p_i the probability that an individual of the population i verifies the characteristic being studied and $y_i=n_i-x_i$, it is usual to refer to such an experience as a *comparative trial* (whose practical importance in all experimental sciences is well known). The aim is to test H_o: $p_1=p_2$ (=p), for which two possible methodologies can be employed: the conditional (Fisher, 1935) and the unconditional (Barnard, 1947). In this paper we adopt the unconditional point of view, but any reader interested in the debate should refer to the revisions by Martín (1991) and Sahai and Khurshid (1995).

Under H_o, the probability of a result (x_1, x_2) is:

$$P(x_1,x_2) = \binom{n_1}{x_1}\binom{n_2}{x_2} p^{a_1}(1-p)^{a_2}, \qquad (1)$$

with $a_1 = x_1+x_2$ and $a_2 = y_1+y_2$. If CR is a critical region formed by different pairs (x_1,x_2), then the type I error of the test (for a given p) is $\alpha(p) = \Sigma_{CR}P(x_1,x_2)$ and its size will be $\alpha^* = \max_{0<p<1} \alpha(p)$. The difficulty of obtaining α^* has delayed the development of the unconditional method for many years, but the diffusion and improvements in computing have allowed us to advance quite far in the last decade. The unconditional method posed two questions:

1) How can we reduce the computation time for the value α^*?; and
2) How can we form the CR to obtain an optimal test?

The first question has been resolved recently by Silva and Martín (1997). The second is the object of this work. A combination of both questions has been analyzed by Berger (1996).

Basically, our objective consists in defining an arrangement criterion for the pairs (x_1,x_2) of the sample space (which will define their order of entry in the CR), and which will greatly influence the computation time. Martín and Silva (1994) revised the procedure of the nine arrangement methods proposed in existing literature, among them method B (Barnard's original, 1947), method B′ (a simplification of method B), method Z (arranging from the largest to the smallest value of the chi-square statistics: Garside and Mack, 1967) and method F1 (arranging from the smallest to the largest value of Fisher's exact test one-tailed p-value: Boschloo, 1970). These authors have shown that the more powerful method is B, closely followed by B′ (which requires 8 or 9 times less computation time). Unfortunately both methods can require an excessive computation time in tables with moderate values of n_i. Searching for alternative methods, Martín and Silva selected the Z and F1 methods as the most powerful among the remaining seven methods in the study. Unfortunately, however, these were not found to be advisable in any given situation: Z was only found to be optimal when $K = \max n_i / \min n_i$ is small and the method F1 when K is large. The advantage is that their computation times are 10 times lower than those of B′ (even more so with moderate or large-sized samples).

The aim of this article is to find an arrangement method comparable in computation time to methods F1 and Z, but with a similar or greater power than any of them (for each value of K). This would allow us to carry out one test (instead of two) without any loss of power. To this end, 13 new arrangement methods are proposed, which will be compared amongst themselves and then against the methods previously stated: B, B′, Z and F1.

II. Old and new simplified arrangement methods

In order to decide easily the arrangement by which the pairs (x_1,x_2) are successively entered in the CR, it has been traditional procedure to employ either the habitual asymptotic statistics (the case of Z) or the exact conditional test (the case of F1). In this paper, other less habitual statistics and conditional versions will be obtained. The following will outline the old optimal F1 and Z methods and the other 13 new methods.

Barnard (1947) pointed out how useful it would be for an arrangement method to verify the properties of *convexity*- if (x_1,x_2) belongs to the CR, this also should be true of (x_1+1, x_2) and (x_1, x_2-1) in the case of H_1: $p_1 > p_2$- and of *symmetry* -if (x_1,x_2) is of the CR, this also should also be true of (y_1,y_2) in the case of H_1: $p_1 \neq p_2$. It can be shown that all the methods defined here verify the above properties (except the F2 and FM2 which do not verify convexity: this then has to be imposed as a priority condition so that they may perform coherently).

2.1 Methods related with the chi-square statistic

It is known that, under H_0, the statistic

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \rightarrow N(0;1)\ , \qquad (2)$$

with $\hat{p}_i = x_i/n_i$ and $q_i = 1-p_i$. Here and from now on, it is assumed that $\hat{p}_1 > \hat{p}_2$. As under H_o is $p_1 = p_2 = p$, it is usual to estimate p for $\hat{p} = a_1/n$ (with $n = n_1 + n_2$), and thus the denominator in (2) becomes $D = \{\hat{p}\hat{q}\,(1/n_1 + 1/n_2)\}^{1/2}$, with $\hat{q} = 1-\hat{p}$ so obtaining the classic statistic Z (pooled) and the Garside and Mack's arrangement Z method (1967). From here onwards, we will place the Z arrangement method in the chi-square format (which is more convenient), by which Z alludes to the arrangement based on:

$$\chi_Z^2 = (x_1 y_2 - x_2 y_1)^2 n / a_1 a_2 n_1 n_2\ . \qquad (3)$$

Goodman (1964) proposed estimating p_i individually, whereby which $D = \{\hat{p}_1\hat{q}_1/n_1 + \hat{p}_2\hat{q}_2/n_2\}^{1/2}$ and the method (known as Z unpooled) will become as G method and be based on:

$$\chi_G^2 = (x_1 y_2 - x_2 y_1)^2 n_1 n_2 / \{n_2^3 x_1 y_1 + n_1^3 x_2 y_2\}. \qquad (4)$$

The method was adopted as an unconditional arrangement by Suissa and Shuster (1985) and studied by Haber (1987) against the classic Z. Here, it is adopted once more, placing it alongside those methods to which it is conceptually close.

For reasons of asymptotic power, Sathe (1982) proposed the permutation of the sample sizes so that $D = \{\hat{p}_1\hat{q}_1 / n_2 + \hat{p}_2\hat{q}_2 / n_1)\}^{1/2}$ and the new S method is determined by the arrangement based on:

$$\chi^2_S = (x_1y_2 - x_2y_1)^2 / \{n_2x_1y_1 + n_1x_2y_2\} \ . \quad (5)$$

A similar possibility consists of permuting the estimations of $\hat{q}_i$ or of $\hat{p}_i$. One option is to associate the largest $\hat{p}_i\ \hat{q}_j$ product with the smallest n_i. Supposing that $n_1 \leq n_2$, and given that it has already been agreed that $\hat{p}_1 > \hat{p}_2$ we would then obtain $D = \{\hat{p}_1\hat{q}_2 / n_1 + \hat{p}_2\hat{q}_1 / n_2\}^{1/2}$ and the new method N1 will then be given as a result of the arrangement based on:

$$\chi^2_{N1} = (x_1y_2 - x_2y_1)^2 / \{n_2x_1y_2 + n_1x_2y_1\}. \quad (6)$$

If, on the other hand, $\hat{p}_i\ \hat{q}_j$ is associated with the largest n_i, the new N3 method is obtained with an arrangement based on:

$$\chi^2_{N3} = (x_1y_2 - x_2y_1)^2 / \{n_2x_2y_1 + n_1x_1y_2\}. \quad (7)$$

It can be seen that if $n_1 = n_2$ then G≡S and N1≡N3.

Shuster (1992), in the case of 2×2 multinomial trials, used the Yates' statistic as an arrangement criterion for the exact test, showing in this way that better results can be obtained than with the classic arrangement Z. In the present case, there are many possible continuity corrections, but one of the most effective, and one which alters the order given by Z (Martín and Silva, 1996), is that of Schouten *et al.* (1980), which results in the following US method:

$$\chi^2_{US} = \frac{\{|x_1y_2 - x_2y_1| - \min(n_1, n_2)/2\}^2 n}{a_1a_2n_1n_2} \quad (8)$$

2.2 Methods related to the Student's statistic

If the random variable X is defined by 0 or 1 depending on whether the characteristic under study IS or IS NOT verified, the test for H_0: $p_1 = p_2$ becomes a test for the comparison of two means ($H_0 \equiv \mu_1 = \mu_2$) and, under H_0, is:

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0;1), \quad (9)$$

with σ_i^2 the populational variances of X in each one of the groups. The pooled parallel solution to (3) was proposed as an asymptotic method by D'Agostino *et al.* (1988), thus giving the chi-square statistic $(x_1y_2 - x_2y_1)^2\{(n-2)/n\}/\{n_2x_1y_1 + n_1x_2y_2\}$ which, being equal to (5) (except for the constant factor), results in the same arrangement as the S method.

The parallel asymptotic solutions to (4) and (5) were quoted by Martín *et al.* (1992) and now result in the arrangement methods L and M defined by:

$$\chi^2_L = \frac{(x_1y_2 - x_2y_1)^2(n_1 - 1)(n_2 - 1)}{n_2^2(n_2 - 1)x_1y_1 + n_1^2(n_1 - 1)x_2y_2} \quad (10)$$

$$\chi^2_M = \frac{(x_1y_2 - x_2y_1)^2(n_1 - 1)(n_2 - 1)}{n_2^2(n_1 - 1)x_1y_1 + n_1^2(n_2 - 1)x_2y_2} \ . \quad (11)$$

The parallel asymptotic solutions to (6) and (7), for the previous L and M methods, obtain the N2 and N4 methods respectively:

$$\chi^2_{N2} = \frac{(x_1y_2 - x_2y_1)^2(n_1 - 1)(n_2 - 1)}{n_1n_2\{(n_2 - 1)x_1y_2 + (n_1 - 1)x_2y_1\}} \quad (12)$$

$$\chi^2_{N4} = \frac{(x_1y_2 - x_2y_1)^2(n_1 - 1)(n_2 - 1)}{n_1n_2\{(n_2 - 1)x_2y_1 + (n_1 - 1)x_1y_2\}} \ . \quad (13)$$

As before, note that when $n_1 = n_2$ then L≡M and N2≡N4.

2.3. Methods related to the arc sine transformation

The arc sine transformation has been used in the present context fundamentally as a way of determining sample sizes. As an arrangement criterion, this transformation would obtain the A1 method based on:

$$\chi^2_{A1} = \left(\text{sen}^{-1}\sqrt{\hat{p}_1} - \text{sen}^{-1}\sqrt{\hat{p}_2}\right)^2 4n_1n_2/n \ , \quad (14)$$

or, using Anscombe's improvement, the A2 method:

$$\chi^2_{A1} = \left(\text{sen}^{-1}\sqrt{\hat{p}'_1} - \text{sen}^{-1}\sqrt{\hat{p}'_2}\right)^2 \frac{(2n_1 + 1)(2n_2 + 1)}{n + 1} \quad (15)$$

with $\hat{p}'_i = (x_i + 3/8) / (n_i + 3/4)$.

2.4. Methods related to the Fisher's exact test

The Fisher's exact test is based on the hypergeometric probability:

$$P(x_1) = \binom{n_1}{x_1}\binom{n_2}{x_2} \Big/ \binom{n}{a_1} , \quad (16)$$

with $r=\max(0;a_1-n_2) \le x_1 \le \min(a_1;n_1)=s$. If $\hat{p}_1 > \hat{p}_2$ then $x_1 > E=a_1n_1/n$ (the mean of x_1), and the one-tailed p-value for the alternative H_1: $p_1>p_2$ is:

$$F1(x_1) = \sum_{i=x_1}^{s} P(i) , \quad (17)$$

which thus results in the classic F1 method. For the two-tailed test (H_1: $p_1 \neq p_2$) the value $x'_1<E$ should first be determined so that $P(x'_1) \le P(x_1)$ and $P(x'_1+1) > P(x_1)$, thus the p-value will be:

$$F2(x_1) = \sum_{i=r}^{x'_1} P(i) + \sum_{i=x_1}^{s} P(i) , \quad (18)$$

which, in a two-tailed test, gives a no more powerful arrangement than the F1 method (Martín and Silva, 1994) and thus will not be considered here.

As has already been mentioned, the F1 method (applied to one- or two-tailed test) produces an arrangement which is relatively effective in high values of K. Haber (1986) proposed the adaptation of Fisher's exact test to the criterion of Lancaster's mid-p-value (1952), thus obtaining Fisher's mid-p value, and Barnard (1989), Routledge (1992) and Upton (1992) supported the opinion. Here, we propose using the mid-p as a criterion of an unconditional arrangement, in the hope that it will improve the performance of the F1 method (making it good for any value of K).

In a general way, the mid-p consist in making the experimental point (x_1) enter the CR with half its probability. Thus, for H_1: $p_1>p_2$, Fisher's mid-p-value for the experimental value $x_1> E$ will be (Haber, 1986):

$$FM1(x_1) \sum_{i=x_1+1}^{s} P(i) + \frac{1}{2}P(x_1) , \quad (19)$$

this will then result in the FM1 arrangement method. For two-tailed test, Fisher's mid-p-value will be given by (Hirji *et al.*, 1991):

$$FM2(x_1) = \sum_{i=r}^{x'_1} P(i) + \sum_{i=x_1}^{s} P(i) + \begin{cases} \frac{1}{2}P(x_1) & \text{if} \quad P(x_1) > P(x'_1) \\ 0 & \text{if} \quad P(x_1) = P(x'_1) \end{cases} \quad (20)$$

thus leading to the arrangement method FM2.

III. Selection of the optimal method

3.1. Procedure for the selection.

The comparison of a test with another (to equality of error α) is made by comparing the power of both. The comparison of a test with another (to equality of error α) is made by comparing the power of both. As the power varies according to the considered alternative (p_1,p_2), Martín and Silva (1994) defined the concept of *long term power* $\theta(\alpha)$ as the mean power attained in all the parametric space, assuming that the parameters p_i will follow a uniform distribution. For the two-tailed test, $\theta(\alpha)$ is proportional to the number of points in the CR; for the one-tailed test the expression is a little more complicated (an average of Fisher's p-values). As the power $\theta(\alpha)$ varies according to the value of the objective error α, the same authors suggested the computation of the *mean power* $\theta(\alpha_1,\alpha_2)$, in determined intervals of the error α (in their article the appropriate formulae are indicated). Finally, considering determined values of n and of (α_1,α_2), Martín and Silva carried out the selection of methods B, B′, Z and F1 in the way that was indicated in the introduction.

In order to obtain comparable results, this article adopts the same work methodology and the same basic values used by Martín and Silva (1994). In this way, the average of the mean power of each method was obtained (for their subsequent comparison): **a)** For a one- or two-tailed test; **b)** For the intervals of α of (0%;1%), (1%;5%), (5%;10%); **c)** For values of n in the intervals of 6-14, 16-24, 27-33, 37-43 y 48-52; **d)** For values of K in the intervals 1; 1-1,25; 1,25-1,50; 1,50-1,75; 1,75-2,25; 2,25-3; 3-4,25 y 4,25-6.

The aim of the following points is two-fold:

A) Select the optimal method amongst 13 news proposed methods; and

B) Compare the above with Martín and Silva's optimals (1994) with comparable computation time (Z and F1) and evaluate the loss of power produced in the comparison with the optimals of greater computation time (B and B′).

3.2. Selection amongst the new methods

The average of the mean power attained in each one of the previous specifications and by the eleven new methods (G, S, US, L, M, N1, N2, A1, A2 and FM1 for one-tailed tests; these same, plus the FM2, for two-tailed tests) are shown in Table 1(a) (the remaining tables are available from the authors). In order to make the comparisons easier, the previous data has been shown graphically. Figure 1(a) is representative of the graph obtained in this way (the remainder are also available from the authors). Methods N3 and N4 performed worse (especially in not small values of K) than their homonyms N1 and N2, and have thus been excluded from the tables and graphics in order not to complicate them further. Methods L, G and A1 have also been excluded from the graphs, as these performed worse than the others.

The following general conclusions can be extracted from the analysis of such data:

1) All the methods (except A1, which performed badly) have a very similar power to those values of K close to 1, increasing their differences with the increase in K.

2) From the one-tailed tests, and in a very general way, four groups of methods can be observed, which, when ordered from best to worst, are: GI (FM1 and A2), GII (N2 and N1), GIII (US, S and M) and GIV (A1, G and L). The GIV group contains extremely bad methods. In the other groups, differences in power increase with K. So, it can be seen that the unpooled methods (G and L) performed very badly, but improve appreciably (moving up from GIII) if their n_i are permuted (S and M methods) and still further (moving up from GII) if their $\hat{q}_i$ are permuted (N2 and N1 methods).

3) For two-tailed tests the situation is more complex, as methods US and S (and the new FM2 method) are incorporated into GI and the methods of GII are now worse than those of GIII.

4) Comparing method S (or the equivalent method D) with Martín and Silva's (1994) data for method Z (shown later), it can be seen that both methods are practically the same: the permutation of the n_i allows Z-unpooled (method G) to be transformed into an S method which is comparable with the classic Z. The continuity correction (US) improves the performance of the classic Z.

With respect to the selection of the optimal method, the following can be concluded:

<u>One-tailed test:</u>

In general, the most powerful method is FM1, although method A2 is close (especially when α is moderate or when n or K is small). Methods N2 and N1 (in that order) follow at a greater distance.

<u>Two-tailed test:</u>

In general, methods S, US, A2, FM1 and FM2 were found to be the best. A more general selection would advise using A2 for values of $K \leq 1.5$; S for $1.5 < K \leq 2.5$ and FM2 for values of $K > 2.5$. Given that method FM2 usually performs appreciably better than the rest for large values of K, and only slightly worse in small and medium values of K, a choice coherent with the above results for one-tailed tests would be method FM2 (although it shows a slight loss power around $K = 2$, especially in high values of α).

Consequently, if there is a method able to compete with Z and F1 for any value of K, it is method FM (in its versions FM1 and FM2 for one and two-tailed test). Particularly, for one-tailed tests, method FM1 is much better than Z (since FM1 is optimal with respect to S and this is equivalent to Z).

3.3. New versus old methods

Table 1(b) shows the average of the mean power attained by the classic optimal methods (B, B′, Z and F1) and the new optimal methods (FM1 or FM2 according to whether it is a one- or two-tailed test). Its graphic representation is shown in Figure 1 (b). Complete data can be requested from the authors. The following conclusions can be extracted from an analysis of such data:

<u>One-tailed test:</u>

1) Method Z is always the worst, although circumstantially (in values of K close to 1) it can perform slightly better than its competitors.
2) Method F is slightly worse than FM1, B′ and B, especially when n or a is large.
3) Methods FM1, B¢ and B have a similar power, although B is systematically better, followed in general by B¢ (except in large values of K, when FM1 performs better).

<u>Two-tailed test:</u>

1) Methods B and B′ (practically the same in themselves) are clearly the best, although B is slightly better than B¢.
2) Of the other three methods, Z is best in K£3 (but only slightly better than FM2) and FM2 is best in K>3 (and appreciably so), although for values α–0%-1%, method FM2 is always better.

Consequently, for one-tailed tests it is clear that method FM1 is the optimal: a) Against its competitors in computation time (Z and F1), due to its greater power; b) Against higher computation time methods (B′ y B), as it is only slightly less powerful than them. For two-tailed tests, method FM2 is the optimal against its competitors possessing the same computation time (Z and F1), but is slightly worse than the other two (B′ and B, which are practically the same).

IV. Discussion and Conclusions

Barnard's unconditional method (B) for comparing two proportions through two independent samples presents the problem of the great computation time required. Martin and Silva (1994) resolved the problem in two senses: a) Proposing an intermediate computation time method (B′), which is almost as powerful as B (especially for two-tailed tests); and b) Selecting two other methods (Z and F1) of a more practicable computation time (from amongst those already published). The disadvantage of methods B and B′ is that in large samples they are impractical. The disadvantage of methods Z and F1 is two-fold: on the one hand, their power is noticeably lower than that of methods B′ and B; on the other hand, their relative performance depends on the value of K (the quotient of the sample sizes), for that reason one of these alone can not be selected as an optimal.

In this paper, 13 new arrangement methods have been studied, all with a low computation time, with the result that the arrangement based on Fisher's exact test mid-p-value (in its one- or two- tailed versions) combines all the characteristics necessary for it to be considered the optimal:

a) Its computation time is comparable to that of methods Z and F1;

b) Its power is greater than that of Z and F1 in one-tailed tests and greater or nearly equal in the two-tailed tests; c) Its power clearly approaches that of methods B and B′ (much more costly to evaluate), being practically the same in the one-tailed test and slightly lower in the two-tailed test. To sum up, the methods known here as FM1 (one-tailed test) and FM2 (two-tailed test) are the optimals, bearing in mind the balance of power/ computation time. The conclusions are reliable as they do not vary with the increase of n and the values of α and K considered here are fairly exhaustive.

Finally, we wish to point out that Berger (1996) has developed a procedure which is intended to improve the power of any arrangement method with a lower computation time. Basically, the procedure consists of obtaining α^*, maximizing $\alpha(p)$ not in the $0<p<1$ interval, but in an exact confidence interval of p (although to compensate for the possible lack of precision, it is necessary to add to α^* the error β of the interval). He applied this procedure to methods Z and F1, but it could be applied to any of the present methods, particularly to those selected as optimals. However Berger has not yet proved that his method verifies the key property of "convexity", and this means that for the present it is of doubtful applicability.

References

BARNARD, G.A. (1947): 'Significance tests for 2×2 tables'. *Biometrika 34*, 123-138.

BARNARD, G.A. (1989): 'On alleged gains in power from lower P-values'. *Stat. in Medicine 8(12)*, 1469-1477.

BERGER, R.L. (1996): 'More powerful tests from confidence interval p values'. *The Amer. Stat. 50(4)*, 314-8.

BOSCHLOO, R.D. (1970): 'Raised conditional level of significance for the 2×2 Table when testing the equality of two probabilities'. *Statistica Neerlandica 24(1)*, 1-35.

D'AGOSTINO, R.B., CHASE, W. and BELANGER, A. (1988): 'The appropriateness of some common procedures for testing the equality of two independent binomial populations'. *The Amer. Stat. 42(3)*, 198-202.

FISHER, R.A. (1935): 'The logic of inductive inference'. *J.R.S.S. A 98*,39-54.

GARSIDE, G.R. and MACK, C. (1967): 'Correct confidence limits for the 2×2 homogeneity contingency table with small frequencies'. *The New Journal of Statistics and Operational Research 3(2)*, 1-25.

GOODMAN, L. (1964): 'Simultaneous confidence intervals for contrast among multinomial populations'. *Annals of Mathematical Statistics 35*, 716-725.

HABER, M. (1986): 'A modified exact test for 2×2 Contingency Tables'. *Biometrical Journal 28(4)*, 455-463.

HABER, M. (1987): 'A comparison of some conditional and unconditional exact tests for 2×2 contingency tables'. *Commun. in Stati. -Simul. 16(4)*, 999-1013.

HIRJI, K.F., TAN, S.J. and ELASHOFF, R.M. (1991): 'A quasi-exact test for comparing two binomial proportions'. *Stat. in Med. 10*, 1137-1153.

LANCASTER, H.O. (1952): 'Statistical control of counting experiment'. *Biometrika 39*, 419-422.

MARTIN ANDRES, A. (1991): 'A review of classic non-asymptotic methods for comparing two proportions by means of independent samples'. *Comm. Stat. -Sim. and Comp. 20(2&3)*, 551-583.

MARTIN ANDRES, A. and SILVA MATO, A. (1994): 'Choosing the optimal unconditioned test for comparing two independent proportions'. *Comp. Stat. and Data Anal. 17*, 555-574.

MARTIN ANDRES, A. and SILVA MATO, A. (1996): 'Optimal correction for continuity and conditions for validity in the unconditional chi-squared test'. *Comput. Statis. & Data Anal 21*, 609-626.

MARTIN ANDRES, A., SILVA MATO, A. and HERRANZ TEJEDOR, I. (1992): 'A critical review of asymptotic methods for comparing two proportions by means of independent samples'. *Comm. Stat.-Sim. and Comp. 21(2)*, 551-586.

ROUTLEDGE, R.D. (1992): 'Resolving the conflict over Fisher's exact test'. *The Canadian Journal of Stat. 20(2)*, 201-209.

SAHAI, H. and KHURSHID. A. (1995): 'On analysis of epidemiological data involving a 2×2 contingency table: an overview of Fisher's exact test and Yates' correction for continuity'. *Jour. of Biopharmaceutical Statist. 5(1)*, 43-70.

SATHE, Y.S. (1982): 'Another test for equality of two proportions'. *Comm. in Stat. Simul. and Comp. 11(3)*, 373-375.

SCHOUTEN, H.J.A., MOLENAAR, I.W., VAN STRIK, R., BOOMSA,A. (1980): 'Comparing two independent binomial proportions by a modified chi-squared test' *Biometrical Journal 22(3)*, 241-248.

SHUSTER, J.J. (1992): 'Exact unconditional tables for significance testing in the 2×2 multinomial trial'. *Statistics in Medicine 11(7)*, 913-922. Correction in 11, 1619 (1992).

SILVA MATO, A. and MARTIN ANDRES, A. (1997): 'Simplifying the calculation of the P-value for Barnard's test and its derivatives'. *Stat. and Comp. 7(2)*, 137-143.

SUISSA, S. and SHUSTER, J.J. (1985): 'Exact unconditional samples sizes for the 2×2 binomial trial'. *Journal of the Royal Statistical Society A 148(4)*, 317-327.

UPTON, G.J.G. (1992): 'Fisher's exact test'. *Journal of the Royal Statistical Society A 155(3)*, 395-402.

Table 1

For each method (first column) and each interval of values for $K = n_2 / n_1 \geq 1$ (first row) within each subtable, the average of the mean powers $\bar{\theta}$ attained in all tables verifying the imposed conditions and for $27 \leq n \leq 33$ y $1\% < \alpha \leq 5\%$ is given. Each interval for K begins at the number that figures to its left and ends in the one indicated by its column.

(Left = One-tailed test; Right = Two-tailed test)

(a)

Power of the new methods

Met/K	-1.00	≤1.25	≤1.50	≤1.75	≤2.25	≤3.00	≤4.25	≤6.00
G	47.74-41.67	47.57-42.36	45.55-41.66	42.22-40.07	37.85-36.41	31.56-30.75	24.75-23.75	19.54-16.89
L	47.74-41.67	47.55-42.35	45.39-41.59	41.95-39.94	37.27-36.02	30.51-29.89	23.71-22.83	18.82-16.62
S	47.74-41.67	47.79-42.46	47.41-42.29	45.97-41.76	43.71-40.48	40.73-38.41	36.86-35.21	28.86-28.88
M	47.74-41.67	47.77-42.47	47.31-42.34	45.77-41.83	43.33-40.56	40.15-38.24	35.91-34.92	27.39-27.42
US	47.69-41.59	47.83-42.19	47.47-41.99	46.54-41.58	44.62-40.23	41.46-38.55	37.73-35.54	30.88-35.58
N1	47.74-41.67	47.94-42.31	47.44-41.82	46.72-40.82	44.82-39.12	41.98-36.19	38.12-32.23	31.96-26.36
N2	47.74-41.67	47.94-42.31	47.44-41.79	46.71-40.75	44.82-39.08	42.01-36.19	38.13-32.22	32.00-26.34
A1	42.34-37.06	40.63-36.56	40.24-36.16	38.20-36.01	36.05-34.49	32.88-31.91	29.20-27.77	25.07-21.34
A2	47.94-42.01	47.88-42.59	47.71-42.40	46.94-41.70	45.54-40.14	43.25-37.67	40.02-34.15	34.82-28.69
FM1	47.69-41.59	47.79-42.30	47.53-41.95	46.93-41.95	45.36-40.01	43.36-37.63	40.21-34.67	35.23-30.25
FM2	-41.59	-42.34	-42.13	-41.59	-40.03	-38.39	-35.63	-31.51

(b)

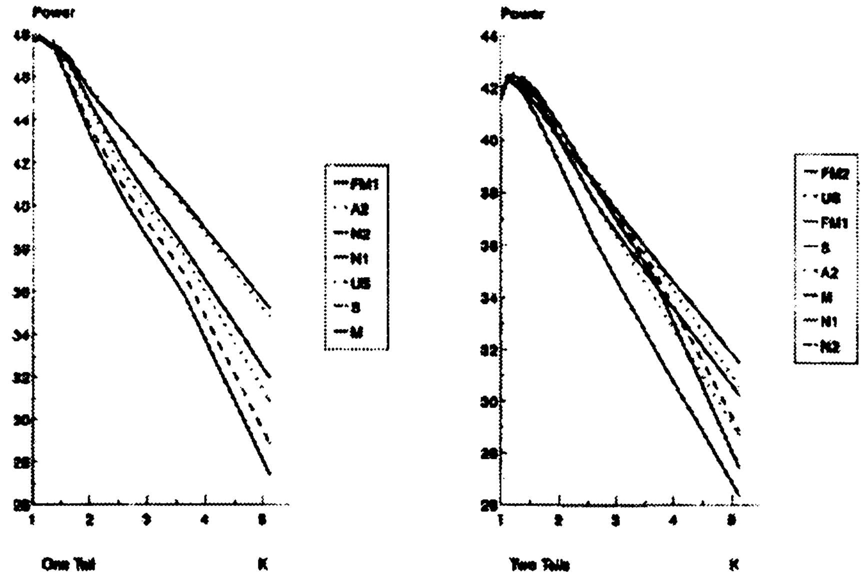
Power of the optimal methods present and traditional

ONE TAIL- TWO TAILS

	- 1.00	≤ 1.25	≤ 1.50	≤ 1.75	≤ 2.25	≤ 3.00	≤ 4.25	≤ 6.00
B	48.11-42.08	48.28-42.87	47.98-42.71	47.29-42.19	45.82-40.88	43.65-39.02	40.37-36.30	35.29-32.20
B'	48.03-42.07	48.18-42.79	47.88-42.69	47.22-42.16	45.73-40.85	43.56-38.95	40.23-36.27	35.10-32.19
Z	47.74-41.67	47.79-42.46	47.41-42.29	45.97-41.76	43.71-40.48	40.73-38.41	36.86-35.21	28.86-28.88
F1	47.68-41.59	47.49-42.00	47.18-41.67	46.72-41.04	45.22-39.84	43.12-37.48	40.01-34.60	35.08-30.33
FM1/2	47.69-41.59	47.79-42.34	47.53-42.13	46.93-41.59	45.36-40.03	43.36-38.39	40.21-35.63	35.23-31.51

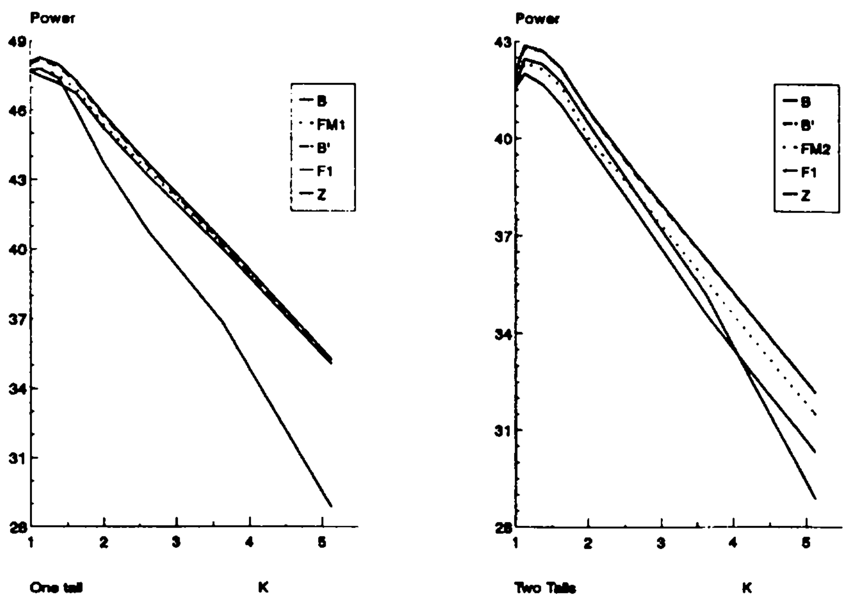
Figure 1

The average of the mean power $\overline{\theta}$ is represented in function of the value of $K = n_2 / n_1 \geq 1$ for each of the tests and methods indicated ($27 \leq n \leq 33$; $1\% < \alpha \leq 5\%$). Their order of entry is that of their appearance, from the right, in the graph).



(a)

Power of the new methods



(b)

Power of the optimal methods present and traditional