UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1993:2
ISSN 0349-8034

EXACT PROPERTIES OF McNEMAR'S TEST IN SMALL SAMPLES

by

Robert Jonsson

Statistiska institutionen

Göteborgs Universitet

Viktoriagatan 13

S-411 25 Göteborg

Sweden

SUMMARY

The exact distribution of McNemar's test statistic is used to determine critical points for two-sided tests of equality of marginal proportions in the correlated 2x2 table. The result is a conservative unconditional test which reduces to the conditional binomial test as a special case. Exact critical points are given for the significance levels 0.05, 0.01 and 0.001 with the sample sizes n=6(1)50. A computer program for tail probabilities makes the calculation of power easy. It is concluded that McNemar's test is never inferior to the conditional binomial test and that much can be gained by using the McNemar test if the main purpose is to detect differences between the marginal proportions in small samples. A further conclusion is that the chi-square approximation of McNemar's test statistic may be inadequate when $n \leq 50$. Especially the 5% critical points are constantly too small.

Key words: Exact unconditional test; Matched 2x2 table;
Nuisance parameter; Power.

1. Introduction

McNemar (1947) introduced a well known test for the null hypothesis of equality of the marginal proportions in the matched 2x2 frequency table. In text-books the reader is often recommended to use this test with large samples, say greater than 20, in which case the asymptotic chi-square distribution with one degree of freedom (d.f.) is believed to be adequate. With smaller samples a conditional binomial test is usually suggested (cf. Conover, 1980).

The present paper originates from an observation which has been made frequently during applied work. Namely, that the conditional binomial test may fail to reject the null hypothesis in cases when the latter is strongly rejected by McNemar's test or other unconditional tests such as the Likelihood ratio test. To find out whether this is due to a lack of agreement with the limiting distribution of the test statistic or to a genuine difference in power, one has to study the exact distribution of the unconditional test statistic.

Bennett and Underwood (1970) compared a few exact significance levels of McNemar's test with those predicted by the chi-square distribution. For three sample sizes and three particular values of the nuisance parameter specified by the null hypothesis, they found that the exact levels may exceed the levels determined from the chi-square distribution. Duffy (1984) made exact power calculations and Connett, Smith and McHugh (1987) performed simulation studies of the power to compare exact results with those based on asymptotic theory. In the latter two studies the critical points of the rejection region were determined from the asymptotic distribution. While the problem of finding exact critical points for unconditional tests has been solved for the unmatched case (Suissa and Shuster, 1985; Storer and Kim, 1990; Shuster, 1992), the same problem has remained unsolved in the matched case.

This paper presents exact critical points for the two-sided McNemar test. The nuisance parameter in the distribution is eliminated by maximizing the null power function over a domain of the nuisance parameter as described in Basu (1977). The result is a conservative unconditional test which contains the conditional binomial test as a special case. Section 2 provides some arguments for McNemar's test. Expressions for the exact distribution of the test statistic are given in Section 3, accompanied by a SAS program for numerical calculations. The critical points are given in Section 4, while Section 5 shows how to use the critical points for power calculations. Section 6 concludes with a discussion on the choice of test statistic when the main purpose is to detect differences between the marginal proportions in small samples.

2. Background on the matched 2x2 table

Consider n independent observations on the pair of random variables (Y_1,Y_2) with probabilities $P(Y_1=i,Y_2=j)=p_{ij}$ for i,j=0,1 in the table below:

		Y_2 (Response 2)		
		1	0	
Y_1 (Response 1)	1	p_{11}	p_{10}	p_{1+}
	0	p_{01}	p_{00}	p_{0+}
		p_{+1}	p_{+0}	1

The marginal probabilities $P(Y_1=1)$ and $P(Y_2=1)$ are $p_{1+}=p_{11}+p_{10}$ and $p_{+1}=p_{11}+p_{01}$, respectively. The observed frequency in cell (i,j) is n_{ij} for i,j=0,1. The marginal frequencies are n_{1+} and n_{+1} where $n_{1+}=n_{11}+n_{10}$, $n_{+1}=n_{11}+n_{01}$ and $n_{1+}+n_{0+}=n=n_{+1}+n_{+0}$.

McNemar's test statistic for $H_0:p_{1+}=p_{+1}$ may be written

$$T_{McN}= (n_{10}-n_{01})^2/(n_{10}+n_{01}) \tag{1}$$

and H_0 is rejected for large values of T_{McN}. The use of (1) can be motivated in several ways:

(i) (1) is identical with the square of $(\hat{p}_{1+}-\hat{p}_{+1})/(\hat{V}_0(\hat{p}_{1+}-\hat{p}_{+1}))^{1/2}$, where '^' is used to denote Maximum Likelihood (ML) estimators and $\hat{V}_0$ is an unbiased estimator of the variance of $\hat{p}_{1+}-\hat{p}_{+1}$ under H_0 (Snedecor and Cochran, 1980).

(ii) (1) is the chi-square goodness-of-fit statistic under H_0 with ML estimators inserted for the cell proportions (Bennett, 1967, 1968).

(iii) To show how (1) is related to the Likelihood ratio statistic, let c be the covariance between Y_1 and Y_2. Then $p_{11}=p_{1+}p_{+1}+c$, $p_{10}=p_{1+}(1-p_{+1})-c$, $p_{01}=p_{+1}(1-p_{1+})-c$ and $p_{00}=(1-p_{1+})(1-p_{+1})+c$. The likelihood is proportional to the product $\Pi(p_{ij})^{n_{ij}}$, where the p_{ij}'s are functions of p_{1+}, p_{+1} and c. The unrestricted ML estimators are $\hat{p}_{1+}=n_{1+}/n$, $\hat{p}_{+1}=n_{+1}/n$ and $\hat{c}=(n_{11}-n_{1+}n_{+1}/n)/n$. Under H_0 the ML estimators are $\hat{p}_{1+}=\hat{p}_{+1}=\hat{p}_0=(n_{1+}+n_{+1})/2n$ and $\hat{c}=(n_{11}-\hat{p}_0^2)/n$. These estimators inserted into the Likelihood ratio, Λ (likelihood under H_0 divided by unrestricted likelihood), yields

$$\Lambda=((n_{10}+n_{01})/2)^{n_{10}+n_{01}}/(n_{10}^{n_{10}}+n_{01}^{n_{01}}).$$

By putting $S=n_{10}+n_{01}$ and $D=n_{10}-n_{01}$ and by using the expansion $\log(1+x) \approx x-x^2/2$ one obtains $-2\log\Lambda=(S+D)\log(1+D/S)+(S-D)\log(1-D/S)\approx D^2/S=T_{McN}$.
(iv) Let z take the value +1 if $Y_1>Y_2$, the value 0 if $Y_1=Y_2$ and the value -1 if $Y_1<Y_2$ and let $\bar{z}$ and s_z denote the sample mean and standard deviation of n independent z's. Then the square of the large-sample statistic $\bar{z}\sqrt{n}/s_z$ is $(1-1/n)T_{McN}/(1-T_{McN}/n)\approx T_{McN}$ for large n.

An alternative to McNemar's test is to use the statistic n_{10}, conditionally on $n_{10}+n_{01}=n'$, with a binomial distribution to test the reformulated hypothesis $H_0':p'=p_{10}/(p_{10}+p_{01})=1/2$. This binomial test is sometimes called the exact version of McNemars test (cf. StatXact User Manual, 1991). This is not a correct description since the binomial test is only a special case of McNemar's test, as will be shown in this paper where the two tests are compared.

The statistic in (1) does not cover the case $n_{10}=0=n_{01}$, which may be likely in small samples. To handle also this case, the following natural extension of (1) is considered:

$$T=\begin{cases}0, & \text{if } n_{10}=0=n_{01}\\ T_{McN}, & \text{otherwise.}\end{cases} \qquad (2)$$

3. Exact Distribution of McNemar's Statistic

McNemar's statistic may be written d^2/s, where $d=|n_{10}-n_{01}|$ and $s=n_{10}+n_{01}$. Let $p(d,s)$ be the joint probability function (p.f.) of d and s. Then $p(d,s)=p(s)p(d|s)$, where $p(s)$ is a binomial p.f. with parameters n and $p_{10}+p_{01}$ while $p(d|s)$ is a conditional p.f.. If $n_{10}(s)$ is the value of n_{10} for given s, then $n_{10}(s)$ has a binomial p.f. with parameters s and $p_{10}/(p_{10}+p_{01})$. The p.f. of $n_{10}(s)$ is related to $p(d|s)$ in the following way:

$$p(d|s)=\begin{cases} P(n_{10}(s)=s/2), & \text{if } d=0 \\ P(n_{10}(s)=(s-d)/2)+P(n_{10}(s)=(s+d)/2), & \text{if } d>0. \end{cases}$$

Multiplication of the two p.f.'s yields

$$p(d,s)=\frac{n!}{(n-s)!\,(\frac{s+d}{2})!\,(\frac{s-d}{2})!}\,(p_{10}p_{01})^{\frac{s-d}{2}}\,(1-p_{10}-p_{01})^{n-s}\,\delta(d), \qquad (3)$$

where $\delta(d)=1$ if $d=0$ and $\delta(d)=p_{10}^{d}+p_{01}^{d}$. In (3) $0\leq d\leq s\leq n$ and for $p(d,s)$ to be nonzero, d and s have to be both even or both odd. This set of values of d and s will be denoted $A(d,s)$ in the sequel.

The p.f. of the statistic T in (2) can now be expressed

$$P(T=0)=\sum_{s\geq 0} p(0,s) \text{ and for } t>0,\ P(T=t)=\sum_{S} p(d,s), \qquad (4)$$

where S is the set of values of d and s such that $d^2/s=t$.

To illustrate how (4) is used, consider an example with $n=4$ and $p_{10}=p_{01}=1/3$. The possible values of $d^2/s=t$ together with the values of $p(d,s)$ (in braces) are shown below. From these the p.f. and tail probabilities are computed.

d \ s	0	1	2	3	4
0	0(1/81)		0(12/81)		0(6/81)
1		1(8/81)		$\frac{1}{3}$(24/81)	
2			2(12/81)		1(8/81)
3				3(8/81)	
4					4(2/81)

t	P(T=t)	$P(T\geq t)$
0	19/81	1.0000
1/3	24/81	0.7654
1	16/81	0.4691
2	12/81	0.2716
3	8/81	0.1235
4	2/81	0.0247

The same calculations are performed by the SAS program in the Appendix 1.

The p.f. in (4) depends on the two parameters p_{10} and p_{01}. When $p_{10}=p_{01}=p$ there is only one nuisance parameter and a simple expression can be obtained for the tail probability. Let t be an observed value of T and let s_r be a value in the main diagonal of the set A(d,s) which is 2r steps below t, i.e. $t=(s_r-2r)^2/s_r$ for $r=0,1,\ldots,[n/2]$. Then (3) gives

$$P(T=s_r)=p(s_r,s_r)=2\binom{n}{s_r}p^{s_r}(1-2p)^{n-s_r}.$$

Because of the relation

$$P(T=t=(s_r-2r)^2/s_r)=p(s_r-2r,s_r)=\binom{s_r}{r}P(T=s_r),$$

the p.f. of T at any argument can be expressed in terms of the p.f. of T at s_r. To find an expression for $P(T\geq t^*)$ one has to identify the set of values of r and s_r for which $(s_r-2r)^2/s_r\geq t^*=(s_{r^*}-2r^*)^2/s_{r^*}$, say $C(r,s_r)$. Then

$$P(T\geq t^*)=\sum_{C(r,s_r)}\binom{s_r}{r}P(T=s_r)=2p^n\sum_{r=0}^{r^*}\binom{n}{r}\sum_{v=0}^{n-s_r}\binom{n-r}{v}\left(\frac{1-2p}{p}\right)^v=$$

$$=2\sum_{r=0}^{r^*}\binom{n}{r}p^r(1-p)^{n-r}\sum_{v=0}^{n-s_r}\binom{n-r}{v}\left(\frac{1-2p}{1-p}\right)^v\left(1-\left(\frac{1-2p}{1-p}\right)\right)^{n-r-v}. \qquad (5)$$

The last expression in (5) is suitable for calculations based on binomial p.f.'s.

For the purpose of illustration, consider the previous example with n=4 and p=1/3. To find $P(T\geq 1)$ one notice that there are two diagonals in the set A(d,s) with values ≥ 1. The first appears in the main diagonal and thus $s_0=1$. The first value ≥ 1 in the diagonal which is two steps above the main diagonal appears in the column with s=4. So, $s_1=4$ and the set $C(r,s_r)$ consists of $(r,s_r)=(0,1)$ and (1,4). Using these values in (5) gives $P(T\geq 1)=0.4691$.

(5) makes it possible to study $P(T\geq t)$ as a function of p. For numerical calculations it may be easier to use (4).

4. Critical Points for McNemar's Test

To test the hypothesis $H_0:p_{10}=p_{01}=p$ one is interested in finding critical points, e.g. the 5% point $t_{.05}$ defined as the smallest t for which $P(T\geq t|p_{10}=p_{01}=p)\leq 0.05$. H_0 is rejected if the observed value of T exceeds the critical point. The latter will in general be dependent on the value of the nuisance parameter p. Here, two cases will be considered: p=1/2 and $0\leq p<1/2$.

4.1 The Case p=1/2

When p=1/2 no observations are possible in the cells (0,0) and (1,1) and $s=n_{10}+n_{01}=n$. Assume n_{10} to be the smallest of n_{10} and n_{01} and therefore $d=n-2n_{10}$. Then (5) gives, with p=1/2 and $r^*=n_{10}$,

$$P(T\geq t^*=(n-2n_{10})^2/n)=(1/2)^{n-1}\sum_{r=0}^{n_{10}}\binom{n}{r}. \qquad (6)$$

(6) is identical with the probability (or p-value) obtained for a two-tailed test of the hypothesis $H_0':p'=p_{10}/(p_{10}+p_{01})=1/2$ using the conditional binomial statistic $n_{10}|n_{10}+n_{01}=n$ (Lehmann, 1959).

To use the binomial test when $n>n_{10}+n_{01}=n'$ is thus the same as using McNemar's test whith reduced sample size, n', and assuming a value of the nuisance parameter, 1/2, which is designed for a very special case.

4.2 The Case $0\leq p<1/2$

When there are at least one observation in the cells (0,0) or (1,1) and without prior information about p, except that $0\leq p<1/2$, it seems reasonable to determine conservative 100α% points from the requirement that

$$\sup_{0 \le p < 1/2} P(T \ge t_\alpha | p_{10} = p_{01} = p) \le \alpha. \qquad (7)$$

Here, such critical points will be determined for n=6(1)50 and α=0.05, 0.01 and 0.001.

Consider the expression in (5) as a function of p for fixed n, say $f_n(p)$. Local maxima of $f_n(p)$ can be obtained by equating the derivative of $f_n(p)$ to zero. But, this gives rise to a polynomial equation in p of degree n-1 and numerical solutions may be inaccurate when n is large

Instead, some properties of the expression in (5) can be exploited. Obviously $f_n(p) \to 0$ as $p \to 0$. For small p an upper bound for $f_n(p)$ can be derived, as shown in the Appendix 2. From the latter it was concluded that $f_n(p) \le \alpha$ when $p \le 0.008$.

For larger values of p $f_n(p)$ was calculated at arguments with increments 10^{-6}. To check that $f_n(p) \le \alpha$ between these arguments, the following inequality, derived in the Appendix 2, was used:
For $h \le 10^{-6}$,

$$|f_n(p+h) - f_n(p)| \le nh(\,f_n(p) - f_{n-1}(p))/p + n(n-1)h^2/p^2. \qquad (8)$$

Finally it remained to check that $\lim_{p \to 1/2} f_n(p) = (1/2)^{n-1} \sum_{r=0}^{r^*} \binom{n}{r} \le \alpha$.

As an illustration, consider the determination of the 5% point $t_{.05}$ when n=50. By first running the SAS program in the Appendix 1 with $p_{10}=p_{01}=p=0.01(0.01)0.49$, the trial value $t_{.05}=169/43 \approx 3.93$ was obtained. To compute $P(T \ge 169/43)$ by means of (5) the following set of values of r and s_r was identified: 4(=s_0),8,11,14,16,19,22,24,27,29,32, 34,36,39,41,43,46,48(=s_{17}). A plot of $f_{50}(p)=P(T \ge 169/3)$, computed at arguments with increments 10^{-6}, is seen in Figure 1. $f_{50}(p)$ attains its largest value, 0.049727$\le$ 0.05, when p=0.424620. By means of (8) it was inferred that $f_{50}(p)$ could not differ from 0.049727 by more than

1.4×10^{-8}. The next lower possible value of $t_{.05}$, $196/50 \approx 3.92$, had to be rejected because $f_{50}(p) > 0.05$ for some values of p.

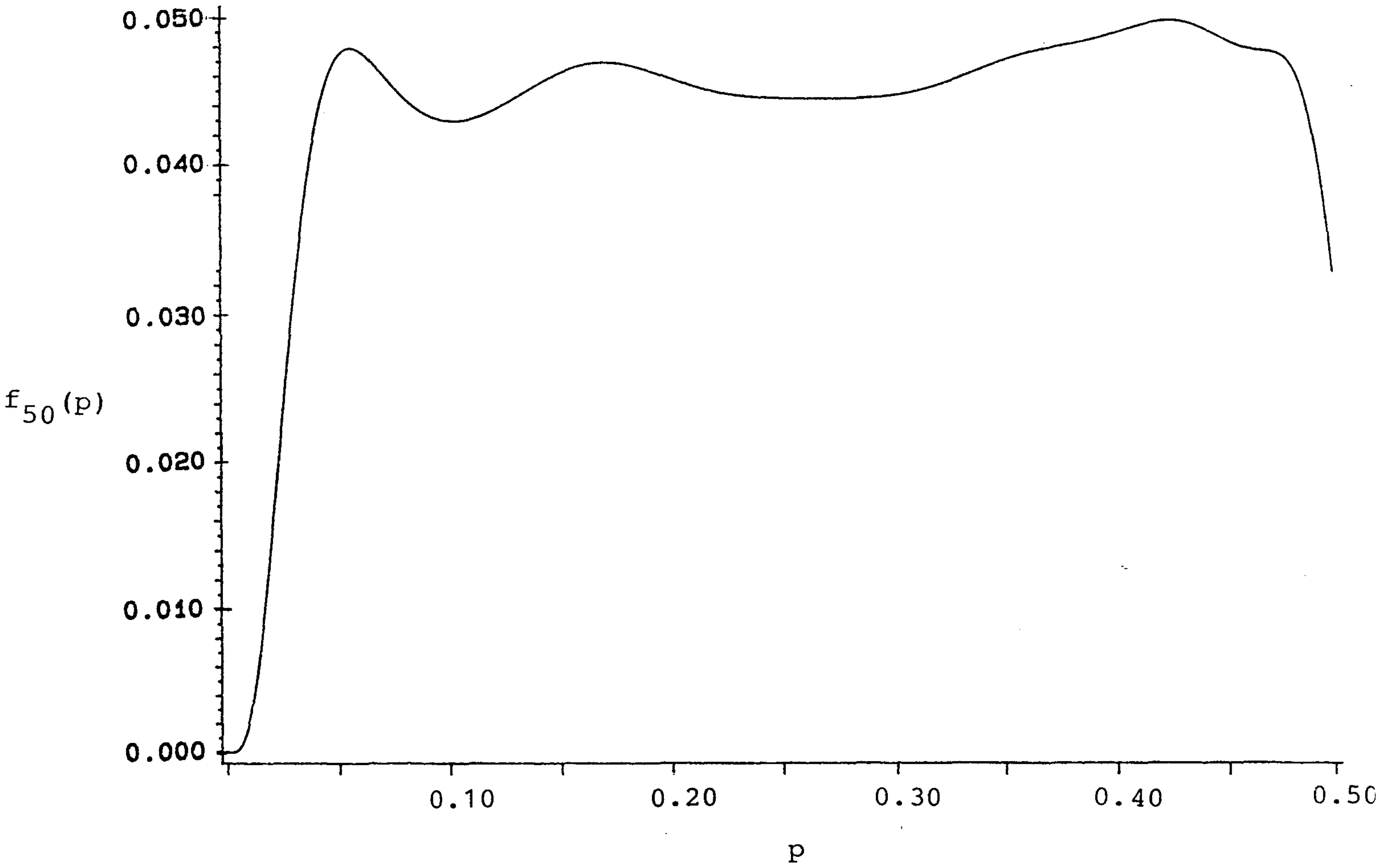

Figure 1. Plot of $f_{50}(p)=P(T\geq 169/43)$.

The critical points of McNemar's test are shown in Table 1. For each value of n the critical point when $p<1/2$ is smaller than or equal to the critical point when $p=1/2$. In 14 cases, among a total of 135, the points are equal. This happens when the supremum of $f_n(p)$ is attained at the boundary value when $p=1/2$. One example is $t_{.001}=11$ when $n=11$. Here (5) gives $f_{11}(p)=2p^{11}$, which is less than 0.001 if $p<0.5011$.

The critical points of McNemar's test can not be compared directly with those of the conditional binomial test based on the conditional sample size $n_{10}+n_{01}=n'\leq n$. But, since the binomial test based on n' observations is identical with McNemar's test based on n' observations and with p=1/2, it is clear from Table 1 that McNemar's test rejects H_0 more easily, possibly with a few exceptions where the tests are identical.

Most of the critical points for McNemar's test in Table 1 exceed the points $t_{.05}=3.84$, $t_{.01}=6.63$ and $t_{.001}=10.83$, determined from the chi-square distribution. Especially $t_{.05}=3.84$ turns out to be constantly too small.

Table 1

Sample size (n) and 5%, 1% and 0.1% critical points for McNemar's test with two choices of the nuisance parameter p

n	$t_{.05}$ p=1/2	$t_{.05}$ p<1/2	$t_{.01}$ p=1/2	$t_{.01}$ p<1/2	$t_{.001}$ p=1/2	$t_{.001}$ p<1/2
6	6	5	-	-	-	-
7	7	5	-	-	-	-
8	8	5	8	7	-	-
9	49/9	9/2	9	7	-	-
10	32/5	9/2	10	7	-	-
11	81/11	9/2	11	8	11	11
12	16/3	49/11	25/3	7	12	10
13	81/13	49/11	121/13	7	13	10
14	50/7	5	72/7	81/11	14	11
15	27/5	49/11	121/15	7	169/15	169/15
16	25/4	49/11	9	7	49/4	72/7
17	81/17	81/17	169/17	50/7	225/17	10
18	50/9	49/11	8	7	128/9	11
19	121/19	49/11	169/19	7	225/19	10
20	5	81/19	49/5	81/11	64/5	10
21	121/21	4	169/21	32/5	289/21	98/9
22	72/11	32/7	98/11	7	128/11	128/11
23	121/23	4	225/23	81/11	289/23	10
24	6	81/19	49/7	7	27/2	75/7
25	121/25	4	9	7	289/25	289/25
26	72/13	4	98/13	121/17	162/13	10
27	169/27	9/2	25/3	169/25	361/27	75/7
28	36/7	4	64/7	121/17	81/7	81/7
29	169/29	81/19	225/29	169/25	361/29	10
30	24/5	4	128/15	169/25	40/3	98/9
31	169/31	4	289/31	169/23	361/31	361/31
32	49/8	50/11	8	169/25	25/2	72/7
33	169/33	4	289/33	7	147/11	11
34	98/17	81/19	128/17	121/17	200/17	72/7
35	169/35	4	289/35	169/25	63/5	32/3
36	49/9	25/6	9	121/17	121/9	169/15
37	169/37	169/37	289/37	169/25	441/37	32/3
38	98/19	25/6	162/19	169/25	242/19	32/3
39	75/13	49/11	289/39	289/39	529/39	289/25
40	49/10	4	81/10	169/25	121/10	32/3
41	225/41	25/6	361/40	64/9	529/41	54/5
42	14/3	4	54/7	128/19	242/21	54/5
43	225/43	4	361/43	128/19	529/43	32/3
44	49/11	49/11	81/11	128/19	144/11	100/9
45	5	4	361/45	128/19	529/45	32/3
46	128/23	13/3	200/23	289/41	288/23	32/3
47	225/47	4	361/47	289/43	625/47	169/15
48	16/3	169/41	25/3	169/25	12	32/3
49	225/49	121/31	361/49	361/49	625/49	98/9
50	128/25	169/43	8	72/11	288/25	288/25

Notes: '-' indicates that no critical point exists. Critical points whith p=1/2 shall only be used when no observations are found in the cells (1,1) and (0,0).

5. Power Calculations

Let t_α be a 100α% critical point in Table 1 and let $P(T \geq t_\alpha)$ be the power of the two-sided test which rejects $H_0: p_{10} = p_{01}$ when observed values of T are greater than or equal to t_α. The power of this test regarded as a function of $p_{10} - p_{01} = p_{1+} - p_{+1}$ can easily be computed by means of the SAS program in Appendix 1.

To make a fair comparison between McNemar's test and the conditional binomial test regarding the ability to detect deviations in $p_{10} - p_{01}$ from zero, one has to compare the power of McNemar's test with the unconditional power of the binomial test,

$$\sum_{n'} P(\text{reject } H_0 | n_{10} + n_{01} = n') P(n_{10} + n_{01} = n'),$$

where $n_{10} + n_{01}$ has a binomial law with parameters n and $p_{10} + p_{01}$. This unconditional power obviously summarizes the performance of the conditional test in the long run.

Table 2 compares the powers of the two tests in two cases: One with a lower turn-over rate ($p_{10} + p_{01} = 0.20$) and one with a higher ($p_{10} + p_{01} = 0.80$). The sample sizes are n=30,40 and 50 while α is 0.05.

It is seen from the table that the power of McNemar's test is always greater than that of the binomial test, which never recovers from the bad start at $p_{10} - p_{01} = 0$. As might be expected, the greater gain by using McNemar's test is obtained when $p_{10} + p_{01}$ is small.

Table 2

Power of McNemar's test with p<1/2 compared to the power of the conditional binomial test

$p_{10}+p_{01}=0.80$

$p_{10}-p_{01}$	n=30 McN	n=30 Bin	n=40 McN	n=40 Bin	n=50 McN	n=50 Bin
.00	.0445	.0295	.0449	.0302	.0488	.0348
.10	.0846	.0605	.0995	.0741	.1212	.0945
.20	.2155	.1687	.2776	.2269	.3506	.2992
.30	.4403	.3728	.5596	.4957	.6728	.6191
.40	.7031	.6386	.8270	.7823	.9097	.8833
.50	.9021	.8663	.9660	.9515	.9901	.9854
.60	.9859	.9773	.9980	.9966	.9998	.9996

$p_{10}+p_{01}=0.20$

$p_{10}-p_{01}$	n=30 McN	n=30 Bin	n=40 McN	n=40 Bin	n=50 McN	n=50 Bin
.00	.0479	.0128	.0443	.0187	.0429	.0221
.02	.0537	.0154	.0520	.0232	.0527	.0285
.04	.0716	.0236	.0758	.0377	.0833	.0491
.06	.1032	.0387	.1183	.0646	.1380	.0878
.08	.1506	.0630	.1827	.1080	.2204	.1501
.10	.2166	.0993	.2718	.1724	.3324	.2412
.12	.3040	.1510	.3869	.2620	.4709	.3631
.14	.4147	.2217	.5250	.3787	.6258	.5120
.16	.5491	.3146	.6788	.5205	.7796	.6754
.18	.7052	.4316	.8341	.6791	.9097	.8313

6. Discussion

Today there is a wide-spread requirement of statistical significance when reporting statistical results. Significance is usually declared when the power under H_0 (p-value) reaches below 5%, 1% or 0.1%. Failure of achieving significance should obviously not merely be a result of using a test which is too weak.

This paper has shown that the binomial test, when testing the hypothesis H_0 of equal marginal proportions in the matched 2x2 table against two-sided alternatives, is a special case of McNemar's test when the nuisance parameter $p_{10}=p_{01}=p$ equals 1/2. For $p<1/2$ McNemar's test has smaller critical values and thus rejects H_0 more easily. The gain in power by using McNemar's test may be considerably, especially when the turn-over rate, $p_{10}+p_{01}$, is small. It was also shown that McNemar's test arises from appealing general test criteria when the purpose is to detect differences between the marginal proportions.

The problem of eliminating the nuisance parameter in finite samples can be solved in several ways (Basu(1977)). The binomial test is a result of eliminating p by conditioning. The popularity of the test may be due to computational convenience, to confusion with the fact that the test is uniformly most powerful for testing the reformulated hypothesis $H_0'{:}\,p'=p_{10}/(p_{10}+p_{01})=1/2$ against two-sided alternatives (Lehmann(1959)), or to statements favouring conditional tests in front of unconditional frequently made during 1980's (cf. Cox(1984) and Yates(1984)).

However, the nuisance parameter can also be eliminated by maximizing the power under H_0 of McNemar's test over a certain domain of p. This means that the worst possible configuration of p is taken into consideration to preserve the size of the test. The choice of a domain of p may sometimes be a problem. Here, the problem can be settled by noticing that p=1/2 implies that no observations are possible in the cells (1,1) and (0,0) under H_0. This naturally leads to the device: "If at least one observation is found in the cells (1,1) or (0,0), then use McNemar's test with the conservative critical values in Table 1 and otherwise the binomial test". Consider for instance an example with $n_{10}=0$, $n_{01}=5$ and $n=6$. Then H_0 can never be rejected at the 5% level by the binomial test, or equivalently by McNemar's test with p=1/2. But, since actually one observation is in one of the cells (1,1) or (0,0) it is hard to see the point in including p=1/2 into the domain of p. Since sup $P(T \geq 5 | p_{10}=p_{01}=p)=0.0387$ for $0 \leq p < 1/2$, obtained at $p=0.46$, it seems reasonable to reject H_0 at the 5% level.

The results in this paper agree with results reported earlier for tests of the same hypothesis using independent samples. Namely, that the unconditional (Z) test is stronger than the conditional (Fisher's exact) test (Suissa and Shuster (1985)). It should be stressed that this concerns the ability to detect differences between p_{1+} and p_{+1}. For other parametrizations, such as $p_{10}/(p_{10}+p_{01})$, $\ln(p_{10}/p_{01})$ or the odds ratio $p_{11}p_{00}/p_{10}p_{01}$ (cf. Frisén(1980)), other results are possible.

APPENDIX 1

A SAS program for the computation of tail probabilities, $P(T\geq t)$, of McNemar's test statistic when n=6 and $p_{10}=p_{01}=1/3$. Below is the corresponding outprint.

```
DATA A;
N=6;
DO V=0 TO FLOOR(N/2);
DO U=0 TO V;
SJ=2*V; DJ=2*U; SU=2*V+1; DU=2*U+1;
OUTPUT; END; END;
DATA BJ; SET A; S=SJ; D=DJ;
DATA BU; SET A; S=SU; D=DU;
DATA B; SET BJ BU;
IF S=0 THEN T=0; ELSE T=D*D/S;
IF S>N OR D>N THEN DELETE;
DATA C; SET B;
P10=1/3; P01=1/3;
G1=GAMMA(N+1); G2=GAMMA(N-S+1); G3=GAMMA(1+(S+D)/2);
G4=GAMMA(1+(S-D)/2); GP=G2*G3*G4; G=G1/GP;
KP=(P10*P01)**((S-D)/2)*(1-P10-P01)**(N-S);
IF D=0 THEN DELT=1; ELSE DELT=P10**D+P01**D;
PDS=G*KP*DELT;
DATA D; SET C;
PROC SORT; BY T; PROC MEANS NOPRINT SUM; VAR PDS; BY T;
OUTPUT OUT=DAT1 SUM=PT;
DATA E; SET DAT1;
SPT+PT; LT=1-SPT;
TAIL=LAG(LT); IF TAIL='.' THEN TAIL=1;
PROC PRINT; VAR T PT TAIL;
```

OBS	T	PT	TAIL
1	0.00000	0.23457	1.00000
2	0.33333	0.29630	0.76543
3	1.00000	0.19753	0.46914
4	2.00000	0.14815	0.27160
5	3.00000	0.09877	0.12346
6	4.00000	0.02469	0.02469

APPENDIX 2

An upper bound for $f_n(p)$ used in Section 4.2

Put $q=(1-2p)/(1-p)$ and let $B(n,q)$ denote a random variable with a binomial p.f. having parameters n and q. Then, from (5)

$$f_n(p) \leq 2 \sum_{r=0}^{r^*} P(B(n-r) \leq n-s_r).$$

In the latter sum the first term, $P(B(n,q) \leq n-s_0)$, is largest since $n-s_r \geq n-s_{r+1}+1$ (c.f. Feller(1968), p.173). Thus,

$$f_n(p) \leq 2(r^*+1)P(B(n,q) \leq n-s_0).$$

REFERENCES

Basu, D.(1977). On the elimination of nuisance parameters. Journal of the Americal Statistical Association 72, 355-366.

Bennett, B.M.(1967). Tests of hypotheses concerning matched samples. Journal of the Royal Statistical Society, Series B 29, 468-474.

Bennett, B.M.(1968). Note on χ^2 tests for matched samples. Journal of the Royal Statistical Society, Series B 30, 368-370.

Bennett, B.M. and Underwood, R.E.(1970). On McNemar's test for the 2x2 table and its power function. Biometrics 26, 339-343.

Connett, J.E., Smith, J.A. and McHugh, R.B.(1987). Sample size and power for pair-matched case-control studies. Statistics in Medicine 6, 53-59.

Conover, W.J. (1980). Practical Nonparametric Statistics, 2nd edition. New York: Wiley.

Cox, D.R.(1984). Discussion of Dr Yates' paper. Journal of the Royal Statistical Society, Series A 147, 451.

Duffy, S.W.(1984). Asymptotic and exact power for the McNemar test and its analogue with R controls per case. Biometrics 40, 1005-1015

Feller, W.(1967). An Introduction to Probability Theory and Its Applications, 3rd edition. New York: Wiley.

Frisén, M.(1980). Consequences of the use of conditional inference in the analysis of a correlated contingency table, Biometrika 67, 22-30.

Lehmann, E.L.(1954). Testing Statistical Hypotheses. New York: Wiley.

McNemar, G.(1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrica 12, 153-157.

Shuster, J.J.(1992). Exact unconditional tables for significance testing in the 2x2 multinomial trial. Statistics in Medicine 11, 913-922.

Snedecor, G.W. and Cochran, W.G.(1980). Statistical Methods, 7th edition. Ames, Iowa: Iowa State University Press.

StatXact, Version 2 (1991). Cambridge, Massachusetts: Cytel Software Corporation.

Storer, B.E. and Kim, C.(1990). Exact properties of some exact test statistics for comparing two binomial proportions. Journal of the American Statistical Association 85, 146-155.

Suissa, S. and Shuster, J.J.(1985). Exact unconditional sample sizes for the 2x2 binomial trial. Journal of the Royal Statistical Society, Series A 148, 317-327.

Yates, F.(1984). Tests of significance for 2x2 contingency tables. Journal of the Royal Statistical Society, Series A 147, 426-463.

1992:1	Teräsvirta, T., Tjøstheim, D. & Granger, C.W.J.	Aspects of Modelling Nonlinear Time Series
1992:2	Palaszewski, B.	A conditional stepwise test for deviating parameters
1992::3	Guilbaud, O.	Exact Semiparametric Inference About the Within--Subject Variability in 2 x 2 Crossover Trails
1992:4	Svensson, E. & Holm, S.	Separation of systematic and random errors in ordinal rating scales
1993:1	Frisén, M & Åkermo, G.	Comparison between two methods of surveillance: exponentially weighted moving average vs cusum