DIMACS Technical Report 2005-16
March 2005

EXTENDING POWER AND SAMPLE SIZE APPROACHES FOR MCNEMAR'S PROCEDURE TO GENERAL SIGN TESTS

Donald R. Hoover[1]
Department of Statistics, Rutgers University, 110 Frelinghuysen Road
Piscataway, NJ 08854

1. Supported by and National Science Foundation Grant EIA 02-05116

ABSTRACT

Current software and textbooks present procedures to estimate power and sample size for sign tests that only apply to settings where positive (i.e., X=1) or negative (i.e., X=-1), but not neutral (i.e., X=0) outcomes occur. However, many studies analyzed by sign tests involve the more general setting where significant amounts of neutral outcomes can occur. This paper illustrates application of existing power / sample size approaches and software that have been developed for matched binary responses (McNemar's discordant pairs) to general sign tests with neutral outcomes occurring. An application is made to a recent study that the author collaborated on.

Key Words: McNemar's Procedure, Power, Sample Size, Sign Test

1. INTRODUCTION

It is often important to evaluate whether an outcome (X) is more likely to be positive (X=1) than negative (X=-1) when neutrality (X=0) is also possible. If p_o, p_{+1} and p_{-1} are probabilities that (X=0), (X=1) and (X=-1), respectively, then irrespective of p_o, the null hypothesis is H_o: $p_{+1} = p_{-1}$. The alternative could either be one sided (i.e., H_a: $p_{+1} > p_{-1}$ or H_a: $p_{+1} < p_{-1}$) or two sided H_a: $p_{+1} \neq p_{-1}$. Without loss of generality, we focus here on the one sided H_a: $p_{+1} > p_{-1}$. For example, a study of couples that use illegal drugs asked each partner how many injection drug users the husband shared needles with. The difference between the numbers reported by each husband and his wife is categorized as: (X=1) if the husband reports more than the wife, (X=-1) if the husband reports less than the wife or (X=0) if both partners report equal numbers. Overall, a woman underestimates the husband's risk to acquire and later transmit to her diseases spread by needles if $p_{+1} > p_{-1}$.

While general sign tests to test H_o when neutrality (i.e., $p_o > 0$) is possible have been developed end evaluated (Coakley & Heise 1996, Radnor 1999) these are not implemented in current software that estimates sample size nor well known to many applied statisticians. Current power and sample size estimation software dedicated to planning studies with sign tests use formulas that restrict to $p_o = 0$ (c.f..; Dixon & Massey 1968, Noether 1987).

Although this may be largely forgotten, it has been shown (Cochran 1937, Dixon & Mood 1946) that McNemar's discordant pairs test (a common statistical procedure), is mathematically equivalent to a general sign test that permits neutral observations. For nearly 40 years, there has been a rich literature and implementation (c.f., Miettinen 1968, Schlesselman 1982, Connett et. al 1987, Dupont 1988, Suissa & Shuser 1991, Sahai & Khurshid 1996), including software (c.f.; Elashoff 2000, Hintze 2000, Dupont & Plummer 2001, O'Brien 2002, Borenstein et. al. 2003, Oloffson 2003) on sample size and power estimation for McNemar's discordant pairs. But as these new approaches and software are developed for McNemar's comparison, application to general sign tests with ties have not been described.

This paper thus extends several common and recent power and sample size estimation approaches developed for McNemar's matched binary responses to the general sign test. Section 2 formulates the matched binary response problem into an analytically equivalent sign test and from this applies existing sample size / power estimates for McNemar's procedure to sign tests. Section 3 gives a real example of using these procedures to estimate sample size and power for general sign tests.

2. EXTENDING MCNEMAR'S METHODS TO GENERAL SIGN TESTS

Some Notation for General Sign Test Power / Sample Size Estimation

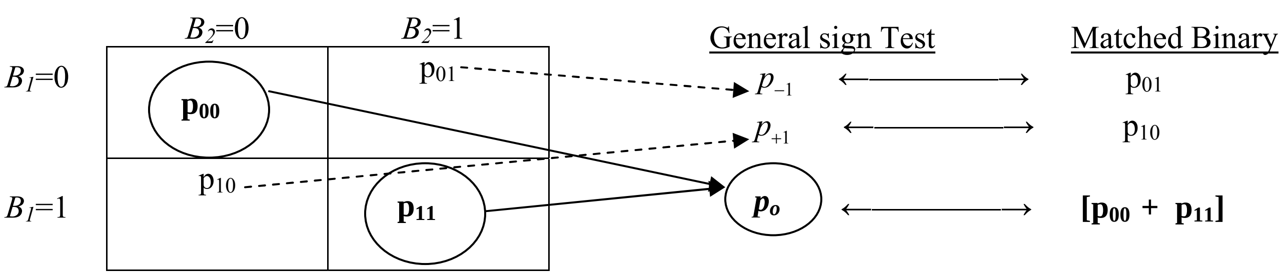
Let $n+$, $n-$ and n_0 be the numbers observed with ($X = 1$), ($X = -1$) and ($X = 0$), respectively with $m = n+ + n-$. Sign tests compare $n+$ to $n-$ with n_0 ignored (Hollander & Wolfe 1973). For a given m the distribution of $n+$ is $B(m, p)$ where $p = p_{+1}/(p_{+1} + p_{-1})$; or p=0.5 when H_o True . Thus H_o is tested by conditioning on m and using exact binomial tests or large sample approximations (c.f. Hollander & Wolfe 1973). For strict Type 1 error levels of α, it is often important to estimate either; (i) β the value of the Type 2 error to not reject H_o with a given N and specified values of p_{+1} and p_{-1} or (ii) the minimum N needed so Type 2 error does not exceed a given β for specified values of p_{+1} and p_{-1}. It is now assumed that the probability of equality (or neutrality) is a fixed value p_o and $w = (1- p_o) = (p_{-1} + p_{+1})$. Under H_o, $p_{-1} = p_{+1} = w/2$. For a specified H_a say $p_{+1} = p_{-1} + \Delta$ for a given $\Delta > 0$, $p_{+1} = (w+\Delta)/2$ and $p_{-1} = (w-\Delta)/2$.

McNemar's Matched Binary Response Design Presented as a General Sign Test

McNemar's approach compares N correlated pairs (B_1, B_2) of binary (i.e. $B_i = 0$ or 1) outcomes, where the first outcome has a generating characteristic (T_1) and the second has a different generating characteristic (T_2). For example, pairs could be generated by couples with B_1 the outcome of the wife (characteristic=T_1) and B_2 that of the husband (characteristic=T_2). It is of interest whether the first or second observation is more likely to be 1 or equivalently whether $X = (B_1 - B_2)$ is more likely to be positive than negative. As outcomes are correlated, statistical tests are performed pairs. Let n_{00}, n_{01}, n_{10} and n_{11} respectively, be observed number of pairs for which; both outcomes are 0, the 1st outcome is 0 and 2nd outcome is 1, the 1st outcome is 1 and 2nd outcome is 0, and both outcomes are 1, respectively. Similarly, let p_{00}, p_{01}, p_{10} and p_{11}, respectively, be probabilities that; both outcomes are 0, the 1st is 0 and 2nd is 1, the 1st is 1 and 2nd is 0, and both outcomes are 1, respectively. Testing whether overall probability of the 1st and 2nd paired outcomes to be 1 are the same reduces to H_o: $p_{10} = p_{01}$ and is done comparing n_{10} to n_{01} (with concordant outcomes n_{00} and n_{11} ignored). As with general sign tests, conditional on m (which now equals $n_{10} + n_{01}$), for H_o true, $n_{10} \sim B(m, 0.5)$ while under H_a, $n_{10} \sim B(m, p)$ for some $p > 0.5$. Thus H_o is tested with exact tests for the binomial or large sample approximations (c.f.; Miettinen 1968, Schlesselman 1982, Connett et. al 1987, Dupont 1988, Suissa & Shuser 1991, Sahai & Khurshid 1996).

Table 1a.- Response Probabilities in McNemar's Matched Binary Study

Table 1b - Probabilities for General Sign Test Reformulation of Matched Binary



As Table 1 illustrates, McNemar's matched binary response model reformulates to a general sign test (c.f. Cochran 1950). For matched binary pairs; "neutral results" (X=0) occur when both outcomes are 1 or both are 0; the 1st outcome is larger (X=1) when it is 1 while the 2nd outcome is 0; and the 2nd outcome is larger (X=1) when it is 1 and the 1st outcome is 0. Thus p_{-1} becomes p_{01}, p_{+1} becomes p_{01}, and combined together $[p_{00} + p_{11}]$ become p_0. Test statistics developed for McNemar's matched pairs and their distributional properties (c.f.; Miettinen 1968, Schlesselman 1982, Connett et. al 1987, Dupont 1988, Suissa & Shuser 1991, Sahai & Khurshid 1996), depend only on sums of probabilities for pair members to be tied, [p_{00} + p_{11}] rather than individually on p_{00} and p_{11} and thus apply to general sign tests in reformulations that sum p_{00} and p_{11}.

Current Statistical Approaches for McNemar's Matched Binary Responses

The range of statistical tests currently used for matched binary responses (all of which extend to general sign tests) was recently reviewed by Sahai and Khurshid (1996). We apply here to the general sign test some commonly used and recommended matched binary response methods. These methods can be categorized by two criteria; 1) "Conditional" as opposed to "Unconditional" and 2) "Exact" as opposed to "Asymptotic". Conditional matched binary approaches condition on the sum of non-equal observations $m = (n_{-1} + n_{+1})$ in the determination of the rejection region to ensure that the rejection probability does not exceed α for any observed value of m. The so called "unconditional" matched binary response approaches do not develop separate rejection regions for each m but rather consider m as a random variable and design an overall rejection region for which the probability (averaged over the distribution of m) of false rejection does not exceed α. In terms of the second classification metric, exact approaches give exact α and β for specified H_o and H_a while asymptotic approaches estimate these from large sample approximations.

Exact Conditional Approach Applied to the General Sign Test

For m non-neutral observations, under H_o, n_{+1} is distributed B(m, 0.5) while under a specified H_a, $n_{+1} \sim$ B$(m,\ 0.5(1+\Delta/w))$. For any fixed m, the exact conditional test of size α (McNemar 1947, Cochran 1950) in general sign test formulation, rejects H_o if $n_{+1} \geq B_{1-\alpha,m}$ where $B_{1-\alpha,m}$ is the smallest integer that exceeds the (1-α) quantile of the B(m, 0.5) cumulative distribution. If m is so small that no integer exceeds this threshold then $B_{1-\alpha,m}$ is undefined and H_o cannot be rejected for that m. Due to discreteness of the binomial, actual probabilities to reject H_o for given m may be < α. The overall probability H_o is rejected by the conditional exact test is the sum over m of {probability to observe m non-neutral observations} x {probability to reject H_o for that m when the specified H_a is True and H_o is tested by an exact test at a given α}; or

$$\beta = \sum_{m=0}^{N}\left[\sum_{n_{+1}=B_{1-\alpha,m}}^{m}\left[\frac{N!}{(N-m)!n_{+1}!(m-n_{+1})!}(1-w)^{N-m}\left((w+\Delta)/2\right)^{n_{+1}}\left((w-\Delta)/2\right)^{m-n_{+1}}\right]\right] \quad (1)$$

where the value of the second summand in (1) is zero if $B_{1-\alpha,m}$ is undefined as described earlier in this paragraph.

Several current software packages (c.f.; Elashoff 2000, Hintze 2000, Dupont & Plummer 2001, O'Brien 2002, Borenstein et. al. 2003, Oloffson 2003) use (1) formulated to the matched binary response problem to either; find β for a given *N*, or find the minimal *N* for which (1) does not exceed a predetermined β. Section 3 of this paper demonstrates how to extend the matched binary power / sample size estimation algorithm in PASS (Hintze 2000) to estimate power & sample size for the general sign test.

Asymptotic Conditional Approach Reformulated to the General sign Test

In the past, computational capacity to implement (1) for McNemar's test did not exist and many persons still do not have software that implements (1). Thus, large sample approximations to the conditional exact test that can be easily implemented with calculators were developed to either test H_o and/or to estimate the required sample size for settings where large sample approximations (or even exact tests) were used to test H_o. McNemar (1947) noted that under H_o, the test statistic

$$Z = (n_{+1} - n_{-1}) / \sqrt{m} \quad (2)$$

has an asymptotic normal distribution and that chi-square tests based on (2) could test H_o. Several authors have developed different large sample approximations to the distribution of (2) and corresponding sample size and power estimates (Miettinen 1968, Schlesselman 1982, Sahai & Khurshid 1996). We use an approximation given by Miettinen (1968) which Lachin (1992) found to be more accurate than three other well known approximations for matched binary responses. More details on Mietinnen's and other asymptotic power / sample size approaches for matched binary responses can be found elsewhere (Miettinen 1968, Schlesselman 1982, Lachin 1992, Sahai & Khurshid 1996).

Briefly, with the number of non-neutral pairs *m* fixed, the value of *m* needed to have given α and β with (2) is

$$m = \left[\left[Z_{1-\alpha}(w) + Z_{1-\beta}\sqrt{(w+\Delta)(w-\Delta)}\right]/\Delta\right]^2 \qquad (3),$$

where Z_x is the standard normal variate with the cumulative distribution function *x*. However, *m* is random $\mathrm{B}(N, w)$ so what is needed is the minimal *N* which generates a distribution of *m* whose weighted power is at least1-β. A first-order conditional power estimate of $N = \left[\left[Z_{1-\alpha}(w) + Z_{1-\beta}\sqrt{(w+\Delta)(w-\Delta)}\right]/\Delta\right]^2 / w$ obtains from the fact that the expected number of discordant pairs *m* for this *N* is the right hand side of (3). This was improved to a second-order unconditional power estimate (Miettinen 1968) that also adjusts for the variability of *m* (for a given *N*) and thus the increased variability of (2) through this randomness or;

$$N = \left[Z_{1-\alpha}\sqrt{w} + Z_{1-\beta}\sqrt{w - \{(\Delta)^2(3+w)\}/(4w)}\right]^2 / [\Delta]^2 \qquad (4).$$

Reformulating this approximation with *N* fixed gives:

$$1-\beta = \Phi\left[\left[|\Delta|\sqrt{N} - Z_{1-\alpha}\sqrt{w}\right] / \left[\sqrt{w - \{(\Delta)^2(3+w)\}/(4w)}\right]\right] \qquad (5),$$

where $\Phi[y]$ is the standard normal cumulative distribution of *y*.

Exact Unconditional Approach Applied to General sign Test

The exact conditional approach (1) rejection region is often quite conservative with the true value of the Type 1 error being far less than the nominal α, which also reduces power when H_o is false. To minimize such loss of power through conservative rejection regions, Suissa & Shuster (1991) developed an "unconditional approach" for McNemar's matched binary responses which obtains a less conservative rejection region from a metric that averages conditional rejection probabilities for each *m* weighted by the probabilities to have *m* non-neutral pairs. For some *m*, the value of the conditional Type 1 error can exceed the overall unconditional nominal α.

The exact unconditional approach is computationally intensive without closed form formulas, so the reader is referred to Suissa & Shuster (1991) for more details.

Briefly, these authors chose (2) as the test statistic for the rejection region (but other test statistics are possible) and used an iterative approach to find the rejection threshold Z_C (i.e. reject H_o if Z from (2) > Z_C) which gave a desired overall α. If C is the set of all possible m and n_{+1} for which (2) > Z_C, then for given N, w and Z_C the Type I error is:

$$\alpha^* = \sum_{m,n_{+1}\in C}\sum\left[\frac{N!}{(N-m)!n_{+1}!(m-n_{+1})!}(w/2)^m(1-w)^{N-m}\right] \qquad (6).$$

The value of w that gives α^{Sup} the maximum possible α^* in (6) for a given N and Z_C, was identified by a numerical approach . But due to problems of numerical stability, Suissa & Shuster (1991) considered only values of w<0.995. However, the restriction w w < 0.995 should not greatly influence application of this method to the general sign test. The likely N needed for the general sign test would be < 200; so if $w \geq 0.995$, then neutral values would only rarely occur. Furthermore, if one believes $w \geq 0.995$, the chance of a neutral observation is so low than existing approaches for the restricted sign test with p_o=0 (c.f.; Dixon & Massey 1968, Noether 1987) might be more appropriate.

From identification of the w which obtains α^{Sup} in (6) for given N and Z_C, it is certain that (for w <0.995) the real α cannot exceed α^{Sup} for that N and Z_C. With N fixed and a specified α (say α=0.025) the value of Z_C (to the second decimal place) was then systematically searched to find the minimum Z_C for which α^{Sup} did not exceed that specified α. This value of Z_C defines the rejection region that for the given N maintains the specified α error. Suissa & Shuster (1991) present these values of Z_C for different N and α = (0.05, 0.025 and 0.01) in Table 2. For example, with N=30 if $\alpha \leq 0.025$ was desired, the minimum Z_C that generated $\alpha \leq 0.025$ was 2.05 (i.e., reject H_o if $(n_{+1}-n_{-1})/\sqrt{m}$ > 2.05).

The power of this Exact Unconditional test for a given N, $w = 1 - p_o$, α, Δ is

$$1-\beta = \sum_{m,n_{+1}\in C}\sum\left[\frac{N!}{(N-m)!n_{+1}!(m-n_{+1})!}(1-w)^{N-m}\left((w+\Delta)/2\right)^{n_{+1}}\left((w-\Delta)/2\right)^{m-n_{+1}}\right] \qquad (7),$$

where again C is the rejection region defined by α, N and Z_C. To obtain the minimal N needed for a predetermined power (1-β), α, w and Δ, (6) was used to find the Minimal Z_C to maintain the specified α for the given N. This Z_C and N were then incorporated with specified w and Δ into (7) to determine if power exceeded (1-β), with N systematically increased until the smallest N with power that exceeded (1-β) for the specified H_a while maintaining the specified α for all possible H_o with w <0.995 was found. These values of N were presented in Tables 3 & 4 of Suissa & Shuster (1991) for α=0.05, 0.025 and 0.001; β=0.2 and 0.1; Δ ranging from 0.10 to 0.60; and w being in a corresponding constrained range.

3. AN APPLICATION

We now apply the three approaches described in Section 2 to a general sign test problem. As described in the Introduction, a study was planned of couples who had used illegal drugs to see if a wife is more likely to underestimate than overestimate her husband's lifetime number of injection drug partners compared to what the husband reports; H_o: $p_{+1} = p_{-1}$, the wife is not more likely to overestimate than underestimate versus H_a: $p_{+1} > p_{-1}$ the husband reports more than the wife estimates, was to be tested with a one sided α=0.025. From pilot data, it is believed the probability that wife and husband will agree is p_0=0.30 or w = 1- p_0=0.70. The investigators wanted to be able to detect a minimal difference of Δ=0.30, which for w =0.70 means $p_{+1} = 0.50$ and $p_{-1} = 0.20$.

Sample Size Estimation for Fixed Power

We first apply the approaches described in Section 2 to estimate minimal sample size needed to achieve a Type 2 error of β=0.20 (80% power). If either a one sample chi-square test or an exact test will evaluate H_o, then a large sample approximation for the conditional test may be appropriate for sample size estimation. Using Miettinen's (1968) approach (4) with α=0.025, β=0.20, Δ=0.30 and $w = 0.70$ estimates that N=57.8 is needed which rounds up to 58 couples.

If exact tests will be used, then an exact conditional sample size estimate is more appropriate and, as described in Section 2, can be obtained from software designed for matched binary responses. For example, with the "McNemar Test" option in the "Proportions" menu of PASS (Hintze 2002), the following parameters will need to be input for our design:

1. In the "*Find (solve for)* frame, select *"N*".
2. In the "*Alternative Hypothesis"* frame, select *"One-Sided"*.
3. In the *"Alpha (General significance Level)* frame, enter "0.025".
4. In the *"Beta (1- Power)* frame, enter "0.20".
5. In the *"Which Parameters* frame, chose *"P12 & p21*" which corresponds to p_{-1} and p_{+1} for our general sign test notation
6. In the *"P21"* frame, enter *"0.50"* (or the value of p_{+1})
7. In the *"P12"* frame, enter *"0.20"* (or the value of p_{-1})
8. All other frames on the PASS screen can be left blank
9. Click on the arrow in the upper left hand side of the PASS screen which performs the estimation

The answer returned is that *N*=64 couples should be interviewed.

If the exact unconditional approach (modified to the general sign test) will test H_o, one needs to use the Tables in Suissa & Shuster (1991) to: First find the minimal *N* that obtains a given β, and next find the rejection value Z_C which attains the specified α error for that N. To determine *N* look at Table 3 with Δ (also Δ in our general sign test notation) = 0.30, ψ (*w* in our notation) = 0.70, and α=0.025. The column for N^* gives the minimal *N* for the exact unconditional approach which is 60 couples. Next, to apply the exact conditional approach once the data has been collected, H_o will be rejected if $(n_{+1} - n_{-1}) / \sqrt{m}$ > 1.99 where 1.99 is obtained from looking at the z^* column under *N*=60 and α=0.025 in Table 2 of Suissa & Shuster (1991).

Power Estimation for Fixed Sample Size

Now assume that for the same study, only *N*=50 couples are available and we wish to estimate power. If either a one sample chi-square test or an exact test will evaluate H_o, then the asymptotic conditional approach with Miettinen's (1968) large sample approximation (5) may provide an appropriate power estimate, which with α=0.025, *N*=50, Δ=0.30 and $w = 0.70$ is that $\beta = 0.26$ or power is 74%.

If the exact test will evaluate H_o, then an exact conditional power estimate is more appropriate and can be implemented by software designed for matched binary responses. For example, with the "McNemar Test" option in the "Proportions" menu of PASS software (Hintze 2000), these parameters will need to be input for our design:

1. In the "*Find (solve for)* frame, select "*Beta and Power*".
2. In the "*Alternative Hypothesis"* frame, select *"One-Sided"*.
3. In the *"Alpha (General significance Level)* frame, enter "0.025".
4. In the *N (Number of Pairs)* frame, enter "50".
5. In the *"Which Parameters* frame, chose *"P12 & P21*" which corresponds to p_{-1} and p_{+1} for our general sign test notation
6. In the *"P21"* frame, enter *"*0.50*"* (or the value of p_{+1})
7. In the *"P12"* frame, enter *"*0.20*"* (or the value of p_{-1})
8. All other frames can be left blank
9. Click on the arrow in the upper left hand side of the PASS screen which performs the estimation

The answer returned is the exact β error 0.32 or power is 68%.

A systematic way to calculate power for fixed sample size using the exact unconditional approach has not been presented for McNemar's matched binary responses. But if this is done, it would readily extend to the general sign test as did estimation of sample size needed for power with the exact unconditional approach.

4. DISCUSSION

While this is not well known, approaches to estimate power and sample size for McNemar's matched binary responses extend to general sign tests when neutral values (p_o >0) are possible. Current power and sample size estimation software do not have procedures dedicated to general sign tests that permit neutral observations (i.e., $p_0 > 0$). Ironically, several software packages (c.f.; O'Brien 2002, Borenstein et. al. 2003, Oloffson 2003) with options to estimate power / sample size for restricted sign tests that do not allow $p_0 > 0$ also have separate McNemar's options that can be modified as described in this paper to estimate power and sample size for general sign tests when $p_0 > 0$ (as well as when $p_0 = 0$).

We presented a notational correspondence between the general sign test and matched binary response problem to enable current formulas, software and tables developed for McNemar's matched binary responses to be used for the general sign test and illustrated this with a specific example. If needed, other analytic approaches to matched binary responses also have extensions to the general sign test. For example, since p_{10} / p_{01} is the matched odds ratio for paired binomial responses (Schlesselman 1982), by the reformulation in Table 3b, procedures to obtain confidence intervals for the matched odds ratio (Schlesselman 1982) will obtain confidence intervals for (p_{+1} / p_{-1}) the risk ratio of being positive to being negative for the general sign test design.

An important question when planning a general sign test study is which (if any) of the three approaches described in Section 2 is better. But this same question has not been fully resolved for McNemar's discordant binary responses (Sahai & Khurshid 1996), and the issues with the sign test are similar. The exact unconditional approach generally yields smaller required sample sizes and/or larger power than do conditional approaches, but may not be practical to use with currently available software. The asymptotic conditional approach, however, is easily implemented with a calculator and the exact conditional approach is implemented by several software packages. Since the asymptotic conditional approach can be anti-conservative (i.e. the true Type I error exceeds the nominal α) while the exact conditional is not, many people prefer to use an exact (rather than an asymptotic) conditional test of H_o once the data has been collected. Therefore, arguments can be made that the exact conditional approach should also be used to estimate power and sample size.

Since w (or 1-p_o) and Δ will be unknown, a sensitivity approach estimating power or sample size over a range of plausible values for w (or 1-p_o) and Δ should be taken, guided by the best available preliminary data and analysis for minimal meaningful differences. As power increases (or sample size decreases) with Δ, smaller values of Δ are conservative. For a given Δ, power decreases (or sample size increases) with *w* making larger values of *w* conservative up to the setting with no ties $w = 1$ or $p_0 = 0$. While conservative power / sample size estimates are usually desirable, when taken to the extreme, conservativeness may prevent undertaking of studies that could find important results.

In conclusion, due to the mathematical equivalency of the two tests described in this paper, as new power / sample size estimation approaches for matched binary responses are developed and implemented, it may be advantageous to make these procedures accessible to general sign tests. For example, some matched binary power / sample size estimation software require the correlation between responses of matched subjects to be positive (i.e. $p_{00} \times p_{11} - p_{01} \times p_{10} > 0$). This restriction makes sense for matched binary responses as matching is inefficient for non-positive correlations, but prevents use of those software for general sign test problems where analogous restrictions do not apply. While *NQuery* (Elashoff 2000) does not directly offer to estimate power & sample size for sign tests, the "exact conditional" estimator tool for matched binary outcomes in this software is referred to as an exact sign test for paired responses. But the formulation and description of this option restricts it to two group matched pairs comparisons as opposed to a single group sign test; for example basing estimates on π_1 the portion of responses in Group 1 and π_2 the portion of responses in Group 2. However, by the reformulation presented here in Table 3, this *NQuery* procedure can be applied to the general sign test by substituting $(p_{+1} - p_{-1})$ for $(\pi_2 - \pi_1)$ and *w* for $(\pi_{01} + \pi_{10})$.

REFERENCES

Borenstein M, Rothstein H, Cohen J. (2003), *SamplePower Version 2.0 SPSS,* Distributed by SPSS Incorprated.

Coakley CW & Heise MA (1996). Versions of the sign test in the presence of ties. *Biometrics* **52**, 1242-51.

Cochran WG (1937), "The Efficiencies of the Binomial Series test of Significance of a Mean and a Correlation Coefficient" *J. R. Statist. Soc.* 100 69-73

Cochran WG (1950), "The Comparison of Percentages in Matched Samples," *Biometrics,* 37, 256-66.

Connett JE, Smith JA, McHugh RB. (1987), "Sample Size and Power for Pair Matched Case-Control Studies," *Statistics in Medicine*, 6(1), 53-9.

Dixon WG , Mood AM (1946), The Statistical Sign Test," *J Amer Statist. Ass.* 41 557-66.

Dixon WJ, Massey FM. (1968), *Introduction to Statistical Analysis, 3rd Edition*, McGraw Hill. New York, NY

Dupont WD. (1988), "Power Calculations for Matched Case-Control Studies," *Biometrics*, 44, 1157-68.

Dupont WD and Plummer WD. (2001), *PS Power and Sample Size Calculations,* Downloaded December 2001 at http://www.mc.vanderbilt.edu/prevmed/ps.htm.

Elashoff JD. (2000) *nQuery Advisor Release 4.0*, Cork, Ireland: Statistical Solutions

Hintze J. (2000), *PASS-2000* Kaysville, UT: Number Cruncher Statistical Systems.

Hollander M, Wolfe DA. (1973), *Nonparametric Statistical Methods*, Wiley & Sons New York.

Lachin JM. (1992), "Power and Sample Size Evaluation for the Mcnemar Test with Application to Matched Case-Control Studies," *Stat Med,* 11(9), 1239-51.

McNemar Q. (1947), "Note on the Sampling Error of Differences Between Correlated Proportions And Percentages," *Psychometrica,* 12, 153-7.

Miettinen OS. (1968), "The Matched Pairs Design in the Case of all or None Responses," *Biometrics*, 24, 339-52.

Noether E. (1987), Sample Size Determination for Some Common Nonparametric Tests," *J Amer Stat Assn*, 82(398), 645-47.

O'Brien R. (2002), *UnifyPow Version 2002.08.17.a SAS Module for Sample-Size Analysis,* Downloaded September 2003 at http://www.bio.ri.ccf.org/power.html

Oloffson B. (2003), *StudySize Version 1.08,* CreoStat HR.

Rayner J. (1999). Modelling ties in the Sign Test. *Biometrics* 55, 663-665.

Sahai H, Khurshid A. (1996), "Formulas and Tables for Determination of Sample Sizes and Power in Clinical Trials for Testing Differences in Proportions for the Matched Pair Design: A Review," *Fundam Clin Pharmacol*, 19(6), 554-63.

Schlesselman JJ. (1982), *Case-Control studies design, conduct and analysis.* Oxford University Press. Oxford UK.

Suissa S, Schuster JJ. (1991), "The 2x2 Matched Pairs Trial: Exact Unconditional Design and Analysis," *Biometrics*, 47, 361-72.