Published in final edited form as:

Contemp Clin Trials. 2011 January ; 32(1): 147–150. doi:10.1016/j.cct.2010.10.007.

The Power of the Sign Test Given Uncertainty in the Proportion of Tied Observations

Paul J. Nietert, PhD and Mary J. Dooley, BS

Division of Biostatistics and Epidemiology, Department of Medicine, Medical University of South Carolina

Abstract

Background—The sign test is a well-known nonparametric approach for testing whether one of two conditions is preferable to another. In medicine, this method may be used when one is interested in testing in the context of a clinical trial whether either of two treatments that are provided to study subjects is favored over the other. When neither treatment outperforms the other within a given individual, a “tie” is said to have occurred. When planning such a trial and estimating statistical power and/or sample size, one should consider the probability of a tie occurring (P_T). This paper quantifies the degree to which uncertainty in P_T affects a study’s statistical power.

Methods—Binomial theory was used to calculate power given varying levels of uncertainty and varying distributional forms (i.e. beta, uniform) for P_T .

Results—Across a range of prior distributions for P_T , power was reduced (i.e. <80%) for 46 (71.9%) of 64 experimental conditions, with large reductions (i.e. power <70%) for 10 (15.6%) of them.

Conclusions—When designing a clinical trial that will incorporate the sign test to compare 2 conditions, ignoring potential variation in the probability of a tie occurring will tend to result in an underpowered study. These findings have implications to the design of any clinical trial for which assumptions are made in calculating an appropriate sample size.

Keywords

statistical power; sign test; non-parametric statistics; uncertainty; binomial distribution; sample size estimation

Introduction

The sign test is a non-parametric technique for testing whether one condition is preferable to another in the context of paired responses [1]. In some study designs, there exists the possibility that the 2 conditions being compared could be equally preferable (a “tie”). While modifications have been developed to incorporate ties into the test statistic [2-5], the

© 2010 Elsevier Inc. All rights reserved.

Please address all correspondence to: Paul J Nietert, PhD, Division of Biostatistics and Epidemiology, Department of Medicine, Medical University of South Carolina, 135 Cannon Street, Suite 303, PO Box 250835, Charleston, SC 29425, nieterpj@musc.edu, P: 843-876-1204, F: 843-876-1126.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

conventional application of the sign test ignores tied responses. Thus in planning a study that ultimately relies on the traditional sign test for hypothesis testing, one must consider how often ties are likely to occur. Although the sign test may not be as commonly used as other traditional statistical tests for 2-group comparisons, it provides an excellent opportunity to demonstrate why study designs need to consider uncertainty in the assumptions underlying sample size and statistical power estimation.

For a given experiment comparing conditions A and B, the sign test involves testing the null hypothesis of $H_0: P_{A:B} = 0.5$ versus $H_1: P_{A:B} \neq 0.5$, where $P_{A:B}$ represents the probability that condition A is preferred over condition B. Using binomial theory, the probability of rejecting H_0 given $P_{A:B}$, which will be abbreviated $\Pi(P_{A:B})$, can be expressed as:

$$\Pi(P_{A:B}|N_{NT}) = \sum_{k=L}^{N_{NT}} \binom{N_{NT}}{k} \cdot P_{A:B}^k \cdot (1 - P_{A:B})^{N_{NT}-k} + \sum_{k=0}^{N_{NT}-L} \binom{N_{NT}}{k} \cdot P_{A:B}^k \cdot (1 - P_{A:B})^{N_{NT}-k} \quad (1)$$

where N_{NT} is the total number of non-tied observations, and L is selected to ensure that α , $\Pi(0.5)$, is as high as possible but less than 0.05. Power is $\Pi(P_{A:B})$, where $P_{A:B}$ is specified under H_1 .

Prior to conducting a study using the sign test to evaluate whether one condition (treatment) outperforms another, a statistical power assessment is typically performed. Certain assumptions are made, including the type I error rate (i.e. alpha level), whether the hypothesis test is 1 or 2-sided, and the magnitude of the difference in the effect of each of the conditions of interest. One should also anticipate the proportion of observations for which neither condition is superior (ties). Different approaches may be taken for handling the *a priori* predictions in the power estimation. One approach is to simply make one or more estimates of what the values are likely to be and calculate the power accordingly. Another approach might be to find the sample size which would provide the required number of non-tied observations some percentage (e.g. 95%) of the time. We demonstrate here an alternative method, incorporating uncertainty in the proportion of tied observations into the statistical power calculation, and show how the magnitude of this uncertainty affects power.

Methods

For a study with a given overall sample size N , the number of potential non-tied observation (N_{NT}) ranges from 0 to N , meaning that power depends on N_{NT} . To calculate unconditional (marginal) power that can be expected for such a study, first use equation (1) for each possible number of non-tied observations (N_{NTi}) (i.e. $i=0, 1, \dots, N$) to determine the cutoff values (L_i). These calculations will use $P_{A:B}=0.5$ (assuming no preference for A or B), and the L_i values are selected in order to obtain alpha levels that are as high as possible but < 0.05 . Next, for a given “effect size” (δ) of interest, use equation [1] to calculate power conditioned on each possible value of N_{NT} . By “effect size”, we mean the probability that condition A would be preferred in favor of condition B. After selecting a distributional form for P_T (described below), allowing for variation in P_T to be introduced, one then assigns probability weights to each of the potential values for N_{NT} , corresponding to potential values for P_T . Finally, summing the products of these probability weights and the conditional power calculations results in obtaining the unconditional (marginal) power.

For example, the power for a study involving 50 subjects for an effect size of 0.722 ranges from 0% to 89% depending on the number of non-tied observations (see Figure 1). If certain values of N_{NT} are more likely than others, as is the case with the beta($\mu=0.2, \sigma=0.1$) prior as

shown in the figure, the marginal power, found by summing the products of the conditional power estimates and their probability weights, will lie between 0% and 89%. In this example the marginal power is 77.83%.

We used this algorithm to calculate power across 4 overall sample sizes (i.e. N=50, 100, 200, 400), 4 different mean P_T values (20%, 40%, 60%, 80%), and 5 different levels of variation for P_T (1 with 0 variance [negative control], 3 with a beta prior, and 1 with a uniform prior). The calculations were conducted using SAS v9.2 (Cary, NC).

In designing a study, one may have prior knowledge about the likelihood of ties occurring. Thus for our power calculations we assumed 2 different prior distributional forms for P_T , the beta and uniform distributions. When P_T was assumed to have a beta prior distribution, pre-specified means (μ 's) included 20%, 40%, 60%, and 80%, and pre-specified standard deviations (σ 's) included 1%, 5%, and 10%. The beta distribution, ideal for modeling proportions since it ranges from 0 to 1, is parameterized using 2 positive shape parameters (α and β). Through algebraic manipulations, the shape parameters can be expressed in terms of the mean and standard deviation using the formulae:

$$\alpha = \frac{(\mu^2 - \mu^3 - \mu\sigma^2)}{\sigma^2} \quad \text{and} \quad \beta = \frac{(\mu - 2\mu^2 + \mu^3 - \sigma^2 + \mu\sigma^2)}{\sigma^2}$$
 This reparameterization allows the prior distribution of P_T to be described in a manner with which non-statisticians would typically be more familiar. When P_T was assumed to have a uniform prior distribution with a given mean, we assumed that the distribution boundaries were equidistant from the mean, as far apart as possible, but within the interval (0,1). For example, if 40% of observations were expected to be tied, we assumed that came from a uniform distribution on the interval (0.0, 0.8).

Our analysis focused on the impact of variability in P_T (not on P_T itself) on power, within a variety of experimental circumstances. Since some investigators may alternatively wish to choose an overall sample size N' to ensure (with 95% confidence) that at least a given number of non-ties occur, we also investigated the effect of using such an approach on power. In doing so, we were able compare the (in)efficiency of this technique relative to our approach.

Results

The findings from our primary analyses are listed in Table 1. When P_T was expected to be low (i.e. 0.2), there was minimal impact on power, regardless of the distributional form of P_T . However, across all values for P_T , power was reduced (i.e. <80%) for 46 (71.9%) of the 64 examples, with large reductions in power (i.e. <70%) for 11 (17.2%) of them. When the beta prior was used, increasing variation in P_T resulted in declining power, although this relationship was more pronounced as P_T became greater. When a uniform prior for P_T was selected, power was reduced 87.5% of the time, with large reductions in power 50% of the time. As the variability in P_T increases, the greatest reductions in power occur when the expected number of non-tied observations is lowest. Figure 2 illustrates the effect of variation in P_T on power for a hypothetical study with an overall sample size of N=50.

Table 2 lists additional power calculations assuming an increased overall sample size N' required to ensure that N non-tied observations would occur 95% of the time. This table illustrates potential inefficiencies in this strategy, as the power estimates are >80% for 50 (78.1%) of the calculations. Interestingly, if a uniform prior is selected, the increased sample size (N') was not sufficient to provide 80% power unless the probability of ties was expected to be low (0.2 or 0.4).

Discussion

When designing experiments that incorporate the sign test, ignoring potential variation in the probability of ties occurring will tend to result in underpowered studies. This impact may be quite large if the overall sample size is 50 or fewer, if the probability of a tie occurring is expected to be 0.6 or greater, or if the prior distribution for this probability is uniform in nature.

While our results demonstrate that power was typically reduced several percentage points, it is important to note that it was not always reduced. This is largely due to the fact that the sign test's power is not uniformly increasing with larger sample sizes (see Figure 1). Chernick and Liu (2002) discuss the phenomenon of this “saw-tooth” behavior of power versus sample size for exact methods involving binomial proportions such as that used with the sign test [6].

In today's clinical trials, sign tests are not typically the analytical methods of choice. There are more sophisticated modeling strategies to account for covariates, correlated outcomes, or missing data. However, one can easily see how not taking into account the uncertainty in any *a priori* assumptions associated with a planned analysis could have similar impacts on statistical power. For example, Ambrosius et al. highlight the importance of the consideration of uncertainty in allele and genotype frequencies when designing genetic association studies [7]. While the impact of the type of variability we describe may be modest, accounting for variability in assumed parameter values may have the potential to shift the paradigm of sample size estimation.

Acknowledgments

Grant Support

The project described was supported by Award Number UL1RR029882 from the National Center for Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health.

References

1. Dixon WJ, Mood AM. The statistical sign test. *Journal of the American Statistical Association* 1946;41:557–566. [PubMed: 20279351]
2. Randles RH. On neutral responses (zeroes) in the sign test and ties in the Wilcoxon-Mann-Whitney test. *The American Statistician* 2001;55:96–101.
3. Fong DYT, Kwan CW, Lam KF, et al. Use of the sign test for the median in the presence of ties. *The American Statistician* 2003;57:237–240.
4. Coakley CW, Heise MA. Versions of the sign test in the presence of ties. *Biometrics* 1996;52:1242–1251.
5. Rayner JCW, Best DJ. Modeling ties in the sign test. *Biometrics* 1999;55:663–665. [PubMed: 11318231]
6. Chernick MR, Liu CY. The saw-toothed behavior of power versus sample size and software solutions: single binomial proportion using exact methods. *The American Statistician* 2002;56:149–155.
7. Ambrosius WT, Lange EM, Langefeld CD. Power for genetic association studies with random allele frequencies and genotype distributions. *American journal of human genetics* 2004;74:683–693. [PubMed: 15024689]

Figure 1.

Left axis: Power (solid “saw-tooth” line) as a function of the number of non-tied observations, assuming an effect size of 0.722, which yields 80.2% power when $N_{NT}=40$. Right axis: probability (dotted line) of N_{NT} occurring, assuming a beta prior for P_T with mean 0.2 and standard deviation 0.1.

Figure 2.
Effect of variation in the probability of a tie (P_T) occurring on statistical power, assuming an overall $N=50$ subjects

Table 1

Effect of variation in the probability of a tie occurring (PT) on the statistical power of the sign test

Overall Sample Size (N)	Mean P _T level	Expected Number of Non-Tied Observations	Beta Prior: Standard Deviation in P _T					
			δ *	0.00 (Negative Control)				
				0.01	0.05	0.10	Uniform Prior	
Statistical Power								
50	0.2	40	0.722	80.17%	78.55%	78.35%	77.83%	77.50%
50	0.4	30	0.750	80.34%	78.39%	75.16%	74.52%	70.01%
50	0.6	20	0.799	80.09%	77.45%	74.33%	72.19%	62.01%
50	0.8	10	0.917	80.10%	80.08%	74.47%	63.56%	60.31%
100	0.2	80	0.664	80.52%	82.00%	81.70%	81.35%	81.09%
100	0.4	60	0.692	80.22%	82.40%	83.09%	82.23%	76.80%
100	0.6	40	0.722	80.17%	77.65%	77.85%	75.81%	66.00%
100	0.8	20	0.799	80.09%	75.04%	72.27%	65.56%	62.01%
200	0.2	160	0.611	80.36%	79.07%	78.43%	78.02%	77.84%
200	0.4	120	0.633	80.15%	81.39%	81.23%	80.44%	75.86%
200	0.6	80	0.664	80.52%	81.93%	81.37%	79.62%	69.38%
200	0.8	40	0.722	80.17%	78.20%	75.84%	69.69%	66.00%
400	0.2	320	0.582	81.00%	81.37%	81.33%	80.88%	80.72%
400	0.4	240	0.592	80.65%	79.83%	79.76%	79.01%	74.73%
400	0.6	160	0.611	80.36%	78.67%	78.04%	76.40%	67.56%
400	0.8	80	0.664	80.52%	81.73%	79.67%	73.41%	69.38%

* Effect size (probability that Condition A is preferable to Condition B) required to ensure 80% power, assuming no variation in P_T.

Table 2

Effect on statistical power of requiring N* observations to achieve N non-tied observations 95% of the time

Mean P _T level	Desired Number (N) of Non-Tied Observations	N* Required*	δ †	Beta Prior: Standard Deviation in P _T		
				0.01	0.05	0.10
				Statistical Power		
0.2	40	56	0.722	84.39%	83.51%	82.89%
						82.78%
0.4	30	60	0.750	84.88%	84.02%	82.93%
						77.40%
0.6	20	65	0.799	86.53%	85.52%	83.47%
						70.37%
0.8	10	75	0.917	94.96%	91.88%	81.16%
						73.04%
0.2	80	109	0.664	84.35%	84.72%	84.27%
						83.98%
0.4	60	114	0.692	88.19%	87.99%	86.99%
						81.44%
0.6	40	121	0.722	86.64%	85.96%	83.67%
						71.71%
0.8	20	135	0.799	87.30%	84.95%	77.19%
						71.43%
0.2	160	212	0.611	81.00%	80.85%	80.43%
						80.17%
0.4	120	220	0.633	85.20%	84.98%	84.09%
						79.17%
0.6	80	229	0.663	87.10%	86.08%	84.17%
						72.78%
0.8	40	249	0.722	87.70%	84.80%	77.55%
						72.53%
0.2	320	417	0.582	84.18%	83.87%	83.39%
						83.20%
0.4	240	427	0.592	82.55%	82.39%	81.58%
						76.99%
0.6	160	441	0.611	82.53%	82.00%	80.23%
						70.42%
0.8	80	468	0.664	87.95%	85.39%	78.64%
						73.65%

* N* represents the number of observations required to ensure that N non-tied observations would occur 95% of the time.

† Effect size (probability that Condition A is preferable to Condition B) required to ensure 80% power, assuming no variation in P_T.