

## Chapter 232

# Tests for Paired Specificities

## Introduction

This procedure gives power or required sample size for comparing two diagnostic tests when the outcome is specificity. In the design covered in this chapter, the two diagnostic tests are each performed on the same subject.

Specifically, a set of  $N$  subjects is selected, and each is given both of the diagnostic tests being considered. Subsequently, a gold standard test is also given and used to obtain the true presence or absence of the disease. The gold standard may be a more expensive test or require the sacrifice of the subject.

The measures of diagnostic accuracy are sensitivity and specificity. *Sensitivity* ( $Se$ ) is the probability that the diagnostic test is positive for the disease, given that the subject actually has the disease. *Specificity* ( $Sp$ ) is the probability that the diagnostic test is negative, given that the subject does not have the disease.

Mathematically,

$$\text{Sensitivity } (Se) = \Pr(+\text{Test}|\text{Disease})$$

$$\text{Specificity } (Sp) = \Pr(-\text{Test}|\text{No Disease})$$

Li and Fine (2004) present sample size methodology for testing sensitivity and specificity using a paired, prospective design. Their methodology is used here. Other useful references are Obuchowski and Zhou (2002), Machin, Campbell, Tan, and Tan (2009), and Zhou, Obuchowski, and McClish (2002).

## Technical Details

In this, two diagnostic tests are administered to each of  $N$  subjects. Of those,  $N_{ND}$  do not have the disease of interest. The  $N_D$  subjects with the disease are set aside, and only the  $N_{ND}$  subjects without the disease are used in the analysis. The results can be displayed in the following 2-by-2 table.

		<u>Test 2</u>		
<u>Test 1</u>	<u>Yes</u>	<u>No</u>		<u>Total</u>
<b>Yes</b>	$a$	$b$		$a + b$
<b>No</b>	$c$	$d$		$c + d$
<b>Total</b>	$a + c$	$b + d$		$N_{ND}$

Pairs with the same response from Test 1 and Test 2 (Yes-Yes and No-No) are called *concordant pairs*. Pairs with different responses (Yes-No and No-Yes) are called *discordant pairs*.

## Tests for Paired Specificities

The specificities for Test 1 and Test 2, respectively, are calculated as

$$Sp_1 = P(c + d) = \frac{c + d}{N_{ND}}$$

$$Sp_2 = P(b + d) = \frac{b + d}{N_{ND}}$$

---

## Specificity Hypotheses

The null and alternative hypotheses that the specificities of the two tests are equal are

$$H_0: Sp_1 = Sp_2 \quad \text{or} \quad H_0: Sp_1 - Sp_2 = 0$$

$$H_1: Sp_1 \neq Sp_2 \quad \text{or} \quad H_1: Sp_1 - Sp_2 \neq 0$$

or equivalently

$$H_0: P(c + d) = P(b + d) \quad \text{or} \quad H_0: P(c + d) - P(b + d) = 0$$

$$H_1: P(c + d) \neq P(b + d) \quad \text{or} \quad H_1: P(c + d) - P(b + d) \neq 0$$

These are equivalent to the hypotheses

$$H_0: P(b) = P(c) \quad \text{or} \quad H_0: P(b) - P(c) = 0$$

$$H_1: P(b) \neq P(c) \quad \text{or} \quad H_1: P(b) - P(c) \neq 0$$

The upper one-sided null and alternative hypotheses are

$$H_0: Sp_1 \leq Sp_2 \quad \text{or} \quad H_0: Sp_1 - Sp_2 \leq 0$$

$$H_1: Sp_1 > Sp_2 \quad \text{or} \quad H_1: Sp_1 - Sp_2 > 0$$

The lower one-sided null and alternative hypotheses are

$$H_0: Sp_1 \geq Sp_2 \quad \text{or} \quad H_0: Sp_1 - Sp_2 \geq 0$$

$$H_1: Sp_1 < Sp_2 \quad \text{or} \quad H_1: Sp_1 - Sp_2 < 0$$

---

## McNemar's Test

These hypotheses may all be tested using McNemar's paired-samples test statistic

$$\chi_1^2 = \frac{(b - c)^2}{b + c},$$

which follows a chi-square distribution with one degree of freedom.

## Power Analysis and Sample Size Calculations

The sample size problem reduces to a study of how many Yes-No's and No-Yes's are needed. Once this has been determined, the overall sample size is found using the proportion of discordant pairs to inflate the sample size appropriately.

Some power analysis programs follow an approximate procedure. Since the McNemar statistic follows the binomial probability distribution for a fixed number of discordant pairs, they use formulas that use the normal approximation to the binomial and then adjust the sample size depending on the proportion of discordant pairs,  $PD = P(b) + P(c)$ . This is called the *conditional* procedure.

One such approximate formula is given by Machin, Campbell, Fayers, and Pinol (1997):

$$N_D = \frac{\left( z_{1-\alpha/s}(\psi + 1) + z_{1-\beta}\sqrt{(\psi + 1)^2 - (\psi - 1)^2 PD} \right)^2}{(\psi - 1)^2 PD}$$

where  $s$  is the number of sides to the test (one or two),  $\psi = \frac{P(b)}{P(c)}$ ,  $PD = P(b) + P(c)$ , and  $\alpha$  and  $\beta$  are the usual type 1 and type 2 error rate probabilities. This sample size formula may be rearranged to compute the power.

However, Schork and Williams (1980) published a formula which provides the exact results for the *unconditional* case using multinomial enumeration of all possible outcomes. This formulation is also available in **PASS**:

$$Power = \sum_{x=r}^{N_{ND}} \sum_{y=0}^I \frac{N_{ND}!}{(N_{ND} - x)! y! (x - y)!} (1 - PD)^{N_{ND} - x} \left( \frac{PD + \delta}{2} \right)^y \left( \frac{PD - \delta}{2} \right)^{x-y}$$

where

$$PD = P(b) + P(c)$$

$$\delta = P(b) - P(c)$$

$N_{ND}$  is total of all four cells (the number of non-diseased subjects)

$r$  is the smallest integer for which  $\left(\frac{1}{2}\right)^r \leq \alpha$

$I$  is the largest integer such that  $\sum_{j=0}^I \binom{x}{j} \left(\frac{1}{2}\right)^x \leq \alpha$

## Tests for Paired Specificities

Finally, to obtain the estimate of the total sample size,  $N$ , we inflate  $N_{ND}$  by one minus the disease prevalence,  $1 - P$  (the proportion of non-diseased subjects in the population), using the equation

$$N = \frac{N_{ND}}{(1 - P)},$$

with fractional  $N$  values always being rounded up to the nearest whole number. This is called Method 0 in the paper by Li and Fine (2004). We can calculate  $N_{ND}$  from  $N$  by rearranging this formula as

$$N_{ND} = N \times (1 - P),$$

with fractional  $N_{ND}$  values always being rounded down to the nearest whole number.

### Discussion – Multinomial Enumeration vs. Normal Approximation

The multinomial enumeration (exact) algorithm works for  $N_{ND} < 2000$ . Above 2000, computing time goes up and the algorithm has numerical problems. **PASS** lets you select either the multinomial enumeration (exact) or the approximate solution. We have found that the approximate solution tends to result a sample size that is about 10% less than the exact solution.

Because of the lengthy computer time necessary to compute the exact answer when  $N_{ND} > 2000$ , we suggest that you use the approximate formula to begin with and then revert to the exact formula when you are ready for your final results.

# Example 1 – Finding the Power

Suppose that diagnosing a certain disease has used a certain diagnostic test which has a specificity of 75%. A new diagnostic test has been developed that is much less expensive and invasive. Researchers want to design a paired, prospective study in which each subject will be given both the old and new diagnostic tests. The data will be analyzed using a two-sided McNemar’s test with a significance level of 0.05. From previous studies, the percent discordant is estimated at 30%.

They want to consider changes in specificity of 5% and 10%. These changes translate to specificities of 78.75% and 82.50%. The prevalence of the disease in the population of interest is 20%. The power will be determined for trials with sample sizes between 300 and 2400 incremented by 300.

## Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For .....	<b>Power</b>
Power Calculation Method .....	<b>Normal Approximation</b>
Alternative Hypothesis .....	<b>Two-Sided</b>
Alpha.....	<b>0.05</b>
N (Number of Subjects) .....	<b>300 to 2400 by 300</b>
P (Disease Prevalence) .....	<b>0.2</b>
Sp1 (Specificity of Test 1).....	<b>0.75</b>
Sp2 (Specificity of Test 2).....	<b>0.7875 0.8250</b>
PD (Proportion Discordant).....	<b>0.3</b>

## Tests for Paired Specificities

## Output

Click the Calculate button to perform the calculations and generate the following output.

## Numeric Reports

## Numeric Results

Solve For: [Power](#)  
 Alternative Hypothesis: Two-Sided ( $H_1: Sp_1 - Sp_2 \neq 0$ )  
 Test Type: McNemar's Test

Power*	Number of Subjects		Disease Prevalence P	Specificity			Proportion Discordant PD	Alpha
	Total N	Non-Diseased NND		Test 1 Sp1	Test 2 Sp2	Difference Sp1 - Sp2		
0.18368	300	240	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.32238	600	480	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.45100	900	720	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.56424	1200	960	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.66009	1500	1200	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.73879	1800	1440	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.80186	2100	1680	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.85141	2400	1920	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.56470	300	240	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.85312	600	480	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.95824	900	720	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.98940	1200	960	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.99752	1500	1200	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.99946	1800	1440	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.99989	2100	1680	0.2	0.75	0.8250	-0.0750	0.3	0.05
0.99998	2400	1920	0.2	0.75	0.8250	-0.0750	0.3	0.05

\* Power was computed using the normal approximation method (conditional).

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.  
 N The total number of subjects in the study.  $N = NND / (1 - P)$ .  
 NND The number of non-diseased subjects in the study.  $NND = N \times (1 - P)$ .  
 P Disease Prevalence. The proportion of individuals in the population that have the disease (or condition).  
 Sp1 The specificity of Test 1.  
 Sp2 The specificity of Test 2.  
 Sp1 - Sp2 The specificity difference to detect.  
 PD Proportion Discordant. The proportion of subjects for whom the results of the two tests do not match.  
 Alpha The probability of rejecting a true null hypothesis.

## Summary Statements

A paired diagnostic test design (both diagnostic tests are performed on each subject) will be used to test the specificity of diagnostic test 1 against the specificity of diagnostic test 2. The comparison will be made using a two-sided, paired-sample McNemar Test with a Type I error rate ( $\alpha$ ) of 0.05. The prevalence of the disease (or condition) in the population is assumed to be 0.2, and the proportion of discordant pairs is assumed to be 0.3. To detect a specificity difference of -0.0375 (test 1 specificity: 0.75, test 2 specificity: 0.7875) with a total sample size of 300, the power is 0.18368.

## Tests for Paired Specificities

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	300	375	75
20%	600	750	150
20%	900	1125	225
20%	1200	1500	300
20%	1500	1875	375
20%	1800	2250	450
20%	2100	2625	525
20%	2400	3000	600

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula $N' = N / (1 - DR)$ , with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. $D = N' - N$ .

**Dropout Summary Statements**

Anticipating a 20% dropout rate, 375 subjects should be enrolled to obtain a final sample size of 300 subjects.

**References**

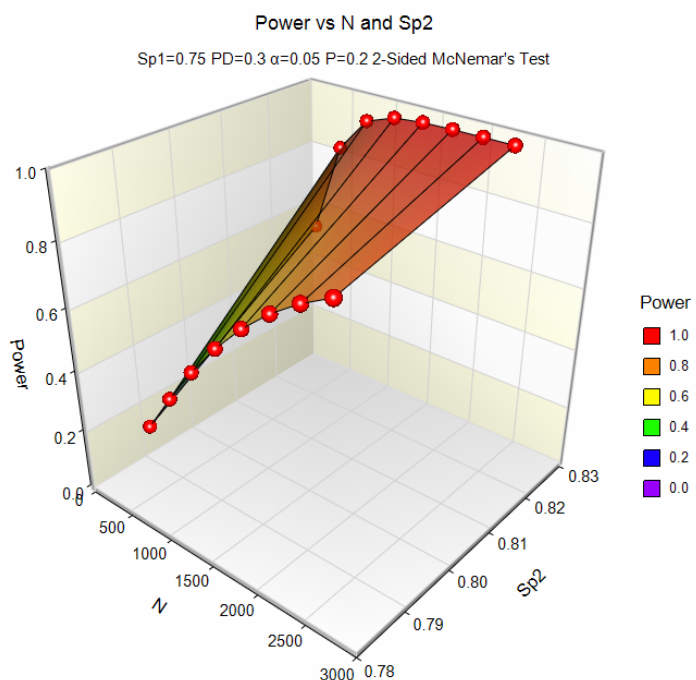
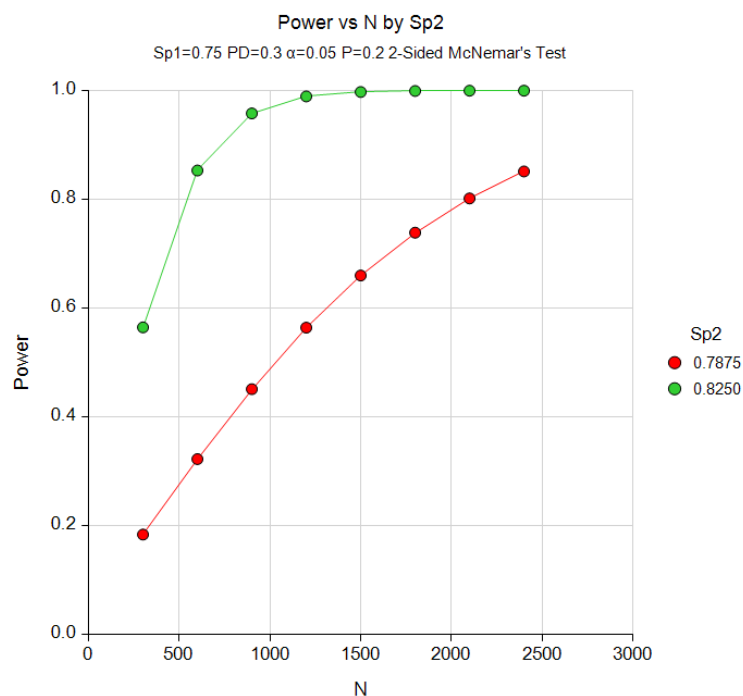
- Obuchowski, N.A., Zhou, X.H. 2002. 'Prospective studies of diagnostic test accuracy when disease prevalence is low,' Biostatistics, Volume 3, No. 4, pages 477-492.
- Li, J., Fine, J. 2004. 'On sample size for sensitivity and specificity in prospective diagnostic accuracy studies,' Statistics in Medicine, Volume 23, pages 2537-2550.
- Machin, D., Campbell, M.J., Tan, S.B., Tan, S.H. 2009. Sample Size Tables for Clinical Studies, Third Edition. Wiley-Blackwell, Chichester, United Kingdom.
- Zhou, X.H., Obuchowski, N.A., McClish, D.K. 2002. Statistical Methods in Diagnostic Medicine. Wiley-Interscience, New York.
- Schork, M. and Williams, G. 1980. 'Number of Observations Required for the Comparison of Two Correlated Proportions.' Communications in Statistics-Simula. Computa., B9(4), 349-357.

This report shows the values of each of the parameters, one scenario per row.

## Tests for Paired Specificities

## Plots Section

## Plots



These plots show the relationship between power, sample size, and Sp2 in this example.



## Example 2 – Finding the Sample Size

Continuing with Example 1, suppose you want to study the impact of two choices for Sp2 on sample size. Using a significance level of 0.05 and 90% power, find the sample size when Sp2 is 78.75% and 82.50%. Assume a two-tailed test is used.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For ..... **Sample Size**  
 Power Calculation Method ..... **Normal Approximation**  
 Alternative Hypothesis ..... **Two-Sided**  
 Power ..... **0.9**  
 Alpha ..... **0.05**  
 P (Disease Prevalence) ..... **0.2**  
 Sp1 (Specificity of Test 1) ..... **0.75**  
 Sp2 (Specificity of Test 2) ..... **0.7875 0.8250**  
 PD (Proportion Discordant) ..... **0.3**

### Output

Click the Calculate button to perform the calculations and generate the following output.

#### Numeric Results

Solve For: [Sample Size](#)  
 Alternative Hypothesis: Two-Sided ( $H_1: Sp_1 - Sp_2 \neq 0$ )  
 Test Type: McNemar's Test

Power*	Number of Subjects		Disease Prevalence P	Specificity			Proportion Discordant PD	Alpha
	Total N	Non-Diseased NND		Test 1 Sp1	Test 2 Sp2	Difference Sp1 - Sp2		
0.90007	2798	2238	0.2	0.75	0.7875	-0.0375	0.3	0.05
0.90040	697	557	0.2	0.75	0.8250	-0.0750	0.3	0.05

\* Power was computed using the normal approximation method (conditional).

The required sample sizes depend heavily on the specificity to detect.

## Example 3 – Validation using Machin et al. (2009)

Machin et al. (2009) page 166-167 gives the results of a sample size determination in which the sensitivity of test 1 is 0.27, the sensitivity of test 2 is 0.66,  $P = 0.25$ ,  $\alpha = 0.05$  (two-sided), and power = 0.80. The resulting sample size is 80 using one formula and 100 using another. The formulas used in their book are based on a conditional argument that does not require the proportion discordant.

The power calculation formulas for specificity are the same as those for sensitivity, except that  $1 - P$  is used to adjust the sample size instead of  $P$ . Therefore, we will match the results in Machin et al. (2009) when entering specificity in place of sensitivity if we enter  $1 - P$  (0.75) for the disease prevalence instead of  $P$  (0.25).

We will compare our results to theirs using PD = 0.4, 0.5, and 0.6 and calculate power using Multinomial Enumeration.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For .....	<b>Sample Size</b>
Power Calculation Method .....	<b>Multinomial Enumeration</b>
Max N <sub>ND</sub> for Multinomial Enumeration .....	<b>2000</b>
Alternative Hypothesis .....	<b>Two-Sided</b>
Power .....	<b>0.8</b>
Alpha .....	<b>0.05</b>
P (Disease Prevalence) .....	<b>0.75</b>
Sp1 (Specificity of Test 1) .....	<b>0.27</b>
Sp2 (Specificity of Test 2) .....	<b>0.66</b>
PD (Proportion Discordant) .....	<b>0.4 0.5 0.6</b>

## Tests for Paired Specificities

## Output

Click the Calculate button to perform the calculations and generate the following output.

### Numeric Results

Solve For: [Sample Size](#)  
 Alternative Hypothesis: Two-Sided ( $H_1: Sp_1 - Sp_2 \neq 0$ )  
 Test Type: McNemar's Test

Power*	Number of Subjects		Disease Prevalence P	Specificity			Proportion Discordant PD	Target Alpha†
	Total N	Non-Diseased N <sub>ND</sub>		Test 1 Sp1	Test 2 Sp2	Difference Sp1 - Sp2		
0.83196	80	20	0.75	0.27	0.66	-0.39	0.4	0.05
0.80961	104	26	0.75	0.27	0.66	-0.39	0.5	0.05
0.81101	128	32	0.75	0.27	0.66	-0.39	0.6	0.05

\* Power was computed using multinomial enumeration of all possible outcomes (unconditional).

† Warning: For small values of N<sub>ND</sub> (i.e., less than 100), power computed by Multinomial Enumeration may be overly optimistic because the discrete nature of the multinomial distribution results in the actual alpha value being higher than its target. To be safe, we recommend that you use the power calculation based on the normal approximation.

**PASS** has also obtained an N of 80 for the case when PD = 0.4. However, the sample size increases if PD is 0.5 or 0.6.