

## Chapter 155

# Tests for Two Correlated Proportions in a Matched Case-Control Design

## Introduction

A 2-by- $M$  case-control study investigates a risk factor relevant to the development of a disease. A population of *case* patients with a disease and *control* patients without the disease is considered. Some of these patients have had exposure to a risk factor of interest. A random sample of  $N$  case patients is selected. Patients are stratified by the levels of a confounding variable (such as age, gender, etc.). For each selected case patient, a random sample of  $M$  matched control patients is drawn from the same strata (group). An estimate of the odds ratio,  $OR$ , of developing the disease in exposed and unexposed patients who have equal values of the confounding variable is desired. This odds ratio is assumed to be constant across all levels of the confounding variables.

To fix these ideas, consider the following fictitious data concerning the relationship between smoking and lung cancer. Suppose that a sample of  $N = 100$  cases of identical twins in which only one twin has lung cancer is selected. The second twin serves as the control. Each pair of twins is surveyed to determine if either, both, or none smoke tobacco. The results are summarized in the following two-way table:

	<u>No Lung Cancer Twin (Control)</u>	
<u>Lung Cancer Twin (Case)</u>	<u>Smokes = Yes</u>	<u>Smokes = No</u>
<u>Smokes = Yes</u>	16	21
<u>Smokes = No</u>	4	59

Note that the values in this table are counts of pairs, not individuals. The proportion of controls who smoke is  $(16+4)/100 = 0.20$ . The proportion of cases who smoke is  $(16+21)/100 = 0.37$ . The odds ratio is  $21/4 = 5.25$ . That is, the twin who smoked is 5.25 times more likely to have lung cancer than the twin who did not.

This procedure is similar to the McNemar procedure also available in *PASS*. It differs from that procedure in three basic ways:

1. The results are based on the normal approximation to the binomial.
2. This procedure lets you have multiple controls for each case. The McNemar procedure only allows one control per case.
3. The input parameters are different.

## Technical Details

The following results summarize the article by Dupont (1988) upon which this module is based. The probabilities that the data fall into various categories are:

1. The probability that a case patient was exposed to the risk factor is  $p_1$ .
2. The probability that a control patient was exposed to the risk factor is  $p_0$ .
3. The probability that a case patient was not exposed to the risk factor is  $q_1 = 1 - p_1$ .
4. The probability that a control patient was not exposed to the risk factor is  $q_0 = 1 - p_0$ .

The odds ratio,  $OR$ , is defined as

$$OR = \frac{p_1 / q_1}{p_0 / q_0}$$

Assume that you use a  $\chi^2$  test for the null hypothesis that  $OR = 1$ , that is, that  $p_0 = p_1$ . Such a test is given by Breslow and Day (1980).

Let  $x_k = 1$  or  $0$  if the  $k^{\text{th}}$  sampled case patient was or was not exposed, respectively. Let  $y_k = 1$  or  $0$  if the corresponding first matched control patient was or was not exposed. Let  $p_{ij} = \Pr(x_k = i \text{ and } y_k = j)$ . For example,  $p_{10}$  is the probability that the case patient was exposed to the risk factor while the corresponding first control patient was not. The relationships between these probabilities are

$$p_1 = p_{11} + p_{10}$$

and

$$p_0 = p_{11} + p_{01}$$

Define  $\phi$  to be the correlation between  $x_k$  and  $y_k$ . It can be shown that

$$\phi = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_1q_1p_0q_0}}$$

A little algebra will show that

$$p_{11} = p_1p_0 + \phi\sqrt{p_1q_1p_0q_0}$$

$$p_{10} = p_1q_0 - \phi\sqrt{p_1q_1p_0q_0}$$

$$p_{01} = q_1p_0 - \phi\sqrt{p_1q_1p_0q_0}$$

$$p_{00} = q_1q_0 + \phi\sqrt{p_1q_1p_0q_0}$$

$$p_{0+} = \frac{p_{01}}{q_1}$$

$$p_{0\cdot} = \frac{p_{01}}{q_1}$$

## Tests for Two Correlated Proportions in a Matched Case-Control Design

$$q_{0+} = 1 - p_{0+}$$

$$q_{0\cdot} = 1 - p_{0\cdot}$$

and

$$t_k = p_1 \binom{M}{k-1} p_{0+}^{k-1} q_{0+}^{M-k+1} + q_1 \binom{M}{k} p_{0\cdot}^k q_{0\cdot}^{M-k}; \quad k = 1, \dots, M$$

Let  $n_{ij}$  represent the number of matched sets of subjects in which the case patient was ( $i = 1$ ) or was not ( $i = 0$ ) exposed and  $j$  of the  $M$  control subjects were exposed. Let

$$y = \sum_{m=1}^M n_{1,m-1}$$

be the number of discordant sets in which the case patient was exposed and let

$$T_m = n_{1,m-1} + n_{0,m}$$

be the number of sets in which  $m$  subjects were exposed. The expected value of  $T_m$  is  $Nt_m$ .

Let

$$e_{OR} = \sum_{m=1}^M \frac{m t_m(OR)}{m(OR) + M - m + 1}$$

$$v_{OR} = \sum_{m=1}^M \frac{m t_m(OR)(M - m + 1)}{(m(OR) + M - m + 1)^2}$$

Dupont (1988) provides the following formula relating  $\alpha$ ,  $\beta$ ,  $p_0$ ,  $\phi$ ,  $OR$ ,  $N$ , and  $M$ .

$$1 - \beta = \Phi\left(\frac{\sqrt{N}(e_1 - e_{OR}) - z_{\alpha/2}\sqrt{v_1}}{\sqrt{v_{OR}}}\right) + 1 - \Phi\left(\frac{\sqrt{N}(e_1 - e_{OR}) + z_{\alpha/2}\sqrt{v_1}}{\sqrt{v_{OR}}}\right)$$

This equation may be used to make power and sample size calculations.

## Estimating the Sample Control Exposure Probability

To calculate power and sample size, a value for the probability that a sample control patient was exposed to the risk factor ( $p_0$ ) must be estimated. Remember that the control sample is not a random sample of the population. Instead, it is matched to a random sample of case patients. Hence, the sample does not necessarily provide an unbiased estimate of  $p_0$ . Care should be taken to provide an accurate estimate of the probability that a matched control patient was exposed, not the probability that someone was exposed in the general population. However, when there is little association between the confounding (matching) variable and the exposure variable in the control population, the baseline probability of the exposure variable may be used.

## Estimating the Correlation, $\phi$

Previous matched 2-by-2 contingency tables can be used to estimate  $\phi$  using the relationship

$$\phi = \sqrt{\frac{\chi_u^2}{N}}$$

where

$$\chi_u^2 = \frac{N(n_{00}n_{11} - n_{01}n_{10})^2}{n_{0\bullet}n_{1\bullet}n_{\bullet 0}n_{\bullet 1}}$$

When no previous data are available about  $\phi$ , Dupont (1988) suggests using a value of 0.2 rather than 0.

## Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

### Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

#### Solve For

##### Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Odds Ratio*, *P0*, *Correlation*, *Sample Size (Cases)*, *Sample Size (Controls per Case)*, *Alpha*, and *Power*. Under most situations, you will select either *Power* or *Sample Size*.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment that has already been run.

#### Power and Alpha

##### Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

##### Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, the test is two-sided.

## Tests for Two Correlated Proportions in a Matched Case-Control Design

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

A single value may be entered here or a range of values such as *0.05 to 0.2 by 0.05* may be entered.

---

### Sample Size

#### **N (Number of Case Patients)**

The number of case patients to be sampled. This is the number of pairs. Since there are  $M$  controls for each case patient, the total number of patients is  $N(M + 1)$ . Values of one or greater are allowed here.

#### **M (Number of Controls per Case)**

The number of control patients matched with each case patient. When  $M$  is one, McNemar's test is used to analyze the data.

The total number of patients in the study is  $N(M + 1)$ . Values of one or greater are allowed here.

---

### Effect Size

#### **P0 (Probability that a Control is Exposed)**

Specify the value of  $p_0$ , the probability that a sample control patient is exposed to the risk factor. If the matching variables are independent of the risk factor, you can use the baseline rate of exposure to the risk factor in the general population.

Since this is a probability, it must be between zero and one.

You can enter a list of values separated by commas or blanks.

#### **OR (Odds Ratio)**

This option sets the value of the odds ratio,  $OR$ , which is the ratio of the disease odds of individuals exposed to the risk factor to the disease odds of individuals not exposed to the risk factor.

For example, an odds ratio of 2.0 means that subjects exposed to the risk factor are twice as likely of developing the disease as are unexposed subjects.

A value greater than one is usually used. The value must be greater than zero. The null hypothesis is that the odds ratio is one. You should use a value that will be of interest to others, such as 1.5 or 2.0.

You can enter a list of values.

#### **Phi (Correlation between Case and Control)**

This is the correlation for exposure between a case subject and the first of the corresponding control subjects. A value of zero here indicates independence between exposure rates for case and controls. Often, assuming complete independence is unrealistic, so when no other information is available, Dupont (1988) suggests using a value of 0.2.

Correlations can range between -1 and 1. However, only positive correlations should be used.

## Tests for Two Correlated Proportions in a Matched Case-Control Design

### Example 1 – Calculating Power

This example will show how to calculate the power of a retrospective study for several sample sizes and odds ratios.

Suppose that a matched case-control study is to be run in which the odds ratios of interest are 1.5, 2.5, or 3.5,  $P_0 = 0.6$ , correlation = 0.2,  $M = 1$ ,  $N = 25\ 50\ 100\ 150\ 200$ , alpha = 0.05, and power is to be found.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions in a Matched Case-Control Design** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions in a Matched Case-Control Design**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
N (Case Patients) .....	<b>25 50 100 150 200</b>
M (Controls Per Case) .....	<b>1</b>
P0 (Prob Control Exposed).....	<b>0.6</b>
OR (Odds Ratio) .....	<b>1.5 2.5 3.5</b>
Phi (Correlation of Case and Control) ....	<b>0.2</b>

### Annotated Output

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

### Numeric Results

Numeric Results								
Power	Cases (N)	Controls Per Case (M)	Odds Ratio (OR)	Probability Exposed (P0)	Correlation (Phi)	Alpha	Beta	
0.08863	25	1	1.50	0.60000	0.20000	0.05000	0.91137	
0.13364	50	1	1.50	0.60000	0.20000	0.05000	0.86636	
0.22622	100	1	1.50	0.60000	0.20000	0.05000	0.77378	
0.31832	150	1	1.50	0.60000	0.20000	0.05000	0.68168	
0.40652	200	1	1.50	0.60000	0.20000	0.05000	0.59348	
0.23067	25	1	2.50	0.60000	0.20000	0.05000	0.76933	
0.44278	50	1	2.50	0.60000	0.20000	0.05000	0.55722	
0.75646	100	1	2.50	0.60000	0.20000	0.05000	0.24354	
0.90966	150	1	2.50	0.60000	0.20000	0.05000	0.09034	
0.97004	200	1	2.50	0.60000	0.20000	0.05000	0.02996	
0.36379	25	1	3.50	0.60000	0.20000	0.05000	0.63621	
0.68570	50	1	3.50	0.60000	0.20000	0.05000	0.31430	
0.95159	100	1	3.50	0.60000	0.20000	0.05000	0.04841	
0.99482	150	1	3.50	0.60000	0.20000	0.05000	0.00518	
0.99956	200	1	3.50	0.60000	0.20000	0.05000	0.00044	

## Tests for Two Correlated Proportions in a Matched Case-Control Design

### Report Definitions

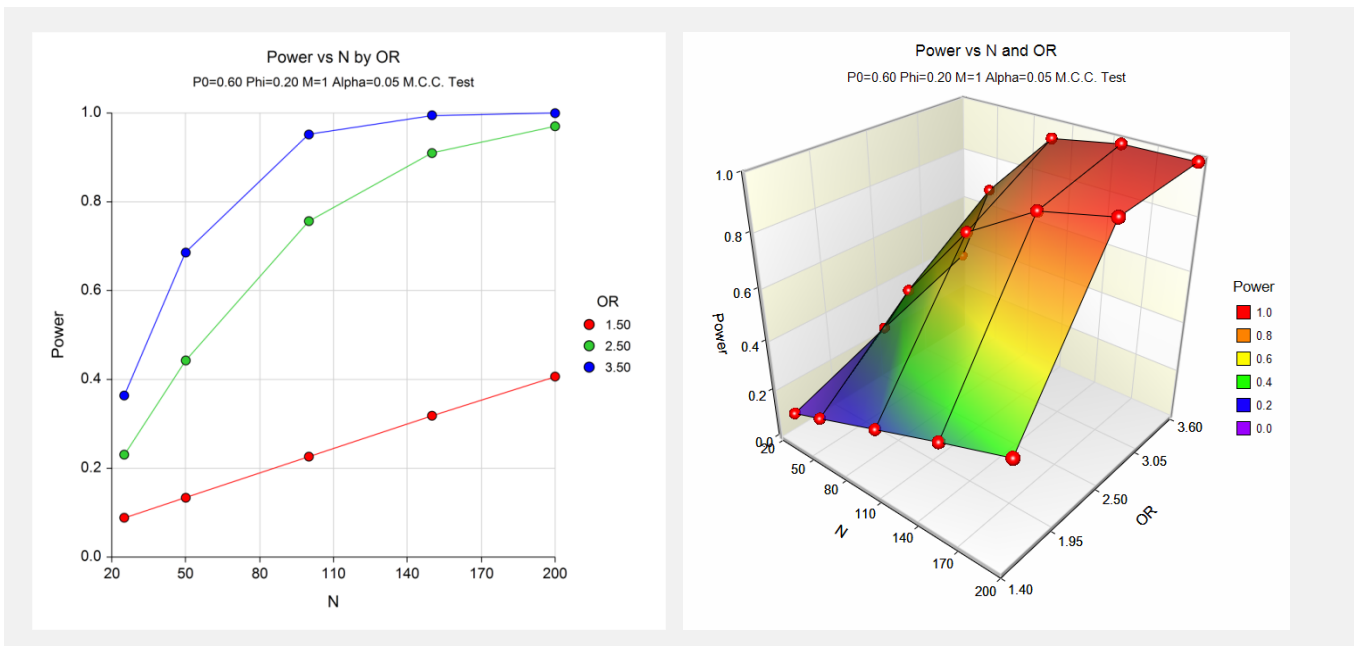
Power is the probability of rejecting a false null hypothesis.  
 N is the size of the sample drawn from the treatment (case) group.  
 M is the number of matching control patients drawn for each case patient.  
 OR is the odds ratio of for subjects exposed to the risk factor.  
 P0 is the probability of exposure among sampled control patients.  
 Phi is the correlation of exposure between matched individuals.  
 Alpha is the probability of rejecting a true null hypothesis.  
 Beta is the probability of accepting a false null hypothesis.

### Summary Statements

In a matched case-control study, the probability of exposure among sampled control patients is 0.60000 and the correlation coefficient for exposure between matched case and control patients is 0.20000. A sample of 25 case patients is obtained. For each case patient, a matching sample of 1 control patient(s) is also obtained. This sample of 50 patients achieves 9% power to detect an odds ratio of 1.50 versus the alternative of equal odds using a Chi-Square test with a 0.05000 significance level.

This report shows the power for each of the scenarios.

### Plots Section



This plot shows the power versus the sample size for the three odds ratios.

## Example 2 – Calculating Sample Size

Suppose that a matched case-control study is planned to study the relationship between smoking and a certain kind of cancer. Researchers want to have a sample large enough to detect an odds ratio of 2.0. During the power analysis, the researchers also want to calculate the required sample size for odds ratios of 1.5 and 2.5.

The probability that a sampled control (non-cancer) patient smokes is estimated at 0.3. The correlation of smoking between cases and controls is 0.2. The researchers want samples sizes large enough to achieve 80% power at the 0.05 significance levels. In an effort to reduce the number of cancer patients that must be enrolled, the researchers want to try several values for the number of controls per case between 1 and 20.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions in a Matched Case-Control Design** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions in a Matched Case-Control Design**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size (Cases)</b>
Power.....	<b>0.80</b>
Alpha.....	<b>0.05</b>
M (Controls Per Case) .....	<b>1 2 3 4 5 10 20</b>
P0 (Prob Control Exposed).....	<b>0.3</b>
OR (Odds Ratio) .....	<b>1.5 2.0 2.5</b>
Phi (Correlation of Case and Control) ....	<b>0.2</b>



Tests for Two Correlated Proportions in a Matched Case-Control Design

# Output

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

## Numeric Results and Plots

Numeric Results							
Power	Cases (N)	Controls Per Case (M)	Odds Ratio (OR)	Probability Exposed (P0)	Correlation (Phi)	Alpha	Beta
0.80064	539	1	1.50	0.30000	0.20000	0.05000	0.19936
0.80021	378	2	1.50	0.30000	0.20000	0.05000	0.19979
0.80025	324	3	1.50	0.30000	0.20000	0.05000	0.19975
0.80038	297	4	1.50	0.30000	0.20000	0.05000	0.19962
0.80076	281	5	1.50	0.30000	0.20000	0.05000	0.19924
0.80009	248	10	1.50	0.30000	0.20000	0.05000	0.19991
0.80044	232	20	1.50	0.30000	0.20000	0.05000	0.19956
0.80100	180	1	2.00	0.30000	0.20000	0.05000	0.19900
0.80249	127	2	2.00	0.30000	0.20000	0.05000	0.19751
0.80077	108	3	2.00	0.30000	0.20000	0.05000	0.19923
0.80220	99	4	2.00	0.30000	0.20000	0.05000	0.19780
0.80087	93	5	2.00	0.30000	0.20000	0.05000	0.19913
0.80261	82	10	2.00	0.30000	0.20000	0.05000	0.19739
0.80125	76	20	2.00	0.30000	0.20000	0.05000	0.19875
0.80156	102	1	2.50	0.30000	0.20000	0.05000	0.19844
0.80262	72	2	2.50	0.30000	0.20000	0.05000	0.19738
0.80028	61	3	2.50	0.30000	0.20000	0.05000	0.19972
0.80327	56	4	2.50	0.30000	0.20000	0.05000	0.19673
0.80559	53	5	2.50	0.30000	0.20000	0.05000	0.19441
0.80351	46	10	2.50	0.30000	0.20000	0.05000	0.19649
0.80672	43	20	2.50	0.30000	0.20000	0.05000	0.19328

This report shows the sample size for each of the scenarios. Notice that the required number of cancer patients ( $N$ ) drops off drastically as more controls are added. However, using more than five controls seems to only moderately reduce the sample size necessary sample size.

Also notice that the difference in sample size is much larger when moving from an odds ratio of 2.0 to 1.5 than from 2.5 to 2.0.

## Example 3 – Validation using Dupont

The formulas used in this module were given in Dupont (1988). He gives an example on page 1164 of the article in which  $P_0$  is 0.6,  $\Phi$  is 0.2,  $OR$  is 3.0,  $\alpha$  is 0.05, and  $\beta$  is 0.2. Dupont finds the sample size for  $M = 1$  to be 80 and for  $M = 3$  to be 50.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Correlated Proportions in a Matched Case-Control Design** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Correlated Proportions in a Matched Case-Control Design**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size (Cases)</b>
Power .....	<b>0.80</b>
Alpha .....	<b>0.05</b>
M (Controls Per Case) .....	<b>1 3</b>
P0 (Prob Control Exposed) .....	<b>0.6</b>
OR (Odds Ratio) .....	<b>3.0</b>
Phi (Correlation of Case and Control) ....	<b>0.2</b>

### Output

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

#### Numeric Results

Numeric Results								
Power	Cases (N)	Controls Per Case (M)	Odds Ratio (OR)	Probability Exposed (P0)	Correlation (Phi)	Alpha	Beta	
0.80149	80	1	3.00	0.60000	0.20000	0.05000	0.19851	
0.80052	50	3	3.00	0.60000	0.20000	0.05000	0.19948	

Note that values of 80 and 50 for N agree exactly with Dupont.