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:

| | No Lung Cancer Twin (Control) | |
|---|---|---|
| **Lung Cancer Twin (Case)** | **Smokes = Yes** | **Smokes = No** |
| **Smokes = Yes** | 16 | 21 |
| **Smokes = No** | 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 $k^{th}$ 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 $p_{ij} = Pr(x_k = i$ 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

$$\phi = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{p_1 q_1 p_0 q_0}}$$

A little algebra will show that

$$p_{11} = p_1 p_0 + \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{10} = p_1 q_0 + \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{01} = q_1 p_0 - \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{00} = q_1 q_0 + \phi\sqrt{p_1 q_1 p_0 q_0}$$

$$p_{0+} = \frac{p_{11}}{p_1}$$

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

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

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

and

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

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, Φ

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

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

where

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

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

# 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, *P0* = 0.6, correlation = 0.2, *M* = 1, *N* = 25 50 100 150 200, alpha = 0.05, and power is to be found.

## 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 ........................................ **Power**
Alpha............................................... **0.05**
N (Case Patients) ............................ **25 50 100 150 200**
M (Controls Per Case) ...................... **1**
P0 (Prob Control Exposed) ................ **0.6**
OR (Odds Ratio) ............................... **1.5 2.5 3.5**
Phi (Correlation of Case and Control)........... **0.2**

## Output

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

### Numeric Reports

**Numeric Results**

Solve For: Power

| Power | Cases N | Controls per Case M | Odds Ratio OR | Probability Exposed P0 | Correlation Phi | Alpha | Beta |
|---|---|---|---|---|---|---|---|
| 0.08863 | 25 | 1 | 1.5 | 0.6 | 0.2 | 0.05 | 0.91137 |
| 0.13364 | 50 | 1 | 1.5 | 0.6 | 0.2 | 0.05 | 0.86636 |
| 0.22622 | 100 | 1 | 1.5 | 0.6 | 0.2 | 0.05 | 0.77378 |
| 0.31832 | 150 | 1 | 1.5 | 0.6 | 0.2 | 0.05 | 0.68168 |
| 0.40652 | 200 | 1 | 1.5 | 0.6 | 0.2 | 0.05 | 0.59348 |
| 0.23067 | 25 | 1 | 2.5 | 0.6 | 0.2 | 0.05 | 0.76933 |
| 0.44278 | 50 | 1 | 2.5 | 0.6 | 0.2 | 0.05 | 0.55722 |
| 0.75646 | 100 | 1 | 2.5 | 0.6 | 0.2 | 0.05 | 0.24354 |
| 0.90966 | 150 | 1 | 2.5 | 0.6 | 0.2 | 0.05 | 0.09034 |
| 0.97004 | 200 | 1 | 2.5 | 0.6 | 0.2 | 0.05 | 0.02996 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.36379 | 25 | 1 | 3.5 | 0.6 | 0.2 | 0.05 | 0.63621 |
| 0.68570 | 50 | 1 | 3.5 | 0.6 | 0.2 | 0.05 | 0.31430 |
| 0.95159 | 100 | 1 | 3.5 | 0.6 | 0.2 | 0.05 | 0.04841 |
| 0.99482 | 150 | 1 | 3.5 | 0.6 | 0.2 | 0.05 | 0.00518 |
| 0.99956 | 200 | 1 | 3.5 | 0.6 | 0.2 | 0.05 | 0.00044 |

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

**Summary Statements**

A matched case-control design will be used to compare two correlated proportions. Assume the probability of exposure among sampled control subjects is 0.6 and the correlation coefficient for exposure between matched case and control subjects is 0.2. To detect an odds ratio of 1.5 versus the alternative of equal odds with 25 case subjects and a matching sample of 1 control subject per case subject (totaling 50 subjects), and with a Type I error rate (α) of 0.05, the power is 0.08863.

**Dropout-Inflated Sample Size**

| Dropout Rate | Sample Size N | Dropout-Inflated Enrollment Sample Size N' | Expected Number of Dropouts D |
|---|---|---|---|
| 20% | 25 | 32 | 7 |
| 20% | 50 | 63 | 13 |
| 20% | 100 | 125 | 25 |
| 20% | 150 | 188 | 38 |
| 20% | 200 | 250 | 50 |

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, 32 subjects should be enrolled to obtain a final sample size of 25 subjects.

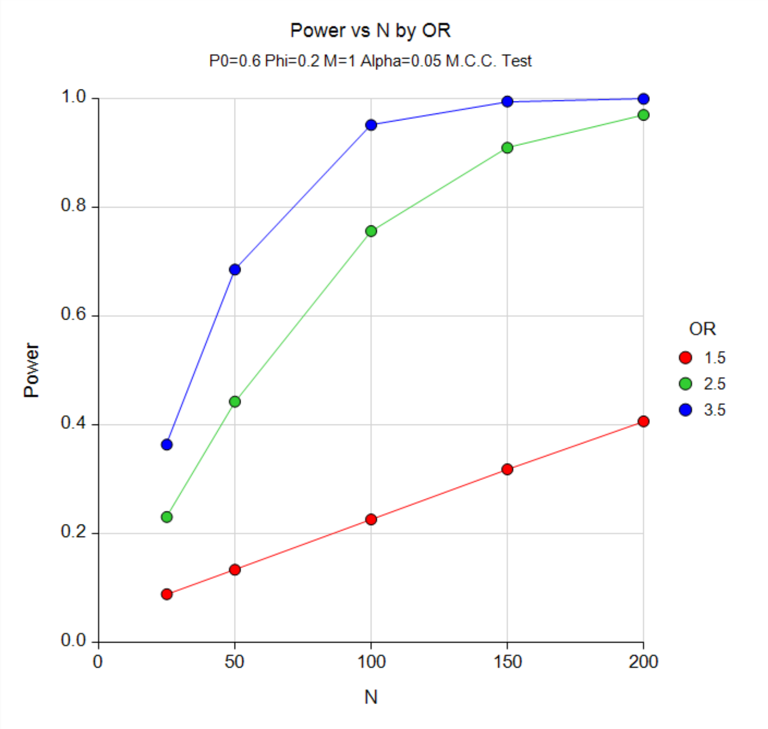
**References**

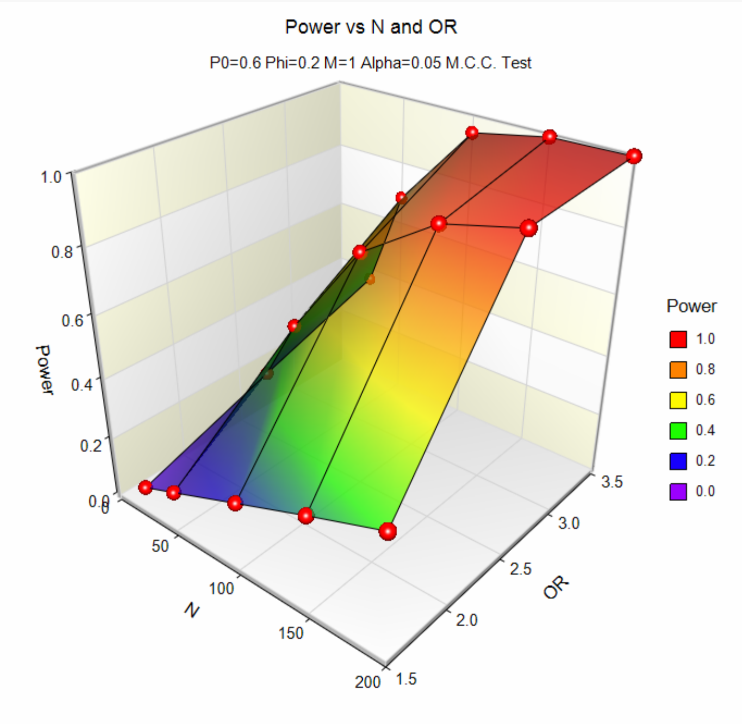
Dupont, William. 1988. 'Power Calculations for Matched Case-Control Studies', Biometrics, Volume 44, pages 1157-1168.

This report shows the power for each of the scenarios.

## Plots Section

### Plots




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

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 (Cases)**
Power........................................................... **0.80**
Alpha............................................................ **0.05**
M (Controls Per Case) ................................... **1 2 3 4 5 10 20**
P0 (Prob Control Exposed) ............................ **0.3**
OR (Odds Ratio) ........................................... **1.5 2.0 2.5**
Phi (Correlation of Case and Control)............ **0.2**

## Output

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

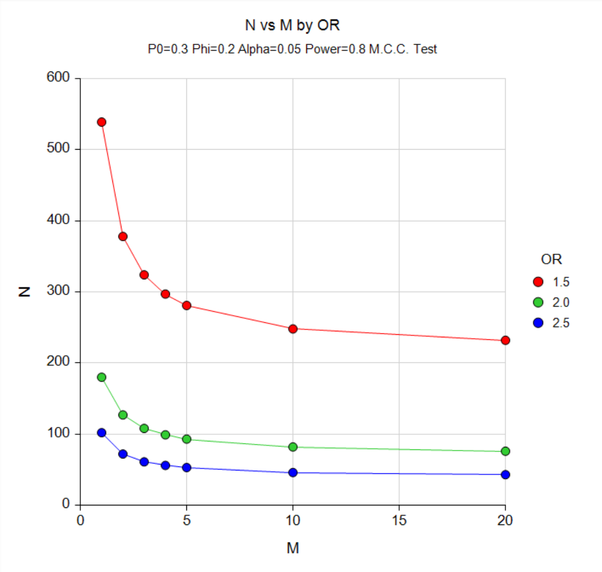
**Numeric Results**

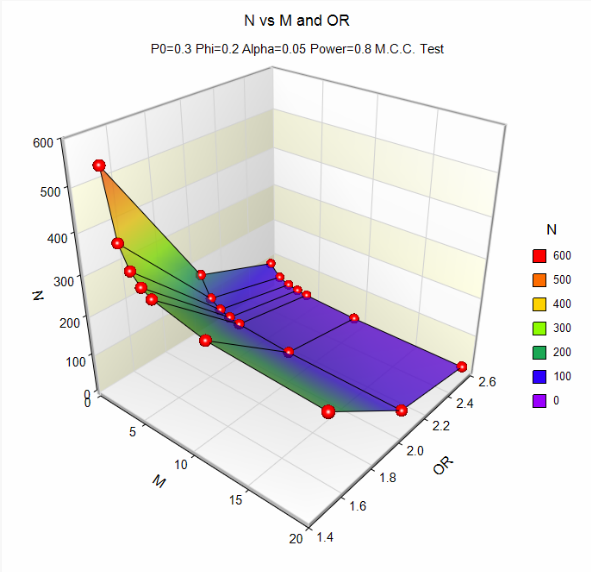
Solve For: Sample Size (Cases)

| Power | Cases N | Controls per Case M | Odds Ratio OR | Probability Exposed P0 | Correlation Phi | Alpha | Beta |
|---|---|---|---|---|---|---|---|
| 0.80064 | 539 | 1 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19936 |
| 0.80021 | 378 | 2 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19979 |
| 0.80025 | 324 | 3 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19975 |
| 0.80038 | 297 | 4 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19962 |
| 0.80076 | 281 | 5 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19924 |
| 0.80009 | 248 | 10 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19991 |
| 0.80044 | 232 | 20 | 1.5 | 0.3 | 0.2 | 0.05 | 0.19956 |
| 0.80100 | 180 | 1 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19900 |
| 0.80249 | 127 | 2 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19751 |
| 0.80077 | 108 | 3 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19923 |
| 0.80220 | 99 | 4 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19780 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.80087 | 93 | 5 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19913 |
| 0.80261 | 82 | 10 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19739 |
| 0.80125 | 76 | 20 | 2.0 | 0.3 | 0.2 | 0.05 | 0.19875 |
| 0.80156 | 102 | 1 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19844 |
| 0.80262 | 72 | 2 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19738 |
| 0.80028 | 61 | 3 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19972 |
| 0.80327 | 56 | 4 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19673 |
| 0.80559 | 53 | 5 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19441 |
| 0.80351 | 46 | 10 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19649 |
| 0.80672 | 43 | 20 | 2.5 | 0.3 | 0.2 | 0.05 | 0.19328 |

## Plots




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 (1988)

The formulas used in this module were given in Dupont (1988). He gives an example on page 1164 of the article in which *P0* 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

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 ....................................................**Sample Size (Cases)**
Power.........................................................**0.80**
Alpha.........................................................**0.05**
M (Controls Per Case) .................................**1 3**
P0 (Prob Control Exposed) ...........................**0.6**
OR (Odds Ratio) ..........................................**3.0**
Phi (Correlation of Case and Control)...........**0.2**

## Output

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

**Numeric Results**

Solve For: Sample Size (Cases)

| Power | Cases N | Controls per Case M | Odds Ratio OR | Probability Exposed P0 | Correlation Phi | Alpha | Beta |
|---|---|---|---|---|---|---|---|
| 0.80149 | 80 | 1 | 3 | 0.6 | 0.2 | 0.05 | 0.19851 |
| 0.80052 | 50 | 3 | 3 | 0.6 | 0.2 | 0.05 | 0.19948 |

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