Chapter 157

Tests for the Odds Ratio in a Matched Case-Control Design with a Quantitative X

Introduction

This procedure calculates the power and sample size necessary in a matched case-control study designed to detect a relationship between the occurrence of a disease and a quantitative risk factor (exposure variable) using an odds ratio computed from a conditional logistic regression. The procedure also provides an adjustment to power for other covariates. Kleinbaum and Klein (2010) provide a detailed discussion of the interpreting the odds ratio in a conditional logistic regression.

Suppose a subject population is to be studied for the relationship between an outcome variable (such as lung cancer) and a quantitative risk factor (such as amount of cigarette smoking). A matched case-control study is planned in which N matched sets will used. Each matched set will consist of M_D **case** subjects which are positive for the outcome (diseased) and M_H **control** subjects that are negative for the outcome (healthy). The subjects in each set are matched according other covariates that are assumed to have a large impact on the probability of the disease such as age and gender. In each matched set a quantitative **exposure** variable is measured. Note that the design may be *retrospective* or *prospective*.

Technical Details

Hypotheses are investigated using a score test of the log odds ratio in a conditional logistic regression. Power and sample size formulas are given in Lachin (2008) and Tang (2009).

It is assumed that a set of N matched sets of cases and controls are available. Suppose a quantitative exposure variable X (the covariate) is measured for each subject. The data can be fit using conditional logistic regression. This is will result in regression coefficients for X and for any other independent variables included in the model. The regression coefficient of X is interpreted as the log odds ratio of a positive outcome for two values of X with a difference equal to one.

This relationship may be written in terms of *OR* as

$$OR = \exp(\beta_X X)$$

The regression coefficient, log(OR), is tested using a normally distributed score test. The power of the score test is calculated using

$$z_{1-\beta} = |\theta| \sqrt{N \sigma_X^2 \left(\frac{M_D M_H}{M_D + M_H} \right)} - z_{1-\alpha}$$

where σ_X^2 is the variance of the X values.

This can be rearranged to obtain the following expression for sample size

$$N = \frac{\left(z_{1-\beta} + z_{1-\alpha} \right)^2}{\sigma_X^2 \theta^2 \left(\frac{M_D M_H}{M_D + M_H} \right)}$$

Note that for two-sided tests, α is replaced by α/2.

Adjusting for Other Covariates

Lachin (2008) provides an adjustment to the power when additional covariates are fit in the conditional logistic regression. Let $R_{X|Z}^2$ represent the coefficient of determination for a (multiple) regression of the exposure variable X on the covariates Z. Note that Z is a vector of 1 or more covariates and that the number of covariates is not needed. The adjustment is made by multiplying σ_X^2 by $1 - R_{X|Z}^2$ in the formulas above. Lachin stresses that in order for this adjustment to be accurate, none of the adjusting covariates can have a strong effect upon the response. He indicates that this assumption should be meet since any covariate with a large effect should be controlled for by the matching.

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. 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.

Test

Alternative Hypothesis

Specify whether the alternative hypothesis of the test is one-sided or two-sided. If a one-sided test is chosen, the hypothesis test direction is chosen based on whether OR is greater than or less than one.

Two-Sided Hypothesis Test

H0: OR = 1 vs. Ha: OR ≠ 1

One-Sided Hypothesis Tests

Upper: H0: OR ≤ 1 vs. Ha OR > 1

Lower: H0: OR ≥ 1 vs. Ha OR < 1

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

Alpha is the probability of rejecting a true null hypothesis. The null hypothesis is that the odds ratio (OR) is 1. This means that the probability of a positive outcome is no different for those that have been exposed to the risk factor and those that have not.

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 Sets or Strata)

N is the number of matched sets in the study. Each set consists of M_D cases and M_H controls. The total number of subjects is N x $(M_D + M_H)$.

The range of possible values is of 3 or more. Multiple values may be entered.

M_D (Number of Cases per Set)

The number of case subjects in each of the N matched sets. The cases are those subjects with a positive outcome, such as diseased. Thus, the total number of subjects is N x $(M_D + M_H)$.

Values of one or greater are allowed. Seldom is this value chosen to be more than five.

Multiple values may be entered.

M (Number of Controls per Set)

The number of control subjects in each of the N matched sets. The controls are those subjects with a negative outcome, such as healthy (non-diseased). The total number of subjects is N x (M_D + M_H).

Values of one or greater are allowed. Seldom is this value chosen to be more than ten.

Multiple values may be entered.

Effect Size

OR (Odds Ratio | 1-Unit Change in X)

The value of the odds ratio to be detected. This is the odds ratio at which the power is calculated. Note that the power or sample size will be identical for 1/OR and OR.

The odds ratio is the ratio of the odds at X = x + 1 to the odds at X = x. The logarithm of the odds ratio is used in the formula. Log (OR) is equal to the regression coefficient of the quantitative covariate X in the conditional logistic regression.

For example, an odds ratio of 2.0 means that subjects for which the covariate value is x+1 have twice the odds of having the disease as do subjects for which the covariate value is x.

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 can enter a single value such as *1.5*, a list of values such as *1.5 2.0 2.5*, or a series of values such as

1.5 to 2.5 by 0.01.

σ_X (Standard Deviation of X)

Enter an estimate of the standard deviation of X, the quantitative variable of interest. Note that as σ_X increases, the sample size decreases. If you don't have an estimate, you can select the *Standard Deviation Estimator* from the *Tools* menu to obtain a window that will help you determine a rough estimate of the standard deviation.

The range of this parameter is greater than zero.

You can enter a single value such as 1.0, a list of values such as 1.1 1.3 1.5, or a series of values such as

1 to 3 by 0.5.

R² (Exposure vs. Covariates)

Enter the R^2 value that would be obtained if the X variable was regressed on the other covariates (independent variables) to be used in the conditional logistic model. This is referring to the R^2 from a regular multiple regression of X on the covariates.

This R^2 value is used to calculate an adjustment to the power and sample size when the covariates are in the model. Note that the number of covariates is not used in the adjustment and so it is not entered.

If you do not have any covariates, enter '0.0'.

The absolute range is from zero to just less than one. However, it is assumed that covariates with large impacts will be used in the matching process.

You can enter a single value such as *0.1*, a list of values such as *0.05 0.1 0.2*, or a range of values such as *0.0 to 0.1 by 0.01*.

Example 1 – Calculating Sample Size

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

Suppose that a matched case-control study is to be run in which the OR = 1.5, 2.0, 2.5, or 3.0, σ_x = 1.3, R^2 = 0.2, M_D = 1, M_H = 1, 2, or 5, power = 0.9, and 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 the Odds Ratio in a Matched Case-Control Design with a Quantitative X** procedure window. 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**.

Option	Value
Design Tab	
Solve For	**Sample Size N**
Alternative Hypothesis	**Two-Sided**
Power	**0.9**
Alpha	**0.05**
M_D (Number of Cases per Set)	**1**
M_H (Number of Controls per Case)	**1 2 5**
OR (Odds Ratio)	**1.5 2 2.5 3**
σ_x (Standard Deviation of X)	**1.3**
R^2 (Exposure vs. Covariates)	**0.2**

Annotated Output

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

Numeric Results

Numeric Results
Hypothesis Type: Two-Sided

Power	Number of Matched Sets N	Cases per Set M_D	Controls per Set M_H	Odds Ratio OR	Standard Deviation of the Covariate X σ_x	Regression of X on other Covariates R^2	Alpha
0.9000	95	1	1	1.5000	1.3000	0.2000	0.0500
0.9000	71	1	2	1.5000	1.3000	0.2000	0.0500
0.9000	57	1	5	1.5000	1.3000	0.2000	0.0500
0.9000	33	1	1	2.0000	1.3000	0.2000	0.0500
0.9000	25	1	2	2.0000	1.3000	0.2000	0.0500
0.9000	20	1	5	2.0000	1.3000	0.2000	0.0500
0.9000	19	1	1	2.5000	1.3000	0.2000	0.0500
0.9000	14	1	2	2.5000	1.3000	0.2000	0.0500
0.9000	12	1	5	2.5000	1.3000	0.2000	0.0500
0.9000	13	1	1	3.0000	1.3000	0.2000	0.0500
0.9000	10	1	2	3.0000	1.3000	0.2000	0.0500
0.9000	8	1	5	3.0000	1.3000	0.2000	0.0500

Tests for the Odds Ratio in a Matched Case-Control Design with a Quantitative X

References
Lachin, John M. 2008. 'Sample size evaluaion for a multiply matched case-control study using the score test from a conditional logistic (discrete Cox PH) regression model.' Statistics in Medicine, Volume 27, Pages 2509-2523.
Lachin, John M. 2011. Biostatistical Methods: The Assessment of Relative Risks, Second Edition. John Wiley & Sons. New York.
Tang, Yongqiang. 2009. 'Comments on 'Sample size evaluation for multiply matched case-control study using the score test from a conditional logistic (discrete Cox PH) regression model.'' Statistics in Medicine, Volume 28, Pages 175-177.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
Number of Matched Sets, N, is the number of sets (strata) in the study. Each set consists in a fixed number of cases and controls.
Cases per Set, M_D, is the number cases in each matched set.
Controls per Set, M_H, is the number of controls in each matched set.
Odds Ratio, OR, is the ratio of the odds of a positive outcome when the covariate is X+1 to the odds of a positive outcome when the covariate is X.
Standard Deviation of the Covariate X, σ_X, is the average SD of X across all subjects in the study.
Regression of X on other Covariates, R^2, is the R^2 that occurs when the covariate variable of interest X is regressed on any other covariates. This adjustment assumes that covariates that have a large correlation with the outcome are used in the matching process and are not included here.
Alpha is the probability of rejecting a true null hypothesis of no association between disease and the exposure variable.

Summary Statements
In a matched case-control study, a sample of 95 matched sets (or strata) is obtained. Each matched set consists of 1 case and 1 control. The standard deviation of X is 1.3000. This sample achieves 90% power to detect an odds ratio of 1.5000 calculated using conditional logistic regression with a 0.0500 significance level.

This report shows the power for each of the scenarios.

Plots Section
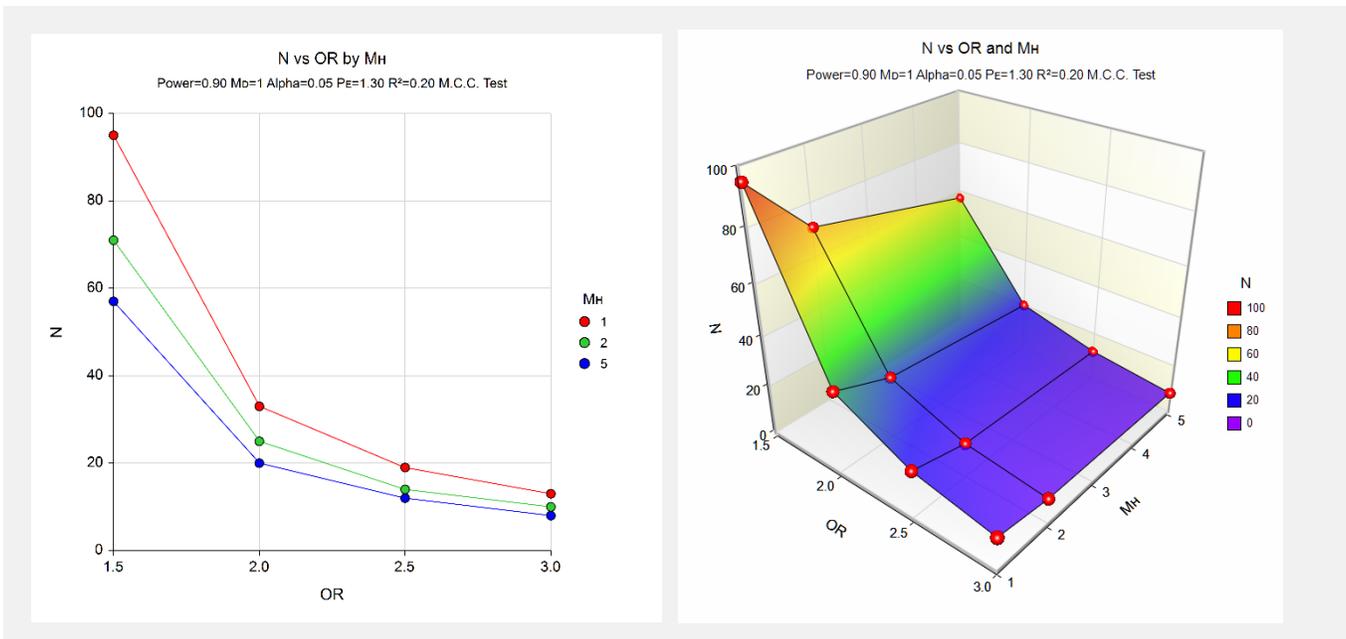
This plot shows the sample size versus the odds ratio for the three M_H's.

Example 2 – Validation using Lachin (2011)

This example will validate this procedure by comparing the results to those in Lachin (2011) on page 351. In this example, OR = 1.4, $\sigma_x = 1$, $R^2 = 0.0$, $M_D = 1$, $M_H = 2$, power = 0.85, and alpha = 0.05. The test is two-sided. The resulting sample size is 119.

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 the Odds Ratio in a Matched Case-Control Design with a Quantitative X** procedure window. 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**.

Option   **Value**

Design Tab
Solve For .... **Sample Size N**
Alternative Hypothesis .... **Two-Sided**
Power **0.85**
Alpha **0.05**
M_D (Number of Cases per Set) **1**
M_H (Number of Controls per Case) **2**
OR (Odds Ratio) **1.4**
σ_x (Standard Deviation of X) **1.0**
R^2 (Exposure vs. Covariates) **0.0**

Output

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

Numeric Results

Numeric Results
Hypothesis Type: Two-Sided

Power	Number of Matched Sets N	Cases per Set M_D	Controls per Set M_H	Odds Ratio OR	Standard Deviation of the Covariate X σ_x	Regression of X on other Covariates R^2	Alpha
0.8500	119	1	2	1.4000	1.0000	0.0000	0.0500

PASS has also calculated N to be 119.