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 be 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 to 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 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 $\alpha/2$.

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 met since any covariate with a large effect should be controlled for by the matching.

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

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 .... **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**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size N
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
0.9014	95	1	1	1.5	1.3	0.2	0.05
0.9004	71	1	2	1.5	1.3	0.2	0.05
0.9014	57	1	5	1.5	1.3	0.2	0.05
0.9056	33	1	1	2.0	1.3	0.2	0.05
0.9083	25	1	2	2.0	1.3	0.2	0.05
0.9083	20	1	5	2.0	1.3	0.2	0.05
0.9072	19	1	1	2.5	1.3	0.2	0.05
0.9023	14	1	2	2.5	1.3	0.2	0.05
0.9206	12	1	5	2.5	1.3	0.2	0.05
0.9027	13	1	1	3.0	1.3	0.2	0.05
0.9096	10	1	2	3.0	1.3	0.2	0.05
0.9096	8	1	5	3.0	1.3	0.2	0.05

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The number of matched sets (strata) in the study. Each set consists in a fixed number of cases and controls.
M_D The number cases in each matched set.
M_H The number of controls in each matched set.
OR 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.
σ_x The average standard deviation of the covariate X across all subjects in the study.
R^2 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 The probability of rejecting a true null hypothesis of no association between disease and the exposure variable.

Summary Statements

A matched case-control design will be used to test the relationship between a binary outcome and a quantitative risk factor (exposure variable) based on the odds ratio. The comparison will be made using a two-sided odds ratio score test from a conditional logistic regression analysis. The assumed standard deviation of the quantitative exposure variable of interest is 1.3. The assumed R-squared when regressing the exposure variable on any other covariates is 0.2. Each set of matched case-controls will consist of one case and one matched control. To detect an odds ratio of 1.5 with 90% power and a Type I error rate (α) of 0.05, 95 matched sets of subjects will be needed, totaling 190 subjects.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	95	119	24
20%	71	89	18
20%	57	72	15
20%	33	42	9
20%	25	32	7
20%	20	25	5
20%	19	24	5
20%	14	18	4
20%	12	15	3
20%	13	17	4
20%	10	13	3
20%	8	10	2

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.

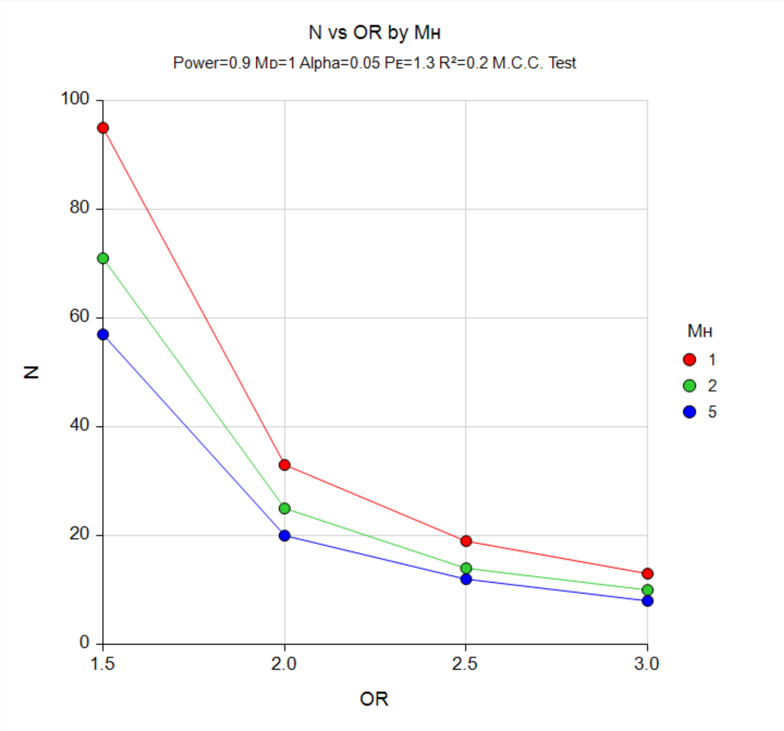
Dropout Summary Statements

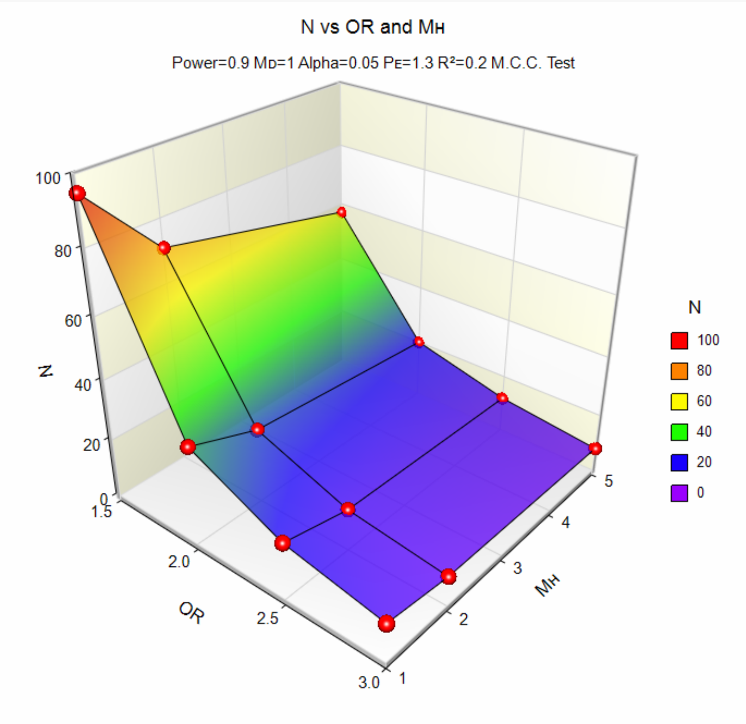
Anticipating a 20% dropout rate, 119 subjects should be enrolled to obtain a final sample size of 95 subjects.

This report shows the power for each of the scenarios.

Plots Section

Plots





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, σ_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

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 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

Solve For: Sample Size N
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
0.8501	119	1	2	1.4	1	0	0.05

PASS has also calculated N to be 119.