Chapter 156

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

Introduction

This procedure calculates the power and sample size necessary in a matched case-control study designed to detect a relationship between the development of a disease and a 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 binary risk factor (such as 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, some subjects are positive for a binary *exposure* variable of interest. 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).

Consider a binary exposure variable E which is set to '1' if the patient has been exposed to the risk factor of interest or to '0' if not. The probability that a case patient was exposed to the risk factor is P_{ED} and the probability that a control patient was exposed to the risk factor is P_{EH}. Furthermore, the probability that any patient has been exposed to the risk factor is P_E. Hence, P_E can be thought of as the population prevalence of exposure. If you don't have better information, you can use $P_E = (P_{ED} + P_{EH})/2$ or, if the disease is rare, $P_E = P_{EH}$.

We consider that a score test will be constructed from a conditional logistic regression to test hypotheses about the regression coefficient corresponding to the exposure variable which we will call θ (see Lachin (2008) for details). For reasonable sample sizes, this score test can be assumed to follow the standard normal distribution.

The regression coefficient θ of the exposure variable in a conditional logistic regression is interpreted as the log odds ratio. Note that in a case-control study, the odds ratio is defined as

$$OR = e^{\theta} = \frac{P_{ED}/(1 - P_{ED})}{P_{EH}/(1 - P_{EH})}$$

It important to note that the numerator in this expression gives the odds that case patient was exposed, and the denominator gives the odds that a control patient was exposed. It is useful to recognize that these are the odds that a subject has been exposed, not the odds that the subject has the disease. Hence, OR is the ratio of odds that a case subject has been exposed to the odds that a control subject has been exposed.

The power of the score test is calculated using

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

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

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

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 $P_E(1 - P_E)$ 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, P_E = 0.3, 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	**0.90**
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**
P_E (Probability of Exposure)	**0.3**
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	Probability of Risk Exposure P_E	Regression of Exposure on other Covariates R^2	Alpha
0.9000	761	1	1	1.5	0.3	0.2	0.05
0.9002	571	1	2	1.5	0.3	0.2	0.05
0.9003	457	1	5	1.5	0.3	0.2	0.05
0.9007	261	1	1	2.0	0.3	0.2	0.05
0.9011	196	1	2	2.0	0.3	0.2	0.05
0.9014	157	1	5	2.0	0.3	0.2	0.05
0.9000	149	1	1	2.5	0.3	0.2	0.05
0.9007	112	1	2	2.5	0.3	0.2	0.05
0.9019	90	1	5	2.5	0.3	0.2	0.05
0.9010	104	1	1	3.0	0.3	0.2	0.05
0.9010	78	1	2	3.0	0.3	0.2	0.05
0.9037	63	1	5	3.0	0.3	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 odds ratio of developing a disease associated with exposure to a certain risk factor.
P_E The probability of exposure to the risk factor in the overall population.
R^2 The R^2 that occurs when the exposure variable 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 binary 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 probability of exposure to the risk factor in the population is 0.3. 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, 761 matched sets of subjects will be needed, totaling 1522 subjects.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	761	952	191
20%	571	714	143
20%	457	572	115
20%	261	327	66
20%	196	245	49
20%	157	197	40
20%	149	187	38
20%	112	140	28
20%	90	113	23
20%	104	130	26
20%	78	98	20
20%	63	79	16

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, 952 subjects should be enrolled to obtain a final sample size of 761 subjects.

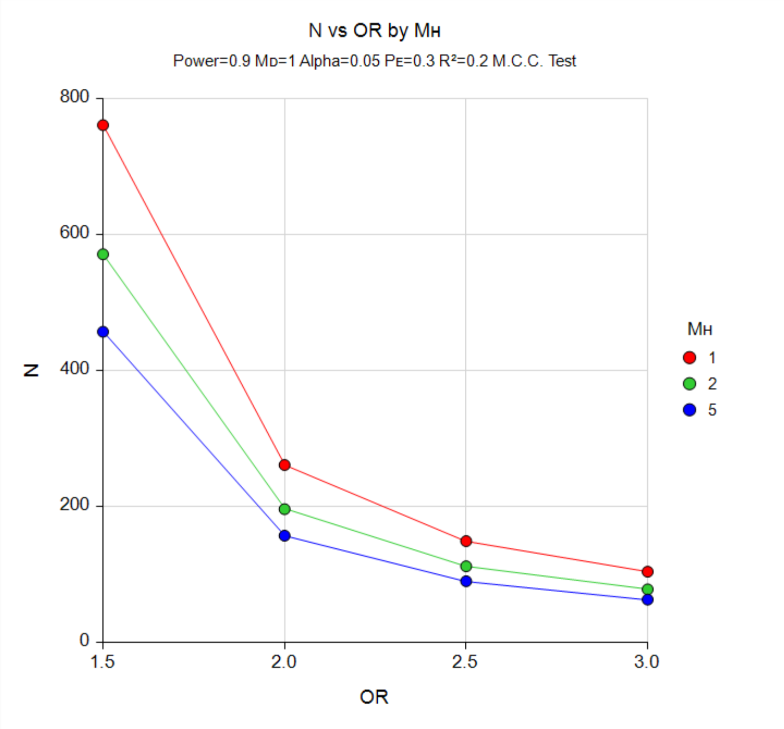
References

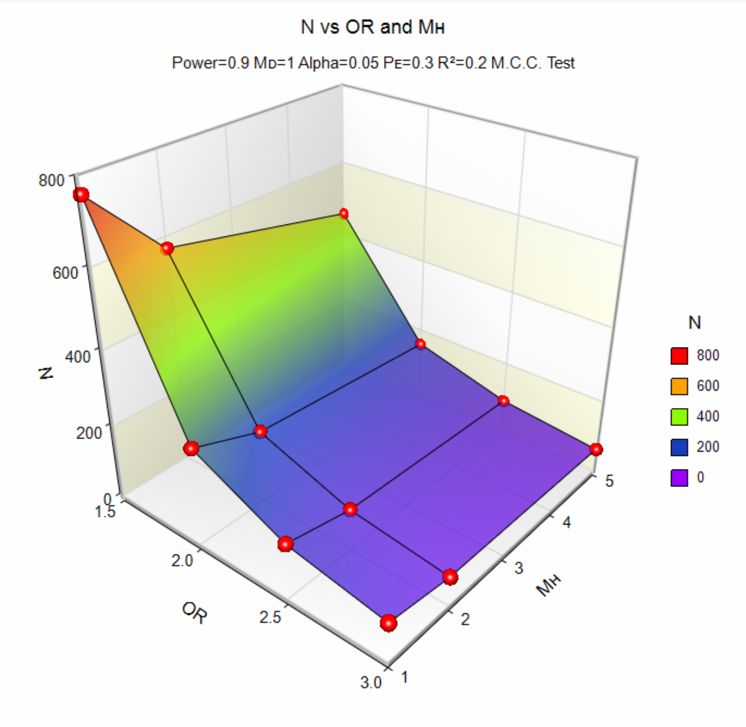
Lachin, John M. 2008. 'Sample size evaluation for a multiply matched case-control study using the score test from a conditional logistic (discrete Cox PH) regression model.' Statistics in Medicine, Volume 27, Pages 2509-2523.
Lachin, John M. 2011. Biostatistical Methods: The Assessment of Relative Risks, 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.

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 = 0.4444, P_E = 0.15, 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 161.

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) **0.4444**
P_E (Probability of Exposure)........................... **0.15**
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	Probability of Risk Exposure P_E	Regression of Exposure on other Covariates R^2	Alpha
0.8509	161	1	2	0.4444	0.15	0	0.05

PASS has also calculated N to be 161.