Chapter 851

# Mendelian Randomization with a Binary Outcome

## Introduction

This module computes the sample size and power of the causal effect in Mendelian randomization studies with a binary outcome. This analysis is often used in observational studies where clinical trials are not possible.

Analogous to randomized clinical trials (RCT), Mendelian randomization (MR) divides subjects into two or more groups. However, MR uses a genetic variable, such as the state of a certain gene, to form the groups. The state of the gene is assumed to be random. Using two-stage least squares and making several assumptions, the causal impact of an exposure variable on the outcome variable can be measured.

For further reading, we recommend the book by Burgess and Thompson (2015) which is completely devoted to this topic. We also recommend the paper by Burgess (2014).

## Technical Details

### Causal Relationship Test

The following details follow closely the results in Burgess and Thompson (2015). Assume that we are interested in assessing the causal relationship between a dichotomous *outcome* variable Y and an *exposure* variable X using a two-stage procedure. A genetic variable G is available to use as an *instrumental* variable. A sample of *n* will be selected. The basic models are

$$\ln(P(Y = 1)) = \beta_1 \hat{X} + e_Y$$

$$X = \beta_{XG} G + e_X$$

An estimate of $\beta_1$ will be obtained using a two-stage procedure. In the first stage, the regression of X on G is fit. This equation is used to create a predicted value of X for each data row. In the second stage, a logistic regression is fit in which the response is Y and the independent variable is the predicted value of X.

## Power Calculation

The power for a two-sided test is given by

$$\text{Power} = \Phi\left(\beta_1\sqrt{n\rho_{GX}^2 P(Y=1)P(Y=0)} - z_{1-\alpha/2}\right)$$

where $z_{1-\alpha/2}$ is the quantile from the standard normal distribution which has 1 - $\alpha$/2 to the left. Also, $\Phi(\mathrm{x})$ is the cumulative standard normal distribution.

# Example 1 – Finding the Sample Size

Researchers are planning an observation study to determine the causal effect of an exposure X on a binary outcome variable Y using an instrumental variable G. They want to determine the sample sizes necessary to have 80% power and 5% significance using a two-sided test when OR is 1.5 to 3.0 by 0.5, PC is 0.5, and $\rho^2(XG)$ is 0.01 or 0.02.

## 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** |
| Alternative Hypothesis | **Two-Sided** |
| Power | **0.80** |
| Alpha | **0.05** |
| Pc (Proportion of Y = 1) | **0.50** |
| OR (Odds Ratio of Y per Change in X) | **1.5 2 2.5 3** |
| $\rho^2(XG)$ | **0.01 0.02** |

# Output

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

## Numeric Reports

**Numeric Results**

Solve For: Sample Size
Alternative Hypothesis: Two-Sided

| Power | Sample Size N | Proportion of Subjects where Y = 1 Pc(Y = 1) | Odds Ratio of Y for One SD Change in X OR | R² of the Regression of X on G ρ²(XG) | Alpha |
|---|---|---|---|---|---|
| 0.8000 | 19097 | 0.5 | 1.5 | 0.01 | 0.05 |
| 0.8000 | 9549 | 0.5 | 1.5 | 0.02 | 0.05 |
| 0.8000 | 6535 | 0.5 | 2.0 | 0.01 | 0.05 |
| 0.8001 | 3268 | 0.5 | 2.0 | 0.02 | 0.05 |
| 0.8001 | 3740 | 0.5 | 2.5 | 0.01 | 0.05 |
| 0.8001 | 1870 | 0.5 | 2.5 | 0.02 | 0.05 |
| 0.8001 | 2602 | 0.5 | 3.0 | 0.01 | 0.05 |
| 0.8001 | 1301 | 0.5 | 3.0 | 0.02 | 0.05 |

| | |
|---|---|
| Y | The continuous outcome variable. |
| X | The exposure variable. |
| G | The genetic variant or instrumental variable used to divide the subjects up into groups. |
| Power | The probability of rejecting a false null hypothesis when the alternative hypothesis is true. |
| N | The sample size, the number of subjects in the study. |
| Pc(Y = 1) | The proportion of subjects in the study for which Y = 1. |
| OR | The parameter of interest. It is the odds ratio of Y for a one standard deviation change in X. |
| ρ²(XG) | The proportion of the variance of X explained by the regression of X on G. |
| Alpha | The probability of rejecting a true null hypothesis. |

**Summary Statements**

A Mendelian randomization study design with a binary outcome variable (Y) will be used to test whether there is a causal relationship between Y and an exposure variable (X), with a genetic variable (G) as an instrumental variable that divides the subjects into groups. The comparison will be made using a two-sided two-stage test involving a regression of X on G (first stage) and a logistic regression of Y on X (second stage). The Type I error rate (α) will be 0.05. The proportion of the variance of X explained by the regression of X on G is assumed to be 0.01. The anticipated proportion of subjects in the study where Y is equal to 1 (a case) is 0.5. To detect a causal effect (odds ratio of Y for a one standard deviation change in X) of 1.5 with 80% power, the number of needed subjects will be 19097.

**Dropout-Inflated Sample Size**

| Dropout Rate | Sample Size N | Dropout-Inflated Enrollment Sample Size N' | Expected Number of Dropouts D |
|---|---|---|---|
| 20% | 19097 | 23872 | 4775 |
| 20% | 9549 | 11937 | 2388 |
| 20% | 6535 | 8169 | 1634 |
| 20% | 3268 | 4085 | 817 |
| 20% | 3740 | 4675 | 935 |
| 20% | 1870 | 2338 | 468 |
| 20% | 2602 | 3253 | 651 |
| 20% | 1301 | 1627 | 326 |

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. 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. After solving for N, 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, 23872 subjects should be enrolled to obtain a final sample size of 19097 subjects.

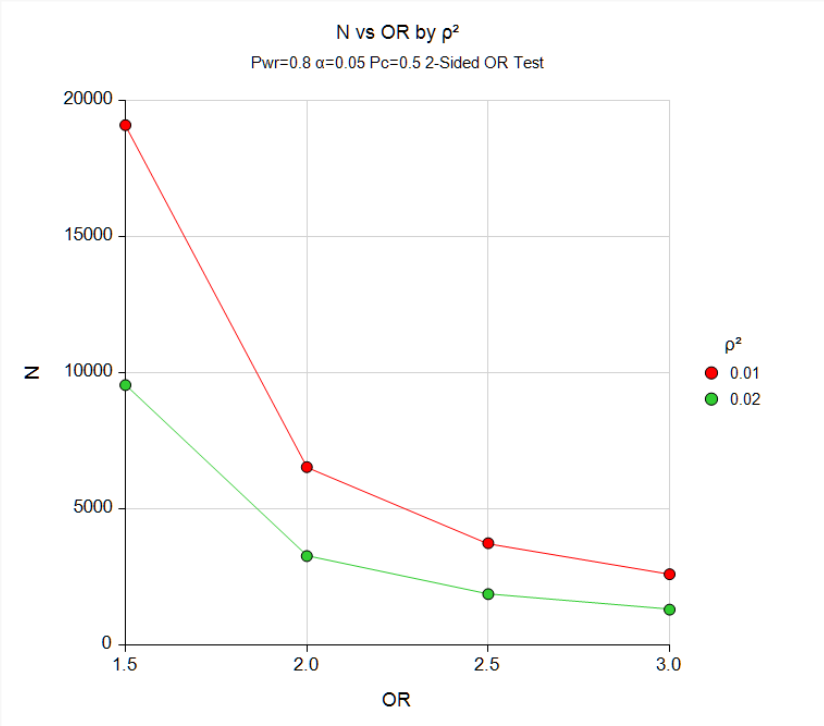
**References**

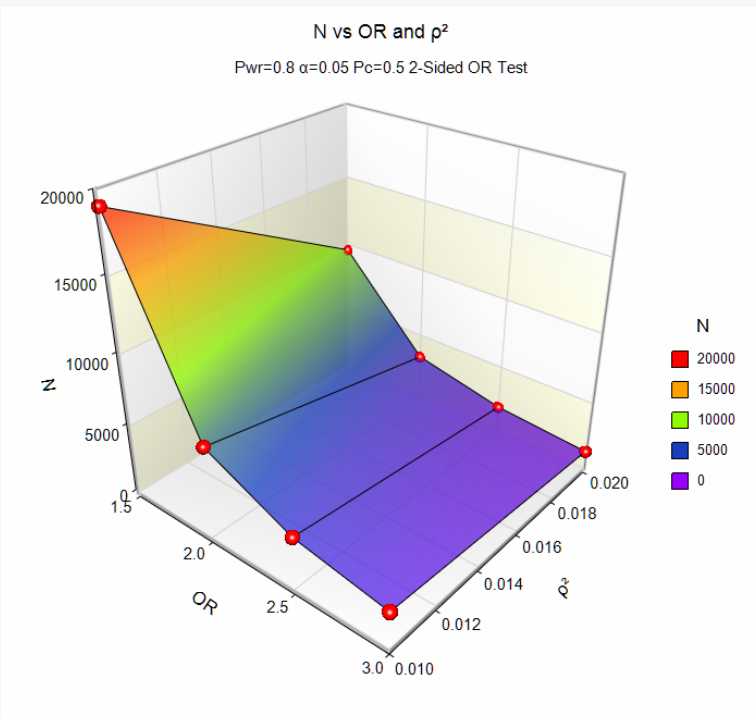
Burgess, S. and Thompson, S.G. 2015. Mendelian Randomization Methods for Using Genetic Variants in Causal Estimation. Chapman & Hall/CRC Press. New York.
Burgess, Stephen. 2014. 'Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome.' International Journal of Epidemiology, 43, pages 922-929.
Brion, M.J.A., Shakhbazov, K., Visscher, P.M. 2013. 'Calculating statistical power in Mendelian randomization studies.' International Journal of Epidemiology, 42, pages 1497-1501.

This report presents the calculated sample sizes for each scenario as well as the values of the other parameters.

## Plots Section

### Plots




These plots show the relationship among the varying parameters.

# Example 2 – Validation using Burgess and Thompson (2015)

Burgess and Thompson (2015) give an example in their book of a sample size calculation. However, they only present graphic results. We obtained numeric results using their online tool (https://sb452.shinyapps.io/power/). The parameter settings for this example where power is 0.80, alpha = 0.05, Pc = 0.50, OR = 1.5, and $\rho^2$(XG) is 0.03. They obtain N = 6400 (which was rounded).

## 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**
Alternative Hypothesis .................................**Two-Sided**
Power...........................................................**0.80**
Alpha...........................................................**0.05**
Pc (Proportion of Y = 1) ...............................**0.50**
OR (Odds Ratio of Y per Change in X)..........**1.5**
$\rho^2$(XG).........................................................**0.03**

## Output

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

**Numeric Results**

Solve For: Sample Size
Alternative Hypothesis: Two-Sided

| Power | Sample Size N | Proportion of Subjects where Y = 1 Pc(Y = 1) | Odds Ratio of Y for One SD Change in X OR | R² of the Regression of X on G ρ²(XG) | Alpha |
|---|---|---|---|---|---|
| 0.8 | 6366 | 0.5 | 1.5 | 0.03 | 0.05 |

**PASS** calculated N as 6366 which matches the 6400 (which was rounded).