

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 β_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 .

Mendelian Randomization with a Binary Outcome

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, $\Phi(x)$ is the cumulative standard normal distribution.

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. This chapter covers four procedures, each of which has different effect size options. However, many of the options are common to all four procedures. These common options will be displayed first, followed by the various effect size options.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power* and *Sample Size*.

Test

Alternative Hypothesis

Specify whether the statistical test is two-sided or one-sided.

- **Two-Sided**
This option gives the two-sided test.
- **One-Sided**
This option gives the approximate result for a one-sided test by making an appropriate substitution for α .

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis. Values must be between zero and one. Historically, the value of 0.80 was used for power. Now, 0.90 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

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when you reject the null hypothesis of equal survival curves when in fact the curves are equal.

Mendelian Randomization with a Binary Outcome

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for a two-sided test and 0.025 has been used for a one-sided test. 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.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Sample Size)

Enter a value for the sample size, N. This is the number of subjects in the study. You can enter one or more positive integers greater than or equal to 3. Note that in these type of studies, very large sample sizes (50000+) are often needed.

You may also enter a range such as “10000 to 50000 by 10000” or a list of values separated by commas or blanks such as “2000 4000 6000 8000.”

Pc (Proportion Y = 1)

Pc is the proportion of subjects in which $Y = 1$ in the study. If the design calls for an equal number of cases and controls, Pc is set to 0.5.

In this definition, we adopt the common convention that if an outcome (Y) is binary, the situation in which $Y = 1$ is called a case and the situation in which $Y = 0$ is called a control. Therefore, we might have used yes, positive, or abnormal for cases and no, negative, or normal for controls.

Range

$0 < Pc < 1$. Often, we plan to have $Pc = 0.5$.

Effect Size

OR (Odds Ratio of Y per Change in X)

Enter one or more values for the odds ratio, the parameter of interest in the study. This is the odds ratio of the binary variable Y per a change in X of one $\sigma(X)$.

A log-linear regression model is used to relate Y to X. The model is

$$\ln(OR) = \beta I(X) + \text{error term}$$

where

$$OR = \text{Prob}(Y = 1/X) / \text{Prob}(Y = 0/X).$$

OR must be a positive number greater than zero. Usually, it will be near one. Typical values are 1.5, 2, 3, 4, etc.

$\rho^2(XG)$

Enter one or more value for $\rho^2(XG)$, the proportion of the variance of X explained by the regression of X on G. In practice, it is estimated by the R^2 of this regression. It is routinely very small. Values of 0.01 and 0.02 are common. This value varies between zero and one. For mathematical reasons, it cannot be exactly 0.

Mendelian Randomization with a Binary Outcome

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Mendelian Randomization with a Binary Outcome** procedure window by expanding **Regression** and then clicking on **Mendelian Randomization with a Binary Outcome**. 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**.

<u>Option</u>	<u>Value</u>
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

Annotated Output

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

Numeric Results

Numeric Results for the Two-Sided, Mendelian Randomization Test of a Binary Outcome

Power	N	Pc(Y=1)	OR	$\rho^2(XG)$	Alpha
0.8000	19097	0.500	1.5000	0.0100	0.0500
0.8000	9549	0.500	1.5000	0.0200	0.0500
0.8000	6535	0.500	2.0000	0.0100	0.0500
0.8000	3267	0.500	2.0000	0.0200	0.0500
0.8000	3739	0.500	2.5000	0.0100	0.0500
0.8001	1870	0.500	2.5000	0.0200	0.0500
0.8001	2602	0.500	3.0000	0.0100	0.0500
0.8001	1301	0.500	3.0000	0.0200	0.0500

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.

Mendelian Randomization with a Binary Outcome

Report Definitions

Y is the continuous outcome variable.

X is the exposure variable.

G is the genetic variant or instrumental variable used to divide the subjects up into groups.

Power is the probability of rejecting a false null hypothesis.

N is the sample size, the number of subjects in the study.

OR (odds ratio) is the parameter of interest. It is the odds ratio of Y for a one standard deviation change in X.

$P_c(Y=1)$ is the proportion of subjects in the study for which $Y = 1$.

$\rho^2(XG)$ is the proportion of the variance of X explained by the regression of X on G.

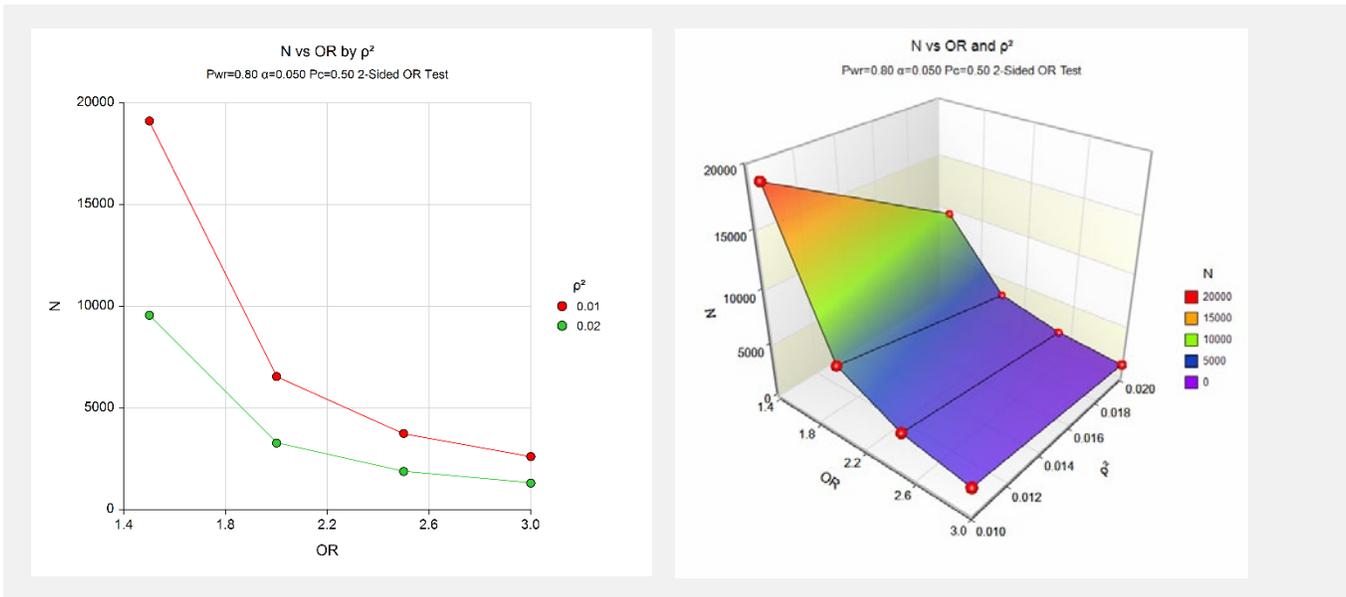
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements

A two-sided test in a Mendelian randomization study for a binary outcome variable that was based on a sample of 19097 subjects achieves 80.0% power at a 0.0500 significance level to detect a causal effect (odds ratio) of 1.5000 when the proportion of $Y = 1$ is 0.500. The proportion of the variance of X explained by its regression on G, the instrumental variable, is 0.0100.

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

Plots Section



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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Mendelian Randomization with a Binary Outcome** procedure window by expanding **Regression** and then clicking on **Mendelian Randomization with a Binary Outcome**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
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

Numeric Results for the Two-Sided, Mendelian Randomization Test of a Binary Outcome

Power	N	Pc(Y=1)	OR	$\rho^2(XG)$	Alpha
0.8000	6366	0.500	1.5000	0.0300	0.0500

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