Shiny Dashboard for Sample Size and Power Calculations

A PROJECT

SUBMITTED TO THE FACULTY

OF THE DIVISION OF HEALTH POLICY & MANAGEMENT

OF THE UNIVERSITY OF MINNESOTA

BY

Matthew F. Partridge

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Plan B Advisor: Julian Wolfson Ph.D.

November 2017

UNIVERSITY OF MINNESOTA

This is to certify that I have examined this copy of a MS Plan B Project by

Matthew F. Partridge

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

James Neaton Ph.D.

Faculty Adviser

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of Faculty Adviser

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date

DIVISION OF HEALTH POLICY & MANAGEMENT

Table of Contents

[0. Abstract 5](#_Toc496477478)

[1. Introduction 5](#_Toc496477479)

[1.1 R and RStudio 5](#_Toc496477480)

[1.2 Packages 5](#_Toc496477481)

[1.2.1 shiny 5](#_Toc496477482)

[1.2.2 shinydashboard 6](#_Toc496477483)

[1.2.3 pwr 6](#_Toc496477484)

[1.2.4 gsDesign 6](#_Toc496477485)

[1.2.5 survival 6](#_Toc496477486)

[2. Using the Application 6](#_Toc496477487)

[2.1 Launching the Application 6](#_Toc496477488)

[2.1.1 RStudio 6](#_Toc496477489)

[2.1.2 shinyapps.io 7](#_Toc496477490)

[2.2 Performing the Calculations 7](#_Toc496477491)

[2.2.1 Calculation Determination 7](#_Toc496477492)

[2.2.2 Dashboard Inputs and Outputs 7](#_Toc496477493)

[2.2.3 Real-Time Interactive Results 7](#_Toc496477494)

[3. Application Tabs 8](#_Toc496477495)

[3.1 One Sample Mean 8](#_Toc496477496)

[3.1.1 Statistical Explanation 8](#_Toc496477497)

[3.1.2 Functional Use 9](#_Toc496477498)

[3.1.3 Inputs (I) and Outputs (O) 9](#_Toc496477499)

[3.1.4 Examples 9](#_Toc496477500)

[3.2 One Sample Proportion 11](#_Toc496477501)

[3.2.1 Statistical Explanation 11](#_Toc496477502)

[3.2.2 Functional Use 13](#_Toc496477503)

[3.2.3 Inputs (I) and Outputs (O) 14](#_Toc496477504)

[3.2.4 Examples 14](#_Toc496477505)

[3.3 Two Sample Means 15](#_Toc496477506)

[3.3.1 Statistical Explanation 15](#_Toc496477507)

[3.3.2 Functional Use 17](#_Toc496477508)

[3.3.3 Inputs (I) and Outputs (O) 18](#_Toc496477509)

[3.3.4 Examples 18](#_Toc496477510)

[3.4 Two Sample Proportions 20](#_Toc496477511)

[3.4.1 Statistical Explanation 20](#_Toc496477512)

[3.4.2 Functional Use 22](#_Toc496477513)

[3.4.3 Inputs (I) and Outputs (O) 23](#_Toc496477514)

[3.4.4 Examples 23](#_Toc496477515)

[3.5 Time to Event 25](#_Toc496477516)

[3.5.1 Statistical Explanation 25](#_Toc496477517)

[3.5.2 Functional Use 28](#_Toc496477518)

[3.5.3 Inputs and Outputs 30](#_Toc496477519)

[3.5.4 Examples 30](#_Toc496477520)

[4. Technical Challenges and Future Improvements 34](#_Toc496477521)

[4.1 User Interface 34](#_Toc496477522)

[4.1.1 Layout 34](#_Toc496477523)

[4.1.2 Calculation Selector 34](#_Toc496477524)

[4.1.3 Slider Units 34](#_Toc496477525)

[4.1.4 Educational Dialog 34](#_Toc496477526)

[4.2 Server 35](#_Toc496477527)

[4.2.1 Functional Programming 35](#_Toc496477528)

[4.2.2 Consolidation 35](#_Toc496477529)

[4.3 Alternative Hypothesis 35](#_Toc496477530)

[4.4 Time to Event 35](#_Toc496477531)

[4.5 New Pages 35](#_Toc496477532)

[5. Acknowledgements 36](#_Toc496477533)

[6. References 37](#_Toc496477534)

# 0. Abstract

Shiny Dashboard for Sample Size and Power Calculations  
Presented by Matthew F. Partridge  
Masters Candidate in Biostatistics  
Plan B Adviser: Julian Wolfson, Ph.D.

Calculating sample sizes and power are vital aspects in the creation and analysis of clinical trials and scientific studies. There are various calculators available at various price points, however very few offer an interactive experience. This Shiny Dashboard attempted to create an interactive sample size and power calculator allowing the user to explore and better understand the relationships between variables through interactivity. R and RStudio were used as the framework to perform the statistical calculations and develop the dashboard. Calculations for the one sample mean, one sample proportion, two sample means, two sample proportions, and time to event scenarios are performed using functions within the "pwr" and "gsDesign" packages. The functions within these packages use the the t-Test, z-Test, or comparison of proportional hazards as a basis for the statistical calculations. The resulting Shiny Dashboard incorporates interactive sliders and text boxes, a dynamic user interface, and robust statistical concepts to create a tool that is both functional and intuitive.

# 1. Introduction

The purpose of this Shiny Dashboard is to create an interactive sample size and power calculation tool that duplicates some of the functionality found in the Piface Java applet created by Russell V. Lenth (2006). As modern computers have continued to support updated versions of Java, the Piface application has not been updated to use newer versions of Java, and now has compatibility issues with most systems. Shiny Dashboards will provide a modernized interface for this updated application. The ability to use interactive sliders and text boxes to enter information allows the user to see in real time how certain inputs (e.g. mean, proportion, significance level, power, etc.) affect outputs (e.g. required sample size or power). This allows the user to explore the relationships among variables and ultimately understand the calculations better.

## 1.1 R and RStudio

This Shiny Dashboard is programmed using R version 3.4.0 (R Core Team, 2017) and RStudio version 1.0.143 (RStudio Team, 2016). In addition to base R, five other R packages, described below, are used for dashboard setup, statistical calculations, and examples.

## 1.2 Packages

### 1.2.1 [shiny](https://cran.r-project.org/web/packages/shiny/shiny.pdf)

The shiny package is a tool to assist users in building the foundations of an interactive web application using R (Chang, Cheng, Allaire, Xie, and McPherson, 2017). It uses R functions to create HTML co de for a web page. There are various inputs, displays, and settings options that can be customized all within the same Shiny application.

### 1.2.2 [shinydashboard](https://cran.r-project.org/web/packages/shinydashboard/shinydashboard.pdf)

The shinydashboard package expands beyond the shiny package to allow a compilation of many Shiny pages in one dashboard. It also adds visual themes as well as other aesthetic options to give the dashboard a more attractive look (Chang, 2016).

### 1.2.3 [pwr](https://cran.rstudio.com/web/packages/pwr/pwr.pdf)

The pwr package calculates sample sizes and power for various scenarios using the calculations found in the book *Statistical Power Analysis for the Behavioral Sciences* (Cohen, 1980) as a basis (Champley, 2017). In this dashboard, the pwr package is used for the calculations of the One Sample Mean, One Sample Proportion, Two Sample Means, and Two Sample Proportions scenarios.

### 1.2.4 [gsDesign](https://cran.r-project.org/web/packages/gsDesign/gsDesign.pdf)

The gsDesign package is used for sample size and power calculations for time to event scenarios (Anderson, 2016). Specifically, it incorporates factors targeting the time, recruitment, and censoring components that are relevant for time to event settings. It uses the calculations of Lachin and Foulkes (1986) and Schoenfeld (1981) as the basis for the nSurvival and nEvents functions, which calculate the required sample size and the expected number of events respectively.

### 1.2.5 [survival](http://cran.irsn.fr/web/packages/survival/survival.pdf)

The example sections in this document use the flchain data set from the survival package. It contains a stratified sample of half of the data collected from a study that examined "the prevalence of monoclonal gammopathy of undetermined significance (MGUS) in Olmsted County, Minnesota" (Therneau, 2015). This subset of willing participants was assayed to collect Kappa, , and Lambda, serum free light chain (FLC) levels in the blood among a few other variables. Assaying the blood examines immunoglobulins made up of heavy chains. Light chains bind to the heavy chains leaving any extra light chains to be considered "free". These free light chains are classified as either or . The levels of and FLC are expressed in units of milligrams per liter (mg/L) and are often presented as the Kappa/Lambda ratio, (Turley, 2017). In the time between the original prevalence study and the secondary assay, some work had been done suggesting serum FLC levels were associated with immune disregulation. The secondary assay collected data necessary to examine the association of FLC levels and death rates.

# 2. Using the Application

## 2.1 Launching the Application

The first step in using the application is to launch it by one of two possible ways.

### 2.1.1 RStudio

The application can be launched by downloading the R code from GitHub (<https://github.com/mattpartridge/ShinySampleSizes>) and executing the ui.R and server.R files in RStudio.

### 2.1.2 shinyapps.io

The application can also be used by navigating to the website version of the application hosted on [shinyapps.io](https://www.shinyapps.io/) at <https://mfpartridge.shinyapps.io/shinysamplesizesdashboard/>.

## 2.2 Performing the Calculations

### 2.2.1 Calculation Determination

The first step in performing the calculations is to determine the appropriate scenario found on the sidebar of the dashboard. After the scenario determination, the desired calculation must be selected using the selection box at the top of the dashboard. The One Sample Mean and One Sample Proportion scenarios can calculate sample size, power, or margin of error while the Two Sample Means, Two Sample Proportions, and Time to Event scenarios can calculate sample size and power. The rest of the dashboard will then update automatically based on the scenario and calculation selections.

### 2.2.2 Dashboard Inputs and Outputs

The layout of the dashboard for each page is very similar. The One Sample Mean, One Sample Proportion, Two Sample Means, and Two Sample Proportions scenarios are all set up with study information on the left-hand panel and calculation information, as well as the calculation itself, on the right-hand panel. The Time to Event scenario is slightly different in that the study information is on the left-hand panel, population information is on the middle panel, and the calculation information, and the calculation itself, is on the right-hand panel. For all of the pages, the final calculation is outlined in green as seen below.



### 2.2.3 Real-Time Interactive Results

On each page, there are numerical and, in the case of the Time to Event scenario, categorical inputs. There are many other sample size and power calculation tools available online or as functions in programming languages, however, the majority of these tools require the user to submit a set of inputs to calculate an output. Then, if the user wants to perform a new calculation, a new set of inputs must be submitted making it difficult to see the direct effect of a change in inputs. This application allows the user to update the set of inputs and see the resulting output changes in real time. This helps the user make clear comparisons between differences in inputs.

# 3. Application Tabs

## 3.1 One Sample Mean

### 3.1.1 Statistical Explanation

The One Sample Mean scenario looks to compare a population mean, , to a fixed reference value, in a two-sided comparison. The null and alternative hypotheses are and respectively. In order to compare the two values, the Student's -Test, as shown below, can be used.

In the equation above, is defined as previously stated. is the sample mean, is the sample standard deviation, and is the size of the sample from the population. Under the null hypothesis, follows a -distribution with degrees of freedom.

**Power:** The power of a study is defined as the probability of correctly rejecting the null hypothesis and can be expressed as the probability that the -statistic, above, is more extreme than the -distribution critical value for a given significance level and degrees of freedom as shown below.

is the true mean of the sampled population, denotes the -distribution quantile function with a significance level of and degrees of freedom, and denotes the central -distribution function.

**Sample Size:** The sample size can be defined as the closest integer value of which, for a given power, , satisfies the equation below. It is typically found using a numerical search method.

**Margin of Error:** The margin of error of an estimate is defined as the amount of random sampling error found in an estimate. It is expressed as for a given significance level, . A smaller margin of error implies a more precise estimate and a larger margin of error implies a less precise estimate. The margin of error is typically expressed as a confidence interval surrounding the estimate from the sample. The confidence interval is defined as . Thus, the confidence interval is as follows:

### 3.1.2 Functional Use

The One Sample Mean scenario uses the pwr.t.test function within the pwr package to perform the power and sample size calculations. The function takes a set of variables as inputs, one of which being NULL, and calculates the NULL variable. The variables below are the relevant inputs for the pwr.t.test function.

* n - Number of observations (in the sample)
* d - Effect size
* sig.level - Significance level
* power - Power of the test
* alternative - Alternative hypothesis (, , )

The sample size, effect size, power, and significance level can all be calculated by the function. In the case of this scenario, only the sample size or power will be calculated in a two-sided, , alternative hypothesis. When solving for the power, pwr.t.test uses the pt function of the -distribution to calculate the power of the given variables. When solving for the sample size, it uses the uniroot function to perform a binary search for the value of which equates the left and right sides of the equation found in section *3.1.1 Statistical Explanation: Sample Size* above.

### 3.1.3 Inputs (I) and Outputs (O)

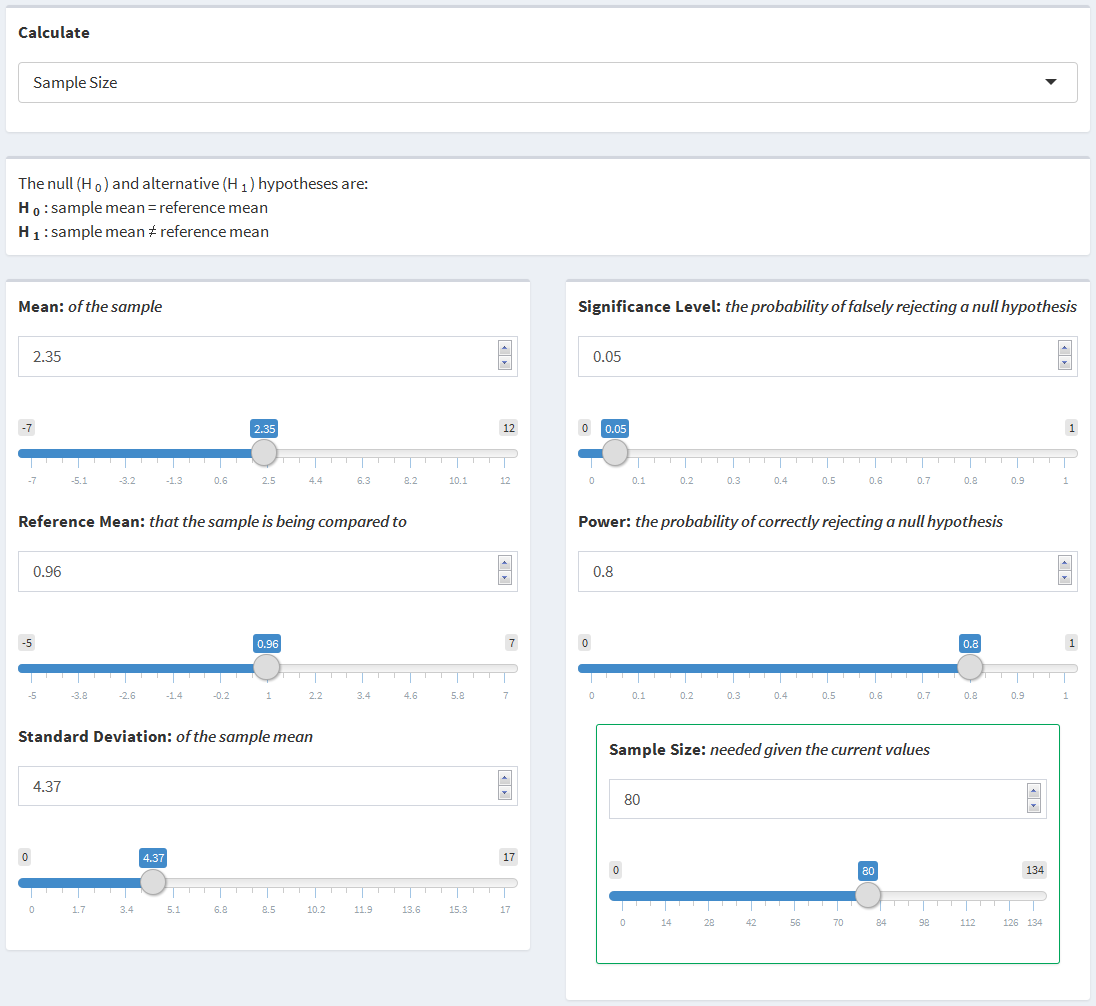
* Mean (I) - The mean of the target sample
* Reference Mean (I) - The mean of the reference population
* Standard Deviation (I) - The standard deviation of the measure for the target sample
* Sample Size (I/O) - The size of the target sample
* Significance Level (I) - The probability of falsely rejecting a null hypothesis
* Power (I/O) - The probability of correctly rejecting a null hypothesis
* Margin of Error (O) - The amount of sampling error found in the sample mean
* Confidence Interval (O) - The interval within which the mean is likely to fall

### 3.1.4 Examples

**Calculating Sample Size**: The flchain data set assayed a set of already-sampled citizens of Olmsted County, Minnesota from a previous study to examine Kappa/Lambda, , rates. The data suggests that there may be a difference in rates for participants with and without monoclonal gammopathy of undetermined significance, MGUS. Suppose a new study is being designed to specifically compare rates across this stratification by sampling participants with MGUS and comparing the average rate to a reference value. A previous study suggested the overall average rate is between 0.26 and 1.65 (Singh, 2017), say 0.96, and using the flchain data set as a pilot of sorts, an expected rate for participants with MGUS is 2.35 with a standard deviation of 4.37. Using a significance level of 0.05 with a desired 80% power, the required size of the sample of participants with MGUS can be calculated as follows:

* Mean: 2.35
* Reference Mean: 0.96
* Standard Deviation: 4.37
* Significance Level: 0.05
* Power: 0.8

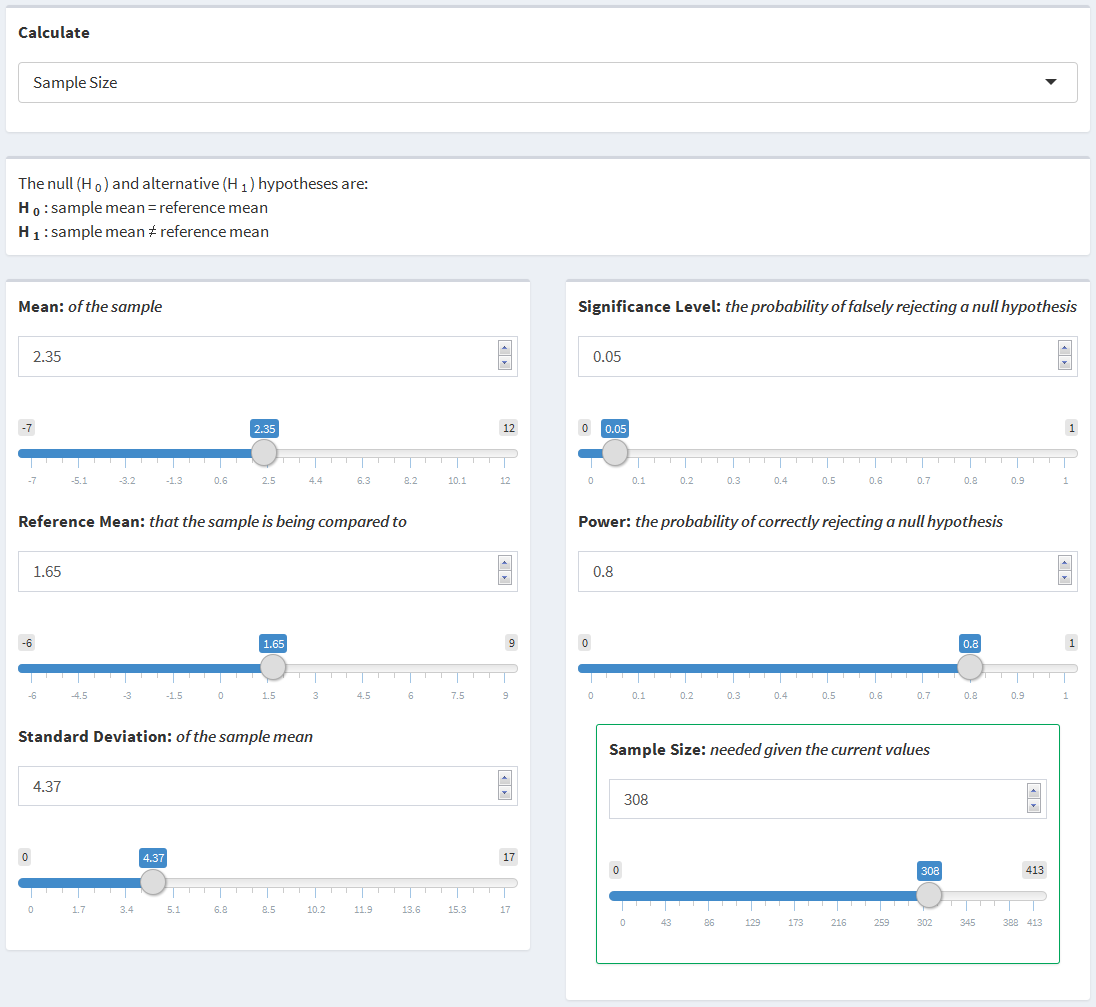
Based on the values entered, a sample size of 80 participants would be required.



Now, suppose instead of using the middle of the normal range of rates, the size of the sample required using the largest normal value, i.e. the smallest difference, is to be determined. The reference value can be changed from 0.96 to 1.65 keeping the rest of the inputs as they are.

* Mean: 2.35
* Reference Mean: Adjust from 0.96 to 1.65
* Standard Deviation: 4.37
* Significance Level: 0.05
* Power: 0.8

The required sample size has now increased from 80 to 308 participants.



## 3.2 One Sample Proportion

### 3.2.1 Statistical Explanation

The One Sample Proportion scenario looks to compare a population proportion, , to a fixed reference proportion, , in a two-sided comparison. In comparing the two values, it follows that the null and alternative hypotheses are and respectively. In order to compare the two proportions, the -Test can be used as follows.

**Power:** The power in this scenario can be expressed as the probability that the -score, above, is more extreme than the standard normal critical value for a given significance level:

where is the sample proportion, is the reference proportion, is the true proportion of the sampled population, denotes the standard normal quantile function with significance level , is the size of the sample, and denotes the standard normal distribution function.

**Sample Size:** The sample size can be defined as the integer of which, for a given power, , satisfies the following equation:

**Margin of Error:** The margin of error of an estimate is defined as the amount of random sampling error found in an estimate. It is expressed as for a given significance level, . The confidence interval is defined as , where is the sample proportion as defined previously. Thus, the confidence interval is as follows:

### 3.2.2 Functional Use

The One Sample Proportion scenario uses the pwr.p.test function within the pwr package to perform the power and sample size calculations. The function takes a set of variables as inputs, one of which being NULL, and calculates the NULL variable. The variables below are the inputs for the pwr.t.test function.

* h - Effect size
* n - Number of observations (in the sample)
* sig.level - Significance level
* power - Power of the test
* alternative - Alternative hypothesis (, , )

The sample size, effect size, power, and significance level can all be calculated by the function. It should be noted that the effect size uses a nonlinear transformation of the proportion. Since the difference between the sample and reference proportions can be the same at different points between zero and one, e.g. 0.7 - 0.4 = 0.5 - 0.2, but the power to detect the difference can be different, the nonlinear transformation, , allows for equal differences to have equal detectability (Cohen, 1988). In the case of this scenario, only the sample size or power will be calculated in a two-sided ,, alternative hypothesis. When solving for the power, pwr.p.test uses the pnorm function of the normal distribution to calculate the power of the given variables. When solving for the sample size, it uses the uniroot function to perform a binary search for the value of which equates the left and right sides of the equation found in section *3.2.1 Statistical Explanation: Sample Size* above.

### 3.2.3 Inputs (I) and Outputs (O)

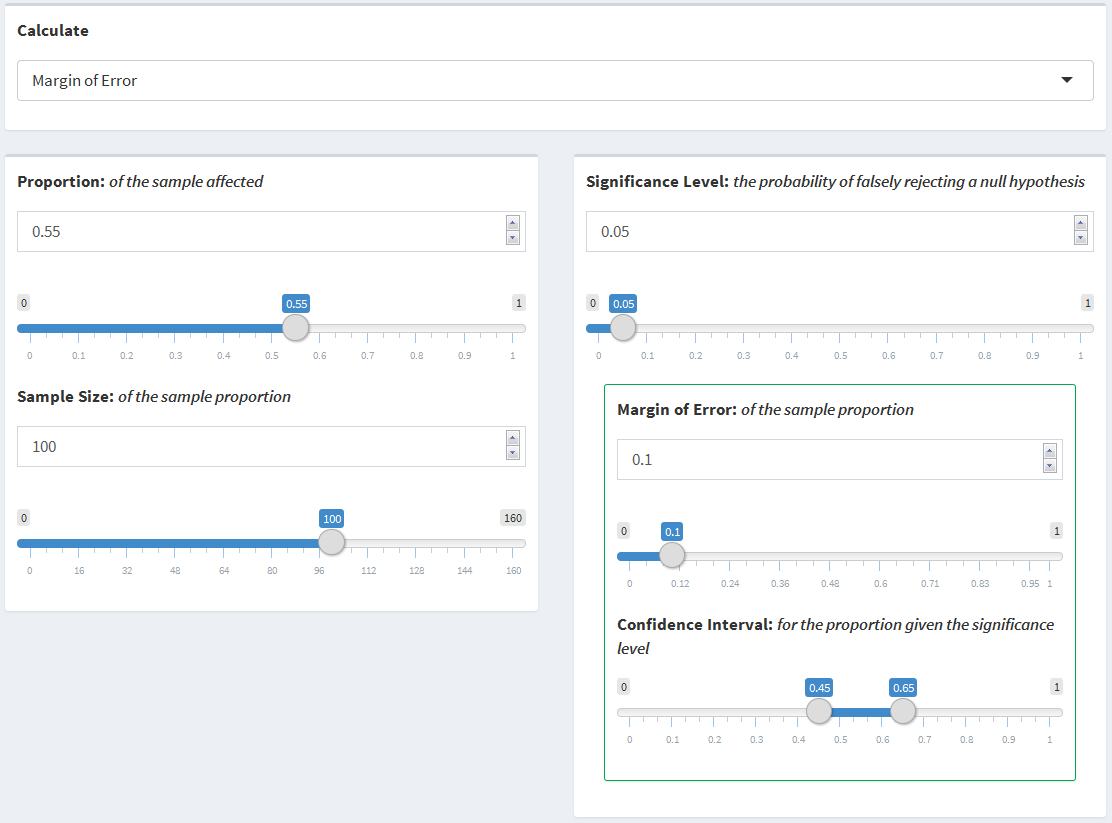
* Proportion (I) - The proportion affected in the target sample
* Reference Proportion (I) - The proportion affected in the reference population
* Sample Size (I/O) - The size of the target sample
* Significance Level (I) - The probability of falsely rejecting a null hypothesis
* Power (I/O) - The probability of correctly rejecting a null hypothesis
* Margin of Error (O) - The amount of sampling error found in the sample proportion
* Confidence Interval (O) - The interval in which the proportion is likely to fall within

### 3.2.4 Examples

**Solving for Margin of Error**: Suppose a simple study is being designed to determine the proportion of citizens in Olmsted County that are female. Specifically, the sample size required to obtain a margin of error of 0.1 on the proportions with a significance level of 0.05 is desired. The flchain data suggests that women make up 55% of Olmsted County. Using that as an estimate, the sample size can be manipulated to determine the required sample size to obtain a margin of error of 0.1.

* Proportion: 0.55
* Sample Size: 100
* Significance Level: 0.05

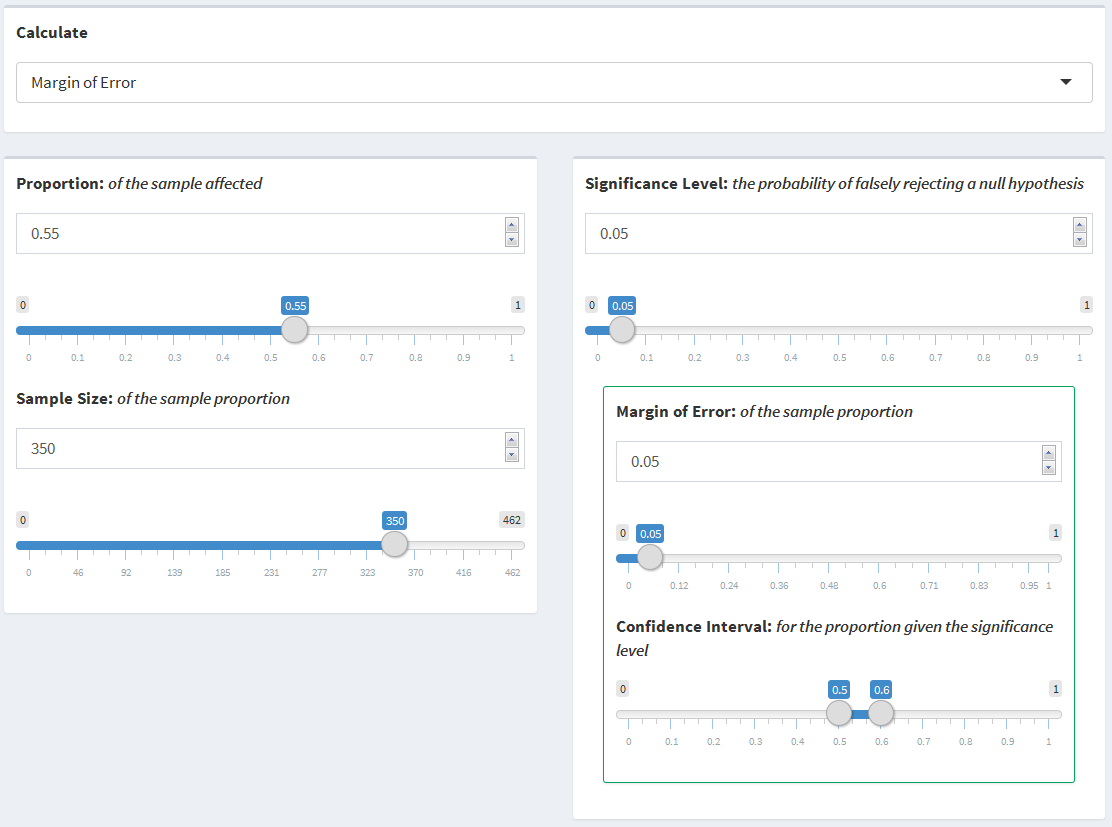
A sample size of roughly 100 participants is required to obtain a margin of error of 0.1.



Suppose a margin of error of 0.05 is desired instead of 0.1. The Sample Size input can be increased from 100 to 350 in order to obtain the desired margin of error.

* Proportion: 0.55
* Sample Size: Adjust from 100 to 350
* Significance Level: 0.05

A sample size of roughly 350 participants is required to obtain a margin of error of 0.05.



## 3.3 Two Sample Means

### 3.3.1 Statistical Explanation

The Two Sample Means scenario looks to compare two population means, and , in a two-sided comparison. The most common way of comparing the populations means is to compare the difference of the two means to a reference value. That is, and are the null and alternative hypothesis respectively, where is the mean from the first population, is the mean from the second population, and is the reference value. Often, the reference value is set to zero to test if the two population means are equal. In order to compare the difference in the population means to the reference value, the Student's -Test, as shown below, can be used.

is defined as the mean from the sample of the first population, is defined as the mean from the sample of the second population, is the size of the first sample, is the size of the second sample, and is defined as the pooled standard deviation for both samples assuming the variance within each population is equal. Variances could alternatively be assumed unequal, though the power and sample size calculations are much more difficult for that case. Under the null hypothesis, has a -distribution with degrees of freedom.

**Power:** The power can then be expressed as the probability that the -statistic, above, is more extreme than the -distribution critical value for a given significance level.

, , , , , and are all defined as above. is the true mean of population one, denotes the -distribution quantile function with a significance level and degrees of freedom, and denotes the central -distribution function.

**Sample Size:** Since it is the case that multiple combinations of and can equate to the same combined sample size, one of the sample sizes must be specified in order to solve for the other. In this case, it will be assumed that is specified and is being calculated. The size of the sample from population two can be defined as the integer of which, for a given power, , and satisfies the following equation.

### 3.3.2 Functional Use

The Two Sample Means scenario uses the pwr.t2n.test function within the pwr package to perform the power and sample size calculations. The function takes a set of variables as inputs, one of which being NULL, and calculates the NULL variable. The variables below are the inputs for the pwr.t2n.test function.

* n1 - Number of observations in the first sample
* n2 - Number of observations in the second sample
* d - Effect size
* sig.level - Significance level
* power - Power of the test
* alternative - Alternative hypothesis (, , )

The effect size, size of sample one, size of sample two, power, and significance level can all be calculated by the function. In the case of this scenario, only the size of the second sample or power can be calculated with a two-sided, , alternative hypothesis. When solving for the power, pwr.t2n.test uses the pt function of the -distribution to calculate the power of the given variables. When solving for the size of the second sample, it uses the uniroot function to perform a binary search for the value of which equates the left and right sides of the equation found in section *3.3.1 Statistical Explanation: Sample Size* above.

### 3.3.3 Inputs (I) and Outputs (O)

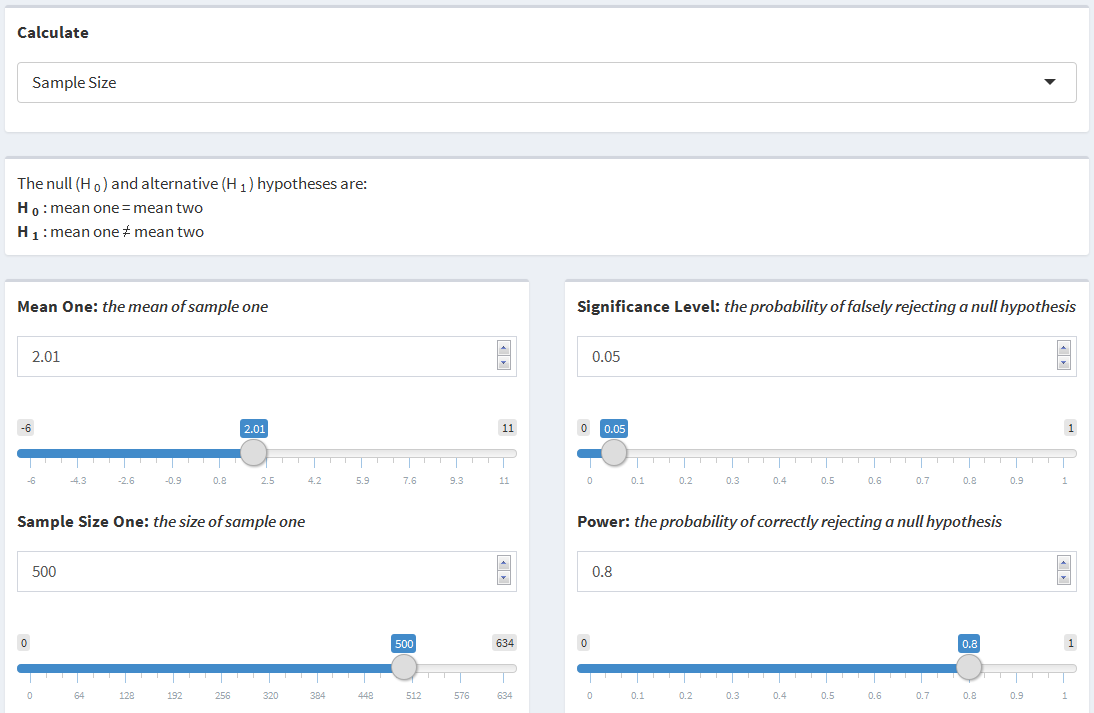
* Mean One (I) - The mean of the first sample
* Sample Size One (I) - The size of the first sample
* Mean Two (I) - The mean of the second sample
* Sample Size Two (I/O) - The size of the second sample
* Standard Deviation (I) - The standard deviation of the measure for the target sample
* Significance Level (I) - The probability of falsely rejecting a null hypothesis
* Power (I/O) - The probability of correctly rejecting a null hypothesis

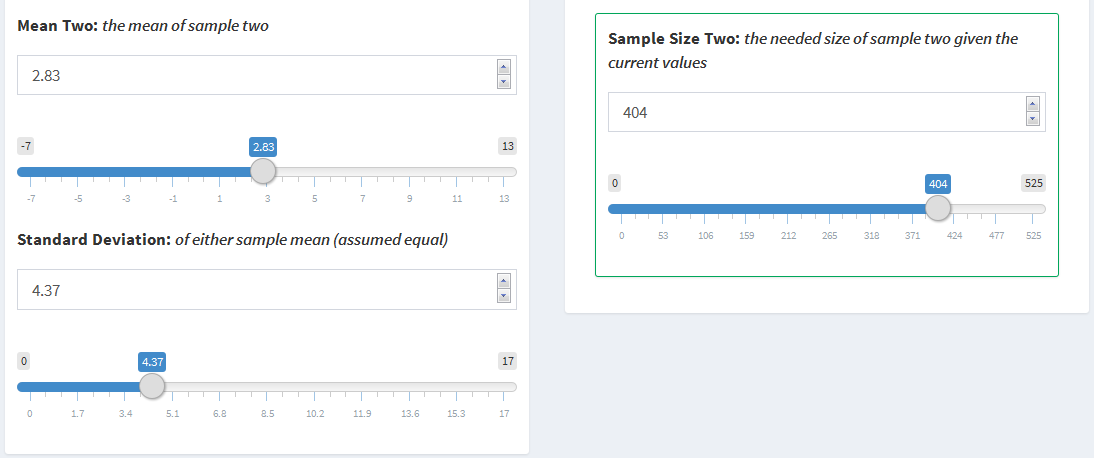
### 3.3.4 Examples

**Calculating Sample Size**: The flchain data suggests that there may be a difference in rates between sexes for participants that have MGUS. Suppose a new study is being designed to specifically compare rates of female and male participants diagnosed with MGUS. The rates from the flchain data can be used as expected rates, 2.01 for female participants and 2.83 for male participants with a standard deviation of 4.37. Suppose there will be 500 female participants in this new study and the number of male participants required needs to be calculated. The size of the sample of male participants with MGUS can be calculated with a significance level of 0.05 and 80% power as follows:

* Mean One: 2.01
* Sample Size One: 500
* Mean Two: 2.83
* Standard Deviation: 4.37
* Significance Level: 0.05
* Power: 0.8

Based on the values entered, a sample size of 404 male participants would be required.

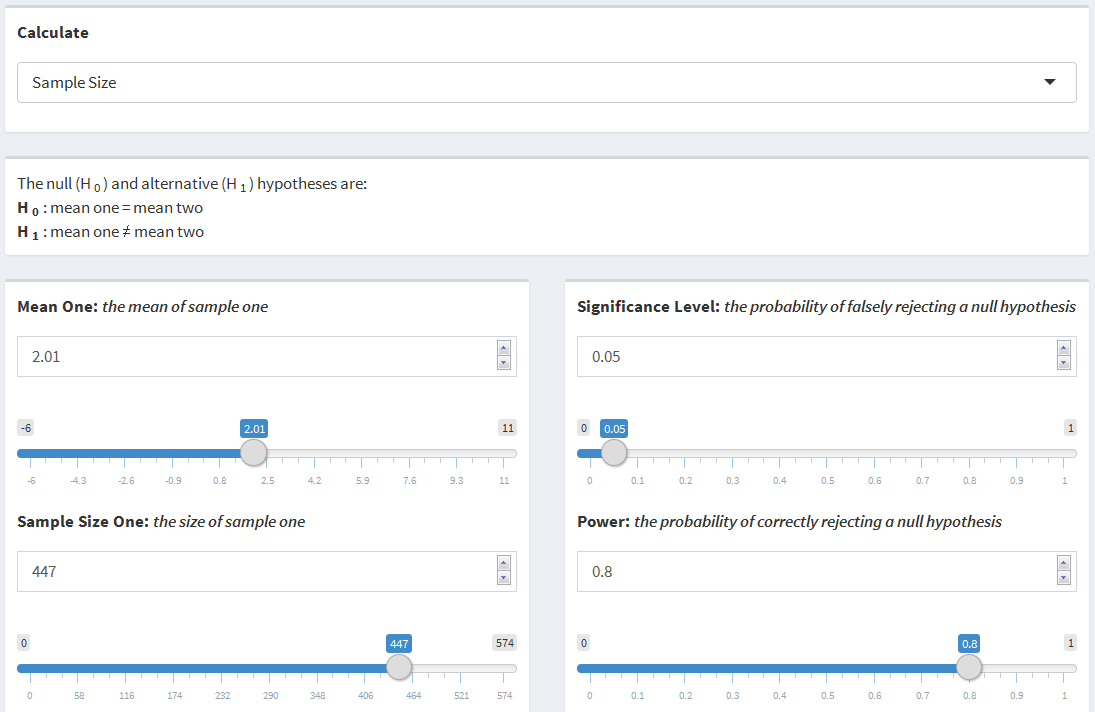


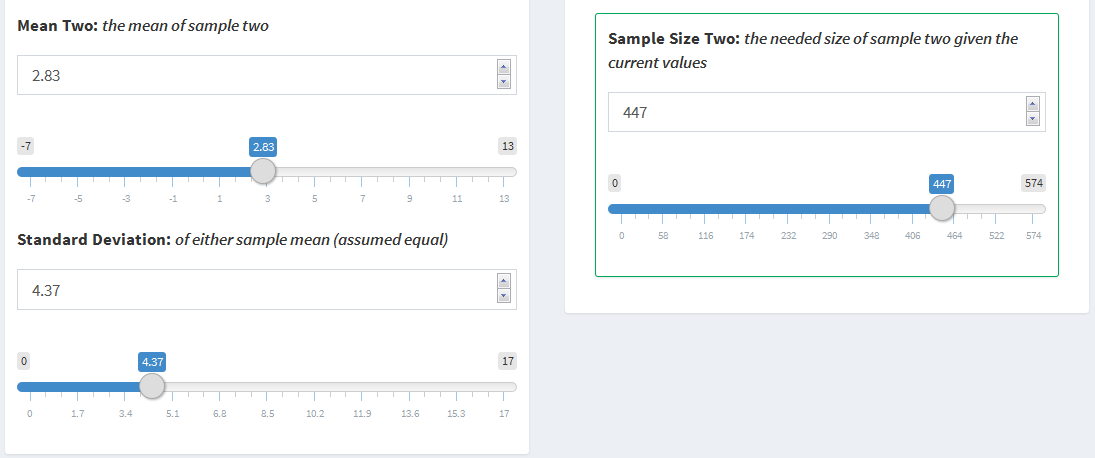


Suppose that instead of knowing that there will be 500 female participants and wanting to calculate the number of male participants required, the study will be designed to have a 1:1 sampling ratio of females to males. Using the arrow buttons or adjusting the slider, the Sample Size One input can be adjusted to find the required size for which both samples are the same. In this case, the number of female participants can be adjusted down from 500, where the number of male participants required would be 404, to 447, where the same number of male participant would be required.

* Mean One: 2.01
* Sample Size One: Adjust from 500 to 447
* Mean Two: 2.83
* Standard Deviation: 4.37
* Significance Level: 0.05
* Power: 0.8

As the Sample Size One input decreases, the Sample Size Two output will increase. At the point the Sample Size One input is at 447 participants, the Sample Size Two output will also be at 447 participants.





## 3.4 Two Sample Proportions

### 3.4.1 Statistical Explanation

The Two Sample Proportions scenario looks to compare two population proportions, and , in a two-sided comparison. The null and alternative hypotheses are and respectively, where is the proportion affected from population one and is the proportion affected from population two. In order to compare the two proportions, the -Test can be used as follows.

is the proportion affected from the sample of population one, is the population affected from the sample of population two, is the size of the first sample, and is the size of the second sample.

**Power:** The power can be expressed as the probability that the -statistic, above, is more extreme than the standard normal distribution critical value for a given significance level as shown below.

, , , and are all defined as above. is the true proportion affected from population one, is the significance level, denotes the standard normal quantile function, and denotes the standard normal distribution function.

**Sample Size:** The size of the sample from population two can be defined as the integer of which, for a given power, , and , satisfies the following equation.

### 3.4.2 Functional Use

The Two Sample Proportions scenario uses the pwr.2p2n.test function within the pwr package to perform the power and sample size calculations. The function takes a set of variables as inputs, one of which being NULL, and calculates the NULL variable. The variables below are the inputs for the pwr.2p2n.test function.

* h - Effect size
* n1 - Number of observations in the first sample
* n2 - Number of observations in the second sample
* sig.level - Significance level
* power - Power of the test
* alternative - Alternative hypothesis (, , )

The effect size, size of sample one, size of sample two, power, and significance level can all be calculated by the function. It should be noted that the effect size uses a nonlinear transformation of the proportion. Since the difference between two proportions can be the same at different points between zero and one, e.g. 0.7 - 0.4 = 0.5 - 0.2, but the power to detect the difference can be different, the nonlinear transformation, , allows for equal differences to have equal detectability (Cohen, 1988). In the case of this scenario, only the size of the second sample or the power can be calculated with a two-sided, , alternative hypothesis. When solving for the power, pwr.2p2n.test uses the pnorm function of the normal distribution to calculate the power of the given variables. When solving for the size of the second sample, it uses the uniroot function to perform a binary search for the value of which equates the left and right sides of the equation found in section *3.4.1 Statistical Explanation: Sample Size* above.

### 3.4.3 Inputs (I) and Outputs (O)

* Proportion One (I) - The proportion affected in sample one
* Sample Size One (I) - The size of sample one
* Proportion Two (I) - The proportion affected in sample two
* Sample Size Two (I/O) - The size of sample two
* Significance Level (I) - The probability of falsely rejecting a null hypothesis
* Power (I/O) - The probability of correctly rejecting a null hypothesis

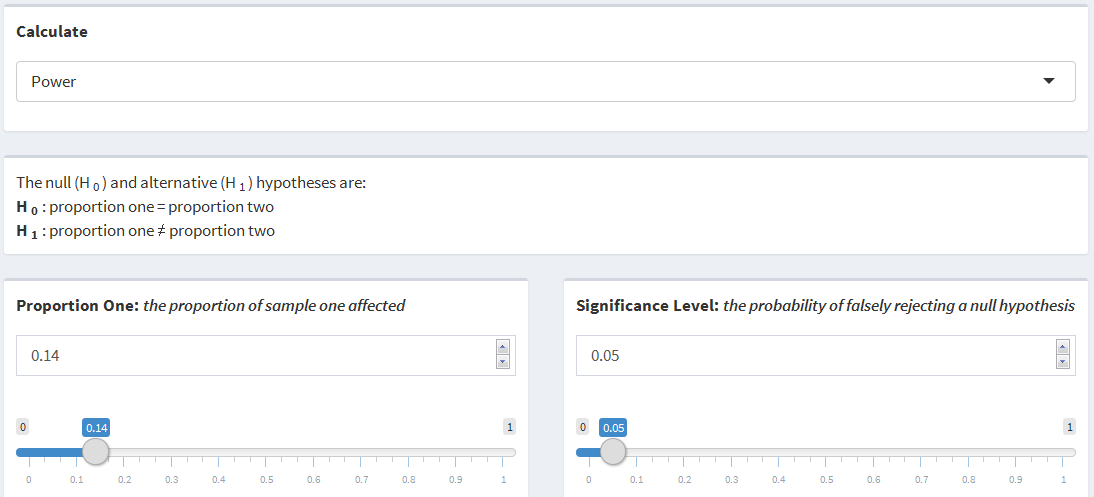
### 3.4.4 Examples

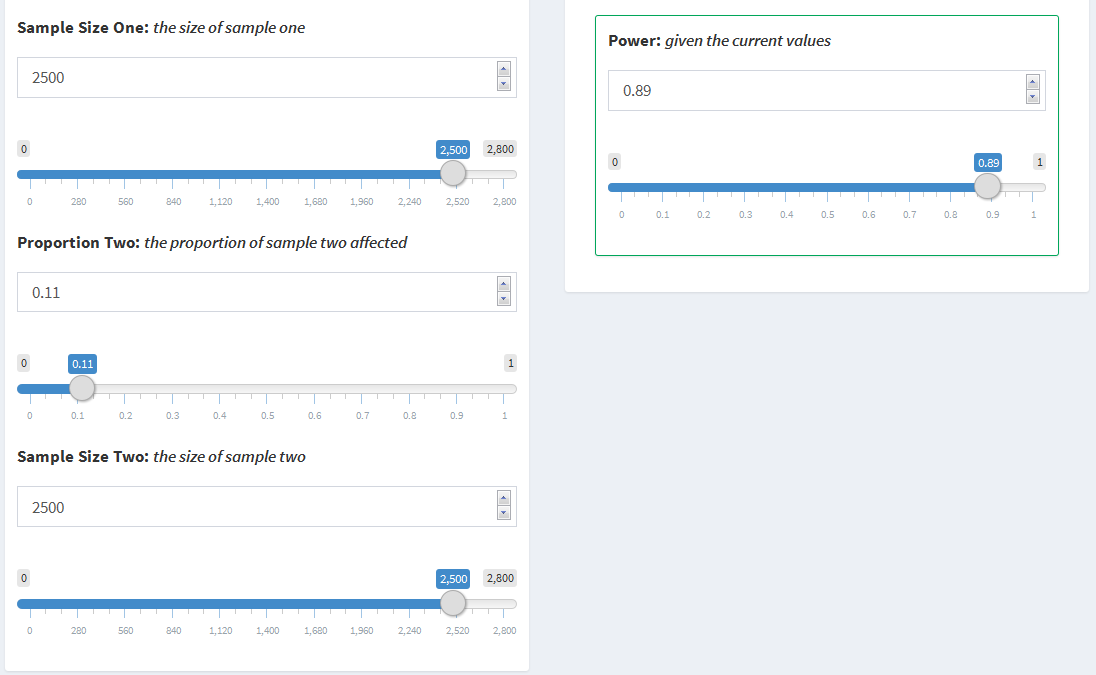
**Calculating Power**: The original study that collected data from Olmsted County, MN collected data only on whether or not the participant had MGUS. Suppose a new 15-year study is being designed to compare the proportion of participants diagnosed with MGUS that die during follow-up in Olmsted County, MN to Hennepin County, MN. The proportion of patients diagnosed with MGUS that died during follow-up in the original study, 0.14, can be used as an estimate for Olmsted County. Suppose that 2500 participants from each county will be entered into the study. The question of interest is, with a significance level of 0.05, how much power the study will have to detect a 20% difference, in either direction, in the proportion of participants with MGUS that die during follow-up. That is, what is minimum power the study will have to detect a difference between 0.14 and 0.14 0.03.

Assuming there is a 20% decrease in the death rate, the following values can be used to calculate the power:

* Proportion One: 0.14
* Sample Size One: 2500
* Proportion Two: 0.11
* Sample Size Two: 2500
* Significance Level: 0.05

The study would have 89% power to detect a 20% decrease in death rate.

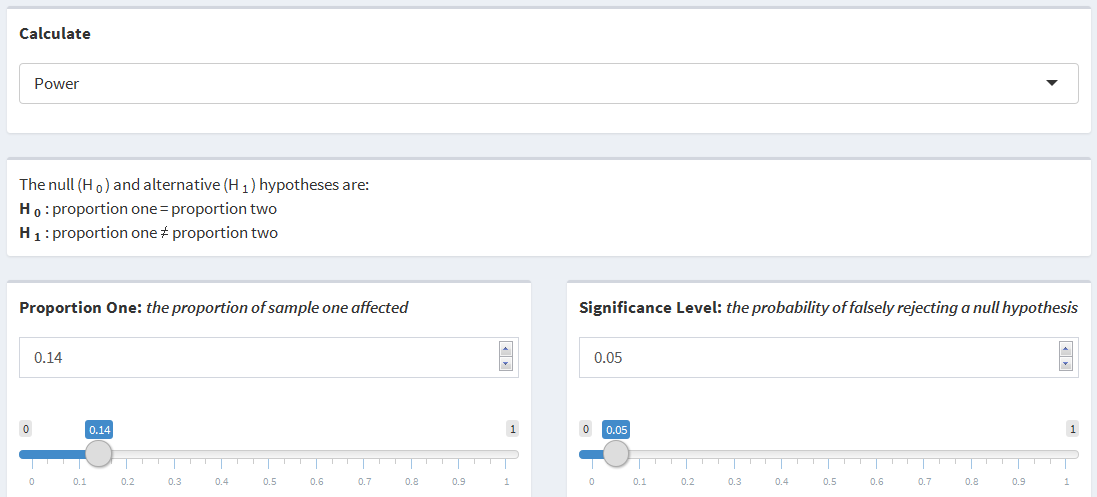


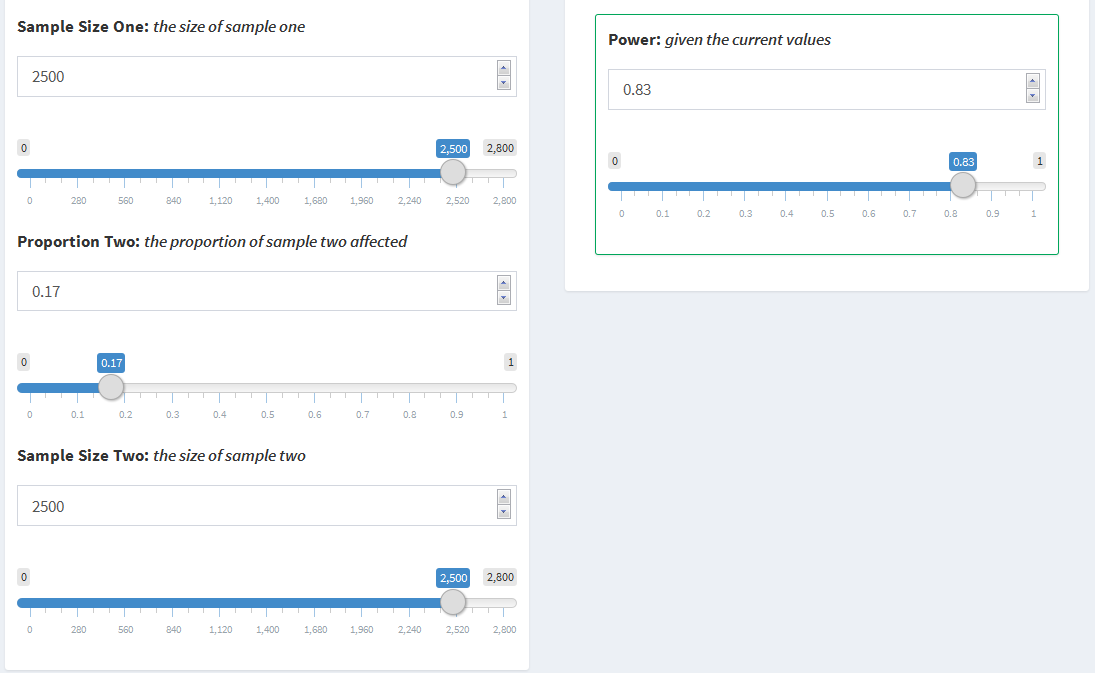


Now, assuming there is a 20% increase in the death rate, the Proportion Two input can be changed from 0.11 to 0.17 to calculate the power of the study as follows:

* Proportion One: 0.14
* Sample Size One: 2500
* Proportion Two: Adjust from 0.11 to 0.17
* Sample Size Two: 2500
* Significance Level: 0.05

The study would have 83% power to detect a 20% increase in death rate. This means that the study will have, at a minimum, 83% power to detect a 20% difference from the Olmsted County death rate.





## 3.5 Time to Event

### 3.5.1 Statistical Explanation

**Sample Size**: The Time to Event scenario looks to compare the event hazard rates of two different samples in a one-sided comparison. Specifically, the event hazard rates, and , are being compared by examining the hazard ratio, , of the two samples. The null and alternative hypotheses in this case are and respectively. Lachin and Foulkes (1981) derived a basic equation relating the sample size with the power when examining the risk ratio as follows:

.

In the equation above, is the proportion of the whole population in the sample, is the weighted event hazard rate, is the expected number of events given the event hazard rate, and is defined as the component of the variance, , independent of the sample size such that . In these types of scenarios, the variance and power are functions of the expected number of events. For a cohort of size , the expected number of events can be expressed as . Let be an indicator of whether the participant experienced the event. Then and (Lachin, 1986). Therefore, for a sample of size N, can be expressed as follows:

The expected number of events component of the equation changes depending on the distribution of entry into the study, in this case, either uniform or exponential. It is often the case that studies will have recruitment periods within the duration of entry into the study. One basic assumption is that entry into the study during the recruitment period will occur uniformly throughout that time. This can be reflected by letting , where is the event hazard rate, is the duration of the study, and is the duration of recruitment. It may also be the case that entry into the study doesn't occur uniformly and instead occurs at an exponential rate, either convex or concave. In this case, let , where is the exponential rate parameter. Entry will be convex, faster than expected, if and entry will be concave, slower than expected, if . The rate at which entry increases or decreases gets faster as gets further from 0. Another common assumption made in these types of studies is that all of the possible events that could occur will be observed. However, it is very likely that some events may not be observed for various reasons such as not being reported to the researchers or participants dropping out of the study early. These types of scenarios result in censoring of participant information.. Each group can have its own censoring rate representing the probability of being censored. Then, the observed event hazard rate, , can be estimated by deflating the expected event hazard rate, , using the censoring rate, , where . The observed event hazard rate can then be substituted into the sample size equation for each instance of its corresponding expected event hazard rate. The final sample size equation is then as shown below.

As stated above, and can be replaced with the observed event hazard rates if censoring is present and can changed depending on they type of entry into the study that is expected.

**Power**: In *The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions* (Schoenfeld, 1981), Schoenfeld derived a test statistic for comparing survival distributions in terms of the expected number of events. The statistic is asymptotically normal with unit variance. He also showed that using the assumptions of the log-rank test, which is commonly used for comparing survival curves, the mean for the test statistic reduces to , where is the event hazard rate for the group, is the sample size, , and is the combined probability of an event occurring. This can be further simplified by multiplying and to arrive at the expected number of events, . The power can then be calculated in terms of the expected number of events following the equation outlined by Schoenfeld below.

### 3.5.2 Functional Use

The Time to Event scenario uses the nSurvival and nEvents functions within the gsDesign package to perform sample size and power calculations. The nSurvival function is used to calculate the sample size and both functions are used to calculate the power.

**Sample Size**: The nSurvival function calculates the sample size required for a time to event study in terms of the event hazard rates for the respective groups. It uses the inputs listed below to perform the calculations.

* lambda1 - Event hazard rate for the first sample
* lambda2 - Event hazard rate for the second sample
* eta - Dropout hazard rate for both samples
* ratio - Randomization ratio between samples one and two
* Ts - Maximum duration of the study
* Tr - Duration of recruitment
* alpha - Type I error rate
* beta - Type II error rate
* sided - One or two-sided test
* type - Risk ratio or risk difference
* entry - Uniform or exponential entry
* gamma - Exponential entry rate parameter

In order to calculate the sample size in this scenario, a few of the inputs must be pre-specified. Since the event hazard rates will each be adjusted using censoring to represent the observed event hazard rate for each group, the equal dropout hazard rate, eta, is set to zero. This scenario is set to use a one-sided, , alternative to examine the risk ratio. The rest of the parameters can be manipulated to calculate the sample size for different situations by following the equation outlined in the *Statisistical Explanation* section above.

**Power**: Both the nSurvival and nEvents functions are used to calculate the power. Calculations for the power are slightly more involved than for the sample size due to the way the gsDesign package is built (Anderson, 2016). Although calculations for sample size and power can theoretically be calculated in terms of either the event hazard rates or the expected number of events, the gsDesign package can only calculate power in terms of the expected number of events. Thus, conversions from the event hazard rates to the number of expected events are performed and then the power is calculated. The first step is to calculate the required sample size in terms of the event hazard rates using any power. The equation below uses a power of to calculate the sample size.

In this equation, all of the parameters are as defined in the *3.5.1 Statistical Explanation* section above. The expected number of events, below, can then be calculated using the sample size estimation from above. That is,

where is the estimated sample size from above, represents the event of interest, and is defined as in the sample size calculations above. By using the sample size estimate, the expected number of events is dependent on the parameter and thus, the power. Dividing the expected number of events, from the initial sample size estimation, by the sample size estimate will result in the number of events per person, where . The events per person equation no longer uses the sample size estimate and is thus free from dependency on the parameter (this is why any power could be used for the initial sample size estimation). The number of expected events is calculated my multiplying the specified sample size by the events per person. The power can then be calculated in terms of the expected number of events following the equation outlined in the *Power* sub-section of the *Statistical Explanation* section above. The same inputs for the nSurvival function outlined above are used for the power calculation, except it disregards the power input (i.e. any power can be used). The following inputs are used by the nEvents function to calculate the power.

* hr - Hazard ratio under the alternative hypothesis
* alpha - Type I error rate
* beta - Type II error rate
* ratio - Randomization ratio between groups one and two
* sided - One or two-sided test
* hr0 - Hazard ratio under the null hypothesis
* n - Number of events

In this scenario, beta is the parameter that is being calculated. The calculations are performed using a one-sided, , alternative where the null hypothesis assumes the risk ratio is one. The rest of the parameters can be manipulated to calculate the power. The nEvents function uses the pnorm function of the normal distribution to calculate the power.

### 3.5.3 Inputs and Outputs

* Enrollment Schedule - The time at which participants are enrolled in the study
* Distribution of Enrollment - The distribution of how participants are enrolled
* Exponential Rate - The rate of growth or decay of the exponential enrollment
* Study Duration - The duration of the study in units of time
* Enrollment Duration - The duration of enrollment in units of time
* Sample Allocation Ratio - The ratio of the target sample size to the reference sample size
* Target Event Rate - The number of events per unit of time for the target sample
* Reference Event Rate - The number of events per unit of time for the reference sample
* Target Censoring Rate - The rate at which events will not be observed in the target sample
* Reference Censoring Rate - The rate at which events will not be observed in the reference sample
* Total Sample Size - The combined size of the target and reference samples
* Significance Level - The probability of falsely rejecting a null hypothesis
* Power - The probability of correctly rejecting a null hypothesis

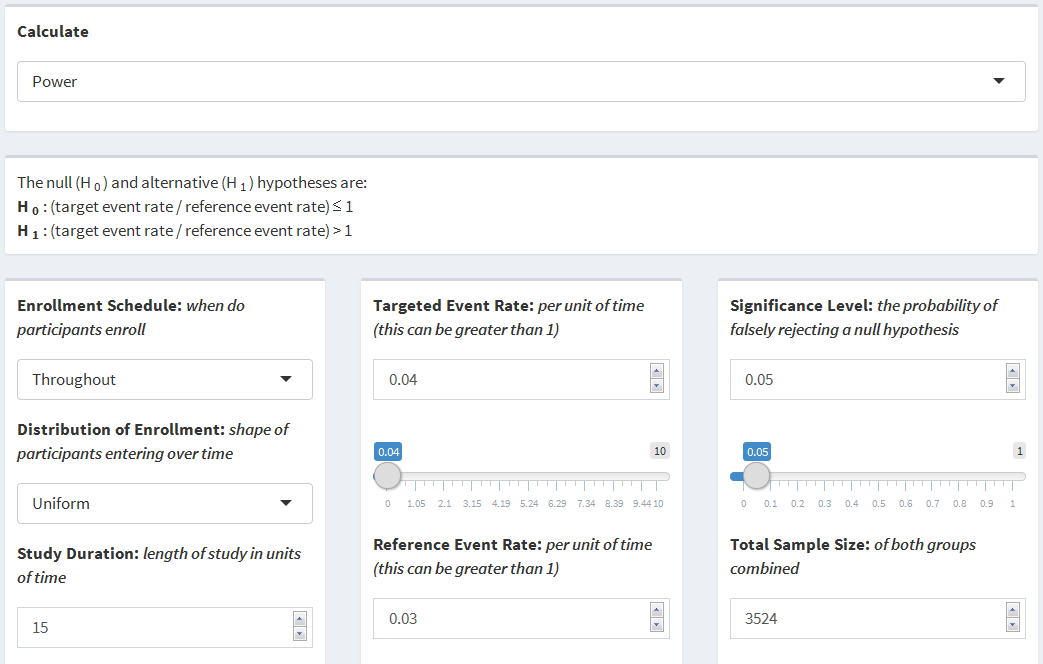
Since certain inputs are dependent on other inputs being active, some inputs will only display with certain conditions. Distribution of Enrollment becomes visible if either "All at Once" or "Throughout" are selected for Enrollment Schedule. Exponential Rate becomes visible if Distribution of Enrollment is visible and if "Exponential" has been selected. Enrollment Duration disappears when "Throughout" is selected for Distribution of Enrollment. The censor rate inputs should be set to 0 if the observed event hazard rate is used for the event hazard rate inputs. When a censor rate is greater than 0, it will change the supplied event hazard rates to reflect the observed event hazard rates.

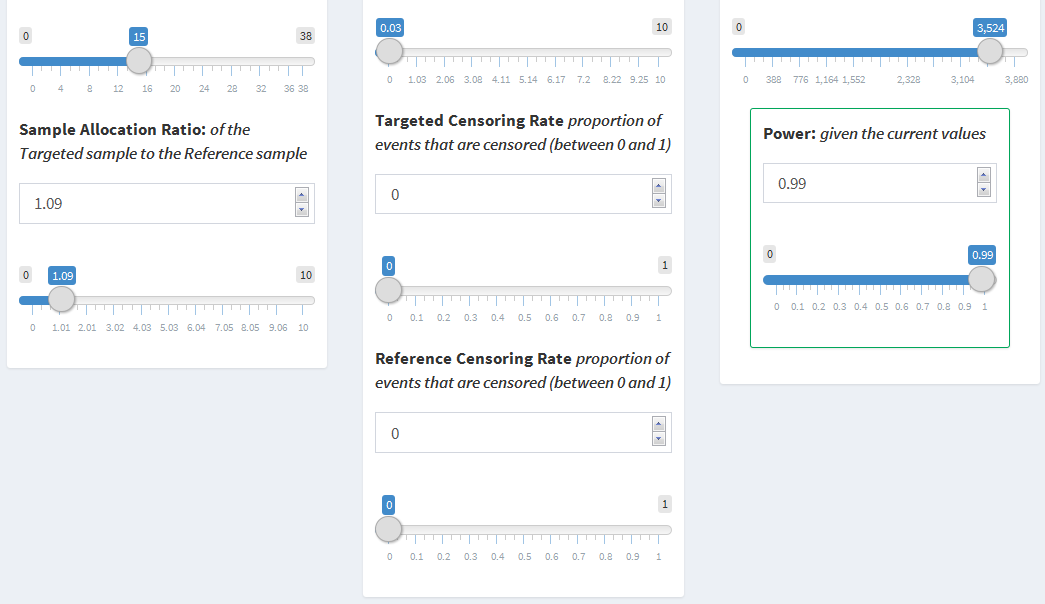
### 3.5.4 Examples

**Calculating Power:** Although the original collection of MGUS data in Olmsted County, MN was not aimed at examining death rates, the secondary assay of serum FLC levels specifically examined the association of FLC levels and death. For the purposes of this example, male participants will be dichotomized by the median rate, , among male participants from the secondary assay ( or ). Suppose the power to detect the difference between hazard rates for participants in the upper half and participants in the lower half of rates is being calculated. Based on the information provided by the study, it lasted about 15 years with recruitment occurring throughout. For the sake of this example, recruitment will be assumed to have occurred uniformly throughout the 15 year study. The sample allocation ratio can be estimated as the number of male participants in the upper half of rates divided by the number of participants in the lower half of rates, . The death rate per year for each group can be estimated as 0.04 for participants in the upper half and 0.03 for participants in the lower half. Since the events have already been observed, the censor rates for both groups is zero. There were 3524 male participants in the study. Assuming a 0.05 significance level, the power of the study can be calculated as follows:

* Enrollment Schedule: Throughout
* Distribution of Enrollment: Uniform
* Study Duration: 15 (years)
* Sample Allocation Ratio: 1.09
* Target Event Rate: 0.04 (death rate per year)
* Reference Event Rate: 0.03 (death rate per year)
* Target Censoring Rate: 0
* Reference Censoring Rate: 0
* Significance Level: 0.05
* Total Sample Size: 3524

Based on the values entered, the study would have 99% power.

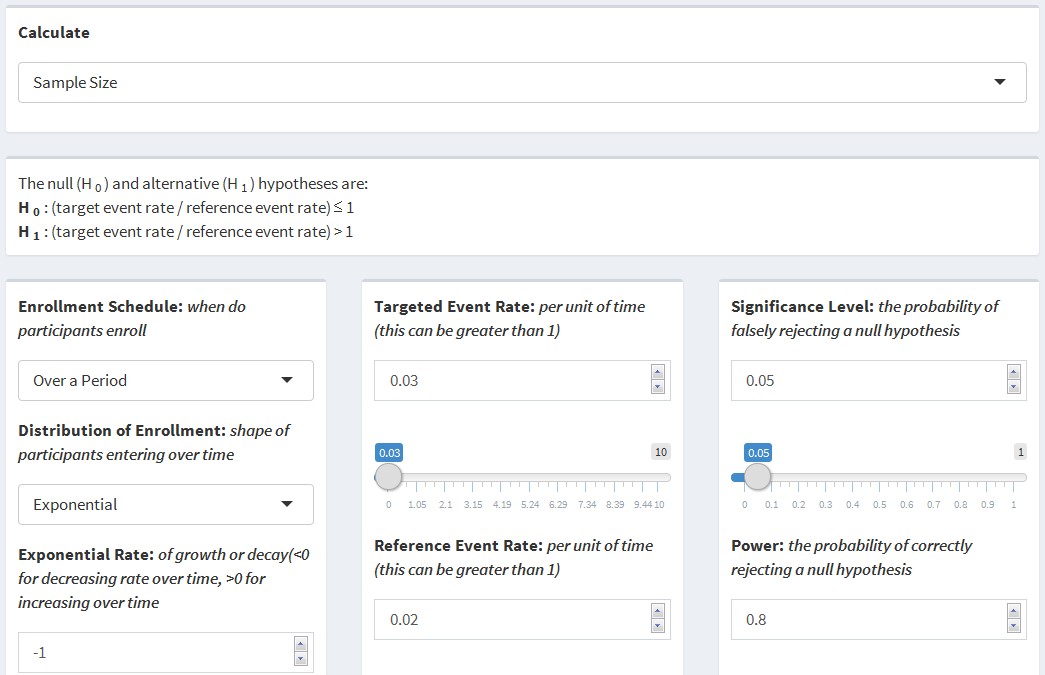


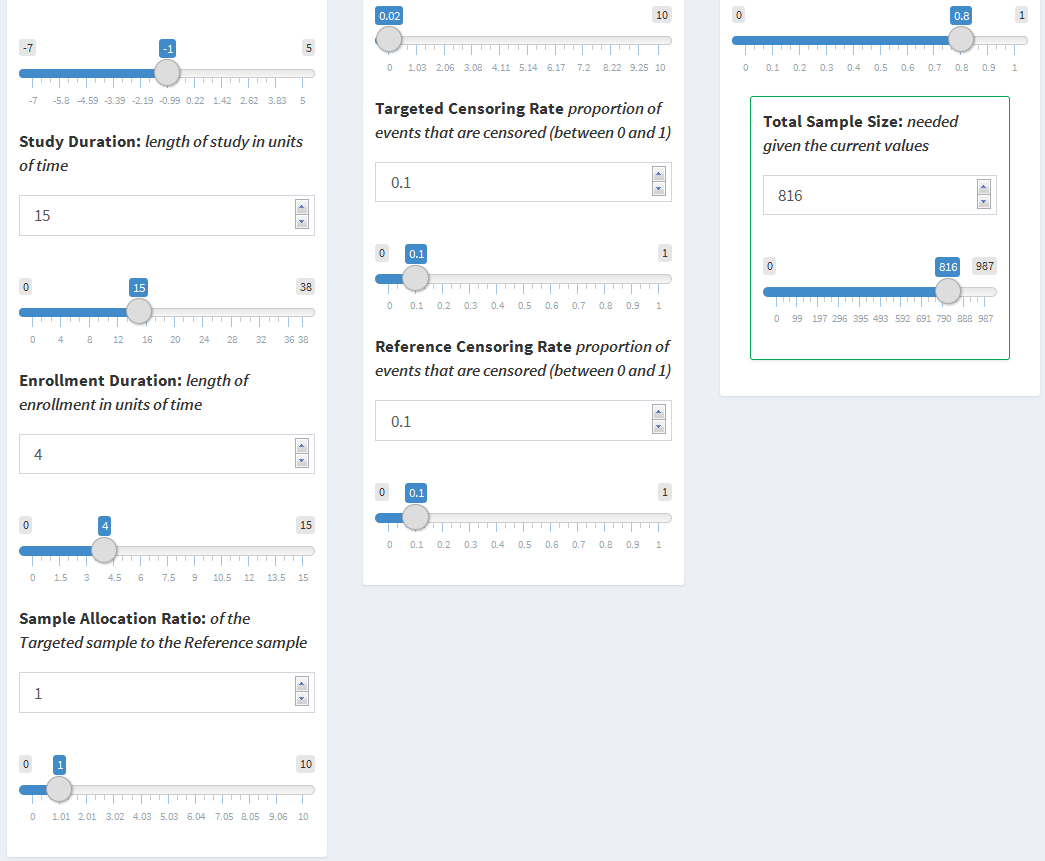


**Calculating Sample Size:** Suppose, instead of examining the death rates among male participants, a new study is being designed to examine the same measures among female participants. The secondary assay median rate, , among female participants can be used to dichotomize the participants, or . It will be a similar 15 year study with recruitment only occurring during the first four years. Instead of recruitment occurring uniformly, suppose recruitment is expected to occur at an exponentially decaying rate, . The death rates can be estimated at 0.03 for women in the upper half and 0.02 for women in the lower half. Suppose 10% of events will be censored in each group. If it is desired that both groups have the same number of participants using a 0.05 significance level and 80% power, the the total number of participants required for the study can be calculated as follows:

* Enrollment Schedule: Over a Period
* Distribution of Enrollment: Exponential
* Exponential Rate: -1
* Study Duration: 15 (years)
* Enrollment Duration: 4 (years)
* Sample Allocation Ratio: 1
* Target Event Rate: 0.03 (death rate per year)
* Reference Event Rate: 0.02 (death rate per year)
* Target Censoring Rate: 0.1
* Reference Censoring Rate: 0.1
* Significance Level: 0.05
* Power: 0.8

Based on the values entered, a total sample size of 816 female participants is required, i.e. 408 participants with and 408 participants with .





# 4. Technical Challenges and Future Improvements

Throughout working on this dashboard, there were many features that were particularly difficult to develop or were noted as possibilities for future improvements. Explained below are those challenges and some possible future improvements that can be made as well as additional features that can be added to improve the dashboard.

## 4.1 User Interface

### 4.1.1 Layout

The most important aspect of the user interface is its layout. The goal of this dashboard was to have the output visible while changing any of the inputs. This allows the user to to see the direct effect each input has on the output. That being said, it is also important for the inputs and outputs to be grouped in such a way that is intuitive. For the most part, the current layout attempts to maximize both of these aspects within the limitations of the environment. Further development can be performed to optimize the layout. One specific space-saving improvement would be to use a checkbox or radio input to toggle displaying the slider input for numeric variables.

### 4.1.2 Calculation Selector

The original goal during the development of the dashboard was for all of the inputs and outputs to be interconnected, similar to an equation. For example, the user would be able thaving to select a specific variable to calculate. However, initial attempts revealed limitations of reactive values in Shiny causing a fully interconnected page to result in infinite loops of inputs updating. Further research into reactive values may achieve a fully interconnected page. o change the power to have the sample size update, and then change the sample size to have the power update, all without

### 4.1.3 Slider Units

A useful feature of Shiny Dashboards is the interactive sliders. Dynamic bounds and intervals appropriate for the value being presented can be further refined to maximize the visual benefits of a slider. There are currently semi-dynamic bounds on most of the sliders, however, more appropriate bounds can be added, especially for numbers near zero.

### 4.1.4 Educational Dialog

One of the next steps for this dashboard is shaping it into an educational tool. While many users of this application will have an in-depth knowledge of the statistical concepts used, many other users will have little to no understanding of these concepts. That being said, it is important that everyone using the dashboard understands how to use it and what the output means. Further improvements in this area could include an introductory page explaining how to use the dashboard, similar to section 2 of this document. In addition, a box at the top of each page explaining the scenario and a box at the bottom of each page summarizing and explaining the results could also be added. Graphics including power and sample size curves would also provide an additional way for users to visualize the output based on the changing of inputs.

## 4.2 Server

### 4.2.1 Functional Programming

Most of the code on the server side of the dashboard is unique to each page, and thus, cannot be simplified by using functional programming. It is possible that the code in the server that updates the inputs can be simplified to be a function. Currently, only the inputs for the specific page and calculation are updated, however this code could be consolidated into one function that updates all of the inputs across the dashboard at the same time. This would save numerous lines of repeating code, but may result in decreased efficiency.

### 4.2.2 Consolidation

As mentioned above, most of the server code is unique throughout and does not allow for much improvement. The updating of inputs has been identified as a possible area that can be improved upon. The inputs for each page could instead be updated once for each page instead of once for each calculation selection. Another code section that can be consolidated is in the solving of each calculation. Currently, the code defines all of the input variables into temporary variables first, then uses the temporary variables in the function call to calculate the output. The input variables could be used directly in the function call to consolidate a considerable amount of code.

## 4.3 Alternative Hypothesis

For simplicity, each page currently has a static alternative hypothesis, two-sided for One Sample Mean, One Sample Proportion, Two Sample Means, and Two Sample Proportions and one-sided for Time to Event. The functions used by the dashboard, however, have the capability to perform both one-sided and two-sided calculations for each scenario. Adding this functionality would allow for non-inferiority, superiority, and equivalence scenarios.

## 4.4 Time to Event

As time to event scenarios are generally more complex than the other scenarios, it follows that there are more parameters that can be used for more unique situations. One parameter that can change is whether the risk difference or the risk ratio should be used as the comparison. The power and sample size equations change slightly depending on which comparison is being used. Although both of the comparisons are examining the same parameters, they are interpreted slightly differently. Having the ability to choose instead of just calculating the risk ratio would be a significant improvement.

## 4.5 New Pages

Although this dashboard covers most of the common scenarios that will be seen in practice, there are still numerous other scenarios that can be incorporated into the dashboard. A couple of possible additions include adding paired and crossover study scenarios. Versions of mean and proportion comparisons where more than two, , means or proportions are being compared could also be added to cover as many of the common study scenarios as possible.

# 5. Acknowledgements

I would first like to thank my adviser on this project Dr. Julian Wolfson of the University of Minnesota School of Public Health. Dr. Wolfson was always available for any questions whenever they arose. His guiding presence helped me through the difficult process of developing an application independently.

I would also like to thank the entire biostatistics faculty at the University of Minnesota for all of the work they do to help students in their pursuit of knowledge.

A special thank you is in order for two of my college professors, Drs. Brian and Mariah Birgen of Wartburg College. The guidance I received by Dr. Brian Birgen was second only to his enthusiasm for mathematics. The same can be said of Dr. Mariah Birgen, who ultimately is the reason I pursued biostatistics.

Finally, I am incredibly thankful to have had the best support system through this long, arduous, and sometimes lonely process in my wife, Cassie Partridge. This project would not have been completed without her encouragement and constant support.

# 6. References

Allaire, JJ & Cheng, Joe & Xie, Yihui & McPherson, Jonathan & Chang, Winston & Allen, Jeff & Wickham, Hadley & Atkins, Aron & Hyndman, Rob & Arslan, Ruben (2017). *rmarkdown: Dynamic Documents for R*. R package version 1.5. <https://CRAN.R-project.org/package=rmarkdown>

Anderson, Keaven (2016). *gsDesign: Group Sequential Design*. R package version 3.0-1. <https://CRAN.R-project.org/package=gsDesign>

Champely, Stephane (2017). *pwr: Basic Functions for Power Analysis*. R package version 1.2-1. <https://CRAN.R-project.org/package=pwr>

Chang, Winston & Cheng, Joe & Allaire, JJ & Xie, Yihui & McPherson, Jonathan (2017). *shiny: Web Application Framework for R*. R package version 1.0.3. <https://CRAN.R-project.org/package=shiny>

Chang, Winston (2016). *shinydashboard: Create Dashboards with 'Shiny'*. R package version 0.5.3. <https://CRAN.R-project.org/package=shinydashboard>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Creating Shiny reactive variable that indicates which widget was last modified. (2015, July 6). Retrieved May 06, 2017, from <http://stackoverflow.com/questions/31250587>

HyLown Consulting LLC. (n.d.). Overview of Power and Sample Size .com Calculators. Retrieved May 06, 2017, from <http://powerandsamplesize.com/Calculators/>

Lachin, J. M., & Foulkes, M. A. (1986). *Evaluation of Sample Size and Power for Analyses of Survival with Allowance for Nonuniform Patient Entry, Losses to Follow-Up, Noncompliance, and Stratification*. Biometrics, 42(3), 507-519. <doi:10.2307/2531201>

Lenth, R. V. (2006-9). Java Applets for Power and Sample Size [Computer software]. Retrieved May 3, 2017, from <http://www.stat.uiowa.edu/~rlenth/Power>

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RStudio. (2014, January 6). Reactivity: An overview. Retrieved May 06, 2017, from <http://shiny.rstudio.com/articles/reactivity-overview.html>

RStudio Team (2016). *RStudio: Integrated Development for R*. RStudio, Inc., Boston MA. URL <http:www.rstudio.com/>.

Schoenfeld, D. (1981). The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. Biometrika, 68(1), 316-319. <doi:10.2307/2335833>

Schoenfeld, D. A. (1983). Sample-Size Formula for the Proportional-Hazards Regression Model. Biometrics, 39(2), 499-503. <doi:10.2307/2531021>

Singh, G. (2016). Serum Free Light Chain Assay and Ratio: Performance in Patients With Monoclonal Gammopathy-High False Negative Rate for Ratio. Journal Of Clinical Medicine Research, 9(1), 46-57. doi: <https://doi.org/10.14740/jocmr2802w>

Turley, R., BSN, MSN, & Walton-Ziegler, O., MS, PA-C. (n.d.). Free Light Chains (Blood). Retrieved September 18, 2017, from <https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=167&contentid=serum_free_light_chains>