Tutorial 4: Power and Sample Size for the Two-sample t-test with Unequal Variances

Preface

Power is the probability that a study will reject the null hypothesis. The estimated probability is a function of sample size, variability, level of significance, and the difference between the null and alternative hypotheses. Similarly, the sample size required to ensure a pre-specified power for a hypothesis test depends on variability, level of significance, and the null vs. alternative difference. In order to understand our terms, we will review these key components of hypothesis testing before embarking on a numeric example of power and sample size estimation for the two independent-sample case with unequal variances.

Overview of Power Analysis and Sample Size Estimation

A *hypothesis* is a claim or statement about one or more population parameters, e.g. a mean or a proportion. A *hypothesis test* is a statistical method of using data to quantify evidence in order to reach a decision about a hypothesis. We begin by stating a *null hypothesis*, H_0, a claim about a population parameter, for example, the mean; we initially assume the null hypothesis to be true. H_0 is where we place the burden of proof for the data; it is usually what an investigator hopes to *disprove* if the evidence in the data is strong enough. In the two-sample means case the goal is to determine if the two samples have the same average. So, for a two-sample test of the mean, the null hypothesis H_0 is a simple statement about the value the difference in the means is expected to have.

As an example, let us say that a randomized two-arm trial in post-surgery/ radiation head and neck cancer patients is planned, with a 2:1 ratio of patients on resveratrol (an antioxidant found in the skins of grapes and other fruits) relative to patients on placebo. Ki-67 is measured in tumors after surgery and before the administration of either resveratrol or placebo. Measurements are obtained from biopsies after 6 months on resveratrol or placebo. The primary endpoint is change in Ki-67, a measure of cell proliferation (% of cells staining for Ki-67). The goal is to compare mean change in Ki-67 after resveratrol vs. after placebo, i.e. $\mu_{\Delta Ki-67, resveratrol}$ vs. $\mu_{\Delta Ki-67, placebo}$. The null hypothesis is that there is no difference in mean change in Ki-67, i.e. H_0: $\mu_{\Delta Ki-67, resveratrol} - \mu_{\Delta Ki-67, placebo} = 0\%$. An opposing, or alternative hypothesis H_1 is then stated, contradicting H_0, e.g. H_1: $\mu_{\Delta Ki-67, resveratrol} - \mu_{\Delta Ki-67, placebo} \neq 0\%$, H_1: $\mu_{\Delta Ki-67, resveratrol} - \mu_{\Delta Ki-67, placebo} < 0\%$, or H_1: $\mu_{\Delta Ki-67, resveratrol} - \mu_{\Delta Ki-67, placebo} > 0\%$. If the mean changes from the two samples are sufficiently different a decision is made to *reject the null* in favor of the

specific alternative hypothesis. If the sample mean change difference is close to zero, the null hypothesis cannot be rejected, but neither can a claim be made that the hypothesis is unequivocally true.

Because of sampling there is inherent uncertainty in the conclusion drawn from a hypothesis test. Either a correct or an incorrect decision will be made, and the goal is to minimize the chances of making an incorrect one. The probability of rejecting a true null hypothesis, called Type I Error, is denoted by α. The probability of failing to rejecting a false null hypothesis, called Type II Error, is denoted by β. The possible outcomes of a hypothesis test are summarized in the table below along with the conditional probabilities of their occurrence.

Hypothesis Testing Decision	State of Nature	
	H_0 True	H_1 True
Fail to reject H_0	True negative A correct decision Prob(True negative) = $1-\alpha$	False negative A Type II Error Prob(False negative) = β
Reject H_0	False positive A Type I Error Prob(False positive) = α	True Positive A correct decision Prob(True positive) = 1-β

A Type I error occurs when H_0 is incorrectly rejected in favor of the alternative (i.e. H_0 is true). A Type II error occurs when H_0 is not rejected but should have been (i.e. H_1 is true). To keep the chances of making a correct decision high, the probability of a Type I error (α, the level of significance of a hypothesis test) is kept low, usually 0.05 or less, and the power of the test (1-β, the probability of rejecting H_0 when H_1 is true) is kept high, usually 0.8 or more. When H_0 is true, the power of the test is equal to the level of significance. For a fixed sample size, the probability of making a Type II error is inversely related to the probability of making a Type I error. Thus, in order to achieve a

desirable power for a fixed level of significance, the sample size will generally need to increase.

Other factors that impact power and sample size determinations for a two-independent sample hypothesis test include variability as summarized by σ_1^2 and σ_2^2, where σ_1 and σ_2 are the standard deviations in the two populations; and the difference between the population means under the null hypothesis and under a specific alternative, e.g. $\mu_{\Delta Ki\text{-}67, resveratrol} - \mu_{\Delta Ki\text{-}67, placebo} = 0\%$ vs. $\mu_{\Delta Ki\text{-}67, resveratrol} - \mu_{\Delta Ki\text{-}67, placebo} = 20\%$. In practice, σ_1^2 and σ_2^2 can be difficult to specify. Previous work reported in the literature is a good source of information about variability, but sometimes a pilot study needs to be carried out in order to obtain a reasonable estimate. Specifying the magnitude of the difference you wish to detect between the null and alternative hypothesis varies by consideration of subject matter. The importance of a difference varies by discipline and context.

The following summarizes the interrelationships of power, sample size, level of significance, variability and detectable difference:

> To increase power, 1-β:
> - Increase: Required Sample Size, n; Detectable Difference, between the mean under H_0 and H_1, $|\mu_0 - \mu_1|$; Level of significance, α;
> - Decrease: Variability, σ_1^2 and σ_2^2
>
> To reduce required sample size, n = n_1 + n_2 -
> - Increase: Detectable Difference, between the mean under H_0 and H_1, $|\mu_0 - \mu_1|$; Level of significance, α
> - Decrease: Variability, σ_1^2 and σ_2^2; power, 1-β

An additional consideration for power analysis and sample size estimation is the direction of the alternative hypothesis. A one-sided alternative hypothesis, e.g. H_1: $\mu_{\Delta Ki\text{-}67, resveratrol} - \mu_{\Delta Ki\text{-}67, placebo} > 0\%$, leads to greater achievable power than a two-sided alternative for a fixed sample size, or a lower required sample size for a fixed power. The use of one-sided tests is controversial and should be well justified based on subject matter and context.

Uncertainty in specifications of mean difference, variability

For a discussion of the role of uncertainty regarding inputs to power analysis and sample size estimation, refer to the tutorial in this series on *Uncertainty in Power and Sample Size Estimation*.

Content A: Features and Assumptions
Study design description: Ki-67 and resveratrol example

In the Ki-67 example, the inference relates to the difference in mean change in Ki-67 between two populations. The outcome variable for each sample is continuous (change in % staining) with an unknown, but finite variance of the change in each population, σ_1^2 and σ_2^2. Since we base the % staining on a large number of tumor cells, we may reasonably assume a normal distribution value for that Ki-67 value. As linear functions of normally distributed quantities are also normally distributed, the change in Ki-67 from before treatment to after treatment is normally distributed. Finally, the *a priori* stated hypothesis regards the mean difference in change of Ki-67 and whether the mean difference is 0%. These are the features that make the example suitable for a two-sample t-test.

In carrying out a two-sample t-test we make the assumption that the individual change values are randomly sampled from one of two well-characterized populations and that the observations within a sample are independent of each other, i.e. that there is no clustering between subjects or units of observation. In most cases, we can easily verify this assumption. As noted above, we assume an approximately normal underlying distribution of change in Ki-67 with resveratrol or placebo. Finally, we make no assumptions about the equality of σ_1^2 and σ_2^2.

Two-sample t-test: inference about difference in means μ_1 and μ_2 when the variances σ_1^2 and σ_2^2 are unknown and unequal

To test the hypothesis about the difference in mean change in Ki-67, the standardized value of the difference between sample mean changes, $\overline{X}_{\Delta Ki-67,resveratrol} - \overline{X}_{\Delta Ki-67,placebo}$, becomes the test statistic:

$$t = \frac{\overline{X}_{\Delta Ki-67,resveratrol} - \overline{X}_{\Delta Ki-67,placebo} - \mu_{diff_0}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

where μ_{diff_0} is the difference in mean change under H_0, s_1^2 and s_2^2 are the two sample variances, and n_1 and n_2 are the sample sizes for the two groups. When the variances are not equal the stated t-statistic does not follow a t distribution with $n_1 + n_2 - 2$ df. This issue is known as the Behrens-Fisher problem and various approximate solutions

have been proposed. The most widely used of these approximations is known as the Satterthwaite approximation or as Welch's t-test. Instead of following the usual t distribution based on assuming equal variances, the *t* value is assumed to follow a t distribution with *v* degrees of freedom:

$$v = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2\big/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2\big/n_2\right)^2}{n_2 - 1}}.$$

Note that the df for this test can take on fractional values. These values pose no problems for power and sample size calculations or data analysis in which probabilities are obtained using numerical integration. When using statistical tables to obtain p-values after data analysis, however, the df are rounded down to the nearest integer. In GLIMMPSE, the numerical integration approximation by DiSantostefano and Muller (1995) is used to obtain power and sample size values.

To test the hypothesis H_0: $\mu_{\Delta Ki\text{-}67, resveratrol}$ - $\mu_{\Delta Ki\text{-}67, placebo}$ = 0% vs. H_1: $\mu_{\Delta Ki\text{-}67, resveratrol}$ - $\mu_{\Delta Ki\text{-}67, placebo}$ ≠ 0%, we reject H_0 if:

$$\bar{X}_{\Delta Ki-67,resveratrol} - \bar{X}_{\Delta Ki-67,placebo} > 0\% + t_{v,1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad or$$

$$\bar{X}_{\Delta Ki-67,resveratrol} - \bar{X}_{\Delta Ki-67,placebo} < 0\% - t_{v,1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $t_{v,1-\alpha/2}$ is the (1-α/2) x 100th percentile of the t distribution with *v* degrees of freedom as defined above.

The two-sample t-test as a <u>G</u>eneral <u>L</u>inear <u>M</u>odel (GLM)

A more general characterization of the two-sample t-test can be made using the general linear model (GLM), $Y = X\beta + \varepsilon$. For the test of the mean change in Ki-67, *Y* is an (n_1 + n_2) x 1 matrix of Ki-67 change values, *X* is an (n_1 + n_2) x 2 matrix of 1s and 0s, β is a 1 x 2 matrix of the unknown regression coefficients, β_1 and β_2 − in this case the mean changes in Ki-67 in the resveratrol and placebo groups, and ε is an (n_1 + n_2) x 1 matrix that represents the random deviation of a single Ki-67 change value from its group mean change. The Y and X matrices are shown below:

Y_{Δ}	X_1	X_2
$Y_{1,1Pre}-Y_{1,1Post}$	1	0
$Y_{1,2Pre}-Y_{1,2Post}$	1	0
.	.	.
.	.	.
$Y_{1,n1Pre}-Y_{1,n1Post}$	1	0
$Y_{2,1Pre}-Y_{2,1Post}$	0	1
$Y_{2,2Pre}-Y_{2,2Post}$	0	1
.	.	.
.	.	.
$Y_{2,n2Pre}-Y_{2,n2Post}$	0	1

The assumptions of the GLM are:

- **Existence**: For any fixed value of the variable *X*, *Y* is a random variable with a certain probability distribution having finite mean and variance.
- **Independence**: The *Y*-values are statistically independent of one another.
- **Linearity**: The mean value of *Y* is a straight-line function of *X*.
- **Homoscedasticity (equal variances)**: The variance of *Y* is the same for any value of *X*. That is,

$$\sigma^2_{Y|X} = \sigma^2_{Y|X=1} = \sigma^2_{Y|X=2} = ... = \sigma^2_{Y|X=x}$$

For the test presented here, we relax the assumption to allow unequal variances between the two groups.

- **Normal Distribution**: For any fixed value of *X*, *Y* has a normal distribution. Note this assumption does *not* claim normal distribution for *Y*.

We obtain the estimate of β, $\hat{\beta}$ (or $\hat{\beta}_1$ and $\hat{\beta}_2$), using the method of least squares. To test the hypothesis H$_0$: $\mu_{\Delta Ki\text{-}67, resveratrol}$ - $\mu_{\Delta Ki\text{-}67, placebo}$ = 0%, we compare $\hat{\beta}_1$ and $\hat{\beta}_2$ by contrasting the elements of $\hat{\beta}$: $t = \dfrac{(\hat{\beta}_1 - \hat{\beta}_2)}{\sqrt{s^2_{\hat{\beta}_1} + s^2_{\hat{\beta}_1} - 2\text{cov}(\hat{\beta}_1, \hat{\beta}_2)}} \sim t_v$, where $s^2_{\hat{\beta}_1}$ and $s^2_{\hat{\beta}_2}$ are the estimated variances of $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively, cov ($\hat{\beta}_1, \hat{\beta}_2$) is the covariance between them, and *v* are the Satterthwaite approximation degrees of freedom. The approximation by Kenward and Rogers can also be applied for this purpose (see SAS

PROC MIXED and PROC GLIMMIX for examples) and is the preferred approximation method for means estimated from longitudinal and multilevel data (Muller et al., 2007).

B.1   Inputs for Power analysis

A power analysis for the difference in mean change consists of determining the achievable power for a specified: difference between means under the stated H_0 and under the stated H_1, sample size, standard deviation of the difference, and α-level. By varying these four quantities a set of power curves can be obtained that show the tradeoffs.

Information about mean change values and variability of change can be obtained from the published literature. Sample size can be varied over a feasible range of values, and various values of α can be selected to illustrate the sensitivity of the results to conservative vs. liberal choices for the Type I error rate.

For the change in Ki-67 expected with resveratrol and placebo, we look to a study examining change in Ki67. A standard deviation of 14% for change in Ki67 with resveratrol was reported.[1] While we can use this estimate for both the resveratrol and the placebo groups in order to estimate power, we will continue to assume that the variances are not necessarily equal in reality.

Even when using information from large studies in the literature, it we cannot assume that the means and standard deviations are known quantities. These values are estimates, and as such lead to uncertainty in estimated power. For further discussion of uncertainty refer to *Tutorials 0: Uncertainty in Power and Sample Size Estimation*.

Specification of the difference between means for change, $\mu_{\Delta 0}$ - $\mu_{\Delta 1}$

As noted above, subject matter considerations dictate the choice of a value for the difference between the null and alternative mean change. The goal is to specify the

---

[1] It should be noted that in some publications, standard error values are reported. Standard errors refer to the precision of a mean, not to the variability in a single sample. To convert a standard error to a standard deviation, multiply the standard error by the square root of the sample size. Often separate standard deviations of the pre- and post- measures are reported instead of the standard deviation of the change. In this case, assumptions must be made about the correlation, ρ, between the pre- and post- measures in order to obtain an estimate for the standard deviation of change. A value of 0 for ρ means the pre- and post- measures are uncorrelated while a value of 1 means they are perfectly correlated. Low values of ρ would lead to conservative power and sample size estimates while large values of ρ would lead to liberal ones. For a specific ρ value, the standard deviation of change is obtained as: $s_{\Delta} = \sqrt{s^2_{pre} + s^2_{post} - 2\rho s_{pre} s_{post}}$, where s^2 and s are the sample variance and standard deviation, respectively. Rarely is the variance reported in a publication. If it is, the standard deviation is obtained as the square root of the variance.

smallest difference that would be considered scientifically important. For the two independent sample Ki-67 example we consider a 20% difference between treatment groups to be biologically important.

B.2   Inputs for sample size estimation

Sample size estimation for a single mean from paired observations consists of determining the required sample size to achieve a specified level of power for a given difference between the mean change under the stated H_0 and under the stated H_1, standard deviation of the change and α-level. By varying these four quantities we can obtain a set of sample size curves that show the tradeoffs.

As with power analysis, information about mean values and variability can be obtained from the published literature. Uncertainty in these estimates should be incorporated into the sample size estimation by either allowing those values to vary over a reasonable range, e.g. from 0.5 times up to twice the standard deviation, or by using sampling variability to obtain confidence intervals for power or sample size. Power can be varied over a desired range of values, and various values of α can be selected to illustrate the sensitivity of the results to conservative vs. liberal choices for the Type I and Type II error rates.

Content C: How to use the software
How to perform the analysis in GLIMMPSE

To start GLIMMPSE 2.0 beta, type http://samplesizeshop.com/calculate-power-and-sample-size-now/ in your browser window or visit www.SampleSizeShop.com and click on the GLIMMPSE tab, then on GLIMMPSE 2.0.0 Beta is Here! Google Chrome is the suggested browser for this application.

C.1   Power analysis using Matrix mode

Guided Study Design mode is suggested for most users. However, for the two-sample case with unequal variances, Matrix mode is needed to specify the design appropriately.

Start Your Study Design

Welcome to GLIMMPSE. The GLIMMPSE software calculates power and sample size for study designs with normally distributed outcomes. Select one of the options below to begin your power or sample size calculation.

Guided Study Design	Matrix Study Design	Upload a Study Design
Build common study designs including ANOVA, ANCOVA, and regression with guidance from the study design wizard. This mode is designed for applied researchers including physicians, nurses, and other investigators.	Directly enter the matrices for the general linear model. This mode is designed for users with advanced statistical training.	If you have previously saved a study design from GLIMMPSE, you may upload it here. Click browse to select your study design file.
Select	Select	Browse_

Selecting *Matrix Study Design* takes you to the *Introduction* page:

Throughout your use of the GLIMMPSE website, use the blue arrows at the bottom of the screen to move forward and backwards. The pencil symbols beside the screen names on the left indicate required information that has not yet been entered. In the above picture, the pencils appear next to the *Solving For* and the *Type I Error* screen names.

Once the required information is entered, the pencils become green check marks. A red circle with a slash through it indicates that a previous screen needs to be filled out before the screen with the red circle can be accessed. Search for a pencil beside a screen name in the previous tabs to find the screen with missing entries.

Once you have read the information on the *Introduction* screen, click the forward arrow to move to the next screen.

The *Solving For* screen allows you to select Power or Sample Size for your study design. For the purposes of this tutorial, click *Power*. Click the forward arrow to move to the next screen.
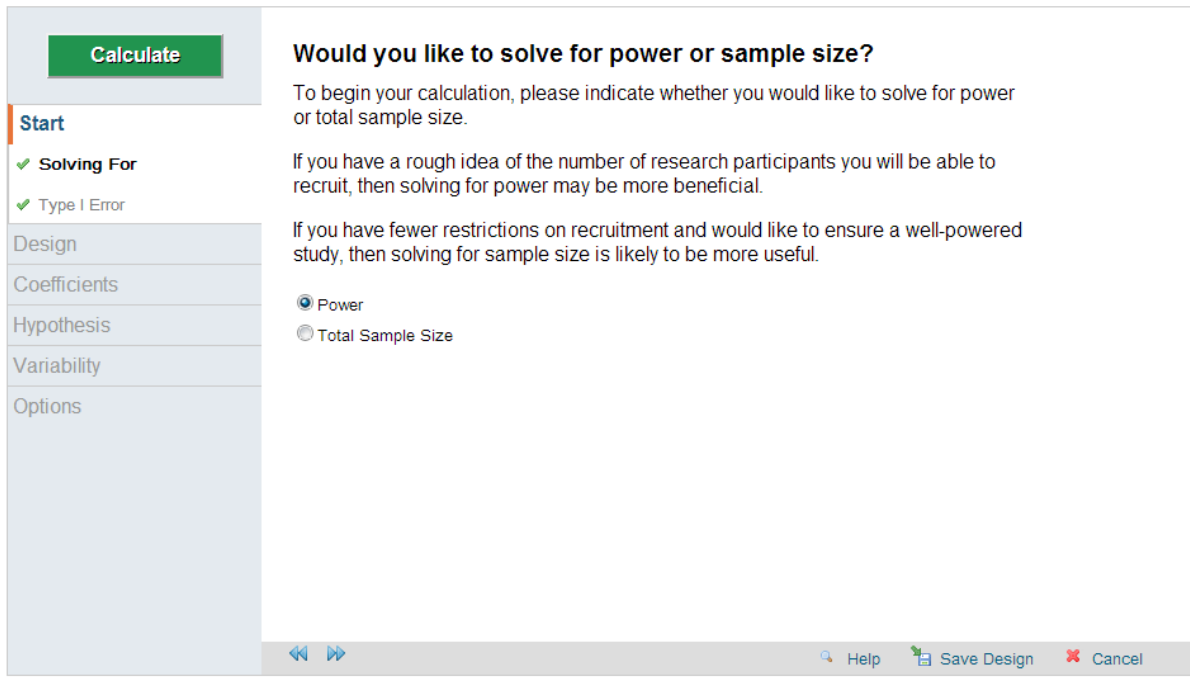
The *Type I Error* screen allows you to specify the fixed levels of significance for the hypothesis to be tested. Once you have entered your values, click the forward arrow to move to the next screen.
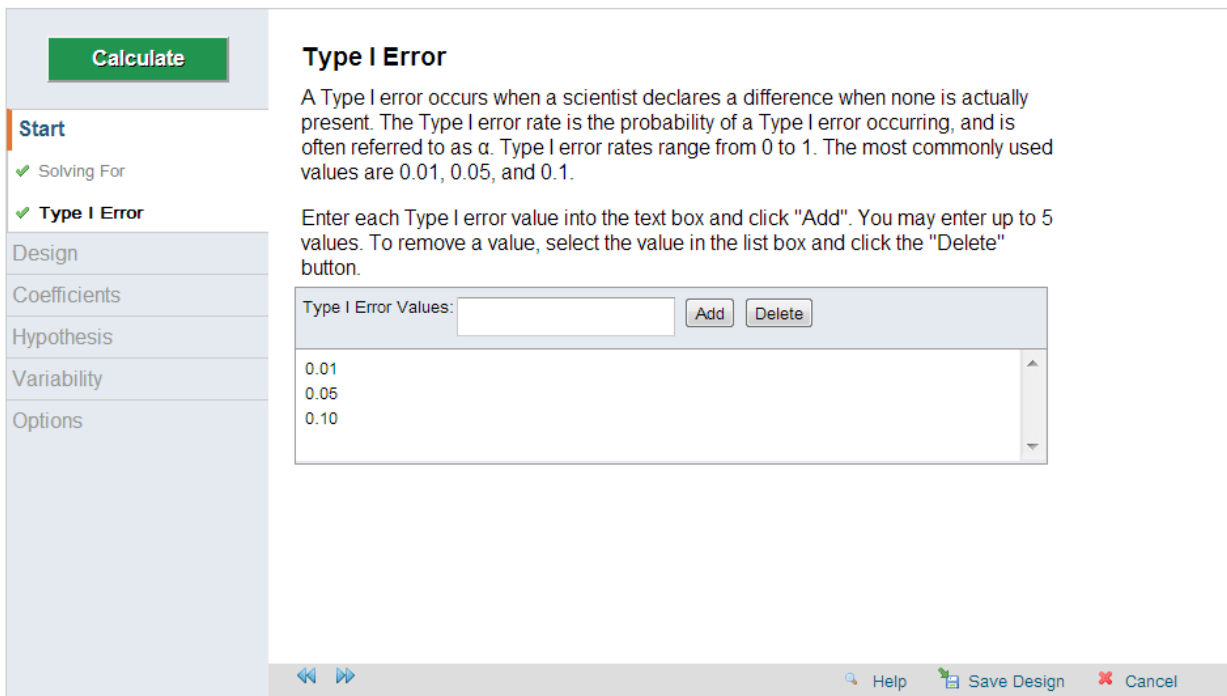
Read the information on the *Design* screen and click next to move to the next screen.

On the *Design Essence* screen, enter 1 x 1 for the dimension of the essence matrix, with a value of 1. This represents one sample. In a later screen you will specify that two responses will be measured in the single sample with a correlation of zero between the responses. This will equate to a single response measured in two independent samples.

Since no covariate will be used, skip the *Covariate* screen.

Use the *Sample Size* screen to specify the size of the smallest group for your sample size(s). Entering multiple values for the smallest group size allows you to consider a range of total sample sizes.

Size of the Smallest Group

Enter the number of independent sampling units (participants, clusters) in the smallest group in the study. If your group sizes are equal, the value is the same for all groups. You may enter multiple values for the smallest group size in order to consider a range of total sample sizes.

Enter one or more sample sizes in the text box below and click "Add". To remove a sample size from the list, highlight it and click the "Delete" button.

Size of the Smallest Group: [_____] [Add] [Delete]

5
10
15
20

Read the *Coefficients* screen, and click the forward arrow when you have finished.

The *Beta Coefficients* screen allows you to enter the means of the response variable of interest. In this example, the response variable is the change in Ki-67 in each of the treatment groups, *Resveratrol* and *Placebo*. Enter 2 in the box (1 x 2) and then enter the expected mean change in Ki-67 in each of the two groups, 20% and 0%. Click the forward arrow when you have finished entering your response variable means.

Regression Coefficients: B Matrix

The **B** matrix contains regression coefficients. Specify the values you expect to see for these coefficients. These values may be determined from pilot data or previous studies. We recommend selecting values which represent scientifically meaningful differences. At least one of the values in the **B** matrix should be non-zero. Otherwise, power will equal the test size.

The number of columns in **B** indicates the number of outcomes in your study (i.e. the number of columns in **Y**). To adjust the number of outcomes in your study, change the column dimension in the text box above the matrix data. The number of rows in **B** must equal the number of columns in Es(**X**), so it cannot be adjusted on this screen.

Enter values for the regression coefficients in the matrix below.

The true mean difference is uncertain. The *Beta Scale Factors* screen allows you to see how power varies with the assumed difference, so that you can allow it to vary over a reasonable range. Suggested values are from 0.5x to 2x the stated difference, e.g. from 10% to 20% to 40%.

Read the *Hypothesis* screen, then click the forward arrow to move to the *Between Participant Contrast* screen.

The *Between-Subject Contrast: C Matrix* screen allows you to specify the comparison between groups. For this example, the matrix should be a scalar (1 x 1) with a value of 1. After entering your known mean values, click the forward arrow.

Between-Subject Contrast: C Matrix

The **C** matrix defines the contrasts for between-subject effects. The number of rows in the **C** matrix is at most one fewer than the number of rows in Es(**X**). The number of columns in **C** must equal the number of columns in **B**. To ensure conforming matrices, the number of columns of **C** cannot be adjusted on this screen.

Enter your between-subject contrast matrix below.

1×1

1

Read the information on the *Within-Subject Contrast: U matrix* screen. This is the screen where the comparison between the *Resveratrol* and *Placebo* groups will be specified. Enter 1 for the 2 x 1 vector, and then enter [1 -1] to specify that the difference between the groups will be tested. Click the forward arrow to move to the *Null Hypothesis Matrix* screen.

Enter the value of zero for the null hypothesis difference between means.

Null Hypotheses: Θ_0 Matrix

For $\Theta = \mathbf{CBU}$, the general linear hypothesis is stated as

$H_0 : \Theta = \Theta_0$.

In most cases, the Θ_0 matrix will contain zeros.

The number of rows in Θ_0 must equal the number of rows in **C**, and the number of columns must match the number of columns in **U**. To ensure conforming matrices, the dimensions of Θ_0 cannot be adjusted on this screen.

Enter your Θ_0 matrix below.

1×1

0

Read the information on the *Variability* screen, then click the forward arrow to continue.

The *Error Covariance* screen allows you to specify the expected variability in terms of standard deviation of the outcome variable, Ki-67. For this example, since the variances are assumed to be unequal, a standard deviation can be entered separately for the Resveratrol and Placebo groups. Enter 14% and 20% as the diagonal elements of the 2x2 matrix, with a covariance of zero since the groups are independent. Click the forward arrow to move to the next screen.

The true variability in Ki-67 is also uncertain. To see how power varies with the assumed standard deviations, you can allow the standard deviations to vary over a reasonable range. Suggested values are from 0.5x to 2x the stated values of standard deviation, e.g. from 10% to 20% to 40% in the *Resveratrol* group.

Read the information on the *Options* screen, then click the forward arrow to continue.

For the test of two independent means, the available tests in GLIMMPSE yield equivalent results. Click on any one of the tests and then click the forward arrow to continue.
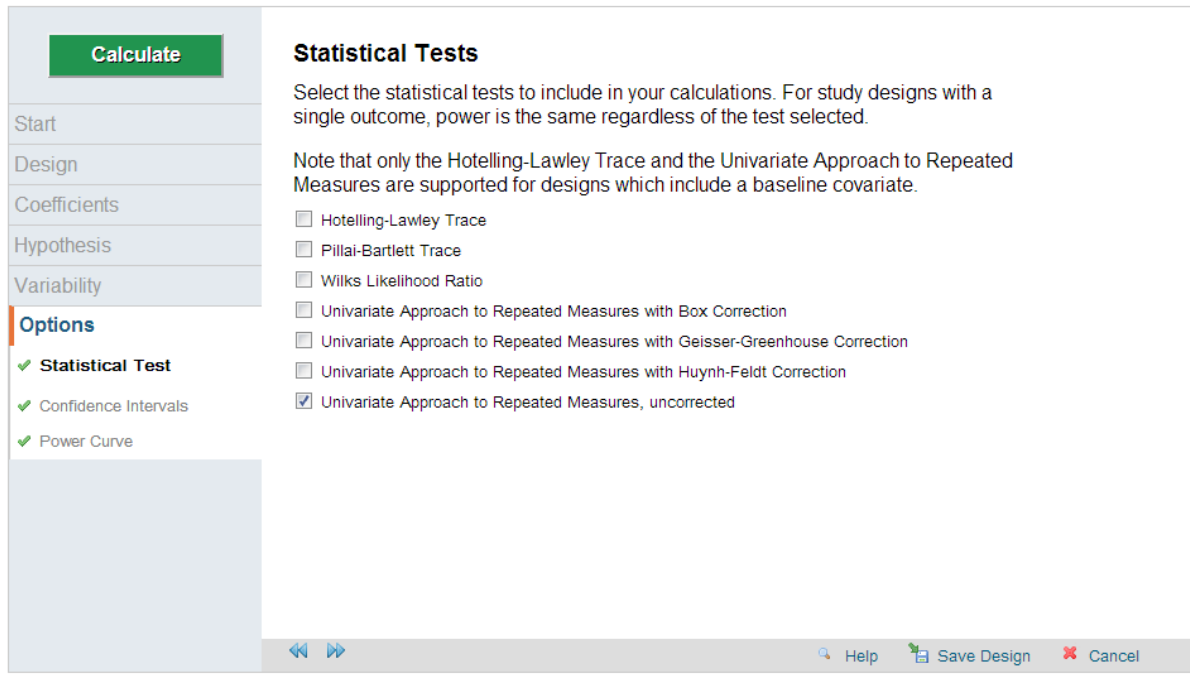
Leave the box checked in the *Confidence Interval Options* screen, and click forward to continue to the next screen.

Power analysis results are best displayed on a graph. To obtain a plot, first uncheck No thanks on the *Power Curve Options* screen.
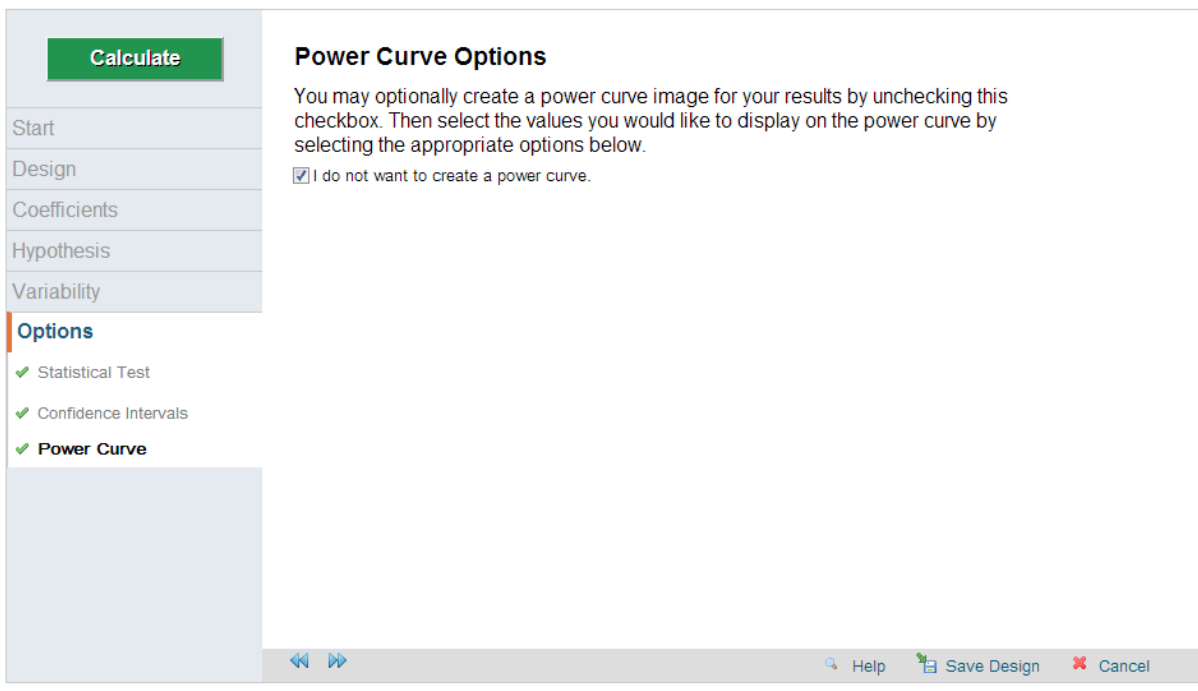
From the pull down menu that appears once you have unchecked the box, select the variable to be used as the horizontal axis (e.g. *Total sample size* or *Variability Scale Factor*). GLIMMPSE will produce one power plot based on specific levels of the input variables that you specify on this page. If you specified more than one value for an input variable, choose the specific level you want GLIMMPSE to use to plot the power cu

After you have made your inputs, click *Add* to populate the screen, then click the forward arrow to see your results.

The power curve and table results for these inputs are shown in Section D below.

C.2 Sample size estimation with GLIMMPSE

To start, click on *Total Sample Size* in the *Solving For* screen. Click on the forward arrow to move to the next screen.

Enter the values for desired power, then click the forward arrow to move to the next screen.

The remaining screens are filled in as above for power.

Content D: Interpret the results

<u>Power analysis interpretation</u>

For the power analysis inputs in Guided Study Design Mode (Section C.1), GLIMMPSE produces a curve showing the relationship between achievable power, mean difference (beta scale), standard deviation (variability scale), and fixed values of total sample size and level of significance. A complete downloadable table of results in Excel.csv format is also produced. The plot below shows achievable power over a range of standard deviations of change in Ki-67 with an α-level of 0.05 (two-sided) and a sample size of 30.

It can be observed in the graph below that above a certain level of variability, power is less sensitive to changes in specified values. The sensitivity of power to assumptions within a certain range of values has been called the "slippery slope of power". In order to avoid the slippery slope, it is best to assume a large amount of variability and/or a large difference between null and alternative means.

Power Curve

— Tutorial 4
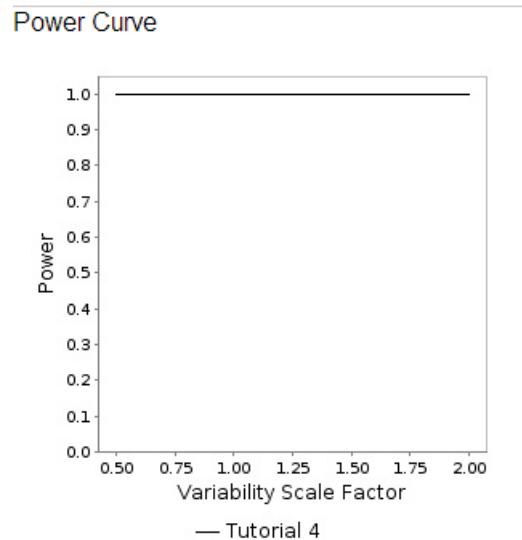
Sample size estimation interpretation

For the sample size estimation inputs in Guided Study Design mode above, GLIMMPSE produces a downloadable table showing the relationship between required sample size, desired power, level of significance, variability, and difference between null and alternative hypothesis means. This can be saved in Excel.csv format and plotted.

Content E: References cited

List of matrices used in calculation:

$$\mathbf{Es(X)} = [1.0000]$$

$$\mathbf{B} = [20.0000 \quad 0.0000]$$

$$\mathbf{C} = [1.0000]$$

$$\mathbf{U} = \begin{bmatrix} 1.0000 \\ -1.0000 \end{bmatrix}$$

$$\mathbf{\Theta_0} = [0.0000]$$

$$\mathbf{\Sigma_E} = \begin{bmatrix} 20.0000 & 0.0000 \\ 0.0000 & 14.0000 \end{bmatrix}$$

For the independent sample t-test, GLIMMPSE works with the matrices listed above in making the computations. Since two means are being tested the Essence matrix is of dimension 2x2. The $\mathbf{\Theta_0}$ matrix represents the mean under H_0 and the Σ_e matrix represents the between-subject variance.

References

DiSantostefano, R.L. and Muller, K.E. (1995). A Comparison of Power Approximations for Satterthwaite's Test. *Communications in Statistics – Simulation and Computation*, 24:583–593.

Muller KE, LaVange LM, Ramey SL, Ramey CT (1992). Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications. *Journal of the American Statistical Association*, 87:1209-1226.

Muller, K. E., Edwards, L. J., Simpson, S. L., & Taylor, D. J. (2007). Statistical tests with accurate size and power for balanced linear mixed models. *Statistics in Medicine,* 26:3639–3660.

Content F: Exhibits, tables, figures

Ki-67 staining
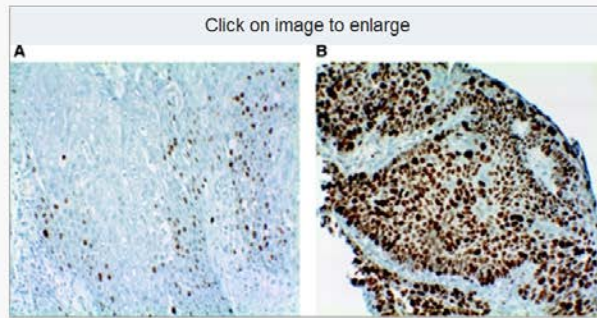

Click on image to enlarge

A   B

Figure 2

Immunohistochemical detection in primary pretreatment biopsies of: (A) a low percentage (17%) of Ki-67-positive nuclear staining; (B) a high percentage (93%) of Ki-67-positive nuclear staining.

From: Br J Cancer. 2008 October 7; 99(7): 1121–1128. Published online 2008 September 2. doi: 10.1038/sj.bjc.6604633