

## **Tutorial 2: Power and Sample Size for the Paired Sample t-test**

**Anna E. Barón, Keith E. Muller, Sarah M. Kreidler, and Deborah H. Glueck**

### **Acknowledgements:**

The project was supported in large part by the National Institute of Dental and Craniofacial Research under award NIDCR 1 R01 DE020832-01A1. The content is solely the responsibility of the authors, and does not necessarily represent the official views of the National Cancer Institute, the National Institute of Dental and Craniofacial Research, nor the National Institutes of Health.

## Tutorial 2: Power and Sample Size for the Paired Sample t-test

### 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 one-sample case.

### 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. The investigator usually hopes that the evidence in the data will *disprove* the null hypothesis. For a one-sample test of the mean, the null hypothesis  $H_0$  is a simple statement about the expected value of the mean if the sample is representative of a given reference population. (Should this sentence be about the paired sample t-Test, and not the one-sample test?)

As an example, let us say that a Phase II one-arm pre vs. post trial design will be used to examine changes in Ki67 (a measure of cell proliferation, % of cells staining for Ki-67) after administration of resveratrol, an antioxidant found in the skins of grapes and other fruits, in persons at high risk of head and neck cancer. Ki-67 is measured in a tissue biopsy before the administration of resveratrol, and again 6 months after. The goal is to compare mean Ki-67 before resveratrol,  $\mu_{\text{Ki-67, pre}}$ , to mean Ki-67 after resveratrol,  $\mu_{\text{Ki-67, post}}$ . The null hypothesis is that there is no change in Ki-67, i.e.  $H_0: \mu_{\Delta\text{Ki-67}} = \mu_{\text{Ki-67, pre}} - \mu_{\text{Ki-67, post}} = 0\%$ . An opposing, or alternative hypothesis  $H_1$  is then stated, contradicting  $H_0$ , e.g.  $H_1: \mu_{\Delta\text{Ki-67}} = \mu_{\text{Ki-67, pre}} - \mu_{\text{Ki-67, post}} \neq 0\%$ ,  $H_1: \mu_{\Delta\text{Ki-67}} = \mu_{\text{Ki-67, pre}} - \mu_{\text{Ki-67, post}} < 0\%$ , or  $H_1: \mu_{\Delta\text{Ki-67}} = \mu_{\text{Ki-67, pre}} - \mu_{\text{Ki-67, post}} > 0\%$ . If the mean change from the sample is sufficiently different from 0% a decision is made to *reject the null* in favor of the specific alternative hypothesis. If the sample mean 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  $\alpha$ . The probability of failing to rejecting a false null hypothesis, called Type II Error, is denoted by  $\beta$ . The possible outcomes of a hypothesis test are summarized in the table below along with the conditional probabilities of their occurrence.

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

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 ( $\alpha$ , the level of significance of a hypothesis test) is kept low, usually 0.05 or less, and the power of the test ( $1-\beta$ , 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 include variability as summarized by  $\sigma_{\Delta}^2$ , where  $\sigma_{\Delta}$  is the standard deviation of change in the population;

and the difference between the mean change under the null hypothesis and the mean change under a specific alternative, e.g.  $\mu_{\Delta Ki-67} = 0\%$  vs.  $\mu_{\Delta Ki-67} = 20\%$ . In practice,  $\sigma_{\Delta}$  can be difficult to specify. Previous work reported in the literature is a good source of information about variability of the change, 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 means 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,  $1-\beta$ :

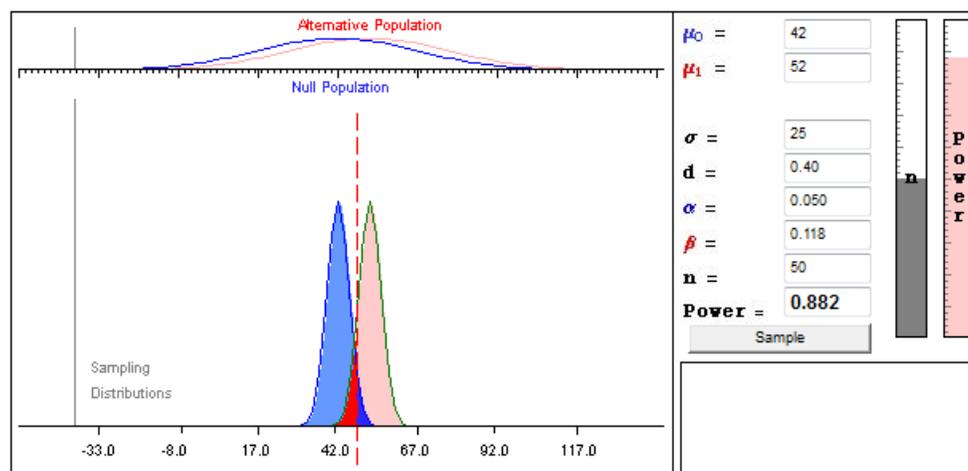
- Increase: Required Sample Size,  $n$ ; Detectable Difference, between the mean under  $H_0$  and  $H_1$ ,  $|\mu_0 - \mu_1|$ ; Level of significance,  $\alpha$ ;
- Decrease: Variability,  $\sigma_1^2$  and  $\sigma_2^2$

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,  $\alpha$
- Decrease: Variability,  $\sigma_1^2$  and  $\sigma_2^2$ ; power,  $1-\beta$

The applet available through the link below illustrates these relationships:

[http://wise.cgu.edu/power\\_applet/power.asp](http://wise.cgu.edu/power_applet/power.asp)



An additional consideration for power analysis and sample size estimation is the direction of the alternative hypothesis. A one-sided alternative hypothesis, e.g.  $H_1: \mu_{\Delta Ki-67} > 20\%$ , 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 mean change in Ki-67 in a single population. The outcome variable is continuous (change in % staining) with an unknown but finite variance of the change,  $\sigma_{\Delta}^2$ . 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. The *a priori* stated hypothesis regards the mean change of Ki-67 and whether the mean is 0%. These are the features that make the example suitable for a paired-sample t-test.

In carrying out a paired-sample t-test we make the assumption that the individual change values are randomly sampled from a well-characterized population and 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. In addition, we assume an approximately normal underlying distribution of change in Ki-67 with resveratrol. This assumption is harder to verify. For large samples the t-distribution will be a good approximation to the t-statistic for the data, but in small to moderate sized samples the approximation will not be as good.

### Paired-sample t-test: inference about the mean change $\mu_{\Delta}$ when the variance of the change $\sigma_{\Delta}^2$ is unknown

To test the hypothesis about the mean change in Ki-67, the standardized value of the sample mean change,  $\bar{X}_{\Delta Ki-67}$ , becomes the test statistic,  $t$ . We assumed the  $t$  value to follow a t distribution with  $n-1$  degrees of freedom, i.e. the sample size minus 1:

$$t = \frac{\bar{X}_{\Delta Ki-67} - \mu_{\Delta_0}}{s_{\Delta Ki-67} / \sqrt{n}} \sim t_{n-1}$$

where  $\mu_{\Delta_0}$  is the mean change under  $H_0$ , and  $s_{\Delta Ki-67}$  is the sample standard deviation of the mean change in Ki-67. To test the hypothesis  $H_0: \mu_{\Delta Ki-67} = 0\%$  vs.  $H_1: \mu_{\Delta Ki-67} \neq 0\%$ , we reject  $H_0$  if:

$$\bar{X}_{\Delta Ki-67} > 0\% + t_{n-1, 1-\alpha/2} \frac{s_{\Delta Ki-67}}{\sqrt{n}} \quad \text{or} \quad \bar{X}_{\Delta Ki-67} < 0\% - t_{n-1, 1-\alpha/2} \frac{s_{\Delta Ki-67}}{\sqrt{n}},$$

where  $t_{n-1, 1-\alpha/2}$  is the  $(1-\alpha/2) \times 100$ th percentile of the t distribution with  $n-1$  degrees of freedom.

## The paired t-test as a General Linear Model (GLM)

### Univariate test

A more general characterization of the paired 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  $n \times 1$  matrix of Ki-67 change values,  $X$  is an  $n \times 1$  matrix of 1s,  $\beta$  is a  $1 \times 1$  matrix of the unknown regression coefficient (in this case the estimated mean change in Ki-67), and  $\varepsilon$  is an  $n \times 1$  matrix that represents the random deviation of a single Ki-67 change value from the mean change. The  $Y$  and  $X$  matrices are shown below:

$Y_{\Delta}$	$X$
$Y_{1Pre} - Y_{1Post}$	1
$Y_{2Pre} - Y_{2Post}$	1
.	.
.	.
$Y_{nPre} - Y_{nPost}$	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,

$$\sigma_{Y|X}^2 = \sigma_{Y|X=1}^2 = \sigma_{Y|X=2}^2 = \dots = \sigma_{Y|X=x}^2$$

- **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  $\beta$ ,  $\hat{\beta}$ , using the method of least squares. To test the hypothesis  $H_0: \mu_{\Delta Ki-67} = 0\%$ , we use the standardized value of  $\hat{\beta}$ , which is distributed

as a t with n-1 degrees of freedom:  $t = \frac{\hat{\beta} - 0\%}{\frac{s_{y/x}}{\sqrt{n}}} \sim t_{n-1}$ , where  $s_{y/x}$  is the square root of

the residual variance from the GLM.

### Test based on repeated measures

Alternatively, still using the GLM, the hypothesis can be tested using a repeated measures design. In this design Y is an n x 2 matrix, where the first column contains the pre-treatment and the second column contains the post-treatment measures. X is an n x 1 matrix of 1s,  $\beta$  is a 1 x 2 matrix of the unknown regression coefficients, in this case the estimated pre- ( $\hat{\beta}_1$ ) and post-treatment ( $\hat{\beta}_2$ ) Ki-67 means, and  $\varepsilon$  is an n x 2 matrix that represents the random deviations of single Ki-67 observations from the pre-treatment mean (column 1) and from the post-treatment mean (column 2). The Y and X matrices are shown below:

Y <sub>pre</sub>	Y <sub>post</sub>	X
Y <sub>1Pre</sub>	Y <sub>1Post</sub>	1
Y <sub>2Pre</sub>	Y <sub>2Post</sub>	1
.	.	.
.	.	.
Y <sub>nPre</sub>	Y <sub>nPost</sub>	1

The hypothesis, e.g.  $H_0: \mu_{\Delta\text{Ki-67}} = \mu_{\text{Ki-67, pre}} - \mu_{\text{Ki-67, post}} = 0\%$ , is tested by contrasting the

elements of  $\beta$ :  $t = \frac{(\hat{\beta}_1 - \hat{\beta}_2)}{\sqrt{s_{\hat{\beta}_1}^2 + s_{\hat{\beta}_2}^2 - 2\text{cov}(\hat{\beta}_1, \hat{\beta}_2)}} \sim t_{n-1}$ , where  $s_{\hat{\beta}_1}^2$  and  $s_{\hat{\beta}_2}^2$  are the estimated

variances of  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , respectively, and  $\text{cov}(\hat{\beta}_1, \hat{\beta}_2)$  is the covariance between them.

For a more general presentation on power and sample size analysis for repeated measures, refer to the tutorial on Repeated Measures.

## Content B: How to obtain inputs for power analysis and sample size estimation

### B.1 Inputs for Power analysis

A power analysis for mean change consists of determining the achievable power for a specified difference between the mean change under the stated  $H_0$  and under the stated  $H_1$ , sample size, standard deviation of the change, and  $\alpha$ -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  $\alpha$  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, we look to a study examining change in Ki-67 which reports a standard deviation of 14% for change in Ki-67. 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,  $\rho$ , between the pre- and post- measures in order to obtain an estimate for the standard deviation of change. A value of 0 for  $\rho$  means the pre- and post- measures are uncorrelated while a value of 1 means they are perfectly correlated. Low values of  $\rho$  would lead to conservative power and sample size estimates while large values of  $\rho$  would lead to liberal ones. For a specific  $\rho$  value, we obtain the standard deviation of change as :

$$s_{\Delta} = \sqrt{s_{pre}^2 + s_{post}^2 - 2\rho s_{pre}s_{post}}$$
, where  $s^2$  and  $s$  are the sample variance and standard deviation, respectively. Rarely is the variance reported in a publication. If reported, the researchers usually obtain the standard deviation as the square root of the variance.

Even when using information from large studies in the literature, 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 about uncertainty refer to the *Tutorial 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 smallest difference that would be considered scientifically important. For the paired-sample Ki-67 example, we consider a difference of 20% 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  $\alpha$ -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  $\alpha$  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](http://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 Guided Study Design mode

Guided Study Design mode is suggested for most users.

##### 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.
<input type="button" value="Select"/>	<input type="button" value="Select"/>	<input type="text"/> <input type="button" value="Browse.."/>

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

**Calculate**

**Start**

- Solving For
- Type I Error
- Sampling Unit
- Responses
- Hypothesis
- Means
- Variability
- Options

### Introduction

The GLIMMPSE wizard will guide you through several steps to calculate power or sample size.

Use the forward and back arrows to navigate through the wizard. You may save your work at any time by clicking the "Save Design" link at the lower right of the screen. The "Cancel" link, also at the lower right of the screen, allows you to cancel your current work and begin a new study design. The help manual may be accessed by clicking the "Help" link.

General steps for a power analysis are listed on the left hand side of the screen. We will ask you to specify:

- The Type I error rate
- The independent and dependent variables
- The primary study hypothesis of interest
- Choices for group means
- Choices for standard deviations and correlations for study outcomes
- The statistical test and additional display options

Click the forward arrow to begin.

Help Save Design Cancel

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.

Calculate

**Start**

- ✓ Solving For
- Type I Error

Sampling Unit

Responses

Hypothesis

Means

Variability

Options

### Would you like to solve for power or sample size?

To begin your calculation, please indicate whether you would like to solve for power or total sample size.

If you have a rough idea of the number of research participants you will be able to recruit, then solving for power may be more beneficial.

If you have fewer restrictions on recruitment and would like to ensure a well-powered study, then solving for sample size is likely to be more useful.

Power  
 Total Sample Size

◀▶    Help    Save Design    Cancel

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.

Calculate

**Start**

- ✓ Solving For
- ✓ Type I Error

Sampling Unit

Responses

Hypothesis

Means

Variability

Options

### Type I Error

A Type I error occurs when a scientist declares a difference when none is actually present. The Type I error rate is the probability of a Type I error occurring, and is often referred to as  $\alpha$ . Type I error rates range from 0 to 1. The most commonly used values are 0.01, 0.05, and 0.1.

Enter each Type I error value into the text box and click "Add". You may enter up to 5 values. To remove a value, select the value in the list box and click the "Delete" button.

Type I Error Values:  Add Delete

- .01
- .05
- .1

◀▶    Help    Save Design    Cancel

Read the information on the *Sampling Unit: Introduction* screen and click next to move to the next screen.

For a one-sample test, there is one fixed predictor with one level. On the *Study Groups* screen, select *One group* and click the forward arrow to move to the next screen.

**Calculate**

**Study Groups**

Describe the predictors which assign independent sampling units into groups, such as gender or treatment. If the study includes only one group, select the "One group" button. If the study includes multiple groups, select the "Multiple groups" button.

One group  
 Multiple groups

Start

**Sampling Unit**

✓ Study Groups  
✓ Covariate  
✓ Clustering  
✓ Sample Size

Responses

Hypothesis

Means

Variability

Options

◀▶ Help Save Design Cancel

Since no covariate and no clustering will be used, skip the *Covariate* and the *Clustering* screens.

Use the *Sample Size* screen to specify the size of the smallest group for your sample size(s). Although only one sample is being used for this example, entering multiple values for the smallest group size allows you to consider a range of total sample sizes.

**Calculate**

Start

**Sampling Unit**

- ✓ Study Groups
- ✓ Covariate
- ✓ Clustering
- ✓ **Sample Size**

Responses

Hypothesis

Means

Variability

Options

### 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 or 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

◀▶ Help Save Design Cancel

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

The *Response Variable* screen allows you to enter the response variable(s) of interest. In this example, the response variable is Ki-67. Click the forward arrow when you have finished entering your value(s).

Calculate

Start

Sampling Unit

**Responses**

- ✓ Response Variables
- ✓ Repeated Measures

Hypothesis

Means

Variability

Options

### Response Variables

Enter the response variables in the table below. For example, in a study investigating cholesterol-lowering medication, the response variable could be HDL, LDL, and total cholesterol.

Note that repeated measurement information will be addressed on the next screen.

Response Variables:

Ki-67

◀▶ Help Save Design Cancel

The *Repeated Measures* screen should be skipped, as there are no repeated measures for this example.

Read the *Hypothesis: Introduction* screen, then click the forward arrow to move to the *Hypothesis* screen.

The *Hypotheses* screen allows you to enter the known mean values for your primary hypothesis. For this example, enter 0.

**Calculate**

Start

Sampling Unit

Responses

**Hypothesis**

✓ Hypothesis

Means

Variability

Options

### Hypotheses

The list below shows the hypotheses which are available for the current study design. Select the hypothesis which most closely resembles your primary study hypothesis. Trends within an interaction hypothesis are specified in the "Interaction" tab. This hypothesis will be used to determine power for your study.

The tab highlighted in "white" indicates the currently selected hypothesis. For more information about the type of hypothesis, click the magnifying glass icon.

Grand mean

Enter the known mean values for each response below.

KI-67 0

Help Save Design Cancel

Read the information on the *Means: Introduction* screen, then click the forward arrow. The *Mean Differences* screen allows you to specify the difference between the null and alternative hypothesis means. For this example, enter 20%, or 20, in the screen below, then click the forward arrow to continue.

**Calculate**

Start

Sampling Unit

Responses

Hypothesis

**Means**

✓ Mean Differences

✓ Beta Scale Factors

Variability

Options

### Means

The table below shows the mean values for each outcome within each study subgroup. The study subgroups are listed along the left hand side of the table, and the outcomes are listed across the top.

Enter the mean values you expect to observe for each outcome within each study subgroup. The table should contain at least one value that is non-zero. Also, at least two subgroups should have means which differ by a scientifically meaningful amount.

Sample	KI-67
Sample	20

Help Save Design Cancel

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. GLIMMPSE allows this to be from 0.5x to 2x the stated difference, e.g. from 10% to 20% to 40%. Click Yes, then click on the forward arrow to continue.

**Calculate**

Start

Sampling Unit

Responses

Hypothesis

**Means**

✓ Mean Differences

✓ **Beta Scale Factors**

Variability

Options

**Flexible Means**

Power and sample size results will change depending on the mean values specified on the previous screen. It is not possible to know exact values for these means until the experiment is observed. To account for this uncertainty, it is common to calculate power for the mean values as specified, the mean values divided by 2, and the mean values multiplied by 2.

Yes, include power calculations for the mean values as entered, the mean values divided by 2, and the mean values multiplied by 2.

Help Save Design Cancel

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

The *Within Participant Variability* screen allows you to specify the expected variability in terms of standard deviation of the outcome variable. For this example, the standard deviation of the change in Ki-67 was estimated to be 14%. Enter 14 and click Next to continue.

**Calculate**

Start

Sampling Unit

Responses

Hypothesis

Means

**Variability**

✓ **Within Participant Variability**

✓ Sigma Scale Factors

Options

**Variability and Correlation within an Individual Research Participant**

For a given research participant, responses vary across response variables and across repeated measurements. The amount of variability can dramatically impact power and sample size. Click on each of the tabs below to describe the variability you expect to observe for the response variables and each within-participant factor.

Responses

Enter the standard deviation you expect to observe for each response. Note that GLIMMPSE currently assumes that the standard deviation is constant across repeated measurements.

KI-67

Help Save Design Cancel

The true variability in change in Ki-67 is also uncertain. To see how power varies with the assumed s.d., you can allow the s.d. to vary over a reasonable range. GLIMMPSE allows this to be from 0.5x to 2x the stated s.d., e.g. from 7% to 14% to 28%. Click on Yes, then click the forward arrow to continue.

**Calculate**

Start

Sampling Unit

Responses

Hypothesis

Means

**Variability**

✓ Within Participant Variability

✓ **Sigma Scale Factors**

Options

**Flexible Variability**

On the previous screens, you entered standard deviations and correlations. GLIMMPSE has used these values to calculate a covariance matrix which describes the overall variability.

Changes in variability can dramatically affect power and sample size results. It is not possible to know the variability until the experiment is observed. To account for this uncertainty, it is common to calculate power or sample size for alternative values for variability.

By clicking the box below, GLIMMPSE will calculate power using the calculated covariance matrix, the covariance matrix divided by 2, and the covariance matrix multiplied by 2.

Yes, include power for the covariance matrix, the covariance matrix divided by 2, and the covariance matrix multiplied by 2.

Help Save Design Cancel

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

For the paired-sample test of mean change, the available tests in GLIMMPSE yield equivalent results. Click on any one of the tests and then click the forward arrow to continue.

**Calculate**

Start

Sampling Unit

Responses

Hypothesis

Means

Variability

**Options**

✓ **Statistical Test**

✓ Confidence Intervals

✓ Power Curve

**Statistical Tests**

Select the statistical tests to include in your calculations. For study designs with a single outcome, power is the same regardless of the test selected.

Note that only the Hotelling-Lawley Trace and the Univariate Approach to Repeated Measures are supported for designs which include a baseline covariate.

Hotelling-Lawley Trace

Pillai-Bartlett Trace

Wilks Likelihood Ratio

Univariate Approach to Repeated Measures with Box Correction

Univariate Approach to Repeated Measures with Geisser-Greenhouse Correction

Univariate Approach to Repeated Measures with Huynh-Feldt Correction

Univariate Approach to Repeated Measures, uncorrected

Help Save Design Cancel

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.

Calculate

Start

Sampling Unit

Responses

### Power Curve Options

You may optionally create a power curve image for your results by unchecking this checkbox. Then select the values you would like to display on the power curve by selecting the appropriate options below.

I do not want to create a power curve.

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 curve.

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

Calculate

Start

Sampling Unit

Responses

Hypothesis

Means

Variability

Options

- ✓ Statistical Test
- ✓ Confidence Intervals
- ✓ **Power Curve**

### Power Curve Options

You may optionally create a power curve image for your results by unchecking this checkbox. Then select the values you would like to display on the power curve by selecting the appropriate options below.

I do not want to create a power curve.

- Select the quantity to display on the horizontal axis of the power curve (the vertical axis will display the power value).  
 Variability Scale Factor ▼
- Add data series to the plot. Select values for each variable below. Click add to include sample size values matching these criteria as a data series on the plot. To remove a data series, highlight it in the list box and click "Remove data series".

Total Sample Size	<span style="border: 1px solid #ccc; padding: 2px;">5 ▼</span>
Regression Coefficient Scale Factor	<span style="border: 1px solid #ccc; padding: 2px;">1 ▼</span>
Statistical Test	<span style="border: 1px solid #ccc; padding: 2px;">Univariate Approach to Repeated Measures, uncorrected ▼</span>
Type I Error	<span style="border: 1px solid #ccc; padding: 2px;">0.05 ▼</span>
Data Series Label	<span style="border: 1px solid #ccc; padding: 2px;">Tutorial 2</span>

Add
Delete

Tutorial 2: Sample Size=5 Test=Univariate Approach to Repeated Measures, uncorrected Regr

⏪ ⏩
Help Save Design Cancel

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.

**Calculate**

**Start**

- ✓ Solving For
- ✎ Desired Power
- ✓ Type I Error

Sampling Unit

Responses

Hypothesis

Means

Variability

Options

### Would you like to solve for power or sample size?

To begin your calculation, please indicate whether you would like to solve for power or total sample size.

If you have a rough idea of the number of research participants you will be able to recruit, then solving for power may be more beneficial.

If you have fewer restrictions on recruitment and would like to ensure a well-powered study, then solving for sample size is likely to be more useful.

Power

Total Sample Size

◀▶ Help Save Design Cancel

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

**Calculate**

**Start**

- ✓ Solving For
- ✓ **Desired Power**
- ✓ Type I Error

Sampling Unit

Responses

Hypothesis

Means

Variability

Options

### Power Values

Enter the desired power values in the list box below. Power values are numbers between 0 and 1. Higher values correspond to a greater likelihood of rejecting the null hypothesis. Common values are 0.8 or 0.9, although 0.9 or higher is usually preferred.

Type each value into the list box and click "Add". To remove an item, highlight the value and click the "Delete" button.

Power Values:  Add Delete

- .8
- .9
- .95

◀▶ Help Save Design Cancel

The remaining screens are filled in as above for power.

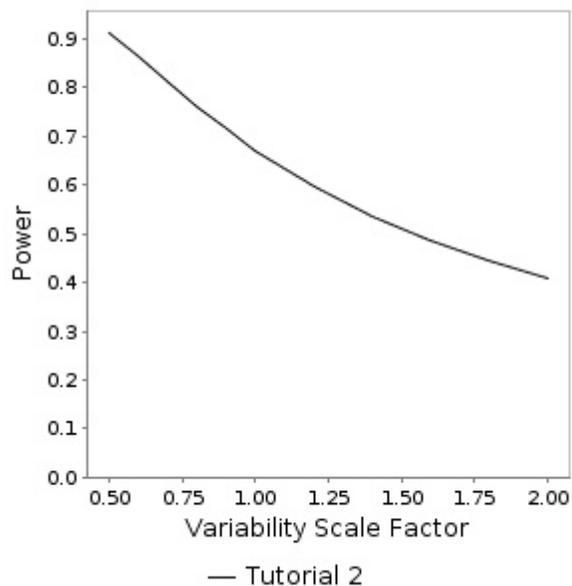
## Content D: Interpret the results

### Power analysis interpretation

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 differences in mean change in Ki-67 and standard deviations of change in Ki-67 with an  $\alpha$ -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 much 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



## 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:

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

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

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

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

$$\Theta_0 = [0.0000]$$

$$\Sigma_E = [196.0000]$$

For the paired-sample t-test, GLIMPSE works with the matrices listed above in making the computations. Since only a single mean is being tested the matrices are all of dimension 1 x 1. The  $\Theta_0$  matrix represents the mean under  $H_0$  and the  $\Sigma_e$  matrix represents the between-subject variance. The remaining matrices  $E_s(\mathbf{X})$ ,  $\mathbf{C}$  and  $\mathbf{U}$  are scalars with a value of 1.

## References

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.

Seoane J, Pita-Fernandez S, Gomez I et al (2010). Proliferative activity and diagnostic delay in oral cancer. *Head and Neck*, 32:1377–1384.

## Content F: Exhibits, tables, figures

### Ki-67 staining

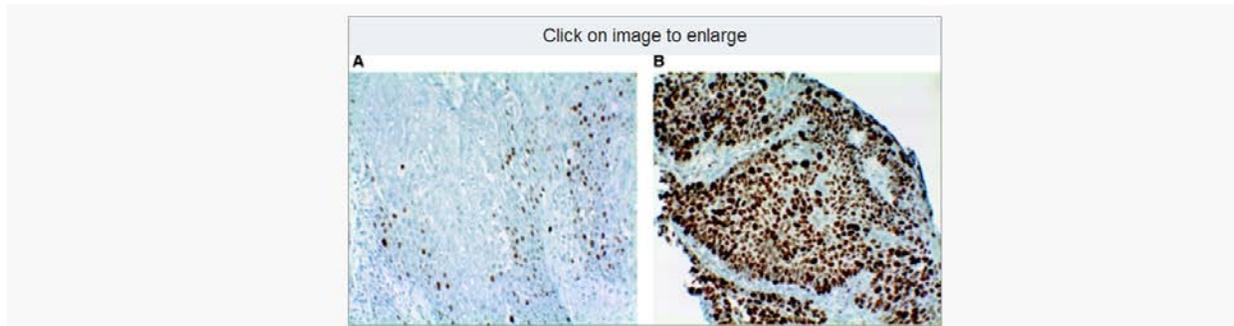


Figure 2

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

[Copyright/LicenseRequest permission to reuse](#)

### Normal curves for power analysis and sample size estimation

Consider the distribution of  $\bar{X}$  if  $\mu = \mu_0$  and if  $\mu = \mu_1$

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{n}\right)$$

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

