

Methods: Mind the Gap

Webinar Series

Choosing Sample Sizes for Multilevel and Longitudinal Studies Analyzed with Linear Mixed Models



Presented by:

Keith E. Muller, Ph.D.

Deborah H. Glueck, Ph.D.



National Institutes of Health
Office of Disease Prevention

Choosing Sample Sizes for Multilevel and Longitudinal Studies Analyzed with Linear Mixed Models

Keith E. Muller

Department of Health Outcomes and Biomedical Informatics
University of Florida College of Medicine

and

Deborah H. Glueck

Department of Pediatrics
University of Colorado School of Medicine

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgement of Support

- R01-GM121081, Methods and Software for Lifecourse Epidemiology Data and Sample Size Analysis, NIGMS, Glueck, Muller, and Dabelea MPI, 08/16 - 06/21.
- R25-GM111901, A Master Course on Power for Multilevel and Longitudinal Health Behavior Studies, OBSSR and NIGMS, Muller and Glueck MPI, 08/14 - 06/18.

Other Team Members

- The version of the software that will be discussed was primarily developed by Alasdair Macleod, Elizabeth Litkowski, Qian Li, and Dr. Jiang Bian. Many others created previous versions (most notably Dr. Sarah Kreidler).
- The educational technology that will be discussed was primarily developed by Dr. Albert Ritzhaupt, Dr. Max Sommer, Dr. Natercia Lourinho Moura do Valle, and Michelle Zamperlini. Many others helped create the original course content (most notably Jessica R. Shaw).

Outline

- An original study is *reproducible* if a new sample of data collected under the original protocol is likely to lead to the same conclusions as the original study.
- To be reproducible, a study must have good statistical power.
- Advances in power theory make it accurate and fast to compute power for linear mixed models.
- Free point-and-click software makes it easy to compute power.
- Planning a longitudinal clinical trial illustrates the process.
- No power software covers everything. There is much room for future methodological research, and extensions to the software.

Ethics of Sample Size

- Scientific knowledge is public and reproducible.
- If the sample size is too small, a study may be inconclusive and waste resources.
- If the sample size is too large, then a study may expose too many participants to possible harm.
- Ethical designs have sample size that is large enough and not too large.

Choosing the Correct Sample Size Requires Alignment

- Conscientious study planners routinely align data analysis with study design and study design with scientific goals.
- Aligning power analysis with data analysis is not routine because it requires:
 - statistical theory,
 - accurate computational forms, and
 - easy to use software.
- Until recently, linear mixed models lacked all three.

Power Theory for Many Linear Mixed Model Hypotheses

- The core statistical theory of power analysis for a broad swath of general linear mixed models used in health research is in Muller et al. (1992) and Chi et al. (2019).
- With no missing data, the mixed models of interest give a Wald statistic that is equivalent to the Hotelling-Lawley trace.
- With no missing data, the power results are exact in many important cases.
- References in the bibliography give a path to the underlying theory and the equally important computational methods.

Why not Run Simulations?

- Computer simulations take longer because they take time to:
write and debug the simulation,
validate the simulation,
iteratively change the design, and
generate enough points for a power curve.
- Single programmers with a deadline are vulnerable to:
mistakes in initial assumptions,
mistakes in coding,
mistakes in modules written by others and
an inability to validate with independent target values.

We Give You Easy-to-Use and Validated Power Software

- We have built a point-and-click application for web browsers.
- The wizard-style software is free, and copyleft under the GNU public license so that it will always be free.
- A sequence of dialog boxes leads a user through well-defined steps.
- The software is validated to industry standards.
- Software development was funded by five NIH grants to our group.

Features of GLIMMPSE Software

- Free access is available to anyone.
- A user only needs a web browser.
- No programming expertise is needed.
- No mathematical notation is used.
- A partial or complete study design can be saved for later use.
- The software works on computers, tablets and smartphones.
- The computations are done on a remote server.

GLIMMPSE Gives Power and Sample Size for Many Designs

- Single level designs
 - Cluster- or group-randomized trials, like the TAAG study.
 - Observational studies with clustering, like the ECHO consortium of cohorts.
- Randomized block (randomized complete block) designs, like the CAPS gardens trial.
- Multilevel observational or randomized designs like the ENRICH randomized controlled trial, with participant within nurse and nurse within clinic.
- Multivariate designs, like a study of cord blood chemokines in children in the Healthy Start cohort.
- Studies with both clustering or grouping and repeated measures, like a study of micronutrients throughout pregnancy for women who are cohort members in the ECHO consortium.

Our Track Record

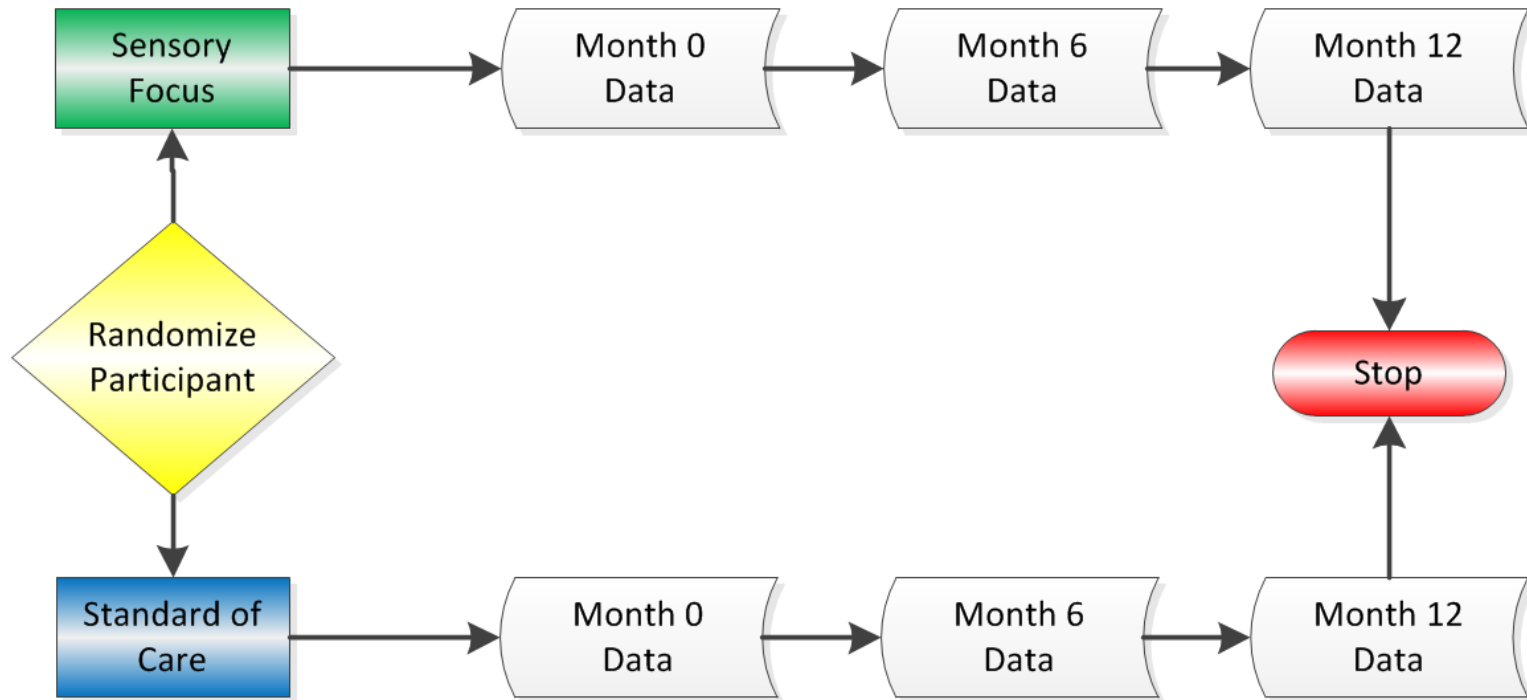
- Users cited our software in 288 peer-reviewed articles since 2013, with exponential growth.
- Our software was cited in manuscripts which also cited over \$100,000,000 in NIH grants.
- Users citing our software are funded by 12 NIH Institutes and Centers.
- Manuscripts citing our software encompass all 16 Medical Subject Headings.
- In the last year, Google Analytics counted more than 10,000 website visits.
- In the last year, some 2,700 power and sample size analyses were completed with GLIMMPSE.
- Over a sampling of 82 days, 819 users created logins, with 138 (17%) from US universities.
- Only 20% of users (N =169) had foreign country codes.

An Example Power Analysis: Proposed Randomized Trial

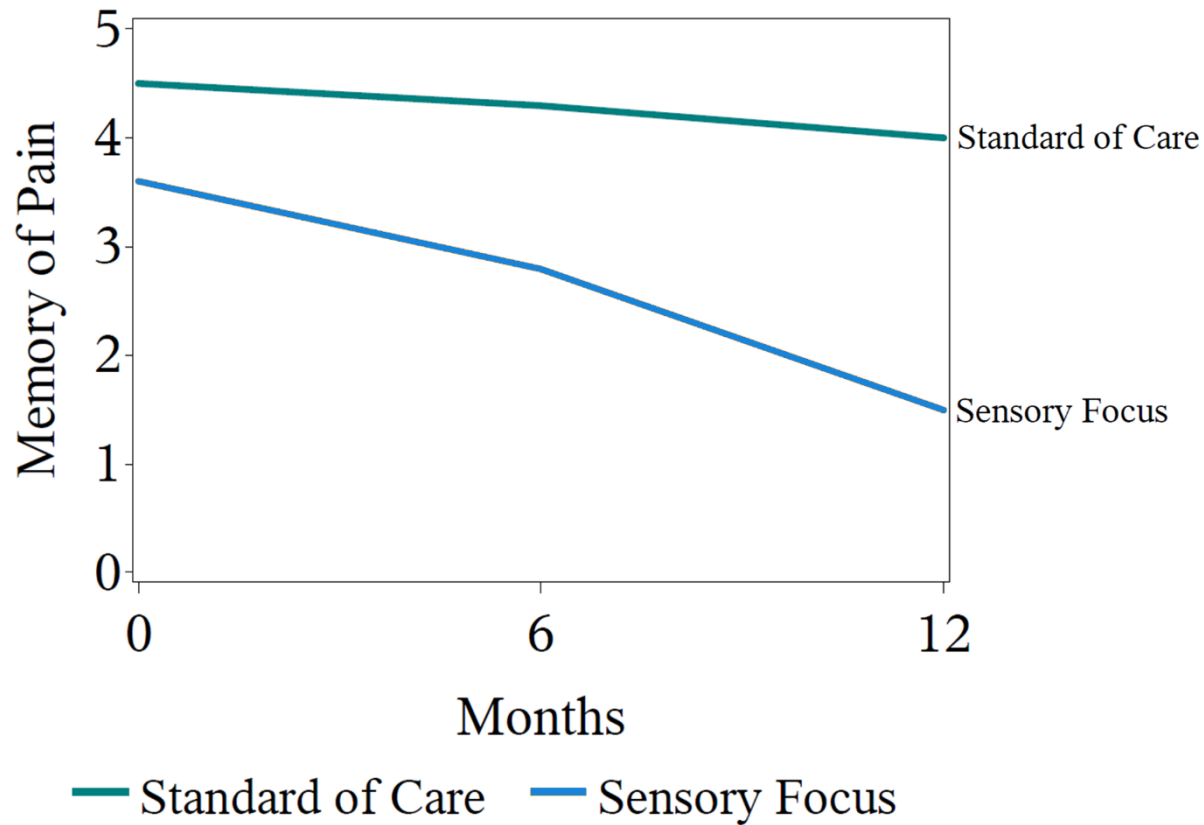
- The primary trial outcome will be the memory of pain reported after a root canal.
- The memory of pain will be recorded immediately, at 6 months and 12 months, so each participant will have 3 repeated measures.
- Participants will be included if they have a coping style with low perceived control and high desired control.
- Participants will be randomized to one of two treatment arms: **sensory focus** or **standard of care**.

(Proposed study based on Logan et al., 1995)

Proposed Randomized Trial



Hypothesized Memory of Pain Pattern



Study Population for Proposed Randomized Trial

- Participants will be included if they have a coping style with low perceived control and high desired control.
- Results from the previous study predict 30 patients / week.
- The previous study had a 40% consent rate.

Inputs for Power Analysis

- Type I error rate:
- Desired power:
- Loss to follow-up:

Inputs for Power Analysis

- Type I error rate: 0.05
- Desired power:
- Loss to follow-up:

Inputs for Power Analysis

- Type I error rate: 0.01
- Desired power: 0.85, 0.90, 0.95
- Loss to follow-up:

Inputs for Power Analysis

- Type I error rate: 0.01
- Desired power: 0.85, 0.90, 0.95
- Loss to follow-up: 25%

Create a Study Design

- Click the “New Study” button to start a new power and sample size analysis, or
- Click the “Upload” button to upload a json file with a previous study design that you have saved.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

Design a Study

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

New Study

Start a new design.

Upload

You have previously used GLIMMPSE and wish to work on a saved design.

For a new study: Add title here.



Please pick a concise title for the study:

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

Untitled Study: Study title

Progre

Solving for

Click the “Power” or “Sample Size” button.
In this case, we are solving for sample size.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Solve for

Progress ○ Help ? Save ↓

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 solve for power.

If you have few restrictions on recruitment then you may wish to solve for sample size.

Power

Sample Size

Desired Power

On the “Target Power” screen, we indicate that we want power values of 0.85, 0.90 and 0.95. Type each desired power value and hit enter after each.

◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size

module2: Target power Progress Help

Please choose one or more power values, for which you wish to calculate minimum sample size.

All target power values must be between 0 and 1, exclusive.

| Target Power | remove |
|--------------|----------------------------------|
| 0.85 | <input type="button" value="x"/> |
| 0.9 | <input type="button" value="x"/> |
| 0.95 | <input type="button" value="x"/> |

You can specify multiple values of interest for the same GLIMMPSE design.


Hypothesis Test

We now move to select the hypothesis test of interest.
We click the Hotelling-Lawley trace (and align with the mixed model Wald test).

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Statistical tests

Progress 

Please choose one or more statistical tests. If you are unsure which to pick, we recommend the Hotelling Lawley Trace test due to its equivalence to a mixed model test.

- Hotelling Lawley Trace
- Pillai-Bartlett Trace
- Wilks Likelihood Ratio
- Box Corrected
- Geisser-Greenhouse Corrected
- Huynh-Feldt Corrected
- Uncorrected

Now we choose the Type I error rate.

We type in 0.05 and click the plus sign to add this value.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Type I error rates

A Type I error occurs when a scientist declares a difference when none is present in the population. The Type I error rate is the probability of a false positive, and is often referred to as α (alpha). A Type I error rate can range from 0 to 1. Although the most common recommendation is 0.05.

Type I Error Rate

remove

0.05

One Outcome Variable

Enter outcome variables here.

Then, click the plus sign to add the outcome to the design.


◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size


module2: Outcomes Progress Help

Enter the name of each outcome variable one at a time in the underlined space below. For example, in a study investigating cholesterol-lowering medication, the outcome variables could be HDL, LDL, and total cholesterol.

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

Please name the one or more outcomes.



| | |
|----------------|---|
| <u>Outcome</u> | remove |
| mem_pain |  |

One Repeated Measures Dimension

This is a longitudinal study with three repeated measurements of the outcomes.

GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measures

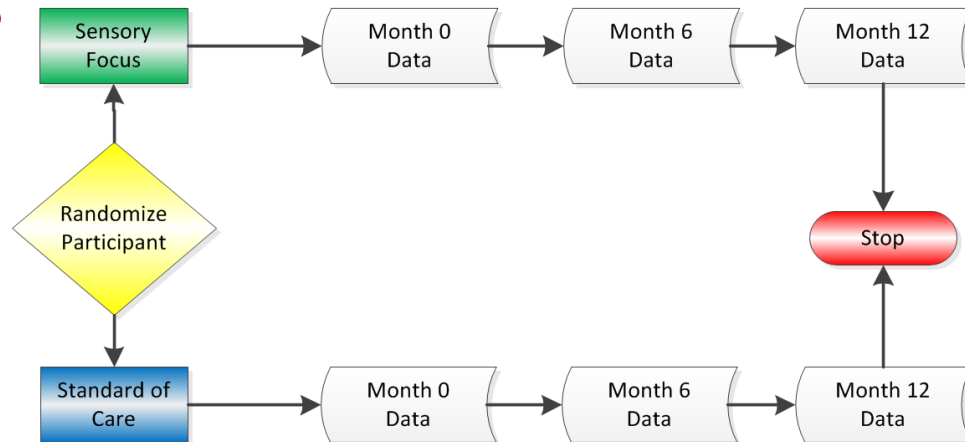
Progress Help Save

GLIMMPSE allows you to define within-participant factors, specified as repeated measures. An independent sampling unit provides one or more observations such that observations from one unit are statistically independent from any other distinct unit while observations from the same unit may be correlated. Repeated measures are present when a response variable is measured on each independent sampling unit on two or more occasions or under two or more conditions. The values of the repeated measures (that is, the levels of the within-participant factors) distinguish the occasions or conditions.

If the study includes repeated measures, click "Add Repeated Measure" and follow the prompts.

You may specify up to 5 repeated measures. Each repeated measure you add will apply to each outcome you specified on the previous page.

Define Repeated Measure



Here, we have named the dimension as time.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measures

Progress  Help 

What is the name of the dimension you will be measuring?

The text entered in the "Dimension" text box indicates the dimension over which measures were taken (e.g. time, days, locations, etc.). The choice of "Type" indicates whether the repeated measures are numeric (e.g. time), or categorical (e.g. arm, leg, hand).

Dimension:

time

Cancel



Next: Type

Here, we define the type of data that our dimension “time” represents.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measures

Progress  Help  Save  Home 

What type of data is time?

Categorical

Numeric

Cancel

Back

Next: No. Measurements

Enter the number of repeated measurements of the response variable.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measures

Progress  Help 

Number of measurements of time?

3

You must have between 2 and 10 repeats (inclusive)

Cancel

Back

Next: Spacing

Here, we enter the spacing values.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measures

Progress Help

Spacing

If the repeated measures are numeric, the spacing values must be unique nonnegative integers, in ascending order.

Set values myself Select values by series

Measurement #1 at

1

Measurement #2 at

2

Measurement #3 at

3

Cancel

Back

Update repeated measure

This screen summarizes the information entered for the repeated measures. GLIMMPSE can measure a given response variable up to 10 times.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measures

Progress  Help 

GLIMMPSE allows you to define within-participant factors, specified as repeated measures. An independent sampling unit provides one or more observations such that observations from one unit are statistically independent from any other distinct unit while observations from the same unit may be correlated. Repeated measures are present when a response variable is measured on each independent sampling unit on two or more occasions or under two or more conditions. The values of the repeated measures (that is, the levels of the within-participant factors) distinguish the occasions or conditions.

If the study includes repeated measures, click "Add Repeated Measure" and follow the prompts.

You may specify up to 5 repeated measures. Each repeated measure you add will apply to each outcome you specified on the previous page.

Define Repeated Measure

| Repeated Measure Dimension | Type | Measurements | Edit | Remove |
|----------------------------|---------|-----------------|---|---|
| time | Numeric | ["1", "2", "3"] |  |  |

No Clustering in the Example

There is no clustering in this design, so we leave the clustering screen blank. GLIMMPSE allows up to 10 levels of clustering, with equal size clusters. Each level may have any size of cluster desired.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Clustering

Progress  Help  S



An independent sampling unit provides one or more observations such that observations from one unit are statistically independent from any other distinct unit while observations from the same unit may be correlated.

In a clustered design, the independent sampling unit is a cluster, such as a community, school, or classroom. Observations within a cluster are correlated. The labels for observations within a cluster must be exchangeable. For example, child "ID" within classroom can be reassigned arbitrarily. In contrast, observations across time cannot be reassigned and should not be considered clustered observations. The common correlation between any pair of cluster members is termed the intraclass correlation or intracluster correlation.

To include clustering in the study, click "Add Clustering" and follow the prompts.

You may specify up to 10 levels of clustering.

Add Clustering



One Fixed Predictor

Here, we define the fixed predictor.

◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size

module2: Fixed predictors Progr

Each independent sampling unit has one or more observations which are statistically independent from observations from any other unit.

GLIMMPSE allows you to define fixed predictors which divide the independent sampling unit into groups. One common example of a fixed predictor is treatment, with values placebo and drug, for which the independent sampling unit is randomized to a placebo group or a drug group. Another is gender, with values male or female.

If the design has no fixed predictors, do not define any here.

Define Fixed Predictor

Now type in "treat" and hit enter. Predictor variables are the same thing as factors between independent sampling units.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Fixed predictors

Progress  Help  Save 

Please name the
predictor:

treat

Cancel **Next: Data Type**

Here, we define what type of data and move on to define the groups.

◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size

module2: Fixed predictors Progress ○ Help ? Save ↓ Home 🏠

What type of data is treat?


Nominal Continuous

Here, we define the groups.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size


module2: Fixed predictors


Progress  Help  S

Please name at
least two groups:

Groups:

 treat_group

 control

Cancel Back: Type Update predictor

Choose the Contrast

The interaction hypothesis is a between-by-within independent sampling unit hypothesis, a time-by-predictor interaction.

◆ GLIMMPSE General Linear Mixed Model Power and Sample Size

module2: Hypothesis choice

Progress Help

Each power or sample size calculation is based on selecting a specific study hypothesis. The options below show the hypotheses which are available for the current study design. Specify the hypothesis that represents your scientific question.

GLIMMPSE chooses sensible contrast matrices based on cell means coding. Should you wish to define your own contrast matrices, pick the highest order interaction and choose from the advanced options in the hypothesis components.

Select a hypothesis from the list.

| | Effects Available for Consideration | Nature of Variation |
|----------------------------------|-------------------------------------|---------------------|
| <input checked="" type="radio"/> | treat x time: Interaction | Between x Within |
| <input type="radio"/> | time: Main Effect | Within |
| <input type="radio"/> | treat: Main Effect | Between |
| <input type="radio"/> | Grand Mean | Between |

Specify means for:

Factors in Hypothesis All Factors

Standard or Custom Coefficients

“All mean differences zero” was selected to indicate the type of contrast desired.

◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size

module2: Hypothesis

Progr

What type of contrast do you wish among the means defined by your groups and repeated measures?

All mean differences zero

A parameter is a characteristic of a population. The parameters of interest are differences between groups at individual repeated measures.

The null hypothesis is that all pairwise differences between groups are the same among all pairs of repeated measures.

Show Advanced Options

This screen gives you the option to select a value different than zero for the contrast comparison constant.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Theta 0

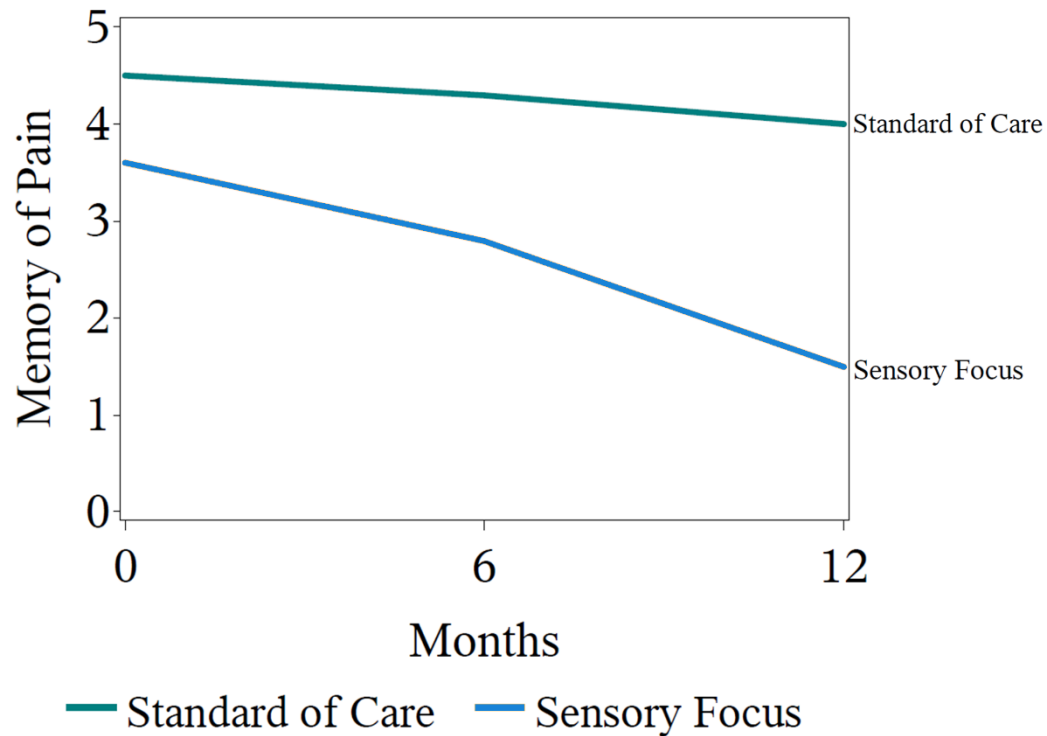
Progress 

A hypothesis compares parameters to a constant, the contrast comparison constant, θ_0 . This is almost always zero. If you choose a value other than zero, be sure that you understand that the hypothesis you define is scientifically meaningful. Also note that the description and interpretation of your hypothesis given when choosing your contrasts will be affected.

▼

Hypothesis

Time by Treatment Interaction



Fixed Predictor

Relative Group Sizes

Study participants are equally randomized to the two levels of treatment of equal size: sensory focus (treat) and standard of care (control).

◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size

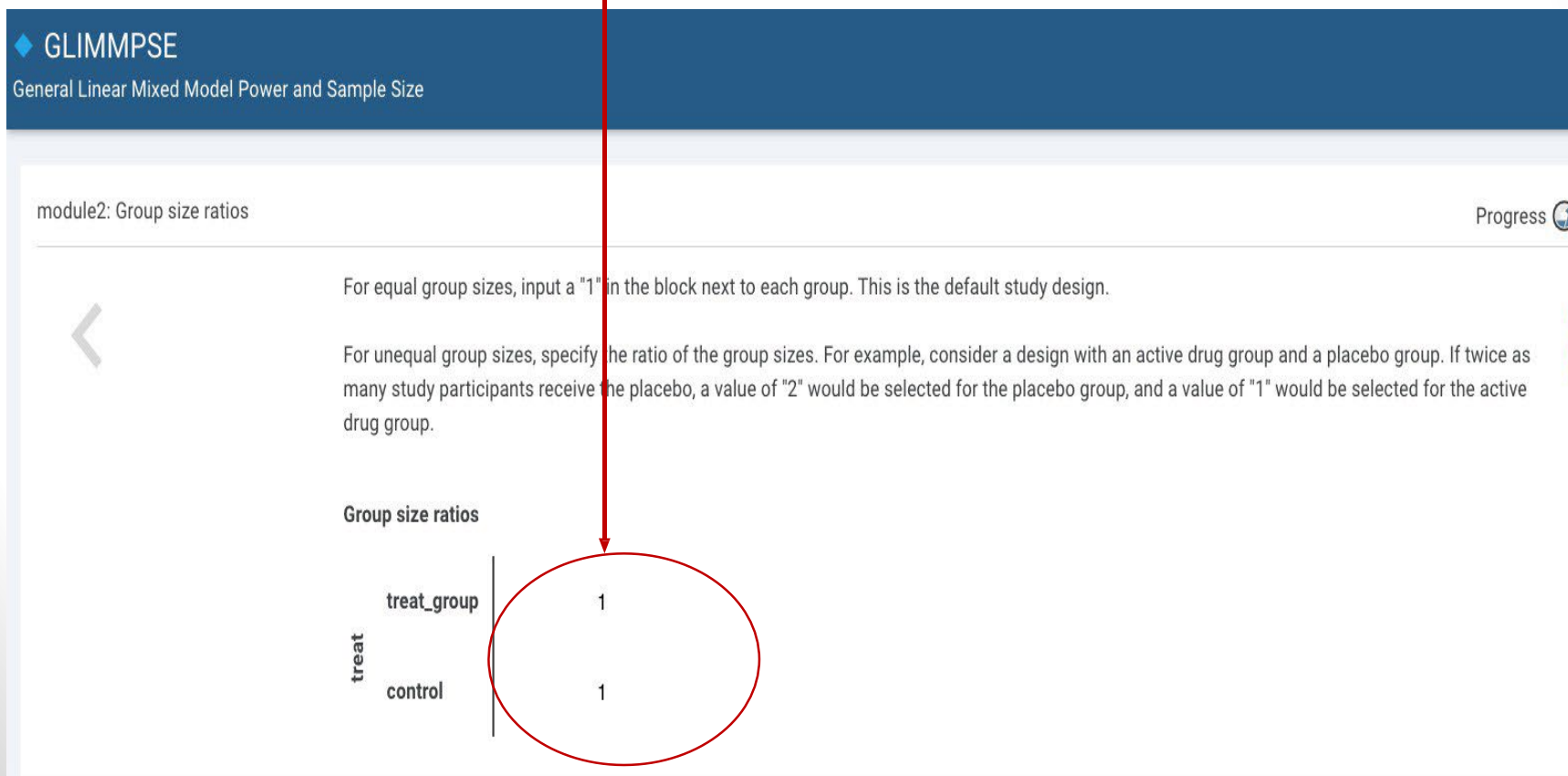
module2: Group size ratios Progress

For equal group sizes, input a "1" in the block next to each group. This is the default study design.

For unequal group sizes, specify the ratio of the group sizes. For example, consider a design with an active drug group and a placebo group. If twice as many study participants receive the placebo, a value of "2" would be selected for the placebo group, and a value of "1" would be selected for the active drug group.

Group size ratios

| | | |
|-------|-------------|---|
| treat | treat_group | 1 |
| | control | 1 |



Means

| Intervention | Baseline | 6 Months | 12 Months |
|---------------------------|----------|----------|-----------|
| Sensory Focus (SF) | 3.6 | 2.8 | 0.9 |
| Standard of Care (SOC) | 4.5 | 4.3 | 3.0 |

Intervention Difference
(SF - SOC)

0.9

1.5

-2.1

Net Difference Over Time
(12 Months - Baseline)

-1.2

Enter the mean values for the outcome variable within each group as shown.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

week2_nv: Marginal means

Progress He

The table below shows the mean values for outcome **mem_pain** within each group in the study. Each group is represented by a row in the table, and each repeated measure dimension is represented by a column.

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

Expected mean values, per group, for *mem_pain*

| | | time | | |
|-------|-------------|------|-----|-----|
| | | 1 | 2 | 3 |
| treat | treat_group | 3.6 | 2.8 | 0.9 |
| | control | 4.5 | 4.3 | 3 |

Set blank values to value

Often it is useful to consider mean differences larger or smaller than our initial guess. Here we type in 1 to indicate that our initial guess is correct. The Scale Factors option is one way of ‘Accounting for Uncertainty.’

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Scale factor for the marginal means

Progress He

In power analysis, it is not possible to know the exact values of means before the experiment is observed. Scale factors allow you to consider alternative values for the means by scaling the values entered on the previous screen.

For example, entering the scale factors 0.5, 1, and 2 would compute power for the mean values divided by 2, the mean values as entered, and the mean values multiplied by 2.

Enter a scale factor:

Scale Factor

remove

1



Variances and Correlations

- Consider the sources of variability and correlation in the study design.
 - Repeated measures **within a given participant** will be correlated.
 - Outcome measurements will vary **between participants**.

Variances and Correlations

Correlation Between Outcomes Over Time

Gedney, Logan, and Baron (2003) identified predictors of the amount of experienced pain recalled over time...One of the findings was that memory of pain intensity at 1 week and 18 months had a correlation of 0.406. We assume that the correlation between measures 18 months apart will be similar to the correlation between measures 12 months apart. Likewise, the correlation between measures 6 months apart will be only slightly greater than the correlation between measures 18 months apart.

Variances and Correlations

Standard Deviation of the Outcome

Logan, et al. (1995) examined whether sensory focus therapy during a root canal procedure could reduce a patient's experienced pain. The investigators assessed experienced pain on a 5 point scale both immediately and at one week following the procedure. The standard deviation of the measurements was 0.90.

Enter the standard deviation as shown.

GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Variability across outcomes

Progress  Hel

Enter the standard deviation you expect to observe for each outcome.

| Outcome | Standard Deviation |
|----------|--------------------|
| mem_pain | 0.9 |

Enter the ratios of standard deviations for time.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measure standard deviation ratios

Progr

Define the ratios of standard deviations for time. One of your values should be 1 and the others should represent the ratio of that value to that value:

For example, if you believe that the standard deviation doubles at each time, enter the values 1, 2, 4, 8... etc.

| time | Standard Deviation Ratio |
|------|--------------------------|
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |

Enter the correlations you expect to observe as shown below.

◆ GLIMMPSE

General Linear Mixed Model Power and Sample Size

module2: Repeated measure correlation

Progress

For a given research participant, responses vary across outcomes and across repeated measurements. The amount of variability can dramatically impact power and sample size.

Define the **time** correlation matrix, by entering correlations you expect to observe among the chosen spacing values of **time**:

Unstructured

LEAR


time

| | 1 | 2 | 3 |
|-------|-------|-----|-------|
| 1 | 1 | 0.5 | 0.406 |
| 0.5 | 0.5 | 1 | 0.5 |
| 0.406 | 0.406 | 0.5 | 1 |

(each off-diagonal correlation must be between -1 and 1, exclusive)

Because we have good evidence from the previous study for the variability and correlation, we enter 1 for the scale factor as shown below.


◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size


module2: Scale factor variance Progress 

<

Changes in variability can dramatically affect power and sample size results. It is not possible to know the variability until the experiment is observed. Scale factors allow you to consider alternative values for variability by scaling the calculated covariance matrix. For example, entering the scale factors 0.5, 1, and 2 would compute power for the covariance matrix divided by 2, the covariance matrix as entered, and the covariance matrix multiplied by 2.

You may add up to 10 scale factors.

Choose a number greater than zero 

| Scale Factor | remove |
|--------------|---|
| 1 |  |

Study Defined; Ready to Compute

After we hit the “Calculate” button, the results will be shown in a table.

◆ GLIMMPSE
General Linear Mixed Model Power and Sample Size

module2: Calculate Progress

Calculate

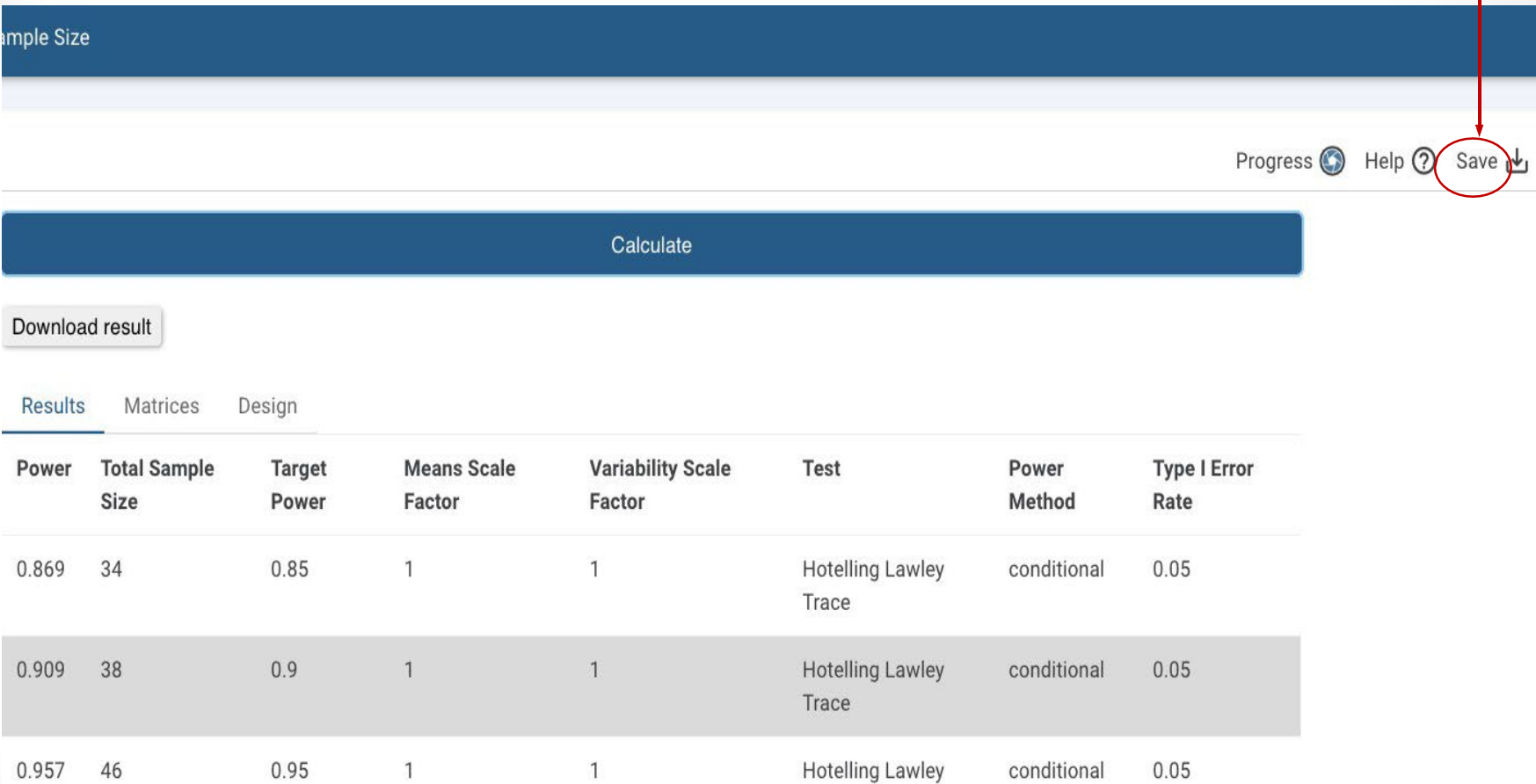
Download result

Results Matrices Design

| | |
|-------------------------|---|
| Design | ✓ |
| Hypothesis | ✓ |
| Design Dimensions | ✓ |
| Parameters | ✓ |
| Optional Specifications | ✓ |

Save Results and .json Files

We can save results and study design inputs by using the “Save” button.



The screenshot shows a software interface with a dark blue header bar. On the right side of the header, there are icons for 'Progress', 'Help', and 'Save'. The 'Save' icon is circled in red, with a red arrow pointing to it from the text above. Below the header is a large blue button labeled 'Calculate'. Underneath that is a 'Download result' button. A table with the following data is displayed:

| Power | Total Sample Size | Target Power | Means Scale Factor | Variability Scale Factor | Test | Power Method | Type I Error Rate |
|-------|-------------------|--------------|--------------------|--------------------------|------------------------|--------------|-------------------|
| 0.869 | 34 | 0.85 | 1 | 1 | Hotelling Lawley Trace | conditional | 0.05 |
| 0.909 | 38 | 0.9 | 1 | 1 | Hotelling Lawley Trace | conditional | 0.05 |
| 0.957 | 46 | 0.95 | 1 | 1 | Hotelling Lawley | conditional | 0.05 |

We sought sample sizes to give “Target” powers of at least 0.85, 0.9, and 0.95. Actual powers are 0.869, 0.909 and 0.957. Sample size is discrete. We need an even sample size for two equal size groups.

Here are the matrices that were used for the calculation.

Results

Matrices

Design

$$E_s(\mathbf{X}) = \begin{bmatrix} 1.00 & 0.00 \\ 0.00 & 1.00 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 3.60 & 2.80 & 0.900 \\ 4.50 & 4.30 & 3.00 \end{bmatrix}$$

$$\mathbf{C} = [1.00 \quad -1.00]$$

$$\mathbf{U} = \begin{bmatrix} 1.00 & 1.00 \\ -1.00 & 0.00 \\ 0.00 & -1.00 \end{bmatrix}$$

$$\Sigma_* = (\mathbf{U}'_o \Sigma_o \mathbf{U}_o) \otimes (\mathbf{U}'_r \Sigma_r \mathbf{U}_r) \otimes (\mathbf{U}'_c \Sigma_c \mathbf{U}_c)$$

$$= [0.810] \otimes \begin{bmatrix} 1.00 & 0.594 \\ 0.594 & 1.19 \end{bmatrix} \otimes [1.00] = \begin{bmatrix} 0.810 & 0.481 \\ 0.481 & 0.962 \end{bmatrix}$$

$$\Theta_0 = [0.00 \quad 0.00]$$

$$\alpha = 0.05$$

$$\Theta = [0.600 \quad 1.20]$$

$$\mathbf{M} = [2.00]$$

$$\nu_e = 2$$

No. of replicated rows in design matrix: 1

$$E_s(\Delta) = \begin{bmatrix} 0.180 & 0.360 \\ 0.360 & 0.720 \end{bmatrix}$$

Sample Size Calculation Summary

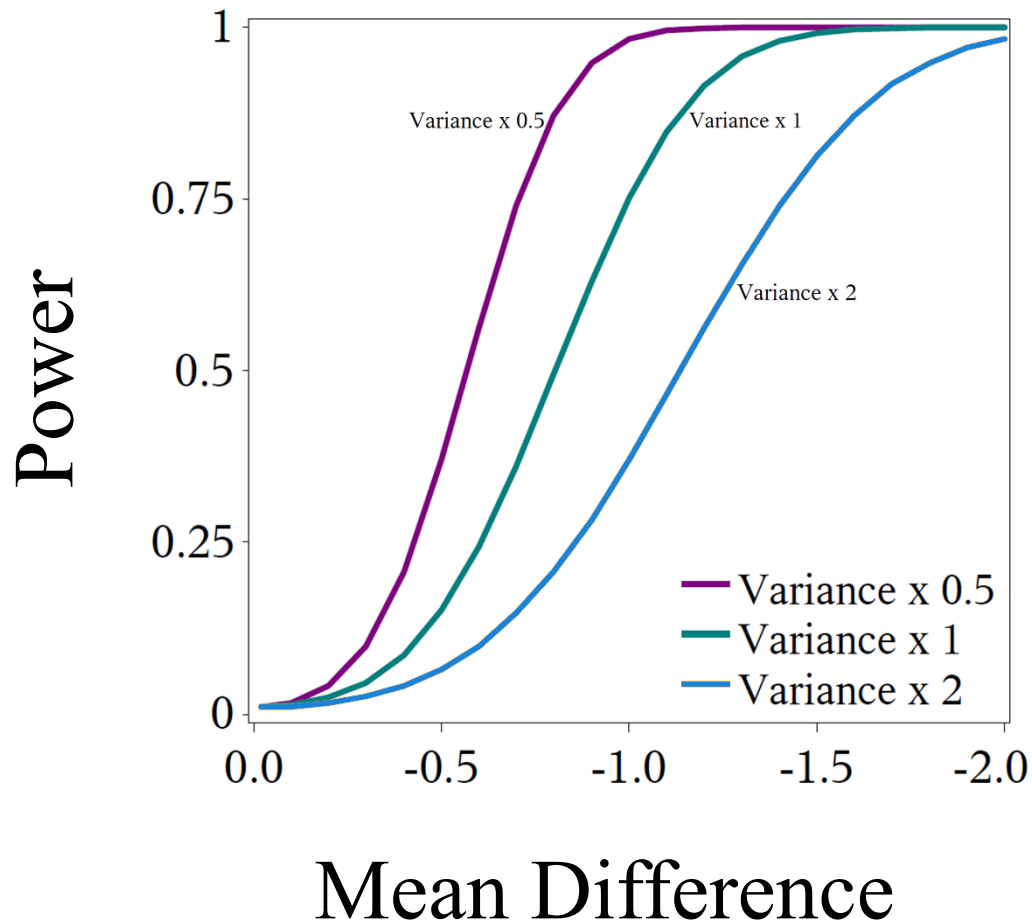
We plan a repeated measures analysis using the Wald (Hotelling-Lawley) statistic to test for a time by treatment interaction.

Based on previous studies, we predict memory of pain measures will have a standard deviation of 0.90 and the correlation between baseline and 6 months will be 0.5.

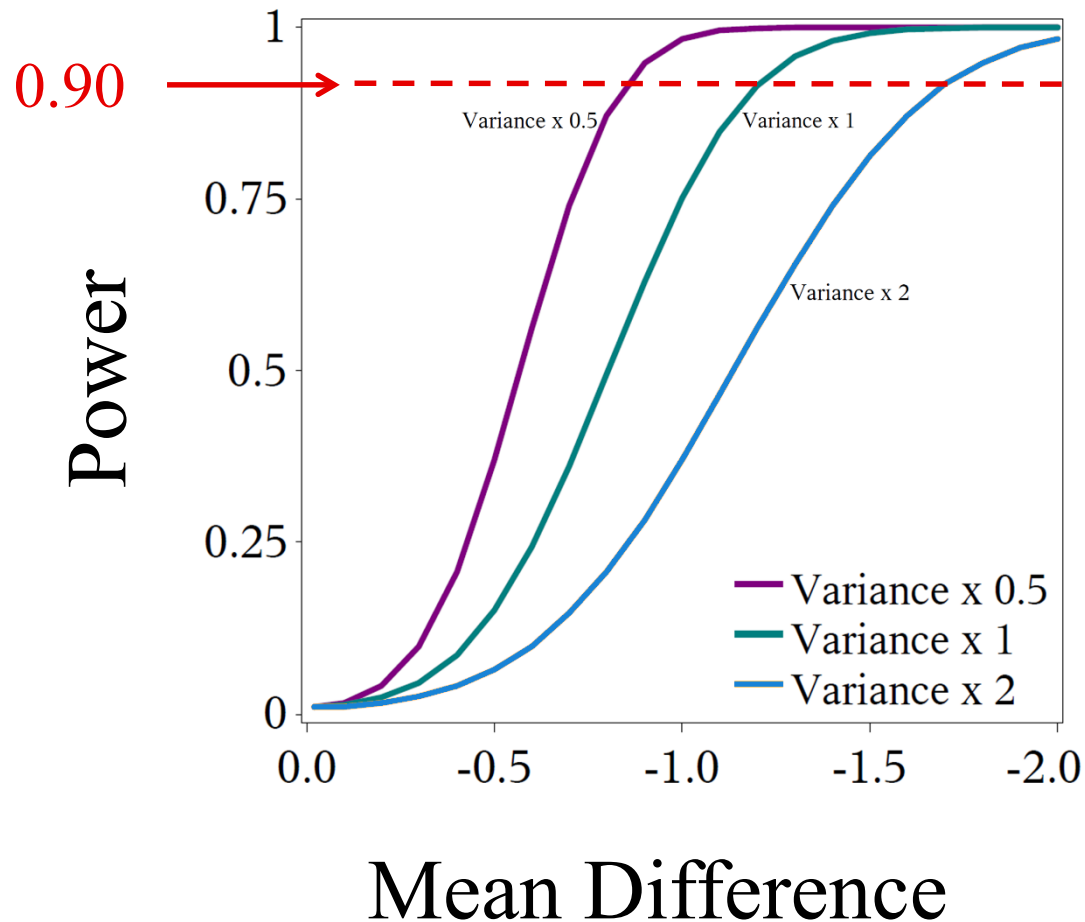
Based on previous studies, we believe the correlation will decrease slowly over time, for a correlation of 0.406 between pain recall measures at baseline and 12 months.

For a desired power of 0.90 and a Type I error rate of 0.05, we estimated that we would need 38 participants to detect a mean difference of 1.2.

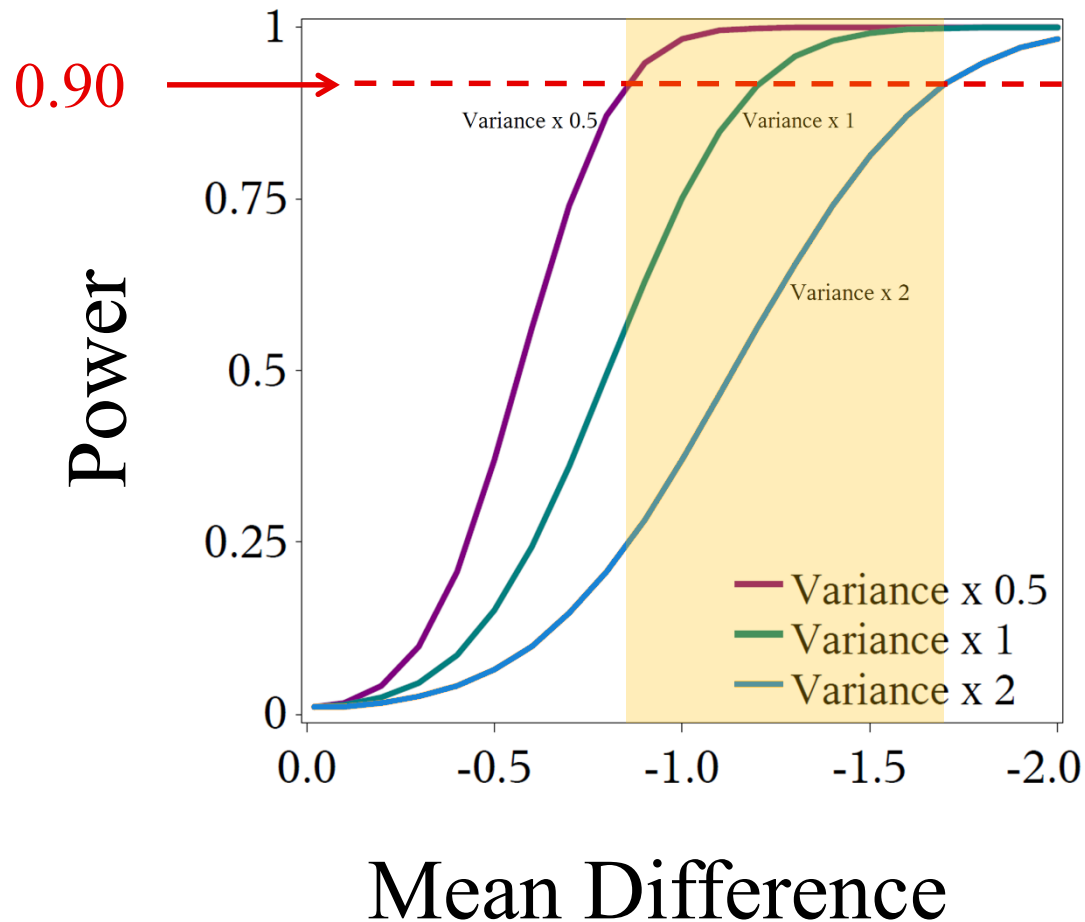
Accounting for Uncertainty



Accounting for Uncertainty



Accounting for Uncertainty



Handling Missing Data

- Assume 25% loss to follow-up.
- Inflate calculated sample size by dividing calculated result:

$$38 \div (1 - 0.25) = 50.7$$

- Round up, to **51**, and increase the result to make it divisible by the number of groups, yielding **52**.

Sample Size Calculation Summary

Over 12 months, we expect **25% loss to follow up**. We will inflate the sample size to account for the attrition, for a total enrollment goal of **52 participants**, or 26 participants per treatment arm.

The clinic treats **30 patients per week** with the high desire/low felt coping style. Based on recruitment experience for previous studies, we expect a **40% consent rate**. At an effective enrollment of 12 participants per week, we will reach the enrollment goal of **52 participants in 5 weeks time**.

FAIR Training Materials at www.SampleSizeShop.org

- A free short course is available, *Power and Sample Size for Multilevel and Longitudinal Study Designs*.
- The course was designed for scientists, and tested with over 200 scientists and statisticians.
- The material was presented as a two-credit course.
- The course materials (~1,000 slides, home works, answers, screenshots, and video lectures) are individually available at “Sample Size Workshop” menu item on www.samplesizeshop.org.
- Tutorials, applied articles, and theory articles are also linked.

What is Not Available

- GLIMMMPSE software covers only a subset of mixed models: incomplete designs are not allowed.
- Continuous predictors are limited to being simple covariates.
- Unequal cluster sizes are not covered.
- Missing repeated measures are not allowed.
- Binary, Poisson and survival outcomes are not allowed.
- High dimensional problems are ignored.
- No power software covers everything.

There is much room for future methodological research.

There is much room for extensions to the software.

We Plan to Add What Users Want

- The preferences of users will guide us.
- We are seeking funding to continue the project.

Outline

- An original study is *reproducible* if a new sample of data collected under the original protocol is likely to lead to the same conclusions as the original study.
- To be reproducible, a study must have good statistical power.
- Advances in power theory make it accurate and fast to compute power for linear mixed models.
- Free point-and-click software makes it easy to compute power.
- Planning a longitudinal clinical trial illustrates the process.
- No power software covers everything. There is much room for future methodological research, and extensions to the software.

Questions?



Bibliography

- Chi YY, Glueck DH and Muller KE (2019) Power and sample size for fixed-effects inference in reversible linear mixed models, *The American Statistician*, 73, 350-359. doi: 10.1080/00031305.2017.1415972, PMID: PMC7009022.
- Gedney JJ, Logan H, Baron RS. (2003) Predictors of short-term and long-term memory of sensory and affective dimensions of pain. *J Pain*. Mar;4(2):47-55. doi: 10.1054/jpai.2003.3. PMID: 14622715.
- Kenward, M. G. and Roger, J. H. (1997) Small sample inference for fixed effects from restricted maximum likelihood, *Biometrics*, 53, 983–997.
- Kreidler SM, Muller KE, Grunwald GK, Ringham BM, Coker-Dukowitz ZT, Sakhadeo UR, Barón AE, and Glueck DH (2013) GLIMMPSE: online power computation for linear models with and without a baseline covariate. *Journal of Statistical Software*, 54(10). PMID: PMC3882200
- Logan HL, Baron RS, Kohout F. (1995) Sensory focus as therapeutic treatments for acute pain. *Psychosom Med.*, 57(5):475–484.
- Muller KE, LaVange LM, Ramey SL, and Ramey CT (1992) Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association*, 87, 1209-1226. PMID: PMC4002049
- Muller KE and Stewart PW (2006) *Linear Model Theory; Univariate, Multivariate, and Mixed Models*. New York: Wiley. DOI: 10.1002/0470052147
- Sommer M, Ritzhaupt AD, Muller KE and Glueck DH (2019) Transformation of a face-to-face workshop into a Massive Open Online Course (MOOC): a design and development case, *Journal of Formative Design in Learning*, 3, 97–110. doi:10.1007/s41686-019-00037-y
- Also see www.SampleSizeShop.org