Finding Power and Sample Size for Mixed Models in Study Designs with Repeated Measures and Clustering

M. M. Maldonado-Molina, PhD[1]
A. E. Barón, PhD[2]
S. M. Kreidler, DPT, MS[2]

1. Health Outcomes and Policy, Institute for Child Health Policy, University of Florida
2. Department of Biostatistics and Bioinformatics, Colorado School of Public Health, University of Colorado Denver

Session Outline

Introduction Dr. Mildred Maldonado-Molina	2:00 – 2:05
Power and Sample Size for the Most Common Hypotheses in Mixed Models Dr. Anna Barón	2:05 – 2:20
Selecting a Covariance Model for Longitudinal and Multilevel Designs Dr. Mildred Maldonado-Molina	2:20 – 2:35
Power Analysis for Mixed Models: Using Free Web-Based Power Software Dr. Sarah Kreidler	2:35 – 3:05
Discussant Dr. Jacinda Dariotis	3:05 – 3:15
Question and Answer	3:15 – 3:30

Collaborators

- Deb Glueck, PhD[1]
- Yi Guo, MSPH, PhD[2]
- Keith Muller, PhD[2]
- Aarti Munjal, PhD[1]
- Brandy Ringham, MS[1]
- Uttara Sakhadeo, BS[1]

[1] Department of Biostatistics and Bioinformatics, Colorado School of Public Health, University of Colorado Denver.
[2] Health Outcomes and Policy, College of Medicine, University of Florida.

Grant Support

The progress described in this talk was supported by the National Institute of Dental and Craniofacial Research under award NIDCR 1 R01 DE020832-01A1, Multilevel and Longitudinal Study Sample Size Tools for Behavioral Scientists (12/09/10-11/30/14), led by Keith Muller and Deborah Glueck. The content is solely the responsibility of the authors, and does not necessarily represent the official views of the National Institute of Dental and Craniofacial Research nor the National Institutes of Health.

The Sample Size Problem

- Every study requires an accurate sample size calculation.
- If sample size is too large, participants are exposed to unnecessary risk.
- If sample size is too small, the study may have insufficient power.
- It is important to match power and sample size analysis to data analysis.
- Repeated measures and multilevel features make power and sample size analysis more challenging.
- Not all studies have a dedicated statistician to assist with design.

Power for the Linear Mixed Model

- No general power methods exist for mixed models.
- Extensive power methods exist for the general linear multivariate model.
- ***Can we use existing results in the linear mixed model?***
- ***How would we implement the methods in day-to-day practice?***

Session Outline

Introduction Dr. Mildred Maldonado-Molina	2:00 – 2:05
Power and Sample Size for the Most Common Hypotheses in Mixed Models Dr. Anna Barón	2:05 – 2:20
Selecting a Covariance Model for Longitudinal and Multilevel Designs Dr. Mildred Maldonado-Molina	2:20 – 2:35
Power Analysis for Mixed Models: Using Free Web-Based Power Software Dr. Sarah Kreidler	2:35 – 3:05
Discussant Dr. Jacinda Dariotis	3:05 – 3:15
Question and Answer	3:15 – 3:30

Power and Sample Size for the Most Common Hypotheses in Mixed Models

Anna E. Barón, PhD

Department of Biostatistics and Bioinformatics
Colorado School of Public Health
University of Colorado Denver

Agenda

- Mixed Model (MM): Clustered and Repeated Measures Data
  - Common Hypothesis Tests in the Linear MM (LMM)
  - The LMM as a General Linear Multivariate Model
- Missing Data
- Summary and Segue to Building Covariance Models

Agenda

- **Mixed Model (MM): Clustered and Repeated Measures Data**
  - Common Hypothesis Tests in the Linear MM (LMM)
  - The LMM as a General Linear Multivariate Model
- Missing Data
- Summary and Segue to Building Covariance Models

LMM Commonly Used for Clustered and Repeated Measures Data

- Linear MM: Laird and Ware, 1982; Demidenko, 2004; Muller and Stewart, 2007
- Studies with Clustering
  - Designed: Cluster randomized studies
  - Observational: Clustered observations
- Studies with Repeated Measures (RM)
  - Designed: Randomized clinical trials
  - Observational: Cohort studies, natural history
- Combination
  - Cluster randomized longitudinal studies

Data Structures

Clustering

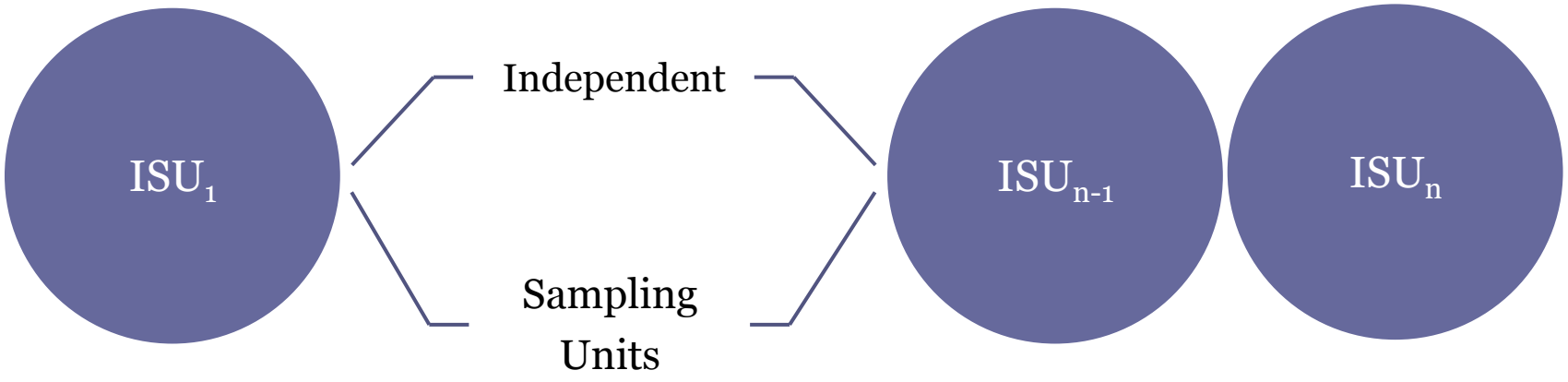
⇔ Restricted Multi-level

Repeated Measures

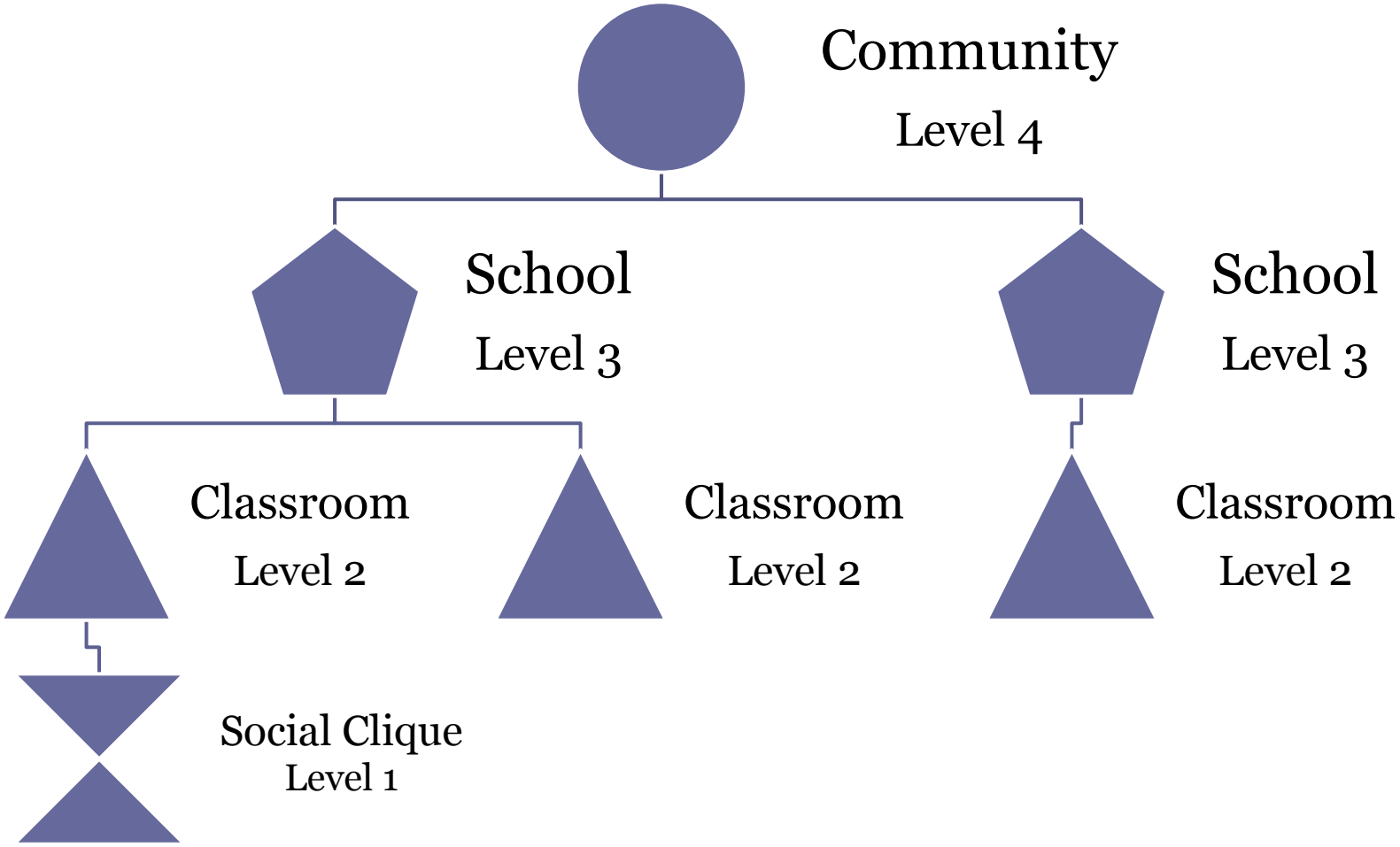
⇔ Restricted Longitudinal

Clustering – Top Level (k)

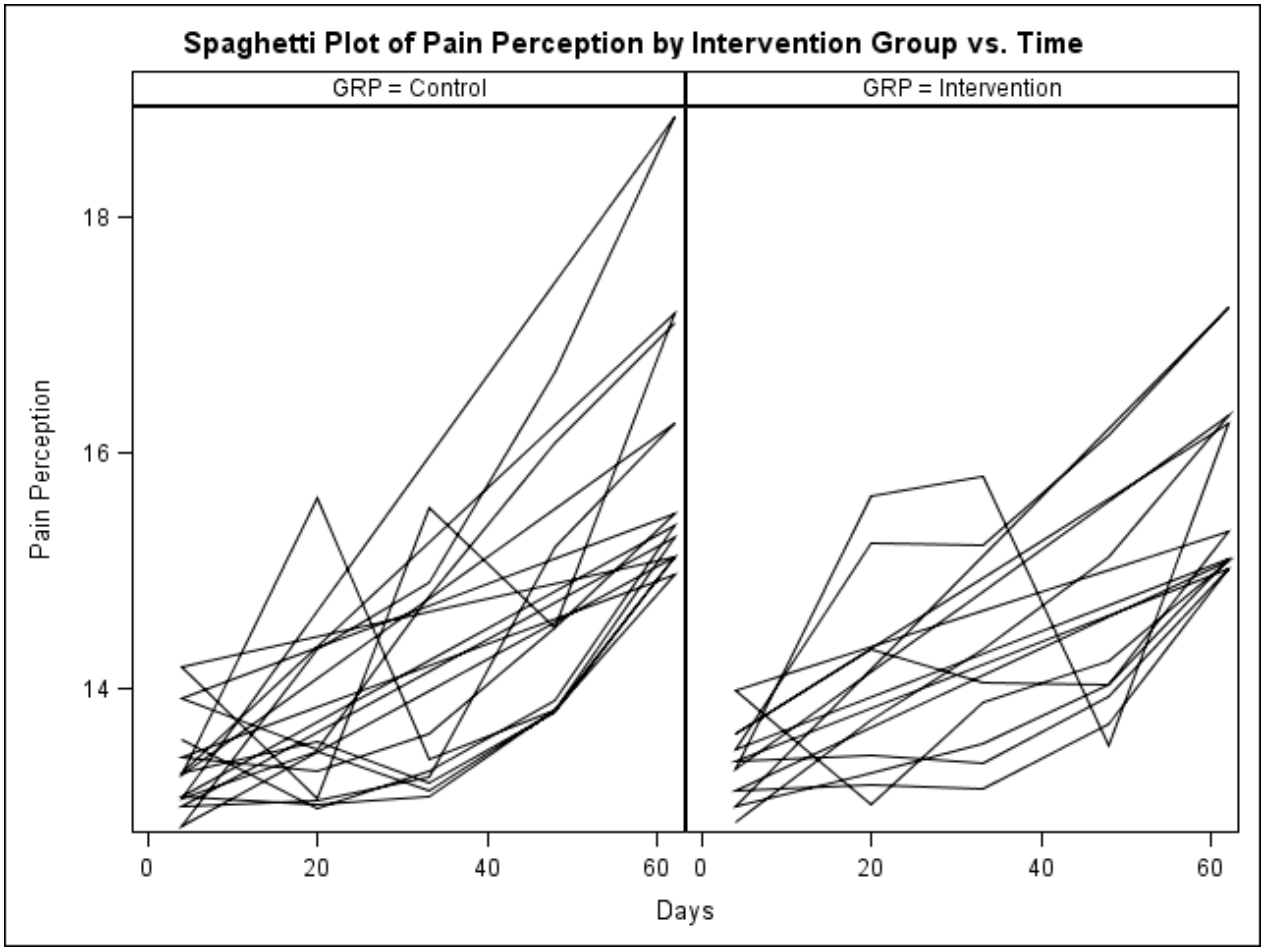
Clusters: Communities as Independent Sampling Units (ISU)



Clustering – additional levels



Repeated Measures



Agenda

- Mixed Model (MM): Clustered and Repeated Measures Data
  - **Common Hypothesis Tests in the Linear MM (LMM)**
  - The LMM as a General Linear Multivariate Model
- Missing Data
- Summary and Segue to Building Covariance Models

Power for the Most Common Hypothesis Tests for the Linear Mixed Model

- ✓ A. Power for testing fixed effects (means)
- x B. Power for testing random effects (covariance)
- x C. Power for testing fixed and random effects

General and accurate power and sample size methodology is not available.

There are, however, good methods for most of class A.

Agenda

- Mixed Model (MM): Clustered and Repeated Measures Data
  - Common Hypothesis Tests in the Linear MM (LMM)
  - **The LMM as a General Linear Multivariate Model**
- Missing Data
- Summary and Segue to Building Covariance Models

Power and Sample Size for Fixed Effects in the Linear Mixed Model

Key idea: Some linear mixed models (LMM) can be recast as general linear multivariate models (GLMM)

- Which ones?
  - No missing data and no mistimed data
  - Unstructured covariance model across responses (a robust, safe, conservative assumption)
  - Typical clinical trial or longitudinal study in which main inference is about time by treatment interaction
- Why do we care?
  - Muller et al. (1992) show how to do power for time by treatment using GLMM framework

Four Specific Requirements for a LMM to be Recast as a GLMM

To be reversible to a General Linear Multivariate Model, a Linear Mixed Model must:

1. Have a balanced design within ISU; no repeated covariates; saturated with regard to between-within effects
2. Have an unstructured covariance model
3. Use Wald test for inference about fixed effects
4. Use Kenward-Roger df approach

Reversibility Requirements – 1.

1. Have a balanced design within ISU; no repeated covariates; saturated with regard to between-within effects
   - **No missing or mistimed data**
   - **Unequal group sizes ok**
   - **Treatment assignment does not change over time**
   - **Factorial design including Interaction between Treatment (between) and Time (within)**

Reversibility Requirements – 2.

2. Have an unstructured covariance model
   - **All variances and covariances unspecified, i.e. they do not follow a pattern or rule, e.g. for three repeated measures –**

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix} = 0.25 \begin{bmatrix} 1 & 0.3 & 0.2 \\ 0.3 & 1 & 0.5 \\ 0.2 & 0.5 & 1 \end{bmatrix}$$

Reversibility Requirements – 3.

3. Use Wald test for inference about fixed effects
   - **Most common test used for LMM analysis by standard packages**

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
trt	1	98	25.43	25.43	<.0001	<.0001
time	3	96	184.48	60.24	<.0001	<.0001
trt*time	3	96	107.79	35.20	<.0001	<.0001

Reversibility Requirements – 4.

4. Use Kenward-Roger df approach
   - **DF approximation method with modified covariance matrix**
   - **Under reversibility, covariance matrix is unstructured and test is equivalent to Hotelling-Lawley Trace test**
   - **Muller et al. (2007) showed it's the best test**

Example Code

- SAS

```
TITLE 'Repeated Measures using Mixed Model';
   PROC MIXED DATA=one;
   CLASS TRT ID TIME;
   MODEL y = trt time trt*time / S CHISQ DDFM=KR;
   REPEATED time / SUBJECT=ID TYPE=UN R RCORR;
RUN;
```

- SPSS – Satterthwaite, but not Kenward-Roger method

- R

```
library(lmerTest)
m <- lmer(y ~ factor(trt) + factor(time) + factor(time):factor(trt) + (1|Subject),
          data=one)
anova(m, ddf="Kenward-Roger")
```

- STATA – none, but promised for future releases

Power and Sample Size for GLMM

- Muller, LaVange, Ramey and Ramey (1992)
- Multivariate approach to repeated measures and MANOVA: Hotelling-Lawley Trace
- Kenward-Roger Wald Test equivalent when LMM is reversible

Agenda

- Mixed Model (MM): Clustered and Repeated Measures Data
  - Common Hypothesis Tests in the Linear MM (LMM)
  - The LMM as a General Linear Multivariate Model
- **Missing Data**
- Summary and Segue to Building Covariance Models

Missing Data Adjustments

- Some useful crude approximations (Catellier and Muller, 2000):
  - Complete data power is an upper bound
  - Power for N = (100% - % missing) x # ISUs appears conservative, requires assuming data are Missing at Random

- Work is in progress to identify better approximations

Agenda

- Mixed Model (MM): Clustered and Repeated Measures Data
  - Common Hypothesis Tests in the Linear MM (LMM)
  - The LMM as a General Linear Multivariate Model
- Missing Data
- **Summary and Segue to Building Covariance Models**

Summary

- Under widely applicable restrictions a LMM can be expressed as a General Linear Multivariate Model for which accurate power and sample size analysis is available.
- Convenient adjustments appear to suffice for simple missing data patterns.
- Next: How to select and build a multilevel covariance model.

Session Outline

Introduction Dr. Mildred Maldonado-Molina	2:00 – 2:05
Power and Sample Size for the Most Common Hypotheses in Mixed Models Dr. Anna Barón	2:05 – 2:20
Selecting a Covariance Model for Longitudinal and Multilevel Designs Dr. Mildred Maldonado-Molina	2:20 – 2:35
Power Analysis for Mixed Models: Using Free Web-Based Power Software Dr. Sarah Kreidler	2:35 – 3:05
Discussant Dr. Jacinda Dariotis	3:05 – 3:15
Question and Answer	3:15 – 3:30

Selecting a Covariance Model for Longitudinal and Multilevel Designs

Mildred M. Maldonado-Molina, PhD

Department of Health Outcomes and Policy
Institute for Child Health Policy
University of Florida

Agenda

- Motivate the need for valid covariance structures
- Identify appropriate covariance structures for multilevel and longitudinal features
- Combine structures for longitudinal and multilevel features into a single covariance model
- Review ongoing research and additional resources

Agenda

- **Motivate the need for valid covariance structures**
- Identify appropriate covariance structures for multilevel and longitudinal features
- Combine structures for longitudinal and multilevel features into a single covariance model
- Review ongoing research and additional resources

Take home

- It is important to:
  - select an appropriate sample size
  - align the research design (including sample size selection), data collection, and statistical analyses.

- Valid covariance structures may be created by layering simpler patterns for each source of correlation.

Why worry about covariance structures?

- Variability affects power and sample size
- Failing to account for correlation during study design may lead to incorrect sample size
- Failing to account for correlation during data analysis may lead to inflated Type I error rates.

Why does sample size selection matter?

- If sample size is too small
  - inadequate power to detect meaningful effects, producing unreliable answers

- If the sample size is too large
  - ethical considerations
  - wasted time and effort in the study

Why align power and data analysis?

- Reversibility makes alignment possible for complex designs with multilevel and longitudinal features

Agenda

- Motivate the need for valid covariance structures
- **Identify appropriate covariance structures for multilevel and longitudinal features**
- Combine structures for longitudinal and multilevel features into a single covariance model
- Review ongoing research and additional resources

Start your study design

Steps for designing a study with a properly aligned power analysis and data analysis are:

1. Specify sampling patterns
2. Model correlation and variance patterns
3. Choose an analysis method

1. Specify sampling patterns

- Independent sampling unit vs. observational unit
  - Independent sampling units (ISU) (N=30, number of schools)
  - Observational unit is the number of observations in each ISU

- Example
  - p = 100, indicating that there 100 students in each school
  - total number of observations (n)
  - number of independent sampling units * number of time points

- When data is multilevel, we need to differentiate:
  - longitudinal
  - cluster
  - consistent spacing

2. Model correlation and variance

- Choose correlation and variance patterns
  - Unstructured
  - Compound symmetric
  - Autoregressive
  - Linear exponent autoregressive (LEAR)
  - Direct-products
- Consider "sources" of correlation separately

Study Design Informs Covariance Model

- Identify features of the study design which lead to correlation:
  - participants between ISU (e.g. schools) are independent
  - participants within schools are correlated
  - participants within classrooms are correlated
  - observations within participant are correlated

Covariance Structures for Clustering and Repeated Measures

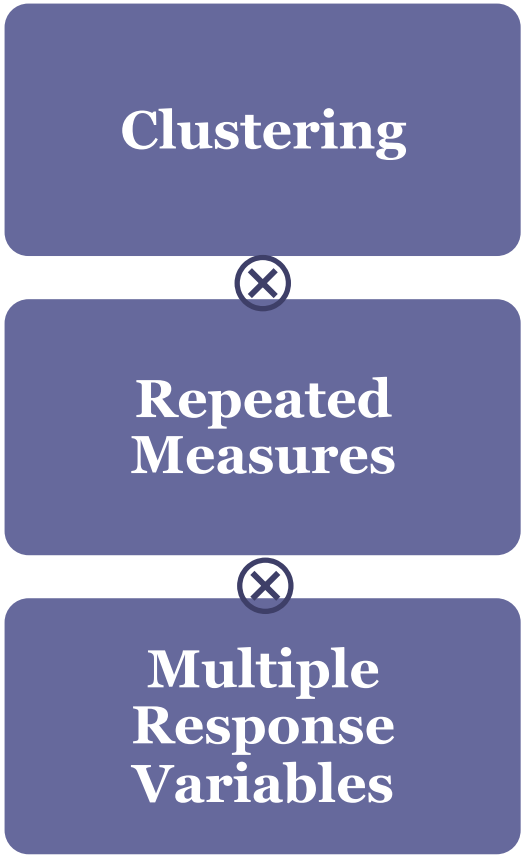
- Clustering
  - exchangeable observations
  - compound symmetric covariance may be used

- Repeated measures
  - unstructured is the most flexible model
  - auto-regressive or LEAR model may be used

Agenda

- ❑ Motivate the need for valid covariance structures
- ❑ Identify appropriate covariance structures for multilevel and longitudinal features
- ❑ **Combine structures for longitudinal and multilevel features into a single covariance model**
- ❑ Review ongoing research and additional resources

Build the Overall Covariance Structure



Example: A design with multilevel, longitudinal, and multivariate features

Variance Clusters | Repeated Measures | Multiple Responses

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_4 \\ \rho_4 & 1 \end{bmatrix}$$

Clusters of Size 3 | 3 Repeated Measures | 2 Response Variables

Agenda

- Motivate the need for valid covariance structures
- Identify appropriate covariance structures for multilevel and longitudinal features
- Combine structures for longitudinal and multilevel features into a single covariance model
- **Review current research and additional resources**

Ongoing Power and Sample Size Research

- Many repeated measures with few independent sampling units
- Unbalanced cluster randomized designs
- Binary or count outcomes
- Improved handling of missing data

Helpful resources

- Muller, LaVange, Ramey, and Ramey (1992)
  - detailed technical review
- Muller and Stewart (2006)
  - methods for univariate, multivariate, and mixed linear models with Gaussian errors
- Muller and Fetterman (2002)
  - Regression and ANOVA: An integrated approach using SAS software (book)
- Online tutorials at **SampleSizeShop.org**

Summary

- Power analysis should be aligned with data analysis.
- Reversibility allows us to select an appropriate sample size for complex multilevel and longitudinal designs.
- Valid covariance structures may be created by layering simpler patterns for each source of correlation.

Thank you

- Questions

- Next presentation
  - Power and sample size calculations with GLIMMPSE, our free web-based software

Session Outline

Introduction Dr. Mildred Maldonado-Molina	2:00 – 2:05
Power and Sample Size for the Most Common Hypotheses in Mixed Models Dr. Anna Barón	2:05 – 2:20
Selecting a Covariance Model for Longitudinal and Multilevel Designs Dr. Mildred Maldonado-Molina	2:20 – 2:35
Power Analysis for Mixed Models: Using Free Web-Based Power Software Dr. Sarah Kreidler	2:35 – 3:05
Discussant Dr. Jacinda Dariotis	3:05 – 3:15
Question and Answer	3:15 – 3:30

Power Analysis for Mixed Models: Using Free Web-Based Power Software

Sarah M. Kreidler, DPT, MS

Department of Biostatistics and Bioinformatics
Colorado School of Public Health
University of Colorado Denver

Agenda

- Motivate the need for GLIMMPSE
- Introduce the GLIMMPSE software
- Present GLIMMPSE validation results
- Example: The Project Northland Chicago (PNC) trial

Agenda

- **Motivate the need for GLIMMPSE**
- Introduce the GLIMMPSE software
- Present GLIMMPSE validation results
- Example: The Project Northland Chicago (PNC) trial

GLIMMPSE Motivation

- Power and sample size calculation is critical for ethical study design.
- Known results are underutilized.
- Our goal: provide a user-friendly tool for calculating power and sample size.

Agenda

- Motivate the need for GLIMMPSE
- **Introduce the GLIMMPSE software**
- Present GLIMMPSE validation results
- Example: The Project Northland Chicago (PNC) trial

What is GLIMMPSE?

GLIMMPSE is a user-friendly, online tool for calculating power and sample size for multilevel and longitudinal studies.

http://glimmpse.samplesizeshop.org/

GLIMMPSE Team

- Software Development:
  - Sarah Kreidler, Tech Lead
  - Aarti Munjal, Senior Software Engineer
  - Uttara Sakhadeo, Software Engineer

- Manual Preparation:
  - Zacchary Coker-Dukowitz
  - Brandy Ringham
  - Yi Guo

Statistical Foundation

- Power for the general linear multivariate model
  - Based on the work of Keith Muller and colleagues.

- Power for designs with fixed predictors
  - Muller and Peterson, 1984
  - Muller and Barton, 1989
  - Muller *et al.*, 1992
  - Muller *et al.*, 2007

- Power for designs with fixed predictors and a Gaussian covariate
  - Glueck and Muller, 2003

Why a Web-based interface?

- Free
- Requires no programming expertise
- Built with industry standard Java technology

GLIMMPSE Features

- Web-based
- Free and open-source
- Designed with an intuitive wizard input style
- Able to produce power curves
- Able to export power results
- Able to save study designs for later use

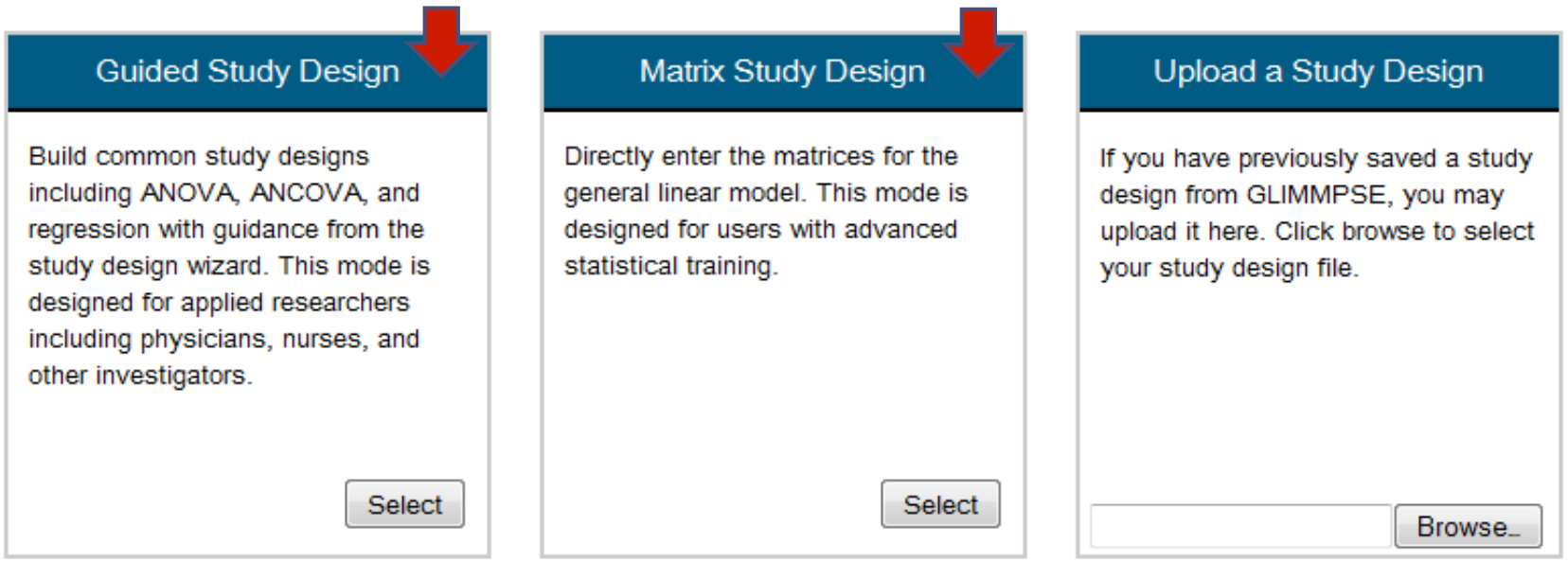
Supported Study Designs

- Cross-sectional studies
- Longitudinal designs
- Multilevel designs
- Designs with a baseline covariate

Current Limitations

- **Binary or count data**
- **Adjustments for missing data**
- **Sample size based on confidence interval width**
- Very high dimensional, low sample size designs
- Certain classes of mixed models

Two interaction modes



Online Resources

www.SampleSizeShop.org

- Documentation
  http://samplesizeshop.org/documentation/glimmpse/
- Tutorials
  http://samplesizeshop.org/education/
- Downloads
  - http://samplesizeshop.org/software-downloads/glimmpse/

Agenda

- Motivate the need for GLIMMPSE
- Introduce the GLIMMPSE software
- **Present GLIMMPSE validation results**
- Example: The Project Northland Chicago (PNC) trial

Validation Results

- Validated against published results and simulation
- Full validation results are available online

http://samplesizeshop.com/documentation/glimmpse-validation-results/

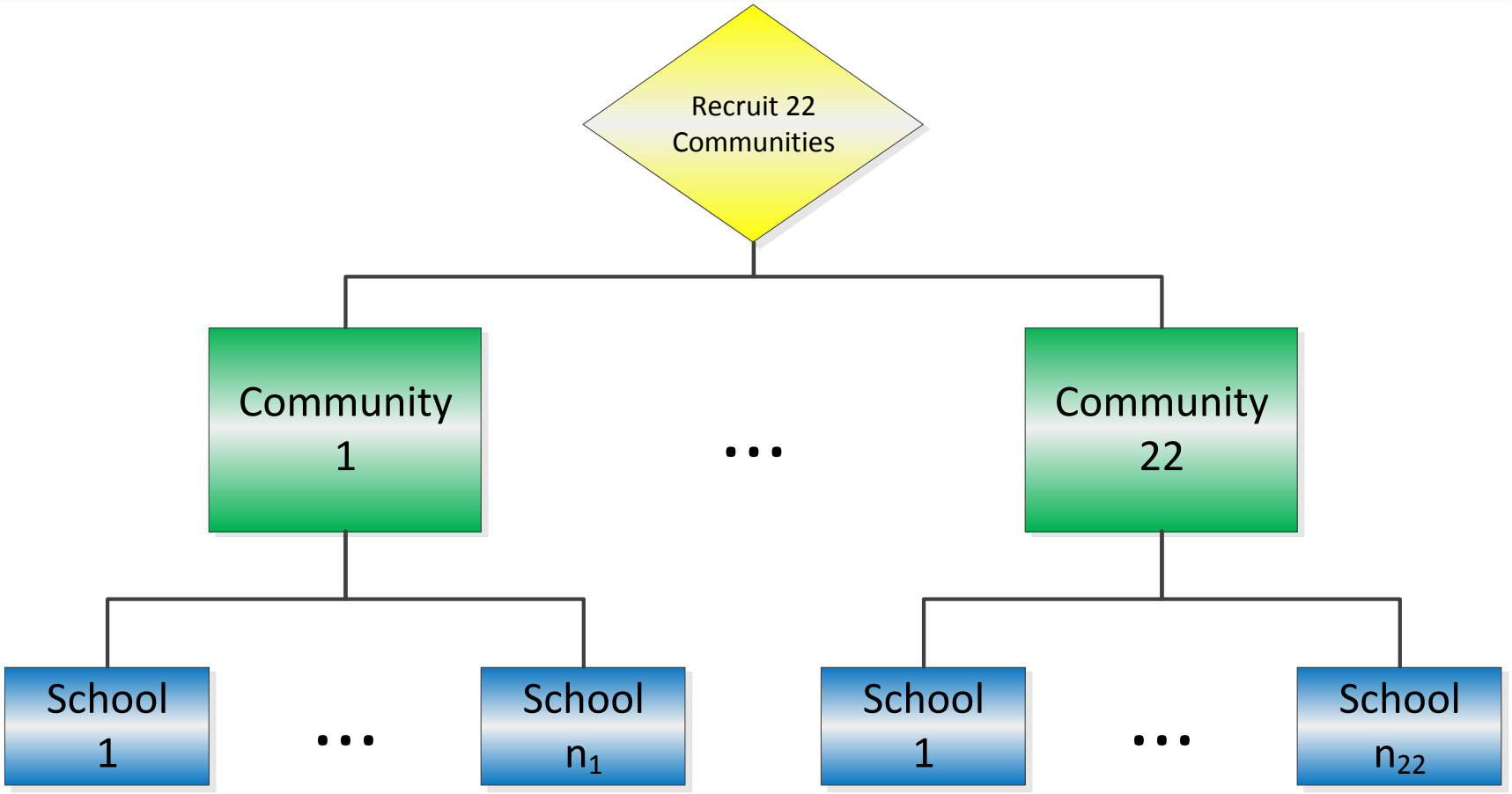
Validation Results

- 6 decimal accuracy against published results.
- 2 decimal accuracy against simulation.
- Worst case error in 1st decimal for complex multivariate designs.

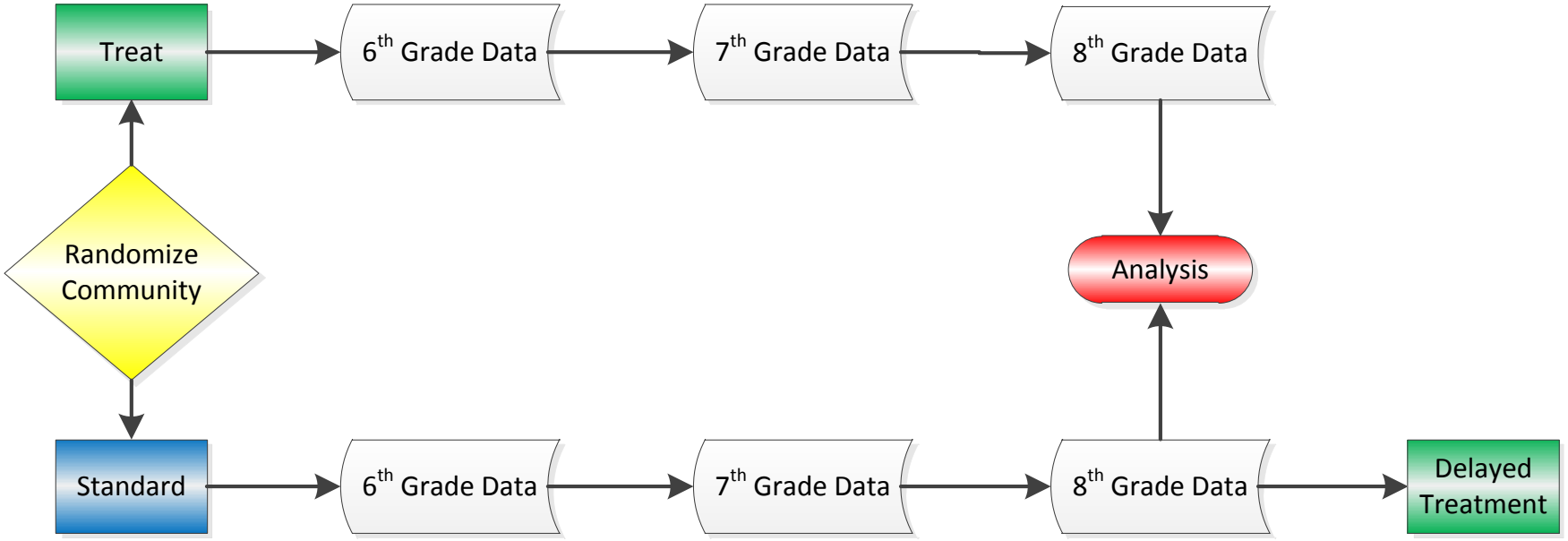
Agenda

- ❑ Motivate the need for GLIMMPSE
- ❑ Introduce the GLIMMPSE software
- ❑ Present GLIMMPSE validation results
- ❑ **Example: The Project Northland Chicago (PNC) trial**

The PNC Trial: Cluster Randomized Design



The PNC Trial: Longitudinal Features



PNC Trial: What is the study design goal?

- We have a fixed sample size, so solving for power is more appropriate.

- We fix the Type I Error rate at 0.05

PNC Trial: What is the sampling scheme?

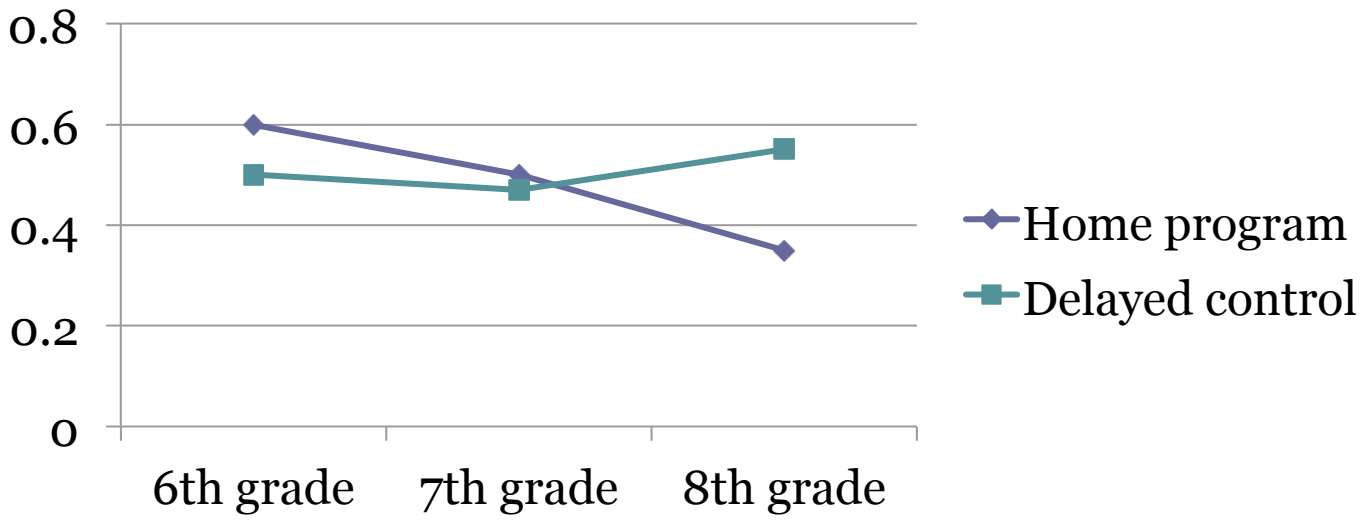
- The independent sampling unit is the community.
- Communities are randomized to receive either the home-based program or delayed program participation.
  - Therefore, treatment is the only predictor variable
  - We will calculate power for 2,3,…,10 communities randomized to each treatment
- The design does not control for a covariate

PNC Trial: What are the responses?

- What responses are measured?
  - Response variable: alcohol behavior scale.
- How often are the responses measured?
  - 3 repeated measures in 6^{th}, 7^{th}, and 8^{th} grade.

PNC Trial: What is the primary hypothesis of interest?

Time trend by treatment interaction



PNC Trial: What are the means?

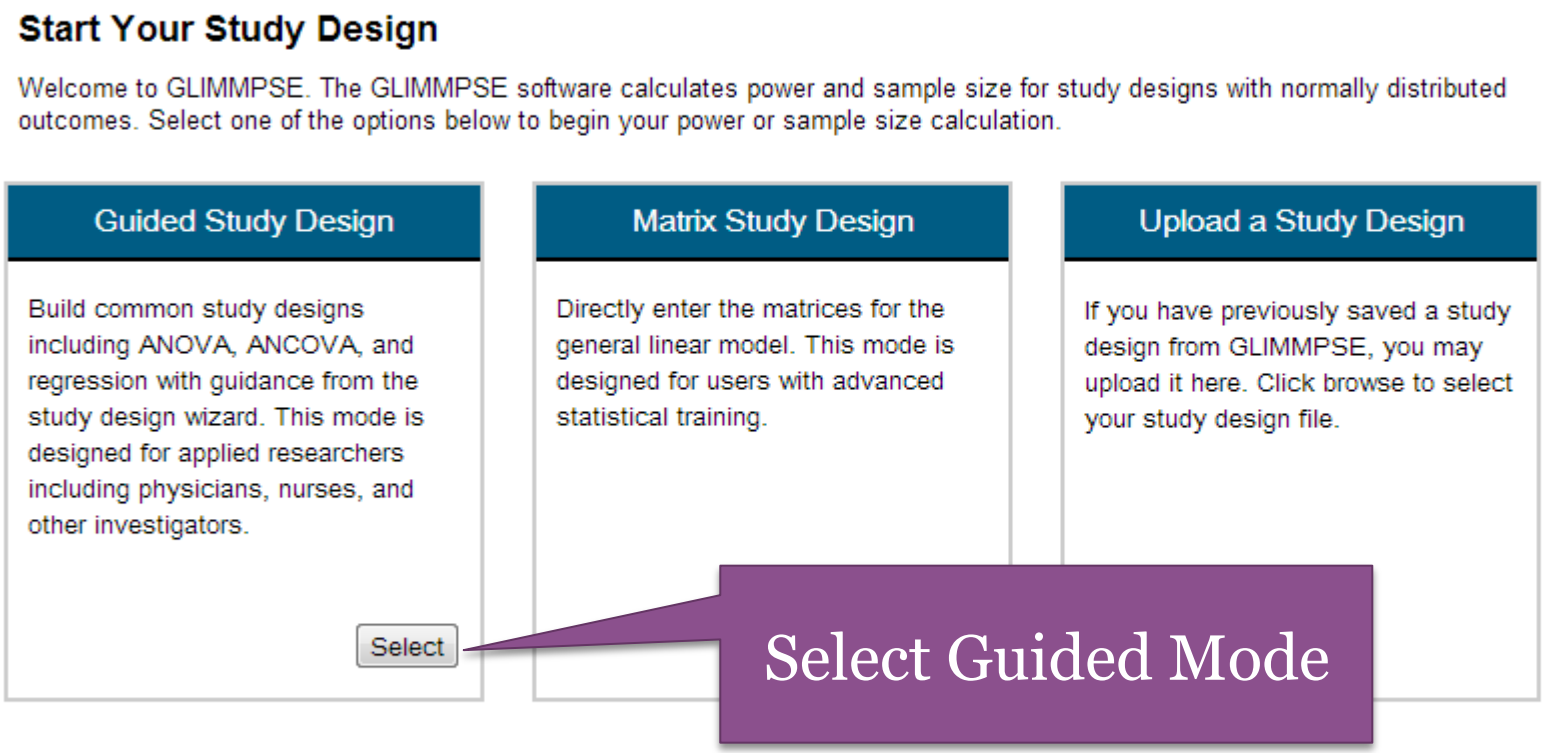
- We wish to detect a reduction in alcohol use in the treatment group in 8th grade

- A reduction of 0.25 on the alcohol behavior scale is considered clinically meaningful.

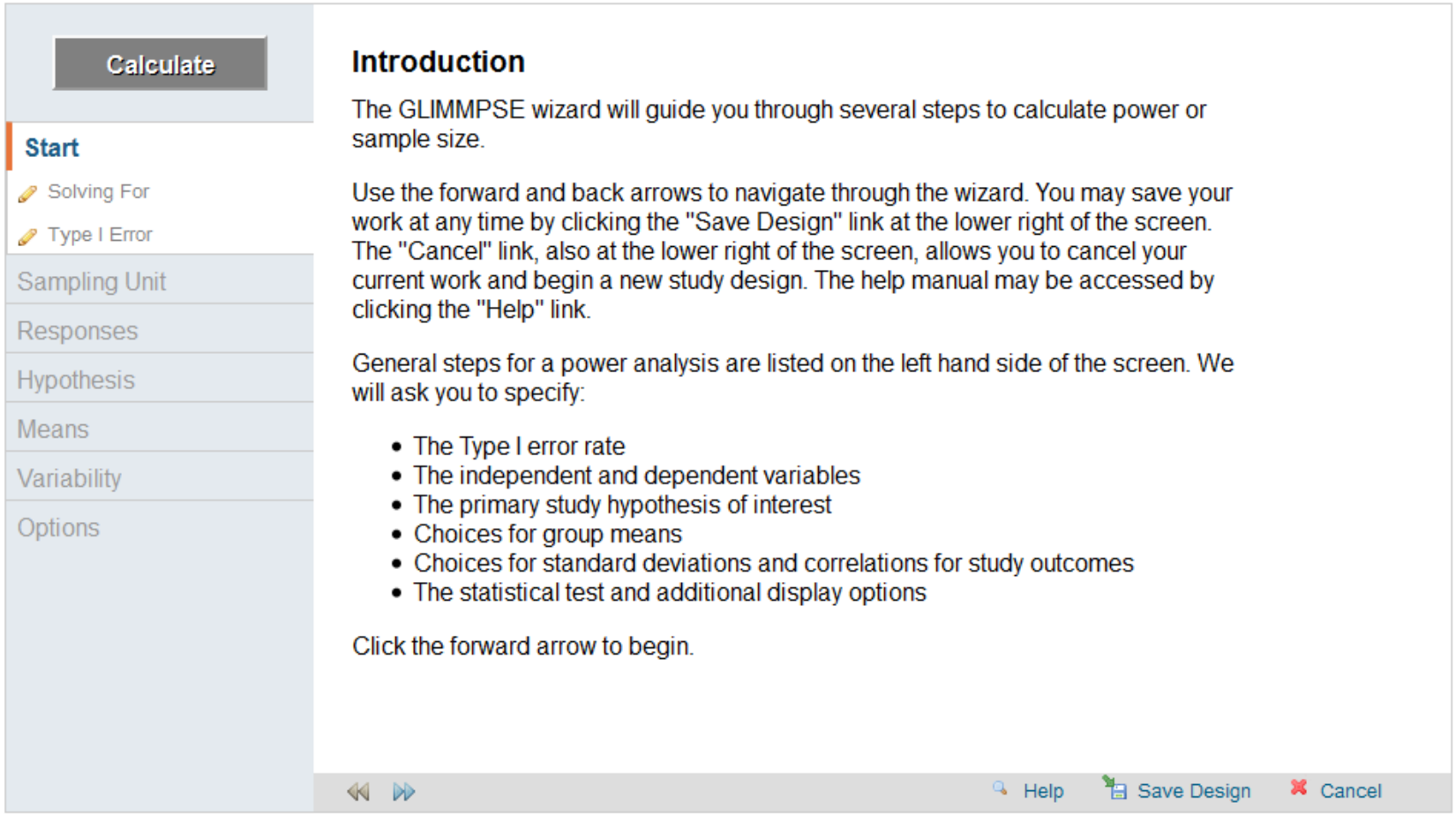
PNC Trial: What is the variance structure?

- Correlation due to clustering and repeated measures
  - Cluster size: 10
  - Standard deviation of alcohol behavior scale: 0.3
- Patterns of variability
  - Clustering
    - Compound symmetry
    - Intraclass correlation: 0.01
  - Repeated Measures:
    - Correlation 1 year apart: 0.3
    - Decay rate: 0.3

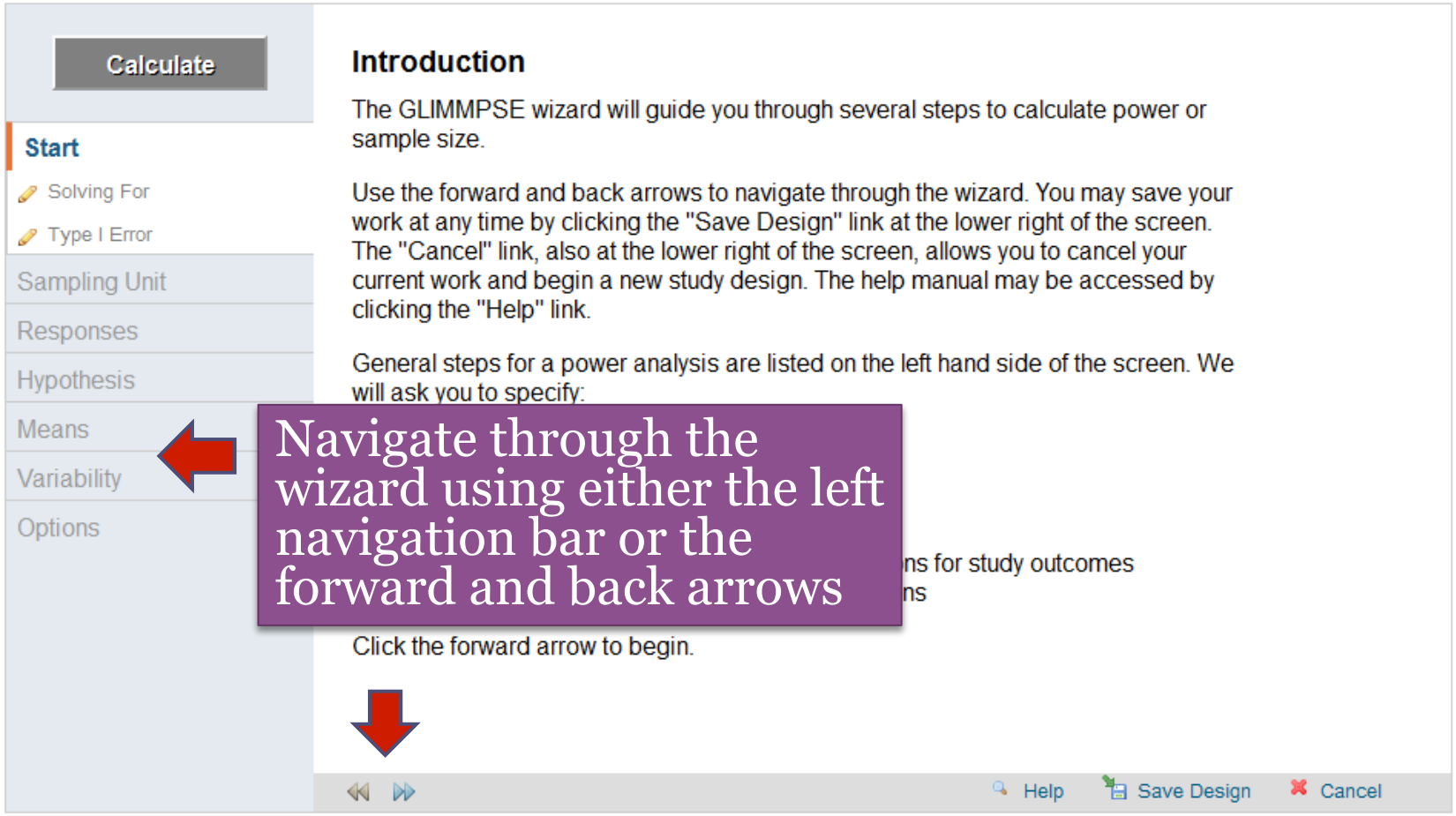
Power with GLIMMPSE



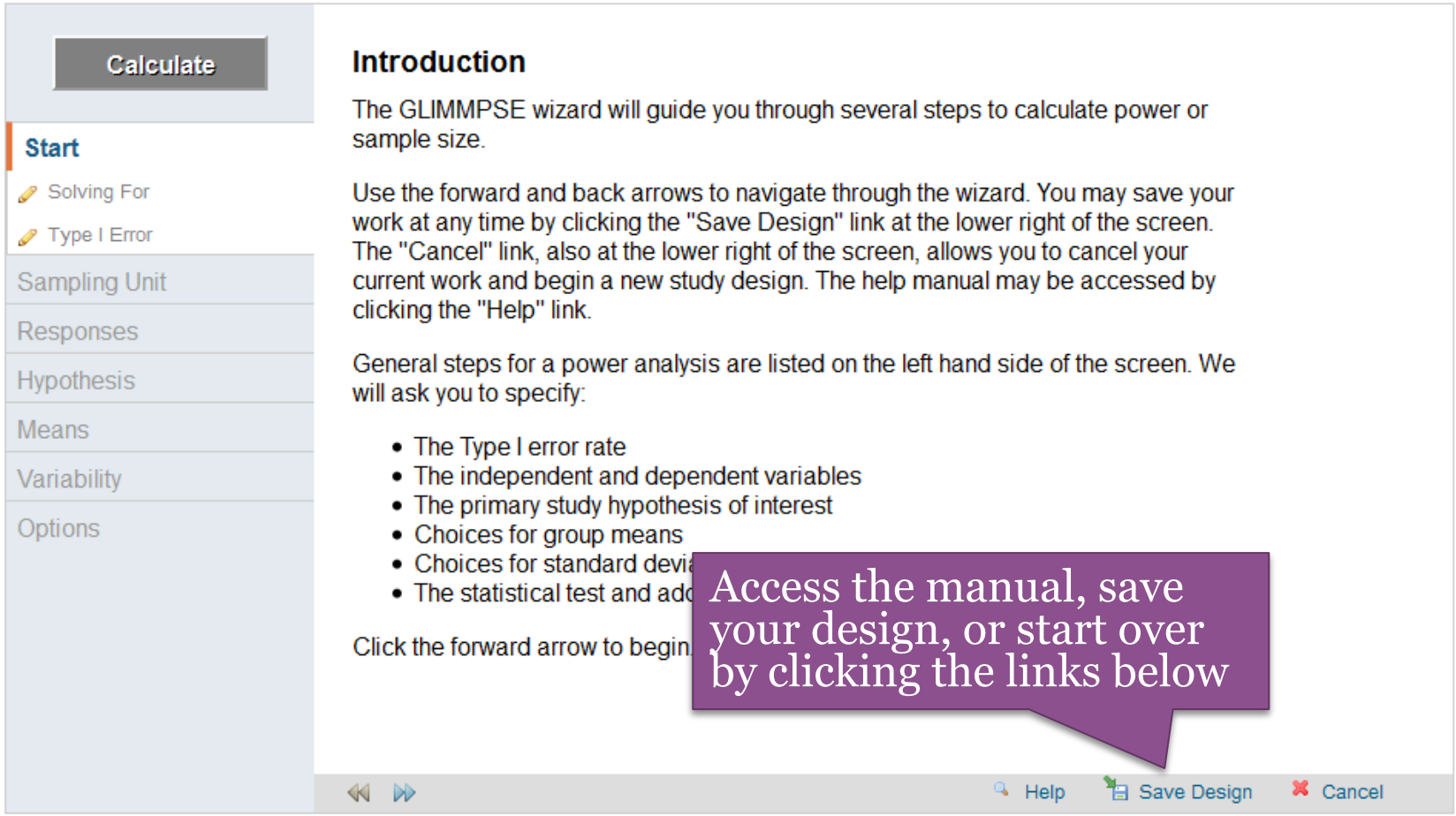
The GLIMMPSE Wizard



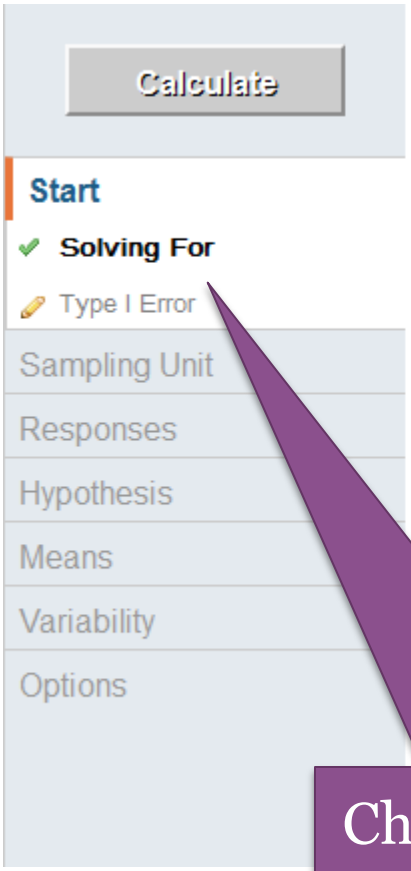
The GLIMMPSE Wizard



The GLIMMPSE Wizard



Solving For



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

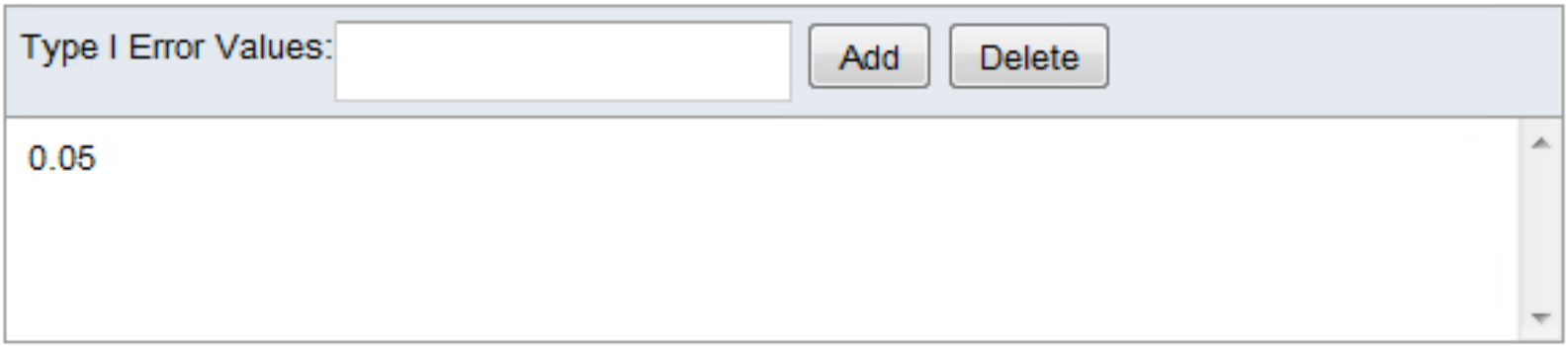
Checkmark = complete
Pencil = incomplete

Type I Error Rate

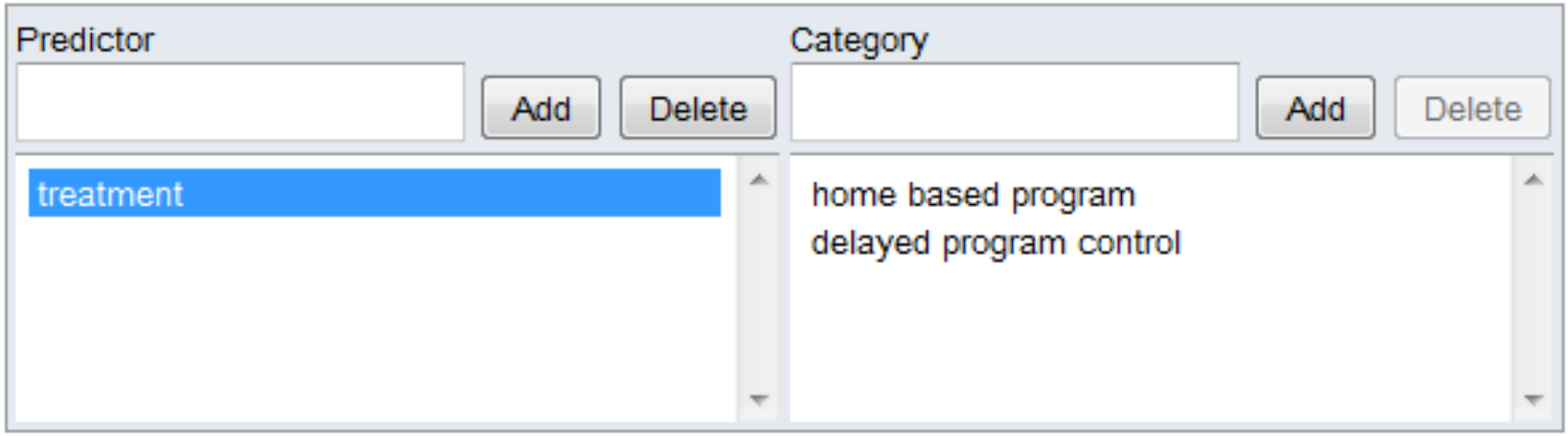
Type I Error

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 α. 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.



GLIMMPSE Predictors



Clustering

- Assumes compound symmetry within a cluster

Remove clustering

Cluster label	community
Number of observations or sub-clusters within each cluster of this type	10
Intra-cluster correlation	0.01

Add subgroup Remove subgroup

Describing Sample Size

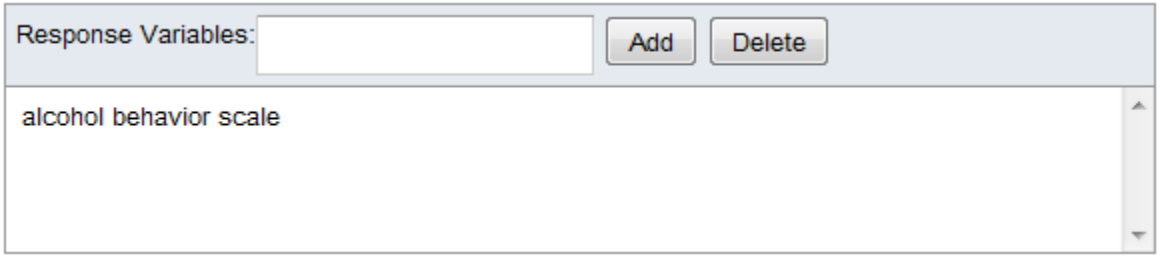
- Relative group size

Relative Group Size	treatment
1	home based program
1	delayed program control

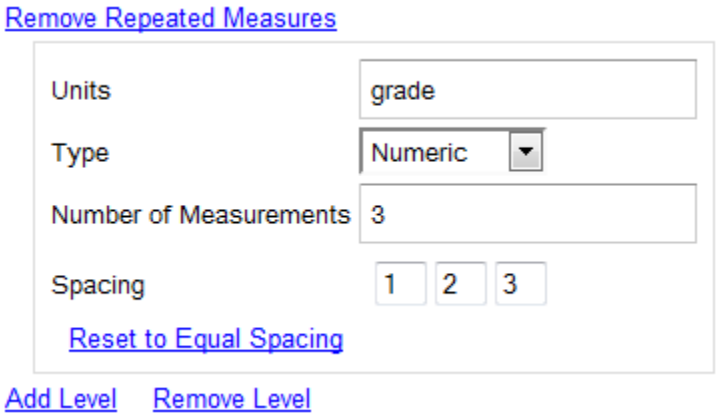
- Smallest group size

Describing Responses

- Response Variables



- Repeated Measures



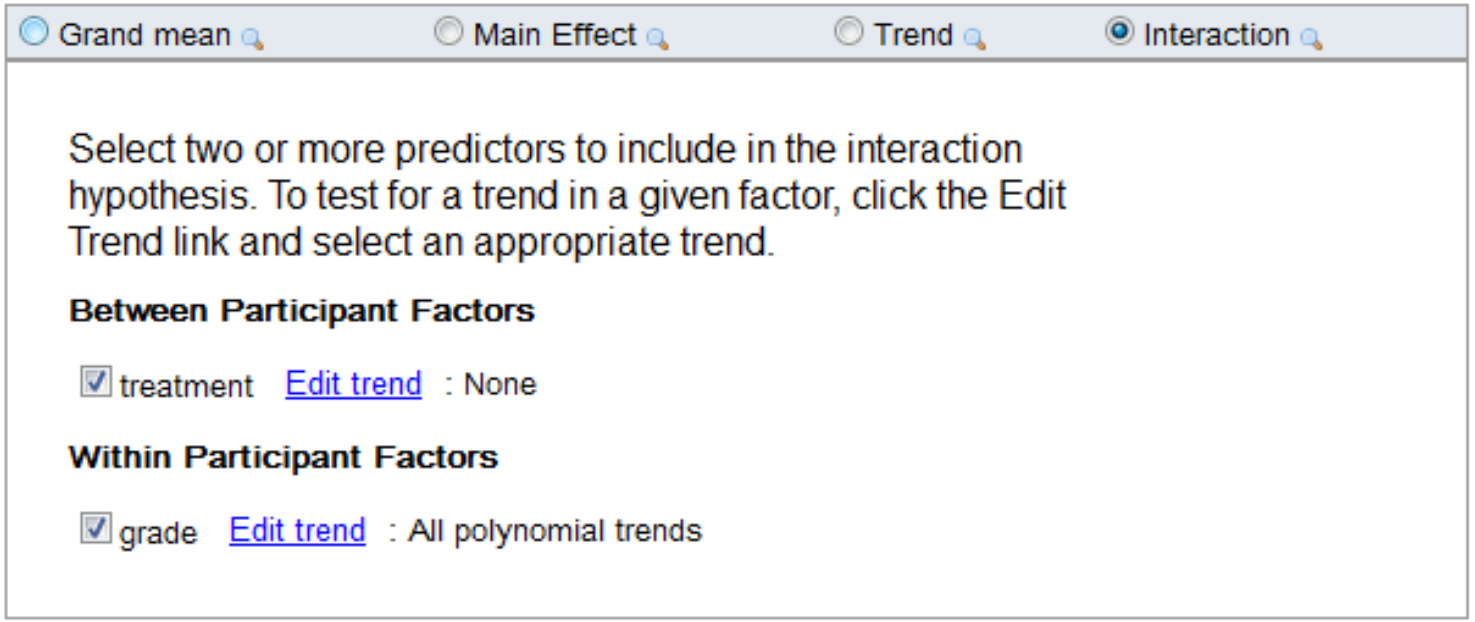
Clustering or Repeated Measures?

- Clustering
  - Same correlation between any two observations
  - Limit of 3 levels
  - Only computational limits on cluster size

- Repeated Measures
  - Allows complex covariance structures such as the Lear model
  - Limit of 3 levels
  - Limit of 10 repeated measures per level

Specifying a Hypothesis

- ❑ Identify the type of hypothesis
- ❑ Select the factors included in the hypothesis



Entering Means

- Enter raw means or the "clinical difference"

treatment	alcohol behavior scale
home based program	-0.25
delayed program control	0

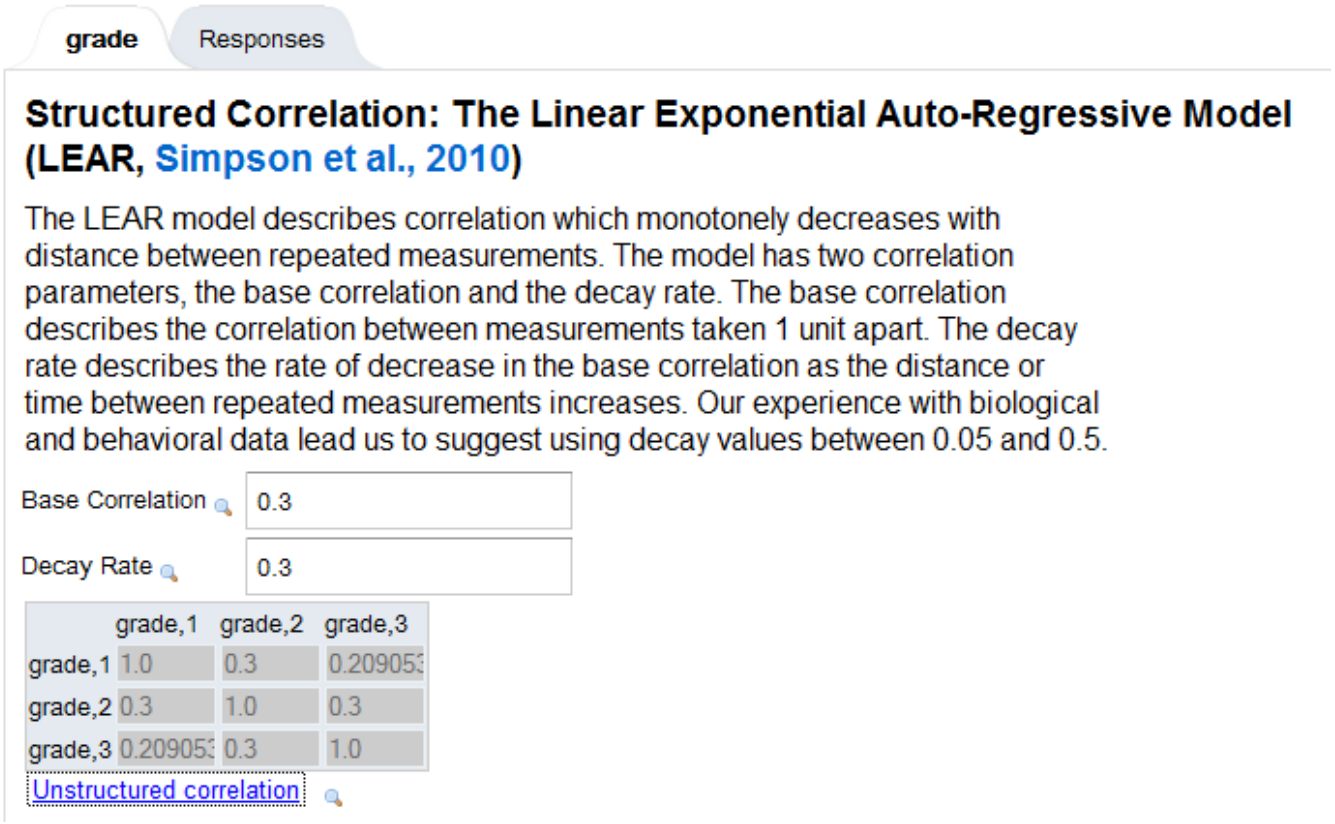
Select the time (location, etc.) from the list(s) below. This will etc.).

grade 3

- For repeated measures, be sure to enter means for each time point!

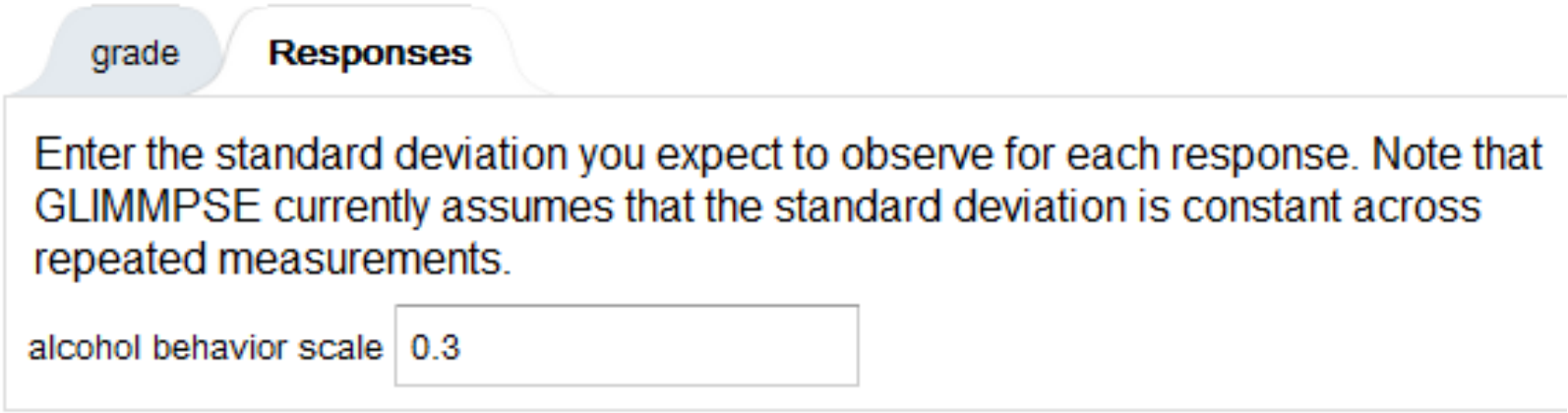
Variability

- Enter the correlation across repeated measures



Variability

- Enter the standard deviation for the response variable



- For multiple response variables, enter the pair-wise correlations

Selecting a Statistical Test

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.

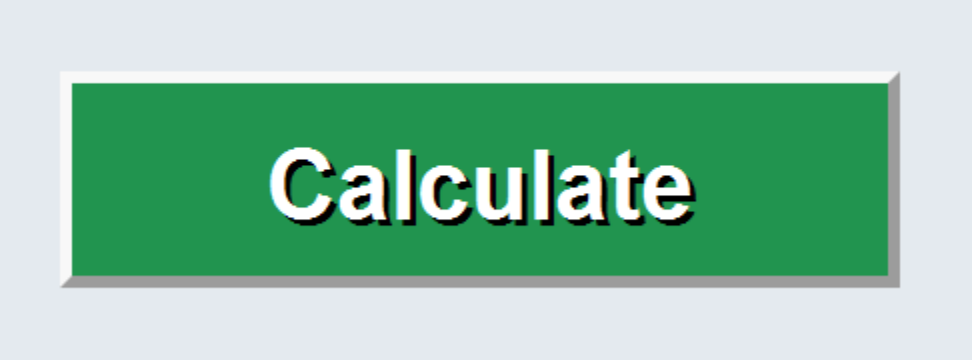
Click here to learn more about selecting an appropriate test.

- [x] 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

Other Options

- Scale factors for means
- Scale factors for variability
- Power curves
- Confidence intervals

GLIMMPSE Calculate Button



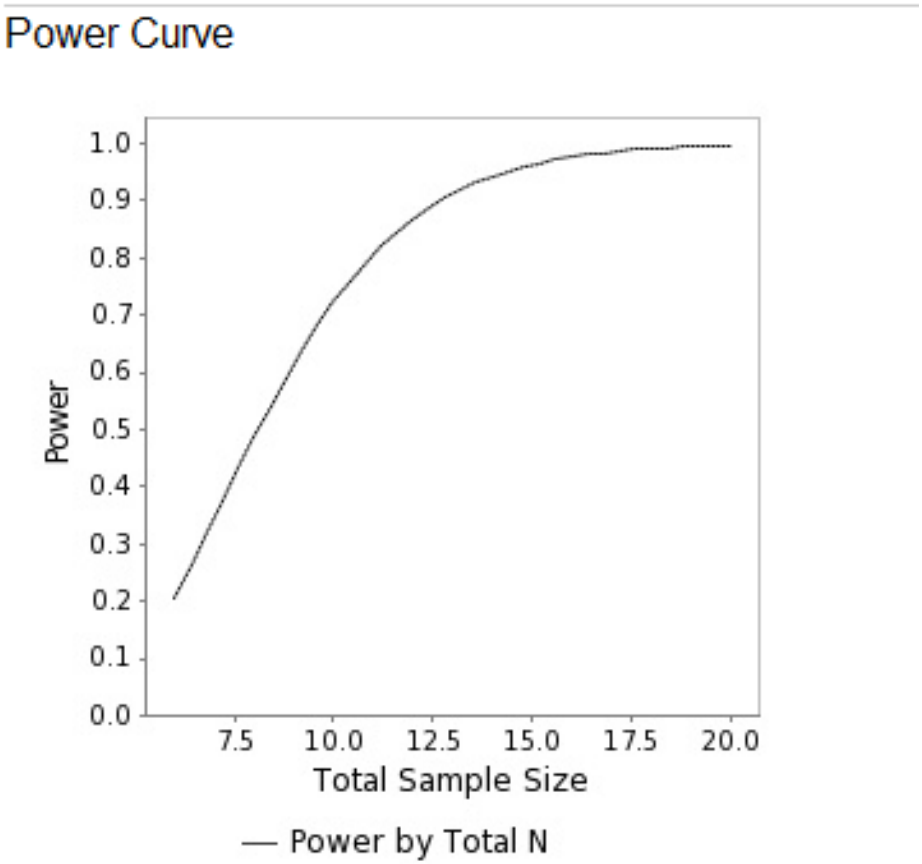
Results

Power Results

Power	Total Sample Size	Test	Type I Error Rate	Means Scale Factor	Variability Scale Fact
0.206	6	HLT	0.05	1	1
0.485	8	HLT	0.05	1	1
0.722	10	HLT	0.05	1	1
0.867	12	HLT	0.05	1	1
0.942	14	HLT	0.05	1	1
0.976	16	HLT	0.05	1	1
0.991	18	HLT	0.05	1	1
0.997	20	HLT	0.05	1	1

Save to CSV View Matrices

Results

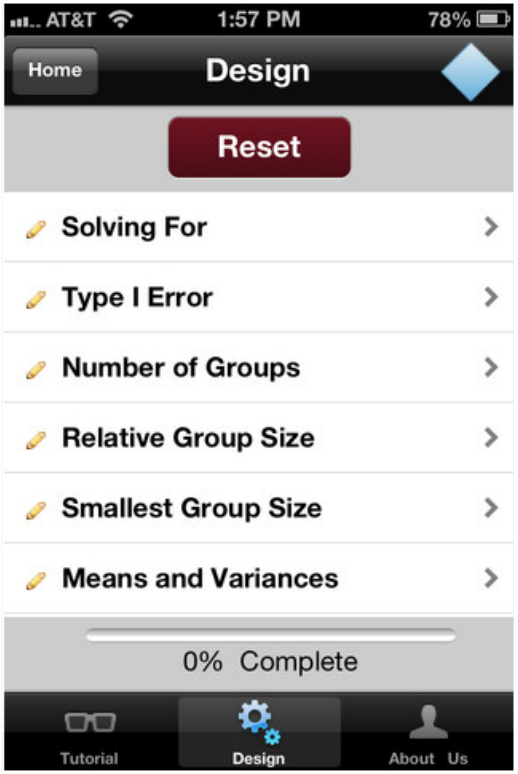

Power Curve
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
7.5
10.0
12.5
15.0
17.5
20.0
Total Sample Size
Power by Total N

Power Calculation Summary

Ten communities were randomized to receive either the home based intervention or delayed intervention. Ten students were recruited from each community. The intracluster correlation within community was assumed to be 0.01. Correlation between repeated alcohol behavior scores within a student was assumed to be 0.3 for measures taken one year apart, with gradual decay over time. Power was calculated for a time by treatment interaction using the Hotelling-Lawley trace test. For a Type I error rate of 0.05, and an assumed standard deviation of 0.3 for alcohol behavior scores, the study had 97.7% power to detect a difference of 0.25 in a time by treatment interaction.

GLIMMPSE Lite for Mobile

- Power for one-way ANOVA on your iPhone or Android device
- Available from Google Play and the Apple Store



https://itunes.apple.com/app/glimmpse-lite/id594924574
https://play.google.com/store/apps/details?id=edu.ucdenver.bios.glimmpseandroid&hl=en

Summary

- Power and sample size calculations are a critical part of study design
- Answers to basic questions about the study design can lead investigators to an appropriate sample size calculation
- GLIMMPSE is a free, web-based tool to aid in calculating power or sample size for a variety of multilevel and longitudinal designs

Session Outline

Introduction Dr. Mildred Maldonado-Molina	2:00 – 2:05
Power and Sample Size for the Most Common Hypotheses in Mixed Models Dr. Anna Barón	2:05 – 2:20
Selecting a Covariance Model for Longitudinal and Multilevel Designs Dr. Mildred Maldonado-Molina	2:20 – 2:35
Power Analysis for Mixed Models: Using Free Web-Based Power Software Dr. Sarah Kreidler	2:35 – 3:05
Discussant Dr. Jacinda Dariotis	3:05 – 3:15
Question and Answer	3:15 – 3:30

POWER, SAMPLE SIZE, MIXED MODELS, REPEATED MEASURES, CLUSTERING, … AND THE KITCHEN SINK

Jacinda K. Dariotis, PhD, MAS

Johns Hopkins Bloomberg School of Public Health

Center for Adolescent Health

21. March 2013

Main Points – Sample Size

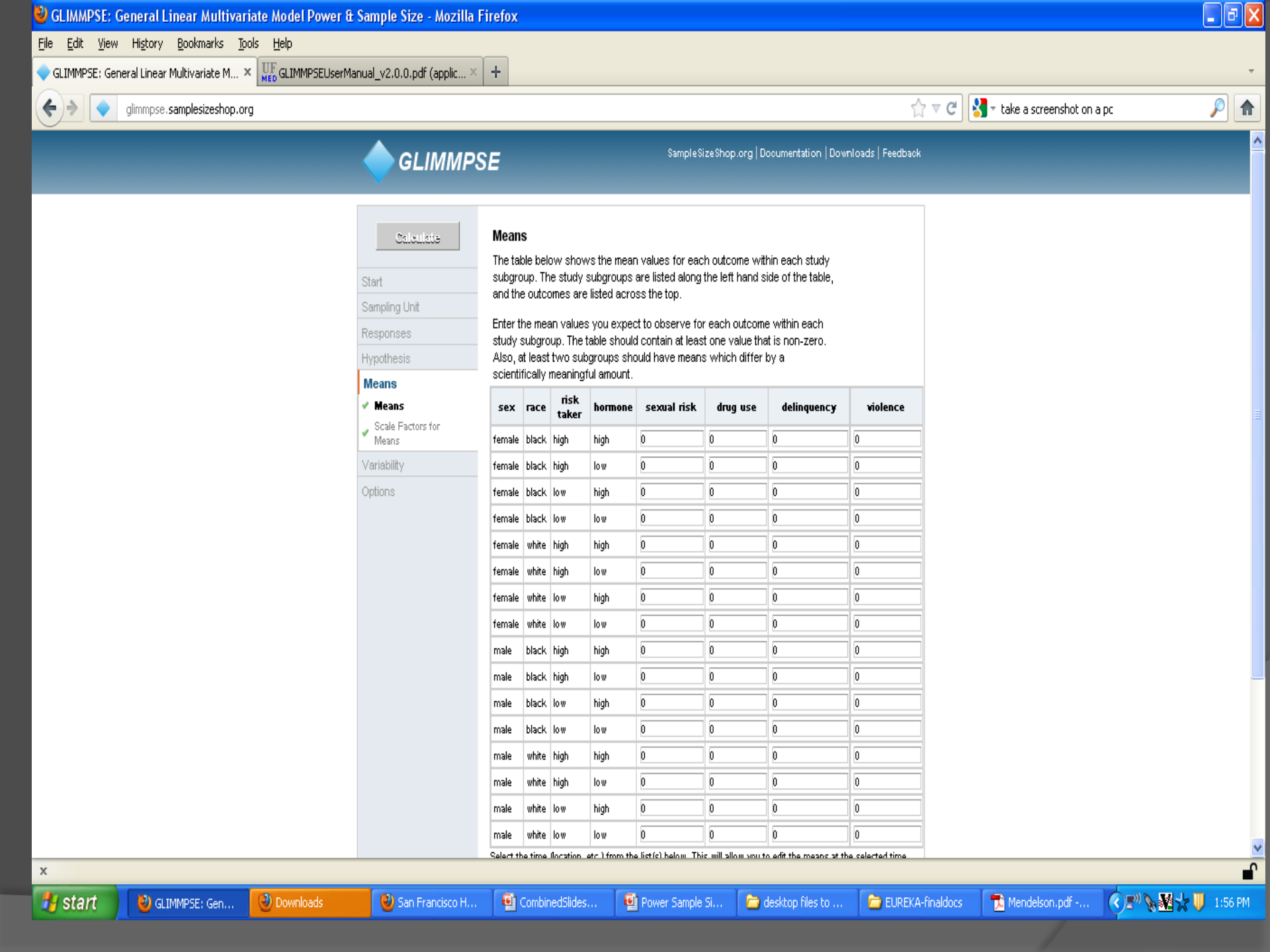
- Ethical considerations
  - Respondent burden - biomedical, biosocial
- Cost
  - Time & money
- Dangers in drawing inferences
  - Overpowered → overinflated, not meaningful
  - Underpowered → not possible to reject null anyway
- Solution
  - Just enough sample size for proper inferences
  - GLIMMPSE

HONESTY Project - example

- Respondent Burden – 8 hours in the lab
  - Biospecimens: swabs, spit, urines
  - Imaging: 2 hours ($1250)
  - Survey: long
  - 12 month follow-up: Weekly text messaging
- At which level do I calculate sample size?
  - Voxels (~ millions)
  - Brain regions (~ 200)
  - Hormones levels (~8 samples; area under curve)
  - Time points (~ 52 weeks or 2 in-person visits)
  - Subgroups (e.g., bio sex; race; risk-takers)

GLIMMPSE

- Purpose: assist with calculations for complex designs
  - “This mode is designed for applied researchers including physicians, nurses, and other investigators.”
  - Especially for non-statistical investigators
- My attempts
  - HONESTY texting example
  - Mindfulness school intervention


GLIMMPSE: General Linear Multivariate Model Power & Sample Size - Mozilla Firefox
File Edit View History Bookmarks Tools Help
GLIMMPSE: General Linear Multivariate M...
GLIMMPSEUserManual_v2.0.0.pdf (applic...
glimmpse.samplesizeshop.org
take a screenshot on a pc
GLIMMPSE
SampleSizeShop.org | Documentation | Downloads | Feedback
Calculate
Start
Sampling Unit
Responses
Hypothesis
Means
Means
Scale Factors for Means
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.
sex race risk taker hormone sexual risk drug use delinquency violence
female black high high 0 0 0 0
female black high low 0 0 0 0
female black low high 0 0 0 0
female black low low 0 0 0 0
female white high high 0 0 0 0
female white high low 0 0 0 0
female white low high 0 0 0 0
female white low low 0 0 0 0
male black high high 0 0 0 0
male black high low 0 0 0 0
male black low high 0 0 0 0
male black low low 0 0 0 0
male white high high 0 0 0 0
male white high low 0 0 0 0
male white low high 0 0 0 0
male white low low 0 0 0 0
start
GLIMMPSE: Gen...
Downloads
San Francisco H...
CombinedSlides...
Power Sample Si...
desktop files to ...
EUREKA-finaldocs
Mendelson.pdf -...
1:56 PM

GLIMMPSE - observations

- Benefits
  - Tries to estimate sample size the right way (taking into account levels, clustering/ correlations)
  - Great user interface
  - Definitions embedded
  - May specify several values of betas and alphas
  - Save your design

GLIMMPSE - observations

- Concerns
  - Requires a lot of information and knowledge researchers may not have
    - …just dangerous enough…
  - May be best targeted for more sophisticated researchers
  - Limited to “simpler” complex designs
  - Assumptions may not be realistic
    - (equal spacing, no missing data, normally distributed data, outcomes as categorical/ordinal/count)

Possible Next Steps…

- Relaxing some of the assumptions
  - Non-normatively distributed data
  - Moving beyond binary or count data
  - Intensive longitudinal data (>10 time points)
- What about novel studies without known confidence intervals, means differences, etc.?
- What about using effect sizes?
- Is the reverse possible (specifying intended sample size and generating estimates of difference)?

THANK YOU.

Session Outline

Introduction Dr. Mildred Maldonado-Molina	2:00 – 2:05
Power and Sample Size for the Most Common Hypotheses in Mixed Models Dr. Anna Barón	2:05 – 2:20
Selecting a Covariance Model for Longitudinal and Multilevel Designs Dr. Mildred Maldonado-Molina	2:20 – 2:35
Power Analysis for Mixed Models: Using Free Web-Based Power Software Dr. Sarah Kreidler	2:35 – 3:05
Discussant Dr. Jacinda Dariotis	3:05 – 3:15
Question and Answer	3:15 – 3:30

References

Adams, G., Gulliford, M. C., Ukoumunne, O. C., Eldridge, S., Chinn, S., & Campbell, M. J. (2004). Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *Journal of clinical epidemiology*, *57*(8), 785-794.

Catellier, D. J., & Muller, K. E. (2000). Tests for gaussian repeated measures with missing data in small samples. *Statistics in Medicine*, *19*(8), 1101-1114.

Demidenko, E. (2004). *Mixed Models: Theory and Applications* (1st ed.). Wiley-Interscience.

Glueck, D. H., & Muller, K. E. (2003). Adjusting power for a baseline covariate in linear models. *Statistics in Medicine*, *22*, 2535-2551.

References

Gedney , J.J., Logan, H.L., Baron, R.S. (2003). Predictors of short-term and long-term memory of sensory and affective dimensions of pain. *Journal of Pain*, 4(2), 47–55.

Gedney, J.J., Logan H.L. (2004). Memory for stress-associated acute pain. *Journal of Pain*, 5(2), 83–91.

Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, *30*(22), 2696-2707. doi:10.1002/sim.4293

Kerry, S. M., & Bland, J. M. (1998). The intracluster correlation coefficient in cluster randomisation. *BMJ (Clinical research ed.)*, *316*(7142), 1455.

References

Kreidler, S.M., Muller, K.E., Grunwald, G.K., Ringham, B.M., Coker-Dukowitz, Z.T., Sakhadeo, U.R., Barón, A.E., Glueck, D.H. (accepted). GLIMMPSE: Online Power Computation for Linear Models With and Without a Baseline Covariate. *Journal of Statistical Software*.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974.

Law, A., Logan, H., & Baron, R. S. (1994). Desire for control, felt control, and stress inoculation training during dental treatment. *Journal of Personality and Social Psychology*, *67*(5), 926-936.

Logan, H.L., Baron, R.S., Keeley, K., Law, A., Stein, S. (1991). Desired control and felt control as mediators of stress in a dental setting. *Health Psychology, 10(5)*, 352–359.

References

Logan, H.L., Baron, R.S., Kohout, F. (1995). Sensory focus as therapeutic treatments for acute pain. *Psychosomatic Medicine*, 57(5), 475–484.

Muller, K. E, & Barton, C. N. (1989). Approximate Power for Repeated-Measures ANOVA Lacking Sphericity. *Journal of the American Statistical Association*, *84*(406), 549-555.

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*(19), 3639-3660.

Muller, K. E, Lavange, L. M., Ramey, S. L., & Ramey, C. T. (1992). Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications. *Journal of the American Statistical Association*, *87*(420), 1209-1226.

References

Muller, K. E, & Peterson, B. L. (1984). Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis. *Computational Statistics and Data Analysis*, *2*, 143-158.

Muller, K.E., & Stewart, P. W. (2006). *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. Hoboken, NJ: Wiley.

Taylor, D. J., & Muller, K. E. (1995). Computing Confidence Bounds for Power and Sample Size of the General Linear Univariate Model. *The American Statistician*, *49*(1), 43-47. doi:10.2307/2684810