Easy Power and Sample Size for Most of the Mixed Models You Will Ever See

Keith E. Muller
Chief, Division of Methodology
Department of Health Outcomes and Policy
University of Florida, KMuller@ufl.edu

Anna E. Barón (Colorado SPH)
Sarah M. Kreidler (Colorado SPH)
Yueh-Yun Chi (Florida Biostatistics)
Deborah H. Glueck (Colorado SPH)

http://samplesizeshop.org/

Sponsorship

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.

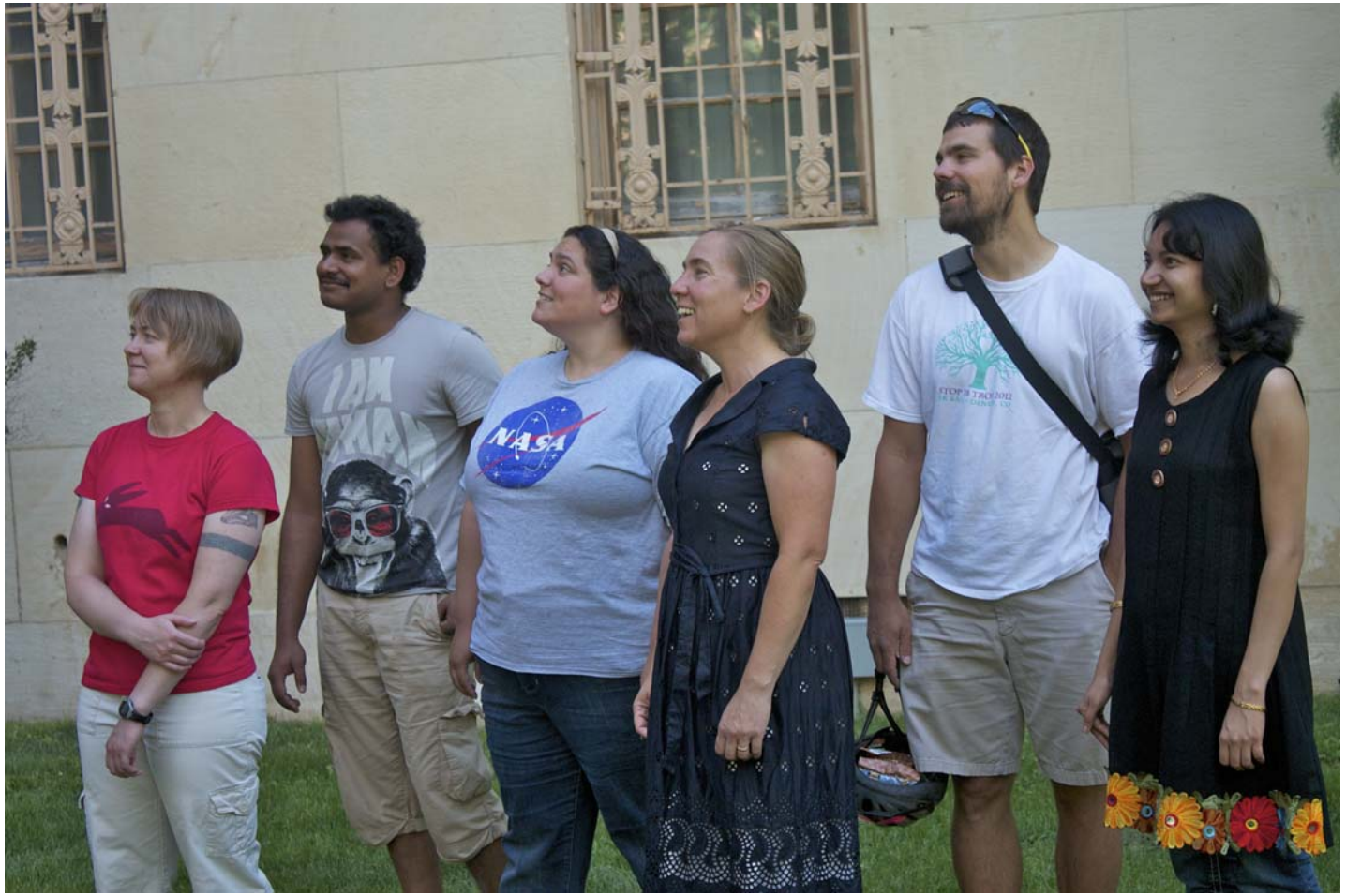
Additional support from NIAAA R01-AA013458-01, NHLBI R01-HL091005, and NIDCR U54-DE019261.

Saga of Sample Size Selection

- We have long needed to select sample size for designs with clusters, repeated measures and multiple outcomes, and now we see combinations.
- Existing approaches: 1) simulations,
  2) exemplary data,
  3) large sample approximations, and
  4) special cases.
- We think the ideas and software we present today make the job easier than ever before.
- The first version of our free power software was written 30 years ago.
- Previous versions matrix based, user hostile.

Now point and click (GUI).

Software Development Team UCD



Outline

1. Characterizing the Linear Models You See
   - Common Tests in the Linear Mixed Model (LMM)
   - The LMM as a General Linear Multivariate Model
2. Six-Step Checklist for Power and Sample Size Analysis
   - Two Real Design Examples
   - Using the Checklist for the Examples
3. Simple Adjustments for Power with Missing Data
4. Free, Web-based Software, GLIMMPSE, and Related Web Resources

1. Characterizing The Linear Models You See

General Linear Mixed Model Commonly Used for Clustered and Repeated Measures Data

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

Data Structures

Clustering

Clustering is a restricted form of multilevel data,
a special simple case of importance that we can cover.

Repeated Measures

Repeated measures is a restricted form of multilevel data,
a special simple case of importance that we can cover.

Distinguishing the Independent Sampling Unit from the Observational Unit

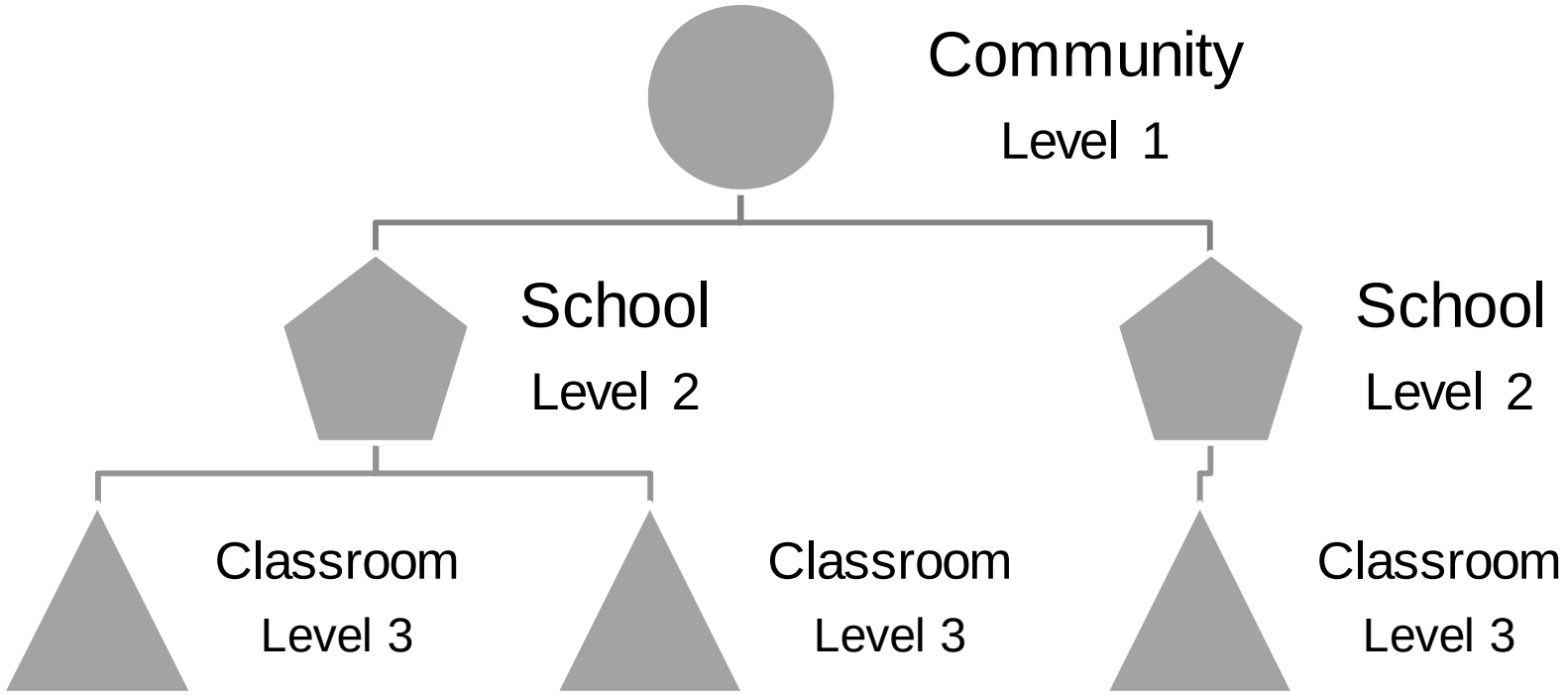
Observational Unit: alcohol use by
child
in a school
in a particular year

Independent Sampling Unit (ISU): the school.
ISU, the "Subject," is not always the participant.

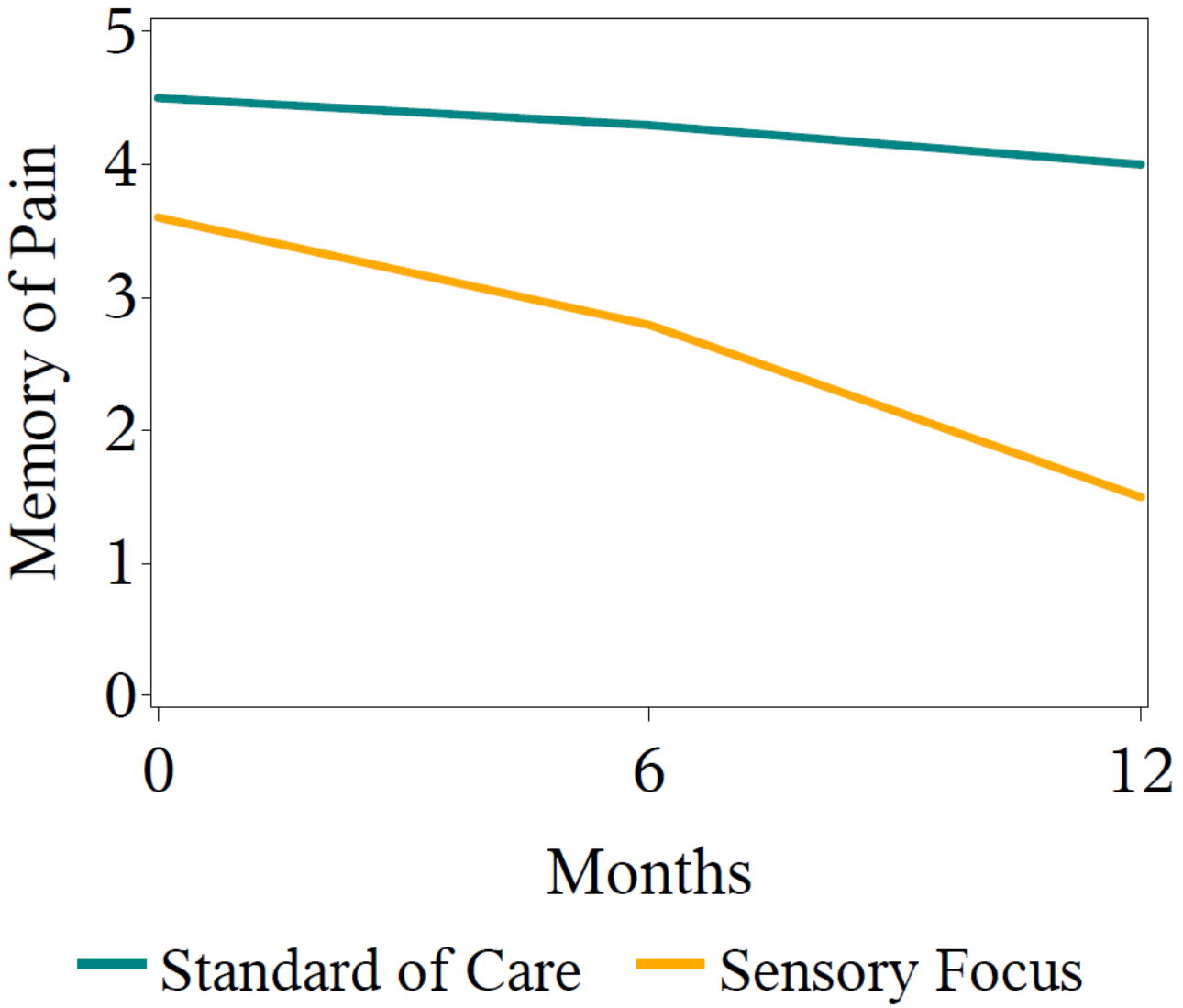
N ISU: How many clusters, here schools.

n total observations sums over ISU, clusters, times.

Clustering in One or More Levels



Repeated Measures: Memory of Pain Trial



Three Classes of Tests in the General Linear Mixed Model

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

General and accurate power and sample size tools are not available.

There are good methods for most common tests in A.

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

- Many General Linear Mixed Model tests can be recast as tests in the General Linear *Multivariate* Model, *GLMM* (Muller and Stewart, 2006; Muller, et al., 2007)

- **Why do we care?**
  - Muller, et al. (1992) show how to do power for time by treatment using *multivariate* framework.

- **We know how to computer power and sample size for a wide class of linear mixed models!**
  - Typical clinical trial or longitudinal study in which main inference is about time by treatment interaction, and others.

Four Requirements for a Reversible LMM Scenario

Requirement	Description
1	"Nice" Design Within ISU
2	Specific Covariance of Responses
3	Wald Test Statistic of Fixed Effects
4	Specific Error df for F Reference Distribution

Four Requirements to Recast a LMM as a GLMM: 1

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

- *Have a "Nice" Design*
  - *No missing or mistimed data, Balanced Within ISU*
  - *Treatment assignment does not change over time; no repeated covariates*
  - *Saturated in time and time by treatment effects*
  - *Unequal ISU group sizes OK*

Four Requirements to Recast a LMM as a GLMM: 2

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

- Fit an Unstructured Covariance Model

Four Requirements to Recast a LMM as a GLMM: 3

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

- Use Wald test for inference about Fixed Effects

- *Most common test used for analysis*

Four Requirements to Recast a LMM as a GLMM: 4

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

- Use Kenward-Roger *F* method

- *df approximation method with modified covariance matrix*
- *With reversibility, covariance matrix is unstructured and test is equivalent to Hotelling-Lawley Trace test*
- *Muller et al. (2007), among many others, showed it gives the best LMM test*

Power and Sample Size for Multivariate Model

- Muller, LaVange, Ramey and Ramey (1992) described power for univariate and multivariate approaches to repeated measures, including Hotelling-Lawley Trace (HLT)
- If data analysis fits unstructured covariance, then Kenward-Roger Wald test equivalent to HLT when reversible (Edwards, et al., 2008)

There is a second path to a reversible scenario.

- If data analysis fits "random intercept" only, then Wald Test with residual method for error df equivalent to *univariate* approach to repeated measures with uncorrected test (Gurka, Edwards, Muller, 2011)

Two Paths to Reversing a LMM Scenario

	Repeated Measures (Test)	
Require	Multivariate (Hotelling)	*Univariate (Uncorrected)*
1	Balanced within ISU	Balanced within ISU
2	Unstructured Covariance	*Compound Symmetric*
3	Wald Test Fixed Effects	Wald Test Fixed Effects
4	Kenward-Roger df	*Residual df*

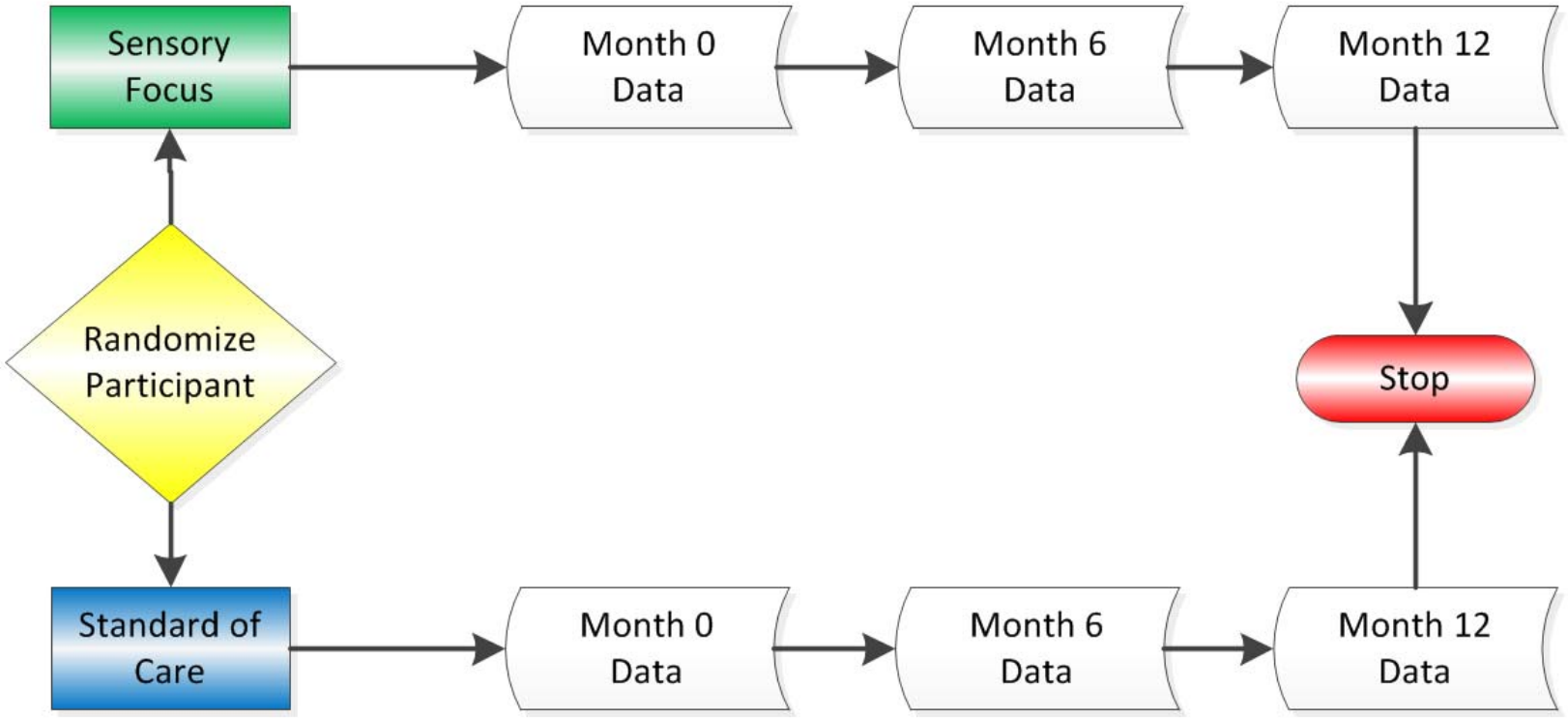
Either column gives an easy path to power.

Outline

1. Characterizing the Linear Models You See
   - Common Tests in the Linear Mixed Model (LMM)
   - The LMM as a General Linear Multivariate Model
2. Six-Step Checklist for Power and Sample Size Analysis
   - Two Real Design Examples
   - Using the Checklist for the Examples
3. Simple Adjustments for Power with Missing Data
4. Free, Web-based Software, GLIMMPSE, and Related Web Resources

First of Two Examples

- Memory of Pain: Proposed study comparing effect of sensory focus intervention to placebo on memory of dental pain (Law et al., 1994; Logan et al., 1995)



Second of Two Examples

- Project Northland Chicago (PNC) Trial: Proposed longitudinal cohort study using data from previous community-randomized controlled trial to test intervention for adolescents (ages 11-14) designed to prevent alcohol use (Komro et al., 2007)

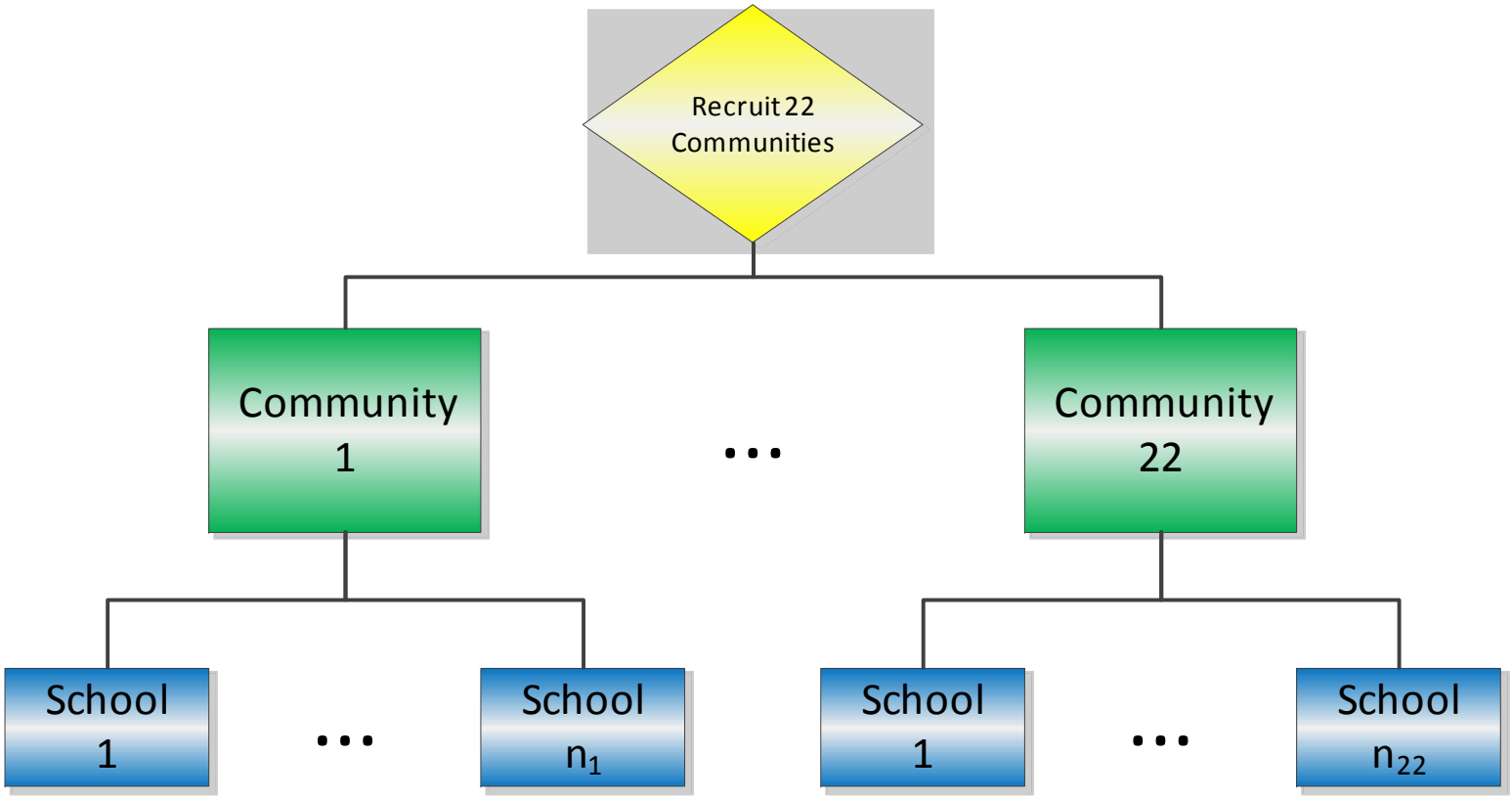
Clustering

Randomization

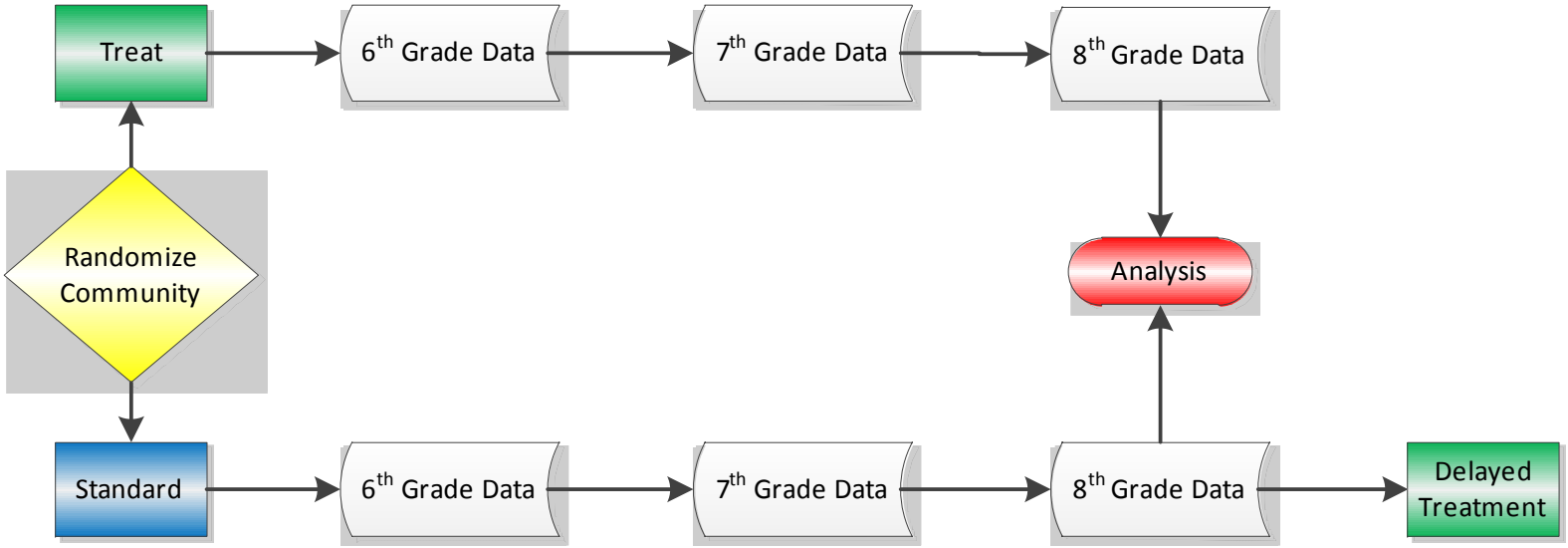
Repeated Measures

The PNC Trial: Cluster Randomized Design

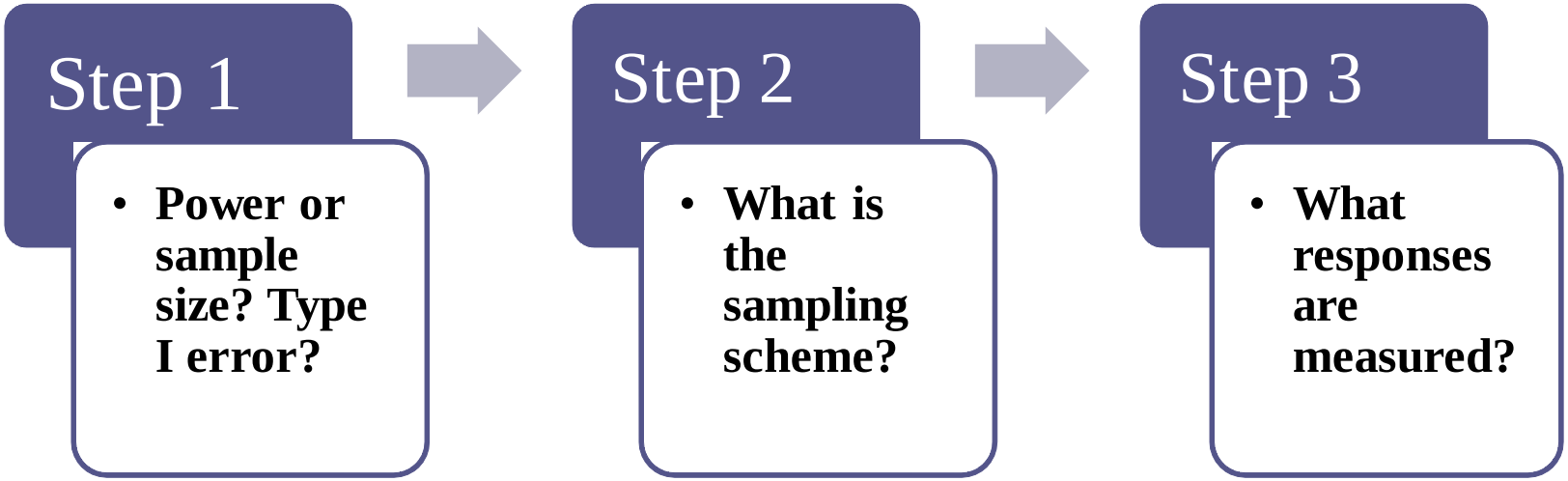
The PNC Trial: Clustering



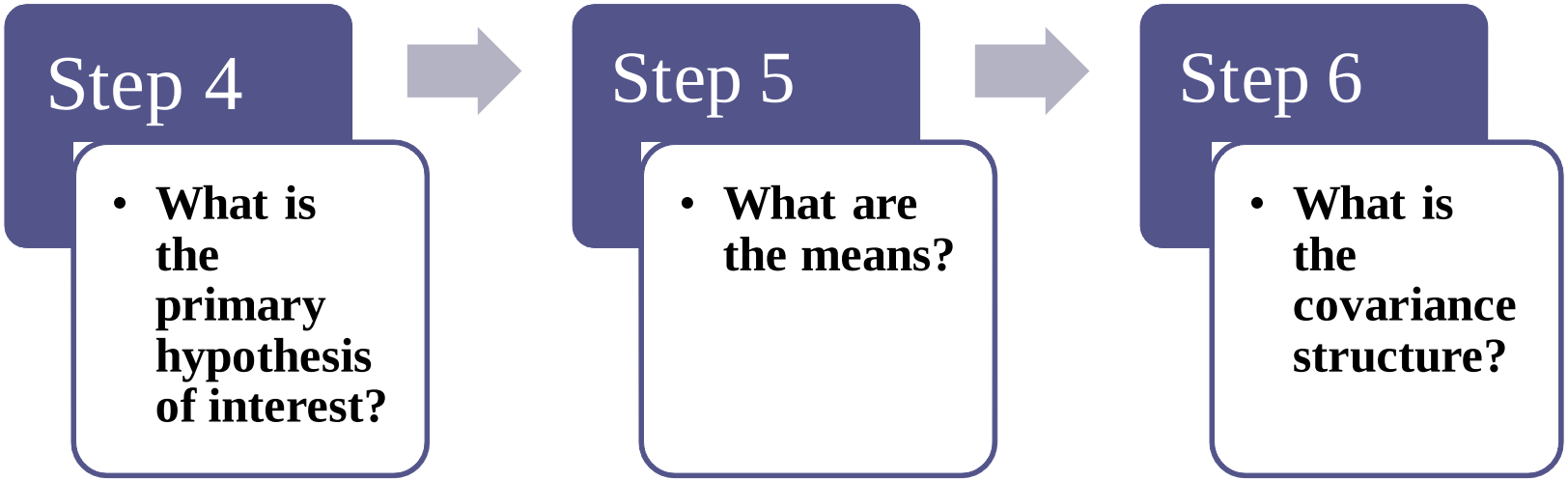
The PNC Trial: Clustering + Randomization



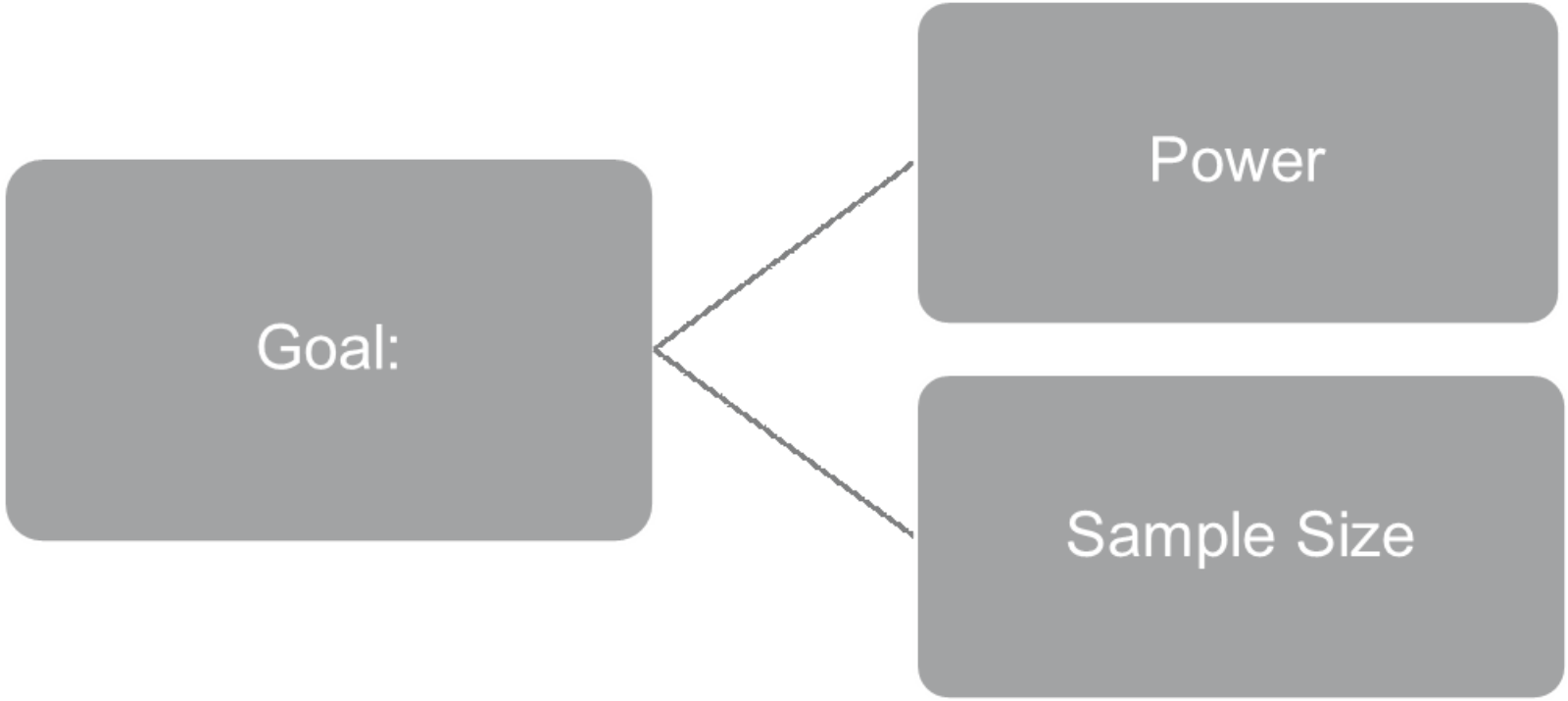
Checklist for Power and Sample Size Analysis



Checklist for Power and Sample Size Analysis



Step 1. What is the Study Design Goal?

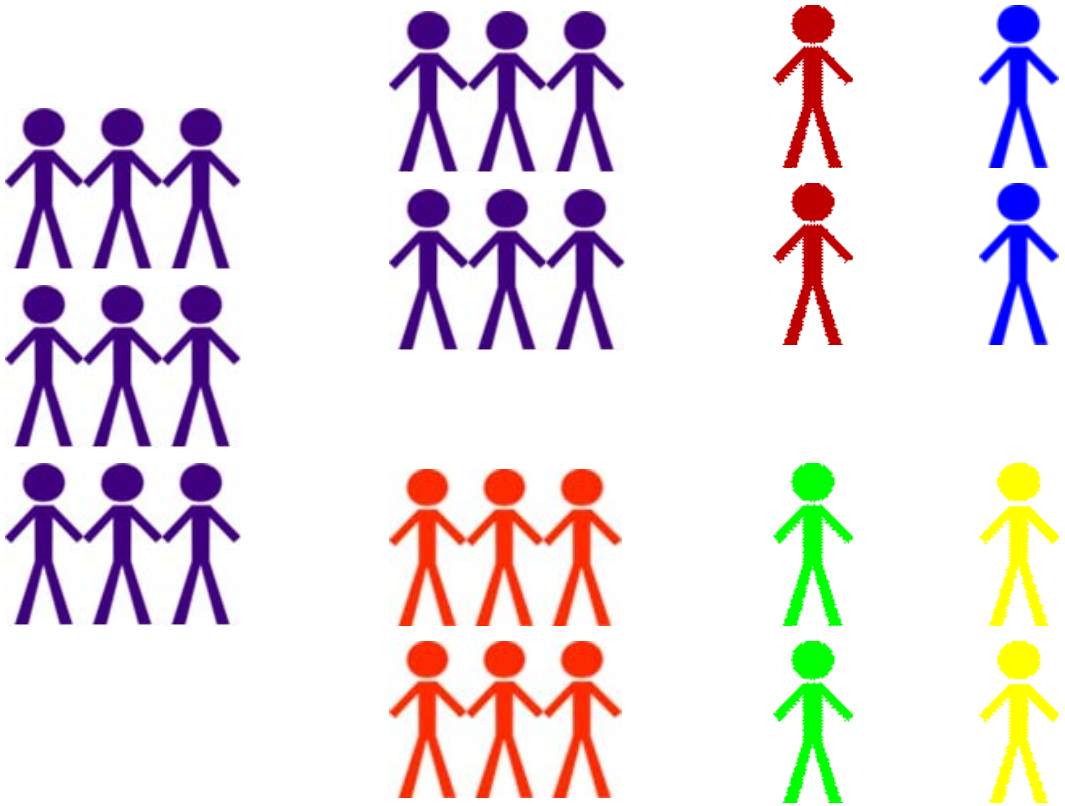


Goal for the Memory of Pain Trial

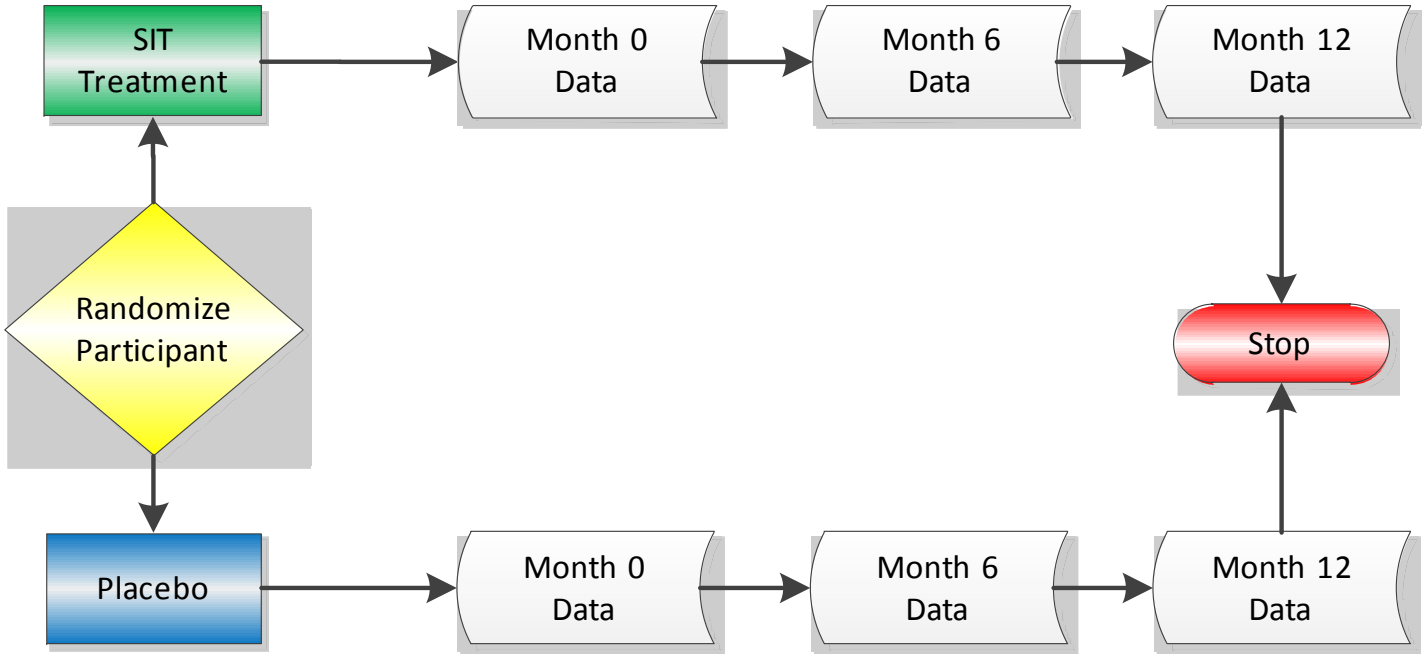
- Determine Sample Size
- Power of 0.9 and (Type I Error Rate) = 0.01
- Primary Hypothesis: Time trend by Treatment Interaction
- Expect the Treated group mean to be 1.2 points lower in Memory of Pain (5-point scale) compared to the Placebo at the last time measurement (12 months)

Step 2a. Specify Study Design Groups

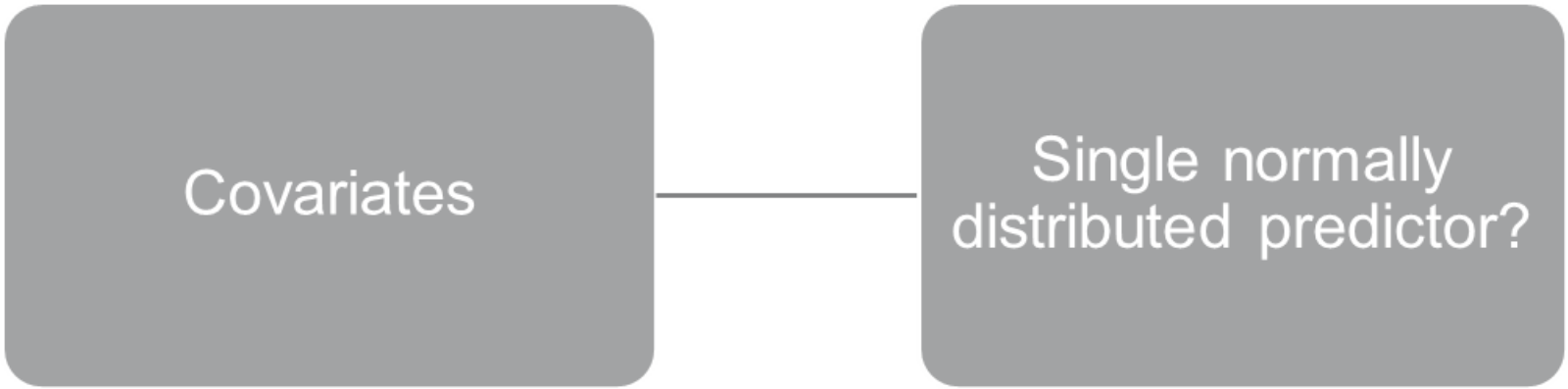
One-sample Two-sample Multi-sample



Two Samples for the Memory of Pain Trial

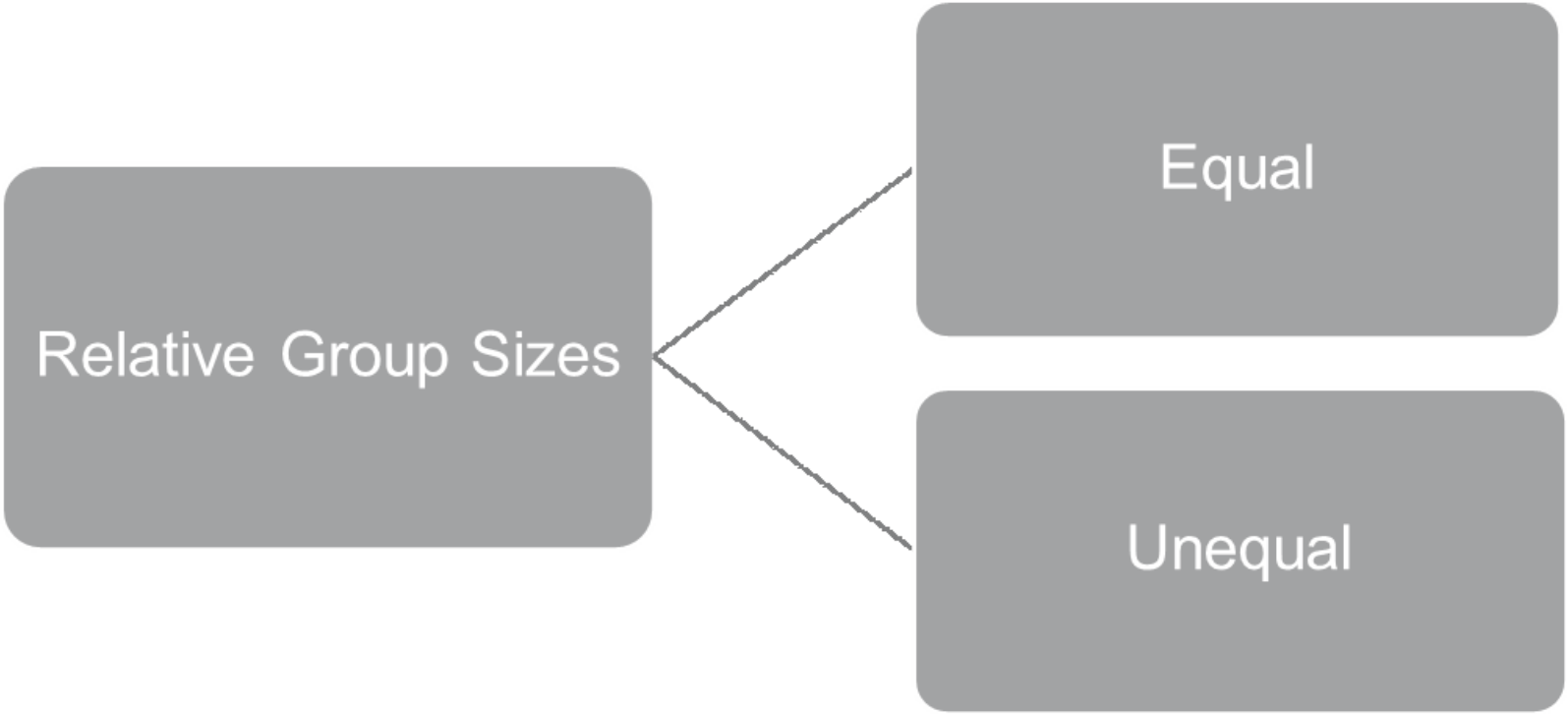


Step 2b. Specify Study Design Covariates

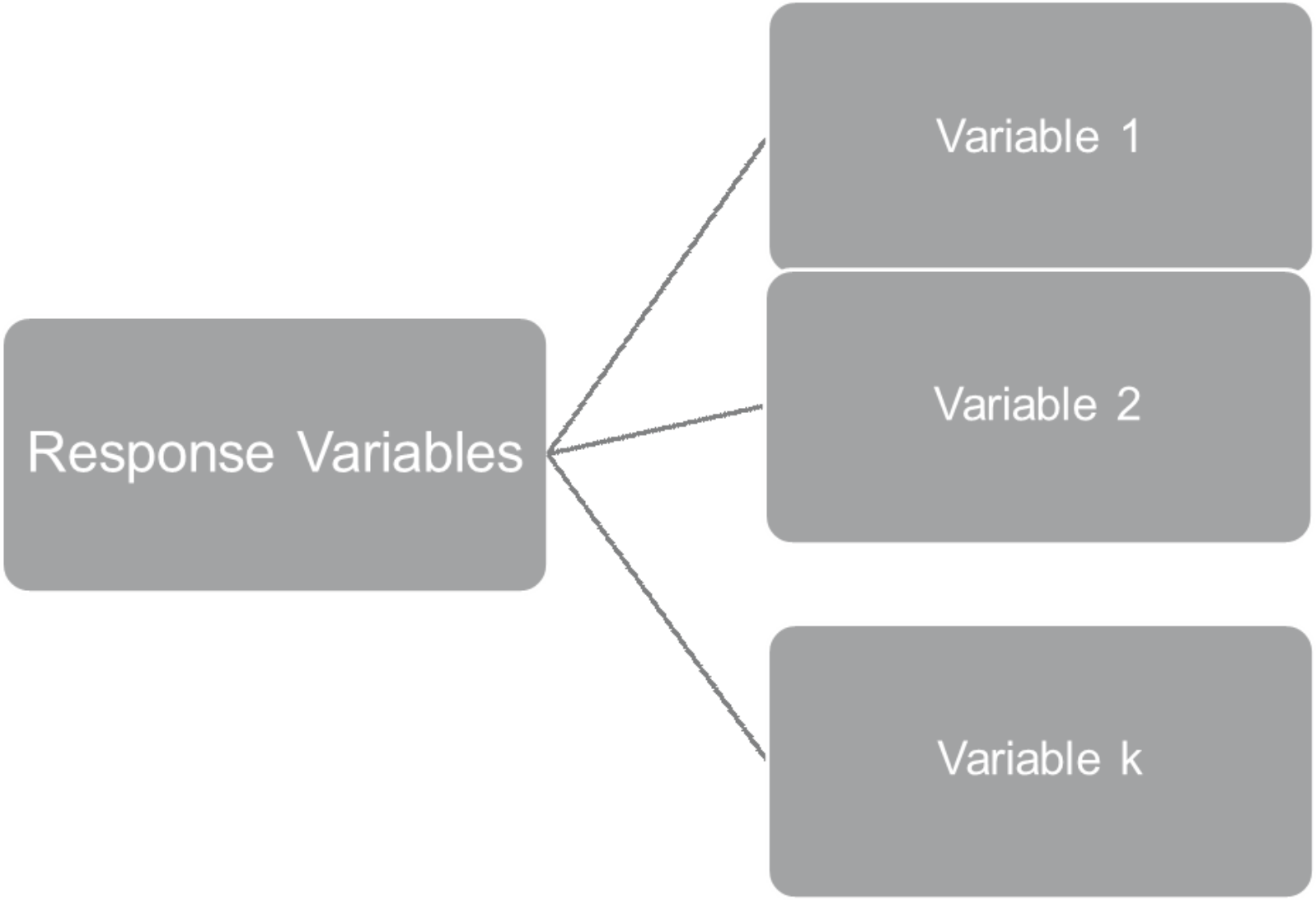


Step 2c. Specify Cluster Sampling Scheme

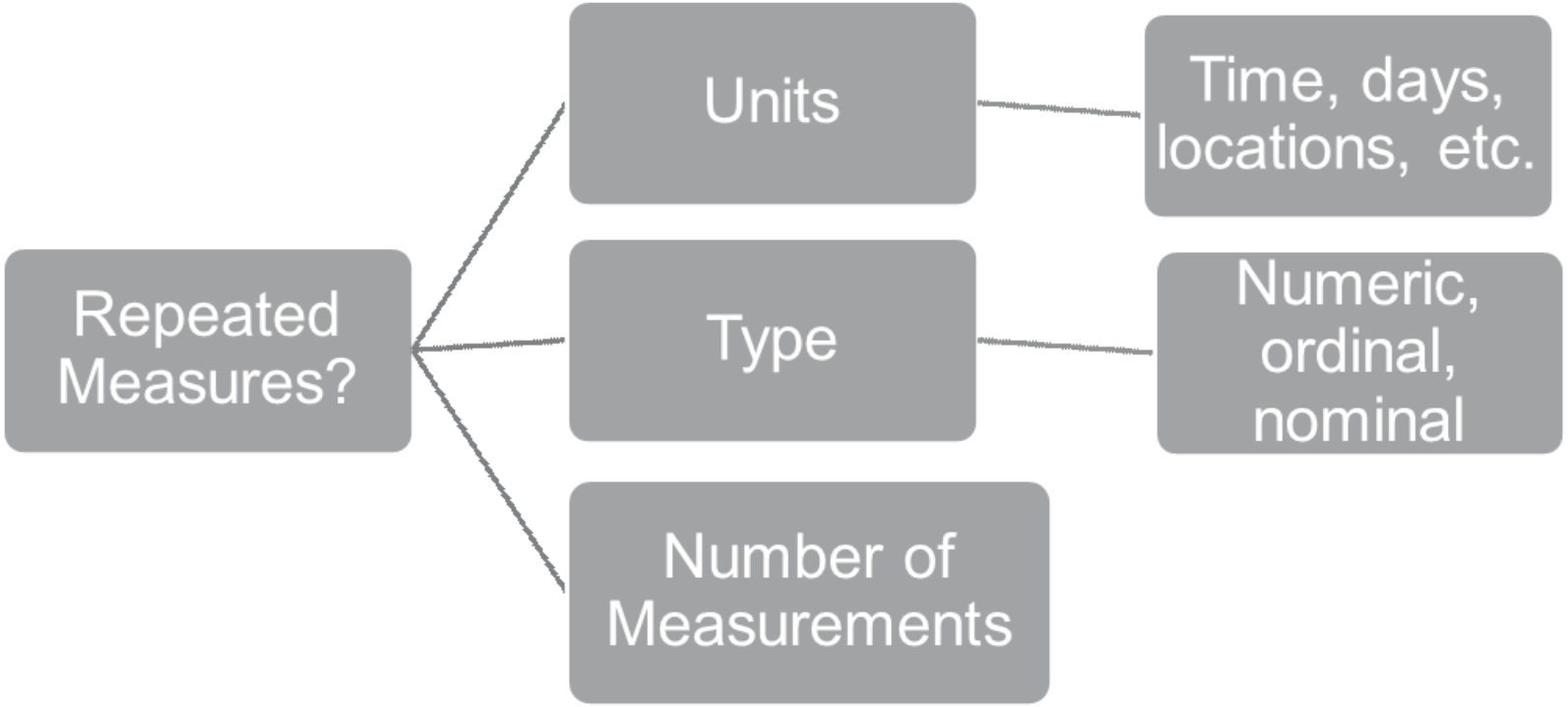
Step 2d. Specify Relative Group Sizes



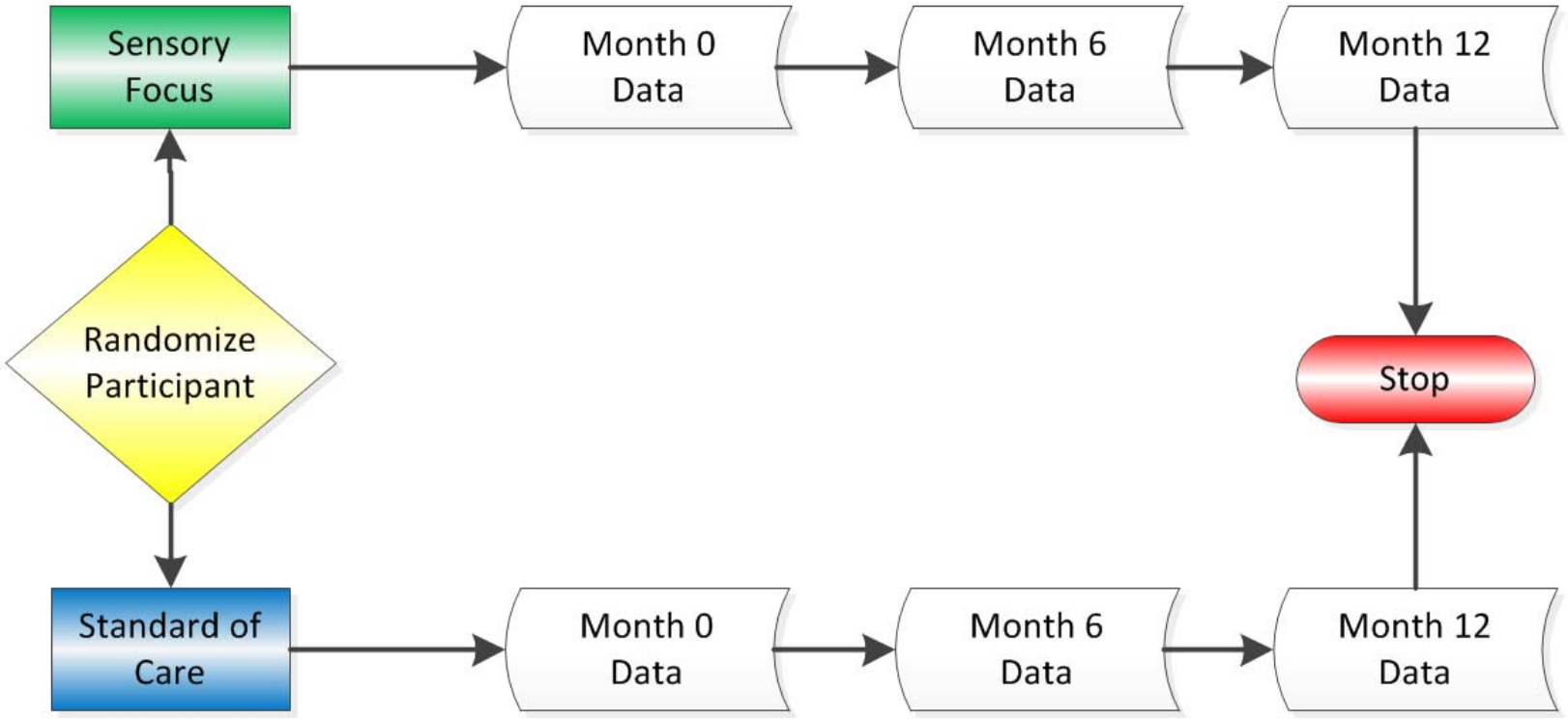
Step 3a. Specify Response Variables



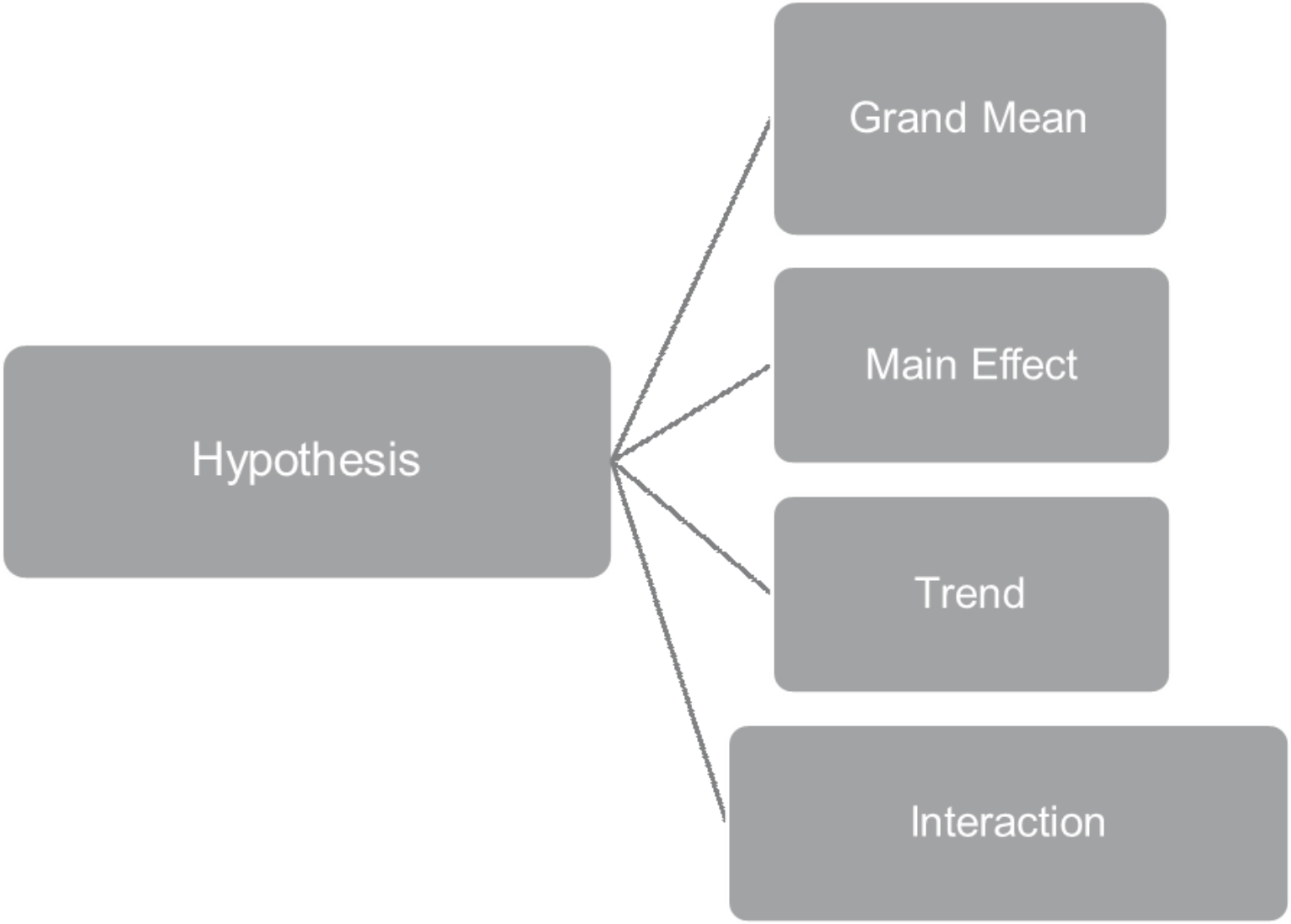
Step 3b. Specify Repeated Measures



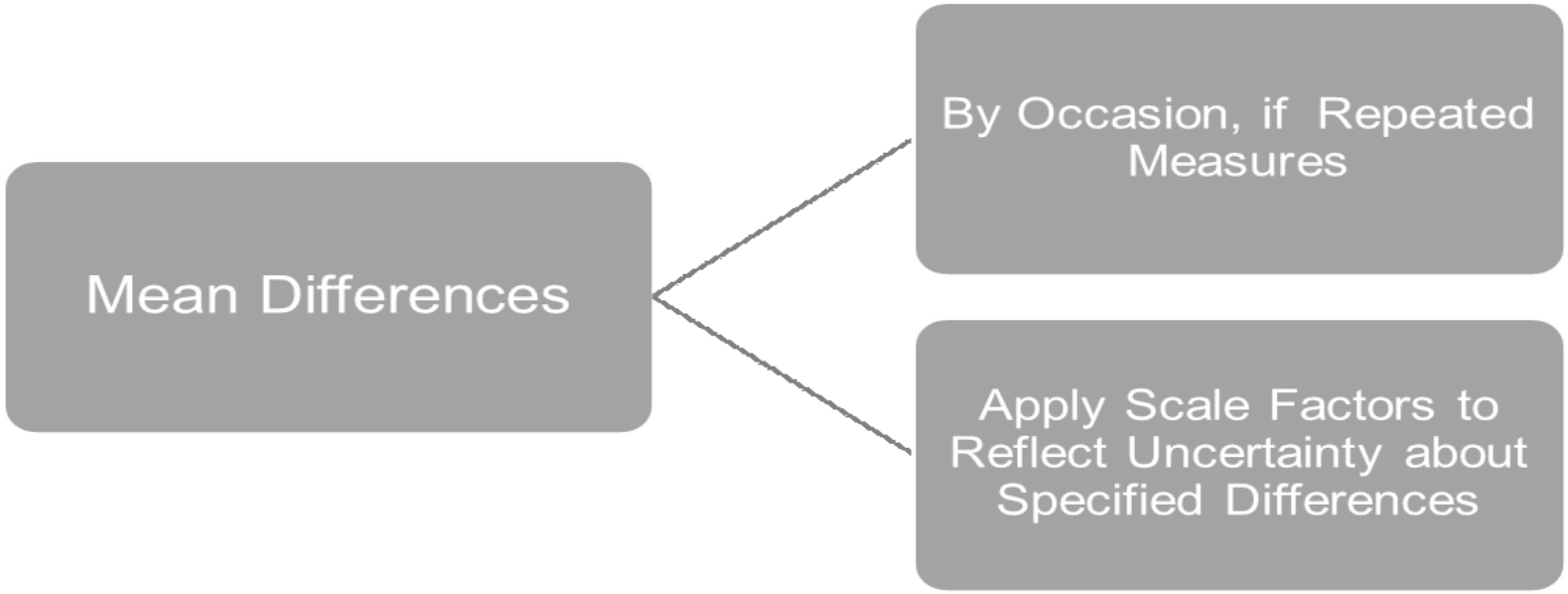
Repeated Measures for the Memory of Pain Trial



Step 4. Specify Primary Hypothesis of Interest



Step 5. Specify Mean Differences Between Groups



Mean Differences for the Memory of Pain Trial

Treatment group mean is 1.2 points lower on Memory of Pain compared to the Placebo group mean at the last time measurement (12 months).

Consider effect sizes of .5 times up to 2 times the stated effect to allow for uncertainty of the input information.

Step 6. Variance Structure: Multi-level Model Sources of Correlation

Clustering Covariance Pattern

For clustering, exchangable sampling induces

Compound Symmetry

$$\sigma^2 \begin{bmatrix} 1 & \rho_0 & \rho_0 \\ \rho_0 & 1 & \rho_0 \\ \rho_0 & \rho_0 & 1 \end{bmatrix} = .25 \begin{bmatrix} 1 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 1 \end{bmatrix}$$

Covariance Patterns Often Used for Repeated Measures

- Unstructured
- AR(1)
- Linear Exponent AR(1) (LEAR allows slower decay)
- "Random Intercept" $\Leftrightarrow$ CS covariance of responses; Gurka, Edwards, Muller (2011) showed danger

Covariance Patterns for Repeated Measures - Unstructured

Unstructured

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix} = .25 \begin{bmatrix} 1 & 0.3 & 0.2 \\ 0.3 & 1 & 0.5 \\ 0.2 & 0.5 & 1 \end{bmatrix}$$

Covariance Patterns for Repeated Measures - AR(1)

First order autoregressive

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_1^2 \\ \rho_1 & 1 & \rho_1 \\ \rho_1^2 & \rho_1 & 1 \end{bmatrix} = .25 \begin{bmatrix} 1 & 0.30 & 0.09 \\ 0.30 & 1 & 0.30 \\ 0.09 & 0.30 & 1 \end{bmatrix}$$

Covariance Patterns for Repeated Measures - LEAR

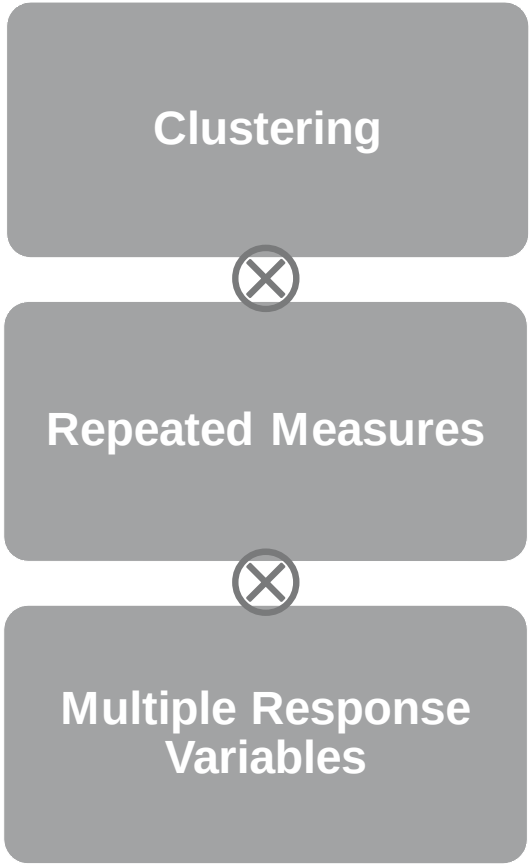
Linear Exponent AR(1) (δ = 0.5)

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_1^{1+\delta} \\ \rho_1 & 1 & \rho_1 \\ \rho_1^{1+\delta} & \rho_1 & 1 \end{bmatrix} = .25 \begin{bmatrix} 1 & 0.30 & 0.16 \\ 0.30 & 1 & 0.30 \\ 0.16 & 0.30 & 1 \end{bmatrix}$$

Commonly Used Covariance Patterns for Multiple Response Variables

- Unstructured observed
- Structure from Structural Equations Model
- Theoretical framework

Building Overall Covariance Structure



Building Overall Covariance Structure

Variance Clusters Time Responses

$$\boldsymbol{\Sigma}_i = \sigma^2 \begin{bmatrix} 1 & \rho_0 & \rho_0 \\ \rho_0 & 1 & \rho_0 \\ \rho_0 & \rho_0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_4 \\ \rho_4 & 1 \end{bmatrix}$$

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

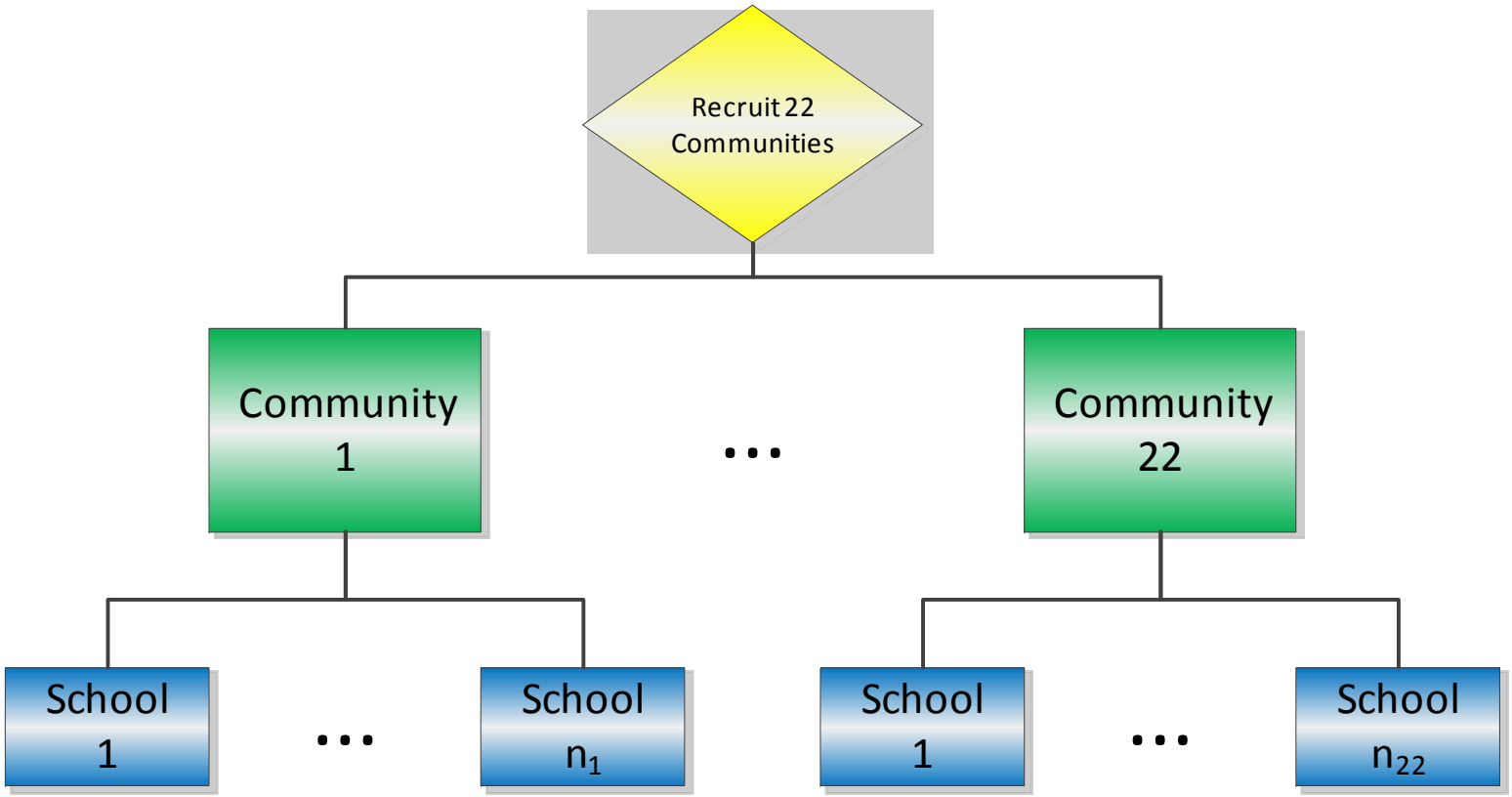
Overall Covariance Model for Memory of Pain Trial

- Variance of Memory of Pain $=$ 0.96
- Correlation of responses 6 months apart $=$ 0.5
- Correlation decays slowly over time, between 0 and 12 months correlation $=$ 0.4

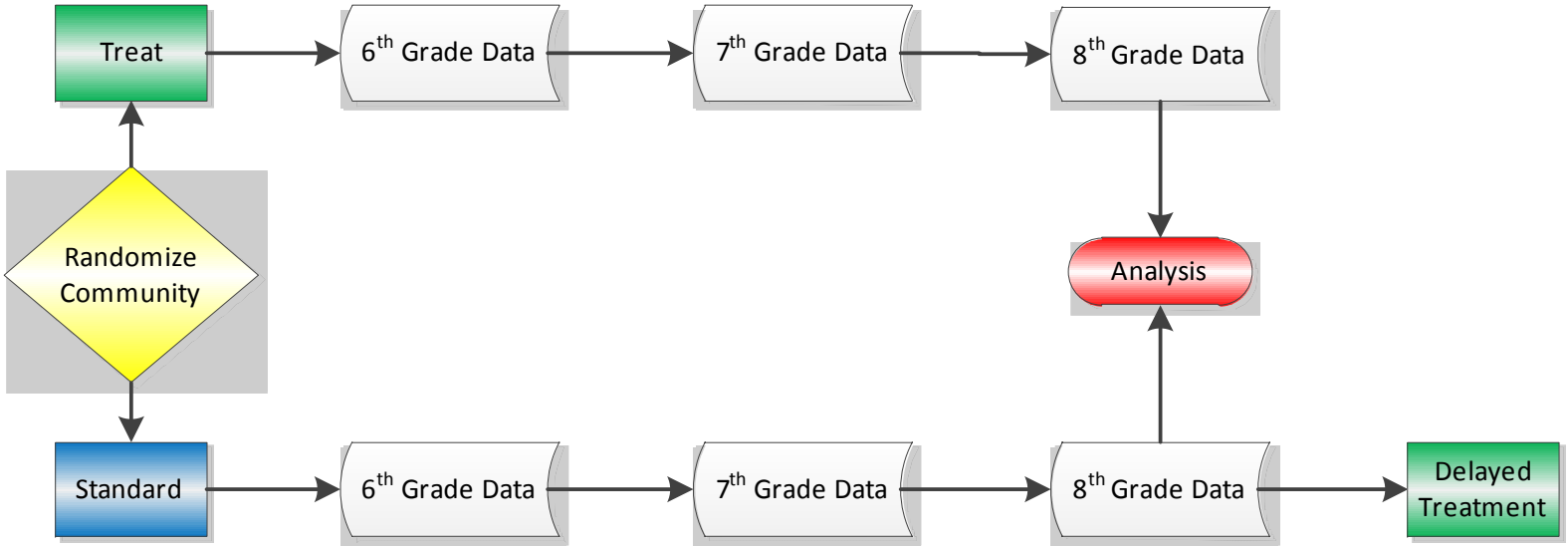
Overall Covariance Model for Memory of Pain Trial

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix} = .25 \begin{bmatrix} 1 & 0.3 & 0.4 \\ 0.5 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{bmatrix}$$

Example 2: PNC, Alcohol Use Prevention Study Power



PNC, Alcohol Use Prevention Study Example for Power



PNC Trial: Study Design Checklist

1. What is the study design goal?

a. Solving for power or sample size

Power

b. Type I error rate

0.05

PNC Trial: Study Design Checklist

- What is the sampling scheme?
- How many groups?

2 treatment groups

- What are the covariates?

None

- Is clustering present?

Yes; one level

- Are group sizes equal or unequal?

Yes, with 10 communities per group

PNC Trial: Study Design Checklist

- What responses are measured?
- What are the response variables?

Alcohol use behavior scale

- Are repeated measures present?

Yes, at 6th, 7th and 8th grades

- What is the primary hypothesis of interest?

Time Trend by Treatment Interaction

PNC Trial: Study Design Checklist

- What are the means?

Mean difference is 0.25 reduction in self reported alcohol use in treatment group vs. control

PNC Trial: Study Design Checklist

- What is the variance structure?
- What are the sources of correlation in the study design?

- Clustering (one level), with clusters of size 10 (# children/cluster)
  - Repeated Measures, 3 occasions, 1 year apart

PNC Trial: Study Design Checklist

- What is the variance structure?
- What is the pattern of variability for each source of correlation?

- Variance: $\sigma^2 = 0.09$
- Intraclass correlation for community: $\rho_0 = 0.01$
- Correlation for responses 1 year apart: $\rho_1 = 0.3$
- Correlation decays slowly with decay rate: $\delta = 0.3$

Overall Covariance Structure for PNC Trial

$$\sigma^2 \begin{bmatrix} 1 & \cdots & \rho_0 \\ \vdots & \ddots & \vdots \\ \rho_0 & \cdots & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_1 & \rho_1^{1+\delta} \\ \rho_1 & 1 & \rho_1 \\ \rho_1^{1+\delta} & \rho_1 & 1 \end{bmatrix} =$$

$$0.09 \begin{bmatrix} 1 & \cdots & 0.01 \\ \vdots & \ddots & \vdots \\ 0.01 & \cdots & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0.30 & 0.21 \\ 0.30 & 1 & 0.30 \\ 0.21 & 0.30 & 1 \end{bmatrix}$$

Outline

1. Characterizing the Linear Models You See
   - Common Tests in the Linear Mixed Model (LMM)
   - The LMM as a General Linear Multivariate Model
2. Six-Step Checklist for Power and Sample Size Analysis
   - Two Real Design Examples
   - Using the Checklist for the Examples
3. Simple Adjustments for Power with Missing Data
4. Free, Web-based Software, GLIMMPSE, and Related Web Resources

3. Simple Adjustments for Power with Missing Data

Missing Data Adjustments

- Some useful approximations from null case results in Catellier and Muller (2000) and better non-null case results in Ringham et al. (*in review*):

- Complete data power is an upper bound
- Power for N = (100% - % missing) x # ISUs appears somewhat liberal
- Power for N = (100% - % missing)^{p} x # ISUs appears somewhat conservative to OK
- Results assuming Missing at Random

- More work is in progress to identify better approximations

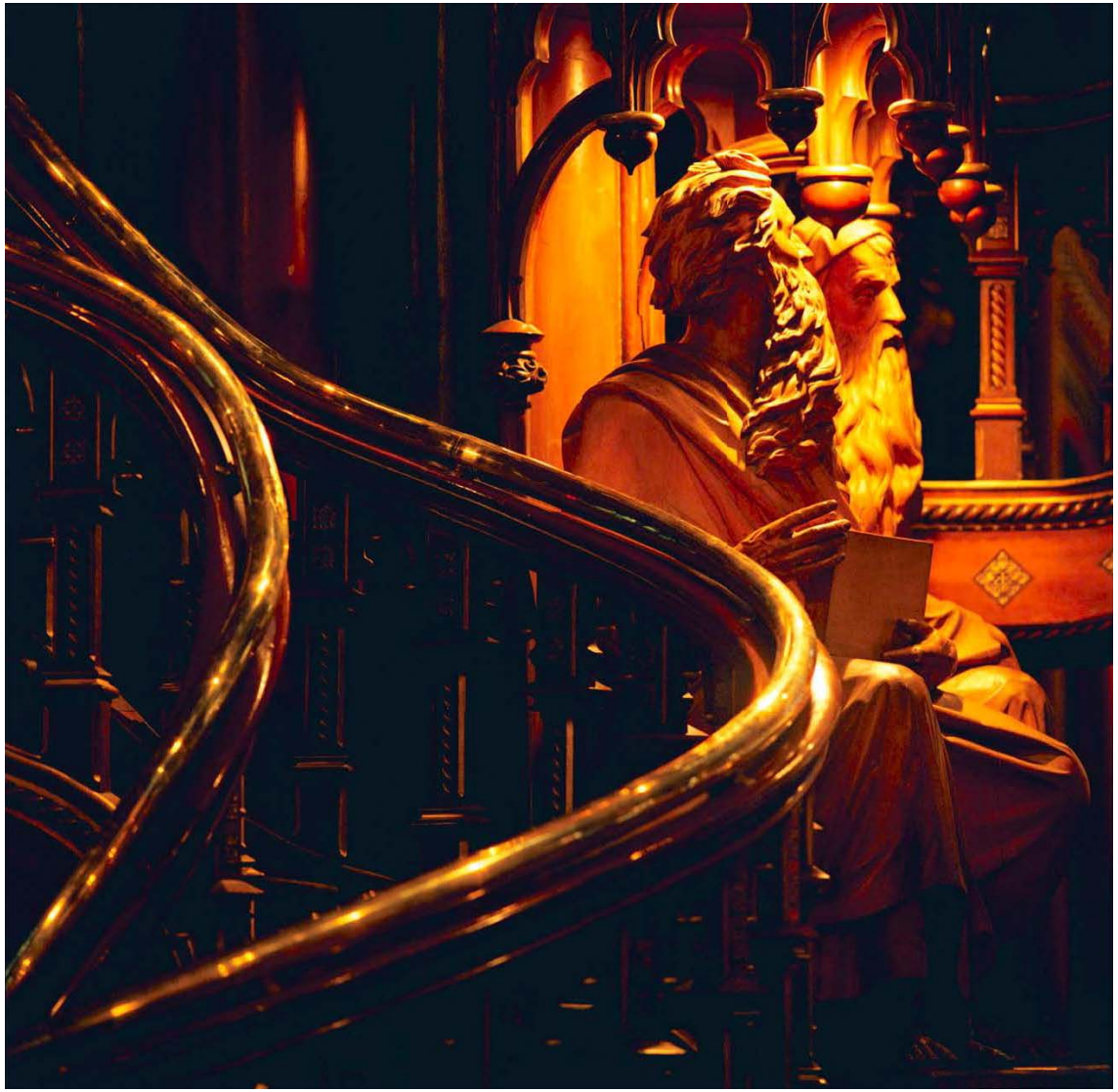
Review

- For widely applicable restrictions a General Linear Mixed Model can be expressed as a General Linear Multivariate Model with accurate power and sample size available.
- Answers to a series of simple questions can completely specify the inputs to a power analysis.
- Convenient adjustments may suffice for simple missing data patterns.

Outline

1. Characterizing the Linear Models You See
   - Common Tests in the Linear Mixed Model (LMM)
   - The LMM as a General Linear Multivariate Model
2. Six-Step Checklist for Power and Sample Size Analysis
   - Two Real Design Examples
   - Using the Checklist for the Examples
3. Simple Adjustments for Power with Missing Data
4. Free, Web-based Software, GLIMMPSE, and Related Web Resources

Any Brief Questions Before We Look at Our Software?



BIBLIOGRAPHY

Books

Kleinbaum DG, Kupper LL, Nizam A, and Muller KE (2007) *Applied Regression Analysis and Other Multivariable Methods.* 4th edition. Boston: Duxbury Press.

Muller KE and Stewart PW (2006) *Linear Model Theory; Univariate, Multivariate, and Mixed Models*. New York: Wiley.

Muller KE and Fetterman BA (2002) *Regression and ANOVA: An Integrated Approach Using SAS® Software*. Cary, NC: SAS Institute.

The following articles can be accessed from *www.samplesizeshop.org* **or** *www.health-outcomes-policy.ufl.edu/muller*

Catellier DJ and Muller KE (2000) Tests for gaussian repeated measures with missing data in small samples, *Statistics in Medicine*, **19**, 1101-1114.

Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, and Muller KE (2010) Real longitudinal data analysis for real people: building a good enough mixed model, *Statistics in Medicine*, **29**, 504-520.87

Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, and Schabenberger O (2008) An R-square statistic for fixed effects in the linear mixed model, *Statistics in Medicine*, **27**, 6137-6157.

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

Guo Y, Logan HL, Glueck DH, and Muller KE (2013) Selecting a sample size for studies with repeated measures, *BMC Medical Research Methodology*, in press.

Gurka MJ, Edwards LJ, Muller KE, and Kupper LL (2006) Extending the Box-Cox transformation to the linear mixed model, *Journal of the Royal Statistical Society Series A, Statistics in Society*, **169**, 255-272.

Gurka, MJ, Muller KE, and Edwards LJ (2011) Avoiding bias in mixed model inference for fixed effects, *Statistics in Medicine*, **30**, 2696-2707. doi:10.1002/sim.4293

Komro KA, Maldonado-Molina MM, Tobler AL, Bonds JR, and Muller KE (2007) Effects of home access and availability of alcohol on young adolescents' alcohol use, *Addiction* (Abingdon, England), **102**, 1597-1608.

Muller KE (2009) Analysis of Variance (ANOVA concepts and computations), *Wiley Interdisciplinary Reviews: Computational Statistics*.

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

Muller KE, Edwards LJ, Simpson SL, and Taylor DJ (2007) Statistical tests with accurate size and power for balanced linear mixed models, *Statistics in Medicine*, **26**, 3639–3660. DOI 10.1002/sim.2827

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.

Muller KE, and Peterson BL (1984) Practical methods for computing power in testing the multivariate general linear hypothesis, *Computational Statistics and Data Analysis*, **2**, 143-158.

Simpson SL, Edwards LJ, Muller KE, Sen PK, and Styner MA (2010) A linear exponent AR(1) family of correlation structures, *Statistics in Medicine*, **29**, 1825–1838.

Taylor DJ, and Muller KE (1995) Computing confidence bounds for power and sample size of the general linear univariate model, *The American Statistician*, **49**, 43-47. doi:10.2307/2684810

Other Sources

Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, and Campbell MJ (2004) Patterns of intra-cluster correlation from primary care research to inform study design and analysis, *Journal of Clinical Epidemiology*, **57**, 785-794.

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

Kerry SM, and Bland JM (1998) The intracluster correlation coefficient in cluster randomisation, *BMJ (Clinical research ed.)*, **316**(7142), 1455.

Laird NM, and Ware JH (1982) Random-effects models for longitudinal data, *Biometrics*, **38**, 963-974.

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

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