ACHIEVING GOOD POWER WITH CLUSTERED AND MULTILEVEL DATA

Keith E. Muller
Department of Health Outcomes and Policy
University of Florida, KMuller@ufl.edu
www.health-outcomes-policy.ufl.edu/muller

Supported in part by NIAAA R01-AA013458-01,
NHLBI R01-HL091005,
NIDCR U54-DE019261, and
NIDCR R01-DE020832-01A1

Disclosures: Muller has been a paid consultant to SAS; his wife worked for SAS for 5 years.

UF UNIVERSITY of FLORIDA

1. Recognizing the Challenge

2. Achieving Good Power

3. Overview of Using GLIMMPSE Software

Bibliography

UF UNIVERSITY of FLORIDA

1. Recognizing the Challenge

1.1 Motivation

Cluster and multilevel sampling can help:

More sensitivity (power)
Lower costs (recruit fewer participants)

Cluster and multilevel sampling can hinder:
Improper analysis usually underestimates variance and inflates type I error rate.

Need to know when and how to use to gain the advantages and avoid the problems.

1.2 Definitions

All of Statistics: estimation and
inference (testing)

Intervention study:

modeling relationships (estimation) and
testing hypotheses

Response $=$ f(predictor)

With random assignment to treatment,

Dependent $=$ f(independent)

Statistical model:

Response $=$ f(predictor) $+$ error

UF UNIVERSITY of FLORIDA

Statistical model:

response $=$ f(predictor) $+$ error

Responses: Y variables

Predictors: X variables

Errors: E

$Y = f(X) + E$

Error Variance:

"job security for statisticians"

UF UNIVERSITY of FLORIDA

1.3 Diagnostic Questions

Question 1: How Many Variables?

# responses	# predictors	Model
1	1	univariate
1	many	multi-variable
many	1	multivariate
many	many	multivariate

Same Y, many times: Multivariate,
and also Repeated Measures (REPM)

Measures between sampling units
(person, machine, hospital . . .) are independent.

Measures within (repeated measures:
“Time," limb, organ, . . .) are not independent
(correlated).

Many distinct Y's, many times:
doubly multivariate, collections of repeated measures

Independent observations

versus $\left\{ \begin{array}{c} \text{Multivariate (many } Y\text{'s)} \\ \text{Repeated Measures} \\ \text{Correlated observations} \\ \text{Non-independent observations} \end{array} \right\}$

Emphasize: statistics different!

Improper analysis can severely bias results

Theory same for multivariate and REPM while interpretation and analysis differ.

UF UNIVERSITY of FLORIDA

Question 2: Variable Types?

Scale	Data (Error Distribution)
Nominal:	Dichotomous Polychotomous Ordered Categorical
Ordinal	Order with infinitely many values.
Interval Ratio	Continuous { Gaussian Non-Gaussian

Inaccuracy in small sample inference for most methods.

UF UNIVERSITY of FLORIDA

Question 3: Sampling Pattern?

Design	#Times	# Indep. Sampling Units	Timing
Cross Sectional	one	N	--

Everything else: not cross sectional, multivariate in some sense

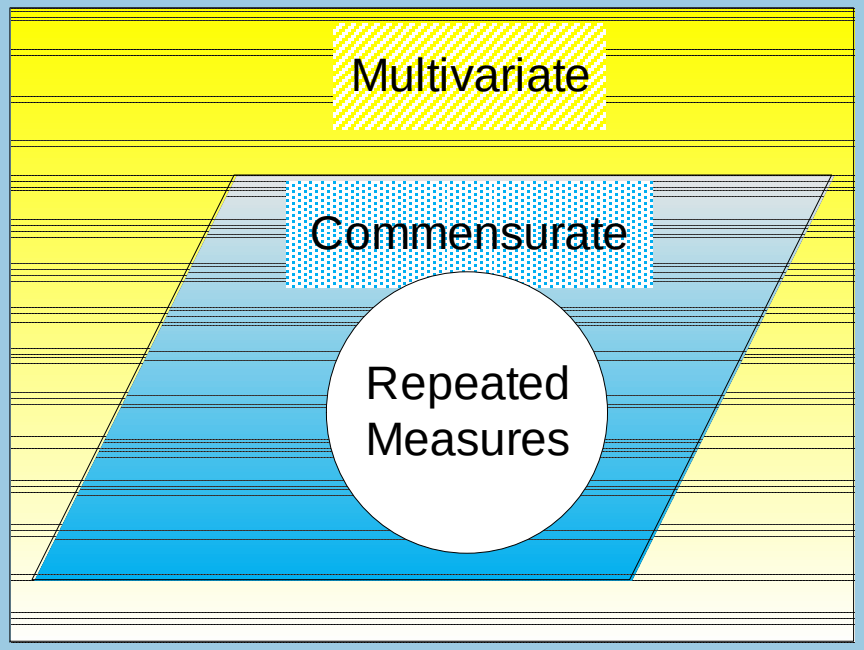

Figure 1. Categories of Multivariate Data

All variables measured in same units, *commensurate*.

What is the independent sampling unit,
in contrast to the observational unit?
observations = # ISU $*$ # Times

$N =$ # independent sampling units (ISU)

$p_i =$ # observations for independent sampling unit i

$n =$ total # observations $= \sum_{i=1}^{N} p_i$

In a cross sectional design $n = N$ because $p_i \equiv 1$

Table 1. Repeated Measures Sampling Patterns

Type	# Obs *per* ISU	#ISU	Timing
REPM	> 1	N	consistent
Crossover	p	N	alternating
Longitudinal	p_i, varies	N	inconsistent
Time series	n	1	regular
Cluster	p_i	# Clust	exchangeable
Survey	p_i	# Clust	exchangeable

Split Plot terminology originated in agriculture

Hierarchal, or multilevel data have "nested" sampling, with combinations of dimensions.

Examples of two or more dimensions of sampling:

teeth within a *person*

teeth within a *person* within a *clinic*

teeth within a *person* within a *clinic*
measured repeatedly over *time*

UF UNIVERSITY of FLORIDA

Question 4: # Independent Sampling Units per Response Variable?

1's?

10's?

100's?

1000's?

more?

High Dimension, Low Sample Size (HDLSS):

Examples of more observations than people:

genomics,
transcripomics,
medical imaging

UF UNIVERSITY of FLORIDA

Question 1: How Many Variables?

Question 2: Variable Types?

Question 3: Sampling Scheme?

Question 4: # ISUs per Response?

Answers to questions

allow recognizing the challenge.

Caution:

Usage of some definitions varies widely!

UF UNIVERSITY of FLORIDA

1.4 Analysis Methods

Choosing a Technique

Accurate estimation (means, proportions)?

Defensible inference (type I error rate)?

Property 1: How Many Variables?

Property 2: Variable Types?

Property 3: Sampling Scheme?

Property 4: # Persons per Response?

UF UNIVERSITY of FLORIDA

$5 \times 6 \times 8 \times 5 = 1200$ answers to 4 questions.
Each a potentially distinct analysis.
Answer requires a career, not a lecture!

Every analysis an approximation.
Seek accurate estimation and
defensible inference (tests).

Answers changing every few years
due to advances in computing.

“Being a statistician means
never having to say you're certain."
(ASA T-shirt)

UF UNIVERSITY of FLORIDA

Shop smart, learn limitations.
Report any limitations in publications.

Accurate small sample tests not available
for many interesting REPM models,
even assuming Gaussian errors.

Worst problem for non-Gaussian data.

Availability of software to fit a model,
even in major packages,
does not guarantee method defensible
for small N studies.

UF UNIVERSITY of FLORIDA

1. Recognizing The Challenge

✓ 1.1 Motivation: *Special Handling*
✓ 1.2 Definitions
✓ 1.3 Diagnosis: *Answer 4 Questions*
✓ 1.4 Analysis Methods

2. Achieving Good Power

2.1 Use the Highest Scale Possible
2.2 Align Power Analysis with Data Analysis
2.3 Conduct Sensitivity Analysis

3. Overview of Using GLIMMPSE

UF UNIVERSITY of FLORIDA

2. Achieving Good Power

2.1 Use the Highest Scale Possible

Nominal, ordinal, (interval, ratio).

RC MacCallum, S Zhang, KJ Preacher, and DD Rucker (2002) On the Practice of Dichotomization of Quantitative Variables

and subsequent related papers attempting to fight back.

Clinical thinking encourages
(I have seen opposite position)

2.2 Align Power Analysis with Data Analysis

Method must match method;
hypothesis must match hypothesis.

Muller, LaVange, Ramey and Ramey (1992) examples of t test power being wrong for repeated measures:
low example (child growth),
high example (kidney disease).

Test what must be tested, but may not have high power.
Seek high power for what you predict.

2.3 Conduct Sensitivity Analysis

Change dimensions, and ratios of dimensions.

How many people, but also
repeated measures,
clusters,
blocks.

Use smart spacing.

a) log2(dose), i.e. doublings: 2, 4, 8 mg/kg, weeks, etc.
b) Use only Min and Max for early science.
Omit middle doses.

Create power curves.

Try different variances for continuous data.

Try difference event rates for binary outcome.

UF UNIVERSITY of FLORIDA

Use confidence intervals for power values based on estimates (Taylor and Muller, 1995).

Available in GLIMMPSE (? now or soon).

1. Recognizing The Challenge

✓ 1.1
✓ 1.2
✓ 1.3
✓ 1.4

2. Achieving Good Power

✓ 2.1 Use the Highest Scale Possible
✓ 2.2 Align Power Analysis with Data Analysis
✓ 2.3 Conduct Sensitivity Analysis

3. **Overview of Using GLIMMPSE**

From APA presentation: slides 70-125

UF UNIVERSITY of FLORIDA

BIBLIOGRAPHY

Most articles listed can be downloaded or linked from *www.health-outcomes-policy.ufl.edu/muller*

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.

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.

Gurka, MJ, Muller KE, and Edwards LJ (2011) Avoiding bias in mixed model inference for fixed effects, *Statistics in Medicine*, in press.

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

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

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.

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.

UF | UNIVERSITY of FLORIDA