Tutorial 6: Tutorial on Translating between GLIMMPSE Power Analysis and Data Analysis

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

Acknowledgements:

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

Tutorial on Translating between GLIMMPSE Power Analysis and Data Analysis

Preface

Aligning power or sample size analysis with planned data analysis helps to avoid the problems of 1) sample size too small to detect important alternative hypotheses and 2) sample size so large that the design squanders precious resources (Muller et al., 1992).

In this tutorial, we describe some of the standard study designs for which the GLIMMPSE software provides power and sample size analysis. We examine the statistical models commonly used for analyzing data collected under those designs. We also provide an illustrative data analysis example for each design. Finally, we show how estimates from data analysis can be used as inputs for power and sample size analysis using GLIMMPSE.

Designs

Cross-sectional (t-test, ANOVA)

In cross-sectional designs we assume that individual observations are randomly sampled from one of two or more (k) well-characterized populations and that the observations within a sample are independent of each other, i.e. that there is no clustering between subjects or units of observation. We further assume that the underlying distribution of the observations is normal or approximately so. Finally, we usually assume that variances across populations are equal, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$. Data can be analyzed with a 2-sample t-tests if the variances are equal, a two sample t-test for unequal variances if the variances are not equal, the one-way ANOVA, or by a General Linear Model approach (see Tutorials on 2-sample t-test with Equal or Unequal Variances, and the Univariate Model – One-way ANOVA).

Repeated Measures (paired t-test, longitudinal)

In repeated measures or longitudinal designs we assume that individual observations are randomly sampled from one or more (k) well-characterized populations and that the observations within a sample are correlated across time or condition within subjects or units of observation. We further assume that the underlying distribution of the observations is normal or approximately so. We usually assume that variances across populations are equal, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$. The covariance model of the observations within a subject over time is assumed to be unstructured. This means that the variances over time can be unequal and the covariances between pairs of observations on a subject can also vary with distance in time. Data can be analyzed

with a General Linear Multivariate Model or a Linear Mixed Model (see Tutorials on Paired t-test and Repeated Measures).

Clustered (hierarchical structure and between group differences)

In clustered or multilevel designs we assume that individual observations are randomly sampled from one or more (k) well-characterized populations and that the observations within a sample are clustered within subjects or units of observation. We further assume that the underlying distribution of the observations is normal, or approximately so, and that the covariance structure of observations within a cluster is compound symmetry, i.e. any two observations in the same cluster have the same correlation, ρ. Finally, we usually assume that variances across populations are equal, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$. Data can be analyzed with a General Linear Multivariate Model or a Linear Mixed Model (see Tutorial on Clustered and Multilevel Designs).

Designs with a (baseline) covariate (ANCOVA, randomized studies, longitudinal)

Both cross-sectional and longitudinal designs can include covariate adjustment. In cross-sectional designs we assume that individual observations are randomly sampled from one of two or more (k) well-characterized populations and that the observations within a sample are independent of each other, i.e. that there is no clustering between subjects or units of observation. In repeated measures or longitudinal designs we assume that individual observations are randomly sampled from one or more (k) well-characterized populations and that the observations within a sample are correlated across time or condition within subjects or units of observation. The role of a covariate can be to reduce bias in a cross-sectional observational study or to increase precision in a (longitudinal) randomized study. We further assume that the underlying distribution of the observations is normal or approximately so. We usually assume that variances across populations are equal, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$. Finally, we assume that the covariate is normally distributed and that its covariance with the longitudinal observations can vary with time. Data can be analyzed with a General Linear Multivariate Model or a Linear Mixed Model (see Tutorials on ANCOVA and ANCOVA with Repeated Measures).

On our website we provide tutorials for a variety of power and sample size scenarios, including scenarios for the four study designs listed above. Below, we illustrate each of the designs with an example. For each example we provide a description of the data, the hypothesis to be tested, an analysis approach, test

statistics and reference distributions, appropriate SAS code and results, and from those results the inputs that would be used to inform the corresponding power analysis using GLIMMPSE. It should be appreciated that issues concerning statistical uncertainty arise when previously collected data are used to inform power and sample size analyses. Taylor and Muller (1995) present a discussion of the uncertainty in power and sample estimates that arise from sampling variability, and these ideas are used in the GLIMMPSE Tutorial on the Univariate Model for a One-way ANOVA. A discussion of bias resulting from using censored study results to inform power and sample size estimation is presented in Taylor and Muller (1996).

Developing a Data Analysis Plan - Examples

Cross-sectional (t-test, ANOVA)

There are many online examples of analyses for cross-sectional data with 2 or more group means to be compared. The two links below include scenarios, data, SAS code and output. The first link presents a two-sample t-test and the second a two-way ANOVA. The data regard exam scores in 200 high school students.

http://www.ats.ucla.edu/stat/sas/output/ttest.htm
http://www.ats.ucla.edu/stat/sas/output/sas_glm_output.htm

- The sampling unit in each scenario is a student in high school.
- When comparing female to male students on the exam writing score a 2-sample t-test or a General Linear Model can be used for analysis. When cross classifying the observations by sex and academic program, a General Linear Model is appropriate.
- If these data were being used as exemplary data to inform a power analysis that would replicate the independent sample t-test design above, the information that would be input into GLIMMPSE using Guided Mode would be the following:
    - Study groups: Multiple (2);Sex – Female, Male,
    - No Covariate
    - No Clustering
    - Relative Group Sizes: Sex – 1:1; Smallest Group Size = 100
    - Response variable: Writing score
    - No Repeated measures
    - Means: Female = 55, Male = 50
    - Scale Factors for Means: Yes

- Variability: SD = 10.3 (the largest value observed across the 2 groups, to be conservative)
- Sigma Scale Factors: Yes
- Statistical test: Any of the tests, as they will yield equivalent results
- Confidence Intervals: No
- Power Curve: No

▪ If these data were being used as exemplary data to inform a power analysis that would replicate a two-way ANOVA with Sex and Program as factors, as above, the information that would be input into GLIMMPSE using Guided Mode would be the following:

- Study groups: Multiple (6); Sex – Female, Male, and Program – 1, 2, 3
- No Covariate
- No Clustering
- Relative Group Sizes: Female – 1:2:1, Male – 1:2:1; Smallest Group Size = 21
- Response variable: Writing score
- No Repeated Measures
- Means: Female – Program 1 = 53.2, Program 2 = 57.6, Program 3 = 51; Males – Program 1 = 49.1, Program 2 = 54.6, Program 3 = 41.8
- Variability: SD = 10.4 (the largest value observed across the 6 groups, to be conservative)
- Statistical test: Any of the tests as they will yield equivalent results
- Confidence Intervals: No
- Power Curve: No

Repeated Measures (paired t-test, longitudinal data)

The data for this example come from a chelation trial for blood lead (mcg/dL) in children (Treatment of Lead-exposed Children (TLC Study Group, 2000; Fitzmaurice, Laird and Ware, 2011). One hundred children were randomly assigned to either chelation treatment vs. placebo in a 1:1 ratio. Blood lead was measured on 4 occasions: Baseline, Week 1, Week 4 and Week 6.

- The independent sampling unit is the child, with repeated blood lead levels clustered within a child.
- No assumptions are made about the correlation between blood lead levels measured over time within a child, i.e. the covariance matrix is Unstructured. For the four repeated blood lead measures in this example the assumed covariance matrix would look like this:

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_4 & \rho_5 \\ \rho_2 & \rho_4 & 1 & \rho_6 \\ \rho_3 & \rho_5 & \rho_6 & 1 \end{bmatrix}$$

- A Linear Mixed Model for repeated measures can be fit to the data as well as a General Linear Multivariate Model.
- The data can be found on the website for the Fitzmaurice, Laird and Ware text: http://www.hsph.harvard.edu/fitzmaur/ala2e/. Click on Datasets and look for Treatment of Lead Exposed Children Trial (N=100). Download the SAS file: tlc.sas7bdat.
- The SAS code below shows how to, first, transpose the data from a wide to a long format and then how to analyze the data using a Linear Mixed Model assuming an Unstructured covariance matrix for the four repeated measures within a child. A response profile approach is illustrated in which the mean blood lead in each group at each time is estimated. This is known as a maximal or saturated model and tests the treatment x time interaction.

  Click here to view the SAS code to analyze these data.
  Click here to view the output generated by this code.

- If these data were being used as exemplary data to inform a power analysis that would replicate the Repeated Measures design above, the information that would be input into GLIMMPSE using Matrix Mode would be the following:
  - Design Essence: 2 x 2, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, for the 2 groups to be compared
  - No Covariate
  - Smallest Group Size: 50
  - Coefficients:

- Beta Coefficients: 2 x 4 matrix, $\begin{bmatrix} 26.6 & 13.5 & 15.5 & 20.7 \\ 26.3 & 24.7 & 24 & 23.6 \end{bmatrix}$; these are the group-specific means of blood lead at each time
- Beta Scale Factors: 0.5, 1, 2

- Hypothesis
  - Between Participant Contrast: 1 x 2 matrix, $\begin{bmatrix} 1 & -1 \end{bmatrix}$, for the difference in means of blood lead between the succimer and placebo groups
  - Within Participant Contrast: 4 x 3 matrix, $\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$, for the 3 contrasts between week 1 and baseline, week 4 and baseline, and week 6 and baseline
  - Null Hypothesis Matrix: 1 x 3 matrix, $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$, for no difference in blood lead between groups at each follow-up time compared with baseline
- Variability
  - Error Covariance, (4 x 4 matrix), $\begin{bmatrix} 25 & 19 & 20 & 22 \\ 19 & 44 & 35 & 30 \\ 20 & 35 & 47 & 30 \\ 22 & 30 & 30 & 58 \end{bmatrix}$, the R matrix from the SAS output, containing the variances and covariances of blood lead across time from baseline to week 6
  - Sigma Scale Factors: 0.5, 1, 2
- Options
  - Statistical Test: Hotelling-Lawley Trace (Muller et al., 1992 provide guidance on when some of the other available tests can be applied. These guidelines are illustrated in the Tutorial on Selecting a Test.)
  - Confidence Intervals: No
  - Power Curve: No

- If these data were being used as exemplary data to inform a power analysis that would replicate the Paired t-test of the change in blood lead between Baseline and Week 1 in the group treated with succimer (only), as above, the information that would be input into GLIMMPSE using Matrix Mode would be the following:
  - Design Essence: 1 x 1, $[1]$, for the single group to be analyzed
  - No Covariate
  - Smallest Group Size: 50
  - Coefficients:
    - Beta Coefficients: 2 x 4 matrix, $[26.6 \quad 13.5]$; these are the succimer group means at baseline and week 1
    - Beta Scale Factors: 0.5, 1, 2
  - Hypothesis
    - Between Participant Contrast: 1 x 1 matrix, $[1]$, i.e. a single group is to be analyzed
    - Within Participant Contrast: 2 x 1 matrix, $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, for the contrast between blood lead means at week 1 and baseline
    - Null Hypothesis Matrix: 1 x 1, $[0]$, i.e. there is no difference in mean blood lead levels between week 1 and baseline
  - Variability
    - Error Covariance, (2 x 2) matrix, $\begin{bmatrix} 25 & 19 \\ 19 & 44 \end{bmatrix}$, the R matrix in SAS output containing the variances and covariance of blood lead at baseline and week 1
    - Sigma Scale Factors: 0.5, 1, 2
  - Options
    - Statistical Test: Hotelling-Lawley Trace (Muller et al., 1992 provide guidance on when some of the other available tests can be applied. These guidelines are illustrated in the Tutorial on Selecting a Test.)
    - Confidence Intervals: No
    - Power Curve: No

Clustered (hierarchical structure and between group differences)

The data for this example are from the state of Georgia's Vital Statistics unit, specifically birthweight for infants born to 200 mothers, with 5 children per mother (Vittinghoff et. al., 2012).

- In this example mothers (Level 2) are the independent sampling unit and babies (Level 1) are clustered within a mother.
- The assumed covariance structure is compound symmetry in which the correlation (ρ) between birthweight values for any two children born to the same mother is the same, and this correlation is the same for all mothers. Since there are five children per mother in the dataset, the covariance matrix would look like this:

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{bmatrix}$$

- Two different, but equivalent, mixed model formulations can be used to fit the data: a Variance Components model with a Random Intercept, or a Repeated Measures model with a Compound Symmetric structure for the covariance matrix of observations within a cluster.
- The data can be downloaded from the website for the Vittinghoff et al. text: http://www.epibiostat.ucsf.edu/biostat/vgsm/data.html. Look for Ch. 8, gababies, and the SAS link. This will download a SAS dataset called gababies.sas7bdat.
- Below is the code needed to fit the Random Intercept model using SAS PROC MIXED and to compare the birthweights for mothers whose first child was born when she was 20 years of age or younger vs. older than 20.

  Click here to view the SAS code to analyze these data.
  Click here to view the output generated by this code.

- If these data were being used as exemplary data to inform a power analysis that would replicate the design above for clustered data with 2 levels, the information that would be input into GLIMMPSE using Guided Mode would be the following:

- Study groups: Multiple (2); Age - $\leq$ 20 years old, > 20 years old
- No Covariate
- Clustering: Mothers, 5 children per mother, intracluster correlation of .39 (from the V correlation matrix in the SAS output)
- Size of the smallest group: 27 (number of mothers over 20 years of age)
- Response variable: Birthweight, no repeated measures
- Hypothesis: Main effect of Age
- Means: 3114, 3273g
- Scale Factors for Means: Yes
- Variability: 450.8g (= $\sqrt{203229 g^2}$, the within group variance)
- Sigma Scale Factors: Yes
- Statistical test: Hotelling-Lawley Trace (Muller et al., 1992 provide guidance on when some of the other available tests can be applied. These guidelines are illustrated in the Tutorial on Selecting a Test.)
- Confidence Intervals: No
- Power Curve: No

Designs with a baseline covariate (ANCOVA, randomized studies)

The data for this example come from a chelation trial for blood lead (mcg/dL) in children (Treatment of Lead-exposed Children (TLC Study Group, 2000; Fitzmaurice, Laird and Ware, 2011). One hundred children were randomly assigned to either chelation treatment vs. placebo in a 1:1 ratio. Blood lead was measured on 4 occasions: Baseline, Week1, Week 4 and Week 6. For this example, baseline blood level will be included as a covariate when looking at post-treatment changes in blood lead from baseline between the treatment groups over time.

- The independent sampling unit is the child, with repeated blood lead levels clustered within a child.
- No assumptions are made about the correlation between changes in blood lead levels measured over time within a child, i.e. the covariance matrix is Unstructured. For the three repeated blood lead change measures in this example the assumed covariance matrix would look like this:

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

- A Linear Mixed Model for repeated measures can be fit the data as well as a General Linear Multivariate Model.
- The data can be found on the website for the Fitzmaurice, Laird and Ware text: http://www.hsph.harvard.edu/fitzmaur/ala2e/. Click on Datasets and look for Treatment of Lead Exposed Children Trial (N=100). Download the SAS file: tlc.sas7bdat.
- The SAS code below shows how to, first, transpose the data from a wide to a long format, create the change scores, center the baseline blood lead values, and then shows how to analyze the data using a Linear Mixed Model assuming an Unstructured covariance matrix for the three repeated change scores in blood lead within a child. A response profile approach is illustrated in which the mean change in blood lead for each group at each time is estimated. This is known as a maximal or saturated model and can be used to tests the treatment x time interaction. A 3 df test of the main effect and interaction is needed to test the effect of succimer on blood lead over time.

  Click here to view the SAS code to analyze these data.
  Click here to view the output generated by this code.

- If these data were being used as exemplary data to inform a power analysis that would replicate the Repeated Measures ANCOVA design above, the information that would be input into GLIMMPSE using Matrix Mode would be the following:
  - Design Essence: 2 x 2 matrix, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, for the 2 groups to be compared
  - Covariate: Check Control for a single, normally distributed predictor
  - Smallest Group Size: 50

- Coefficients:
  - Beta Coefficients: 2 x 4 matrix, $\begin{bmatrix} -13 & -11 & -5.8 \\ -1.6 & -2.2 & -2.6 \end{bmatrix}$; these are the group-specific mean changes in blood lead at each time vs. the baseline
  - Beta Scale Factors: 0.5, 1, 2
- Hypothesis
  - Between Participant Contrast: 1 x 2 matrix, $\begin{bmatrix} 1 & -1 \end{bmatrix}$, for the difference in means of blood lead between the succimer and placebo groups
  - Within Participant Contrast: 3 x 3 matrix, $\begin{bmatrix} .33 & 1 & 1 \\ .33 & -1 & 0 \\ .33 & 0 & -1 \end{bmatrix}$, for the main effect of treatment at week 1 and the 2 contrasts between week 4 and week 1, and between week 6 and week 1
  - Null Hypothesis Matrix: 1 x 3, $\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$, i.e. no difference in the change in blood lead levels from baseline between groups at each follow-up time
- Variability
  - Error Covariance, (3 x 3 matrix), $\begin{bmatrix} 30 & 20.9 & 13 \\ 20.9 & 32 & 13.5 \\ 13 & 13.5 & 40 \end{bmatrix}$; this is the R matrix from SAS output, containing the variances and covariances of changes in blood lead levels from baseline for week 1, week 4, and week 6
  - Variance of Covariate: 25
  - Covariance of Outcomes and Covariate: $\begin{bmatrix} 18 \\ 19 \\ 22 \end{bmatrix}$
  - Sigma Scale Factors: 0.5, 1, 2
- Options
  - Statistical Test: Hotelling-Lawley Trace (Muller et al., 1992 provide guidance on when some of the other available tests can be applied. These guidelines are illustrated in the Tutorial on Selecting a Test.)

- Power Method: Unconditional (the Tutorial on ANCOVA provides more detail on the Quantile Power Method)
- Confidence Intervals: No
- Power Curve: No

References:

Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. *Applied Longitudinal Analysis, 2nd Edition*. New York, Wiley, 2011.

Muller, K.E. and Benignus, V.A. (1992). Increasing Scientific Power with Statistical Power, *Neurotoxicology and Teratology*, 14, 211–219.

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

Taylor D.J. and Muller K.E. (1995). Computing confidence bounds for power and sample size of the general linear univariate model, *American Statistician*, 49, 43-47.

Taylor, D.J. and Muller, K.E. (1996). Bias in Linear Model Power and Sample Size Calculation Due to Estimating Noncentrality, *Communications in Statistics –Theory and Methods*, 25, 1595–1610.

Treatment of Lead-exposed Children (TLC) Trial Group. (2000). Safety and Efficacy of Succimer in Toddlers with Blood Lead Levels of 20-44 µg/dL, *Pediatric Research*, 48, 593-599.

Vittinghoff, E., Glidden, D.V., Shiboski, S.C. and McCulloch, C. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, 1st Edition or 2nd Edition.* Springer-Verlag, 2005 or 2012.