# Tips, Tricks, and Traps on Longitudinal Data Analysis with Discrete and Continuous Times

**Brandy is a statistician and database programmer at the University of Michigan Medical School. She has worked on a wide variety of research projects. She holds a bachelors degree in engineering, and masters degrees in mathematics and statistics. The American Public Health Association awarded Brandy a student research award for her statistics masters thesis on path analysis. She has also received "Advanced Analytics" and "Data for Good" awards from the Midwest SAS Users Group conference.**

**Outside of work, she enjoys playing the piano, composing music, and singing in choir.**


# Tips, Tricks, and Traps on Longitudinal Data Analysis with Discrete and Continuous Times

Brandy R. Sinco, MS, University of Michigan, Ann Arbor, MI
Edith Kieffer, PhD, University of Michigan, Ann Arbor, MI
Michael Spencer, PhD, University of Michigan, Ann Arbor, MI
Gray Ficker, MDP, CHASS Center, Detroit, MI
Gretchen Piatt, PhD, University of Michigan, Ann Arbor, MI
Michele Heisler, MD, University of Michigan, Ann Arbor, MI

# Road Map (Outline)

1)_Introduction to Linear Mixed Model

2)_Empirical Correlation and Covariance Matrices, Descriptive Statistics

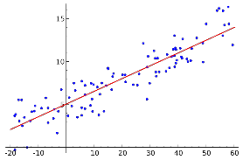
3)_Exploratory Graphics

4)_Covariance Structures in SAS® Proc Mixed

5)_Linear Mixed Models with Discrete Time Points

6)_Linear Mixed Models when Time is Continuous

7)_Graphical Presentation of Results



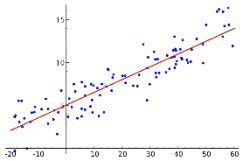
# Linear Regression Review

- Recall the linear regression model with a simple random sample of size N.

- Y = outcome vector, dimension N rows × 1 column; aka dependent variable.
- X = matrix of m predictor variables, dimension N × (m+1), aka m independent variables + intercept.

- Let $\beta_0$ to $\beta_m$ = Linear regression coefficients, where $\beta_0$ = intercept and $\beta_j$ = mean increase in Y for a unit increase in the jth X variable, $X_j$.
- β vector has dimensions (m+1) rows × 1 column.

- Let ε = Vector of error terms. $\varepsilon \sim N(0, \sigma^2)$; error to have constant variance.
- The linear regression equation will be $Y = X\beta + \varepsilon$; solution is found from least squares.



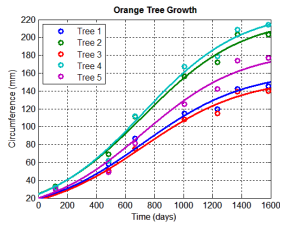
# Linear Regression Example

- For example, let Y = height in feet and $X_1$ = age in years ($X_1$ 10 to 17) and $X_2$ = gender (1=female, 0=male).

- For this model, assume that age is a fixed effect, meaning that age is being used to estimate a population average slope.

- The model would be written as $Y = \beta_0 + \beta_1 age + \beta_2 gender + \varepsilon$.

- If the solution to the regression were $\beta_0 = 4$, $\beta_1 = 0.1$, and $\beta_2 = -0.25$, this would increase that the population average rate of growth was 0.1 feet/year and that, on average, females were an average of ¼ foot shorter than males.

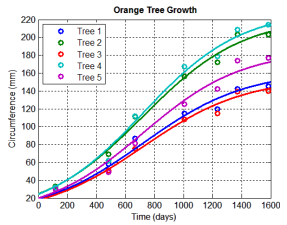



# Introduction to Linear Mixed Model (LMM) 1 of 3

- LMMs are linear statistical models for continuous outcomes, residuals are normally distributed, but not necessarily independent or not having constant variance.

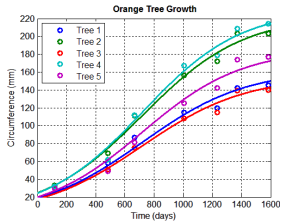
- LMM is an expansion of the linear regression model. X's and β's are the same as above, except that X's are designated as fixed effects.

- An additional matrix of random effects, Z, are added to the model, where $Z_1$ to $Z_p$ are random effects with regression coefficients $b_1$ to $b_p$.




# Introduction to Linear Mixed Model (LMM) 2 of 3

- The equation for the linear mixed model is: $Y = X\beta + Zb + \varepsilon$.
- Instead of $\varepsilon \sim N(0, \sigma^2)$ in linear regression, $\varepsilon \sim N(0,\Sigma)$ because the residuals can be correlated in a LMM.
- $\varepsilon$ is always a random effect; there is an estimate for each subject.
- $b \sim N(0, G)$, where G = covariance matrix of the random effects, other than $\varepsilon$.
- Finally, b and $\varepsilon$ are assumed to be independent.

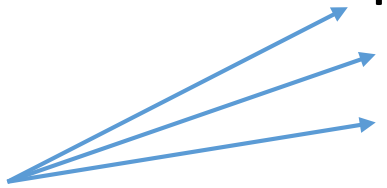


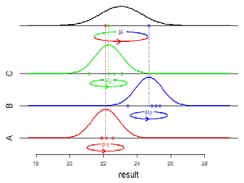

# Introduction to Linear Mixed Model (LMM) 3 of 3

- if age were modeled as both a fixed and a random effect, the linear regression model would become a linear mixed model:

$$Y = \beta_0 + \beta_1 age + \beta_2 gender + b_1 age + \varepsilon.$$

- This equation would produce a population average slope for age ($\beta_1$), in addition to an individual-level estimate for each person ($b_1$), analogous to the way that the random effect, $\varepsilon$, generates an error estimate for each person (or subject).
- The figure below illustrates random slope estimates for 3 subjects.



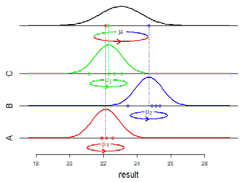



# Difference Between Fixed and Random Effects 1 of 2

- In 1960, Green and Tukey wrote “When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population the corresponding variable is random.”

- If we have an experiment with a random sample of washing machines of 3 different brands and the outcome is reduction in dirt, we are interested in the effect of the brand of the washing machine, not the individual washing machine.
- The brand of washing machine is a fixed effect and the effect of the individual washing machine is a random effect.

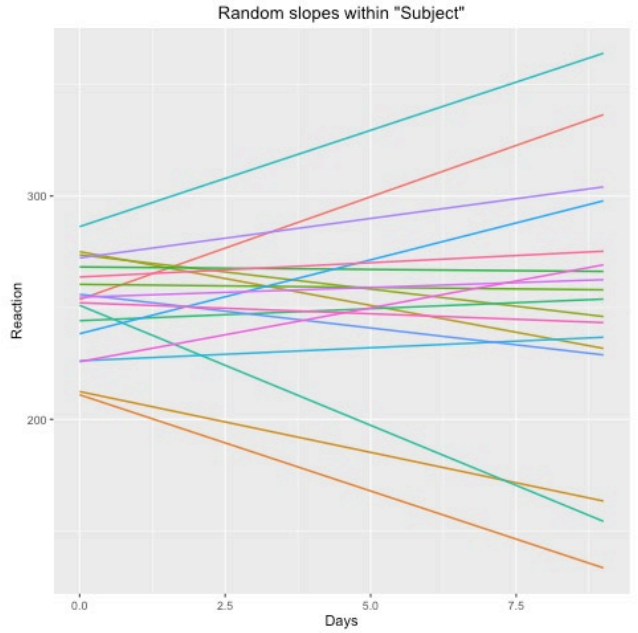
- Consider clinic. If the experiment were repeated, would you choose the same clinics or is the individual clinic site just a random sample of many clinics? If the clinic site would be chosen again, then clinic site is fixed. On the other hand, if we select a random sample of clinics from a huge list, then clinic site would be a random effect.




# Difference Between Fixed and Random Effects 2 of 2

- If we are only interested in population-average effects, then the effect is fixed, such as age in the linear regression model.

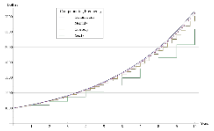
- On the other hand, if we are interested in individual-level estimates, such as estimating pregnancy weight at the end of the first trimester for individual women, then gestational age would be a random effect.



# Empirical Correlation Matrices + Descriptive Statistics

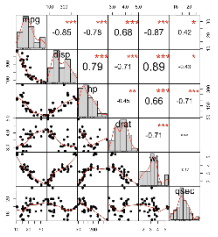
- First steps in longitudinal analysis are descriptive statistics and graphics.
- Means and variances at each time point need to be examined, along with correlation matrix. Straight-forward with Proc Corr.
- If the variances are not equal between time points, choose covariance structure that allows for unequal variances. A guideline for standard deviations is whether any standard deviation is twice another standard deviation at a different time point, because this would correspond to an F statistic of 4 ($F = s_1^2/s_2^2$).

```
/* Generate table of baseline (BL), 6 month (M6), 12 month
(M12) means, variances, correlation matrix */
Proc Corr Data=Across;
Var BL_HbA1c M6_HbA1c M12_HbA1c;  Run;
```



# What to Do if Time Points Are Not Discrete?

```
/* Add a timepoint counter to the dataset by ID */
Proc Sort data=LabData;
By ID HbA1cDate;  Run;

Data LabData;
Set LabData;  by ID;
TimePoint+1; /* Increment counter for each measurement */
If (First.ID) then TimePoint=1; /* Reset counter if new ID */

/* Then, use Proc Transpose to transpose the dataset. */
proc transpose data=LabData out=LabData_XP prefix=HbA1c;
var HbA1c;
id TimePoint;  by ID;  run;

/* View Raw Correlation Structure */
proc corr data=AcrossTime_XP;
var HbA1c1-HbA1c10;  run;
```




# Correlation Matrices

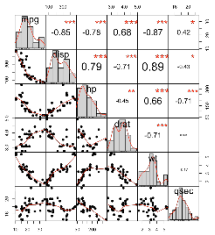
- Interested in whether the correlations appear to be constant over time or attenuating over time. Constant correlation over time indicates a compound symmetry correlation and attenuating correlation suggests an autoregressive correlation structure.

**Compound Symmetry with 3 Time Points**

| | | |
|---|---|---|
| 1 | ρ | ρ |
| ρ | 1 | ρ |
| ρ | ρ | 1 |

**Auto-Regressive Correlation with 3 Time Points**

| | | |
|---|---|---|
| 1 | ρ | $\rho^2$ |
| ρ | 1 | ρ |
| $\rho^2$ | ρ | 1 |

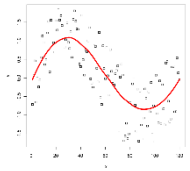


# Correlation Matrix Example from SAS Proc Corr

| Pearson Correlation Coefficients | | | |
|---|---|---|---|
| Prob > \|r\| under H0: Rho=0 | | | |
| Number of Observations | | | |
| | **BL_HbA1c** | **M6_HbA1c** | **M12_HbA1c** |
| **BL_HbA1c** | 1 | 0.257 | 0.184 |
| | | 0.0003 | 0.0001 |
| | 211 | 192 | 176 |
| **M6_HbA1c** | 0.257 | 1 | 0.283 |
| | 0.0003 | | <.0001 |
| | 192 | 192 | 166 |
| **M12_HbA1c** | 0.184 | 0.283 | 1 |
| | 0.0001 | <.0001 | |
| | 176 | 166 | 176 |


# Counting Measurements Per Subject

```
/* Discrete Time – N function counts non-missing entries. */
NumHbA1c=N(BL_HbA1c, M6_HbA1c, M12_HbA1c);

/* Continuous Time: Use Proc Freq. */
/* NumHbA1c = number of HbA1c measurements per person */
proc freq data=LabData;
Tables ID/out=HbA1cCount;
run;

Data HbA1cCount;
Set HbA1cCount(Keep=ID Count);
Rename Count=NumHbA1c;
Run;
```

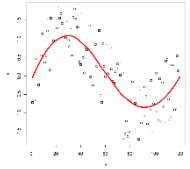


# Exploratory Graphics – Part 1 (LOESS)

- Whether time is discrete or continuous, plotting the LOESS smoothing curve is a helpful starting point. LOESS stands for Locally Weighted Scatterplot Smoothing.

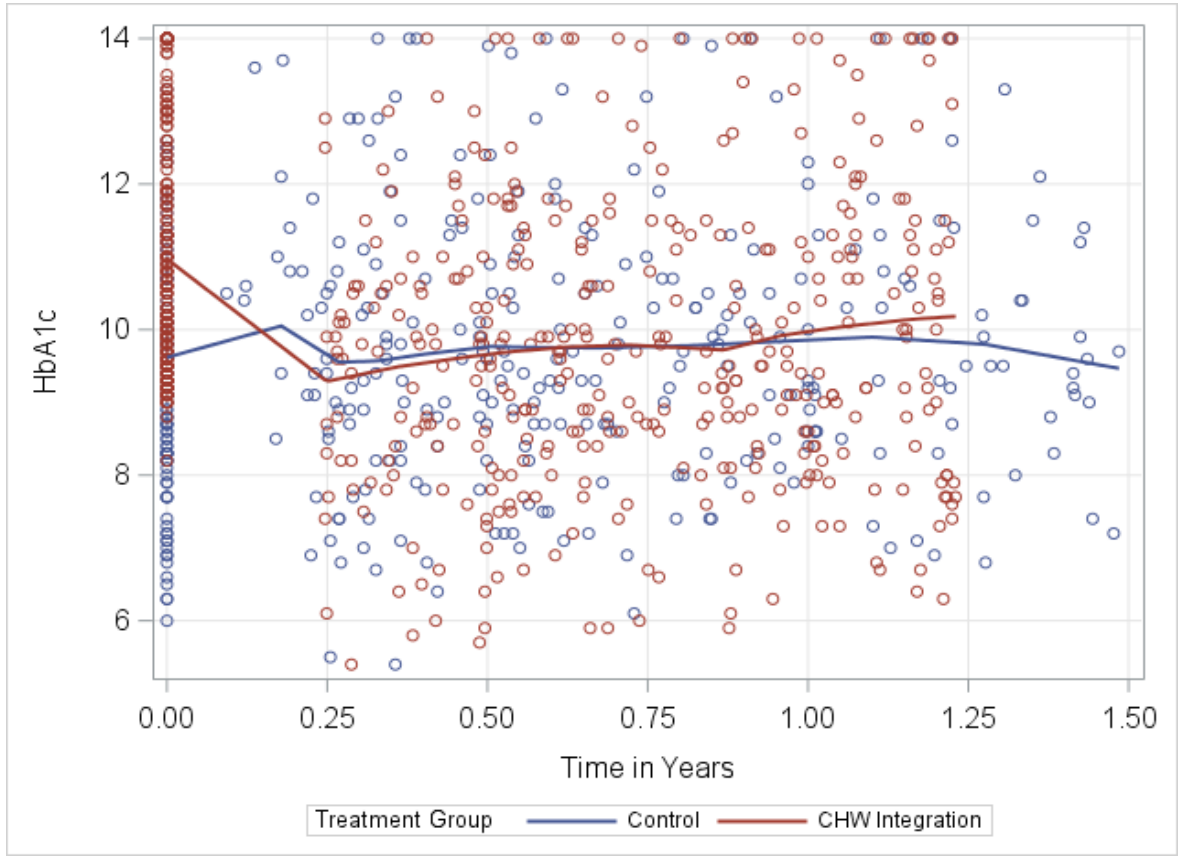
- The LOESS technique estimates local regression polynomials over subsets of the data, given a smoothing parameter, α. The smoothing parameter, α, is also called the bandwidth and indicates the percentage of the data that is used to fit each of the local polynomials.
- LOESS was developed by William Cleveland and Susan Devlin.

```
/* First, select LOESS smoothing parameter */
proc loess data=Across_long_chwpluscontrol;
*where treat=0; /* smooth=.482 */
 where treat=1; /* smooth=.5 */
 ods select fitplot;  run;
```



# Exploratory Graphics – Part 1 (LOESS)

```
Proc SGPlot Data=Across_long_chwpluscontrol;
loess x=TimeInYears Y=HbA1c/group=treat degree=2 smooth=.5;
xaxis grid LABELATTRS=(Size=12) VALUEATTRS=(Size=12) label="Time in Years";
yaxis grid LABELATTRS=(Size=12) VALUEATTRS=(Size=12); Run;
```



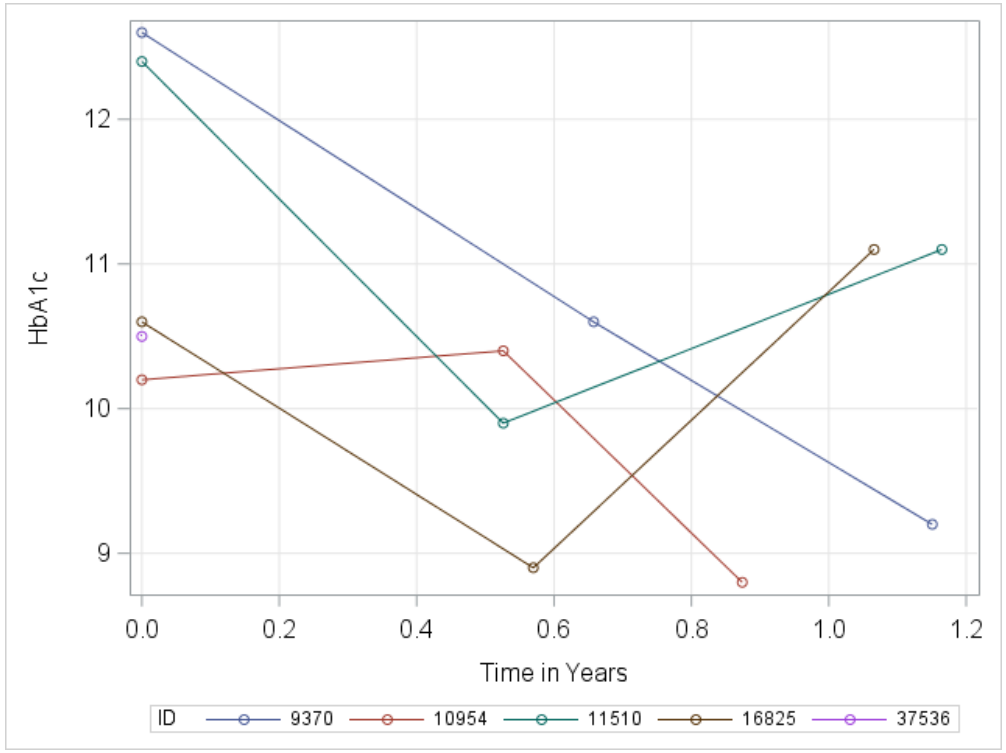
# Exploratory Graphics – Part 2 (Spaghetti Plot)

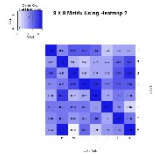
- **Spaghetti Plot =** plot of a random sample of the outcomes across time for a small number of IDs.

```
/* Random selection of 5 people from baseline data */
proc surveyselect data=Baseline method=SRS N=5 Out=FiveRand
seed=8152018;
run;
/* Tip: Use seed.  Otherwise, SAS will choose different 5
for procedure re-runs */

Data FiveRandLong;
Merge FiveRand(Keep=ID IN=X) LongA1c;
by ID;
if X=0 then delete; Run;
```

# Exploratory Graphics – Part 2 (Spaghetti Plot)

```
proc sgplot data=FiveRandLong;
  series x=TimeInYears y=HbA1c / group=ID markers;
xaxis grid LABELATTRS=(Size=12) VALUEATTRS=(Size=12) label="Time in
Years";
yaxis grid LABELATTRS=(Size=12) VALUEATTRS=(Size=12) Label="HbA1c";
```





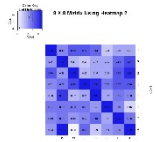
# Covariance Structures in SAS® Proc Mixed

- While SAS offers a wide range of covariance structures, the most useful ones in my work are unstructured, compound symmetry, and autoregressive.

- Covariance structures are compared by using the AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria).

- The covariance structure that fits the data best will have the smallest AIC and BIC.

| Fit Statistics | |
|---|---|
| **-2 Log Likelihood** | 419.5 |
| **AIC (Smaller is Better)** | 447.5 |
| **AICC (Smaller is Better)** | 452.0 |
| **BIC (Smaller is Better)** | 465.6 |

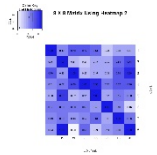


# Tips on Unstructured Covariance

| | Y(time = 1) | Y(time = 2) | Y(time = 3) |
|---|---|---|---|
| **Y(time = 1)** | $\sigma_1^2$ | $\sigma_{12}$ | $\sigma_{13}$ |
| **Y(time = 2)** | $\sigma_{12}$ | $\sigma_2^2$ | $\sigma_{23}$ |
| **Y(time = 3)** | $\sigma_{13}$ | $\sigma_{23}$ | $\sigma_3^2$ |

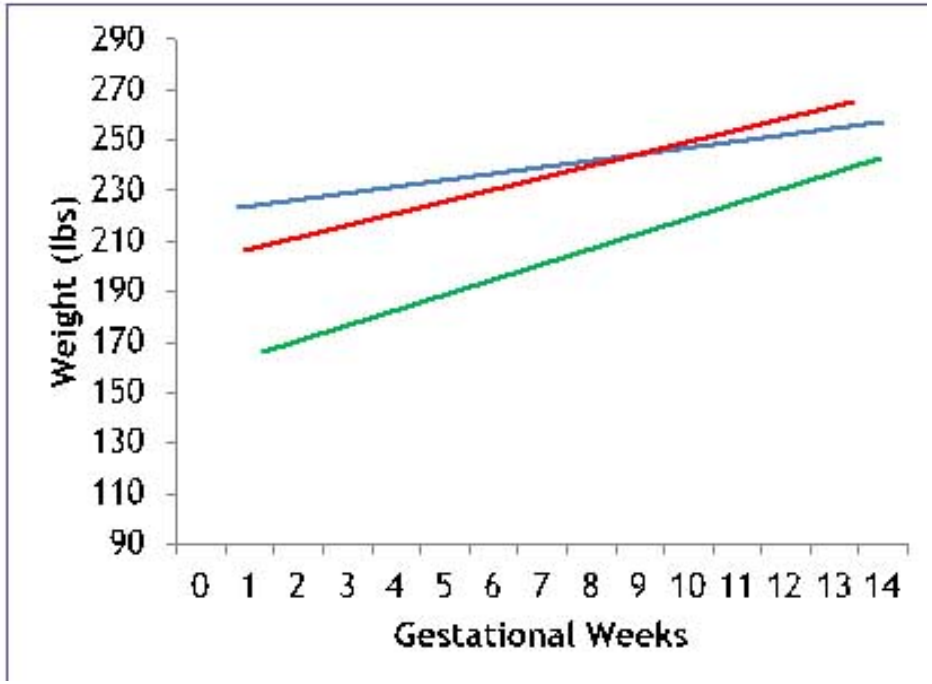
- Convergence problems when time is continuous.
- For discrete time, AIC and BIC sometimes criteria indicate that unstructured covariance is the best choice.

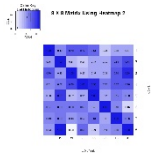
• For models with a random slope and intercept, use type=UN on the random statement. This will estimate variances for the intercept and slope, along with covariances between the intercept and slope. If “type=UN” is omitted on the random statement, the SAS default is type=VC, variance components, which assumes that the intercept and slope are independent, and estimates covariance to be zero.



# Tips on Unstructured Covariance

```
/* Example: Gestational weights by week of pregnancy */
/* OutPred = pdat outputs person-level estimates */
PROC MIXED DATA = GestWt method = REML NOCLPRINT;
Class ID;
model wt = gestwk parity YearUS height triceps anthrowk /
outpred = pdat solution ddfm=kr;
random int gestwk / subject = id type = un g gcorr;  run;
```





# Tips on Compound Symmetry Covariance Structure

| | Y(time = 1) | Y(time = 2) | Y(time = 3) |
|---|---|---|---|
| **Y(time = 1)** | $\sigma^2$ | $\rho\sigma^2$ | $\rho\sigma^2$ |
| **Y(time = 2)** | $\rho\sigma^2$ | $\sigma^2$ | $\rho\sigma^2$ |
| **Y(time = 3)** | $\rho\sigma^2$ | $\rho\sigma^2$ | $\sigma^2$ |

- Compound Symmetry = Equal correlation between any two time points.
- Type = CS or CSH on repeated statement. H = Unequal variances; heterogenous.
- Option for both continuous and discrete time models.
- Choose compound symmetry if AIC and BIC criteria are minimized.

# Tip: Compound Symmetry ⟷ Random Intercept Model

| | | |
|---|---|---|
| $\sigma^2$ | $\rho\sigma^2$ | $\rho\sigma^2$ |
| $\rho\sigma^2$ | $\sigma^2$ | $\rho\sigma^2$ |
| $\rho\sigma^2$ | $\rho\sigma^2$ | $\sigma^2$ |

| | | |
|---|---|---|
| $\sigma_e^2 + \sigma_b^2$ | $\sigma_b^2$ | $\sigma_b^2$ |
| $\sigma_b^2$ | $\sigma_e^2 + \sigma_b^2$ | $\sigma_b^2$ |
| $\sigma_b^2$ | $\sigma_b^2$ | $\sigma_e^2 + \sigma_b^2$ |

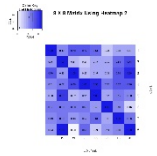
- Compound Symmetry equivalent to random intercept model, $\rho = \sigma_b^2/(\sigma_e^2 + \sigma_b^2)$.

- AIC and BIC will be identical between compound symmetry & random intercept model.

- Random intercept model defined by the equation, $Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_0 + \varepsilon_{ij}$.
- $Var(Y_{ij}) = Var(b_0 + \varepsilon_{ij})$; Recall from model assumptions that $Cov(b_0, \varepsilon_{ij}) = 0$.

- Diagonal elements will be $Var(b_0) + Var(\varepsilon_{ij}) = \sigma_e^2 + \sigma_b^2$.

- When $i \neq j$ or $k \neq l$, $Cov(Y_{ij}, Y_{kl}) = Cov(b_0 + \varepsilon_{ij}, b_0 + \varepsilon_{kl}) = Var(b_0) = \sigma_b^2$.



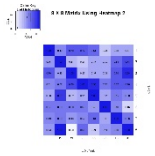
# Tips on Autoregressive and Spatial Covariance

| | Y(time = 1) | Y(time = 2) | Y(time = 3) |
|---|---|---|---|
| **Y(time = 1)** | $\sigma^2$ | $\rho\sigma^2$ | $\rho^2\sigma^2$ |
| **Y(time = 2)** | $\rho\sigma^2$ | $\sigma^2$ | $\rho\sigma^2$ |
| **Y(time = 3)** | $\rho^2\sigma^2$ | $\rho\sigma^2$ | $\sigma^2$ |

- Rho is raised to the difference between time points.

- Discrete time ρ ^ (difference in time index).
- Example for times 1 and 3 $\rho^{(3-1)} = \rho^2$; for times 2 and 4 $\rho^{(4-2)} = \rho^2$.
- Type = AR(1).

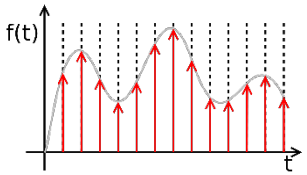
- Continuous time ρ ^ (difference in time units); Spatial covariance.
- Example for TimeInYears. If time 1 = .7 year and time 3 = 1.9 years, then correlation function = $\rho^{(1.9-.7)} = \rho^{1.2}$.
- Type = SP(POW)(TimeInYears).



# Toeplitz Covariance Structure

| | Y(time = 1) | Y(time = 2) | Y(time = 3) |
|---|---|---|---|
| **Y(time = 1)** | $\sigma^2$ | $\sigma_1$ | $\sigma_2$ |
| **Y(time = 2)** | $\sigma_1$ | $\sigma^2$ | $\sigma_1$ |
| **Y(time = 3)** | $\sigma_2$ | $\sigma_1$ | $\sigma^2$ |

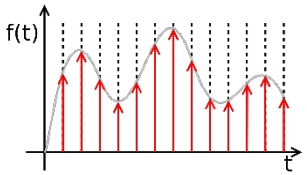
- Toeplitz covariance is a special case of auto-regressive covariance.
- For homogenous Toeplitz covariance, the variance on the diagonal is constant, $\sigma^2$, while the covariances for the differences between two times, j and k, are equal.
- $Cov(Y_{ij}, Y_{ik}) = \sigma_{|j-k|}$. For example, $Cov(Y_{i1}, Y_{i4}) = Cov(Y_{i2}, Y_{i5}) = \sigma_3$.
- SAS syntax Type=TOEP homogenous, Type=TOEPH heterogenous.




# Linear Mixed Models with Discrete Time Points

- $Y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 t_1 + \beta_3 t_2 + \beta_4 G_i t_1 + \beta_5 G_i t_2 + \varepsilon_{ij}$.
- $G_i$ = treatment group (1 = intervention; 0 = control) for the ith participant.
- $t_1$ = first follow-up time, often 6 months (1 = 2nd follow-up time, 0 = other time).
- $t_2$ = second follow-up time, often 12 months (2 = 2nd follow-up time, 0 = other time).
- $\varepsilon_{ij}$ = Random error. $\varepsilon_{ij} \sim N(0, \sum)$.

```
/* SAS Data Step – First Merge from the Various Time Points */
Data Across;
Merge Demographics BaseLab.BL_Lab  /* Baseline */
M6Lab.M6_Lab /* Month 6 */
M12Lab.M12_Lab; /* Month 12 */
by ID;
```




# Linear Mixed Models with Discrete Time Points

```
/* Compute Deltas (Change scores in HbA1c */
M6BL_HbA1c = M6_HbA1c - BL_HbA1c; M12BL_HbA1c = M12_HbA1c -BL_HbA1c;
```

Convert the “across” dataset to a long dataset for SAS Proc Mixed.
TimePointN is used, because SAS sets the largest value to the reference by default.
If time points 1, 2, 3, etc. are -1, -2, -3, etc., SAS will automatically set 0 to the default.

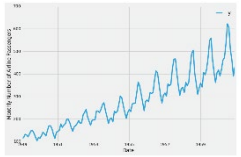
```
Data Time_Long;  Set AcrossTime;
TimePoint=0; TimePointN=0; /* Baseline */
HbA1c=BL_HbA1c;   Delta_HbA1c=0;   Output;

TimePoint=1; TimePointN=-1; /* 6 Months */
HbA1c=M6_HbA1c;  Delta_HbA1c=M6BL_HbA1c;  Output;

TimePoint=2; TimePointN=-2; /* 12 Months */
HbA1c=M12_HbA1c;  Delta_HbA1c=M12BL_HbA1c;  Output;
Run;
```

# Proc Mixed + Proc PLM (Discrete Time Points)

```
Proc Mixed Data=HbA1cLong_Discrete Method=REML NOCLPRINT
plots(only)=(StudentPanel(conditional box));
Class ID TreatN TimePointN;
Model HbA1c= TreatN TimePointN TimePointN*TreatN/ Solution
Influence(effect=ID Est) ddfm=KR;
Repeated / type=AR(1) Subject=ID R RCorr;
Store MixARDiscrete;  Run;

/* Adjust Multiple Comparisons with Proc PLM.  */
PROC PLM Restore=MixARDiscrete;
Estimate 'Control BL' Int 1 TreatN 0 1 TimePointN 0 0 1
TimePointN*TreatN 0 0 0 0 0 1 /adjust=simulate(NSAMP=10000
SEED=8132018);
Estimate 'CHWInt BL' Int 1 TreatN 1 0  TimePointN 0 0 1
TimePointN*TreatN 0 0 1 0 0 0 /adjust=simulate (NSAMP=10000…;
```

# Proc PLM (Continued, Time Discrete)

```
Estimate 'Control M6' Int 1 TreatN 0 1 TimePointN 0 1 0
TimePointN*TreatN 0 0 0 0 1 0 / adjust=simulate(NSAMP=10000 …);

Estimate 'CHWInt M6' Int 1 TreatN 1 0  TimePointN 0 1 0
TimePointN*TreatN 0 1 0 0 0 0 / adjust=simulate(NSAMP=10000 …);

Estimate 'Control M6-BL'  TimePointN 0 1 -1  TimePointN*TreatN 0 0 0
0 1 -1 / adjust=simulate(NSAMP=10000 SEED=8132018);

Estimate 'CHWInt M6-BL'  TimePointN 0 1 -1  TimePointN*TreatN 0 1 -1
0 0 0 / adjust=simulate(NSAMP=10000 SEED=8132018);

Estimate 'Int Eff M6'  TimePointN*TreatN 0 1 -1 0 -1 1 / adjust …;
ODS Output Estimates=PLMEstARDiscrete;  Run;

Proc Print Data=PLMEstARDiscrete; Run;
```

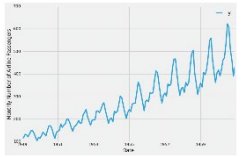


# Linear Mixed Models with Continuous Time Points

- $Y_{ij} = \beta_0 + \beta_1 G_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 G_i t_{ij} + \beta_5 G_i t_{ij}^2 + \varepsilon_{ij}$.

- $G_i$ = treatment group (1 = intervention; 0 = control) for the ith participant.
- t = time for ith participant at jth repetition.
- $\varepsilon_{ij}$ = Random error. $\varepsilon_{ij} \sim N(0, \sum)$.

```
/* Compute Time Differences in SAS data step */
BL_TimeInDays = 0;
M6_TimeInDays = datdif(BL_HbA1cDate, T1_HbA1cDate, 'act/act');
M12_TimeInDays = datdif(BL_HbA1cDate, T2_HbA1cDate, 'act/act');

M6_TimeInMonths = T1_TimeInDays/30;
M12_TimeInMonths = T2_TimeInDays/30;

M6_TimeInYears = T1_TimeInDays/365;
M12_TimeInYears T2_TimeInDays/365;
```



# Linear Mixed Models with Continuous Time Points

- **Convert the "across" dataset to a "long" dataset.**

```
/* Baseline */
TimeInYears=0; TimeInDays=0; TimeInMonths=0;
HbA1c=BL_HbA1c; Delta_HbA1c=0;  Output;

/* T1*/
TimeInYears=T1_TimeInYears;
TimeInDays=T1_TimeInDays;
TimeInMonths=T1_TimeInMonths;
HbA1c=T1_HbA1c;
Delta_HbA1c=T1_HbA1c;
Output;

/* T2*/
Same formulas as above with T2 substituted for T1.
```

# Proc Mixed + Proc PLM (Continuous Time)

```
Proc Mixed Data=LongA1c.Hba1c_long_chwpluscontrol Method=REML
NOCLPRINT plots(only)=(StudentPanel(conditional box));
Class ID TreatN;
Model HbA1c= TreatN TimeInYears TimeInYears2 TreatN*TimeInYears
TreatN*TimeInYears2 AgeGE55
  / Solution Influence(effect=ID Est) ddfm=KR;
Repeated / type=SP(Pow)(TimeInYears) Subject=ID R RCorr;
Where HbA1c NE .; Store MixContinuous;  Run;

/* Adjust Multiple Comparisons with Proc PLM.  */
PROC PLM Restore=MixContinuous;
Estimate 'Control BL' Int 1 TreatN 0 1 TimeInYears 0 TimeInYears2 0
AgeGE55 .426/cl adjust=simulate(NSAMP=10000 SEED=8132018);

Estimate 'CHWInt BL' Int 1 TreatN 1 0 TimeInYears 0 TimeInYears2 0
AgeGE55 .426/cl adjust=simulate(NSAMP=10000 SEED=8132018);
```

# Proc PLM (Continued, Time Continuous)

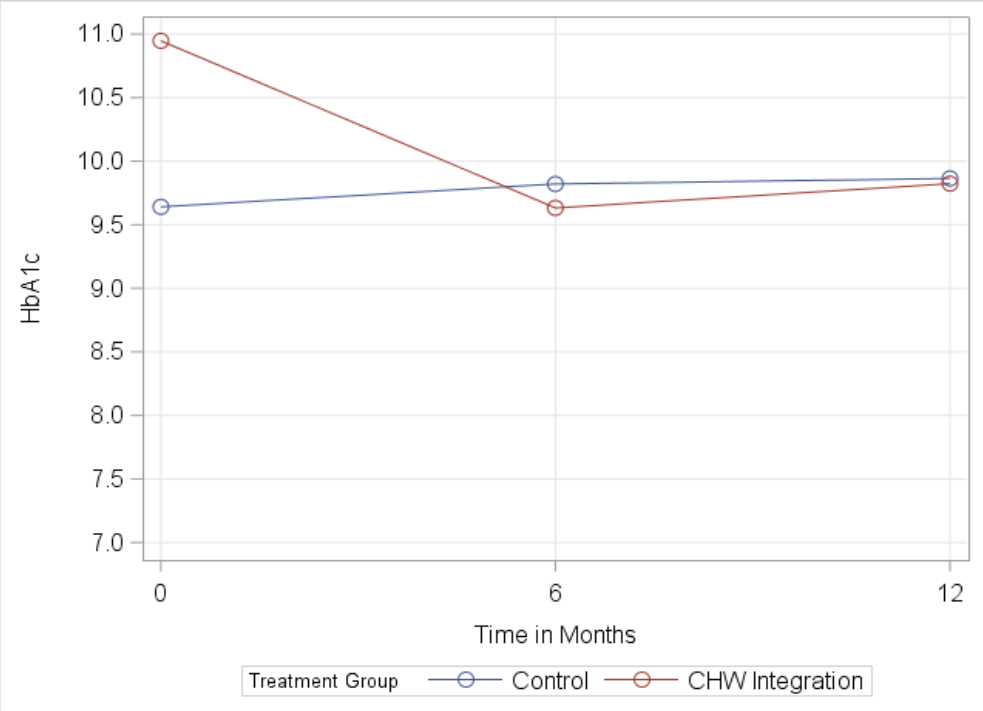
```
Estimate 'Control M6' Int 1 TreatN 0 1 TimeInYears .5 TimeInYears2
.25 TreatN*TimeInYears 0 .5 TreatN*TimeInYears2 0 0.25 AgeGE55
.426/cl adjust=simulate(NSAMP=10000 SEED=8132018);

Estimate 'CHWInt M6' Int 1 TreatN 1 0 TimeInYears .5  TimeInYears2
.25 TreatN*TimeInYears .5 0 TreatN*TimeInYears2 0.25 0 …;

Estimate 'Control M6-BL' TimeInYears .5  TimeInYears2 .25
TreatN*TimeInYears 0 .5 TreatN*TimeInYears2 0 /…;
Estimate 'CHWInt M6-BL'  TimeInYears .5  TimeInYears2 .25
TreatN*TimeInYears .5 0 TreatN*TimeInYears2 .25 0/cl …;

Estimate 'Int Eff M6' TreatN*TimeInYears .5 -.5 TreatN*TimeInYears2
.25 -.25/cl adjust=simulate(NSAMP=10000 SEED=8132018);
ODS Output Estimates=PLMEstTimeCont;  Run;
Proc Print Data=PLMEstTimeCont; Run;
```


# Continuous Time Point Units

- **Tip – Inestimable Error on Estimate Statements.**
- Model with HbA1c data over 5 years, originally coded time in months and the model had a 3rd degree polynomial with SP(POW)(TimeInMonths) covariance structure.
- Model converged, but “Inestimable” errors from the estimate statements.
- SAS tech support advised me that the large (Time in Months) values combined with small coefficient values can cause inestimable errors.
- Model to use TimeInYears. The model with TimeInYears produced the same AIC and BIC values, but the estimate statements no longer produced the inestimable errors.

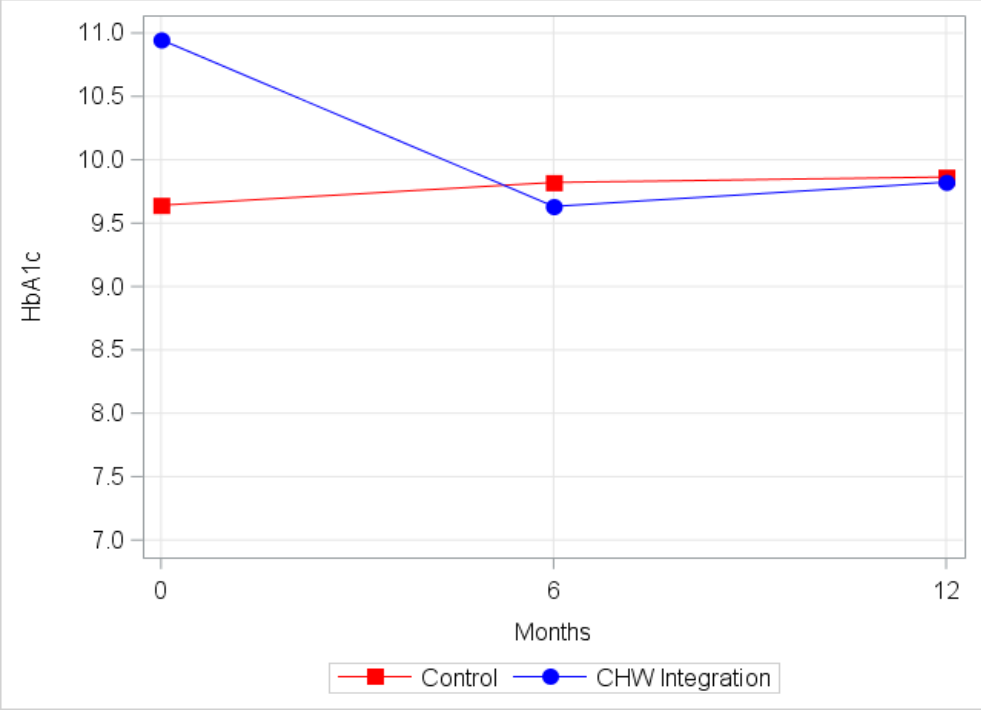
# Tips and Tricks: Means Across Time, Proc Mixed Output

- Tip.  If the data is in the format: Y_Group1, Y_Group2, Time, the graph options are easier to control than if the data is in the format: Y, Group, Time.
- Example 1:  Data format: Y, Group, Time.  Group= option triggers SAS defaults, which are modifiable by editing the graph template – a bit complicated.
- `proc sgplot data=HbA1cAdjTime;`
- `Series x=TimeInMonths y=HbA1c/markers  markerattrs=(size=12) Group=Treat;`

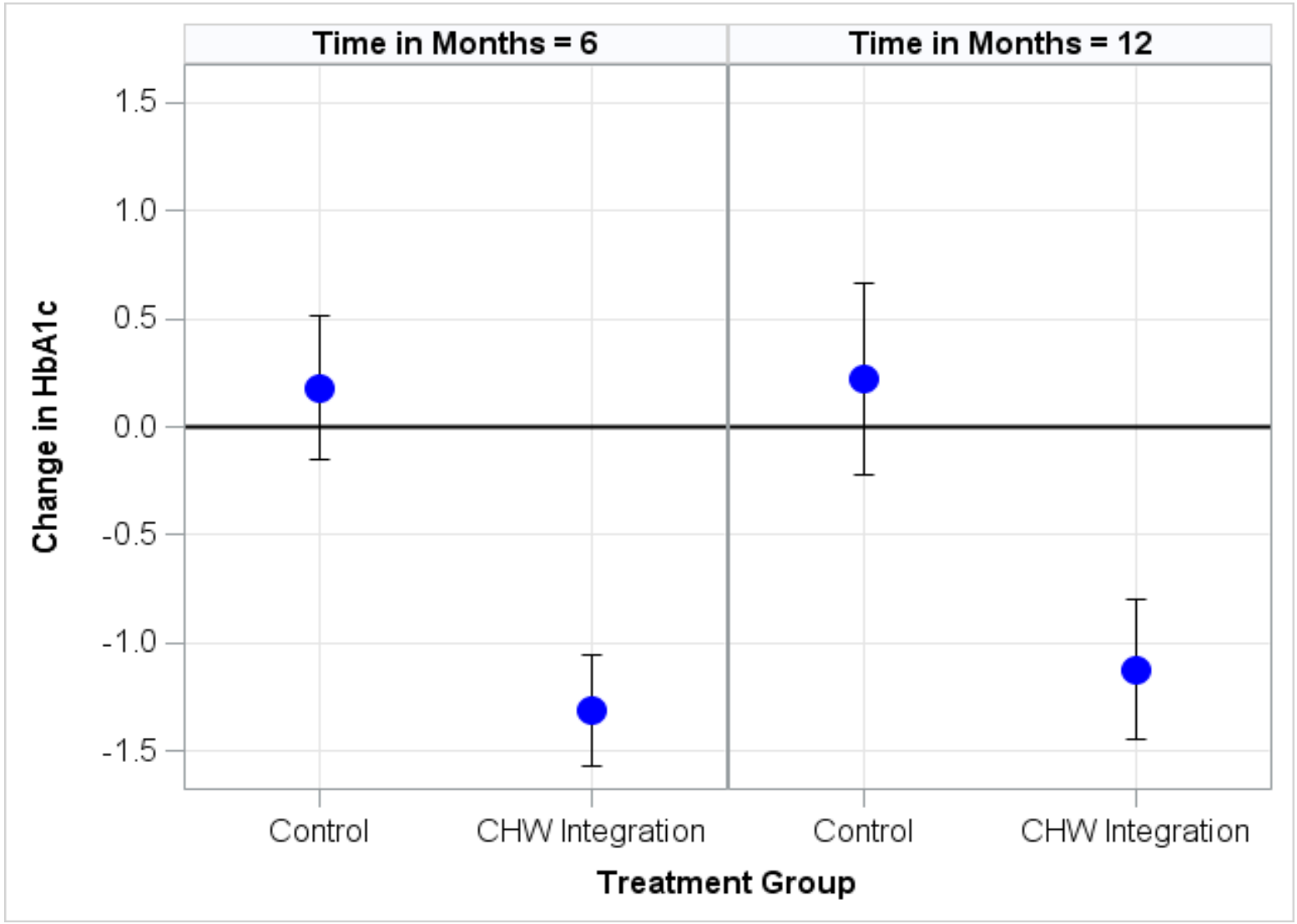


# Tips and Tricks: Means Across Time, Proc Mixed Output

- Example 2:  Easier to customize graph with options on the Series statement.
- `proc sgplot data=HbA1cAdjTimeBoth;`
- `Series x=TimeInMonths y=HbA1cCtrl/markers lineattrs=(color=red) markerattrs=(size=12 symbol=SquareFilled color=red);`
- `Series x=TimeInMonths y=HbA1cTrt/markers lineattrs=(color=blue) markerattrs=(size=12 symbol=CircleFilled color=blue);`



# Tips and Tricks: Changes by Group Time, Proc Mixed Output



# Tips and Tricks: Changes by Group Time, Proc Mixed Output

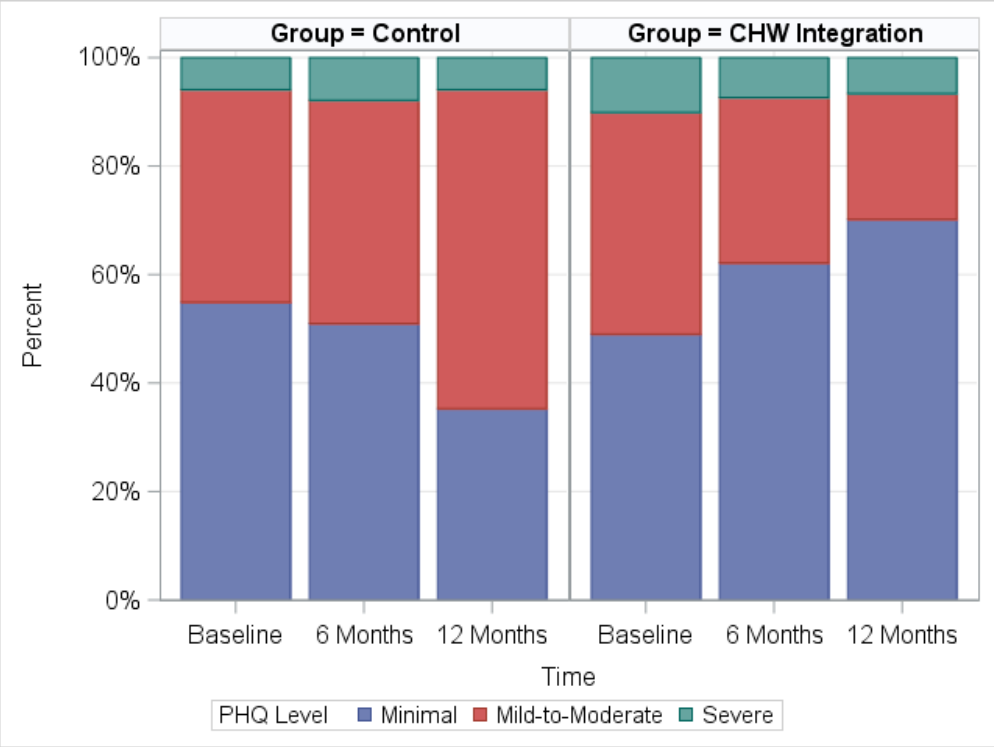
- This SAS code will produce a nice graph of ΔY at 6 and 12 month follow-up.

```
proc sgpanel data=DeltaHbA1cAdjTime;
panelby TimeInMonths / rows=1 columns=2 HEADERATTRS=(Color=Black
Size=12 Weight=Bold);
scatter x=Treat y=DeltaHbA1c /  ERRORBARATTRS=(color=black)
yerrorlower=Lower yerrorupper=Upper
markerattrs=(symbol=circlefilled color=blue size=20);
colaxis grid LABELATTRS=(Size=12 weight=bold)
VALUEATTRS=(Size=12) values=(0 1) type=discrete offsetmin=.25
offsetmax=.25;
rowaxis grid LABELATTRS=(Size=12 weight=bold)
VALUEATTRS=(Size=12) values=(-1.5 to 1.5 by .5);
/* Reference line at Y=0 */
refline 0/ axis=y lineattrs=(Color=Black Thickness=2); run;
```

# Tips and Tricks: Mosaic Plots

- Proc Freq – plots = mosaic option on Tables statement.  Settings w/Proc Template.

```
Proc SGPanel Data=Phq_long_chwpluscontrol pctlevel=group;
panelby Treat/rows=1 columns=2 headerattrs=(Size=12 weight=bold);
Vbar TimePoint/Group=PHQ3Cat stat=percent;  run;
```



# Conclusions

- Begin with LOESS and spaghetti plots, whether time points are discrete or continuous.
- Empirical correlation matrices and descriptive statistics with Procs Corr and Means.

- SAS syntax and correlation structures differ between linear mixed models with discrete and continuous times.
- “Inestimable” errors on estimate statement can sometimes be fixed by changing time unit.

- Use Proc PLM to adjust for multiple comparisons with Monte Carlo simulation.
- Use the seed option for any procedures that draw a random sample.

- When creating graphics, best to choose graphics where options are easily changeable without modifying the graph template.

# Acknowledgements


- University of Michigan Schools of Social Work and Public Health.

- Community Health and Social Services (CHASS) of Detroit, MI.

- Funding: National Institute for Health Care Reform (nihcr.org).


# Contact Information


- Brandy R. Sinco, Statistician and Programmer/Analyst
- Michigan Medicine
- NCRC, CHOP (Center for Healthcare Outcomes and Policy)
- Ann  Arbor, MI 48109-2800
- E-Mail: brsinco@med.umich.edu
- Phone: 734-232-5422