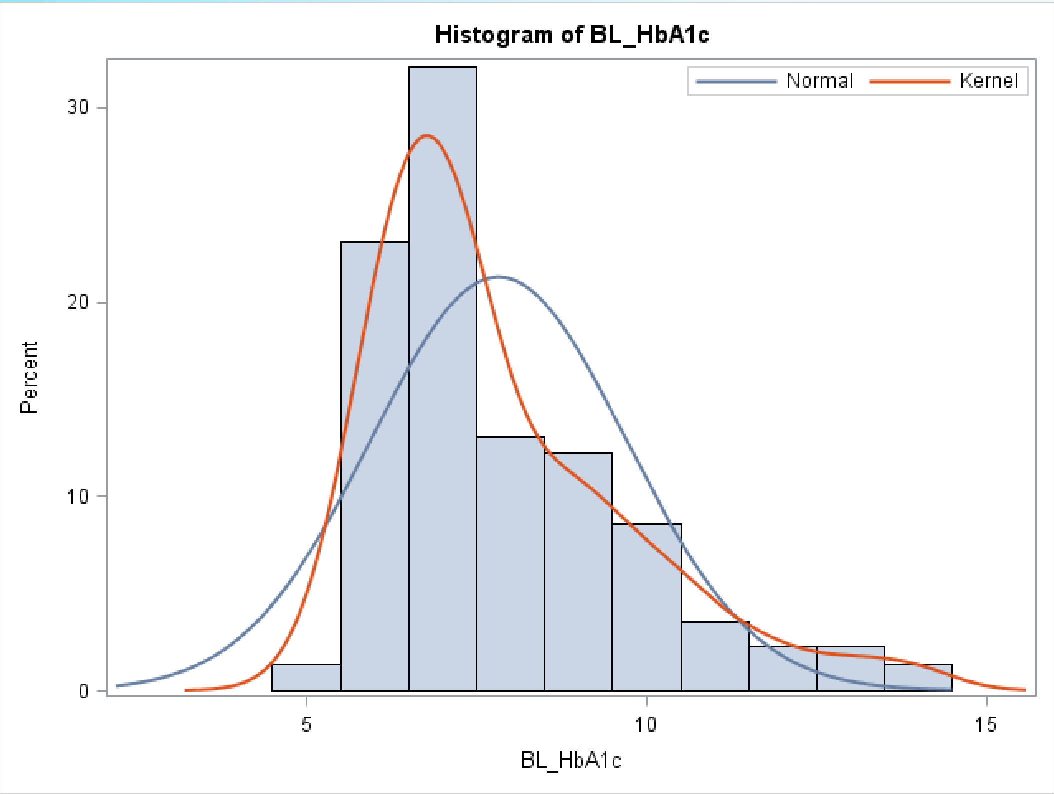
Using SAS for the Longitudinal Analysis of Difference Scores

Brandy R. Sinco, MS, University of Michigan, Ann Arbor, MI
Edith Kieffer, PhD, University of Michigan, Ann Arbor, MI
Michael S. Spencer, PhD, University of Michigan, Ann Arbor, MI
Gloria Palmisano, BS, MA, CHASS Center, Detroit, MI
Gretchen Piatt, PhD, University of Michigan, Ann Arbor, MI
Michele Heisler, M.D., University of Michigan, Ann Arbor, MI

Outline (Road Map)

- Why Differencing Helps to Reduce Skewness.
- Baseline Adjustment in the Pre-Post Model with Proc GLM + Macro.
- Introduction to the Linear Mixed Model for Longitudinal Data with Proc Mixed,
- Two Time Points and Two Treatment Groups.
- Longitudinal Difference Data at 4 Time Points with Three Treatment Groups + Macros.
- Adjustment for Multiple Comparisons.
- Graphics and Tables from Publications.
- Conclusions.

Example of Skewed Outcome Variable




```
proc sgplot data=MyData;
    histogram BL_HbA1c;
    density BL_HbA1c;
    density BL_HbA1c / type=kernel;
     keylegend / location=inside
     position=topright;
run;
```

Example variable is Hemoglobin A1c (HbA1c), an important measurement of blood sugar for people with diabetes.

Why Difference Scores Often Reduce Skewness (1 of 2)

- Skewness, γ, of random variable, Y, with mean μ, $\gamma = E(Y - \mu)^3$.
- Skewness is a measure of asymmetry. Positive values indicate a longer tail to the right. Negative values indicate a longer tail to the left.
- Let Y_0= Outcome at baseline and Y_1 = Outcome at follow-up.
- Let Y_0 and Y_1 be random variables with means μ_0 and μ_1, variances σ_0^2 and σ_1^2, skewnesses γ_0 and γ_1, and Pearson correlation coefficient ρ.
- Then, the skewness of $(Y_1 – Y_0) = E((Y_1 – \mu_1) – (Y_0 – \mu_0))^3 =$
- $E(Y_1 – \mu_1)^3 - 3E((Y_1 – \mu_1)^2(Y_0 – \mu_0)) + 3E((Y_1 – \mu_1)(Y_0 – \mu_0)^2) - E(Y_0 – \mu_0)^3$.
- If Y_0 and Y_1 are independent, the skewness of $(Y_1 – Y_0) = \gamma_1 - \gamma_0$, because the two expectation terms in the center will be zero.

Why Difference Scores Often Reduce Skewness (1 of 2)

- Skewness of $(Y_1 – Y_0) = E((Y_1 – \mu_1) – (Y_0 – \mu_0))^3 =$
- $E(Y_1 – \mu_1)^3 - 3E(Y_1 – \mu_1)^2(Y_0 – \mu_0)) + 3E((Y_1 – \mu_1)(Y_0 – \mu_0)^2) - E(Y_0 – \mu_0)^3$.

- While there is no known formula for

$E((Y_1 – \mu_1)(Y_0 – \mu_0)^2) - E(Y_1 – \mu_1)^2(Y_0 – \mu_0))]$, when the correlation coefficient is ρ, taking the difference score between follow-up and baseline often helps to reduce skewness, because baseline and follow-up measures are usually in the same family of distributions.

- Some examples are baseline and follow-up depression scores, blood pressures, and cholesterol measurements.

Δ Baseline Adjustment in the Pre-Post Model (2 of 2)

- Let N = number of participants in a study, j = index of participant from 1 to N.
- Let G_k = group indicator, k = 0 control, 1 for treatment;
- $G_k = 1$ if person is a member of the group; 0 otherwise.
- Let Y_{j0} = outcome for the jth participant at baseline, aka pre-intervention or time 0.
- Let Y_{j1} = outcome for the jth participant at follow-up, or time 1.
- Let Δ_j = change score from baseline to follow-up = $Y_{j1} - Y_{j0}$.
- Let ε_j = error term.
- Then, the change score model for a pre-post design would be

➢ **Let $\Delta j = \beta_0 + \beta_1 G_0 Y_{j0} + \beta_2 G_1 Y_{j0} + \beta_3 G_1 + \varepsilon_j$.**

Baseline Adjustment in the Pre-Post Model (2 of 2)

$\Delta j = \beta_0 + \beta_1 G_0 Y_{j0} + \beta_2 G_1 Y_{j0} + \beta_3 G_1 + \varepsilon_j$

- Let $\bar{Y}_{00}$= baseline mean of control group, $\bar{Y}_{10}$= baseline mean of treatment group.
- Estimated mean change score for the control group, adjusted for baseline = $\beta_0 + \bar{Y}_{00}\beta_1$.
- Estimated mean change score of the treatment group, adjusted for baseline = $\beta_0 + \bar{Y}_{10}\beta_2 + \beta_3$.
- Intervention effect = Difference in change scores between treatment group, adjusted for baseline difference = $-\beta_1\bar{Y}_{00} + \beta_2\bar{Y}_{10} + \beta_3$.
- When there is no missing data at baseline, the difference score model is an “intent-to-treat” model because it includes all available data at each time point.

Advantages to Model with Δ_j (change) as the Outcome

- 1) The difference score often has a **symmetrical error distribution**, even when Y_{j1} (value at time 1) does not.
- 2) To evaluate the effectiveness of a study, **researchers are interested in the amount of change in the outcome**.
- 3) This model produces the average **change within a treatment group, adjusted for the baseline value within that treatment group**. Similarly, the intervention effect can be calculated to account for baseline difference between treatment groups.
- Next: Example with Hemoglobin A1c (HbA1c), the state of the art measure of blood sugar among people with diabetes.

SAS Code for Proc GLM

```
M6BL_HbA1c = M6_HbA1c – BL_HbA1c; /* Compute difference score in data step */

RandomizationN = -Randomization; /* treatment = -1; control =0. Largest value = reference in SAS.

/* Baseline mean HbA1c’s = 7.9 treatment, 7.7 control */

ods html; ods graphics on;

Proc GLM Data= AcrossTimeHorizontal PLOTS =(RESIDUALS DIAGNOSTICS);

Class RandomizationN;  /* reference will be 0 = control */
```

- Model M6BL_HbA1c=RandomizationN RandomizationN*BL_HbA1c / Solution;
- Estimate 'Treat M6 - BL' Intercept 1 RandomizationN 1 0 RandomizationN*BL_HbA1c 7.9 0;
- Estimate 'Ctl M6 - BL‘ Intercept 1 RandomizationN 0 1 RandomizationN*BL_HbA1c 0 7.7;
- Estimate 'Int Effect M6 - BL' RandomizationN 1 -1 RandomizationN*BL_HbA1c 7.9 -7.7;
- Run; Quit;
- ods graphics off; ods html close;

Proc GLM Output Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2.17	0.75	2.91	0.0041
RandomizationN -1	0.29	0.89	0.33	0.7427
RandomizationN 0	0	.	.	.
BL_HbA1c*Randomizati -1	-0.38	0.06	-6.09	<.0001
BL_HbA1c*Randomizati 0	-0.28	0.10	-2.94	0.0037

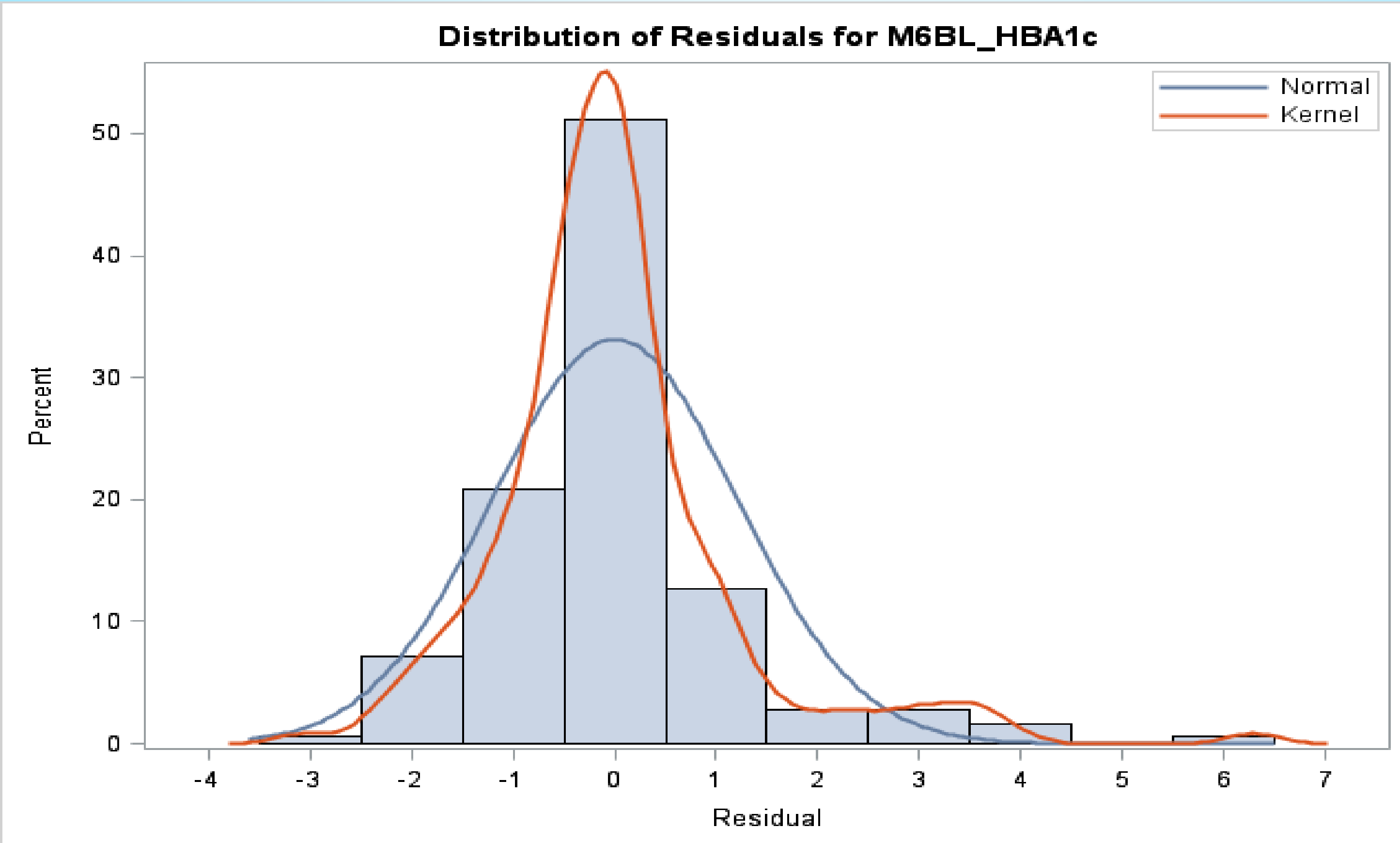
- Note: RandomizationN, by itself, is non-significant.
- Not a problem because the outcome of interest is the change score with baseline adjustment.
- While the estimate for RandomizationN is 0 for the control group, SAS estimates interaction effects between the baseline value and RandomizationN for both the treatment and control groups.

Proc GLM Estimates for Change Scores

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Treatment M6 - BL	-0.53	0.11	-4.76	<.0001
Control M6 - BL	0.00	0.16	-0.01	0.9959
Intervention Effect M6 - BL	-0.53	0.19	-2.76	0.0064

- In a diabetes intervention, a drop of at least 0.5 in HbA1c is considered clinically significant.
- The above table indicates that the treatment group experienced a significant drop in HbA1c from baseline to 6-month follow-up, while the control group stayed approximately the same.
- The effect for the treatment group is significant within-group (1st line) and in comparison to the control group (3rd line).

Diagnostic Residual Plot for Month 6 – Baseline Difference Score



SAS Macro for Difference Score, Post - Pre

```
%Macro DiffM6(VarName, BLName, BLTreat, BLControl);
ods html; ods graphics on;
Proc GLM Data= AcrossTimeHorizontal PLOTS(only)=(RESIDUALS DIAGNOSTICS);
Class RandomizationN;
Model &VarName=RandomizationN RandomizationN*&BLName  / Solution;
Estimate 'Treat M6 - BL'  Intercept 1  RandomizationN 1 0 RandomizationN*&BLName &BLTreat 0;
Estimate 'Control M6 - BL'  Intercept 1  RandomizationN 0 1 RandomizationN*&BLName 0
&BLControl;
Estimate 'Int Effect M6 - BL' RandomizationN 1 -1 RandomizationN*&BLName &BLTreat -&BLControl;
Run; Quit;
ods graphics off; ods html close;
%MEnd DiffM6;

%DiffM6(M6BL_HbA1c, BL_HbA1c, 7.9, 7.7);
%DiffM6(M6BL_TotalCholesterol, BL_TotalCholesterol, 182.3, 182.9);
%DiffM6(M6BL_LDLCholesterol, BL_LDLCholesterol, 96.1, 95.4);
```

Introduction to the Linear Mixed Model (1 of 2)

- The general form of the linear mixed model is $Y = X\beta + Zb + \varepsilon$, where X = matrix of fixed effects and Z = matrix of random effects,
- β = fixed effect estimates, b = random effects estimates, and ε = error terms.
- $b \sim N(0, D)$, $\varepsilon \sim N(0, \sum)$; b and ε are independent D and $\sum$ are covariance matrices.
- Consider two time points and two treatment groups.
- Let Y_{ijk} = outcome variable; (i, j, k) = (randomization, time point, subject).
- i = 0 for control and 1 for treatment; j = 1 for pre-intervention and 2 for post-intervention. k = kth subject.
- Let R = 0 for control and 1 for treatment, Let T = 0 for pre-intervention and 1 for post-intervention.
- Linear Mixed Model (LMM) for $Y_{ijk} = \beta_0 + \beta_1 R + \beta_2 T + \beta_3 RT + \varepsilon_{ijk}$, where ε_{ijk} = error term, $\varepsilon_{ijk} \sim N(0, \sum)$, $\sum$ is 2 x 2 for this example.

Introduction to the Linear Mixed Model (2 of 2)

- Estimated Mean: $E(Y_{ij}) = \beta_0 + \beta_1 R + \beta_2 T + \beta_3 RT$.
- Control group means are β_0 at pre-intervention and $(\beta_0 + \beta_2)$ at post-intervention.
- Treatment group means are $(\beta_0 + \beta_1)$ at pre-intervention and $(\beta_0 + \beta_1 + \beta_2 + \beta_3)$ at post-intervention.
- Change scores are β_2 for the control group and $(\beta_2 + \beta_3)$ for the treatment group.
- The intervention effect is the difference in change scores for the treatment and control groups = β_3.
- If the data is missing at random (MAR), the linear mixed model will provide unbiased estimates.
- All available data at each time point will be used, even if a participant does not have data at each time point. Because all available data is used, the linear mixed model is the "intent to treat" model (West, Welch, Galecki, 2007). Whereas, the difference score regression model in the previous slides would omit participants with incomplete data.

SAS Code for Mixed Model (1 of 4)

First, convert dataset from across (one subject ID per row) to long. In the long dataset, each row contains ID, time point, and the outcome at that time point. In this example, there are multiple time points. This same technique can be used for any number of time points ≥ 2.

```
Data Across;
Merge Baseline Month6Data Month12Data Month18Data;
by ID;

/* Compute difference scores, begin with HbA1c (blood sugar) */
M6BL_HbA1c=M6_HbA1c-BL_HbA1c;
M12BL_HbA1c=M12_HbA1c-BL_HbA1c;
M18BL_HbA1c=M18_HbA1c-BL_HbA1c;

M6BL_TotalCholesterol=M6_TotalCholesterol-BL_TotalCholesterol;
M12BL_TotalCholesterol=M12_TotalCholesterol-BL_TotalCholesterol;
M18BL_TotalCholesterol=M18_TotalCholesterol-BL_TotalCholesterol;
Run;
```

SAS Code for Mixed Model (2 of 4)

```
/* Create Across_Long for Proc Mixed (3 time points in this example) */
Data Across_Long;
Set Across;
/* Baseline */
Timepoint=0; TimepointN=0; HbA1c=BL_HbA1c;  TotalCholesterol=BL_TotalCholesterol;
Output; /* Need output statement at each time point */

/* 6 months */
Timepoint=1; TimepointN=-1;  HbA1c=M6_HbA1c;  TotalCholesterol=M6_TotalCholesterol;
DeltaHbA1c=M6BL_HbA1c;   DeltaTotalCholesterol=M6BL_TotalCholesterol;
Output; /* Need output statement at each time point */

/* 12 Months */
Timepoint=2; TimepointN=-2; HbA1c=M12_HbA1c;  TotalCholesterol=M12_TotalCholesterol;
DeltaHbA1c=M12BL_HbA1c;   DeltaTotalCholesterol=M12BL_TotalCholesterol;
Output; /* Need output statement at each time point */
Run;
```

SAS Code for Mixed Model (3 of 4)

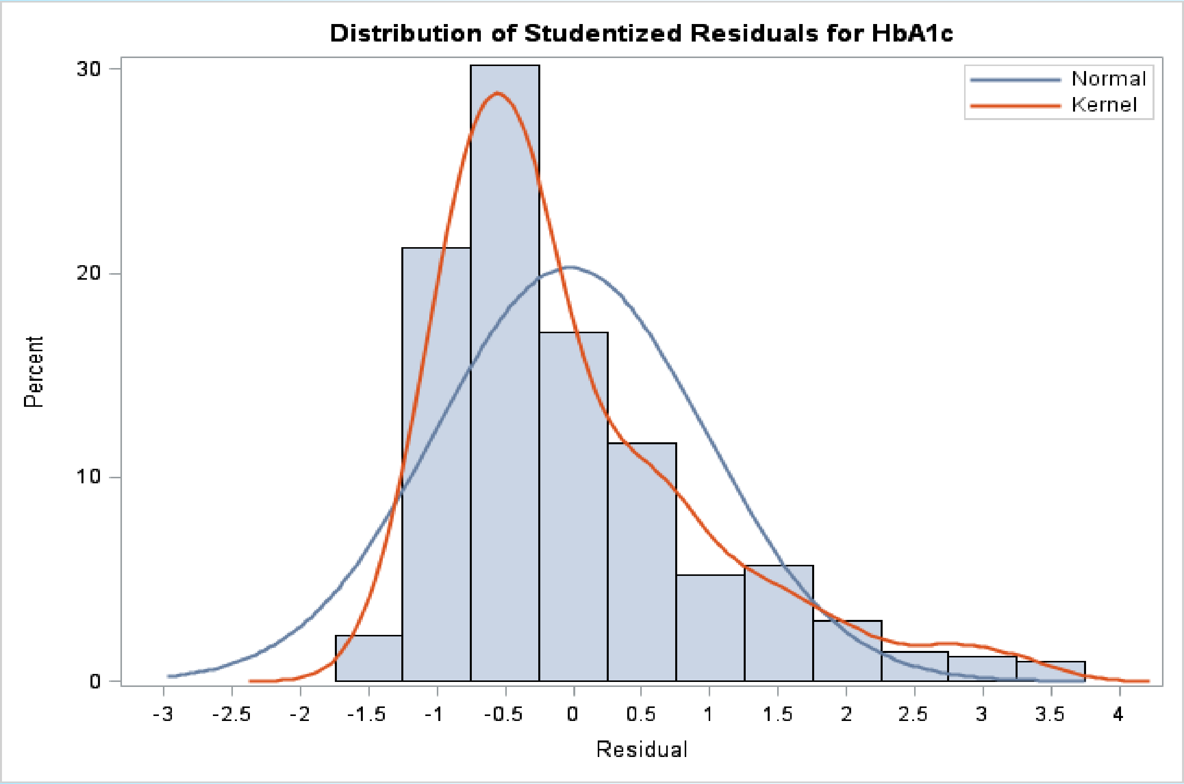
```
/* TimePointN = 0 for baseline and -1 for 6 months, coded so that SAS will set 0 as the reference */
ods html; ods graphics on;
Proc Mixed Data= Across_Long Method=REML NOCLPRINT plots =(StudentPanel(conditional box));
 Class ID TimepointN RandomizationN;
Model HbA1c=TimepointN RandomizationN TimepointN*RandomizationN / Solution
Influence(effect=ID Est) ddfm=KR;

Repeated TimepointN / Type=UN Subject=ID R RCorr;

Estimate 'Treatment Baseline'
Intercept 1 RandomizationN 1 0 TimePointN 0 1 TimepointN*RandomizationN 0 0 1 0;

Estimate 'Control Baseline'
Intercept 1 RandomizationN 0 1 TimePointN 0 1 TimepointN*RandomizationN 0 0 0 1;

Estimate 'Treatment-Control Baseline'
RandomizationN 1 -1 TimepointN*RandomizationN 0 0 1 -1;
```

SAS Code for Mixed Model (4 of 4)

```
Estimate 'Treatment Month 6'
Intercept 1 RandomizationN 1 0 TimePointN 1 0 TimepointN*RandomizationN 1 0 0 0;

Estimate 'Control Month 6'
Intercept 1 RandomizationN 0 1 TimePointN 1 0 TimepointN*RandomizationN 0 1 0 0;

Estimate 'Treatment M6 - BL'
TimePointN 1 -1 TimepointN*RandomizationN 1 0 -1 0;

Estimate 'Control M6 - BL'
TimePointN 1 -1 TimepointN*RandomizationN 0 1 0 -1;

Estimate 'Intervention Effect M6 - BL'
TimepointN*RandomizationN 1 -1 -1 1;
Run;
ods graphics off; ods html close;
```

Residual Plot

With the this model, the mean of any treatment group at any time point can be estimated. However, the residual plot indicates departure from the normality assumption, as seen in the figure.



Difference Model at 4 Times, 3 Treatment Group

- Why? Study on whether effects 6-month diabetes education program, conducted by community health workers, can be sustained to 18 months by adding peer support between 6 months and 18 months.
- 3 Groups – Control Group, Diabetes Education, Diabetes Education + Peer Support.
- Success measured by whether difference between baseline to a given time point continues to be significant after 6 months. I.E., is blood sugar significantly lower at 6 months, 12 months, and 18 months, compared to baseline?

Δ Difference Model at 4 Times, 3 Treatment Group (1 of 3)

- Let G_k = group indicator, k = 0 control, 1 for treatment 1, 2 for treatment.
- G_k = 1 if person is a member of the group; 0 otherwise.
- Let m = time point; 0 = baseline, 1 = 6 months, 2 = 12 months, 3 = 18 months.
- Let t_m = time point indicator; t_1 =1 if 6 months, 0 otherwise, t_2 = 1 if 12 and t_3 = 1 if 18 months.
- Let Y_{j0} = outcome for the jth participant at baseline, aka pre-intervention or time 0.
- Let Y_{jm} = outcome for the jth participant at follow-up time m.
- Let Δ_{jm} = change score from baseline to follow-up = Y_{jm}- Y_{j0}.
- Let ε_{jm} = error term.

Difference Model at 4 Times, 3 Treatment Group (2 of 3)

$$\Delta_{jm} = \beta_0 + \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} + \begin{bmatrix} \beta_3 & \beta_4 & \beta_5 \end{bmatrix} \begin{bmatrix} G_0 Y_{j0} \\ G_1 Y_{j0} \\ G_2 Y_{j0} \end{bmatrix} + \begin{bmatrix} \beta_6 & \beta_7 & \beta_8 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} +$$

$$\begin{bmatrix} \beta_9 & \beta_{10} & \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \end{bmatrix} \begin{bmatrix} G_1 t_1 \\ G_1 t_2 \\ G_1 t_3 \\ G_2 t_1 \\ G_2 t_2 \\ G_2 t_3 \end{bmatrix} + \varepsilon_{jm}$$

Difference Model at 4 Times, 3 Treatment Group (3 of 3)

- For example, consider the change in treatment group 1 from baseline to 12 months. Δ(group 1, time 2) = $\beta_0 + \beta_1 + \beta_4\bar{y}_{10} + \beta_7 + \beta_{10}$.
- Change in group 2 from baseline to 12 months: $\beta_0 + \beta_2 + \beta_5\bar{y}_{20} + \beta_7 + \beta_{13}$.
- Change in control group from baseline to 12 months = $\beta_0 + \beta_3\bar{y}_{00} + \beta_7$.
- Intervention effect for group 1 at 12 months = $\beta_1 + \beta_4\bar{y}_{10} - \beta_3\bar{y}_{00} + \beta_{10}$.
- Average intervention effect for groups 1 and 2, compared to control =
- $.5\beta_1 + .5\beta_4\bar{y}_{10} + .5\beta_{10} + .5\beta_2 + .5\beta_5\bar{y}_{20} + .5\beta_{13} - \beta_3\bar{y}_{00}$.

SAS Code for Mixed Model of Difference Scores

```
Proc Mixed Data= Across_Long Method=REML NOCLPRINT
  plots(only)=(StudentPanel(conditional box));  /* Diagnostic Plots */
Class ID TimepointN RandomizationN;  /* Group = RandomizationN */

/* Outcome is delta (change score).  Adjust for baseline x group. */
Model Delta_HbA1c = BL_HbA1c*RandomizationN TimepointN RandomizationN
TimepointN*RandomizationN / Solution Influence(effect=ID Est) ddfm=KR;

Repeated TimepointN / Type=UN Subject=ID;
Where TimePointN NE 0;  /* Be sure to exclude baseline row in dataset.  */
Run;
```

Macro with Estimate Statements (1 of 2)

```
%Macro LMMDiff3(VarName, CorMat, BLName, BLRef, BLGrp1, BLGrp2);
/* VarName=outcome, CorMat=Correlation structure */
/* BLName=Name of baseline variable */
/* BLRef=Baseline mean for the control or reference group */
/* BLGrp1=Baseline mean for treatment 1 group, BLGrp2=Baseline mean for treatment 2 group */

ods html; ods graphics on;
 *Compute average of treatment groups 1 & 2 to get combined intervention estimates;
%Let BLTrtAve=%sysfunc(mean(&BLGrp1, &BLGrp2));

*Use %syseval because SAS won’t allow .5*macro variable in an estimate statement;
%Let BLTrtAve2=%sysevalf(&BLTrtAve/2);

Proc Mixed Data= Across_Long Method=REML NOCLPRINT plots(only)=(StudentPanel(conditional
box));
Class ID TimepointN RandomizationN;
Model &VarName=&BLName*RandomizationN TimepointN RandomizationN
TimepointN*RandomizationN / Solution Influence(effect=ID Est) ddfm=KR;
Repeated TimepointN / Type=&CorMat Subject=ID R RCorr;
```

Macro with Estimate Statements (2 of 2)

```
/* Estimate Examples at Month 12 */
Estimate 'Ref M12 - BL'  Intercept 1  RandomizationN 0 0 1 TimePointN 0 1 0
TimepointN*RandomizationN 0 0 0 0 0 1 0 0 0 &BLName*RandomizationN 0 0 &BLRef;

Estimate 'Trt1 M12 - BL' Intercept 1  RandomizationN 0 1 0 TimePointN 0 1 0
TimepointN*RandomizationN 0 0 0 0 1 0 0 0 0 &BLName*RandomizationN 0 &BLGrp1 0;

Estimate 'Trt2 M12 - BL' Intercept 1  RandomizationN 1 0 0 TimePointN 0 1 0
TimepointN*RandomizationN 0 0 0 1 0 0 0 0 0 &BLName*RandomizationN &BLGrp2 0 0;

Estimate 'AveTrt1,2-RefM12-BL' RandomizationN .5 .5 -1 TimepointN*RandomizationN 0 0 0 .5 .5 -1 0
0 0 &BLName*RandomizationN &BLTrtAve2 &BlTrtAve2 -&BLRef;

Estimate 'Int Effect1 M12 - BL' RandomizationN 0 1 -1 TimepointN*RandomizationN 0 0 0 0 1 -1 0 0 0
&BLName*RandomizationN 0 &BLGrp1 -&BLRef;

Estimate 'Int Effect2 M12 - BL' RandomizationN 1 0 -1 TimepointN*RandomizationN 0 0 0 1 0 -1 0 0 0
&BLName*RandomizationN &BLGrp2 0 -&BLRef;
```

Effect	Time pointN	Random izationN	Estimate	Std Err	Pr > |t|
BL_HbA1c*Randomizati		-2	-0.42	0.09	<.0001
BL_HbA1c*Randomizati		-1	-0.34	0.08	<.0001
BL_HbA1c*Randomizati		0	-0.43	0.08	<.0001
TimepointN	-3		0.28	0.24	0.2513
TimepointN	-2		0.03	0.16	0.8438
TimepointN	-1		0.00	.	.
RandomizationN		-2	-0.48	0.98	0.6252
RandomizationN		-1	-1.02	0.90	0.2621
RandomizationN		0	0.00	.	.
Timepoint*Randomizat	-3	-2	-0.11	0.34	0.7458
Timepoint*Randomizat	-3	-1	0.10	0.33	0.7638
Timepoint*Randomizat	-3	0	0.00	.	.
Timepoint*Randomizat	-2	-2	0.09	0.24	0.6895
Timepoint*Randomizat	-2	-1	0.21	0.22	0.326
Timepoint*Randomizat	-2	0	0.00	.	.
Timepoint*Randomizat	-1	-2	0.00	.	.
Timepoint*Randomizat	-1	-1	0.00	.	.
Timepoint*Randomizat	-1	0	0.00	.	.

Label	Estimate	Std Err	Pr > |t|
Reference M6 - BL	-0.06	0.16	0.691
Trt1 M6 - BL	**-0.36**	**0.15**	**0.014**
Trt2 M6 - BL	**-0.74**	**0.18**	**<.0001**
Int Effect1 M6 - BL	-0.30	0.21	0.166
Int Effect2 M6 - BL	**-0.68**	**0.24**	**0.005**
Reference M12 - BL	-0.03	0.17	0.853
Trt1 M12 - BL	-0.12	0.16	0.472
Trt2 M12 - BL	**-0.62**	**0.20**	**0.002**
Int Effect1 M12 - BL	-0.09	0.24	0.718
Int Effect2 M12 - BL	**-0.59**	**0.26**	**0.027**
Reference M18 - BL	0.21	0.24	0.380
Trt1 M18 - BL	0.02	0.23	0.946
Trt2 M18 - BL	**-0.58**	**0.24**	**0.020**
Int Effect1 M18 - BL	-0.20	0.34	0.559
Int Effect2 M18 - BL	**-0.79**	**0.34**	**0.023**

- BL = baseline, M6 = month 6, M12 = month 12, M18 = month 18.
- Trt = treatment group.
- Int Effect = Intervention Effect.
- Longitudinal analysis of difference scores indicates that drop in HbA1c continues to be significant out to 18 months in the peer support group.

Multiple Comparisons

- Let α = probability of a type 1 error, i.e., probability of rejecting the null hypothesis when the null hypothesis is true.
- Example: Concluding that there is a significant difference between baseline and 6-month HbA1c in a treatment group, when no difference exists.
- If $\alpha = .05$ and two comparisons are made, such as group 1 to control and group 2 to control from baseline to 6-months, the probability of not making a type 1 error on either test = $(1 - \alpha)^2 = (1 - .05)^2 = .9025$. The probability of making a type 1 error on at least one of the two tests = $1 - (1 - \alpha)^2 = .0975$.
- If n tests are done, the type 1 error = $1 - (1 - \alpha)^n$.
- For $\alpha = .05$ and n = 5, $1 - (1 - \alpha)^n = .2262$. α' = probability of at least one type one error from .05 to .2262.
- Multiple comparisons increase α'. Goal of adjusting for multiple comparisons is to set $\alpha' = \alpha$.

Adjustment for Multiple Comparisons

- In SAS Proc Logistic, Estimate statements can be written with the “Adjust=” option, which adjusts for multiple comparisons.
- In both Procs GLM and Mixed, the Adjust option is not available on the Estimate statement.
- In Proc GLM, need to use LSMeans to adjust for multiple comparisons.
- In Proc Mixed, use either LSMeans or LSMEstimate.
- For both LSMeans and LSMEstimate, all effects must be class variables.
- To adjust for multiple comparisons, need to use model in which outcome at time, point (i.e., HbA1c), rather than difference score (i.e., delta_HbA1c) is the outcome.
- Adjust with multiple comparisons by Adjust=Simulate option on the LSMEANS statement. This technique uses Monte Carlo simulation and performs well according to the literature (Westfall and Young,1993).

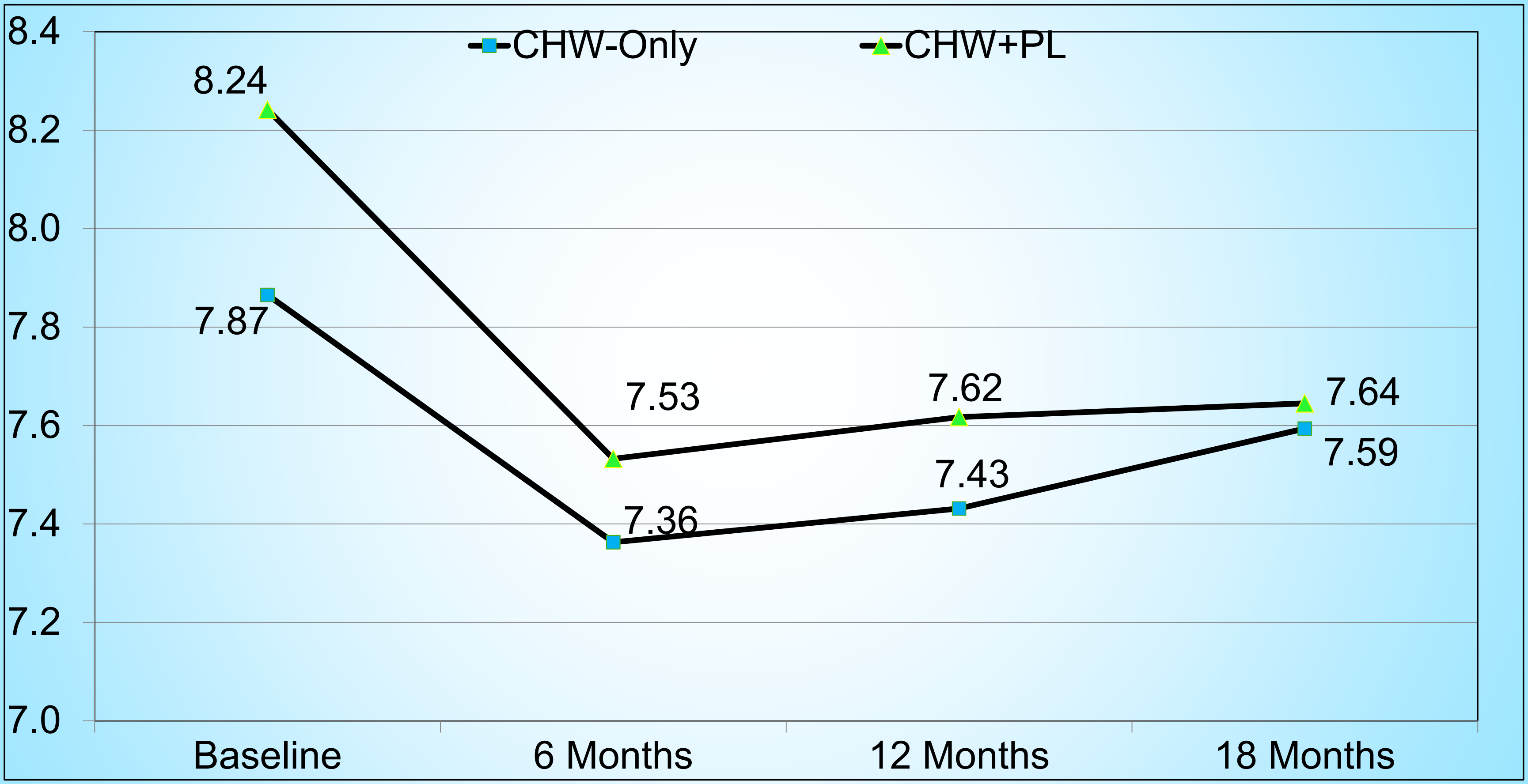
Responses from Reviewers to Longitudinal Analysis with Difference Scores

- The longitudinal model described in this presentation, with difference scores as the outcome and adjustment for baseline, has been used to publish two articles in medical journals.
- “Peer-led, empowerment-based, approach to self-management efforts in diabetes (PLEASED): A randomized controlled trial in the African-American community” was recently published in the Annals of Family Medicine.
- “Comparative Effectiveness of Peer Leaders and Community Health Workers in Diabetes Self-management Support: Results of a Randomized Controlled Trial” was published Diabetes Care.

Example Table from Diabetes Care Publication

Outcome	Time Point	Baseline	6 Months – Baseline	12 Months – Baseline	18 Months – Baseline
HbA1c	CHW+PL Group	8.2 (7.7, 8.8)	-0.7 (-1.0, -0.4) p<.0001	-0.6 (-0.9, -0.3) p=0.001	-0.6 (-1.0, -0.2) p=0.009
	CHW-Only Group	7.8 (7.4, 8.3)	-0.5 (-0.8, -0.3) p=0.0004	-0.4 (-0.7, -0.1) p=0.011	-0.3 (-0.7, 0.2) p=0.234
	p-value difference between groups	0.253	0.883	0.867	0.725

Example Graph from Diabetes Care Publication: Trajectory of HbA1c



Conclusions

- When there are no missing values in the outcome at baseline and the distribution is skewed, longitudinal analysis of differences scores can be a useful analysis technique.
- Taking the difference score between skewed variables from the same family of distributions often produces a result that is more symmetrical.
- Longitudinal analysis of difference scores can be implemented in SAS Proc GLM for two time points and in Proc Mixed for more than two time points.
- Analyses of longitudinal difference scores, with baseline adjustment, have been accepted for publication by major journals.
- Longitudinal analysis of difference scores is a useful technique in the data analyst's tool kit.

Acknowledgements

- REACH-Detroit staff community health workers.
- Peers for Progress and peer support staff.
- This research was supported by a Peers for Progress grant from the American Association of Family Physicians Foundation (grant P30DK092926) (Michigan Center for Diabetes Translational Research), by the National Institute of Diabetes and Digestive and Kidney (NIDDK) Diseases (grant R18DK0785501A1), and by the Centers for Disease Control and Prevention (cooperative agreement no. U50/CCU417409).

Contact Information

- Brandy R. Sinco, Statistician and Programmer/Analyst
- University of Michigan School of Social Work
- 1080 S. University St.
- Box 183
- Ann Arbor, MI 48109-1106
- E-Mail: brsinco@umich.edu
- Phone: 734-763-7784