Paper 2400-2015

Modeling Effect Modification and Higher-Order Interactions: Novel Approach for Repeated Measures Design using the LSMESTIMATE Statement in SAS® 9.4

Pronabesh DasMahapatra, MD, MPH, PatientsLikeMe Inc.; Ryan Black, PhD, NOVA Southeastern University

ABSTRACT

Effect modification (moderation) occurs when the association between a predictor of interest and the outcome, is differential across levels of a third variable – the modifier. Effect modification is statistically tested as the interaction effect between the predictor and the modifier. In repeated measures studies (with more than 2 time points), higher-order (three-way) interactions must be considered to test effect modification by adding time to the interaction terms. Custom fitting and constructing these repeated measures models are difficult and time consuming, especially with respect to estimating post-fitting contrasts. With the advancement of LSMESTIMATE statement in SAS®, a simplified approach can be used to custom test for higher-order interactions with post-fitting contrasts within a repeated measures framework. The purpose of this illustration is to provide a simulated example with tips and techniques to use an application of the nonpositional syntax of the LSMESTIMATE statement to test effect modification in repeated measures studies. This approach is applicable to exploring modifiers in Randomized Controlled Trials (RCTs) which goes beyond the treatment effect on outcome to a more functional understanding of the factors that can enhance, reduce or change this relationship. Using this technique we can easily identify differential changes for specific subgroups of individuals or patients which subsequently impact treatment decision making. We provide examples of conventional approach to higher-order interaction and post-fitting tests using the traditional positional syntax of ESTIMATE statement and compare and contrast this to the nonpositional syntax of the LSMESTIMATE statement.

INTRODUCTION

Effect modification (moderation) tests whether the association between the predictor variable and the outcome is differential across levels of a third variable – the modifier (Fairchild & MacKinnon, 2010). In a RCT, the presence of effect modification indicates that the therapeutic effect changes across different levels of a factor, which may be either manipulated in an experimental setting or a naturally occurring variable such as gender i.e., “for whom or under what conditions the treatment works” (Baron & Kenny, 1986) (Kraemer, Wilson, Fairburn, & Agras, 2002). Therefore, identification of effect modification is important in treatment decision making and has been explored in several studies (Judd, Kenny, & McClelland, 2001).

Effect modification is evaluated as an interaction effect, where the effects of the treatment on the outcome variable depend on levels of the putative modifier. In RCTs with repeated measures design, this is examined by testing difference between treatment arms over time and determining whether the putative modifier predicts that difference. Furthermore, post-fitting contrasts of the levels of the modifier identify the differential changes for specific subgroups of individuals or patients. However, this models are difficult to construct and time consuming, especially when fitting higher-order interactions.

With the advancement of LSMESTIMATE statement in SAS®, a simplified approach can be used to custom fit modifier effects. The purpose of this article is to illustrate a simplistic approach using an application of the nonpositional syntax of the LSMESTIMATE statement to explore effect modification in RCTs in a simulated example.

EXAMPLE FROM A SIMULATED RANDOMIZED CONTROLLED TRIAL DESIGN

We illustrate our approach for a 2 x 4 factorial design RCT with two condition arms (treatment vs. control); over 4 time points (baseline, 1st follow-up, 2nd follow-up and 3rd follow-up) and age as the modifier (three levels: 18-39 years, 40-59 years and 60-79 years).

Data were simulated in SAS® by applying the following specification: (1) 10,000 subjects with 4 assessments (baseline, 1st follow-up, 2nd follow-up and 3rd follow-up), (2) random effect component simulated to account for within-subject correlation, (3) 1:1 ratio of subject allocation to each condition arms i.e., treatment and control arms, (4) a third of subjects randomly assigned to each age group, (5) regression coefficients for interaction terms (age*condition*time) parameterized to zero for age groups 18-39 years and 40-59 years; and non-zero linear change over time for 60-79 years. The simulated code is shown below:

```
PROC FORMAT; * the FORMAT procedure;
value age
   0='18-39'
   1='40-59'
   2='60-79';
value condition
   1='Treatment'
   0='Control';

data modifier;
   ***** set seed for random number generator *****;
   seed = 9876543;
```

```
   ***** fix parameters ****;
   b0 = 0.50; b1 = 0.75; b2 = 1.00; b3 = 1.25; b4 = 1.50; b5 = 1.75;
   b6 = 2.00; b7 = 2.25; b8 = 2.50; b9 = 2.75; b10 = 3.00; b11 = 3.25;
   b12 = 3.50; b13 = 3.75; b14 = 4.00; b15 = 4.75; b16 = 0.00; b17 = 0.00;
   b18 = 0.00; b19 = 0.00; b20 = 0.00; b21 = 0.00; b22 = 5.00; b23 = 5.25;
   b24 = 5.50;
   ***** first do loop *****;
      do subject=1 to 10000;
          random_effect = sqrt(.50)*rannor(seed); *add random effect for within-subject
correlation;
         condition = (subject>5000); * half of observations assigned to each condition;
         age = ranbin(seed,2,.33); * randomly assign age to three groups *
            ***** second do loop *****;
             do time=1 to 4;
            ***** construct linear predictor *****;
               y = b0 + b1*(age=0)
                    + b2*(age=1)
                     + b3*(age=2)
                     + b4*(age=0)*(condition=1)
                     + b5*(age=1)*(condition=1)
                     + b6*(age=2)*(condition=1)
                     + b7*(age=0)*(time=2)
                     + b8*(age=0)*(time=3)
                  + b9*(age=0)*(time=4)
                     + b10*(age=1)*(time=2)
                     + b11*(age=1)*(time=3)
                  + b12*(age=1)*(time=4)
                  + b13*(age=2)*(time=2)
                     + b14*(age=2)*(time=3)
                  + b15*(age=2)*(time=4)
                  + b16*(age=0)*(condition=1)*(time=2)
                     + b17*(age=0)*(condition=1)*(time=3)
                  + b18*(age=0)*(condition=1)*(time=4)
                  + b19*(age=1)*(condition=1)*(time=2)
                     + b20*(age=1)*(condition=1)*(time=3)
                  + b21*(age=1)*(condition=1)*(time=4)
                  + b22*(age=2)*(condition=1)*(time=2)
                     + b23*(age=2)*(condition=1)*(time=3)
                  + b24*(age=2)*(condition=1)*(time=4)
                     + random_effect
                     + rannor(seed);
             output;
            end;
           end;
   keep y condition time subject age;
   format age age. condition condition. ;
run;
quit;
```

TEST OF TREATMENT EFFECTS

Using the GLIMMIX procedure in SAS® 9.4, first, we test the main treatment effects on outcome (y) for the statistical model:

$Y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + e_{ijk}$

where, $\mu + \alpha_i + \beta_k + (\alpha\beta)_{ik}$ is the mean for treatment i at time k, containing effects for treatment, time, and treatment X time interaction.

e_{ijk} is the random error associated with the measurement at time k on the j^{th} subject assigned to treatment i: The model treated time as a random residual to allow for covariation across time points. Statistical focus was the interaction effect condition X time, as this effect tests differences between conditions over time. The sample code is as follows:

```
PROC GLIMMIX data=modifier;
   CLASS condition time subject;
   MODEL y = condition|time / s;
   RANDOM time / subject=subject type=cs residual;
   LSMEANS condition*time;
   ESTIMATE
        'cont-exp 2-1' condition*time 1 -1 0 0 -1 1 0 0,
        'cont-exp 3-1' condition*time 1 0 -1 0 -1 0 1 0,
        'cont-exp 4-1' condition*time 1 0 0 -1 -1 0 0 1
                       / ADJUST=simulate (seed=121211 report) ADJDFE=row cl;
```

```
run;
```

This above model with Restricted Maximum Likelihood (RMLE) estimation method, a variance-covariance matrix blocked by subject (compound symmetric covariance matrix), and treating time as a random residual yielded a -2 Res Log Likelihood (-2 X LL) of 140522.

The full factorial model shows that all Type III tests of fixed effects in the model, namely, condition ($F_{1, 9999} = 4219.91$), time ($F_{3, 29994} = 17612.80$), and condition X time ($F_{3, 29994} = 141.17$) are significant. This is expected as the data is generated to demonstrate results as such.

The post-fitting contrast of interest is to compare the change over time on outcome measure between conditions. The ESTIMATE statement processes results for this comparison. As simulated, the treatment arm showed greater change in outcome score at all follow up periods (using baseline or time=1 as the reference) compared to the control arm after correcting for Type I error rate by multiplicity adjustment. In effect, this output is analogous to showing that the treatment is efficacious over time.

Fit Statistics	
-2 Res Log Likelihood	150273.1
AIC (smaller is better)	150277.1
AICC (smaller is better)	150277.1
BIC (smaller is better)	150291.5
CAIC (smaller is better)	150293.5
HQIC (smaller is better)	150282
Generalized Chi-Square	62523.12
Gener. Chi-Square / DF	1.56

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9998	4219.91	<.0001
time	3	29994	17612.8	<.0001
condition*time	3	29994	141.17	<.0001

Estimates Adjustment for Multiplicity: Simulated											
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
cont-exp 2-1	0.5349	0.03537	29994	15.12	<.0001	<.0001	0.05	0.4656	0.6042	0.4526	0.6172
cont-exp 3-1	0.6083	0.03537	29994	17.2	<.0001	<.0001	0.05	0.539	0.6776	0.526	0.6906
cont-exp 4-1	0.6242	0.03537	29994	17.65	<.0001	<.0001	0.05	0.5549	0.6935	0.5419	0.7065

Output 1. Output of the GLIMMIX Procedure showing model fit, test of fixed effects and post-fitting contrasts for main treatment effects.

THREE-WAY INTERACTION (EFFECT MODIFICATION) USING THE POSITIONAL SYNTAX OF THE ESTIMATE STATEMENT

Once the main treatment effect on outcome is established, it is imperative to construct exploratory models with the modifier as the third variable in higher-order interaction. In our paper, this three-way interaction tested whether the magnitude of treatment effects were different for age groups 18-39 years, 40-59 years and 60-79 years age groups. The statistical model is as follows:

$$Y_{ijkt} = \mu + \alpha_i + \beta_k + \gamma_t + (\alpha\beta)_{ik} + (\alpha\gamma)_{it} + (\beta\gamma)_{kt} + (\alpha\beta\gamma)_{ikt} + e_{ijkt}$$

where, $\mu + \alpha_i + \beta_k + \gamma_t + (\alpha\beta)_{ik} + (\alpha\gamma)_{it} + (\beta\gamma)_{kt} + (\alpha\beta\gamma)_{ikt}$ is the mean for treatment i at time k for age group t, containing effects for treatment, time, age and treatment X time X age interaction.

e_{ijkt} is the random error associated with the measurement at time k on the j^{th} subject in the t^{th} group assigned to treatment i.

First, post-fitting contrasts for each level of the modifier is computed by using the ESTIMATE statement which requires calculation of the coefficient of L matrix. The sample code is shown below:

```
PROC GLIMMIX data=modifier;
   CLASS condition time age subject;
   MODEL y = condition|time|age / s;
   RANDOM time / subject=subject type=cs residual;
   LSMEANS condition*time*age;
   ESTIMATE
   /* Age 18-39 */
      '18-39: cont-exp 2-1' condition*time 1-1 0 0-1 1 0 0 condition*time*age 1 0 0-1 0 0 0 0 0 0
0 0-1 0 0 1 0 0 0 0 0 0 0 0,
```

```
      '18-39: cont-exp 3-1' condition*time 1 0-1 0-1 0 1 0 condition*time*age 1 0 0 0 0 0-1 0 0 0
0 0-1 0 0 0 0 0 1 0 0 0 0 0,
      '18-39: cont-exp 4-1' condition*time 1 0 0-1-1 0 0 1 condition*time*age 1 0 0 0 0 0 0 0 0-1
0 0-1 0 0 0 0 0 0 0 0 1 0 0,
   /* Age 40-59 */
      '40-59: cont-exp 2-1' condition*time 1-1 0 0-1 1 0 0 condition*time*age 0 1 0 0-1 0 0 0 0 0
0 0 0-1 0 0 1 0 0 0 0 0 0 0,
      '40-59: cont-exp 3-1' condition*time 1 0-1 0-1 0 1 0 condition*time*age 0 1 0 0 0 0 0-1 0 0
0 0 0-1 0 0 0 0 0 1 0 0 0 0,
      '40-59: cont-exp 4-1' condition*time 1 0 0-1-1 0 0 1 condition*time*age 0 1 0 0 0 0 0 0 0
0-1 0 0-1 0 0 0 0 0 0 0 0 1 0,
   /* Age 60-79 */
      '60-79: cont-exp 2-1' condition*time 1-1 0 0-1 1 0 0 condition*time*age 0 0 1 0 0-1 0 0 0 0
0 0 0 0-1 0 0 1 0 0 0 0 0 0,
      '60-79: cont-exp 3-1' condition*time 1 0-1 0-1 0 1 0 condition*time*age 0 0 1 0 0 0 0 0-1 0
0 0 0 0-1 0 0 0 0 0 1 0 0 0,
      '60-79: cont-exp 4-1' condition*time 1 0 0-1-1 0 0 1 condition*time*age 0 0 1 0 0 0 0 0 0 0
0-1 0 0-1 0 0 0 0 0 0 0 0 1
                        / ADJUST=simulate (seed=121211 report) ADJDFE=row cl;
run;
```

The model fit, type III test of fixed effects and post-fitting contrasts are shown in the output tables below. The full-factorial model shows that all fixed effects in the model are significant. Of interest, the three way interaction effect of condition X time X age ($F_{6, 9995}$ = 841.38) is statistically significant denoting that age may be a modifier of the relationship between treatment and outcome over time. As the utility of exploring effect modification is to understand the functional relationships among these variables, specifically the effect of treatment on outcome across different levels of age, post-fitting contrasts are required to do so. In our example, the treatment effect on outcomes (over time) is significant only for the oldest age group 60-79 and not for other age levels (see output 2).

However, computation of post fitting contrasts with the conventional ESTIMATE statement is difficult to construct, lengthy and prone to errors. Computing the coefficient of L also requires a working knowledge of linear algebra. As an example, from the above code, consider the post fitting ESTIMATE to test the change in condition (treatment) effect from baseline to 1st follow-up for the age group 18-39 years.

'18-39: cont-exp 2-1' condition*time 1-1 0 0-1 1 0 0 condition*time*age 1 0 0-1 0 0 0 0 0 0 0 0-1 0 0 1 0 0 0 0 0 0 0 0

In the next section, the same coefficient matrix is shown using the nonpositional LSMESTIMATE statement.

Fit Statistics	
-2 Res Log Likelihood	124248.4
AIC (smaller is better)	124252.4
AICC (smaller is better)	124252.4
BIC (smaller is better)	124266.8
CAIC (smaller is better)	124268.8
HQIC (smaller is better)	124257.3
Generalized Chi-Square	39510.31
Gener. Chi-Square / DF	0.99

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
condition	1	9994	19847.2	<.0001
time	3	29982	29014.2	<.0001
condition*time	3	29982	1241.45	<.0001
age	2	9994	9029.03	<.0001
condition*age	2	9994	2921.63	<.0001
time*age	6	29982	2068.11	<.0001
condition*time*age	6	29982	788.76	<.0001

Estimates
Adjustment for Multiplicity: Simulated

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
18-39: cont-exp 2-1	-0.00966	0.0419	29982	-0.23	0.8176	1	0.05	-0.09178	0.07246	-0.1246	0.1052
18-39: cont-exp 3-1	0.07103	0.0419	29982	1.7	0.09	0.5184	0.05	-0.01109	0.1531	-0.04386	0.1859
18-39: cont-exp 4-1	0.0251	0.0419	29982	0.6	0.5491	0.9984	0.05	-0.05701	0.1072	-0.08979	0.14

40-59: cont-exp 2-1	-0.00916	0.04215	29982	-0.22	0.8279	1	0.05	-0.09177	0.07345	-0.1247	0.1064
40-59: cont-exp 3-1	0.03711	0.04215	29982	0.88	0.3786	0.9767	0.05	-0.0455	0.1197	-0.07847	0.1527
40-59: cont-exp 4-1	0.04404	0.04215	29982	1.05	0.296	0.9342	0.05	-0.03857	0.1267	-0.07154	0.1596
60-79: cont-exp 2-1	4.985	0.08705	29982	57.26	<.0001	<.0001	0.05	4.8144	5.1556	4.7463	5.2238
60-79: cont-exp 3-1	5.1375	0.08705	29982	59.02	<.0001	<.0001	0.05	4.9669	5.3081	4.8988	5.3763
60-79: cont-exp 4-1	5.4226	0.08705	29982	62.29	<.0001	<.0001	0.05	5.252	5.5932	5.1839	5.6613

Output 2. Output of the GLIMMIX Procedure showing model fit, test of fixed effects and post-fitting contrasts of higher-order interaction models.

THREE-WAY INTERACTION (EFFECT MODIFICATION) USING THE NONPOSITIONAL SYNTAX OF THE LSMESTIMATE STATEMENT

The following syntax illustrates a simpler way to use the nonpositional syntax of LSMESTIMATE statement to code the same three-way interaction. With more complicated models, such as the one shown in this paper, this new statement can greatly shorten the length of the statement needed to represent the post-fitting contrasts and minimize the chances of coding incorrectly. The nonpositional syntax that is available in the LSMESTIMATE statements enables the following: (1) ignore the underlying ordering of parameters and (2) define directly only the nonzero coefficients that are involved in the post-fitting test (Kiernan, Tobias, Gibbs, & Tao, 2011)

```
PROC GLIMMIX data=modifier;
   CLASS condition time subject age;
   MODEL y = condition|time|age / s;
   RANDOM time / subject=subject type=cs residual;
   LSMEANS condition*time*age;
   LSMESTIMATE condition*time*age
   /* Age 18-39 */
       '18-39: cont-exp 2-1 NonPositional'[1, 1 1 1] [-1, 1 2 1] [-1, 2 1 1] [1, 2 2 1],
       '18-39: cont-exp 3-1 NonPositional'[1, 1 1 1] [-1, 1 3 1] [-1, 2 1 1] [1, 2 3 1],
       '18-39: cont-exp 4-1 NonPositional'[1, 1 1 1] [-1, 1 4 1] [-1, 2 1 1] [1, 2 4 1],
   /* Age 40-59 */
       '40-59: cont-exp 2-1 NonPositional'[1, 1 1 2] [-1, 1 2 2] [-1, 2 1 2] [1, 2 2 2],
       '40-59: cont-exp 3-1 NonPositional'[1, 1 1 2] [-1, 1 3 2] [-1, 2 1 2] [1, 2 3 2],
       '40-59: cont-exp 4-1 NonPositional'[1, 1 1 2] [-1, 1 4 2] [-1, 2 1 2] [1, 2 4 2],
   /* Age 60-79 */
       '60-79: cont-exp 2-1 NonPositional'[1, 1 1 3] [-1, 1 2 3] [-1, 2 1 3] [1, 2 2 3],
       '60-79: cont-exp 3-1 NonPositional'[1, 1 1 3] [-1, 1 3 3] [-1, 2 1 3] [1, 2 3 3],
       '60-79: cont-exp 4-1 NonPositional'[1, 1 1 3] [-1, 1 4 3] [-1, 2 1 3] [1, 2 4 3]
                       / adjust=simulate (seed=121211 report) adjdfe=row cl;
run;
```

The above LSMESTIMATE statements, each bracketed term in the nonpositional syntax defines a coefficient of L, where the first argument (typically a 1 or -1, indicating summation or difference) in brackets defines the coefficient and the second argument defines the level of the effect. In these post-fitting models, there are three effects (condition, time and age) with two levels for condition (treatment, control), four levels for time (baseline, 1st follow-up, 2nd follow-up and 3rd follow-up) and three for age (18-39, 40-59, 60-79). If we consider the statement '18-39: cont-exp 2-1 NonPositional' [1, 1 1 1] [-1, 1 2 1] [-1, 2 1 1] [1, 2 2 1], the values after the comma are the levels of the specified effect. Therefore, the statement computes the difference of differences i.e., the interaction between treatment and control arms from baseline to 1st follow-up for the age group 18-39 years. For explanatory purposes, the levels within each bracket after the comma is written out as follows:

{[control baseline 18-39 years] - [control 1st follow up 18-39 years]} – {[treatment baseline 18-39 years] - [treatment 1st follow up 18-39 years]}

The first set within braces (curly brackets) denote the difference in outcome variable y from baseline to 1st follow up in the control arm and the second set denote the same difference in the treatment arm. The difference of the two sets represent the interaction in the age group 18-39. Comparing this across different age groups as shown in the subsequent statements is a clear way of comparing models with multiple parameters. Further, the nonpositional syntax of LSMESTIMATE statement does not require zeros to occupy positions which must be specified in the positional syntax. In contrast, if we test the same post-fitting parameter using the positional ESTIMATE statement the same model requires complicated computation as shown in this example '18-39: cont-exp 2-1' condition*time 1-1 0 0-1 1 0 0 condition*time*age 1 0 0-1 0 0 0 0 0 0 0 0-1 0 0 1 0 0 0 0 0 0 0 0.

It should be noted that, the two procedures use the same parametrization hence the results are the same.

DISCUSSION

In RCTs, effect modification goes beyond the treatment effect on outcome to a more functional understanding of the factors that can enhance, reduce or change this relationship. Thus modifiers are of considerable importance in informing treatment decisions. Let us consider the hypothetical example simulated in this article. Longitudinal analysis of the data demonstrates that the therapeutic intervention is effective in changing the outcome. This relationship assumes that the treatment effect is homogenous across the patient population. However, in most circumstances, there is a complex interplay between patient characteristics, treatment and outcome. Further examination of putative modifiers show that this existing relationship is differential by age. Specifically, the treatment effect on outcome is significant only for older subjects 60 – 79 years of age. In our simulated study, the treatment target should be a subset of older individuals.

The nonpositional syntax in LSMESTIMATE statement provides a simplified framework for exploration of effect modification with multiple levels of modifiers. To our knowledge, in SAS® 9.4, this nonpositional approach can be used for 12 procedures and is also compatible with the ESTIMATE statement in SAS® versions 9.2 and thereafter. The possibility of exploring modifiers as a continuous variable is beyond the scope of this paper and has not been discussed.

CONCLUSION

Custom testing effect modification with multiple levels of modifiers can be performed using simplified code in the nonpositional syntax of LSMESTIMATE statement in SAS®. This approach can be applied in repeated measures RCT studies for quick exploration of variables that may alter the treatment outcome relationship.

REFERENCES

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Pyschological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 1173-1182.

Fairchild, A. J., & MacKinnon, D. P. (2010). A general model for testing mediation and moderation effects. *Prevention Science*, 87-99.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, 115-134.

Kiernan, K., Tobias, R., Gibbs, P., & Tao, J. (2011). CONTRAST and ESTIMATE Statements Made Easy: the LSMESTIMATE Statement. *SAS Global Forum* (p. 351). Las Vegas: SAS Institute.

Kraemer, H. C., Wilson, T., Fairburn, C., & Agras, S. (2002). Mediators and Moderators of Treatment Effcets in Randomized Clinical Trials. *JAMA Psychiatry*, 877-883.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Pronabesh DasMahapatra
PatientsLikeMe Inc.
(617) 499-4003
pdasmahapatra@patientslikeme.com