

Hands-on Tutorial for Piecewise Linear Mixed-effects Models Using SAS® PROC MIXED

Qinlei Huang, St Jude Children's Research Hospital, Memphis, TN

ABSTRACT

Clinical trials and public health studies are focused on studying changes associated with interventions, events, or critical periods in human development. Evaluating the effect of critical or high-risk events/periods in longitudinal studies of growth may provide clues to the long-term effects of life events and efficacies of preventive/therapeutic interventions. Conventional linear longitudinal models typically involve a single growth profile to represent linear changes in an outcome variable across time, which sometimes does not fit the empirical data. The piecewise linear mixed-effects models allow different linear functions of time corresponding to the pre- and post-critical time point trends. This step-by-step, hands-on tutorial first introduces the reader to how to perform piecewise linear mixed effects models by using SAS PROC MIXED in the context of a clinical trial with 2 intervention arms and a predictive covariate-of-interest. Second, it illustrates how to obtain the slopes and corresponding p-values of intervention and control groups during pre- and post-critical periods, conditional on different values of the predictive covariate. Third, it explains how to make meaningful comparisons and present results in a scientific manuscript. Illustrative SAS commands are provided to fit piecewise linear mixed-effects models and to generate the summary tables, assisting in the interpretation of the results.

Keywords: Longitudinal data, Piecewise Linear Mixed-effects Model, PROC MIXED

INTRODUCTION

Evaluating the effect of critical or high-risk periods in longitudinal studies of growth pay provides clues to the optimal timing of preventive or therapeutic intervention. This paper presents a hands-on tutorial to fit piecewise linear mixed-effects models by using PROC MIXED. It first introduces a step-by-step procedure to perform piecewise linear mixed-effects models using SAS PROC MIXED, in the context of a clinical trial with two-arm interventions and a predictive covariate of interest. Next, it illustrates how to obtain the slopes and corresponding p-values for intervention and control groups during pre- and post-critical periods, conditional on different values of the predictive covariate. Last, it explains how to make meaningful comparisons and present results in a scientific manuscript.

LINEAR MIXED-EFFECTS MODELS

Linear mixed-effects models are an extension of general linear models. Both model types work with continuous response variables and model the linear relationships between responses and explanatory variables. The general linear model assumes independent and identically distributed normal random errors. The linear mixed-effects model includes additional random-effects parameters and allows for a more flexible covariance matrix of the random errors. It accommodates both correlated errors and errors with heterogeneous variances. The mixed model is written as

$$y = X\beta + Z\gamma + \varepsilon$$

Where y denotes the vector of *observed* outcomes; X is the known matrix of fixed effects; β is the unknown vector of fixed-effects parameters; Z is the known matrix of random effects; γ is the unknown vector of random-effects parameters; and ε is the unobserved vector of random errors. Both γ and ε are normally distributed with

$$E \begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{var} \begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

The variance of y is, therefore, $V = ZGZ' + R$. Because of its flexibility, the linear mixed-effects model is a very useful tool for longitudinal and hierarchical data.

CRITICAL PERIOD AND LINEAR SPLINES

Many randomized clinical trials and follow-up observational studies show that the patterns of change in the mean response over the duration of the longitudinal study is not simply linear. One solution is to introduce higher-order polynomials in time. However, the parameters are difficult to interpret. As an alternative, linear spline models are a very useful and flexible way to accommodate many of the non-linear trends. For example, intensive changes in the mean response often occur in a short period, called "critical period", immediately after an intervention or a critical event (Figures 1, 2). The pre-event changes and long-term post-treatment effects are often relatively minimal. By

dividing the time axis into 2 or more segments and fitting a linear model in each of the segments, linear spline models sufficiently accommodate many of the non-linear trends.

The break point, also called knot, is either decided by theory-driven hypothesis or data-driven graphical representations. Once the knot is set, a time spline variable should be created based on time and knot to fit the piecewise linear mixed-effects model (Fitzmaurice et al., 2004).

$$time\ spline = \begin{cases} 0 & \text{if } t \leq k \\ t - k & \text{if } t > k \end{cases}$$

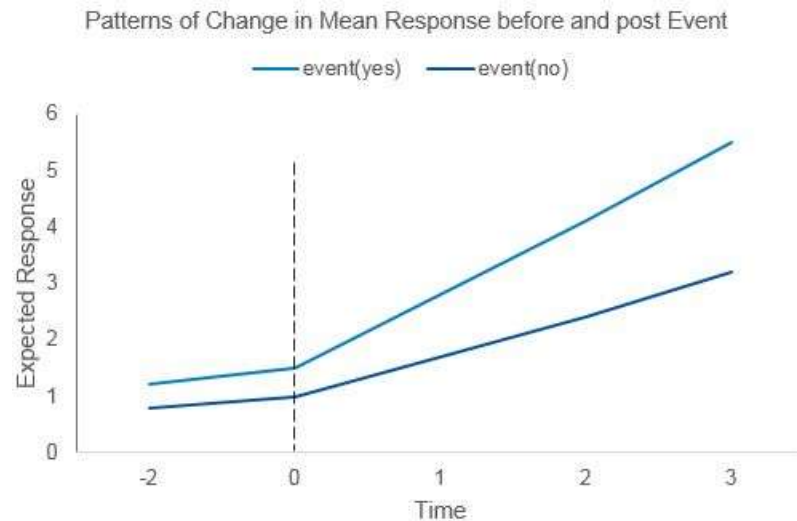


Figure 1. Graphical representation of model with linear splines for observational studies, with knot at time = 0

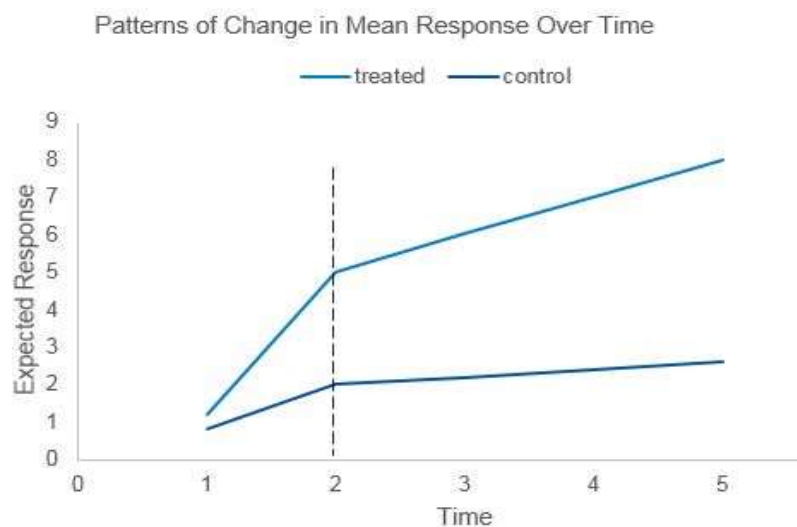


Figure 2. Graphical representation of model with linear splines for clinical trials, with knot at time = 2

STRATEGIES FOR MODELING AND PARAMETER INTERPRETATIONS

STRATEGIES FOR MODELING: TIME, TREATMENT, AND COVARIATE EFFECTS

To build a meaningful piecewise linear mixed-effects model, I first fit a line with a spline at time k to explore the mean response profile over time (i.e., the pure time effect). I next separately fit 2 lines with splines at time k for the treatment and control groups to investigate the effect of treatment on the mean response profiles over time. I then checked the additive effect of potential confounders or important characteristics (e.g., age, SES, time since diagnosis) on the mean response profiles over time. Lastly, I probed whether the treatment effect on mean response profiles differed by covariate levels (i.e., the interactive treatment and covariate effect). For each step, the difference

covariance structures and random effects are tested. Model fit statistics AIC (Akaike Information Criterion) and BIC (Sawa Bayesian Information Criterion) are compared to inform model selection. The piecewise linear mixed-effects models can be written as follows, with fixed effects specified.

Step 1: time effect before and after critical time point k

$$y = \beta_0 + \beta_1 \text{time} + \beta_2 \text{time spline} + Z\gamma + \varepsilon$$

Step 2: treatment effect at baseline and before and after time point k

$$y = \beta_0 + \beta_1 \text{time} + \beta_2 \text{time spline} \\ + \beta_3 \text{trt} + \beta_4 \text{trt} * \text{time} + \beta_5 \text{trt} * \text{time spline} + Z\gamma + \varepsilon$$

Step 3: additive treatment and covariate effect at baseline and before and after time point k

$$y = \beta_0 + \beta_1 \text{time} + \beta_2 \text{time spline} \\ + \beta_3 \text{trt} + \beta_4 \text{trt} * \text{time} + \beta_5 \text{trt} * \text{time spline} \\ + \beta_6 \text{cov} + \beta_7 \text{cov} * \text{time} + \beta_8 \text{cov} * \text{time spline} + Z\gamma + \varepsilon$$

Step 4: interactive treatment and covariate effect at baseline and before and after time point k

$$y = \beta_0 + \beta_1 \text{time} + \beta_2 \text{time spline} \\ + \beta_3 \text{trt} + \beta_4 \text{trt} * \text{time} + \beta_5 \text{trt} * \text{time spline} \\ + \beta_6 \text{cov} + \beta_7 \text{cov} * \text{time} + \beta_8 \text{cov} * \text{time spline} \\ + \beta_9 \text{trt} * \text{cov} + \beta_{10} \text{trt} * \text{cov} * \text{time} + \beta_{11} \text{trt} * \text{cov} * \text{time spline} + Z\gamma + \varepsilon$$

	Model 1	Model 2	Model 3	Model 4
Time Effect				
Intercept	β_0	β_0	β_0	β_0
Time	β_1	β_1	β_1	β_1
Time spline	β_2	β_2	β_2	β_2
Treatment Effect				
Treatment		β_3	β_3	β_3
Treatment * time		β_4	β_4	β_4
Treatment * time spline		β_5	β_5	β_5
Additive Covariate Effect				
Covariate			β_6	β_6
Covariate * time			β_7	β_7
Covariate * time spline			β_8	β_8
Interactive Treatment and Covariate Effect				
Treatment * Covariate				β_9
Treatment * Covariate * time				β_{10}
Treatment * Covariate * time spline				β_{11}

Table 1. Piecewise Linear Mixed-effects Models: the Time, Treatment, and Covariate Effects

Please note that the random effect matrix Z and covariance matrices G and R should be specified and tested by using PROC MIXED REPEATED and RANDOM statements in each step of your choice. If the between-subject variability is large, then random effects, especially those of time and time spline, should be tested. If within-subject correlations dominate, then matrix R should be defined and compared. If both between-subject variability and within-subject correlation exist, then random effects and covariance matrices of G and R should be tested simultaneously.

PARAMETER INTERPRETATION

After a set of model-building steps, you will get the parameter estimates and corresponding p-values. When many novice SAS users first read the PROC MIXED outputs, they wonder what these numbers mean and how these estimates can help them answer their questions. More often than not, the collaborative researchers feel frustrated

when they are given the original results. The model itself is complicated, and in fact, further efforts are needed to provide meaningful parameter estimates and interpretations.

Table 2 lists the parameter estimates (1) to (12) to be calculated based on the fixed-effects coefficient estimates in the raw PROC MIXED output. For example, parameter (2) equals $\beta_0 + \beta_3$, based on model 2; and parameter (8) equals $\beta_1 + \beta_2 + \beta_4 + \beta_5 + (\beta_7 + \beta_8)c$, based on model 4.

	Parameter	Model 1	Model 2	Model 3	Model 4
Baseline					
Control group	(1)	β_0	β_0	$\beta_0 + \beta_6c$	$\beta_0 + \beta_6c$
Treatment group	(2)		$\beta_0 + \beta_3$	$\beta_0 + \beta_3 + \beta_6c$	$\beta_0 + \beta_3 + (\beta_6 + \beta_9)c$
Δ (trt – control)	(3)		β_3	β_3	$\beta_3 + \beta_9c$
Slope before k					
Control group	(4)	β_1	β_1	$\beta_1 + \beta_7c$	$\beta_1 + \beta_7c$
Treatment group	(5)		$\beta_1 + \beta_4$	$\beta_1 + \beta_4 + \beta_7c$	$\beta_1 + \beta_4 + (\beta_7 + \beta_{10})c$
Δ (trt – control)	(6)		β_4	β_4	$\beta_4 + \beta_{10}c$
Slope after k					
Control group	(7)	$\beta_1 + \beta_2$	$\beta_1 + \beta_2$	$\beta_1 + \beta_2 + (\beta_7 + \beta_8)c$	$\beta_1 + \beta_2 + (\beta_7 + \beta_8)c$
Treatment group	(8)		$\beta_1 + \beta_2 + \beta_4 + \beta_5$	$\beta_1 + \beta_2 + \beta_4 + \beta_5 + (\beta_7 + \beta_8)c$	$\beta_1 + \beta_2 + \beta_4 + \beta_5 + (\beta_7 + \beta_8)c$
Δ (trt – control)	(9)		$\beta_4 + \beta_5$	$\beta_4 + \beta_5$	$\beta_4 + \beta_5 + (\beta_{10} + \beta_{11})c$
Δ slope_(after k-before k)					
Control group	(10)	β_2	β_2	$\beta_2 + \beta_8c$	$\beta_2 + \beta_8c$
Treatment group	(11)		$\beta_2 + \beta_5$	$\beta_2 + \beta_5 + \beta_8c$	$\beta_2 + \beta_5 + (\beta_8 + \beta_9)c$
Δ (trt – control)	(12)		β_5	β_5	$\beta_5 + \beta_9c$

Table 2. Piecewise Linear Mixed-effects Models: Parameter Interpretations

Because most clinical trials and pharmaceutical studies focus on treatment effects over time, let's start with the second model. To decide whether an intervention is effective and whether the effectiveness is transient or permanent, the following questions are of interest: 1) Do the treatment and control groups distribute similarly enough at baseline? 2) Before the critical point k, does response change significantly in each group, and do these patterns of change differ by group? 3) After the critical point k, does the pattern stay the same or change in some way, and do these patterns of change differ by group?

In the context of a longitudinal randomization study, the second model is sufficient to answer these questions if no confounding factor exists. For instance, parameter (3) evaluates the difference in response between groups at baseline and, thus, answers the first question. Parameters (4), (5), and (6) assess and compare the changes before k for each group and, therefore, answer the second question. Parameters (7), (8), and (9) measure and compare the patterns after k for each group and, hence, address the third question. Parameters (10), (11), and (12) reveal the difference in the patterns of change before and after k for each group and make a comparison across groups. It is recommended that you reorganize the raw results and provide the calculated parameters (1) to (12) to help your audience interpret the results.

PIECEWISE LINEAR MIXED-EFFECTS MODEL: FIT AND INTERPRET

This section explains how to fit the piecewise linear mixed-effects model by using PROC MIXED.

EXAMPLE DATA

These examples are based on an artificial dataset called `example`. It is created to be the longitudinal data of a two-arm randomization clinical trial. Subjects are randomly assigned to the treatment or control groups. Response y is collected 5 times (at baseline and 1, 2, 3, and 4 years after baseline) for each subject. The subject's age at baseline is recorded. Dataset `example` has 5 variables: `id`, `y`, `trt`, `time`, and `age`.

- `id` is the identification variable of each subject
- `y` is a continuous response variable
- `trt` is a binary variable of treatment groups

- `time` represents the 5 time points
- `age` is the subject's age at baseline

Table 3 shows the first 10 observations of the dataset.

Id	y	trt	time	age
1	5.4	0	1	2.8
1	4.9	0	2	2.8
1	5.0	0	3	2.8
1	4.3	0	4	2.8
1	4.7	0	5	2.8
2	5.0	1	1	5.6
2	5.0	1	2	5.6
2	3.8	1	3	5.6
2	4.0	1	4	5.6
2	5.1	1	5	5.6

Table 3. Dataset for Fitting Piecewise Linear Mixed-effects Model (first 10 observations)

GRAPHICAL REPRESENTATION

For longitudinal data, a spaghetti plot is very useful. It provides a straightforward graphical representation of the response profiles over time. The red line represents the mean response over time. The grey lines depict the change in response for each individual. The spaghetti plot shows an obvious decrease in `y` from time 1 to 2 for both groups. Changes after time 2 are relatively negligible for both groups and an increase from time 4 to time 5 is seen in the treatment group. The graph implies that time 2 is a break point for both groups.

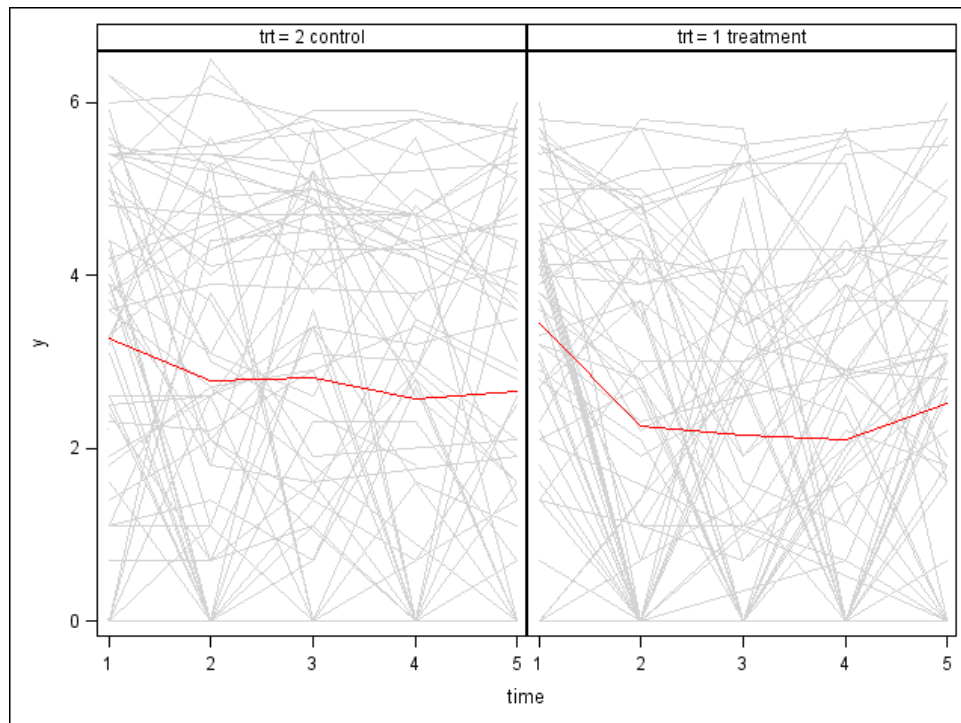


Figure 3. Spaghetti Plot of Response Profiles over Time by Group

Here, I calculate the mean response using PROC SQL and generate the spaghetti plots by group using PROC SGPANEL.

```
proc sql;
  create table example as
```

```

select * , mean(y) as mean_y format 8.2
  from example
  group by trt, time;
quit;

proc sgpanel data = example ;
  panelby trt / columns = 2;
  series x = time y = y / group = id
          lineattrs = (pattern = 1 color = "lightgrey");
  colaxis integer;
  series x = time y = mean_y / lineattrs = (pattern = 1 color = "red");
  format trt trtfmt.;
run;
quit;

```

SPLINE VARIABLE

I next created a spline variable using the following SAS codes. `time spline` equals 0 when `time` is less than or equal to `k`. `time spline` equals `(time - k)` given `time` is greater than `k`

```

data example;
  set example;
  k = 2;
  if time <= k then timespl1 = 0;
  if time > k then timespl1 = time - k;
run;

```

I then fit the 4 proposed linear mixed-effects models with splines and compared different covariance structures for each model.

MODEL 1: TIME EFFECT

The first model examines the pure time effect on response `y`. It has 3 fixed-effects parameters to estimate: intercept, time, and time spline. The model is written as

$$y = \begin{cases} \beta_0 + \beta_1 t + Z\gamma + \varepsilon & \text{if } t \leq k \\ (\beta_0 - \beta_2 k) + (\beta_1 + \beta_2)t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

I first used the following SAS codes to fit the linear mixed-effects model with splines at time 2. The REPEATED statement defines a repeated time effect and an unstructured covariance R matrix. Please note that a new variable `t`, which equals the time variable `time`, was created for the repeated effect. `t` is treated as a classification variable for the repeated effect, and `time` is assumed to be continuous in this model definition. The default estimation method REML was used for the covariance parameters. The Kenward Roger method was used to compute the degrees-of-freedom for the tests of fixed effects.

```

proc mixed data = example order = formatted ;
  class id t ;
  model y = time timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;
run ;

```

The solution for fixed effects shows that both `time` and `timespl1` have p-values less than 0.05. A significant `time` effect indicates that the change before `k` is significant. The `time spline` effect tells the difference between the slope before `k` and the slope after `k`. As explained in the previous section, you need to calculate one more parameter for model interpretation: the slope after `k`.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	4.2021	0.2758	129	15.23	<.0001

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	4.2021	0.2758	129	15.23	<.0001
time	-0.8804	0.1662	126	-5.30	<.0001
timespl1	0.9024	0.1923	127	4.69	<.0001

Table 4. PROC MIXED Output for Model 1: Solution for Fixed Effects

I used the ESTIMATE statement of PROC MIXED to obtain the parameter estimate and corresponding p-value. The ESTIMATE statement provides a table called `estimates`.

```
proc mixed data = example order = formatted ;
  class id t ;
  model y = time timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;
  estimate "before k" time 1 timespl1 0 ;
  estimate "after k" time 1 timespl1 1 ;
run ;
```

The `estimates` table shows that the estimate for slope before k is the same as the estimate for `time` effect (-0.8804) and that the estimate for slope after k is the sum of parameter estimates for `time` and `timespl1` (-0.8804+0.9024=0.022).

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
before k	-0.8804	0.1662	126	-5.30	<.0001
after k	0.02207	0.05073	118	0.44	0.6643

Table 5. PROC MIXED Output for Model 1: Estimates

Please note that the current model assumes that within-subject serial correlation is the primary source of variability in response. The unstructured R matrix allows for heterogeneous variance and covariance parameters (i.e., $[5 \times 6]/2=15$).

To obtain the model best representing the data, different structures of the error covariance matrix should be tested and compared. For example, models with random effects may be considered to examine the between-subject variability. Covariance structures with more constraints could be tested for parsimonious purposes. Because of space limits, I only show one example here. Please refer to other SAS papers for discussions of model comparisons.

```
proc mixed data = example order = data ;
  class id t ;
  model y = time timespl1 / solution ddfm = kr ;
  random int time timespl1 / type = vc sub = id ;
run ;
```

The SAS commands specify random effects of `time` and `time spline` by using the RANDOM statement and specify an independent G matrix by using the TYPE= option. You can try different specifications according to your preferences. It is recommended that you use the default REML estimation method to compare the random part of your models. For nested models, a likelihood ratio test can be established. For non-nested models, model fit statistics AIC and BIC should be compared.

MODEL 2: TREATMENT EFFECT

The second model investigates whether there is any treatment effect on the response profiles over time. The model for the control group is written as:

$$y = \begin{cases} \beta_0 + \beta_1 t + Z\gamma + \varepsilon & \text{if } t \leq k \\ (\beta_0 - \beta_2 k) + (\beta_1 + \beta_2)t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

The model for the treatment group is written as:

$$y = \begin{cases} (\beta_0 + \beta_3) + (\beta_1 + \beta_4)t + Z\gamma + \varepsilon & \text{if } t \leq k \\ (\beta_0 + \beta_3) - (\beta_2 + \beta_5)k + (\beta_1 + \beta_2 + \beta_4 + \beta_5)t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

There are 6 parameters to estimate. The intercept and main treatment estimate give you the baseline information of response for the control and treatment groups. The time and time spline estimates provide information for the change in response before and after k. The interaction terms of treatment with time and time spline assess the difference in response profiles of the treatment and control groups. I used the following SAS codes to fit the linear mixed-effects model with spline at time 2 for both groups.

```
proc mixed data = example order = formatted ;
  class trt id t ;
  model y = time timespl1
         trt trt*time trt*timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;
run ;
```

The PROC MIXED output provides the parameter estimates and corresponding p-values for the fixed-effects in Table 6. As explained in the previous section, more efforts are needed to generate parameters for model interpretation.

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		3.6607	0.3898	128	9.39	<.0001
time		-0.4401	0.2317	125	-1.90	0.0598
timespl1		0.3783	0.2666	125	1.42	0.1583
trt	1 treatment	1.0576	0.5452	128	1.94	0.0546
trt	2 control					
time*trt	1 treatment	-0.8563	0.3241	125	-2.64	0.0093
time*trt	2 control					
timespl1*trt	1 treatment	1.0116	0.3741	126	2.70	0.0078
timespl1*trt	2 control					

Table 6. PROC MIXED Output for Model 2: Solution for Fixed Effects

The ESTIMATE statement is used to obtain the estimates and corresponding p-values of these parameters:

- Mean response at baseline: control, treatment, and treatment - control
- Mean response changes before k: control, treatment, and treatment - control
- Mean response changes after k: control, treatment, and treatment - control
- Differences between slopes before and after k: control, treatment, and treatment - control

```
proc mixed data = example order = formatted ;
  class trt id t ;
  model y = time timespl1
         trt trt*time trt*timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;

  estimate "Intercept: control"    intercept 1 trt 0 1 ;
  estimate "Intercept: treatment"  intercept 1 trt 1 0 ;
  estimate "Intercept: trt - con"  intercept 0 trt 1 -1 ;
  estimate "Slope(time) before k: control "
        time 1 timespl1 0 trt*time 0 1 trt*timespl1 0 0 ;
  estimate "Slope(time) before k: treatment"
        time 1 timespl1 0 trt*time 1 0 trt*timespl1 0 0 ;
  estimate "Slope(time) before k: trt - con"
        time 0 timespl1 0 trt*time 1 -1 trt*timespl1 0 0 ;
  estimate "Slope(time) after k: control"
        time 1 timespl1 1 trt*time 0 1 trt*timespl1 0 1 ;
  estimate "Slope(time) after k: treatment"
        time 1 timespl1 1 trt*time 1 0 trt*timespl1 1 0 ;
```

```

estimate "Slope(time) after k: trt - con"
      time 0 timespl1 0 trt*time 1 -1 trt*timespl1 1 -1 ;
estimate "Slope(time) change k: control"
      time 0 timespl1 1 trt*time 0 0 trt*timespl1 0 1 ;
estimate "Slope(time) change k: treatment"
      time 0 timespl1 1 trt*time 0 0 trt*timespl1 1 0 ;
estimate "Slope(time) change k: trt - con"
      time 0 timespl1 0 trt*time 0 0 trt*timespl1 1 -1 ;

format trt trtfmt. ;
run ;

```

PROC MIXED provided the `estimates` table below in the output. This format is a better presentation for your audience to understand what the model means. For example, the estimates table shows that there is no significant difference in response at baseline between the 2 groups ($\beta = -1.06$, $p > 0.5$). The reductions in response before k are significant for the treatment group ($\beta = -1.30$, $p < .001$) and the difference between the treatment and control group is significant ($\beta = -.86$, $p < .01$).

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
Intercept: control	3.6607	0.3898	128	9.39	<.0001
Intercept: treatment	4.7183	0.3811	128	12.38	<.0001
Intercept: trt - con	1.0576	0.5452	128	1.94	0.0546
Slope(time) before k: control	-0.4401	0.2317	125	-1.90	0.0598
Slope(time) before k: treatment	-1.2964	0.2267	126	-5.72	<.0001
Slope(time) before k: trt - con	-0.8563	0.3241	125	-2.64	0.0093
Slope(time) after k: control	-0.06173	0.07115	115	-0.87	0.3874
Slope(time) after k: treatment	0.09357	0.07123	119	1.31	0.1915
Slope(time) after k: trt - con	0.1553	0.1007	117	1.54	0.1256
Slope(time) change k: control	0.3783	0.2666	125	1.42	0.1583
Slope(time) change k: treatment	1.3900	0.2625	127	5.30	<.0001
Slope(time) change k: trt - con	1.0116	0.3741	126	2.70	0.0078

Table 7. PROC MIXED Output for Model 2: Estimates

MODEL 3: ADDITIVE TREATMENT AND AGE EFFECT

The third model examines the additive effects of treatment and age on the response profiles over time. The model for the control group is written as:

$$y = \begin{cases} (\beta_0 + \beta_6 c) + (\beta_1 + \beta_7 c)t + Z\gamma + \varepsilon & \text{if } t \leq k \\ (\beta_0 + \beta_6 c) - (\beta_2 + \beta_8 c)k + [\beta_1 + \beta_2 + (\beta_7 + \beta_8)c]t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

The model for the treatment group is written as:

$$y = \begin{cases} (\beta_0 + \beta_3 + \beta_6 c) + (\beta_1 + \beta_4 + \beta_7 c)t + Z\gamma + \varepsilon & \text{if } t \leq k \\ (\beta_0 + \beta_3 + \beta_6 c) - (\beta_2 + \beta_5 + \beta_8 c)k + [\beta_1 + \beta_2 + \beta_4 + \beta_5 + (\beta_7 + \beta_8)c]t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

I first used the following SAS codes to fit the linear mixed-effects model with spline at time 2

```

proc mixed data = example order = formatted ;
  class trt id t ;
  model y = time timespl1
        trt trt*time trt*timespl1
        age age*time age*timespl1 / solution ddfm = kr ;

```

```
repeated t / type = un sub = id ;
```

The PROC MIXED output provides the parameter estimates and corresponding p-values for the fixed-effects in Table 8.

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		4.2972	0.5938	127	7.24	<.0001
time		-0.9928	0.3476	125	-2.86	0.0050
timespl1		1.1406	0.3967	126	2.88	0.0047
trt	1 treatment	1.1083	0.5432	127	2.04	0.0434
trt	2 control	0	.	.	.	.
time*trt	1 treatment	-0.9069	0.3200	124	-2.83	0.0054
time*trt	2 control	0	.	.	.	.
timespl1*trt	1 treatment	1.0812	0.3666	125	2.95	0.0038
timespl1*trt	2 control	0	.	.	.	.
age		-0.07714	0.05325	127	-1.45	0.1499
time*age		0.06728	0.03134	124	2.15	0.0337
timespl1*age		-0.09281	0.03580	125	-2.59	0.0107

Table 8. PROC MIXED Output for Model 3: Solution for Fixed Effects

Because age is a continuous covariate, you cannot get a single response profile over time for all age values. For model interpretation, it is good practice to pick a few critical values of the continuous covariate and estimate the response profiles based on these values. In the example, I picked age 5, 10, and 15 to represent very young, school-aged, and teenaged children, respectively.

I used the following SAS codes to obtain parameter estimates and corresponding p-values for age 5. You may revise the codes easily to get the estimates for ages 10 and 15.

```
proc mixed data = example order = formatted ;
  class trt id t ;
  model y = time timespl1
          trt trt*time trt*timespl1
          age age*time age*timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;

  estimate "Intercept (age 5): control"    intercept 1 age 5 trt 0 1 ;
  estimate "Intercept (age 5): treatment"  intercept 1 age 5 trt 1 0 ;
  estimate "Intercept (age 5): trt - con"  intercept 0 age 0 trt 1 -1 ;
  estimate "Slope(time) before k (age 5): control"
          time 1 timespl1 0
          trt*time 0 1 trt*timespl1 0 0
          age*time 5 age*timespl1 0 ;
  estimate "Slope(time) before k (age 5): treated"
          time 1 timespl1 0
          trt*time 1 0 trt*timespl1 0 0
          age*time 5 age*timespl1 0 ;
  estimate "Slope(time) before k (age 5): trt-con"
          time 0 timespl1 0
          trt*time 1 -1 trt*timespl1 0 0
          age*time 0 age*timespl1 0 ;
  estimate "Slope(time) after k (age 5): control"
          time 1 timespl1 1
          trt*time 0 1 trt*timespl1 0 1
          age*time 5 age*timespl1 5 ;
  estimate "Slope(time) after k (age 5): treated"
```

```

time 1 timespl1 1
  trt*time 1 0 trt*timespl1 1 0
  age*time 5 age*timespl1 5 ;
estimate "Slope(time) after k (age 5): trt-con"
time 0 timespl1 0
  trt*time 1 -1 trt*timespl1 1 -1
  age*time 0 age*timespl1 0 ;
estimate "Slope(time) change k (age 5): control"
time 0 timespl1 1
  trt*time 0 0 trt*timespl1 0 1
  age*time 0 age*timespl1 5 ;
estimate "Slope(time) change k (age 5): treated"
time 0 timespl1 1 trt*time 0 0
  trt*timespl1 1 0
  age*time 0 age*timespl1 5 ;
estimate "Slope(time) change k (age 5): trt-con"
time 0 timespl1 0
  trt*time 0 0 trt*timespl1 1 -1
  age*time 0 age*timespl1 0 ;
format trt trtfmt. ;
run ;

```

The estimates table provides the following estimates and p-values based on age 5:

- Mean response at baseline: control, treatment, and treatment - control
- Mean response changes before k: control, treatment, and treatment - control
- Mean response changes after k: control, treatment, and treatment - control
- Differences between slopes before and after k: control, treatment, and treatment - control

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
Intercept (age 5): control	3.9115	0.4293	127	9.11	<.0001
Intercept (age 5): treatment	5.0198	0.4300	128	11.67	<.0001
Intercept (age 5): trt - con	1.1083	0.5432	127	2.04	0.0434
Slope(time) before k (age 5): control	-0.6564	0.2516	124	-2.61	0.0102
Slope(time) before k (age 5): treated	-1.5632	0.2547	124	-6.14	<.0001
Slope(time) before k (age 5): trt-con	-0.9069	0.3200	124	-2.83	0.0054
Slope(time) after k (age 5): control	0.02013	0.07638	115	0.26	0.7926
Slope(time) after k (age 5): treated	0.1945	0.07857	116	2.48	0.0147
Slope(time) after k (age 5): trt-con	0.1744	0.09838	116	1.77	0.0790
Slope(time) change k (age 5): control	0.6765	0.2871	125	2.36	0.0200
Slope(time) change k (age 5): treated	1.7578	0.2926	124	6.01	<.0001
Slope(time) change k (age 5): trt-con	1.0812	0.3666	125	2.95	0.0038

Table 9. PROC MIXED Output for Model 3: Estimates

MODEL 4: INTERACTIVE TREATMENT AND AGE EFFECT

The fourth model studies the interactive effects of treatment and age on the response profiles over time. The model for the control group is written as:

$$y = \begin{cases} (\beta_0 + \beta_6 c) + (\beta_1 + \beta_7 c)t + Z\gamma + \varepsilon & \text{if } t \leq k \\ (\beta_0 + \beta_6 c) - (\beta_2 + \beta_8 c)k + [\beta_1 + \beta_2 + (\beta_7 + \beta_8)c]t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

The model for the treatment group is written as:

$$y = \begin{cases} [\beta_0 + \beta_3 + (\beta_6 + \beta_9)c] & \text{if } t \leq k \\ +[\beta_1 + \beta_4 + (\beta_7 + \beta_{10})c]t + Z\gamma + \varepsilon & \\ [\beta_0 + \beta_3 + (\beta_6 + \beta_9)c] - [\beta_2 + \beta_5 + (\beta_8 + \beta_{11})c]k & \\ +[\beta_1 + \beta_2 + \beta_4 + \beta_5 + (\beta_7 + \beta_8 + \beta_{10} + \beta_{11})c]t + Z\gamma + \varepsilon & \text{if } t > k \end{cases}$$

I first used the following SAS codes to fit the linear mixed-effects model with spline at time 2

```
proc mixed data = example order = formatted ;
  class id trt t ;
  model y = time timespl1
         trt trt*time trt*timespl1
         age age*time age*timespl1
         trt*age trt*age*time trt*age*timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;
run ;
```

The PROC MIXED output provides the parameter estimates and corresponding p-values for the fixed-effects in Table 10.

Solution for Fixed Effects						
Effect	trt	Estimate	Standard Error	DF	t Value	Pr > t
Intercept		4.2064	0.7586	126	5.54	<.0001
time		-1.0851	0.4446	124	-2.44	0.0161
timespl1		1.2399	0.5070	124	2.45	0.0159
trt	1 treatment	1.2854	1.0710	126	1.20	0.2323
trt	2 control	0	.	.	.	.
time*trt	1 treatment	-0.7260	0.6308	123	-1.15	0.2520
time*trt	2 control	0	.	.	.	.
timespl1*trt	1 treatment	0.8862	0.7215	123	1.23	0.2217
timespl1*trt	2 control	0	.	.	.	.
age		-0.06663	0.07712	127	-0.86	0.3892
time*age		0.07858	0.04556	123	1.72	0.0870
timespl1*age		-0.1050	0.05201	123	-2.02	0.0458
age*trt	1 treatment	-0.02068	0.1070	126	-0.19	0.8471
age*trt	2 control	0	.	.	.	.
time*age*trt	1 treatment	-0.02106	0.06299	124	-0.33	0.7387
time*age*trt	2 control	0	.	.	.	.
timespl1*age*trt	1 treatment	0.02260	0.07197	124	0.31	0.7540
timespl1*age*trt	2 control	0	.	.	.	.

Table 10. PROC MIXED Output for Model 4: Solution for Fixed Effects

I next run PROC MIXED 3 times to obtain parameter estimates and corresponding p-values for ages 5, 10, and 15. Here, I present the codes for age 15.

```
proc mixed data = example order = formatted ;
  class id trt t ;
  model y = time timespl1
         trt trt*time trt*timespl1
         age age*time age*timespl1
         trt*age trt*age*time trt*age*timespl1 / solution ddfm = kr ;
  repeated t / type = un sub = id ;

  estimate "Intercept (age15): control"
           intercept 1 age 15 trt 0 1 trt*age 0 15 ;
  estimate "Intercept (age15): treatment"
```

```

        intercept 1 age 15 trt 1 0   trt*age 15 0 ;
estimate "Intercept (age15): trt - con"
        intercept 0 age 0 trt 1 -1 trt*age 15 -15 ;
estimate "Slope(time) before k (age15): control"
        time 1      timespl1 0
          trt*time 0 1      trt*timespl1 0 0
          age*time 15 age*timespl1 0
          trt*age*time 0 15 trt*age*timespl1 0 0 ;
estimate "Slope(time) before k (age15): treated"
        time 1      timespl1 0
          trt*time 1 0      trt*timespl1 0 0
          age*time 15 age*timespl1 0
          trt*age*time 15 0 trt*age*timespl1 0 0 ;
estimate "Slope(time) before k (age15): trt-con"
        time 0      timespl1 0
          trt*time 1 -1      trt*timespl1 0 0
          age*time 0 age*timespl1 0
          trt*age*time 15 -15 trt*age*timespl1 0 0 ;
estimate "Slope(time) after k (age15): control"
        time 1      timespl1 1
          trt*time 0 1      trt*timespl1 0 1
          age*time 15 age*timespl1 15
          trt*age*time 0 15 trt*age*timespl1 0 15 ;
estimate "Slope(time) after k (age15): treated"
        time 1      timespl1 1
          trt*time 1 0      trt*timespl1 1 0
          age*time 15 age*timespl1 15
          trt*age*time 15 0 trt*age*timespl1 15 0 ;
estimate "Slope(time) after k (age15): trt-con"
        time 0      timespl1 0
          trt*time 1 -1      trt*timespl1 1 -1
          age*time 0 age*timespl1 0
          trt*age*time 15 -15 trt*age*timespl1 15 -15 ;
estimate "Slope(time) change k (age15): control"
        time 0      timespl1 1
          trt*time 0 0      trt*timespl1 0 1
          age*time 0 age*timespl1 15
          trt*age*time 0 0 trt*age*timespl1 0 15 ;
estimate "Slope(time) change k (age15): treated"
        time 0      timespl1 1
          trt*time 0 0      trt*timespl1 1 0
          age*time 0 age*timespl1 15
          trt*age*time 0 0 trt*age*timespl1 15 0 ;
estimate "Slope(time) change k (age15): trt-con"
        time 0      timespl1 0
          trt*time 0 0      trt*timespl1 1 -1
          age*time 0 age*timespl1 0
          trt*age*time 0 0 trt*age*timespl1 15 -15 ;
format trt trtfmt. ;
run ;

```

The estimates table provides the following estimates and p-values based on age 15.

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr > t
Intercept (age15): control	3.2069	0.6388	127	5.02	<.0001

Intercept (age15): treatment	4.1822	0.5972	125	7.00	<.0001
Intercept (age15): trt - con	0.9753	0.8744	126	1.12	0.2668
Slope(time) before k (age15): control	0.09366	0.3797	122	0.25	0.8056
Slope(time) before k (age15): treated	-0.9482	0.3474	124	-2.73	0.0073
Slope(time) before k (age15): trt-con	-1.0419	0.5146	123	-2.02	0.0451
Slope(time) after k (age15): control	-0.2407	0.1154	111	-2.09	0.0392
Slope(time) after k (age15): treated	-0.05732	0.1081	118	-0.53	0.5971
Slope(time) after k (age15): trt-con	0.1834	0.1581	114	1.16	0.2486
Slope(time) change k (age15): control	-0.3343	0.4340	121	-0.77	0.4426
Slope(time) change k (age15): treated	0.8909	0.3974	128	2.24	0.0267
Slope(time) change k (age15): trt-con	1.2253	0.5884	124	2.08	0.0394

Table 11. PROC MIXED Output for Model 4: Estimates

FINAL MODEL

The fixed-effect estimates of model 4 indicate that the age and treatment effects are not interactive (no significant interaction terms). Also, the fit statistics AIC and BIC of model 3 (AIC: 2396.2, BIC: 2439.8) show a better fit to the data than do those of model 4 (AIC: 2409.1, BIC: 2452.7). Using the same strategy, I compared model 3 vs. model 2 and model 3 vs. model 1. All fit statistics indicate that model 3 has the best AIC and BIC. Also, fixed-effect estimates of model 3 are meaningful.

Using the strategy I discuss in the section for model 1, I tested different covariance structures for model 3 and concluded that the repeated time effect with an unstructured R matrix delivers the best data fit. Therefore, model 3 is the final model. To present the results in an interpretable way, I refit model 3 and used the ESTIMATE statement to get parameter estimates based on ages 10 and 15. The table below is organized on the basis of the *estimates* table for model interpretation.

	Age 5			Age 10			Age 15		
Label	Est.	SE	p	Est.	SE	p	Est.	SE	p
Baseline									
Control group	3.9115	0.4293	<.0001	3.5258	0.3970	<.0001	3.1401	0.5223	<.0001
Treatment group	5.0198	0.4300	<.0001	4.6341	0.3847	<.0001	4.2484	0.5029	<.0001
Δ (trt – control)	1.1083	0.5432	0.0434	1.1083	0.5432	0.0434	1.1083	0.5432	0.0434
Slope before k									
Control group	-0.6564	0.2516	0.0102	-0.3200	0.2342	0.1743	0.01640	0.3090	0.9578
Treatment group	-1.5632	0.2547	<.0001	-1.2269	0.2260	<.0001	-0.8905	0.2940	0.0030
Δ (trt – control)	-0.9069	0.3200	0.0054	-0.9069	0.3200	0.0054	-0.9069	0.3200	0.0054
Slope after k									
Control group	0.02013	0.07638	0.7926	-0.1076	0.07123	0.1338	-0.2352	0.09419	0.0139
Treatment group	0.1945	0.07857	0.0147	0.06683	0.07044	0.3447	-0.06085	0.09116	0.5057
Δ (trt – control)	0.1744	0.09838	0.0790	0.1744	0.09838	0.0790	0.1744	0.09838	0.0790
Δ slope_(after k-before k)									
Control group	0.6765	0.2871	0.0200	0.2124	0.2675	0.4286	-0.2516	0.3533	0.4776
Treatment group	1.7578	0.2926	<.0001	1.2937	0.2596	<.0001	0.8296	0.3365	0.0150
Δ (trt – control)	1.0812	0.3666	0.0038	1.0812	0.3666	0.0038	1.0812	0.3666	0.0038

Table 12. Parameter Estimates for Final Model

CONCLUSION

This paper offers a hands-on tutorial for piecewise linear mixed-effects modeling using SAS PROC MIXED. It presents the underlying concepts of linear mixed-effects models with splines and illustrates how to build and interpret models step by step. Four types of models and examples are discussed: pure time effect models, treatment effect models, additive treatment and covariate effects models, and interactive treatment and covariate effects models.

REFERENCES

Bozdogan H. Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions. *Psychometrika*. 1987;52(3):345-370.

Fitzmaurice GM, Laird N, and Ware JH. *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley & Sons; 2004.

ACKNOWLEDGMENTS

My sincere gratitude goes to Dr. James Boyett and the Department of Biostatistics at St. Jude Children's Research Hospital for providing an exceptionally supportive working environment and great learning resources; to Dr. Guolian Kang for her continuous support of my work; and to Dr. Kumar Srivastava and experienced biostatisticians Catherine Billups, Yinmei Zhou, Wei Liu, and Deqing Pei for their valuable comments on my work.

RECOMMENDED READING

- Longitudinal Data Analysis with Discrete and Continuous Responses Course Notes
- SAS® Certification Prep Guide: Advanced Programming for SAS® 9

CONTACT INFORMATION

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

Name: Qinlei Huang
Enterprise: St. Jude Children's Research Hospital
Address: 262 Danny Thomas Place
City, State ZIP: Memphis, TN 38105
Work Phone: 901.595.2027
Fax: 901.595.4585
E-mail: Qinlei.huang@stjude.org
Web: www.stjude.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.