BIO 226: APPLIED LONGITUDINAL ANALYSIS

LECTURE 8

Modelling the Mean: Parametric Curves

Modelling the Mean: Parametric Curves

Fitting parametric or semi-parametric curves to longitudinal data can be justified on substantive and statistical grounds.

Substantively, in many studies true underlying mean response process changes over time in a relatively smooth, monotonically increasing/decreasing pattern.

Fitting parsimonious models for mean response results in statistical tests of covariate effects (e.g., treatment $\times$ time interactions) with greater power than in analysis of response profiles.

"Constant Effect" Model

In previous lecture, we discussed a simple model where an exposure or treatment might cause a shift in the mean response that remains constant across measurement occasions

To fit such a model, we can create a new variable for time:

$\text{Posttime}_{ij} = 0$ if baseline ($\text{Time}_{ij} = 0$),
$\text{Posttime}_{ij} = 1$ if post-baseline ($\text{Time}_{ij} > 0$).

Then, in two group setting, the model is:

$$E(Y_{ij}) = \beta_1 + \beta_2 \text{Group}_i + \beta_3 \text{Posttime}_{ij} + \beta_4 \text{Posttime}_{ij} \times \text{Group}_i.$$

This model tests whether the differences between the group means, averaged over the $(n-1)$ post-baseline measurement occasions, are significantly different from the corresponding differences at baseline.

That is, the hypothesis of no group effect on longitudinal change corresponds to the test of no group by post-baseline interaction.

In general, this test has $(G-1)$ d.f., where G is the number of groups.

Polynomial Trends in Time

Describe the patterns of change in the mean response over time in terms of simple polynomial trends.

The means are modelled as an explicit function of time.

This approach can handle highly unbalanced designs in a relatively seamless way.

For example, mistimed measurements are easily incorporated in the model for the mean response.

Linear Trends over Time

Simplest possible curve for describing changes in the mean response over time is a straight line.

Slope has direct interpretation in terms of a constant rate of change in mean response for a single unit change in time.

Consider two-group study comparing *treatment* and *control*, where changes in mean response are approximately linear:

$$E\left(Y_{ij}\right) = \beta_1 + \beta_2 \text{Time}_{ij} + \beta_3 \text{Group}_i + \beta_4 \text{Time}_{ij} \times \text{Group}_i,$$

where $\text{Group}_i = 1$ if i^{th} individual assigned to treatment, and $\text{Group}_i = 0$ otherwise; and Time_{ij} denotes measurement time for the j^{th} measurement on i^{th} individual.

Model for the mean for subjects in control group:

$$E\left(Y_{ij}\right) = \beta_1 + \beta_2 \text{Time}_{ij},$$

while for subjects in treatment group,

$$E\left(Y_{ij}\right) = \left(\beta_1 + \beta_3\right) + \left(\beta_2 + \beta_4\right) \text{Time}_{ij}.$$

Thus, each group's mean response is assumed to change linearly over time (see Figure 1).

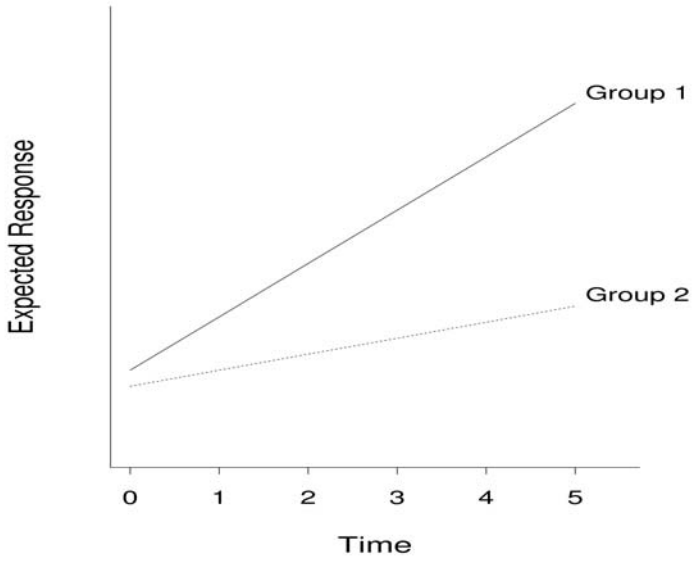

Figure 1: Graphical representation of model with linear trends for two groups.

Quadratic Trends over Time

When changes in the mean response over time are not linear, higher-order polynomial trends can be considered.

For example, if the means are monotonically increasing or decreasing over the course of the study, but in a curvilinear way, a model with quadratic trends can be considered.

In a quadratic trend model the rate of change in the mean response is not constant but depends on time.

Rate of change must be expressed in terms of two parameters.

Consider two-group study example:

$$\begin{aligned} E(Y_{ij}) &= \beta_1 + \beta_2 \text{Time}_{ij} + \beta_3 \text{Time}_{ij}^2 + \beta_4 \text{Group}_i \\ &\quad + \beta_5 \text{Time}_{ij} \times \text{Group}_i + \beta_6 \text{Time}_{ij}^2 \times \text{Group}_i. \end{aligned}$$

Model for subjects in control group:

$$E(Y_{ij}) = \beta_1 + \beta_2 \text{Time}_{ij} + \beta_3 \text{Time}_{ij}^2;$$

while model for subjects in treatment group:

$$E(Y_{ij}) = (\beta_1 + \beta_4) + (\beta_2 + \beta_5) \text{Time}_{ij} + (\beta_3 + \beta_6) \text{Time}_{ij}^2.$$

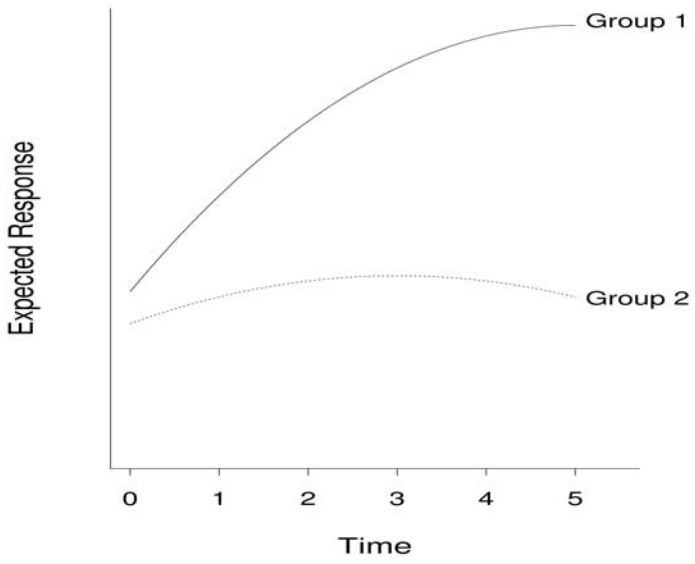

Figure 2: Graphical representation of model with quadratic trends for two groups.

Note: mean response changes at different rate, depending upon Time_{ij}.

Rate of change in control group is $\beta_2 + 2\beta_3\text{Time}_{ij}$
(derivation of this instantaneous rate of change straightforward with calculus).

Thus, early in the study when $\text{Time}_{ij} = 1$, rate of change is $\beta_2 + 2\beta_3$; while later in the study, say $\text{Time}_{ij} = 4$, rate of change is $\beta_2 + 8\beta_3$.

Regression coefficients, $(\beta_2 + \beta_5)$ and $(\beta_3 + \beta_6)$, have similar interpretations for treatment group.

"Centering"

To avoid problems of collinearity it is advisable to "center" Time_j on its mean value prior to the analysis.

Replace Time_j by its deviation from the mean of (Time_1,Time_2,...,Time_n).

Note: centering of Time_{ij} at individual-specific values (e.g., the mean of the n_i measurement times for i^{th} individual) should be avoided, as the interpretation of the intercept becomes meaningless.

Linear Splines

If simplest curve is a straight line, then one way to extend the curve is to have sequence of joined line segments that produces a piecewise linear pattern.

Linear spline models provide flexible way to accommodate many non-linear trends that cannot be approximated by simple polynomials in time.

Basic idea: Divide time axis into series of segments and consider piecewise-linear trends, having different slopes but joined at fixed times.

Locations where lines are tied together are known as "knots".

Resulting piecewise-linear curve is called a spline.

Piecewise-linear model often called "broken-stick" model.

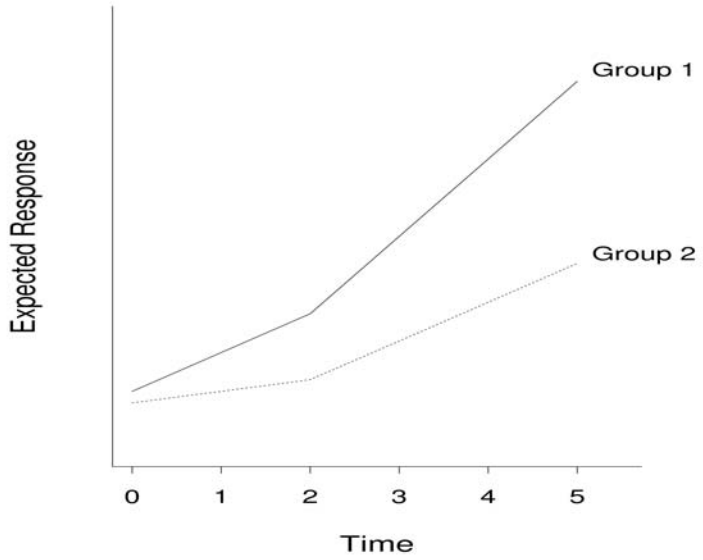

Figure 3: Graphical representation of model with linear splines for two groups, with common knot.

The simplest possible spline model has only one knot.

For two-group example, linear spline model with knot at t^*:

$$\begin{aligned} E\left(Y_{ij}\right) &= \beta_1 + \beta_2 \text{Time}_{ij} + \beta_3 (\text{Time}_{ij} - t^*)_+ + \beta_4 \text{Group}_i \\ &\quad + \beta_5 \text{Time}_{ij} \times \text{Group}_i + \beta_6 (\text{Time}_{ij} - t^*)_+ \times \text{Group}_i, \end{aligned}$$

where $(x)_+$ is defined as a function that equals x when x is positive and is equal to zero otherwise.

Thus, $(\text{Time}_{ij} - t^*)_+$ is equal to $(\text{Time}_{ij} - t^*)$ when $\text{Time}_{ij} > t^*$ and is equal to zero when $\text{Time}_{ij} \leq t^*$.

Model for subjects in control group:

$$E\left(Y_{ij}\right) = \beta_1 + \beta_2 \text{Time}_{ij} + \beta_3 (\text{Time}_{ij} - t^*)_+.$$

When expressed in terms of mean response prior/after t^*:

$$\begin{aligned} E\left(Y_{ij}\right) &= \beta_1 + \beta_2 \text{Time}_{ij}, && \text{Time}_{ij} \leq t^*; \\ E\left(Y_{ij}\right) &= (\beta_1 - \beta_3 t^*) + (\beta_2 + \beta_3)\text{Time}_{ij}, && \text{Time}_{ij} > t^*. \end{aligned}$$

Slope prior to t^* is β_2 and following t^* is $(\beta_2 + \beta_3)$.

Model for subjects in treatment group:

$$E\left(Y_{ij}\right) = (\beta_1 + \beta_4) + (\beta_2 + \beta_5)\text{Time}_{ij} + (\beta_3 + \beta_6)(\text{Time}_{ij} - t^*)_+.$$

When expressed in terms of mean response prior/after t^*:

$$\begin{aligned} E\left(Y_{ij}\right) &= (\beta_1 + \beta_4) + (\beta_2 + \beta_5)\text{Time}_{ij}, && \text{Time}_{ij} \leq t^*; \\ E\left(Y_{ij}\right) &= [(\beta_1 + \beta_4) - (\beta_3 + \beta_6)t^*)] \\ &\quad + (\beta_2 + \beta_3 + \beta_5 + \beta_6)\text{Time}_{ij}, && \text{Time}_{ij} > t^*. \end{aligned}$$

Case Study 1: Vlagtwedde-Vlaardingen Study

Epidemiologic study on prevalence of and risk factors for chronic obstructive lung disease.

Sample participated in follow-up surveys approximately every 3 years for up to 21 years.

Pulmonary function was determined by spirometry: FEV_1.

We focus on a subset of 133 residents aged 36 or older at their entry into the study and whose smoking status did not change over the 19 years of follow-up.

Each study participant was either a current or former smoker.

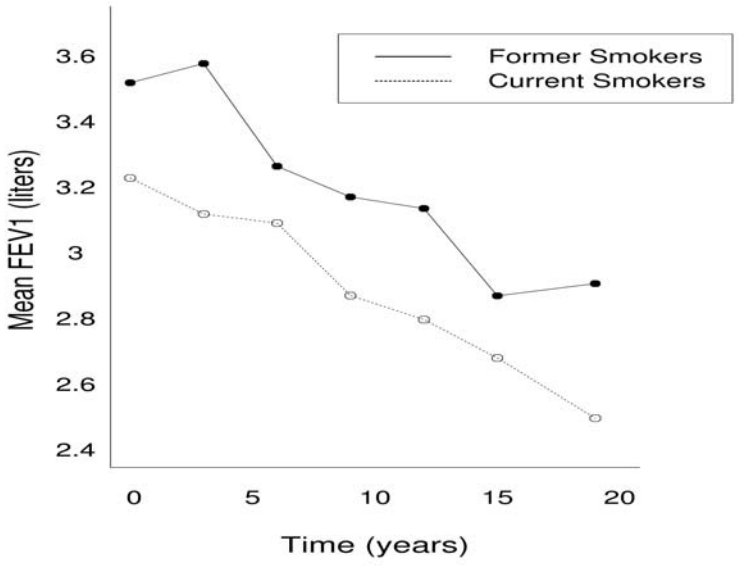

Figure 4: Mean FEV_1 at baseline (year 0), year 3, year 6, year 9, year 12, year 15, and year 19 in the current and former smoking exposure groups.

First we consider a linear trend in the mean response over time, with intercepts and slopes that differ for the two smoking exposure groups.

We assume an unstructured covariance matrix.

Based on the REML estimates of the regression coefficients in Table 1, the mean response for former smokers is

$$E\left(Y_{ij}\right) = 3.507 - 0.033\ \texttt{Time}_{\texttt{ij}},$$

while for current smokers,

$$\begin{aligned} E\left(Y_{ij}\right) &= (3.507 - 0.262) - (0.033 + 0.005)\ \texttt{Time}_{\texttt{ij}} \\ &= 3.245 - 0.038\ \texttt{Time}_{\texttt{ij}}. \end{aligned}$$

Table 1: Estimated regression coefficients for linear trend model for FEV_1 data from the Vlagtwedde-Vlaardingen study.

Variable	Smoking Group	Estimate	SE	Z
Intercept		3.5073	0.1004	34.94
$\texttt{Smoke}_{\texttt{i}}$	Current	−0.2617	0.1151	−2.27
$\texttt{Time}_{\texttt{ij}}$		−0.0332	0.0031	−10.84
$\texttt{Smoke}_{\texttt{i}} \times \texttt{Time}_{\texttt{ij}}$	Current	−0.0050	0.0035	−1.42

Thus, both groups have a significant decline in mean FEV_1 over time.

But there is no discernible difference between the two smoking exposure groups in the constant rate of change.

That is, the $\texttt{Smoke}_\texttt{i} \times \texttt{Time}_\texttt{ij}$ interaction (i.e., the comparison of the two slopes) is not significant, with $Z = -1.42$, $p > 0.15$.

But is the rate of change constant over time?

Adequacy of linear trend model can be assessed by including higher-order polynomial trends.

For example, we can consider a model that allows quadratic trends for changes in FEV_1 over time.

Recall that linear trend model is nested within the quadratic trend model.

The maximized log-likelihoods for the models with linear and quadratic trends are presented in Table 2.

LRT test statistic can be compared to a chi-squared distribution with 2 degrees of freedom (or 6, the number of parameters in the quadratic trend model, minus 4, the number of parameters in the linear trend model).

Note: Likelihood ratio test is based on the ML, not REML, log-likelihood.

Table 2: Maximized (ML) log-likelihoods for models with linear and quadratic trends for FEV_1 data from the Vlagtwedde-Vlaardingen study.

Model	-2 (ML) Log-Likelihood
Quadratic Trend Model	237.2
Linear Trend Model	238.5
$-2 \times$ Log-Likelihood Ratio: $G^2 = 1.3$, 2 df $(p > 0.50)$	

LRT comparing quadratic and linear trend models, produces $G^2 = 1.3$, with 2 degrees of freedom ($p > 0.50$).

Thus, when compared to quadratic trend model, linear trend model appears to be adequate.

Finally, for illustrative purposes, we can make a comparison with a cubic trend model.

This produces LRT statistic, $G^2 = 4.4$, with 4 degrees of freedom ($p > 0.35$), indicating again that the linear trend model is adequate.

Case Study 2: Treatment of Lead-Exposed Children Trial

Recall data from TLC trial:

Children randomized to placebo or Succimer.

Measures of blood lead level at baseline, 1, 4 and 6 weeks.

The sequence of means over time in each group is displayed in Figure 5.

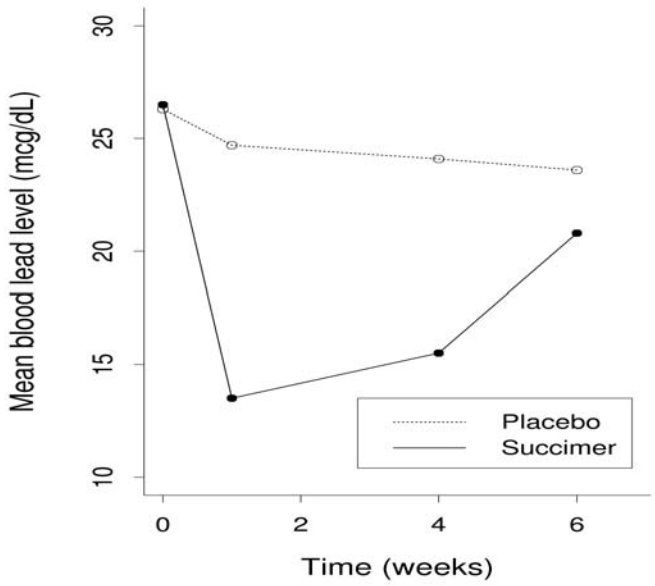

Figure 5: Mean blood lead levels at baseline, week 1, week 4, and week 6 in the succimer and placebo groups.

Given that there are non-linearities in the trends over time, higher-order polynomial models (e.g., a quadratic trend model) could be fit to the data.

Alternatively, we can accommodate the non-linearity with a piecewise linear model with common knot at week 1,

$$\begin{aligned} E\left(Y_{ij}\right) &= \beta_1 + \beta_2 \, \texttt{Week}_{\texttt{ij}} + \beta_3 \left(\texttt{Week}_{\texttt{ij}} - 1\right)_+ + \beta_4 \, \texttt{Group}_{\texttt{i}} \times \texttt{Week}_{\texttt{ij}} \\ &\quad + \beta_5 \, \texttt{Group}_{\texttt{i}} \times \left(\texttt{Week}_{\texttt{ij}} - 1\right)_+, \end{aligned}$$

where $\texttt{Group}_{\texttt{i}} = 1$ if assigned to succimer, and $\texttt{Group}_{\texttt{i}} = 0$ otherwise.

Because of randomization, model does not contain a main effect of `Group`.

That is, we assume a common mean blood lead level at baseline.

In this piecewise linear model, means for subjects in placebo group are

$$E\left(Y_{ij}\right) = \beta_1 + \beta_2 \, \texttt{Week}_{\texttt{ij}} + \beta_3 \left(\texttt{Week}_{\texttt{ij}} - 1\right)_+,$$

while in the succimer group

$$E\left(Y_{ij}\right) = \beta_1 + \left(\beta_2 + \beta_4\right) \texttt{Week}_{\texttt{ij}} + \left(\beta_3 + \beta_5\right) \left(\texttt{Week}_{\texttt{ij}} - 1\right)_+.$$

Table 3: Estimated regression coefficients and standard errors based on a piecewise linear model, with knot at week 1.

Variable	Group	Estimate	SE	Z
Intercept		26.3422	0.4991	52.78
$\texttt{Week}_{\texttt{ij}}$		−1.6296	0.7818	−2.08
$(\texttt{Week}_{\texttt{ij}} - \texttt{1})_+$		1.4305	0.8777	1.63
$\texttt{Group} \times \texttt{Week}_{\texttt{ij}}$	A	−11.2500	1.0924	−10.30
$\texttt{Group} \times (\texttt{Week}_{\texttt{ij}} - \texttt{1})_+$	A	12.5822	1.2278	10.25

When expressed in terms of mean response prior to/after week 1, estimated means in the placebo group are

$$\widehat{\mu}_{ij} = \widehat{\beta}_1 + \widehat{\beta}_2 \, \texttt{Week}_{\texttt{ij}}, \qquad \texttt{Week}_{\texttt{ij}} \leq 1;$$

$$\widehat{\mu}_{ij} = (\widehat{\beta}_1 - \widehat{\beta}_3) + (\widehat{\beta}_2 + \widehat{\beta}_3) \, \texttt{Week}_{\texttt{ij}}, \qquad \texttt{Week}_{\texttt{ij}} > \texttt{1}.$$

Thus, in the placebo group, slope prior to week 1 is $\widehat{\beta}_2 = -1.63$ and following week 1 is $(\widehat{\beta}_2 + \widehat{\beta}_3) = -1.63 + 1.43 = -0.20$.

Similarly, when expressed in terms of the mean response prior to and after week 1, the estimated means for subjects in the succimer group are given by

$$\begin{aligned}
\widehat{\mu}_{ij} &= \widehat{\beta}_1 + (\widehat{\beta}_2 + \widehat{\beta}_4)\,\texttt{Week}_{\texttt{ij}}, && \texttt{Week}_{\texttt{ij}} \leq 1; \\
\widehat{\mu}_{ij} &= \widehat{\beta}_1 - (\widehat{\beta}_3 + \widehat{\beta}_5) \\
&\quad + (\widehat{\beta}_2 + \widehat{\beta}_3 + \widehat{\beta}_4 + \widehat{\beta}_5)\,\texttt{Week}_{\texttt{ij}}, && \texttt{Week}_{\texttt{ij}} > 1.
\end{aligned}$$

The estimates of the mean blood lead levels for the placebo and succimer groups are presented in Table 4.

The estimated means from the piecewise linear model appear to adequately fit the observed mean response profiles for the two treatment groups.

Note that piecewise linear and quadratic trend models (with common intercept for two groups) are not nested.

They both have the same number of parameters and therefore their respective log-likelihoods can be directly compared.

The maximized log-likelihoods indicate that piecewise linear model fits these data better than quadratic trend model (-2 ML log-likelihood $= 2436.2$ for piecewise linear model versus -2 ML log-likelihood $= 2551.7$ for quadratic trend model).

Table 4: Estimated mean blood lead levels for placebo and succimer groups from linear spline model (knot at week 1). Observed means in parentheses.

Group	Week 0	Week 1	Week 4	Week 6
Succimer	26.3 (26.5)	13.5 (13.5)	16.7 (15.5)	19.1 (20.8)
Placebo	26.3 (26.3)	24.7 (24.7)	24.1 (24.1)	23.7 (23.2)

Parametric Curves using PROC MIXED in SAS

Table 5: Illustrative commands for a linear trend model using PROC MIXED in SAS.

```
PROC MIXED;
  CLASS id group t;
  MODEL y=group time group*time / SOLUTION CHISQ;
  REPEATED t / TYPE=UN SUBJECT=id R RCORR;
```

Note that the CLASS statement includes a variable t. This variable is an additional copy of the variable *time*.

The difference is that while t is declared as a categorical variable on the CLASS statement, *time* is not and is treated as a quantitative covariate in the MODEL statement.

It is good practice to include, wherever possible, a REPEATED effect.

This ensures covariance is estimated correctly when the design is balanced but incomplete due to missingness or when repeated measures are not in same order for each subject in data set.

Table 6: Illustrative commands for a quadratic trend model using PROC MIXED in SAS.

```
PROC MIXED;
  CLASS id group t;
  MODEL y=group time timesqr group*time group*timesqr /S CHISQ;
  REPEATED t / TYPE=UN SUBJECT=id R RCORR;
```

Table 7: Illustrative commands for a spline model, with knot at $time = 4$, using PROC MIXED in SAS.

```
PROC MIXED;
  CLASS id group t;
  MODEL y=group time time_4 group*time group*time_4 /S CHISQ;
  REPEATED t / TYPE=UN SUBJECT=id R RCORR;
```

The MODEL statement includes $time$ and $time\_4$, where $time\_4$ is a derived variable for $(time - 4)_+$.

The latter variable can easily be computed in SAS as

$$time\_4 = max(time - 4, 0);$$