10.3 Covariates

Mixed models handle a wide range of problems. In particular, it is straightforward to compare performance of different groups and to include covariates to assess the how background characteristics possibly influence change over time.

To demonstrate, consider again the National Longitudinal Study of Youth data. As part of that project an experimental reading program, See Me Soar, was administered to some children in the sample. One objective was to assess the effectiveness of the treatment. A 30% random sample of records are shown in Figure 3. In this view it is not obvious whether the reading program has helped.

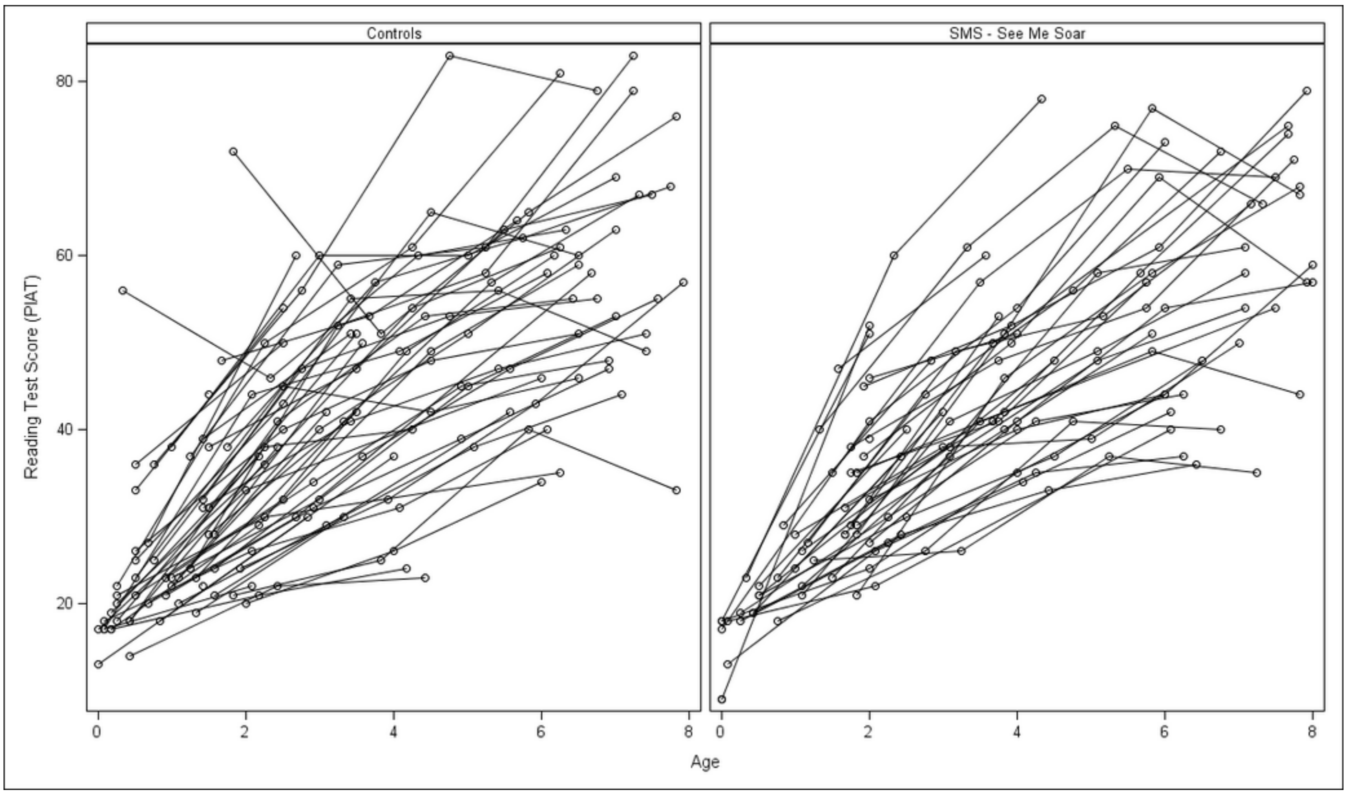

Figure 3. A 30% random sample of repeated measures on PIAT Reading from the NLSY study. Data for those who participated in the reading program See Me Soar are shown in the right-hand panel.

Consequently an important issue is deciding whether the treatment affects reading over time. I also want to check for differences between girls and boys. In previous analyses with these data, there was some evidence that change in Reading is not simply linear but rather that performance is more rapid during some time periods than others. To handle all these issues, set up a nonlinear model with covariates. It has several inter-connected components and incorporates information about three variables. On this page is the most elaborate form.

- As we did earlier, shift the time variable so that the intercept is interpretable as individual Reading performance at age 6. Define

$$Age_{ij} = ActualAge_{ij} - 6$$

where $ActualAge_{ij}$ is calendar age in years for child i at occasion $j = 1, ..., n_i$ and $1 \le n_i \le 4$.

- The two covariates are

$$\begin{array}{ll} T_i & = (1/0) \text{ if a child participated in the reading program or not} \\ boy_i & = (1/0) \text{ for (boy/girl)} \end{array}$$

- The mixed model is quadratic in Age

$$Read_{ij} = \beta_{i0} + \beta_{i1} Age_{ij} + \beta_{i2} Age_{ij}^2 + e_{ij},\ j = 1, ..., n_i$$

Because there are individual differences on each of the three coefficients, the particular way that Reading changes can be markedly different for some children than others.

- We want to know whether the covariates affect initial performance (β_{i0}) or rate of change (β_{i1} and β_{i2}). Thus individual coefficients are the sum of fixed effects plus covariates plus random effects

$$\begin{array}{rcl} \beta_{i0} & = & \beta_0 + \gamma_1 T_i + \gamma_4 boy_i + b_{i0} \\ \beta_{i1} & = & \beta_1 + \gamma_2 T_i + \gamma_5 boy_i + b_{i1} \\ \beta_{i2} & = & \beta_2 + \gamma_3 T_i + \gamma_6 boy_i + b_{i2} \end{array}$$

- The covariance matrix of random effects is a full (3 x 3) matrix,

$$\mathbf{\Phi} = cov(b_{i0}, b_{i1}, b_{i2}) = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix}$$

- Assume that residuals around a child's individual trajectory at any occasion are uncorrelated with residuals at other occasions but that variances at each occasion are the same. This is the sphericity structure. For a case with $n_i = 4$ measurements, for example,

$$\mathbf{\Lambda}_i = cov(e_{i1}, e_{i2}, e_{i3}, e_{i4}) = \begin{pmatrix} \sigma_e^2 & & & \\ 0 & \sigma_e^2 & & \\ 0 & 0 & \sigma_e^2 & \\ 0 & 0 & 0 & \sigma_e^2 \end{pmatrix}$$

The i subscript indicates that this matrix varies person to person, although in this case $\mathbf{\Lambda}_i$ varies only in size. It is a scalar for all who have $n_i = 1$, but has more rows and columns for children with more measurements.

The general mixed model is

$$\mathbf{y}_i = \mathbf{X}_i \, \boldsymbol{\beta} + \mathbf{Z}_i \, \mathbf{b}_i + \mathbf{e}_i$$

For this problem there are nine parameters associated with fixed effects

$$(\beta_0, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6)$$

plus three random effects

$$(b_{i0}, b_{i1}, b_{i2})$$

The setup for a child with a full complement of $n_i = 4$ measurements and all covariates is below. Note that the first three variables of $\mathbf{X}_i$ are the same as those of $\mathbf{Z}_i$

$$\begin{pmatrix} Read_{i1} \\ Read_{i2} \\ Read_{i3} \\ Read_{i4} \end{pmatrix} = \begin{pmatrix} 1 & Age_{i1} & Age_{i1}^2 & T_i & T_i Age_{i1} & T_i Age_{i1}^2 & boy_i & boy_i Age_{i1} & boy_i Age_{i1}^2 \\ 1 & Age_{i2} & Age_{i2}^2 & T_i & T_i Age_{i2} & T_i Age_{i2}^2 & boy_i & boy_i Age_{i2} & boy_i Age_{i2}^2 \\ 1 & Age_{i3} & Age_{i3}^2 & T_i & T_i Age_{i3} & T_i Age_{i3}^2 & boy_i & boy_i Age_{i3} & boy_i Age_{i3}^2 \\ 1 & Age_{i4} & Age_{i4}^2 & T_i & T_i Age_{14} & T_i Age_{i4}^2 & boy_i & boy_i Age_{i4} & boy_i Age_{i4}^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & Age_{i1} & Age_{i1}^2 \\ 1 & Age_{i2} & Age_{i2}^2 \\ 1 & Age_{i3} & Age_{i3}^2 \\ 1 & Age_{i4} & Age_{i4}^2 \end{pmatrix} \begin{pmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \end{pmatrix}$$

There are two main strategies for fitting a complicated regression model: Start with a simple model and add variables, or begin with the most complicated model and eliminate terms that are not needed. Working forward from simple to complicated, or going backward from complicated to simple, does not always arrive at the same place. My personal preference is to start small.

Model A: Three random effects, no covariates (10 parms, dev = 9060, AIC = 9078, BIC = 9114)

The goal is to obtain information as to whether the full quadratic model is needed and whether random effects on each element of β_{ij} are useful.

$$Read_{ij} = \beta_{i0} + \beta_{i1} Age_{ij} + \beta_{i2} Age_{ij}^2 + e_{ij}$$

$$\begin{aligned} \beta_{i0} &= \beta_0 + b_{i0} \\ \beta_{i1} &= \beta_1 + b_{i1} \\ \beta_{i2} &= \beta_2 + b_{i2} \end{aligned} \qquad \mathbf{\Phi} = \begin{pmatrix} \varphi_{00} & & \\ \varphi_{10} & \varphi_{11} & \\ \varphi_{20} & \varphi_{21} & \varphi_{22} \end{pmatrix} \qquad \mathbf{\Lambda}_i = \sigma_e^2 \mathbf{I}_{n_i}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 17.4(.37) \\ 9.30(.28) \\ \text{-}0.494(.03) \end{pmatrix} \qquad \hat{\mathbf{\Phi}} = \begin{pmatrix} 0(na) & & \\ 11.0 & 7.00(2.2) & \\ -1.23 & -0.494 & 0.041(.03) \end{pmatrix} \qquad \hat{\sigma}_e^2 = 25.9(1.4)$$

About the fixed effects, $\hat{\beta}_2 = -0.494$ implies that the quadratic term is needed for the mean function and therefore the trajectory is nonlinear and decelerating. But for the random effects, $\hat{\varphi}_{00} \approx 0$ (which surprised me but probably should not), and also $\hat{\varphi}_{22} = 0.041$ is not significant.

Model B: Two random effects, no covariates (7 parms, dev = 9093, AIC = 9107, BIC = 9135)

Because $var(b_{i0}) \approx 0$, eliminate this random effect from the intercept and also eliminate the first row and column of $\mathbf{\Phi}$. Model B has two random effects on coefficients β_{i1} and β_{i2}. The intercept is a fixed parameter without individual differences.

The idea is that at age six, all children perform at the same level on the reading test. As long as φ_{11} and φ_{22} are large, the children vary in reading performance in terms of their rates of change. Based on Figure 3, it seems reasonable to suppose there is essentially no variability at age six, but lots of variability in rates of change.

$$Read_{ij} = \beta_{i0} + \beta_{i1} Age_{ij} + \beta_{i2} Age_{ij}^2 + e_{ij}$$

$$\begin{array}{l} \beta_{i0} = \beta_0 \\ \beta_{i1} = \beta_1 + b_{i1} \\ \beta_{i2} = \beta_2 + b_{i2} \end{array} \qquad \mathbf{\Phi} = \begin{pmatrix} \varphi_{11} & \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \qquad \mathbf{\Lambda}_i = \sigma_e^2 \mathbf{I}_{n_i}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 17.4(.38) \\ 9.31(.33) \\ \text{-}0.497(.04) \end{pmatrix} \qquad \hat{\mathbf{\Phi}} = \begin{pmatrix} 19.0(1.9) & \\ -1.82 & 0.187(.03) \end{pmatrix} \qquad \hat{\sigma}_e^2 = 26.2(1.4)$$

When reviewing this issue regarding removing random effects in the previous section, the recommended procedure is to test the fit of Model A, the full model, against Model B, the reduced model, by means of the likelihood ratio test. In this instance, we have

$$\begin{aligned} T &= D_R - D_F \\ &= 9093 - 9060 = 33 \end{aligned}$$

which is χ^2 with degrees of freedom equal to the difference in number of parameters

$$\begin{aligned} df &= q_F - q_R \\ &= 10 - 7 = 3 \end{aligned}$$

The critical point at $\alpha = 0.05$ is also χ^2 with $df = 3$: $\chi^2_{df=3} = 7.81$.

Rats! Model B fits significantly worse than Model A. However Model A is unacceptable because $var(b_{i0}) = 0$, so there is no reason to keep b_{i0}.

Here is one of those situations in which, without question, we conclude that Model B is the better of the two because Model A is inadmissible. In the choice between a better fitting model and an interpretable model, always select the latter.

Model C: Two random effects, boy/girl covariate (10 parms, dev = 9091, AIC = 9111, BIC = 9152)

Now include boy_i as a covariate to answer the question of whether there are sex differences, and if so, how they affect the learning process. The intercept does not need random effects; however the specification, $\beta_{i0} = \beta_0 + \gamma_1 boy_i$, allows for a difference in initial mean level between boys and girls.

$$Read_{ij} = \beta_{i0} + \beta_{i1} Age_{ij} + \beta_{i2} Age_{ij}^2 + e_{ij}$$

$$\begin{array}{l} \beta_{i0} = \beta_0 + \gamma_1 boy_i \\ \beta_{i1} = \beta_1 + \gamma_2 boy_i + b_{i1} \\ \beta_{i2} = \beta_2 + \gamma_3 boy_i + b_{i2} \end{array} \qquad \mathbf{\Phi} = \begin{pmatrix} \varphi_{11} & \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \qquad \mathbf{\Lambda}_i = \sigma_e^2 \mathbf{I}_{n_i}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 17.3(.54) \\ 9.65(.46) \\ \text{-}0.540(.05) \end{pmatrix} \qquad \hat{\boldsymbol{\gamma}} = \begin{pmatrix} 0.135(.75) \\ \text{-}0.691(.65) \\ 0.089(.08) \end{pmatrix} \qquad \hat{\mathbf{\Phi}} = \begin{pmatrix} 18.9(1.9) & \\ -1.81 & 0.185(.03) \end{pmatrix} \qquad \hat{\sigma}_e^2 = 26.2(1.4)$$

None of the parameter estimates associated with boy_i are significant. Both AIC and BIC are larger for this model compared to Model B. Consequently conclude no detectable boy/girl differences.

Model D: Two random effects, treatment covariate (9 parms, dev = 9085, AIC = 9103, BIC = 9139)

Children in this sample were randomly assigned to the experimental reading program, See Me Soar. Does the program improve reading over time and if so, how does it work? The covariate T_i is a treatment indicator. We dearly hope there are no intercept differences on T_i because if there are it implies that the randomization process failed. If there is an effect for T_i we want to see it as an improvement in rate over time. Here is the setup

$$Read_{ij} = \beta_{i0} + \beta_{i1}Age_{ij} + \beta_{i2}Age_{ij}^2 + e_{ij}$$

$$\begin{aligned}\beta_{i0} &= \beta_0 + \gamma_1 T_i\\ \beta_{i1} &= \beta_1 + \gamma_2 T_i + b_{i1}\\ \beta_{i2} &= \beta_2 + b_{i2}\end{aligned} \qquad \mathbf{\Phi} = \begin{pmatrix} \varphi_{11} & \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \qquad \mathbf{\Lambda}_i = \sigma_e^2 \mathbf{I}_{n_i}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 17.9(.48) \\ 9.05(.34) \\ \text{-}0.450(.04) \end{pmatrix} \qquad \hat{\boldsymbol{\gamma}} = \begin{pmatrix} \text{-}1.03(.61) \\ 0.524(.20) \end{pmatrix} \qquad \hat{\mathbf{\Phi}} = \begin{pmatrix} 19.0(1.9) & \\ -1.82 & 0.186(.03) \end{pmatrix} \qquad \hat{\sigma}_e^2 = 26.1(1.4)$$

Compared to Model B, this model fits a little better according to AIC (9103 versus 9107) but a little worse compared to BIC (9139 versus 9135). The parameter estimates associated with T_i are in the desired direction, non-significant for γ_1 and respectable for γ_2: the 95% interval estimate for slope is

$$0.524 \pm 1.96 \cdot 0.20 \Rightarrow (0.132 \leq \gamma_2 \leq 0.916)$$

So we conclude that See Me Soar improved reading performance compared to those in the control condition. To be concrete, it increased the rate of reading by $\hat{\gamma}_2 = 0.524$ points per year on the Peabody Reading test over and above the rate of $\hat{\beta}_2 = 9.05$ for control children.

Postscript: Why not include fixed effects on all three coefficients to examine whether T_i also influences β_{i2}? Why not set it up as

$$\begin{aligned}\beta_{i0} &= \beta_0 + \gamma_1 T_i\\ \beta_{i1} &= \beta_1 + \gamma_2 T_i + b_{i1}\\ \beta_{i2} &= \beta_2 + \gamma_3 T_i + b_{i2}\end{aligned}$$

It turns out that both fit measures become worse, AIC = 9105, BIC = 9145, and that γ_3 is non-significant: $\hat{\gamma}_3 = -0.016(.08)$. So the simpler model as presented above is the more efficient of these two.

Setting the random effects to zero and evaluating $\hat{\boldsymbol{\beta}}_i$ for $T_i = 0$ and $T_i = 1$ over ages 6 to 14 gives the curves for the means of the two groups. The calculations are

$$\begin{pmatrix} \hat{\beta}_{i0} \\ \hat{\beta}_{i1} \\ \hat{\beta}_{i2} \end{pmatrix} = \begin{pmatrix} 17.9(.48) \\ 9.05(.34) \\ \text{-}0.450(.04) \end{pmatrix} + \begin{pmatrix} \text{-}1.03(.61) \\ 0.524(.20) \end{pmatrix} T_i$$

$$\widehat{Read}_{ij} = \hat{\beta}_{i0} + \hat{\beta}_{i1} Age_{ij} + \hat{\beta}_{i2} Age_{ij}^2$$

These are in Figure 4

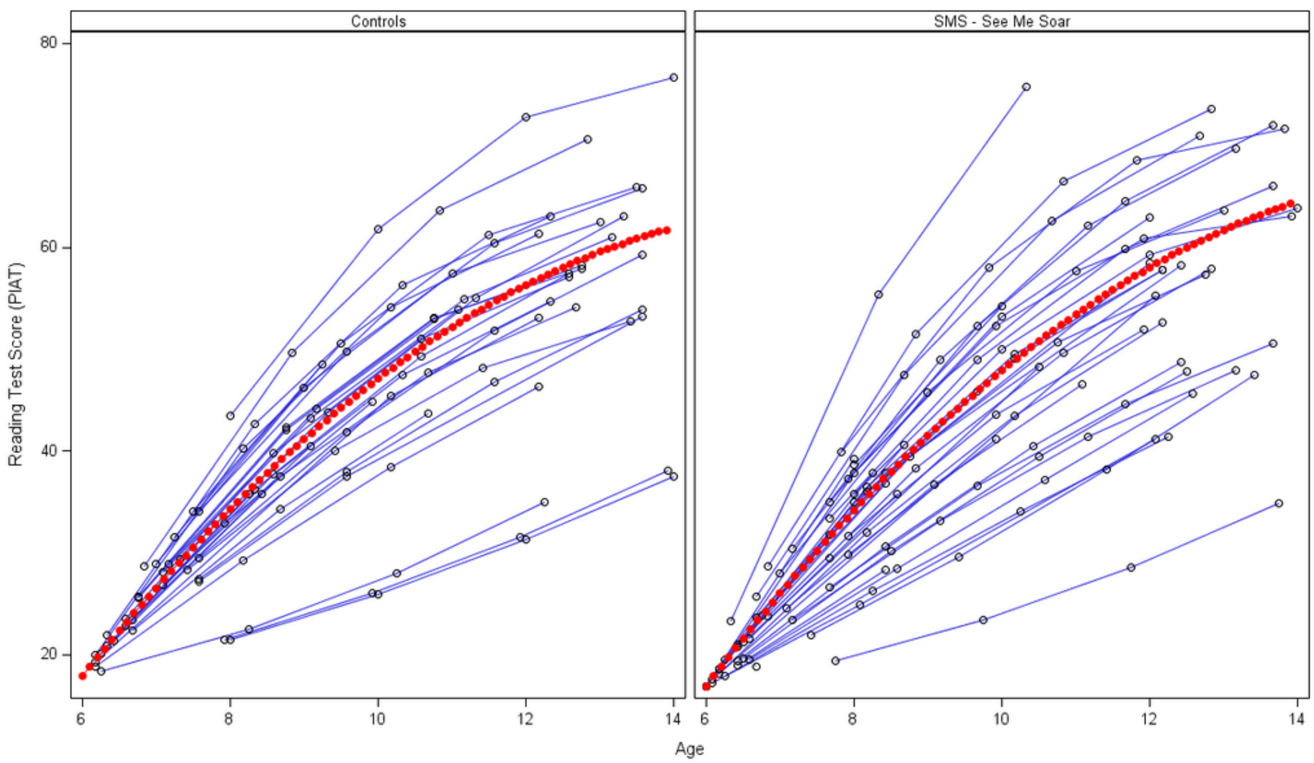

Figure 4. The mean curves for Control students (left) and those from the reading treatment See Me Soar (right), together with 20% random sample of fitted curves for individual students. Based on this approach, the reading program modifies the rate of improvement on β_{i1} of the quadratic function.

After presenting parameter estimates, the curves of the typical subjects in a graph such as Figure 4, and interpreting results, its always interesting to examine individuals. In a large sample with thousands to scan it can seem a waste of time to page though many graphs. However computers are fast, graphs are easy. The idea is to get educated about what the data and statistical analysis have produced. Take random samples. Order participants according to good fitting individuals, poor fitting individuals, fast or slow learners. The mixed model accommodates a very wide range of learning styles. If a person is not fit reasonably well, she or he is an unusual character.

As one example, a trellis display of 16 randomly selected children is below. Things look good here.

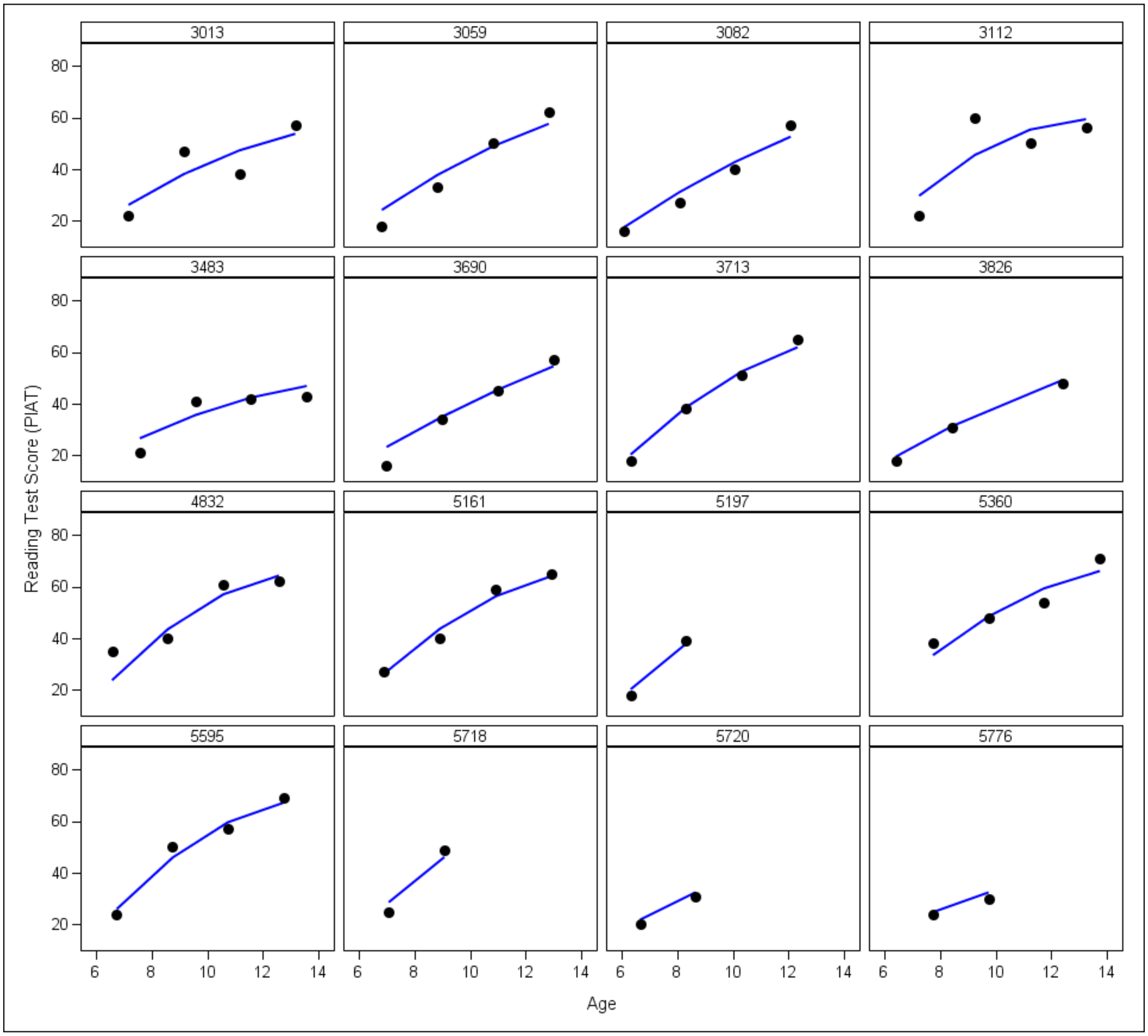

Figure 5. Sixteen randomly selected individuals with fitted functions.

Another interesting view is obtained by selecting random subsets of participants and plotting in groups. This shows similarities and differences in ways that could not even be imagined using numerical methods only. In this sample, $N = 405$. I created $N_g = 35 + 1$ subsets with the instruction

$$b_i = 1 + int(N_g U_i)$$

where U_i is a uniform random number, $0 < U_i < 1$, and $int(x)$ is a function that rounds the decimal value x to the nearest integer. For example, $int(17.58) = 18$. The resulting values are integers in the range $1 \leq b_i \leq N_g + 1$. There are then $N_g + 1$ small clusters of randomly grouped participants.

Here are the first six of 36 groups in a trellis display showing only the fitted functions. This view definitely highlights the flexibility of the quadratic function which adapts to such a diversity of patterns.

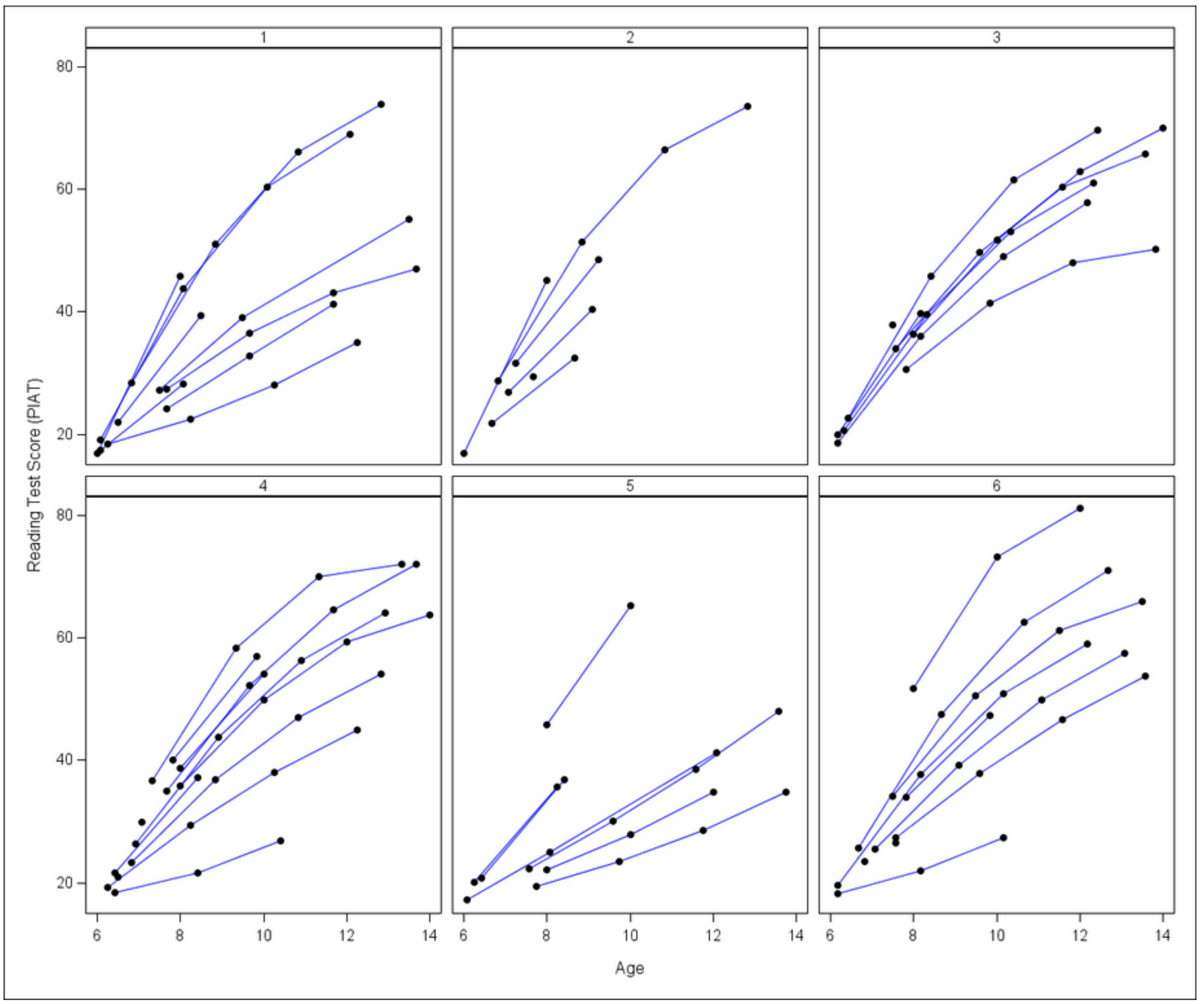

Figure 6. Six subsets of randomly selected individuals, including fitted functions but no data.

Data step and regression setup for Model D in sas

1) For the time variable, define Age = ActualAge - 6. In this data set, "childage" is ActualAge at the first occasion.
2) Up to four repeated measures on the Peabody Reading test, piat1 - piat4
3) Covariates:
   boy = 1 for a boy, boy = 0 for girls
   SmeS = 1 for children in See Me Soar, SmeS = 0 for controls

```
data tall;

  input id piat1-piat4 boy momage childage homecog homeemot SmeS ;

  occas=1; read=piat1;  age=childage-6;  if piat1 ne . then output;
  occas=2; read=piat2;  age=childage-4;  if piat2 ne . then output;
  occas=3; read=piat3;  age=childage-2;  if piat3 ne . then output;
  occas=4; read=piat4;  age=childage;    if piat4 ne . then output;

datalines;
   22  21  39   .   .  0  28  6.08  12  10  0
   34  21  29  45  45  1  28  6.83   8   9  0
   58  23  45  42  46  0  28  6.50   8   6  0
  122  37  80   .   .  1  28  7.83  12  10  0
  125  23  38  43  62  0  29  7.42   9   8  0
  133  18  26  41  40  1  28  6.75   7   9  1
  163  35  48  58  75  1  28  7.17   9  10  1
  190  29  61   .   .  0  28  6.67   8   8  0
  227  18  38  40   .  0  29  6.25  11  10  0
  248  35  57  70  69  0  28  7.50   8   9  1
          < Other records omitted >
      ;


proc mixed data=tall method=ml covtest cl;
  model   read = age age*age   SmeS SmeS*age / s cl;          <- Fixed effects
  random  age age*age / g gcorr type=un subject=id;           <- Random effects
  repeated            / type=toep(1) R=2 Rcorr=2 subject=id;  <- Residuals
run;
```