Topics for today:

- *Modeling random effects in an LMM*

Associated reading: *Sections 1 of 'LMM: modeling random effects and error covariance structure' course notes), Verbeke (with a focus on Ch. 7), Hedeker (Chapters 4-7)*

*1 Modeling random effects (**G** matrix)*

1.1 Adding random slopes (and more) to mixed models

- Choosing whether to use random intercept, slope, both or neither for a particular data set can be based on several criteria.
    - Often, the AIC measure is used to decide which random terms to include. However, this should not be the sole criteria.
    - The researcher should have a sense of whether a random intercept or slope (or both) makes sense to use with the data.
    - If the variance of one of the terms is very low, or the fitting of the model does not converge, then one of the random terms may need to be dropped. But admittedly, it is often not clear what to use until the data have been examined.

- For models involving a random slope for time, one can either include the random intercept or not, but it often makes sense to include the intercept, for many of the same reasons that the y-intercept is generally included in the fitting of a simple linear regression equation.

- But if theoretically it makes sense that all subjects should start at the same point at time 0, then the random intercept term can be removed.

- In the following sections, when I say 'linear random effect model' or 'quadratic random effect model', it will include lower-order random effect terms, unless specifically mentioned otherwise. For example, a quadratic random effect model will have intercept, first order and second order random effect terms.

1.2 Linear random effect regression models

- <u>The Sleep data</u>: One project I have been involved in at National Jewish Health involves a clinical trial for subjects with sleep apnea (Aloia et al., 2013):

    - "…Positive airway pressure (PAP) is considered the standard of care for moderate to severe OSA, generating pressurized air through a pneumatic pump.

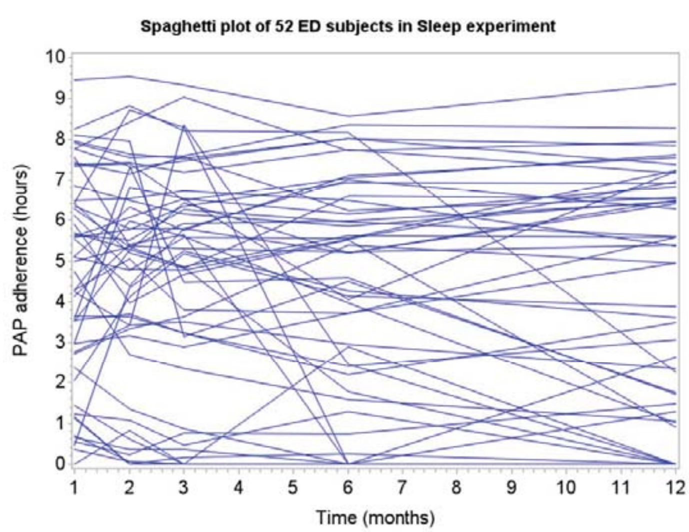
    - The pressurized air is delivered into the upper airway through flexible tubing connected typically to a nasal interface. Once heavy, cumbersome, and uncomfortable, PAP devices are now lightweight and portable, with a wide assortment of interface options from which patients can choose (e.g., nasal masks, full face masks covering the nose and the mouth, varying sizes and shapes of masks)."

- The sleep experiment was conducted in Providence, RI (at a hospital affiliated with Brown University), with the primary purpose of determining how good adherence to use of PAP machines were over time. Adherence was specifically defined as the length of time in hours they used the PAP machine at the prescribed pressure per day while they slept.

- A total of 227 subjects were enrolled and randomized to receive one of 3 treatments aimed at helping them understand sleep apnea and the benefits of treatment. One particular treatment was 'Education' (ED). ED participants were educated regarding the pathophysiology of apnea, its medical and behavioral consequences, and the benefits of treatment.

- Although one of the aims of the actual analysis was to compare adherence means over time between the ED treatment group and other treatment groups, here we will just model the ED group over time in order to better understand different ways linear mixed models can be written and their impacts on estimates of interest. Average adherence responses were computed for each subject for the 1^{st}, 2^{nd}, 3^{rd}, 6^{th} and 12^{th} months.

- The spaghetti plot of the data follows for the 52 ED subjects that completed the experiment (i.e., had no missing data) follows. The data illustrate that between-subject variability was far larger than within –subject variability over time.



- In modeling these data, I will take a few different approaches.
  - First, we will model the clustered data by only including random effects, but keeping the simple independent form for the **R** matrix.
  - We will then compare such model fits with one that specifies a non-independent **R** matrix but no random effects (later in these notes). This will better illustrate that there is no one 'correct' way to model the data, and that correlated responses can be modeled in various ways. To begin with, let's consider statistical models for these data.

Subject form:

$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$, where $\mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 6 \\ 1 & 12 \end{pmatrix}$ for subjects without missing data,

$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, $\mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$, $\mathbf{R}_i = \sigma^2 \underset{n_i \times n_i}{\mathbf{I}}$, $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{G}_i)$,

and where $\mathbf{G}_i = \begin{pmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix}$ (Case I) or $\mathbf{G}_i = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}$ (Case II).

Observation form:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 t_j + b_{0i} + b_{1i} t_j + \varepsilon_{ij} \\ &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) t_j + \varepsilon_{ij} \\ &= u_{0i} + u_{1i} t_j + \varepsilon_{ij} \end{aligned}$$

Here is the SAS code and partial output in 2 cases, one where we allow covariance between the two random terms (specified by the TYPE=UN statement) and one where we do not allow it (the default, TYPE=VC statement).

```
*linear, VC structure for G;
proc mixed data=sleep_nc; class id;
 model y=  time / solution;
 random intercept time / type=vc
subject=id v g; run;
```

Abbreviated output:

```
Covariance Structure      Variance Components
Subject Effect            id
Estimation Method         REML
Degrees of Freedom Method Containment

Dimensions
Covariance Parameters          3
Columns in X                   2
Columns in Z Per Subject       2
Subjects                      52
Number of Observations Used  260

Estimated G Matrix
Row Effect    id      Col1    Col2
1   Intercept 1021    5.7752
2   time      1021            0.03742
```

```
*linear, UN structure for G;
proc mixed data=sleep_nc; class id;
 model y=  time / solution;
 random intercept time / type=un
subject=id v g; run;
```

Abbreviated output:

```
Covariance Structure Unstructured
Subject Effect id
Estimation Method REML
Degrees of Freedom Method Containment

Dimensions
Covariance Parameters          4
Columns in X                   2
Columns in Z Per Subject       2
Subjects                      52
Number of Observations Used  260

Estimated G Matrix
Row Effect    id     Col1     Col2
1   Intercept 1021   6.1687  -0.1485
2   time      1021  -0.1485   0.04083
```

```
Estimated V Matrix for id 1021
Row Col1   Col2   Col3   Col4   Col5
1   6.9293 5.8501 5.8875 5.9998 6.2243
2   5.8501 7.0416 5.9998 6.2243 6.6734
3   5.8875 5.9998 7.2287 6.4489 7.1225
4   5.9998 6.2243 6.4489 8.2392 8.4698
5   6.2243 6.6734 7.1225 8.4698 12.2810

Covariance Parameter Estimates
Cov Parm  Subject Estimate
Intercept id      5.7752
time      id      0.03742
Residual          1.1167

Fit Statistics
-2 Res Log Likelihood     1010.2
AIC (smaller is better)  1016.2
AICC (smaller is better) 1016.3
BIC (smaller is better)  1022.1

Solution for Fixed Effects
Effect    Estimate Std Error DF t Value Pr > |t|
Intercept 4.7892   0.3488    51 13.73   <.0001
time     -0.06290  0.03150   51 -2.00   0.0512

Type 3 Tests of Fixed Effects
Effect Num DF Den DF F Value Pr > F
time   1      51     3.99    0.0512
```

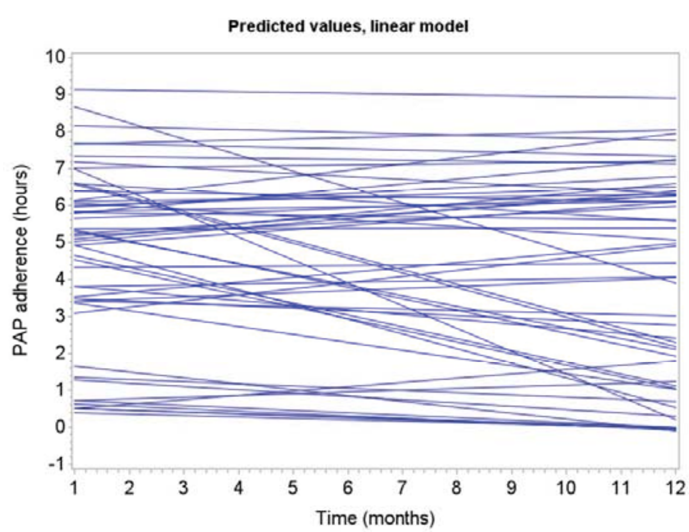
```
Estimated V Matrix for id 1021
Row Col1   Col2   Col3   Col4   Col5
1   7.0071 5.8048 5.6971 5.3740 4.7279
2   5.8048 6.8325 5.6711 5.4705 5.0693
3   5.6971 5.6711 6.7396 5.5669 5.4107
4   5.3740 5.4705 5.5669 6.9508 6.4349
5   4.7279 5.0693 5.4107 6.4349 9.5778

Covariance Parameter Estimates
Cov Parm Subject Estimate
UN(1,1)  id      6.1687
UN(2,1)  id     -0.1485
UN(2,2)  id      0.04083
Residual         1.0946

Fit Statistics
-2 Res Log Likelihood     1006.9
AIC (smaller is better)  1014.9
AICC (smaller is better) 1015.1
BIC (smaller is better)  1022.8

Solution for Fixed Effects
Effect    Estimate Std Error DF t Value Pr > |t|
Intercept 4.7892   0.3592    51 13.33   <.0001
time     -0.06290  0.03244   51 -1.94   0.0580

Type 3 Tests of Fixed Effects
Effect Num DF Den DF F Value Pr > F
time   1      51     3.76    0.0580
```

- The results indicate an almost significant (at the 0.05 level) average drop in adherence over time. From the 1^{st} to 12^{th} month, average adherence is expected to change by an average of 11*(-0.0629)=-0.692 hours. I.e., it drops by an average of about 40 minutes over 11 months, or nearly 4 minutes per month.

- Putting the random slope in the model does appear to improve model fit. The AIC for the same model as above but excluding the random slope yields AIC=1054.1, much higher than those above. (With only one parameter in **G**, there is no difference between UN and VC approaches.) Comparing the models with different forms for the **G** matrix (but that both have the random slope term), there is a slight improvement using the unstructured **G** (i.e., allowing a covariance term between intercept and slope), rather the 'variance components' structure.

- How do we interpret the intercept – slope covariance term in the **G** matrix? What does it mean to have positive covariance? Negative covariance?

- Here is a graph of the predicted values for subjects based on the UN structure. You see three types of subjects here: those that drop in adherence, those that maintain (generally higher up), and those that steadily increase. In a sense, we have straightened out the noodles from the raw spaghetti plot.



1.3 Models that use a random slope term for variables other than time

- Up till now we have examined models with random intercept and random slopes for terms involving time. It is also possible to fit a linear mixed model using random terms for variables other than time.

- For example, in an analysis I did at NJH, I fit a model for personal exposure fine particulate matter ($PM_{2.5}$) from ambient (outdoor) sources, as a function of pollution from a fixed outdoor monitor.

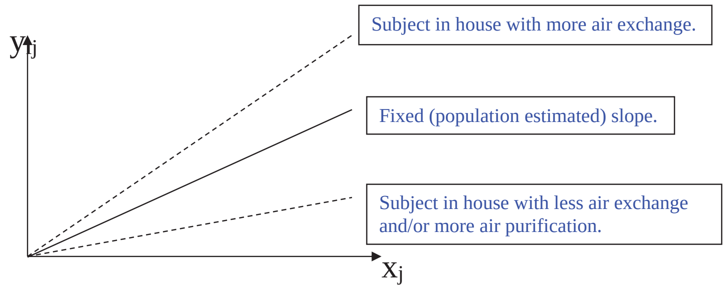
  - We assume that the ambient pollution is fairly homogenous here.
  - Random term for the pollutant used – subjects living in housing with more air exchange (e.g., windows open more) may have higher slopes than those in more closed quarters. [Note: air pollution usually enters the home at a certain rate.]
  - Not including fixed or random intercept terms makes sense since in the absence of outdoor air pollution (measured by the stationary monitor), there is not expected to be ambient pollution measured by the personal monitors. [Generally, be careful before setting the y-intercept to 0 – we can discuss more in class…]

<u>The model</u>

$$\begin{aligned} Y_{ij} &= \beta_1 X_j + b_{i1} X_j + \varepsilon_{ij} \\ &= (\beta_1 + b_{i1}) X_j + \varepsilon_{ij} \end{aligned}$$

where X_j = ambient $PM_{2.5}$ concentration from fixed monitor, Y_{ij} = personal ambient $PM_{2.5}$ concentration, i indexes subject, j indexes day, $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$, $b_{i1} \sim N(0, \sigma^2_{b_1})$. Again, here note that the fixed and random intercepts are not included; I want to stress that this is only done in special cases. In general, lower order terms should also be included unless there are specific reasons to remove them (other than just because they are insignificant).

Illustration and notes



Y_{ij} is actually estimated:

$\Rightarrow Y_{ij}^{*} = Y_{ij} + w_{ij}$, where $w_{ij} \sim N(0, \sigma_w^2)$ (a simple error model)

$\Rightarrow Y_{ij}^{*} - w_{ij} = (\beta_1 + b_{i1})X_j + \varepsilon_{ij} \Rightarrow Y_{ij}^{*} = (\beta_1 + b_{i1})X_j + (\varepsilon_{ij} + w_{ij}) = (\beta_1 + b_{i1})X_j + \varepsilon'_{ij}$

where $\varepsilon'_{ij} \sim N(0, \sigma_\varepsilon^2 + \sigma_w^2)$ if errors ε and w are independent.

Generally, when Y is measured with error, the error gets absorbed into the general error term. When X (a covariate) is measured with error, we need to implement measurement error model methods (such as regression calibration) for correct inference. Otherwise, estimators may be biased.

1.4 Quadratic random effect regression models

- In some cases we may want to examine patterns for subjects over time, and not have the restriction that subject's patterns follow straight lines.

    - For example, with the Kunsberg kids at NJH, we measured FEV_1 and FVC (measures of lung function) daily over the school year (roughly October through April).
    - Due to seasonal events such as allergies, subjects may have ups and downs in their lung function trends over time. Specifically, some kids may start with higher lung function, then dropped in the colder winter months, then improved again (concave up), and some kids may have peaks in the colder months (concave down).
    - Of course, some kids might exhibit steady improvements or declines over the school year, or simply plateau; but all of the aforementioned patterns could be modeled using the mixed quadratic regression model. Those without quadratic trends will simply have much weaker quadratic term components.

- One caution: there needs to be sufficient data in order to fit multiple random terms in a LMM. With limited data, it may only be possible to fit one random term. But if the data set is large enough, it may be possible to fit 2 or 3 random terms, if they are deemed worthy. Any given random term's significance can be assessed by determining whether its addition contributed enough to the model.

- We can model the Sleep data using intercept, linear and quadratic random terms for time, so that subject quadratic trends can be examined. In addition, we have the same polynomial terms for fixed effects. The SAS code follows. Can you write the statistical model?

- Using the UN structure for **G** adds 6 covariance parameters. There is also the residual error variance, making 7 total parameters, but one parameter in **G** is estimated to be 0, indicating a redundancy, so for practical purposes there are 6 covariance parameters.

```
proc mixed data=sleep_nc;
 class id;
 model y=time time*time / solution outp=quad_pred;
 random intercept time time*time / type=un subject=id v;
 estimate 'mean at 1 mo' intercept 1 time 1 time*time 1;
 estimate 'mean at 12 mo' intercept 1 time 12 time*time 144;
 estimate 'diff 1mo - 12mo' time -11 time*time -11; run;
```

```
The Mixed Procedure

Dimensions
Covariance Parameters          7
Columns in X                   3
Columns in Z Per Subject       3
Subjects                      52
Max Obs Per Subject            5
Number of Observations Used  260

Estimated V Matrix for id 1021
Row   Col1  Col2  Col3  Col4  Col5
1     6.85  5.68  5.60  5.29  4.32
2     5.68  6.80  5.70  5.59  4.93
3     5.60  5.70  6.87  5.86  5.48
4     5.29  5.59  5.86  7.57  6.81
5     4.32  4.93  5.48  6.81  9.09
```

```
Fit Statistics
-2 Res Log Likelihood    1012.7
AIC (smaller is better)  1024.7
AICC (smaller is better) 1025.0
BIC (smaller is better)  1036.4

Solution for Fixed Effects
Effect      Estimate Std Error DF t Value Pr > |t|
Intercept   4.8287   0.3811    51 12.67   <.0001
time       -0.08387  0.08864   51 -0.95   0.3485
time*time   0.001575 0.005736  51  0.27   0.7848

Estimates
Label          Estimate Std Error DF t Value Pr > |t|
mean at 1 mo    4.7464    0.3521 51 13.48     <.0001
mean at 12 mo   4.0491    0.4179 51 9.69      <.0001
diff 1mo - 12mo 0.9052    0.9164 51 0.99      0.3279
```

- This fit yields an AIC of 1024.6 and insignificant quadratic effects (similar results for the VC structure). Since there was no pre-hypothesized quadratic pattern, there is no real reason to keep this model over the previous one with only an intercept and random slope for time. Interestingly for this new model, the drop from 1st to 12th month is much less significant. However given that the model had a worse fit, I would probably not put this result into a 'final report'.

- Despite the slightly worse fit, using more covariance parameters allows us to get a more accurate fit for $Var(\mathbf{Y})$. But the AIC statistic indicates that this better fit is not worth the extra covariance parameters added to the model. The issue of modeling covariance structures will be discussed more a few sections later. Later we will also revisit the data and show how the error variance matrix can be modeled using spatial covariance structures.

- <u>The Mt. Kilimanjaro data</u>:

  - Oxygen saturation, or SAO_2, can be measured as the percentage of hemoglobin molecules which are oxgenated (oxyhemoglobin) in arterial blood.

  - The normal range is >95%, however at higher altitudes this percentage tends to go down.

  - This measure was taken on hundreds of subjects that climbed Mt. Kilimanjaro (the tallest mountain on the continent of Africa).

  - The following graph shows SAO_2 versus altitude, along with a quadratic fit using a linear mixed model.

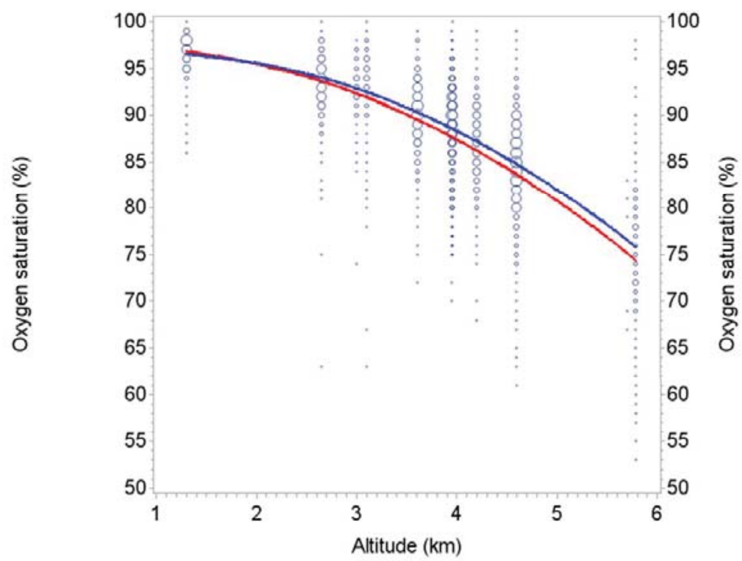
  - The bubble plot was used because many values occurred on the same (x,y) location; bubbles indicate how many subjects occurred at each point, the bigger the bubble, the more subjects at that location.

- Although these are repeated measures data, lines connect points are suppressed due to the large amount of data.

- Superimposed on the bubble plot are two curves, one showing subjects that were taking a medication to help prevent symptoms of high-altitude sickness (blue), and those that were not (not).

- Those taking the medication are able to maintain slightly higher oxygen levels (which may also help reduce symptoms of high altitude sickness), with greater differences at higher altitudes.

- These differences are statistically significant after about 3km, but it may be somewhat subjective as to whether the small differences as worth taking the medication.



- The bubble plot was generated by using the 'bubble' statement in PROC GPLOT. I then overlayed the fitted curves from the fitted linear mixed model in a subsequent 'plot2' statement.

- Below is the SAS code used to fit the mixed model:

```
proc mixed data=alldata; class id recnum;
model oxygen_sat= x x*x diamox_ever x*diamox_ever x*x*diamox_ever
 / outpm=outypm outp=outyp solution;
random intercept x x*x / subject=id v solution g type=un;
estimate 'diamox, alt=5km'
   intercept 1 x 5 x*x 25 diamox_ever 1 x*diamox_ever 5 x*x*diamox_ever 25;
estimate 'no diamox, alt=5km' intercept 1 x 5 x*x 25;
estimate 'diff at alt=1km' diamox_ever 1 x*diamox_ever 1 x*x*diamox_ever 1;
estimate 'diff at alt=2km' diamox_ever 1 x*diamox_ever 2 x*x*diamox_ever 4;
estimate 'diff at alt=3km' diamox_ever 1 x*diamox_ever 3 x*x*diamox_ever 9;
estimate 'diff at alt=4km' diamox_ever 1 x*diamox_ever 4 x*x*diamox_ever 16;
estimate 'diff at alt=5km' diamox_ever 1 x*diamox_ever 5 x*x*diamox_ever 25;
estimate 'diam intercept' intercept 1 diamox_ever 1;
estimate 'diam x term' x 1 x*diamox_ever 1;
estimate 'diam x*x term' x*x 1 x*x*diamox_ever 1;
contrast 'interaction' x*diamox_ever 1, x*x*diamox_ever 1;
contrast 'curve comparison'
              diamox_ever 1, x*diamox_ever 1, x*x*diamox_ever 1; run;
```

Abbreviated output:

```
The Mixed Procedure

Dependent Variable          Oxygen_Sat
Covariance Structure        Unstructured
Subject Effect              id
Estimation Method           REML
Residual Variance Method    Profile
Fixed Effects SE Method     Model-Based
Degrees of Freedom Method   Containment

Dimensions

Covariance Parameters            7
Columns in X                     6
Columns in Z Per Subject         3
Subjects                       916
Max Obs Per Subject             20
Number of Observations Used  13369

Estimated G Matrix

Row Effect id   Col1   Col2   Col3
1   Int.    1        -6.434  1.572
2   x       1 -6.434  9.116 -1.790
3   x*x     1  1.572 -1.790  0.353

Residual Variance Estimat:    8.8320
```

```
Solution for Fixed Effects

Effect          Est. Std Err    DF t Value  Pr>|t|
Intercept       97.062  0.737  914 131.74  <.0001
x                0.954  0.555  914   1.72  0.0857
x*x             -0.840  0.093  914  -9.01  <.0001
diamox_ever     -1.096  0.775 11E3  -1.41  0.1575
x*diamox_ever    0.663  0.584 11E3   1.14  0.2561
x*x*diamox_ever -0.040  0.098 11E3  -0.41  0.6840

Solution for Random Effects
Effect     Id Estimate Std Err DF t Value  Pr>|t|
                      Pred
Intercept  1  -6.2686  0      11E3 -Infty  <.0001
x          1   3.5771  0      11E3  Infty  <.0001
x*x        1  -0.8219  0.320  11E3 -2.57   0.0102
Intercept  2  -2.8734  0      11E3 -Infty  <.0001
x          2   1.5678  0      11E3  Infty  <.0001
x*x        2  -0.1304  0.119  11E3 -1.09   0.2744
Intercept  3   1.9668  0      11E3  Infty  <.0001
x          3  -1.2063  0      11E3 -Infty  <.0001
x*x        3   0.1812  0.119  11E3  1.52   0.1290
. . .
Intercept 921 -8.8468  0      11E3 -Infty  <.0001
x         921  5.2530  0      11E3  Infty  <.0001
x*x       921 -1.1837  0.135  11E3 -8.75   <.0001
```

```
Fit Statistics

-2 Res Log Likelihood    69592.3
AIC (smaller is better)  69604.3
AICC (smaller is better) 69604.3
BIC (smaller is better)  69633.2

Type 3 Tests of Fixed Effects

Effect      Num DF Den DF F Value Pr>F
x                1    914   2.96 0.0857
x*x              1    914  81.13 <.0001
diamox_ever      1   11E3   2.00 0.1575
x*diamox_ever    1   11E3   1.29 0.2561
x*x*diamox_ever 1    11E3   0.17 0.6840
```

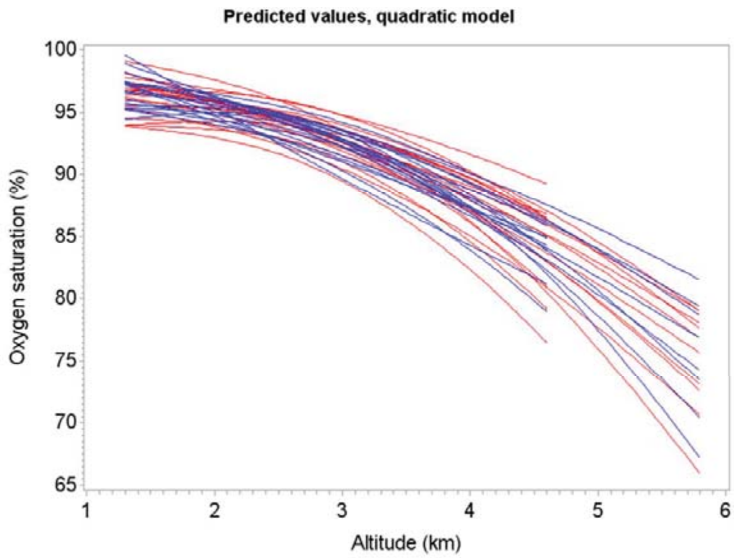
```
Estimates

Label               Est. Std Err DF t Value Pr>|t|
Diamox, alt=5km   82.060 0.152 11E3 541.43 <.0001
no diamox, alt=5km 80.839 0.466 914 173.49 <.0001
diff at alt=1km   -0.472 0.335 11E3  -1.41 0.1580
diff at alt=2km    0.071 0.204 11E3   0.35 0.7281
diff at alt=3km    0.534 0.252 11E3   2.12 0.0341
diff at alt=4km    0.918 0.290 11E3   3.16 0.0016
diff at alt=5km    1.221 0.490 11E3   2.49 0.0127
diam intercept    95.967 0.240 11E3 399.32 <.0001
diam x term        1.617 0.182 11E3   8.87 <.0001
diam x*x term     -0.880 0.031 11E3 -28.66 <.0001

Contrasts

Label            Num DF Den DF  F Value Pr > F
interaction         2   11E3     6.21   0.0020
curve comparison    3   11E3     4.37   0.0044
```

- Some interesting things to point out from the output:
    - The variance of the intercept was estimated to be 0. However, no penalty was added for this in the AIC; essentially, that parameter is removed from the model. You can tell this is the case because the difference between -2 Restriced log Likelihood and the AIC is 12, so 6 parameters are accounted for (5 in G matrix, plus residual variance).

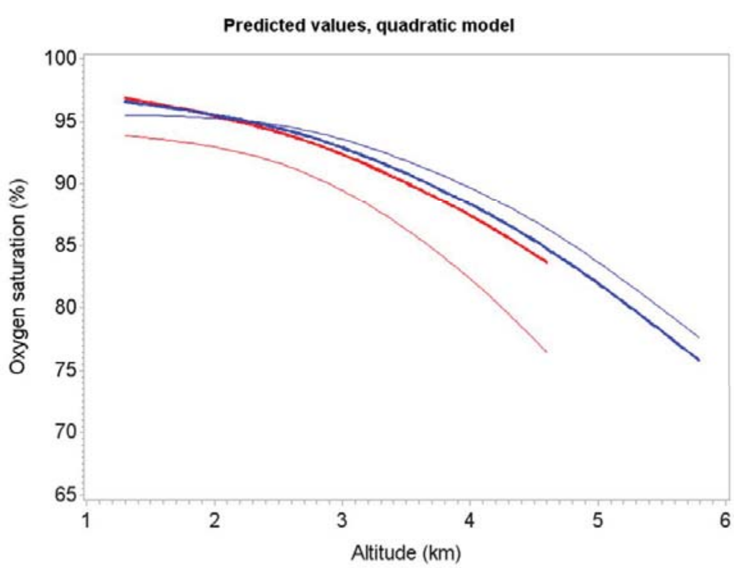
    - Notice also that subject estimates of intercept and some linear terms have predicted standard errors of 0; our interpretation should be that these standard errors could not be estimated, rather than that they were true 0's.

    - Estimates included demonstrate that although differences between medication users and non-users appeared to be minor, visually, they were statistically significant, with greater significance at higher elevations.

    - The contrasts indicate that there were differences between curves that could not be accounted for by intercept differences alone (see 'interaction' test, p=0.0020), and that the curves were not the same (including both 'interaction' and y-intercept differences, p=0.0044).



- The graph above shows predicted values for subjects, from the mixed model, using only 20 per group (no medication – red / medication – blue) are plotted. Differences for subjects are due to the use of the intercept, linear and quadratic random effect terms. Notice that the predicted curves tend to fan out at higher altitudes, just like the raw data.



- This graph shows population-averaged estimates (thick red and blue) and two subject curves (thin red and blue: see subject ID's 1 (red) and 2 (blue) on previous SAS output.

- Recall that random effect estimates are deviations from fixed effect estimates. Notice how the red subject curve has much greater curvature than the thick red curve, compared with the thin and thick blue lines. This is echoed in the tests on the previous output, where the t-tests indicate that subject 1 has significant difference in quadratic effect compared with its population counterpart (p=0.0102), while the blue does not (0.2744).

- Both subject curves have lower intercepts and higher coefficients for the first order term, although we are unable to conduct tests to see if they are significantly different than population counterparts, since SEs could not be determined. Accounting for serial correlation in the data improves the fit even more. This will be discussed more in the following sections.