Longitudinal analysis of AIDS Clinical Trial Data

1. Purpose

To use SAS to fit different models for the mean and correlation of longitudinal data from a four-group study design.

2. Data

The data are downloaded from *Applied Longitudinal Analysis, 2nd Edition*: AIDS Clinal trial group (ACTG) study 193A (cd4-data.txt). The data are from a randomized, double-blind study of AIDS patients with advanced immune suppression (CD4 counts of less than or equal to 50 cells/mm^3).

Patients in AIDS Clinical Trial Group (ACTG) Study 193A were randomized to dual or triple combinations of HIV-1 reverse transcriptase inhibitors. Specifically, patients were randomized to one of four daily regimens containing 600mg of zidovudine: 1 = zidovudine alternating monthly with 400mg didanosine, 2 = zidovudine plus 2.25mg of zalcitabine, 3 = zidovudine plus 400mg of didanosine, and 4 = zidovudine plus 400mg of didanosine plus 400mg of nevirapine.

Measurements of CD4 counts were scheduled to be collected at baseline and at 8-week intervals during follow-up for up to 40 weeks. The response variable is the log transformed CD4 counts, log(CD4 counts + 1), available on 1309 patients.

3. Descriptive data analysis

3.1 Treatment

The mean level of logCD4 over time in each of the four treatment groups are summarized in table 1. The level of CD4 were collected at baseline (time 0) and scheduled to be collected every 8 weeks for five times (time 1 to time 5). However, the CD4 count data are unbalanced due to mistimed measurements and missing data that resulted from skipped visits and dropout, for example, the observation of treatment 1 at time 0 are 335 while the observation decreased to 57 at the last visit time 5.

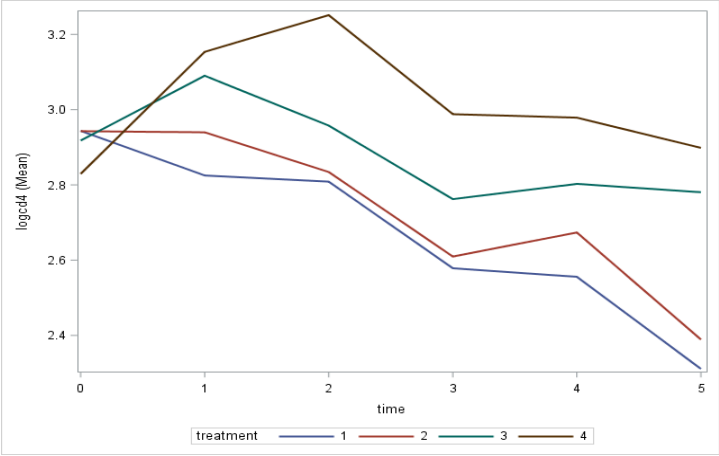
Table 1. Mean response of logCD4 over time in each treatment

Analysis Variable : logcd4					
treatment	time	N Obs	Mean	Std Dev	Variance
1	0	335	2.9435870	0.8283428	0.6861517
	1	225	2.8250139	1.0251331	1.0508978
	2	242	2.8085042	0.9500774	0.9026471
	3	177	2.5785915	0.9304648	0.8657647
	4	203	2.5556553	0.9105806	0.8291571
	5	57	2.3109993	1.1231374	1.2614376
2	0	341	2.9427134	0.9305374	0.8658999
	1	223	2.9397451	1.0849009	1.1770099
	2	245	2.8341586	1.0555707	1.1142295
	3	187	2.6096796	1.1079948	1.2276526
	4	183	2.6735905	1.0387971	1.0790995
	5	72	2.3890957	1.2686565	1.6094894
3	0	346	2.9178908	0.9406495	0.8848214
	1	225	3.0900827	1.1894003	1.4146730
	2	255	2.9575089	1.1953125	1.4287719
	3	180	2.7622109	1.1538690	1.3314136
	4	189	2.8027486	1.2012494	1.4430001
	5	59	2.7805243	1.1275526	1.2713749
4	0	350	2.8292688	0.9604452	0.9224550
	1	232	3.1536976	1.1893406	1.4145310
	2	252	3.2510915	1.1990537	1.4377297
	3	179	2.9879549	1.1832512	1.4000834

Analysis Variable : logcd4					
treatment	time	N Obs	Mean	Std Dev	Variance
	4	208	2.9787630	1.2407587	1.5394821
	5	71	2.8983923	1.1710144	1.3712747

The mean level of logCD4 over time in each of the four treatment groups are visualized in figure 1. The level of logCD4 are the same at baseline (time 0) which means it is a well-designed clinical trial. The Level of logCD4 of each treatment getting different after the first follow up visit, more specifically, treatment 4 has the highest logCD4 value, followed by treatment 3, 2, and 1. This order continued over the entire follow up period. The difference of logCD4 between each treatment getting bigger after time 0 until week 1; the difference between treatment 4 and other groups continued to get bigger until week 2; the difference of logCD4 between each group remain constant between time 2 and time 4, after then the difference getting bigger between group 3,4 and group 1,2 till time 5.

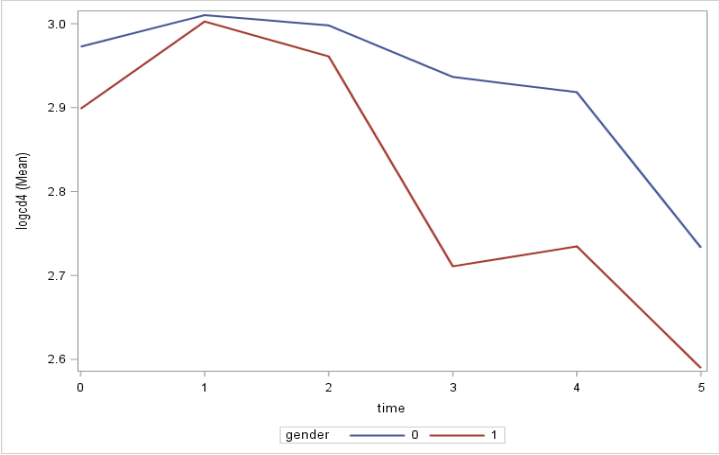
Figure 1. Mean response of logCD4 over time in each treatment



3.1 gender

The mean level of logCD4 over time for two genders (1=M, 0=F) are shown in figure 2, the overall level of logCD4 in female are higher than male, the difference between them gets smaller after time 0 then gets bigger after time 1 until time 3 followed by constant difference between time.

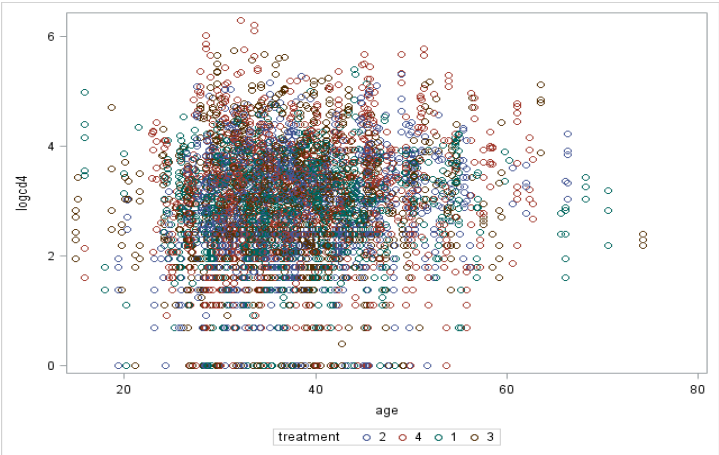
Figure 2 The mean level of logCD4 over time for different genders



3.2 Ages

The age of each patient was also collected during the trial and the figure 3 showed that there is no strong correlation of age and level of logCD4.

Figure 3 Scatter plot of relationship between age and logCD4



4. Model selection

The repeated measures for the same subject are correlated, and this correlation must be taken into account in a repeated measures analysis. We need to specify a covariance structure for the repeated measurements of an individual subject. The regression model for each subject can be assumed to be a random deviation from some population regression model; the standard random coefficient model involves a random intercept and slope for each subject. Random coefficient models are sensible whenever the data arise from independent subjects.

4.1 Model with linear trend in time with different covariance

For the maximal model, random intercept and slope model for the mean response are fitted for the following models of covariance: a) unstructured covariance; b) compound symmetry; c) heterogeneous compound symmetry; d) autoregressive; and e) heterogeneous autoregressive. The best model is selected based on AIC which is the small the better.

Table 2. AIC value of different models of covariance

Covariance	unstructured	compound symmetry	heterogeneous compound symmetry	autoregressive	heterogeneous autoregressive
AIC	12110.7	13722.8	12110.7	13722.8	12110.7

As shown in table 2, model with unstructured, heterogeneous compound symmetry and heterogeneous autoregressive covariance have the smallest AIC 12110.7. Thus unstructured covariance was chosen for the following model.

4.2 Model with linear trend in time with unstructured covariance

Visit weeks were treated as continuous variable and fit a model that includes the effects of age, gender, treatment, week, and the interaction of treatment and

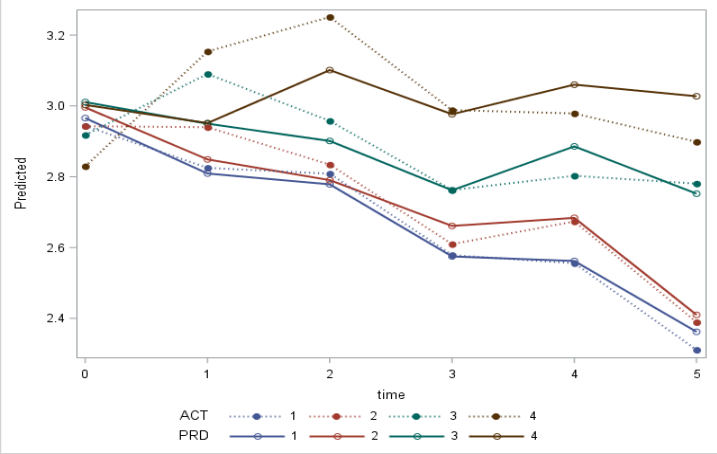
week. The results of test of fixed effects are shown in table 3. The results showed that the pattern of change over time is different for treatments($p<0.0001$).

Table 3. Tests of fixed effects on linear weeks model

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
week	1	1183	150.61	150.61	<.0001	<.0001
treatment	3	2540	0.06	0.02	0.9962	0.9962
age	1	2540	11.94	11.94	0.0006	0.0006
gender	1	2540	0.96	0.96	0.3281	0.3282
week*treatment	3	2540	50.86	16.95	<.0001	<.0001

The time plot that displays the estimated and actual mean logCD4 versus time for different treatment from the model are shown in figure 4. The model predicted well in treatment 1 and 2 while bad in treatment 3 and 4.

Figure 4. Predicted and actual mean logCD4 in linear time model



4.3 Model with quadratic trend in time with unstructured covariance

Since model's performance was not so good when treat time as linear trend, in this part time effect was treated as quadratic trend. Visit weeks were treated as continuous variable and fit a model that includes the effects of age, gender, treatment, week, week*week, the interaction of treatment with week and the

interaction of week*week with treatment. The results of test of fixed effects are shown in table 4. The results showed that the pattern of change over quadratic time is different for treatments($p<0.0001$). Moreover, the AIC of the model is 12086.7, which is smaller than linear time model (12110.7), thus model with quadratic time trend explained much better than linear time model. Thus the final model of this project is based on quadratic time trend.

Table 4. Tests of fixed effects on quadratic weeks model

Type 3 Tests of Fixed Effects						
Effect	**Num DF**	**Den DF**	**Chi-Square**	**F Value**	**Pr > ChiSq**	**Pr > F**
week	1	1183	5.45	5.45	0.0195	0.0197
treatment	3	2536	1.87	0.62	0.6006	0.6007
age	1	2536	11.98	11.98	0.0005	0.0005
gender	1	2536	1.07	1.07	0.3020	0.3021
week*week	1	2536	57.20	57.20	<.0001	<.0001
week*treatment	3	2536	61.20	20.40	<.0001	<.0001
week*week*treatment	3	2536	30.24	10.08	<.0001	<.0001

5. Final model

The estimated coefficients of the final model are shown in table 5.

Table 5. Solution of fixed effect model

Solution for Fixed Effects							
Effect	**gender**	**treatment**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept			2.4700	0.1297	1303	19.05	<.0001
week			0.03314	0.004451	1183	7.44	<.0001
treatment		1	0.09328	0.07596	2536	1.23	0.2195
treatment		2	0.08495	0.07587	2536	1.12	0.2630
treatment		3	0.05253	0.07562	2536	0.69	0.4873
treatment		4	0	.	.	.	.
age			0.01067	0.003082	2536	3.46	0.0005

Solution for Fixed Effects							
Effect	**gender**	**treatment**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
gender	0		0.07959	0.07712	2536	1.03	0.3021
gender	1		0	.	.	.	.
week*week			-0.00101	0.000122	2536	-8.25	<.0001
week*treatment		1	-0.04617	0.006408	2536	-7.20	<.0001
week*treatment		2	-0.03909	0.006354	2536	-6.15	<.0001
week*treatment		3	-0.02609	0.006366	2536	-4.10	<.0001
week*treatment		4	0	.	.	.	.
week*week*treatment		1	0.000892	0.000177	2536	5.04	<.0001
week*week*treatment		2	0.000754	0.000174	2536	4.33	<.0001
week*week*treatment		3	0.000484	0.000176	2536	2.75	0.0060
week*week*treatment		4	0	.	.	.	.

Based on table 5, the regression model could be written as:

logCD4 =

*2.47 + 0.3314*week + 0.09328*I(treatment=1) + 0.08495*I(treatment=2) + 0.05253* I(treatment=3) + 0.01067*age – 0.07959*I(gender=F) – 0.00101*week^2 – 0.04617*week* I(treatment=1) – 0.03909*week* I(treatment=2) -0.02609*week* I(treatment=3) + 0.000892*week^2*I(treatment=1) + 0.000754*week^2*I(treatment=2) + 0.000484*week^2*I(treatment=3) + e*

The implementation of the regression model could be explained in the following example, if we want to estimate the level of logCD4 for patient with ID 3 (treatment 1; age 60.2875; gender 1; week 0.0000) with actual logCD4 value 3.73766961830; it could be calculated as:

logCD4(week=0, age=60.2875, treatment=1, gender=1) =

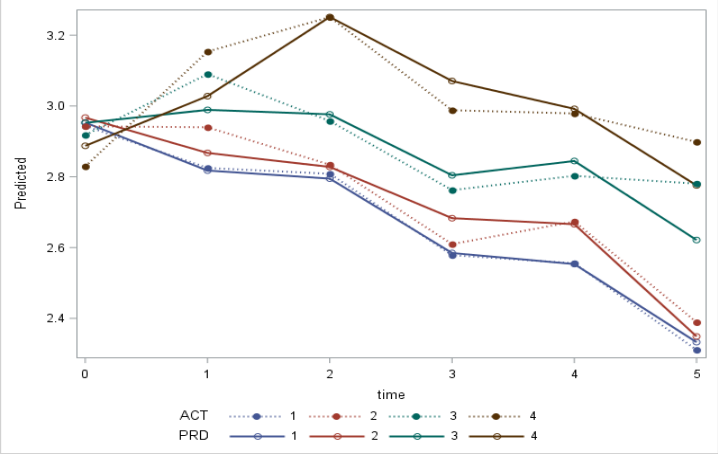
*2.47 + 0.3314*0 + 0.09328*1 + 0.08495*0 + 0.05253*0+ 0.01067*60.2875 - 0.07959*0 - 0.00101*0^2 - 0.04617*0* 1 - 0.03909*0* 1 -0.02609*0* 0 + 0.000892*0^2*1 + 0.000754*0^2*1 + 0.000484*0^2*1 =*

3.206548

The estimated logCD4 of patient with ID 3 at week 0 is 3.206548 which is close to the actual value 3.73766961830.

All the values in the dataset could be predicted similarly and the time plot that displays the estimated and actual mean logCD4 versus time for different treatment from the model are shown in figure 5. The predicted value and actual value ally much better than figure 4.

Figure 5. Predicted and actual mean logCD4 in quadratic time model



6. Conclusion

Compare and contrast the results from linear and quadratic time trend models, the model with only a linear in time could not adequately account for the pattern of change in the four treatment groups, and model with a quadratic in time explains much better than only linear time trend.

Based on the final model, treatment 4 has the best treatment effect on AIDS over time followed by treatment 3, 2 and 1. Female has higher level of logCD4 compare to male patients, while this difference is not significant.

Appendix SAS code

```
ods rtf file="\\Mac\Home\Desktop\biostat\final project\final project-
output.rtf";
libname proj "\\Mac\Home\Desktop\biostat\final project";
data proj.cd4;
infile "\\Mac\Home\Desktop\biostat\final project\cd4-data.txt";
input id$ treatment$ age gender$ week logcd4;
time=0;
if week>5 and abs(week-8)<abs(week-16)then time=1;
if week>12 and abs(week-16)<abs(week-24)then time=2;
if week>20 and abs(week-24)<abs(week-32)then time=3;
if week>28 and abs(week-32)<abs(week-40)then time=4;
if week>30 and abs(week-32)>abs(week-40)then time=5;
run;

/**EDA***/
*count of each visit  by treatment-table;
proc freq data=proj.cd4;
table treatment*time/ nopercent nocum norow nocol out=proj.freqout;  run;
*count of each visit  by treatment-plot;
proc sgplot data=proj.freqout;
hbarparm category=treatment response=count/group=time;
keylegend /opaque across=6 position=bottomright location=outside;
run;
                                /*table 1*/

*mean table: Means across visit by treatment;
proc means mean std var data=proj.cd4;
var logcd4;
class treatment time;
run;
                                /*figure 1*/

*mean plot: Means across visit by treatment;
proc sgplot data=proj.cd4;
vline time/response=logcd4 group=treatment stat=mean
lineattrs=(thickness=2);
run;

                                /*figure 2*/

*mean plot: Means across visit by gender;
proc sgplot data=proj.cd4;
vline time/response=logcd4 group=gender stat=mean
lineattrs=(thickness=2);
run;

                                /*figure 3*/
```

```
*scatter plot age;
proc sgplot data=proj.cd4;
scatter x=age y=logcd4 /group=treatment; run;
```

/*table 2*/

```
/*fit models*/
*M1 unstructured covariance aic 12110.7;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*treatment/s chisq;
random intercept week/subject=id type=un g gcorr;
run;

*M2 compound symmetry aic 13722.8;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*treatment/s chisq;
random intercept week/subject=id type=cs g gcorr;
run;

*M3 heterogeneous compound symmetry aic 12110.7;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*treatment/s chisq;
random intercept week/subject=id type=csh g gcorr;
run;

*M4 autoregressive aic 13722.8;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*treatment/s chisq;
random intercept week/subject=id type=AR(1) g gcorr;
run;

*M5 heterogeneous autogressive aic 12110.7;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*treatment/s chisq;
random intercept week/subject=id type=ARH(1) g gcorr;
run;
```

/*table 3*/

```
*Mixed Effects Model for log CD4 adjusting for Age and Gender;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*treatment/s chisq
outpred=proj.yhat;
random intercept week/subject=id type=un g gcorr;
run;
```

/*figure 4*/

```
*plot predicted vs.actual;
proc sgplot data=proj.yhat;
vline time/response=Pred group=treatment stat=mean lineattrs=(thickness=2)
markers markerattrs=(symbol=circle)name="predict";
vline time/response=logcd4 group=treatment stat=mean
lineattrs=(thickness=2)
markers markerattrs=(symbol=circlefilled)
```

```
lineattrs=(pattern=dot) name="actual";
keylegend "actual" / title="ACT" position=bottom noborder;
keylegend "predict" / title="PRD" position=bottom noborder;
run;

                                    /*table 4 5*/
*week^2;
proc mixed data=proj.cd4 method=reml noclprint=10 covtest;
class id gender treatment;
model logcd4=week treatment age gender week*week week*treatment
week*week*treatment/s chisq outpred=proj.yhat2;
random intercept week/subject=id type=un g gcorr;
run;

                                    /*figure 5*/
*plot predicted vs.actual;
proc sgplot data=proj.yhat2;
vline time/response=Pred group=treatment stat=mean lineattrs=(thickness=2)
markers markerattrs=(symbol=circle)name="predict";
vline time/response=logcd4 group=treatment stat=mean
lineattrs=(thickness=2)
markers markerattrs=(symbol=circlefilled)
lineattrs=(pattern=dot) name="actual";
keylegend "actual" / title="ACT" position=bottom noborder;
keylegend "predict" / title="PRD" position=bottom noborder;
run;


ods rtf close;
```