Analysis of Clustered and Longitudinal Data (2020-21)

Relationship between NO and Flow

Project 2.3.

By Dereje Mengist Belete

June 11, 2021

1. Descriptive Analysis

The dataset contains a few missing observations, and it can be seen from the plot that the missingness is due to attrition or dropout (appendix). The variability in the data is both explained by different children having different no of concentration overall, as well as by different growth rates. The variability in the Nitric Oxide (NO) concentration decrease with increasing flow of exhalation. This is suggestive of the presence of so-called 'random intercepts' (overall differences between subjects across different flow of exhalation and the presence of 'random slopes' (differences in growth rate between subjects). The variation in asthma group is higher compared to healthy group. Further, the smoother plot indicates that a decrease pattern in the concentration of NO for both groups as the flow of exhalation increase. However, we can't determine if one group has higher NO concentration than the other from the plot, they vary across different measurement.

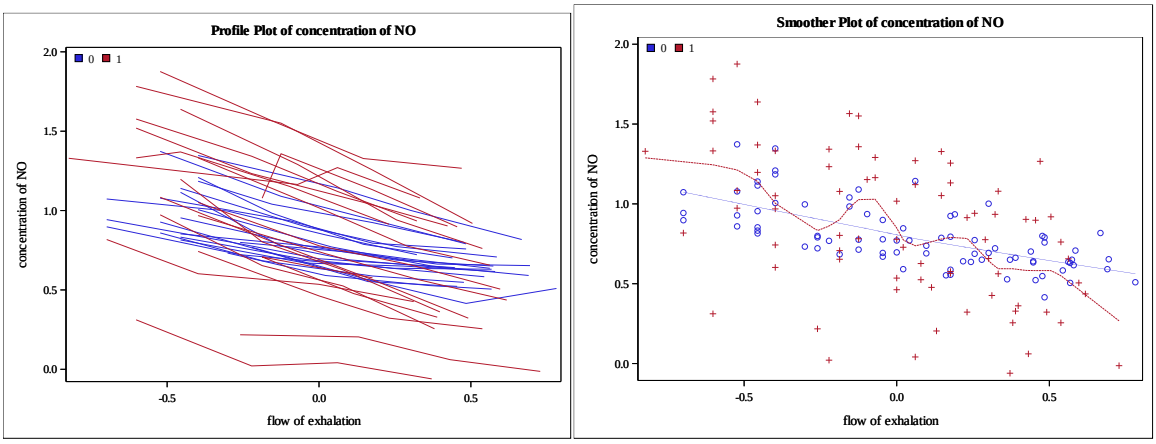


Fig 1. profile plot and Loess smoothers of concentration of NO

2. Model Building

To explore the concentration of NO exhaled over time, the model was built by fitting a saturated linear mixed model for the mean response that comprised the major effects of flow, diagnose, and the interaction between flow and diagnose. The model was built using the entire dataset, which included both treatment groups. I only evaluate UN (Unstructured), CS (Compound Symmetry), and Heterogeneous Compound Symmetry (CSH) as feasible covariance model choices since the time intervals in our data are not equal. As it can be seen from table 1, the AIC for the model with the UN covariance pattern was the smallest, in addition to this the results of the Likelihood ratio

test comparing UN vs CS and UN vs CSH revealed there is sufficient evidence to reject the null hypothesis of reduced model with p<0.001 for both model comparisons.

Table 1. Model Comparison (left) and type 3 test (right)

Description	CS	CSH	UN
-2 Res Log Likelihood	-158.5	-176.8	-244.1
AIC (Smaller is Better)	-154.5	-166.8	-224.1
AICC (Smaller is Better)	-154.4	-166.4	-222.5
BIC (Smaller is Better)	-151.2	-158.5	-207.5

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
diagnose	1	37	0.18	0.6756
flow	1	37	279.13	<.0001
flow*diagnose	1	37	18.85	0.0001

The main impact of treatment (diagnose) is not statistically significant in the model with a p-value of 0.67 at the 5% significance level, but the impact of flow and the interaction term are both statistically significant with a p<0.0001. Since the covariance structure was different amongst children in two group, the diagnosis variable was not removed from the model.

Further, in order to check whether the covariance structure does not differ between the two groups likelihood ratio tests was used. The result showed that there is sufficient evidence to reject the null hypothesis and it can be deduced that the covariance structure differs between two group (LR=29.7, df=3, p<0.001). In addition, a likelihood ratio test was used to compare a model with just the intercept vs. a model with both the intercept and the slope, with the null hypothesis stating the model with only the random intercept is appropriate and the alternative stating that the model with both intercept and slope is appropriate. There is sufficient evidence to reject the null hypothesis and conclude that the model with random intercept and random slope better fits the data structure with p <0.001.

3. Random Effect Model for Asthma Group

In order to fit a model with random intercept and slope for the asthma group alone, new dataset that has only children with asthma were created. So, this model is fitted only for children with asthma and the model doesn't include the healthy group. Also, the treatment variable (diagnose with asthma children) wasn't included in the model.

Table 2 Fixed Effect Model and Covariance Parameter for Asthma Group

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	0.8518	0.09065	18	9.40	<.0001
flow	-0.5849	0.04855	18	-12.05	<.0001

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
UN(1,1)	num	0.1549
UN(2,1)	num	-0.02879
UN(2,2)	num	0.03525
Residual	num	0.004468

The estimated variance of random intercept is $g_{11} = var(b_1) = UN(1,1) = 0.155$

The estimated variance of random slope is $g_{22} = var(b_2) = UN(2,2) = 0.035$

The Conditional model for asthma group is

$$E\left[\log(NO_{ij}) \middle| \log(flow_{ij}), b1_i, b2_i\right] = \overline{\beta}_1 + \overline{\beta}_2 \log(flow_{ij}) + b_1 i + b_2(logflow_{ij})$$

The mean intercept is equal to $= \overline{\beta}_1 + b_1 i = 0.85$

The mean slope is equal to $= (\overline{\beta}_2 + b_2) logflow_{ij} = -0.5849$

The 95% confidence interval for the random intercept variance of asthma group can be calculated as $\overline{\beta}_1 \pm 1.96 * \text{sqrt}(g_{11})$=0.85±1.96*0.39=(0.078, 1.62), which implies that for 95% of children with asthma, the average nitric oxide concentration at baseline measurement varies between (0.078, 1.62). The 95% confidence interval for the random slope variance of asthma group can be calculated as $\overline{\beta}_2 \pm 1.96 * \text{sqrt}(g_{22})$=-0.58±1.96*0.18= (-0.94, 0.21), which implies for 95% of asthma children, the average flow change in concentration of nitric oxide varies between (-0.94, 0.21)

4. Random Effect Model for Healthy Children

Similarly, a random intercept and slope for the only healthy group were fitted separately, it doesn't include the asthma group.

Table 3 Fixed Effect Model and Covariance Parameter for Healthy Group

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	0.8108	0.03203	19	25.31	<.0001
flow	-0.3455	0.02798	19	-12.35	<.0001

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
UN(1,1)	num	0.01987
UN(2,1)	num	-0.01238
UN(2,2)	num	0.01124
Residual	num	0.002456

The estimated variance of random intercept for healthy chideren is $g_{11} = var(b_1) = UN(1,1) = 0.019$
The estimated variance of random slope healthy chideren is $g_{22} = var(b_2) = UN(2,2) = 0.011$

$The\ Conditional\ model\ is\ E[NO_{ij}|flow_{ij}, b1_i, b2_i] = \overline{\beta}_1 + \overline{\beta}_2 \log(flow_{ij}) + b_1 i + b_2 \log(flow_{ij})$

The mean intercept is equal to $= \overline{\beta}_1 + b_1 i = 0.81$
The mean slope is equal to $= (\overline{\beta}_2 + b_2) logflow_i = -0.34$

The 95% confidence interval for the random intercept variance of healthy group can be calculated as $\overline{\beta}_1 \pm 1.96 *$sqrt($g_{11}$)=0.81±1.96*0.13= (0.54, 1.08), which implies for 95% of healthy children, the average nitric oxide concentration at baseline measurement varies between (0.54, 1.08).
The 95% confidence interval for the random slope variance of healthy group can be calculated as $\overline{\beta}_2 \pm 1.96 *$sqrt($g_{22}$)=-0.34±1.96*0.1= (-0.54, -0.13), which implies for 95% of healthy children, the average flow change in concentration of nitric oxide varies between (-0.54, -0.13).

5. Random effect Model for both Group

In order to fit a model for the two groups combined, allowing a completely different covariance structure for both groups I have used the full dataset with all observation. Furthermore, likelihood ratio tests were conducted to see if the covariance structure differed between the two groups. The results demonstrated that there is enough evidence to reject the null hypothesis, and that the covariance structure differs between the two groups (LR=29.7, df=3, p<0.001).

The estimated variance of random intercept for healthy group is g_{11}

$$= var(b_1) = UN(1,1) = 0.019$$

The estimated variance of random slope is healthy group g_{22}

$$= var(b_2) = UN(2,2) = 0.0089$$

The estimated variance of random intercept for asthma group is g_{11}

$$= var(b_1) = UN(1,1) = 0.15$$

The estimated variance of random slope is asthma group g_{22}

$$= var(b_2) = UN(2,2) = 0.036$$

The residual or within-subject is 0.0035

Covariance Parameter Estimates			
Cov Parm	**Subject**	**Group**	**Estimate**
UN(1,1)	num	diagnose 0	0.01967
UN(2,1)	num	diagnose 0	-0.01217
UN(2,2)	num	diagnose 0	0.008941
UN(1,1)	num	diagnose 1	0.1554
UN(2,1)	num	diagnose 1	-0.02856
UN(2,2)	num	diagnose 1	0.03696
Residual	num		0.003471

Table 4. Fixed Effect Model for both Group

Solution for Fixed Effects									
Effect	**diagnose**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**
Intercept		0.8518	0.09072	37	9.39	<.0001	0.05	0.6680	1.0356
diagnose	0 (healthy)	-0.04058	0.09623	37	-0.42	0.6756	0.05	-0.2356	0.1544
diagnose	1 (asthma)	0	.	.	.	.	.	.	.
flow		-0.5854	0.04835	37	-12.11	<.0001	0.05	-0.6834	-0.4875
flow*diagnose	0 (healthy)	0.2415	0.05563	37	4.34	0.0001	0.05	0.1288	0.3542
flow*diagnose	1 (asthma)	0	.	.	.	.	.	.	.

The interaction effect estimates 0.2415 (with 95% confidence interval 0.1288, 0.3542) suggests that the average nitric oxide per target values of the flow of exhalation is 0.2415 higher for healthier children than children with asthma. However, the increase or decrease is in logarithm form.

The marginal mean model is

$$E\left[\log(NO_{ij}) \middle| \log(flow_{ij}), diagnose_i\right] = \overline{\beta}_1 + \overline{\beta}_2 \log(flow_{ij}) + \overline{\beta}_3 diagnose_i + \overline{\beta}_4 \log(flow_{ij}) * diagnose_i$$

For healthy group

$$E\left[\log(NO_{ij}) \middle| \log(flow_{ij}), diagnose_i\right] = \overline{\beta}_1 + \overline{\beta}_2 \log(flow_{ij}) + \overline{\beta}_3 diagnose_i + \overline{\beta}_4 \log(flow_{ij}) * diagnose_i$$

$$= \overline{\beta}_1 + \overline{\beta}_3 + (\overline{\beta}_2 + \overline{\beta}_4) \log(flow_{ij})$$

The mean intercept is $\overline{\beta}_1 + \overline{\beta}_3$=0.8518-0.585=0.267 and mean slope is $(\overline{\beta}_2 + \overline{\beta}_4)$ = (-0.041-0.5854)=-0.6264

For asthma group

$$E\left[\log\left(NO_{ij}\right)\middle|\log\left(flow_{ij}\right), diagnose_i\right] = \overline{\beta}_1 + \overline{\beta}_2 \log\left(flow_{ij}\right)$$

The mean intercept is $\overline{\beta}_1$ =0.8518 and the mean slope is $\overline{\beta}_2$ =-0.041

The conditional mean model is

$$E\left[\log\left(NO_{ij}\right)\middle|\log\left(flow_{ij}\right), diagnose_i, b_{1i}, b_{2i}\right] = \overline{\beta}_1 + \overline{\beta}_2 \log\left(flow_{ij}\right) + \overline{\beta}_3 diagnose_i + \overline{\beta}_4 \log\left(flow_{ij}\right) * diagnose_i + b_{1i} + b_{2i} \log\left(flow_{ij}\right)$$

For Asthma group

$$E\left[\log\left(NO_{ij}\right)\middle|\log\left(flow_{ij}\right), diagnose_i, b_{1i}, b_{2i}\right] = \overline{\beta}_1 + \overline{\beta}_3 + b_{1i} + (\overline{\beta}_2 + \overline{\beta}_4 + b_{2i})\log\left(flow_{ij}\right)$$

The conditional mean intercept is $\overline{\beta}_1 + \overline{\beta}_3 + b_{1i}$ *where* b_{1i} *is the deviation of ith individual intercept from the population intercept* $B_1 + B_3$

The conditional mean slope is $\overline{\beta}_2 + \overline{\beta}_4 + b_{2i}$ *where* b_{2i} *is the deviation of ith individual slope from the population slope* $B_2 + B_4$

For Healthy group

$$E\left[\log\left(NO_{ij}\right)\middle|\log\left(flow_{ij}\right), diagnose_i, b_{1i}, b_{2i}\right] = \overline{\beta}_1 + b_{1i} + (\overline{\beta}_2 + b_{2i})\log\left(flow_{ij}\right)$$

The conditional mean intercept is $\overline{\beta}_1 + b_{1i}$ *and condtional mean slope* $\overline{\beta}_2 + b_{2i}$

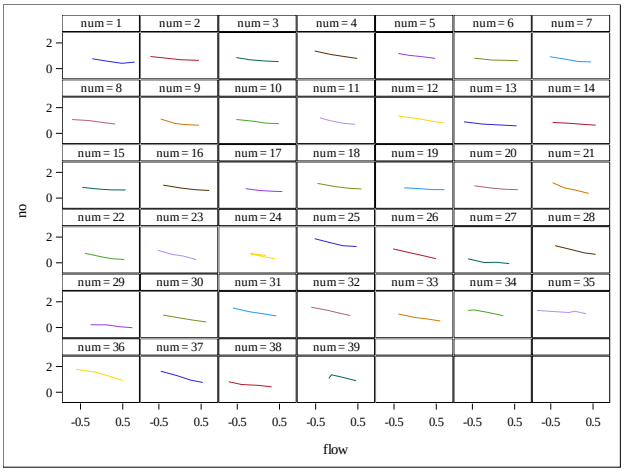
The 95% confidence interval for the random intercept variance of healthy group can be calculated as $\overline{\beta}_1 + \overline{\beta}_3 \pm 1.96 * \text{sqrt}(g_{11})$= (-0.003, 0.53), which implies for 95% of healthy children, it can be inferred that the average nitric oxide concentration at baseline measurement varies between (-0.003, 0.53).

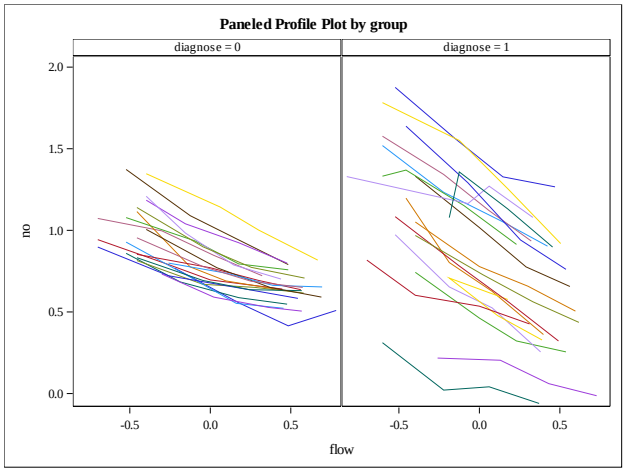
The 95% confidence interval for the random slope variance of healthy group can be calculated as $\overline{\beta}_2 + \overline{\beta}_4 \pm 1.96 * \text{sqrt}(g_{22})$ = (0.43, -0.13), which implies for 95% of healthy children, the average flow change in concentration of nitric oxide varies between (0.43, -0.13).

Similarly, the 95% confidence interval for the random intercept variance of asthma group can be calculated as $\overline{\beta}_1 \pm 1.96 * \text{sqrt}g_{11}$=(0.2, 1.5), which implies for 95% of children with asthma, the average nitric oxide concentration at baseline measurement varies between (0.2, 1.5).

The 95% confidence interval for the random slope variance of asthma group can be calculated as $\overline{\beta}_2 \pm 1.96 *$sqrt($g_{22}$)= (-0.41, 0.32), which implies for 95% of asthma children, the average flow change in concentration of nitric oxide varies between (-0.41, 0.32).

Appendix


num = 1
num = 2
num = 3
num = 4
num = 5
num = 6
num = 7
num = 8
num = 9
num = 10
num = 11
num = 12
num = 13
num = 14
num = 15
num = 16
num = 17
num = 18
num = 19
num = 20
num = 21
num = 22
num = 23
num = 24
num = 25
num = 26
num = 27
num = 28
num = 29
num = 30
num = 31
num = 32
num = 33
num = 34
num = 35
num = 36
num = 37
num = 38
num = 39
no
flow


Paneled Profile Plot by group
diagnose = 0
diagnose = 1
no
flow

```sas
LIBNAME ACLD 'H:/2020-21/ACLD/Project2';

data ACLD.new;
    set ACLD.no;
    cflow=flow;
run;

proc print data=ACLD.new;
run;

*Separate dataset for the healthy and asthma group;

DATA OnlyHealthy;
    SET ACLD.new;

    IF (diagnose=1) then
        delete;
run;

DATA OnlyAsthma;
    SET ACLD.new;

    IF (diagnose=0) then
        delete;
run;

proc print DATA=OnlyHealthy;
run;

proc print DATA=OnlyAsthma;
run;

title 'Profile Plot of concentration of NO';

PROC SGPLOT DATA=ACLD.new NOAUTOLEGEND;
    format diagnose diagnosefmt.;
    SERIES X=flow Y=no / GROUP=num GROUPLC=diagnose NAME='diagnose' BREAK
        LINEATTRS=(PATTERN=1);
    LABEL no='concentration of NO' flow='flow of exhalation';
    KEYLEGEND 'diagnose' / NOBORDER TYPE=linecolor LOCATION=inside;
RUN;

title;
*profile Plot by group;
title 'Paneled Profile Plot by group';

proc sgpanel data=ACLD.new noautolegend;
    panelby diagnose / columns=2;
    series X=flow Y=no / group=num break lineattrs=(pattern=1);
    LABEL pi='concentration of NO' time='flow of exhalation';
run;

title;
*smoother plot;
title 'Smoother Plot of concentration of NO ';

PROC SGPLOT DATA=ACLD.new;
    LOESS X=flow Y=no / GROUP=diagnose NAME='diagnose';
    LABEL no='concentration of NO' flow='flow of exhalation';
    KEYLEGEND 'diagnose' / NOBORDER TYPE=linecolor LOCATION=inside;
RUN;

```

```
62
63  title;
64  *checking missing data;
65
66  proc sgpanel data=ACLD.new noautolegend;
67      panelby num / columns=7 rows=6;
68      series x=flow y=no / group=num break lineattrs=(pattern=1);
69      label day="flow of exhalation" y="concentration of NO";
70  run;
71
72  ***********
73  Model building covariance structure
74  *********;
75  ************************************************************************
76
77  *unstructured;
78
79  PROC MIXED DATA=ACLD.new;
80      CLASS num;
81      MODEL no=diagnose flow flow*diagnose /s chisq;
82      REPEATED / TYPE=UN SUBJECT=num;
83      ODS OUTPUT Fitstatistics=fit_un(rename=(value=UN));
84      ODS OUTPUT Dimensions=Parm_UN(rename=(value=Num_UN));
85  RUN;
86
87  * Compund symmetry;
88
89  PROC MIXED DATA=ACLD.new;
90      CLASS num;
91      MODEL no=diagnose flow flow*diagnose /s chisq;
92      REPEATED / TYPE=CS SUBJECT=num;
93      ODS OUTPUT Fitstatistics=fit_CS(rename=(value=CS));
94      ODS OUTPUT Dimensions=Parm_CS(rename=(value=Num_CS));
95  RUN;
96
97  * Heterogeneous Compound Symmetry;
98
99  PROC MIXED DATA=ACLD.new;
100     CLASS num;
101     MODEL no=diagnose flow flow*diagnose /s chisq;
102     REPEATED / TYPE=CSH SUBJECT=num;
103     ODS OUTPUT Fitstatistics=fit_CSH(rename=(value=CSH));
104     ODS OUTPUT Dimensions=Parm_CSH(rename=(value=Num_CSH));
105 RUN;
106
107 * combine all loglikelihood, AIC and BIC between models and compare;
108
109 Data compare;
110     merge Fit_CS Fit_CSH Fit_UN;
111 run;
112
113 title 'Output Summary comparing the models';
114
115 proc print data=compare label noobs;
116 run;
117
118 title;
119 *From the AIC's UN does best compared to the others;
120 * Just a review of the number of parameter estimate per model;
121
122 Data all_parmnr;
        merge Parm_UN Parm_CS Parm_CSH;
123
```

```
124 run;
125
126 title "Number of parameter estimate per model";
127
128 proc print data=all_parmnr (obs=1) label noobs;
129 run;
130
131 title;
132
133 /* Now we construct LRT tests for nested models to see what
134 model does better; */
135 *a. Unstructured (full) vs compound symmetry (reduced);
136
137 data UN_CS;
138     merge Fit_UN Fit_CS Parm_UN Parm_CS;
139
140     if _n_=1 then
141         do;
142             Chi_UN_CS=CS - UN;
143             df_UN_CS=Num_UN - Num_CS;
144             p_UN_CS=1-probchi(Chi_UN_CS, df_UN_CS);
145             output;
146             stop;
147         end;
148 run;
149
150 title 'Likelihood Ratio Test: Unstructured vs Compound symmetry';
151
152 proc print data=UN_CS label noobs;
153     var UN CS Chi_UN_CS df_UN_CS p_UN_CS;
154     label UN="-2 loglik UN" CS="-2 loglik CS" Chi_UN_CS="Chi-Square" df_UN_CS="DF"
155         p_UN_CS="Pr > ChiSq";
156 run;
157
158 title;
159 *b. Unstructured (full) vs heterogeneous compound symmetry (reduced);
160
161 data UN_CSH;
162     merge Fit_UN Fit_CSH Parm_UN Parm_CSH;
163
164     if _n_=1 then
165         do;
166             Chi_UN_CSH=CSH - UN;
167             df_UN_CSH=Num_UN - Num_CSH;
168             p_UN_CSH=1-probchi(Chi_UN_CSH, df_UN_CSH);
169             output;
170             stop;
171         end;
172 run;
173
174 title 'Likelihood Ratio Test: Unstructured vs Hetero Compound symmetry';
175
176 proc print data=UN_CSH label noobs;
177     var UN CSH Chi_UN_CSH df_UN_CSH p_UN_CSH;
178     label UN="-2 loglik UN" CSH="-2 loglik CSH" Chi_UN_CSH="Chi-Square"
179         df_UN_CSH="DF" p_UN_CSH="Pr > ChiSq";
180 run;
181
182 title;
183 ******************************************************************
184  combine all loglikelihood, AIC and BIC between models and compare;
185
```

```
186 Data compare;
187     merge Fit_CS Fit_CSH Fit_UN;
188 run;
189 
190 title 'Table 1. Output Summary comparing the models';
191 
192 proc print data=compare label noobs;
193 run;
194 
195 title;
196 * testing if covariance structure for treatment groups differ or not ;
197 
198 proc mixed data=ACLD.new;
199     class num  diagnose ;
200     MODEL no=diagnose flow flow*diagnose /s chisq;
201     random Intercept flow / SUBJECT=num G GCORR TYPE=un V VCORR GROUP=diagnose;
202     repeated / type=simple subject=num;
203     ods output Fitstatistics=fit_f;
204 run;
205 
206 *reduced;
207 
208 proc mixed data=ACLD.new;
209     class num  diagnose ;
210     MODEL no=diagnose flow flow*diagnose /s chisq;
211     random Intercept flow / SUBJECT=num G GCORR TYPE=un V VCORR;
212     repeated / type=simple subject=num;
213     ods output Fitstatistics=fit_r;
214 run;
215     *LR test  use REML, Ho: no different covariance structure for treatment groups ;
216 
217 data lrt;
218     set fit_f(obs=1);
219     if Descr='-2 Res Log Likelihood' then
220         lrf=Value;
221     set fit_r(obs=1);
222     if Descr='-2 Res Log Likelihood' then
223         lrr=Value;
224     lr=lrr-lrf;
225     df=3;
226     pvalue=1-probchi(lr, df);
227     keep lr df pvalue;
228 run;
229 
230 proc print data=lrt;
231     title 'LR Test';
232 run;
233 
234 *LR test Ho: H0: random intercepts only vs
235 H1: random intercepts and random slopes;
236 
237 proc mixed data=ACLD.new;
238     class num  diagnose;
239     model no=diagnose flow flow*diagnose/s;
240     random Intercept/ SUBJECT = num G GCORR TYPE= un V VCORR GROUP = diagnose;
241     repeated / type=simple subject=num;
242     ods output Fitstatistics=fit1(rename=(value=Random_int));
243 run;
244 
245 proc mixed data=ACLD.new;
246     class num  diagnose;
    model no=diagnose flow flow*diagnose/s;
247 
```

```
248     random Intercept flow/ SUBJECT = num G GCORR TYPE= un V VCORR GROUP = diagnose;
249     repeated / type=simple subject=num;
250     ods output Fitstatistics=fit2(rename=(value=Random_intslope));
251 run;
252 
253 data lrt;
254 set fit1(obs=1);
255 if Descr='-2 Res Log Likelihood' then lrr=Random_int;
256 set fit2(obs=1);
257 if Descr='-2 Res Log Likelihood' then lrf=Random_intslope;
258 lr=lrr-lrf;
259 df1=3;
260 df2=7;
261 df=df2-df1;
262 p1=1-probchi(lr,df1);
263 p2=1-probchi(lr,df2);
264 p=0.5*p1+0.5*p2;;
265 keep lr df p;
266 run;
267 proc print data=lrt;
268 title 'LR Test';
269 run;
270 
271 *Final Model;
272 
273 proc mixed data=ACLD.new;
274     class num  diagnose;
275     model no=diagnose flow flow*diagnose/s;
276     random Intercept flow/ SUBJECT = num G GCORR TYPE= un V VCORR GROUP = diagnose;
277     repeated / type=simple subject=num;
278 run;
279 
280 *Random effect model for healthy group;
281 
282 proc mixed data=OnlyHealthy;
283     class num;
284     model no=flow  / s;
285     random intercept flow / type=un subject=num g gcorr v=2 vcorr=2;
286     repeated / type=simple subject=num r=2;
287 run;
288 
289 *Random effect model for Asthma group;
290 proc mixed data=OnlyAsthma;
291     class num;
292     model no=flow  / s;
293     random intercept flow / type=un subject=num g gcorr v=2 vcorr=2;
294     repeated / type=simple subject=num r=2;
295 run;
296 
297 
298 * model for both healthy and asthma groups;
299 
300 proc mixed data=ACLD.new;
301     class num  diagnose;
302     model no=diagnose flow flow*diagnose/s cl;
303     random Intercept flow/ SUBJECT = num g gcorr type= un v vcorr group = diagnose;
304     repeated / type=simple subject=num;
```