

Using Model Fitting Information and PROC PLOT to Select the Type of Variance- Covariance Structure for Repeated Measures

Radhi Abdulnabi, Clinical Biostatistics, Pfizer, Inc
Ann Arbor, MI

ABSTRACT

The statistical analysis of repeated measures data must address the issue of covariation between measures on the same unit, because it is usually plausible to assume that observations on the same unit are correlated. The selection of an appropriate type of covariance structure can substantially improve the efficiency of inferences made about the mean structure and provide better estimates of standard errors of estimated mean parameters.

The REPEATED statement in the MIXED procedure provides a rich selection of covariance structures. Selecting the appropriate covariance structure is a major hurdle in the use of PROC MIXED.

One of the methods used to choose an appropriate covariance matrix is using Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC) to compare models with the same fixed effects but different covariance structures.

In the paper we will try to use PROC PLOT and SAS library macros, to run one program, include the same model with many types of covariance structure, and then compare the values of AIC and BIC for each structure, also we will show the differences between SAS 6.12 and SAS 8.2 when we applied this method.

INTRODUCTION

Repeated measures are response outcomes measured on the same experimental unit. Usually, these measurements are made over a period of time, such as pain score measured once a day for a week. The experimental units are most often called subjects. Time is called a within-subject effect because there are different measurements at different times on the same subject. Explanatory variables such as treatment or sex are called between-subject effects because their values change only from subject to subject; there is not a different value for them at different times for the same subject. Questions we may want to answer in a repeated measures analysis may be whether treatment influences the response, whether time, the repeated factor, influences the response, and whether there is a treatment by time interaction.

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.

One of the problems, which face the statisticians and researchers when they try to apply the repeated measures analysis, is the selection of variance covariance structure. Until recently, analysis techniques available in computer software only offered the user limited and inadequate choices. One choice was to ignore covariance structure and make invalid assumptions. Another was to avoid the covariance structure issue by analyzing transformed data or making adjustments to otherwise inadequate analyses. Ignoring covariance structure may result in erroneous inference, and avoiding it may result in inefficient inference.

Recently available mixed model methodology permits the covariance structure to be incorporated into the statistical model. The MIXED procedure of the SAS System provides a rich selection of covariance structures through the RANDOM and

REPEATED statements. Modeling the covariance structure is a major hurdle in the use of PROC MIXED.

The goal of this paper is to present one of the methods used to select the appropriate covariance structure depend on model fitting information results using SAS library macros, also to show the effects of using different covariance structure on the final results, and to highlight the differences between SAS 6.12 and SAS 8.2 when we applied this method.

REPEATED MEASURES AND LONGITUDINAL DATA

Repeated measures are response outcomes measured on the same subject. Most of the times, these measurements are made over a period of time, such as pain score measured once a day for a week. However, repeated measures can also refer to multiple measurements on an experimental unit under different conditions or in different places.

Time is called a within-subject effect because there are different measurements at different times on the same subject.

Explanatory variables such as race or sex are called between-subject effects because their values change only from subject to subject; there is not a different value for them at different times for the same subject.

When a response variable of interest is measured repeatedly over time or space from the same subject, longitudinal data arises. Since the repeated measures on the same subject are correlated, longitudinal data analysis ought to specify covariation among measures on the same subject.

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 same covariance structure is used for all subjects.

Questions we may want to answer in a repeated measures analysis may be whether treatment influences the response, whether time, the repeated factor, influences the response, and whether there is a treatment by time interaction. This task uses the mixed models approach for analyzing repeated measures.

THE GENERAL LINEAR MIXED MODEL

D. A. Harville defined the general linear mixed model (GLMM) as:

$$Y_i = X_i\beta + Z_i\gamma + \epsilon_i$$

where

Y_i represents the observed data for subject i , $1 \leq i \leq N$, N is the number of subjects.

X_i and Z_i are $(n \times p)$ and $(n \times q)$ the known design matrices

β represents the fixed effects

γ represents the random effects

ϵ_i represents the residual error

The mixed model includes both fixed-effects parameters, β , and random-effects parameters, γ .

It extends the general linear model by allowing a more flexible specification of the covariance structure of ϵ .

Assuming that γ and ϵ are independent, and are normally distributed with

$$E \begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

So the variance of y is, $\text{Var}(y) = \mathbf{ZGZ}' + \mathbf{R}$ where

- $\mathbf{G} = \text{Var}(\gamma)$
- $\mathbf{R} = \text{Var}(\epsilon)$, representing the within-patient portion of the covariance structure
- $\mathbf{ZGZ}'$ represents the between-patient portion

Covariance Structures

$\text{Var}(y)$ is modeled by creating the random-effects design matrix $\mathbf{Z}$ and by stipulating covariance structures for $\mathbf{G}$ and $\mathbf{R}$.

Functional specification of the covariance structure in the mixed model is made through $\mathbf{G}$ and $\mathbf{R}$, frequently in terms of $\mathbf{R}_{ij}$.

Covariance structure models hypothesize that a covariance matrix has a particular form. For example, you can test the hypothesis that a set of variables all have equal variances with this procedure, that's mean the structure, which assumes the same covariance between any two measurements for the same subject and the same variance for each measurement called *compound symmetry*. Often, it seems reasonable that the correlations between measurements close together in time are higher than that for measurements further apart in time. A covariance structure that has this property is called *first order autoregressive*, or *AR(1)*. Another possible covariance structure is called *unstructured*, which means that no particular structure is placed on the covariances and variances. The SAS PROC MIXED procedure supplies many options of covariance structures through the RANDOM and REPEATED statements. For a longitudinal data, modeling covariance structure, which entails modeling variation between subjects and covariation between measures at different time points on the same subject.

Some of the common covariance structures are presented in the following table.

Table 1: Features and Examples of Covariance Structures

Variance Components (TYPE=VC) / SAS default

$$\begin{bmatrix} \sigma_D^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AD}^2 \end{bmatrix}$$

Compound Symmetry (TYPE=CS)

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

Unstructured (TYPE=UN)

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

First-Order Autoregressive (TYPE=AR(1))

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Toeplitz (TYPE=TOEP)

$$\begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix}$$

First-Order Ante-dependence (TYPE=ANTE(1))

$$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 \\ \sigma_1\sigma_1\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 \\ \sigma_3\sigma_1\rho_1\rho_1 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 \end{bmatrix}$$

There are another structures designed for multivariate repeated measures called direct product or (Kronecker product). These structures are constructed by taking the Kronecker product of an unstructured matrix (modeling covariance across the multivariate observations) with an additional covariance matrix (modeling covariance across time or another factor). The upper left value in the second matrix is constrained to equal 1 to identify the model. One of these structures is:

Direct Product AR(1)

[TYPE=UN@AR(1)]

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix} =$$

$$\begin{bmatrix} \sigma_1^2 & \sigma_1^2\rho & \sigma_1^2\rho^2 & \sigma_{21} & \sigma_{21}\rho & \sigma_{21}\rho^2 \\ \sigma_1^2\rho & \sigma_1^2 & \sigma_1^2\rho & \sigma_{21}\rho & \sigma_{21} & \sigma_{21}\rho \\ \sigma_1^2\rho^2 & \sigma_1^2\rho & \sigma_1^2 & \sigma_{21}\rho^2 & \sigma_{21}\rho & \sigma_{21} \\ \sigma_{21} & \sigma_{21}\rho & \sigma_{21}\rho^2 & \sigma_2^2 & \sigma_2^2\rho & \sigma_2^2\rho^2 \\ \sigma_{21}\rho & \sigma_{21} & \sigma_{21}\rho & \sigma_2^2\rho & \sigma_2^2 & \sigma_2^2\rho \\ \sigma_{21}\rho^2 & \sigma_{21}\rho & \sigma_{21} & \sigma_2^2\rho^2 & \sigma_2^2\rho & \sigma_2^2 \end{bmatrix}$$

To use these structures in the REPEATED statement, you must specify two distinct REPEATED effects, both of which must be included in the CLASS statement. The first effect indicates the multivariate observations, and the second identifies the levels of time or some additional factor. Note that the input data set must still be constructed in "univariate" format; that is, all dependent observations are still listed observation-wise in one single variable. Although this construction provides for general modeling possibilities, it forces you to construct variables indicating both dimensions of the Kronecker product.

For example, suppose your observed data consist of heights and weights of several children measured over several successive years. Your input data set should then contain variables similar to the following:

- Y, all of the heights and weights, with a separate observation for each

- Var, indicating whether the measurement is a height or a weight
- Year, indicating the year of measurement
- Child, indicating the child on which the measurement was taken

Your PROC MIXED code for a Kronecker AR(1) structure across years would then be

```
proc mixed;
  class Var Year Child;
  model Y = Var Year Var*Year;
  repeated Var Year /type=un@ar(1) subject=Child;
run;
```

You should nearly always want to model different means for the multivariate observations, hence the inclusion of Var in the MODEL statement. The preceding mean model consists of cell means for all combinations of VAR and YEAR. This is subtopic for the above. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

SELECTING AN APPROPRIATE COVARIANCE STRUCTURE FOR ANALYZING LONGITUDINAL DATA

Analysis of longitudinal data using a general linear mixed model requires the specification of a form for the covariance matrix of within-subject observations. The selection of an appropriate covariance structure can substantially improve the efficiency of inferences made about the mean structure and provide better estimates of standard errors of estimated mean parameters.

The problem is not a shortage of possible structures. For example REPEATED statement in SAS PROC MIXED, allows the user to select from more than 20 distinct covariance structures. Rather the problem is to select the appropriate structure to fit to the data.

The objectives of the paper are:

- Fit repeated measures data using Mixed Procedure
- Fit several within-subject variance-covariance matrices.
- Determine the appropriate variance-covariance structure.
- Evaluate the significance and parameters estimate of the fixed effects.

GOODNESS OF FIT

The Goodness of Fit explains how well a sample distribution matches a hypothetical or theoretical population distribution. The likelihood and log-likelihood functions are the basis for deriving estimators for parameters, given data; the likelihood of a set of data is the probability of obtaining that particular set of data given the chosen probability distribution model. This expression contains the unknown parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimates.

SAS version 8 and higher use the following fit statistics:

Log-Likelihood ratio:

Log (L) is the log-likelihood evaluated at the maximum likelihood estimates of the model parameters, and calculated as

$$L = 2[SO_i \ln(O_i / E_i)]$$

Where:

O_i is the Observed frequency

E_i is the Expected frequency

$\ln$ is the natural log function (log base e)

Degrees of freedom are $k-1$ (same as Chi Square)

L follows a Chi Squared Distribution

Akaike's Information Criterion (AIC)

The criterion is calculated as

$$L - K$$

Where:

L - log-likelihood

K - number of covariance parameters

Second-order Akaike's Information Criterion (AAIC)

Calculated as

$$L - K - (K(K+1))/(n - K - 1)$$

Where:

L - log-likelihood

K - number of covariance parameters

n - Sample size

Schwarz's Bayesian Criterion (BIC)

Calculated as

$$L - K/2 \log(N)$$

Where:

L - log-likelihood

K - number of covariance parameters

n - Sample size

To choose the appropriate variance-covariance structure we will try to develop the following method:

Using some of sphericity testing with MIXED procedure like Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC) to choose the best structure. We will try to use some of SAS Macros, which enable us to specify a variety of REPEATED statements also allow us to include as many as we need of variance-covariance structure in one program, and then display the AIC and BIC graphically using PLOTIT macro, which attempt to display points (the values of AIC across the values of BIC) and point label (the name of variance-covariance structure), and allows to select the best structure, which provide the best fit.

SAS PROGRAMS AND MACROS

We will use the following SAS code to test the sphericity using MIXED procedure:

```
%macro runmixed(method=,z=,comment=,out=, group=);
%let _print_=off;
  proc mixed data =&dsname &method;
    class &class;
    model &model;
    &z;
    make 'Fitting' out=fitinfo; * using SAS 6.12;
    or make 'FitStatistics' out=fitinfo; * using SAS 8.2;

  run;

  data fitinfo;
    length structr $20;
    set fitinfo;
    structr="&comment";
  run;

  proc append base=&out data=fitinfo force;
  run;

%let _print_=on;
%mend;
```

And we will use the following SAS statements to fit the model with several variance covariance structures:

```
%let dsname =temp1;
%let class = treatmen center week ptid;
%let model = endscore=treatmen center week0
week;
* first-order autoregressive;
```

```

%runmixed(method = method=rem1, z=repeated /
type = ar(1) subject=ptid, comment=ar1,
out=criteri);

* first-order autoregressive - different R for
each treatment group;
%runmixed(method =,z=repeated / type = arh(1)
subject=ptid, group=treatment, comment=arh1,
out=criteri);

* Compound Symmetry;
%runmixed(method =,z=repeated / type = cs
subject=ptid, comment=cs, out=criteri);

* Toeplitz;
%runmixed(method =,z=repeated / type = toep
subject=ptid, comment=toep, out=criteri);

* block diagonal Toeplitz - different R for each
treatment group;
%runmixed(method =,z=repeated / type = toeph
subject=ptid, group=treatment, comment=toeph,
out=criteri);

* Huynh Feldt;
%runmixed(method =,z=repeated / type = hf
subject=ptid, comment=hf, out=criteri);

* Heterogeneous Compound Symmetry;
%runmixed(method =,z=repeated / type = csh
subject=ptid, comment=csh, out=criteri);

* Unstructure;
%runmixed(method =,z=repeated / type = un
subject=ptid, comment=un, out=criteri);

* Ante-Dependence;
%runmixed(method = method=rem1, z=repeated /
type = ante(1) subject=ptid, comment=antel,
out=criteri);

* first-order autoregressive Moving-Average;
%runmixed(method =,z=repeated / type = arma(1,1)
subject=ptid, group=treatment, comment=armal1,
out=criteri);

* first-order Factor Analytic;
%runmixed(method =,z=repeated / type = fa(2)
subject=ptid, comment=fa2, out=criteri);

* No Diagonal Factor Analytic;
%runmixed(method =,z=repeated / type = fal(2)
subject=ptid, comment=fal2, out=criteri);

* Banded Toeplitz ;
%runmixed(method =,z=repeated / type = toep(2)
subject=ptid, group=treatment, comment=toep2,
out=criteri);

* Banded Heterogeneous Toeplitz;
%runmixed(method =,z=repeated / type = toeph(2)
subject=ptid, comment=toeph2, out=criteri);

* Banded Main Diagonal Unstructure;
%runmixed(method =,z=repeated / type = un(1)
subject=ptid, comment=un1, out=criteri);

* Unstructure Correlations;
%runmixed(method =,z=repeated / type = unr
subject=ptid, comment=unr, out=criteri);

* Banded Unstructure Correlations;
%runmixed(method =,z=repeated / type = unr(1)
subject=ptid, comment=unr1, out=criteri);

* Direct Product CS;
%runmixed(method =,z=repeated treatment week/
type = un@cs subject=ptid, comment=unacs,
out=criteri);

* Direct Product AR(1);
%runmixed(method =,z=repeated treatment week/
type = un@ar(1) subject=ptid, comment=unar1,
out=criteri);

```

```

* Direct Product UN;
%runmixed(method =,z=repeated treatment week/
type = un@un subject=ptid, comment=unun,
out=criteri);

```

```

* Variance Compound;
%runmixed(method =,z=repeated / type = vc
subject=ptid, comment=vc, out=criteri);

```

Then we sort the data and listing the values of AIC and BIC by structure

```

proc sort data = criteri out = aicbic;
  where descr = 'AIC' or descr = 'BIC';
  by descending value;
run;

```

```

proc print data = aicbic;
run;

```

```

data aicbic2;
set aicbic;
if descr='AICC (smaller is better)' then delete;
run;
proc print data = aicbic2;
run;

```

```

proc sort data = aicbic2;
  by structr descr value;
run;

```

```

data aicbic3;
set aicbic2;
by structr descr value;
if first.value;
run;
proc print data = aicbic3;
run;

```

```

proc transpose data = aicbic3 out=taicbic (drop=_name_ _label_
rename= (col1=aic col2=bic));
  var value;
  by structr;
run;

```

```

proc print data = taicbic;
run;

```

To find the effect of using different types of covariance structures on the final results of PROC MIXED, we run the following SAS macro codes:

```

%macro runmixed(method=,z=,comment=,out=, group=);
%let _print_=off;
  proc mixed data =&dsname &method;
    class &class;
    model &model;
    lsmeans rxgrp /pdiff=control cl adjust=dunnett;
    &z;
    make 'Diffs' out=fitinfo;
  run;

  data fitinfo;
    length structr $20;
    set fitinfo;
    structr="&comment";
  run;

  proc append base=&out data=fitinfo force;
  run;

```

```
%let _print_=on;
%mend;
```

THE RESULTS

The following tables show the final values of AIC and BIC when we apply the method on the available longitudinal data using SAS 6.12 and SAS 8.2

Table 1. The values of AIC and BIC (SBC)
SAS 6.12

OBS	STRUCTR	AIC	SBC
1	ante1	-2702.65	-2737.39
2	ar1	-2779.92	-2784.55
3	arh1	-2773.83	-2794.67
4	arma11	-2778.96	-2785.91
5	cs	-2936.84	-2941.48
6	cs	-2922.24	-2943.09
7	fa12	-2783.49	-2820.55
8	fa2	-2728.60	-2781.87
9	hf	-2922.18	-2943.02
10	toep	-2775.99	-2794.52
11	toep2	-3018.86	-3023.49
12	toeph	-2773.07	-2807.81
13	toeph2	-3019.35	-3040.20
14	un	-2679.14	-2762.52
15	un1	-3271.92	-3290.45
16	unr	-2679.14	-2762.52

Table 2. The values of AIC and BIC (SBC)
Output SAS 8.2

Obs	STRUCTR	AIC	BIC
1	ante1	5405.3	5444.1
2	ar1	5559.8	5565.0
3	arh1	5547.7	5570.9
4	arma11	5557.9	5565.7
5	cs	5873.7	5878.9
6	cs	5844.5	5867.8
7	fa12	5565.0	5603.8
8	fa2	5453.5	5507.8
9	hf	5844.4	5867.6
10	toep	5552.0	5572.7
11	toep2	6037.7	6042.9
12	toeph	5546.1	5584.9
13	toeph2	6038.7	6062.0
14	un	5358.3	5451.3
15	un1	6543.8	6564.5
16	unar1	5560.9	5579.0
17	uncs	5869.4	5887.5
18	unr	5358.3	5451.3
19	unr1	6543.8	6564.5
20	unun	5367.7	5473.7
21	vc	6552.7	6555.3

Also we have different results when we use LSMEANS statement to compare different groups table 3 shows the results using SAS 8.2.

Table 3. The differences between treatment group and placebo, stander deviation and P-value, using different covariance structures, SAS 8.2.

Structr	treatment	estimate	stderr	probt
ar1	TRT A	-10.0490	3.6334	0.0070
ar1	TRT B	-2.6521	3.7852	0.4855
Arh1	TRT A	-6.8693	3.3018	0.0406
arh1	TRT B	-0.6123	3.4395	0.8591
cs	TRT A	-9.8674	3.9422	0.0143
cs	TRT B	-2.5389	4.1082	0.5383
toep	TRT A	-10.0178	3.6991	0.0083
toep	TRT B	-2.4161	3.8532	0.5324
toeph	TRT A	-7.8164	3.4176	0.0248
toeph	TRT B	-1.0448	3.5595	0.7699
hf	TRT A	-8.7094	3.0222	0.0051
hf	TRT B	-3.2390	3.1433	0.3059
cs	TRT A	-10.8023	3.4042	0.0021
cs	TRT B	-6.4434	3.5463	0.0729
un	TRT A	-9.0193	3.0172	0.0037
un	TRT B	-3.3543	3.1435	0.2891
ante1	TRT A	-8.0276	2.9959	0.0089
ante1	TRT B	-2.2569	3.1221	0.4718
armall	TRT A	-10.1280	3.7364	0.0082
armall	TRT B	-2.7203	3.8927	0.4867
fa2	TRT A	-7.3523	2.9847	0.0159
fa2	TRT B	-1.5636	3.1103	0.6165
fa12	TRT A	-6.1222	2.6877	0.0254
fa12	TRT B	-0.9322	2.8006	0.7401
toep2	TRT A	-9.7261	1.7274	<.0001
toep2	TRT B	-2.4807	1.7988	0.1717
vc	TRT A	-9.8198	1.4973	<.0001
vc	TRT B	-2.5396	1.5589	0.1072

At the end we will display the AIC and BIC graphically using PLOTIT610 Macro for SAS 6.12 and PLOTIT7 Macro for SAS 8.2, which will help us to select the appropriate Variance-Covariance structure by selecting the structure with highest AIC and BIC for SAS 6.12 and the lowest values of AIC and BIC for SAS 8.2, see the Output of PLOTIT610 and PLOTIT7.

```
%include 'M:\SASPrograms\plotit610.sas';
%plotit(data=taicbic, plotvars= aic bic,labelvar=structr);

%include 'M:\SASPrograms\plotit7.sas';
%plotit(data=taicbic, plotvars= aic bic,labelvar=structr);
```

The following graphs show the covariance structures distributed by the value of AIC and BIC.

If we use SAS 6.12 the structure in the bottom left corner of the graph provide the poorest fit and the structure in the upper right corner provide the best fit. We have to select the opposite corner to have the best fit if we use SAS 8.2, because AIC and BIC values in SAS 8.2 were multiplied by -2.

Figure 1: covariance structures distributed by the value of AIC and BIC using SAS 6.12.

inferences made about the mean structure and provide better estimates of standard errors of estimated mean parameters. Modern statistical software enables the user to incorporate the covariance structure into the statistical model. The problem is not a shortage of possible structures. Rather the problem is to select the appropriate structure to fit to the data.

Phone: (734) 622-3768
Fax: (734) 622-1255
Radhi.Abdulnabi@pfizer.com

To choose the appropriate variance-covariance structure we have tried to use the following two of sphericity testing with MIXED procedure: Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC) to choose the best structure. Also we have used SAS Macros, which enable us to specify a variety of REPETED statements to allow us to include as many as we need of variance-covariance structure in one program, and then display the AIC and BIC graphically using PLOTIT macro, which attempt to display points (the values of AIC across the value of BIC) and point label (the name of variance-covariance structure), and allows to select the best structure, which provide the best fit. The structure in the bottom left corner of the graph provide the poorest fit and the structure in the upper right corner provide the best fit using SAS 6.12 and vice versa when we use SAS 8.2.

REFERENCES

1. Akaike H. A new look at the statistical model identification. IEEE Transaction on Automatic control 1974; AC-19:716-723.
2. Diggle PJ, Liang K-Y, Zeger SL. Analysis of Longitudinal Data. Oxford University Press: New York, 1994.
3. Harville, D.A. (1977), "Maximum likelihood approaches to variance component estimation and to related problems", Journal of the American Statistical Association 72, 320-340.
4. Milliken GA, Johnson DE. Analysis of Messy Data, Volume 1: Designed Experiments. Chapman and Hall: New York, 1992.
5. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. SAS System for Mixed Models. SAS Institute Inc.: Cary, NC, 1996.
6. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. Statist. Med. 2000; 19:1793-1819
7. SAS Institute Inc. Advanced General Linear Models with an Emphasis on Mixed Models. SAS Institute Inc.: Cary, NC, 1996.
8. Schwarz G. Estimating the dimension of a model. Annals of Statistics 1978; 6:461-464.
9. Verbeke G, Molenberghs B. Linear Mixed Models in Practice. Springer: New York, 1997.
10. Wolfinger RD. Covariance structure selection in general mixed models. Communications in Statistics, Simulation and Computation 1993; 22(4):1079-1106.

CONTACT INFORMATION

Radhi Abdulnabi, Ph.D.
Pfizer Inc.
2800 Plymouth Road, Ann Arbor, MI 48105