PROC MIXED and early phase trials

Shelley Fordred

Savvy Stats Ltd, Heston, UK

PROC MIXED is commonly being used to compare treatment or other differences in phase 1 crossover trials. In such trials there is variation between subjects and also variation within subjects — these two sources of variation can be described by random effects. PROC MIXED is used because it can accommodate for random effects and as opposed to other SAS regression procedures, subjects with missing observations can be handled without removing all of their data from the analysis. A fixed effect is a parameter which is modelled in the same way as in PROC GLM — there are pre-specified levels of that effect, e.g. treatment group which is pre-defined in a trial because the aim is to compare responses among the fixed groups. In contrast a random effect is a parameter whose values cause a random variability within a trial and whose values are not known pre-trial, e.g. the subjects' responses in a trial. So commonly, subject is declared as a 'random' parameter in PROC MIXED to account for this random variation in the statistical analysis. The purpose of this paper is to provide an introduction to the PROC MIXED procedure by describing some code that has been used to carry out an analysis on a phase 1 crossover trial.

Keywords: PROC MIXED, PK, Convergence criteria, Cmax, Infinite Likelihood, Model checking

Background

Pharmacokinetics (PK) is the investigation of what the body is doing to the drug[1] and is an integral part of assessing the safety of a new compound. There are many parameters which are used to assess the PK of a compound for example, AUC (area under the curve for a concentration time plot) (see Fig. 1) — if a compound has a large AUC, then this indicates that the subject's exposure to the drug is high and thus, there may be a lot of adverse events for that drug. Similarly for the PK parameter Cmax (see Fig. 1), which is the maximum concentration of the drug reached in the body, a high Cmax may lead to an increase in the number of adverse events that a subject experiences. Conversely a low Cmax is an indication of a slow absorption and may obscure the distribution of drug in the body; hence, any factors that affect the value of Cmax are of interest.

This paper will look at how the two factors of food status and the formulation of the drug affects the Cmax of a drug in a phase 1 trial using the SAS PROC MIXED procedure. PROC MIXED compares the means of the data using the least square means statistical procedure to measure any differences between groups of data.

One of the primary assumptions of PROC MIXED is that the data are normally distributed (Gaussian), and so data which do not exhibit a normal distribution are transformed, such as logarithmic transformation, to convert it to a normal distribution (see Fig. 2).

Correspondence to: S Fordred, Savvy Stats Ltd, 6 Orchard Avenue, Heston TW5 0DU, UK. Email: fordreds@aol.com

Introduction

In this paper, the syntax and the meaning of some PROC MIXED code will be described that has been produced and used to assess if there is a statistically significant effect of the fed state of a subject, i.e. if they have had food or not, and the formulation of the drug on the value of the PK parameter Cmax. The example provided here is taken from a First Time Into Human PK study. The design of this study is that of a typical food effect study with a balanced complete design where all of the subjects receive all of the treatments available in the trial — this is a three-period crossover trial where the subjects would be randomized to receive each of the 'treatments' — fasted, fed milled, and fed micronized over the three periods. The latter is important in that the parameter Period then needs to be included in the model specified below, and the proceeding text provides a definition of the variable names and a description of the PROC MIXED code:

The SAS® code

```
/* To carry out an analysis of Cmax */
  proc mixed data=cmax;
  class subjid period food formul;
  model logpppar=period food formul/ddfm=kr;
  random subjid;
  lsmeans food/diff cl alpha=0.10;
  lsmeans formul/diff cl alpha=0.10;
```


Published by Maney
DOI 10.1179/175709311X13166801334073

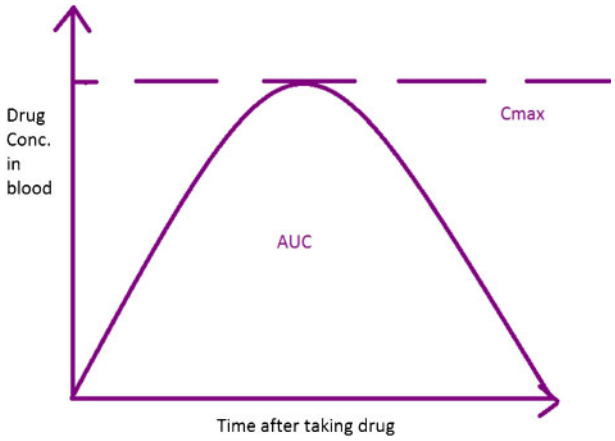

Figure 1 PK parameters

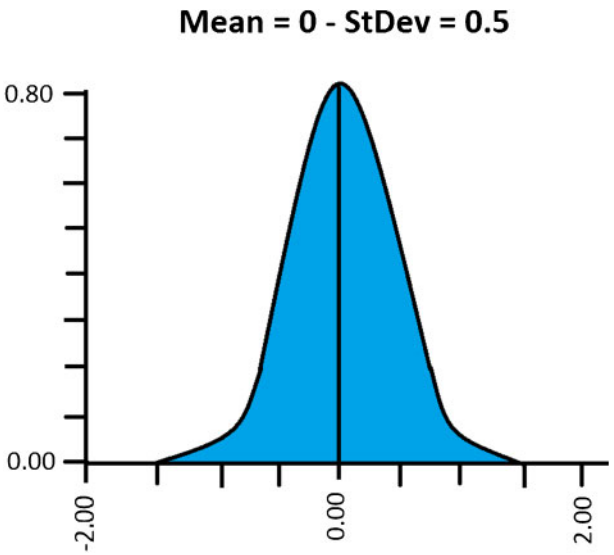

Figure 2 A normal distribution

ods output LSmeans=lsmncm diffs=diffcm;
run;

Definition of the variable names in the PROC MIXED code

subjid=patient

period=the period in which the drug was taken

food=the fed status of the patient, i.e. fed or fasted

formul=the formulation of the drug, i.e. milled or micronized

logpppar=the logged value of Cmax.

Description of the PROC MIXED code[2,3]

Data=the dataset to be analysed.

Class=classification variables to be used in this analysis, i.e. the categorical variables ***subject period food formul***.

Model=the statement used to specify the model for the analysis. The first variable ***logppar*** is the response variable Cmax logarithmically transformed to normalize the data. After the '=' are the explanatory variables, i.e. those variables which maybe affecting the value of Cmax, e.g. for 'food' we are looking to see if the type of food is significantly affecting the Cmax value. The analysis contained an investigation to determine if food has a different effect for different formulations. Therefore, an interaction term **food*formul** was originally included in the model statement; however, this interaction was removed as it was not significant. For most repeated measures or random effects models, the Kenward–Roger method is recommended, i.e. ***ddfm=kr***, this specifies how the degrees of freedom are calculated for the analysis. Depending on the design of the study and the variance–covariance structure, other options for the degrees of freedom are ddfm=satterthwaite, ddfm=residual, ddfm=contain, and ddfm=betwithin.

Random=specifies the variable which is causing the random variability within the study, in this case ***subjid***.

Lsmeans=the statistical method used to test the differences between treatments, i.e. comparing the estimated adjusted mean value of the response variable between the different treatments. In this study lsmeans is testing whether the treatments, i.e. the factors food and formulation are having an effect on the mean values of Cmax adjusting for all of the other parameter in the model, i.e. ***period***. For example, for '***lsmeans foodldiff cl alpha=0.10***', the ***diff*** part is testing whether the difference in the Cmax adjusted estimated mean values between the fed and fast status —— (estimate value in the 'Differences of Least Squares' output) is statistically significant. '***cl***' produces the confidence intervals and ***alpha*** specifies the significance level of the test in this case at the 10% level. As a matter of interest ***lsmeans*** outputs the results of the lsmeans procedure to the log and list file so that the order in which the different values of food status and formulation can be determined, e.g. for 'food' the lsmeans value for the fasted state (fast) is produced first and then the lsmeans value for the fed state (fed) (see Appendix, Output 2). This order is important when using the ***diff*** option in the ***lsmeans*** statement which will order the variables alphanumerically, e.g. the ***lsmeans foodldiff*** statement will subtract the *fed* lsmeans value from the *fast* value as the latter comes out first in the procedure. If the statistical analysis requires the opposite, then manipulation of the dataset produced from the ***diff*** analysis needs to be carried out. ***Ods output*** outputs the results of the individual lsmeans to the ***lsmncm*** dataset, and the difference of the lsmeans from the ***diff*** statement is output to the ***diffcm*** dataset.

The analysis being described is a simple analysis where only one measurement of Cmax is being taken from each subject. However, if more than one measurement of Cmax was taken from each subject over fixed timepoints, then this would become a

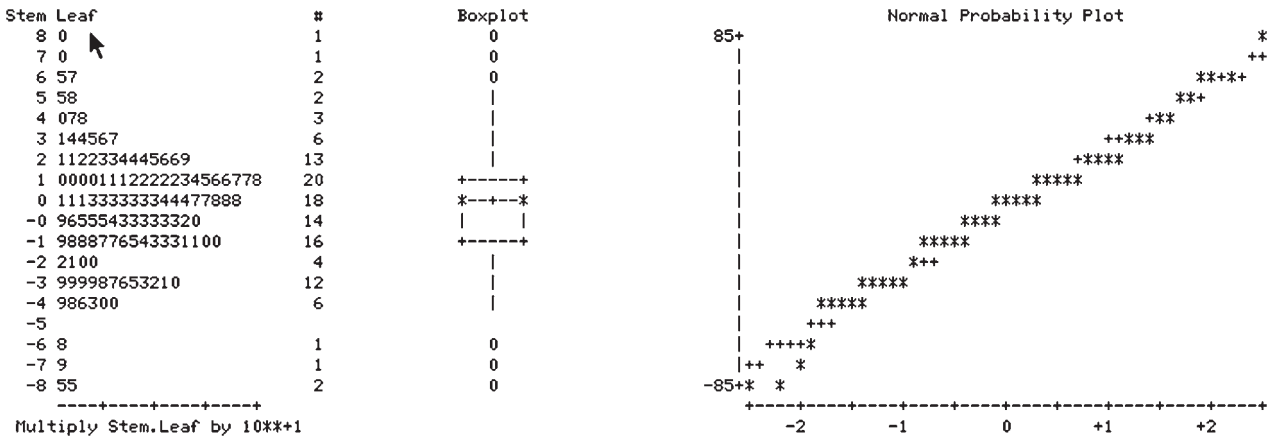

Figure 3 Output from PROC univariate

repeated measures analysis and there would be a ***Repeated*** statement in the PROC MIXED code, e.g. ***repeated visit\|subject=subjid type=UN***.

Model Checking

In order to check that the model is correct, the residuals from the analysis should also follow a normal distribution. Residuals are the predicted value of Cmax, obtained from running the PROC MIXED code, subtracted from the actual value of Cmax from the data. The normality of the residuals can be checked by being output to the predicted dataset using the outpredm=after/in the model statement and normality plots produced using PROC univariate:

```
proc univariate data=cmax plot;
var residual;
run;
```

This code produces a stem and leaf plot, i.e. the horizontal bar chart, a box plot, and a normal probability plot (Fig. 3). As well as checking the normality of the residuals, these plots are also useful for identifying outliers, i.e. data which do not follow the same pattern as the other data and may have some effect on the analysis. These outliers may be true outliers or incorrect due to data collection issues. The reasons require further investigation including data verification by data management.

It is also common to check the adjusted means produced from lsmeans with the raw mean values produced from the raw summary statistics, for example, using PROC means. If the adjusted means are very different from the raw means, then this could mean that the model is incorrect.

What to Look for in the Log

If the model is correct, then there will be a note in the log '**Convergence criteria met**' (see Appendix, Output 1), if there is a note to say otherwise, then the model needs to be changed or the assumption of normality of the data is not met. When the model is incorrect, then this is where the problems occur with the PROC MIXED procedure — it can be very challenging to see why there is a problem with the model. An investigation into the latter by putting the ***Solution*** option in the model statement (after/in the code) may reveal that there is a problem with one of the subject's data. A common error message encountered is '**Estimated G matrix is not positive definite**' and this is because the mean square within subject is greater than the mean square between subjects. The value of the G matrix can be obtained by putting the option '**g**' at the end of the random statement and also a NOBOUND option, and if this is a small negative value relative to the size of the residual, e.g. −0.001, then there is nothing to worry about, if not then the model statement may have to be changed. The warning message '**Warning: Stopped because of infinite likelihood**' should also be investigated, possible causes could be the existence of duplicate observations for some subjects in the data or if there are not enough data to estimate the particular covariance structure selected in the repeated statement. The message '**Convergence criteria met but final hessian is not positive definite**' should also be treated with caution and the cause should be investigated with the procedures mentioned.

When convergence has not been met, then this could be due to the size of the data, if data are very small or very large, then this can cause a problem, When data are too small, then the data can be multiplied by a factor which allows the data to converge and the result modified appropriately, and conversely large data can be scaled down.

Sometimes when running repeated measures PROC MIXED analyses, SAS can run out of memory when running the code interactively and this can be solved by running the code in batch and increasing the memory in SAS using the $ ***sas memsize=300***. With repeated measures 'Infinite likelihood Assumed' can also appear in the log, which is not always a problem when the data have converged but is a problem when the data have not converged. The latter can be investigated using the PARMS statement which allows you to input initial values for

the covariance parameters., but must be specified in the order they appear in the 'Covariance Parameters Estimates table'.

Interpretation of the PROC MIXED Output

When you introduce a random statement, the way that the least square means are calculated is different from a pure fixed effects analysis, i.e. when there is no random statement. To see if there is a statistically significant difference between the fed status of the patient and the drug formulation on Cmax, then you need to look at the 'Differences of Least Squares Means output' (see Appendix, Output 2). In order for the variables ***food*** or ***formul*** to individually have a statistically significantly difference on Cmax, then both Pr>|*t*| values should be less than 0.01. In this case, the Pr>|*t*| values are <0.0001 for both variables and so it can be concluded that there is evidence to suggest that the fed status of the patient and the drug formulation have a statistically significant effect on the Cmax value.

Conclusion

This paper has shown an example of how the PROC MIXED procedure has been used successfully to carry out a food effect analysis with random effects for a PK First Time Into Human clinical trial. The principles of the code provided in this paper apply to other types of analysis in particular in smaller early development studies where a random component is often included.

Acknowledgements

I would like to thank GSK for letting me use their data in this paper and James Roger and Amy Newlands for their technical input.

References

1 Rowland M, Tozer TN. Clinical pharmacokinetics and pharmacodynamics: concepts and applications. London: Lippincott Williams & Wilkins; 2010.
2 Walker GA. Common statistical methods for clinical research with SAS examples. 2nd ed. Cary, NC: SAS Publishing; 2002.
3 SAS OnlineDoc, V8. Cary, NC: SAS Institute. Available from: http://support.sas.com/onlinedoc/913/docMainpage.jsp [accessed September 2006].

Appendix

Output 1 — the log from PROC MIXED

```
/* To carry out an analysis for cmax */
3                 The SAS System                      11:33 Thursday, May 11, 2006
56
57                proc mixed data=cmax;
58                class subjid period food formul;
59                model logppar=period food formul /solution ddfm=kr;
60                random subjid;
61                lsmeans food/diff cl alpha=0.10;
62                lsmeans formul/diff cl alpha=0.10;
63                ods output LSmeans=lsmncm diffs=diffcm ;
64                run;
```

```
WARNING: Length of CLASS variable PERIOD truncated to 16.
WARNING: Length of CLASS variable food truncated to 16.
WARNING: Length of CLASS variable formul truncated to 16.
NOTE: Convergence criteria met.
NOTE: The data set WORK.DIFFCM has 2 observations and 13 variables.
NOTE: The data set WORK.LSMNCM has 4 observations and 11 variables.
NOTE: The PROCEDURE MIXED printed pages 1-2.
NOTE: PROCEDURE MIXED used:
  real time              0.27 seconds
  cpu time               0.03 seconds
65
66
67
NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
  real time            0.46 seconds
  cpu time             0.16 seconds
```

Output 2 — the output from PROC MIXED

Least Squares Means

Effect	food	formul	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
food	Fast		3.9674	0.08208	15.7	48.33	<.0001	0.1	3.8240	4.1109
food	Fed		4.9342	0.08208	15.7	60.11	<.0001	0.1	4.7907	5.0777
formul		Micro	4.9848	0.08208	15.7	60.73	<.0001	0.1	4.8413	5.1282
formul		Mill	3.9169	0.08208	15.7	47.72	<.0001	0.1	3.7734	4.0604

Differences of Least Squares Means

Effect	food	formul	_food	_formul	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
food	Fast		Fed		-0.9668	0.06737	31	-14.35	<.0001	0.1	-1.0810	-0.8526
formul		Micro		Mill	1.0679	0.06737	31	15.85	<.0001	0.1	0.9536	1.1821