

Proc mixed – Trials and Tribulations

Shelley Fordred, GlaxoSmithKline, Greenford, UK

ABSTRACT

At GSK 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. Proc Mixed is used because random effects can be used to describe these two sources of variation. Also 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 subject responses in a trial. So commonly, subject is declared as a ‘random’ parameter within Proc Mixed to account for this randomness in the statistical analysis.

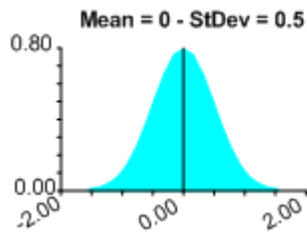
The purpose of this paper is to provide a whistle-stop tour of the Proc Mixed procedure by describing some code that has been used to carry out an analysis on a phase 1 crossover trial in at GlaxoSmithKline.

BACKGROUND

Pharmacokinetics (PK) is the investigation of what the body is doing to the drug and so 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) – if a compound has a large AUC then this indicates that the subject’s exposure to the drug is high and thus there maybe a lot of adverse events for that drug. Similarly for the PK parameter C_{max}, which is the maximum concentration of the drug reached in the body, a high C_{max} may lead to an increase in the number of adverse events that a subject experiences. Conversely a low C_{max} 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 C_{max} are of interest.

So, this paper will look at how the two factors of food status and the formulation of the drug affects the C_{max} 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 does not exhibit a normal distribution is transformed, such as logging, to convert it to a normal distribution.

INTRODUCTION

In this paper I will describe the syntax and the meaning of some Proc Mixed code I have produced and used to see 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 in a First Time Into Human Study (FTIH) PK study. The design of this study is that of a typical food study i.e. a balanced complete design where all of the subjects receive all of the treatments available in the trial. The latter is important in that the parameter Period then needs to be included in the model, the latter being specified in the statistical analysis plan.

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;
  ods output LSmeans=lsmnem 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 micronised
 logpppar= the logged value of Cmax.

Description of the Proc Mixed code

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 logged so as to normalise the data. After the '=' are the explanatory variables i.e. those variables which may be 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 did also require to see if food has a different effect for different formulations so an interaction term **food*formul** was originally included in the model statement, however this interaction was removed as it was not significant. **ddfm=kr** specifies how the degrees of freedom are calculated for the analysis.

Random= specifies the variable which is causing the random variability within the study, in this case **subjid**. The random parameters for the analysis are specified by the statistician.

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 parameters in the model i.e. **period, food, and formul**. For example for '**lsmeans food/diff 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). This order is important when using the **diff** option in the **lsmeans** statement e.g. the **lsmeans food/diff** statement will subtract the **fed** **lsmeans** value from the **fast** value. 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.

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 log output on page 4), if there is a note to say otherwise then the model needs to

be changed or the assumption of normality of the data is incorrect. When the model is incorrect then this is where the *tribulations* 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 -ve 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.

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 output on page 5). In order for the variables *food* or *formul* to individually have a statistically significant difference on Cmax then both $Pr > |t|$ values should be less than 0.01. In this case the $Pr > |t|$ values are $<.0001$ 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

Hence 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 FTIH clinical *trial*, there being no *tribulations* with using Proc Mixed as the model specified in the statistical plan was correct.

REFERENCES

1. **Common Statistical Methods for Clinical Research with SAS Examples – Glenn A.Walker second edition.**
2. **SAS OnlineDoc, V8.**

CONTACT INFORMATION

Shelley Fordred
GlaxoSmithKline GlaxoSmithKline
Research & Development
Greenford Road
Greenford
Middx
UB6 0HE
Work Phone: +44(0)2089662771
Fax: +44(0)208 9663793
Email: Shelley.X.Fordred@gsk.com

PhUSE 2006

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 logpppar=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
----------	--------------

PhUSE 2006

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