PhUSE 2013

Paper SP03

Using PROC GENMOD to Analyse Ratio to Placebo in Change of Dactylitis

Irmgard Hollweck, UCB Biosciences GmbH, Monheim, Germany
Meike Best, UCB Biosciences GmbH, Monheim, Germany

ABSTRACT

A common symptom of psoriatic arthritis is a sausage-like swelling in the fingers or toes, known as dactylitis. In a placebo controlled clinical trial the number of digits affected was counted ranging from 0 to 20. The number of affected digits was assumed to follow a binomial (20,p) distribution, where p is the probability of an individual digit being affected. The presence of dactylitis was to be calculated in combination with a ratio of active treatment versus placebo and the corresponding confidence intervals and p-values. The outcome of interest was the ratio to placebo in change, where change is expressed as the odds ratio of post-baseline vs. baseline based on repeated measures. This paper explains how such a complex logistic regression model with provision of p-values can be built using the SAS Procedure PROC GENMOD.

INTRODUCTION

Dactylitis is associated with many different diseases (e.g. Tuberculosis, Sickle-cell)[1], but in connection with Psoriatic Arthritis it is mentioned most frequently[3]. The reason is that it might be a signal for the onset of this disease. Together with other symptoms it contributes to the diagnosis of Psoriatic Arthritis (CASPAR Criteria=Classification Criteria for Psoriatic Arthritis).

Usually dactylitis involves few fingers and/or toes asymmetrically[2] and shows swelling and pain along the flexor tendons. If occurring in a finger, the patient often cannot flex the finger. Once the underlying health condition that led to the swelling and inflammation is diagnosed (e.g. Psoriatic Arthritis) and directly treated, it will often result in the disappearance of the swelling as well as ease the pain[3]. Dactylitis is referred to as “sausage digit” and is combined with inflammation.

Clinical studies record circumferences and tenderness[4] of the affected digit. 20 digits (10 fingers and 10 toes) might be affected. Due to the fact that the investigator evaluates these measurements and not the patient, the data are stored in the FA-domain (Findings About) of Study Data Tabulation Model (SDTM).

In general the Leeds Dactylitis Instrument-score (LDI-score)[5] is the main focus for dactylitis in a clinical study. The LDI-score is one objective, validated outcome per time-point which has to be calculated from all of the input data. Based on the LDI the different study specific analysis like imputations methods will be performed. The Analysis Data Model (ADaM) contains these derived variables as well as the raw data from the SDTM.FA domain for the tables, figures, listings (TFLs).

Using the “Number of affected digits’ (range 0-20) instead of LDI brings up a few points which are easily forgotten. Therefore, this paper introduces background knowledge about the statistic and basic knowledge about the SAS-procedure PROC GENMOD. Finally, the translation of the Statistical Analysis Plan (SAP) into different solutions of a PROC GENMOD invocation is described The target of this paper is to give enough information to understand the implementation of the solutions, not to explain every detail of the procedure and the underlying statistics.

DEFINITIONS/STATISTICAL ANALYSIS

The statistical analysis described in the last section is based on following facts and definitions:

TRIAL

- Treatment Groups: Active 200mg(Q2W) Active 400mg(Q4W) Placebo 0mg.
- Trial Design Parallel for this 3 treatment groups
- Assessment of Dactylitis exist on: Visit: 20, 40, 50, 70, 90, .. from 90 on visit =x +20.
  Baseline = Visit 20

DEFINITIONS:

- Number of affected digits: Can be 0 to 20 per time point, depending on how many digits are affected.
- Subject with dactylitis: LDI-score criterion met on Baseline = dactylitis exists according to LDI.
- Presence of dactylitis: At least one affected digit at the given visit.
  Note: in articles this term is used without consistent meaning, i.e. it is rarely explained what 'Presence of dactylitis' represents (general investigator assessment or affected finger or LDI-score).

STATISTIC:

- 95% significance level with a 2-sided p-value is used.
- 'Number of affected digits' assumes binary values since each digit of a patient is affected (1) or not (0). A 20 time repetition of these digits results in a binomial (20,p) distribution.
- Logistic Regression is defined as (for binary logit model based on Generalized Linear Model)[6,7]:
  - Probability (= non-linear equation):
    $$\pi = \frac{\exp(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)}{1+\ \exp(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)}$$
  - ODDs: $\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = e^{\beta_0} * e^{\beta_1 x_1} * \ldots * e^{\beta_k x_k} = \frac{P(Y=1|X)}{1-P(Y=1|X)}$ with $\pi = Probability\ of\ the\ dependent\ variable$
  - Logit: $LN\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$
    The linearization (=logarithmic) of the equation from the probability is necessary to get a linear connection between response and independent variable. This is the basic equation for the logistic regression.
  - $Odds\ Ratio_i$: $e^{\beta_i} = \exp(estimate_i)$
    $math.\ explanation\ for\ case\ e^{\beta_0} * e^{\beta_1 x_1}$:

    $$x_1 = 1,\ \ e^{\beta_0} = \frac{P(Y=1|X=0)}{1-P(Y=1|X=0)}, \qquad e^{\beta_0} * e^{\beta_1} = \frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}$$
    $$=>\ e^{\beta_1} = \frac{\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}}{\frac{P(Y=1|X=0)}{1-P(Y=1|X=0)}}$$

    The Odds Ratio is calculated to compare the odds across groups, example
    $$Odds\ Ratio = \left(\frac{ODDs_1}{ODDs_2}\right) = \left(\frac{ODDs_{Visit}}{ODDs_{Baseline}}\right)$$
- SAS uses the generalized linear model to generate the initial values for the Generalized Estimating Equations (GEE) Logistic Regression, see next paragraph. That means the above formulas should only show the basic principles of the statistical background.

INTRODUCTION TO PROC GENMOD[8]

The SAS procedure PROC GENMOD fits a generalized linear model to the data by maximum likelihood estimation of the parameter vector β. It estimates the parameters of the model numerically through an iterative fitting process as generally no closed form solution is available. A number of frequently used link functions and their respective probability distributions are built-in in the GENMOD procedure, among them the logit link function and binomial distribution needed for the presented model.

The general procedure invocation for a generalized linear model with assumed binomial distribution reads:

```
proc genmod;
   model resp = eff1 eff2 … / dist=bin;
run;
```

The response can be specified either in the form of a single variable *resp* or as a ratio of two variables *events/trials.* Effect variables can be either categorical or continuous. Interactions of the effect variables can be specified using the crossing operator (*) or the nesting operator (var1(var2)).

A CLASS statement can be added prior to the MODEL statement to define the categorical effect variables . Categorical effect variables identify explanatory classification levels in the model. They can be either character or numeric and are also referred to as classifying or qualitative variables. Any variable in the model not included in the CLASS statement is assumed to be continuous. A reference statement can be used in the CLASS statement to set a fixed reference level of the classifying variables, e.g. treatment group Placebo.

The logit link function is used with the binomial distribution by default.

The GENMOD procedure can fit models of correlated data due to repeated measurements using GEEs. The REPEATED statement specifies the covariance structure of multivariate responses in the model. In addition, it controls the iterative fitting algorithm and specifies optional output.

`repeated subject=`*`subject-effect`* `/ <options>;`

`subject=` identifies subjects in the input dataset

`option type=` specifies the structure of the working correlation matrix used to model the correlations of the responses from subjects. The default working correlation is independent. For the models studies here `TYPE=UN` will be used.

`option withinsubject=` can be used to define an effect specifying the order of measurements within each subject. This is especially required if some measurements are not present for each subject or if the measurements are not in proper order. All variables used in the *withinsubject effect* must be present in the CLASS statement.

A *`subject-effect`* must be used and all variables used in the *`subject-effect`* statement must be present in the CLASS statement.

To obtain a test for a specified hypothesis concerning the model parameters, either the CONTRAST or the ESTIMATE statement can be used. Both most appear after the MODEL statement. The hypothesis is tested by specifying a fixed model matrix L and testing L'β=0. The CONTRAST statement is more general while the ESTIMATE statement only allows for a test matrix L with one single row.

option *`E`* in the CONTRAST or ESTIMATE statement requests that the L matrix should be displayed.

option `EXP` in the ESTIMATE statement requests that estimates for exp(L'β), along with standard errors and confidence limits, will be additionally computed and output.

FROM STATISTICAL ANALYSIS PLAN TO SAS CODE

REQUIREMENTS

The following extract of the SAP describes the analysis.

Roman numbers on the right site are reference numbers which will referred to, later on:
(I) Introduction, (II) + (X) Additional analysis, (III) + (VII) Model description, (IV) – (VIII) Additional details for the analysis, (IX) Confidence Interval

SAP EXTRACT:

"Presence of dactylitis (post Baseline) and the number of affected digits are newly defined variables. ...
The latter analyses will be done descriptively by visit as mentioned earlier in this section. The number and % of subjects with dactylitis will be given. (I)
Descriptive statistics will be presented for the number of affected digits in all subjects and in those that have dactylitis at that visit separately. (II)
The difference between each active treatment group (and both combined) and PBO in the probability of dactylitis will be analyzed by a repeated measures logistic regression model in which the logit of probability of dactylitis is assumed to depend on treatment group, visit (Baseline vs. the respective post-Baseline visit), and the interaction of treatment group and visit. (III)
Within-subject correlation will be taken into account by allowing an unstructured covariance structure between Baseline and the respective post-Baseline visit. (IV)
The model will be fitted for each post-Baseline visit separately. The number of affected digits will be analyzed for each post-Baseline visit separately. (V)
The model will be a repeated measures logistic regression model in which the number of affected digits is assumed to follow Binomial (20, p) distribution, where p is the probability of an individual digit being affected. (VI)
The model will contain treatment group, visit (Baseline vs. the respective post-Baseline visit), and the interaction of treatment group and visit. Within-subject correlation will be taken into account by allowing an unstructured covariance structure between Baseline and the respective post-Baseline visits. (VII)
For presence of dactylitis and the number of affected digits, the difference between each active treatment groups (and both combined) and PBO will be estimated as the ratio of odds ratios between post-Baseline and Baseline visits. (VIII)
The confidence interval and p-value will be provided. (IX)
The descriptive statistics will be presented for presence of dactylitis and the number of affected digits (in all subjects and in those that have dactylitis at that visit separately). (X)

MOCK-SHELL:

The mock shell also shows what is required:

Table xxx: Presence of Dactylitis and Number of Affected Digits by Visit

Visit	Treatment	Presence of dactylitis n (%)	Ratio to PBO (95% CI), p-value	Number of digits affected among those with dactylitis at visit: Mean	SD	Median	Min	Max	Number of digits affected among all subjects: Mean	SD	Median	Min	Max	Ratio to PBO (95% CI), p-value [a]
WK0 (V2, BL)	PBO N=XXX	XX (xx)		XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	
	ACT200mg N=XXX	XX (xx)		XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	
	ACT400mg N=XXX	XX (xx)		XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	
	ACT 200mg+400mg N=XXX	XX (xx)		XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	
WK1 (V5)	PBO N=XXX	XX (xx)		XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	
	ACT200mg N=XXX	XX (xx)	X.XX (X.XX, X.XX) X.XXX	XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	X.XX (X.XX, X.XX) X.XXX
	ACT400mg N=XXX	XX (xx)	X.XX (X.XX, X.XX) X.XXX	XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	X.XX (X.XX, X.XX) X.XXX
	ACT 200mg+400mg N=XXX	XX (xx)	X.XX (X.XX, X.XX) X.XXX	XX.X	XX.XX	XX.X	XX	XX	XX.X	XX.XX	XX.X	XX	XX	X.XX (X.XX, X.XX) X.XXX
…														

Note: Presence of dactylitis is defined as at least 1 digit affected at the given visit.

[a] Ratio to PBO in change, where change is expressed as the odds ratio of post-Baseline vs. Baseline for whether a digit is affected. Each subject has 20 digits available for evaluation at each visit. The estimate, confidence interval, and p-value are based on a repeated measures logistic regression model with treatment group, visit, and their interaction in the model.

FIRST STEP: UNDERSTANDING OF THE SAP

The essence of this SAP is described below. The explanations handles only 'Number of Digit' analysis (which is circled in the mock shell) because the 'Presence of dactylitis' analysis is analogous and descriptive statistics is trivial. Recognition of the two repetitions (X) +(VIII) in the SAP avoids confusion.

COVARIANCE STRUCTURE (IV)

An unstructured covariance structure is the best structure to fit data and has the least amount of assumptions behind it compared with other covariance structures. Each value is estimated uniquely which means enough data must be available.

MODEL FIT FOR EACH POST-BASELINE VISIT (V)

The interest is to look into the effect for each visit in comparison to the baseline consequently for each visit there is a separate model run. Example: for 30 visits, there are 30 models to be generated.

REPEATED MEASURES LOGISTIC REGRESSION MODEL (VII) + (VIII) + (IX)

- The response variable is 'Number of Digits', the independent variables are treatment, visit, interaction treatment and visit. (VII)
- Repeated measurements (i.e. the measurement of Dactylitis) are used per subject per visit.
- A confidence interval and p-value is requested (X)
- Ratio of the Odds Ratios (VIII):

$$Ratio\ of\ the\ Odds\ Ratio = \frac{Change\ from\ Baseline\ in\ Active}{Change\ from\ Baseline\ in\ Placebo} = \frac{\frac{Mean\ of\ Active\ ([visit]}{Mean\ of\ Active[Baseline]}}{\frac{Mean\ of\ Placebo\ ([visit]}{Mean\ of\ Placebo[Baseline]}}$$

since the model for a logistic regression is a logarithmic function this can be written as:

⇨ Change from Baseline in Active - Change from Baseline in Placebo

⇨ (Mean of Active[visit] - Mean of Active[Baseline]) - (Mean of Placebo[visit] - Mean of Placebo[Baseline])

⇨ Mean of Active[visit] - Mean of Active[Baseline] - Mean of Placebo[visit] + Mean of Placebo[Baseline]

⇨ +1 -1 - 1 +1

which are the weights for our contrasts (see solution 1 below)

DISTRIBUTION (VI)

When the outcome is binomial, the values of the dependent (or outcome, response) variable can be given "events/trials". In the current example, events is the number of affected digits and trials is the number of assessed digits (always assumed to be 20)

SECOND STEP: PROGRAMMING

Repeated measurements and p-values with a logistic regression can be done by PROC GENMOD. Therefore it was chosen for this analysis.

Proc Genmod is a quite powerful procedure and there are different solutions possible to cover this analysis. The obvious one is mentioned first, then a second solution is introduced and in the end an outlook for simplification or generalization is given.

The input dataset DS has the structure one row per subject (= USUBJID) and visit.

Extract of DS:

VIEWTABLE: Work.Ds

	USUBJID	AVISIT	AVISIT	PARAM	AVAL	N	TREAT	TRT_NO
1	kkk-rrr-123	20	WK0 (V2, BL)	Number of Digits Affected	8	20	0mg	4
2	kkk-sss-456	20	WK0 (V2, BL)	Number of Digits Affected	0	20	200mg	5
3	kkk-ttt-789	20	WK0 (V2, BL)	Number of Digits Affected	2	20	400mg	6

SOLUTION 1

This solution provides the separate models (V) per visits via a macro call and is working with the estimate statements to generate the ratio of the odds ratio. In the repeated statement the covariance structure (IV) is mentioned. The distribution adaptation (VI) can be seen in the response variable. Proc Genmod recognizes a binomial distribution with the definition of the response variable automatically (i.e. we do not have to specify the link option explicitly in such cases).

```
proc sort data = ds;
     by trt_no avisitn avisit;
run;

%macro gm_dact(vis = );

 proc genmod data = ds(where = (trt_no in(4 5 6) and avisitn in(20 &vis)));
 class trt_no avisitn usubjid;
 model     aval/n = trt_no avisitn trt_no*avisitn;
 repeated subject = usubjid / corr = unstr;

 estimate "ACT 200 Vs PBO BL &vis" trt_no*avisitn 1 -1 -1  1  0  0 / e exp;
 estimate "ACT 400 Vs PBO BL &vis" trt_no*avisitn 1 -1  0  0 -1  1 / e exp;
 estimate "ALL ACT Vs PBO BL &vis" trt_no*avisitn 1 -1 -0.5  0.5 -0.5 0.5 / e exp;

  ods output estimates = num_&vis;
run;

%mend gm_dact;

%gm_dact(vis = 30);
%gm_dact(vis = 40);
%gm_dact(vis = 50);
...
data dactest;
  set dact_30 dact_40 dact_50 ...;
run;
```

For ACT 200 the output shows:

Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept			-3.1841	0.2223	-3.6197	-2.7484	-14.32	<.0001
TRT_NO	4		0.0359	0.3282	-0.6074	0.6791	0.11	0.9130
TRT_NO	5		-0.1954	0.2947	-0.7729	0.3821	-0.66	0.5073
TRT_NO	6		0.0000	0.0000	0.0000	0.0000		
AVISITN	20		0.4391	0.1152	0.2132	0.6649	3.81	0.0001
AVISITN	40		0.0000	0.0000	0.0000	0.0000		
TRT_NO*AVISITN	4	20	-0.2990	0.1297	-0.5533	-0.0447	-2.30	0.0212
TRT_NO*AVISITN	4	40	0.0000	0.0000	0.0000	0.0000		
TRT_NO*AVISITN	5	20	-0.1507	0.1691	-0.4822	0.1808	-0.89	0.3731
TRT_NO*AVISITN	5	40	0.0000	0.0000	0.0000	0.0000		
TRT_NO*AVISITN	6	20	0.0000	0.0000	0.0000	0.0000		
TRT_NO*AVISITN	6	40	0.0000	0.0000	0.0000	0.0000		

Coefficients for Contrast ACT 200 Vs PBO 40

	Prm1	Prm2	Prm3	Prm4	Prm5	Prm6	Prm7	Prm8	Prm9	Prm10	Prm11	Prm12
ACT 200 Vs PBO vis 40	0	0	0	0	0	0	1	-1	-1	1	0	0

The basis for the odds-ratios of the change from baseline is the estimates above. When we calculate the model

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

with the estimates, then we get the following equation for Active 200mg:
0*(-3.1841) + 0*0.0359+0*(-0.1954)+0*0+0*0.4391+0*0+1*(-0.2990)+(-1)*0+(-1)*(-0.1507)+0*0+0*0+0*0
= -0.1484. This is the estimate for the ratio of the odds ratio (result see next table) with the logit function. To get the required ratio we have still to apply the exponential-function (e^{β})..Analogously, the values for the other treatment groups are generated.

The dataset NUM_&vis contains the results of the model. Below is an example for visit 40. The circled values are the Ratio of the Odds Ratio (VIII) i.e. the exponentiated estimates. The confidence intervals are marked with an arrow and the p-value is marked with a star in front of the value (IX).

VIEWTABLE: Work.Num_a40 (ESTIMATE Statement Results)

	LABEL	MEANESTIMATE	MEANLOWERCL	MEANUPPERCL	LBETAESTIMATE	STDERR	ALPHA	LBETALOWERCL	LBETAUPPERCL	CHISQ	PROBCHIS
1	ACT 200 Vs PBO vis 40	0.4630	0.3971	0.5302	-0.1484	0.1374	0.05	-0.4177	0.1209	1.17	0.2802
2	Exp(ACT 200 Vs PBO vis 40)	_	_	_	0.8621	0.1185	0.05	0.6586	1.1286	_	_
3	ACT 400 Vs PBO vis 40	0.4258	0.3651	0.4888	-0.2990	0.1297	0.05	-0.5533	-0.0447	5.31	0.0212
4	Exp(ACT 400 Vs PBO vis 40)	_	_	_	0.7415	0.0962	0.05	0.5750	0.9563	_	_
5	ACT ALL Vs PBO vis 40	0.4443	0.3950	0.4948	-0.2237	0.1035	0.05	-0.4265	-0.0209	4.67	0.0306
6	Exp(ACT ALL Vs PBO vis 40)	_	_	_	0.7996	0.0827	0.05	0.6528	0.9793	_	_

Setting all individual visit results together leads to one dataset which contains the information for the table column 'Ratio to PBO'.

SOLUTION 2

The separate models (V) per visits are generated this time within a macro loop. Instead of setting the different model solutions outside of the macro, the setting is shortly before the end of the loop (data &outds). The macro variable &vistot contains the maximal number of visits and the macro variable avisitn&visno contains the visit number of the specific visit. The speciality of this solution is the utilization of a reference in the CLASS statement. Note: the reference in the model is prior to any sorting order in the data options, i.e. the descending in the second line of the Proc Genmod statement has no effect i.e. the reference is for trt_no=4 and for avisitn=20. The estimate effects have to be specified only for the treatment groups other than the reference treatment because everything else are already known by the system. The implicite options from solution1: withinsubject option in the repeated statement and the dist and link in the model statement, is now specified.

```
%DO visno=1 %TO &vistot;

    PROC GENMOD DATA=ds (WHERE=( trt_no IN (4,5,6) AND avisitn IN
                                (20,&&avisitn&visno))) DESCENDING;

      CLASS usubjid trt_no(REF=FIRST) avisitn (REF=FIRST) / PARAM=REF;
      MODEL    aval/n  = trt_no avisitn trt_no*avisitn / DIST=BIN LINK=LOGIT;
      REPEATED SUBJECT = usubjid /withinsubject=avisitn TYPE=UN;

      ESTIMATE "ACT 200 Vs PBO"       trt_no*avisitn 1   0   / EXP;
      ESTIMATE "ACT 400 Vs PBO"       trt_no*avisitn 0   1   / EXP;
      ESTIMATE "ALL ACT Vs PBO" trt_no*avisitn 0.5 0.5 / EXP;

      ODS OUTPUT ESTIMATES=_estimatesa;

    RUN;

    DATA _estimates;
      SET _estimatesa;
      avisitn=&&avisitn&visno;

    RUN;

    DATA &outds;
      SET %IF &visno^=1 %THEN &outds; _estimates;
    RUN;
%END;
```

For ACT 200 the output shows:

Parameter			Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept			-3.0082	0.2184	-3.4362	-2.5801	-13.77	<.0001
TRT_NO	5		-0.0829	0.2713	-0.6147	0.4489	-0.31	0.7600
TRT_NO	6		0.2632	0.3039	-0.3325	0.8589	0.87	0.3866
AVISITN	40		-0.1400	0.0596	-0.2569	-0.0232	-2.35	0.0188
TRT_NO*AVISITN	5	40	-0.1484	0.1374	-0.4177	0.1209	-1.08	0.2802
TRT_NO*AVISITN	6	40	-0.2990	0.1297	-0.5533	-0.0447	-2.30	0.0212

When we calculate the model with the estimates, then we get the following equation for Active 200mg:
0 + 0 + 0 + 0 + 1*(-0.1484) + 0 = -0.1484

The rest is according to solution1 i.e. the results are the same.

SIMPLIFICATIONS OR GENERALIZATION OF SOLUTIONS:

SIMPLIFICATION OF SOLUTION 1

Using SAS 9.3, the estimate statements in solution 1 can be replaced by LSMESTIMATE, a new SAS statement combining features of both the LSMEANS and the ESTIMATE statement. In this case the confidence interval has to be stated explicitly in the options (cl)

```
lsmestimate trt_no*avisitn "ACT 200 Vs PBO vis &vis" 1 -1 -1  1  0  0 ,
                           "ACT 400 Vs PBO vis &vis" 1 -1  0  0 -1  1 ,
                           "ALL ACT Vs PBO vis &vis" 1 -1 -0.5  0.5  -0.5  0.5
                                / e exp cl;
```

The output looks more streamlined (instead of 6 rows we are getting now 3 rows with the same information, the circled are our required result):

Coefficients for TRT_NO*AVISITN Least Squares Means Estimates

Parameter	TRT_NO	Analysis Visit (N)	Row1	Row2	Row3
TRT_NO 4 * Analysis Visit (N) 20 4	20	1	1	1	
TRT_NO 4 * Analysis Visit (N) 40 4	40	-1	-1	-1	
TRT_NO 5 * Analysis Visit (N) 20 5	20	-1		-0.5	
TRT_NO 5 * Analysis Visit (N) 40 5	40	1		0.5	
TRT_NO 6 * Analysis Visit (N) 20 6	20		-1	-0.5	
TRT_NO 6 * Analysis Visit (N) 40 6	40		1	0.5	

Least Squares Means Estimates

Effect	Label	Estimate	Standard Error	z Value	Pr > |z|
TRT_NO*AVISITN	ACT 200 Vs PBO vis 40	-0.1484	0.1374	-1.08	0.2802
TRT_NO*AVISITN	ACT 400 Vs PBO vis 40	-0.2990	0.1297	-2.30	0.0212
TRT_NO*AVISITN	ALL ACT Vs PBO vis 40	-0.2237	0.1035	-2.16	0.0306

Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
0.05	-0.4177	0.1209	0.8621	0.6586	1.1286
0.05	-0.5533	-0.04473	0.7415	0.5750	0.9563
0.05	-0.4265	-0.02091	0.7996	0.6528	0.9793

GENERALIZATION OF CODE USING LSMEANS AND DIFF

Generalization of code is sometimes needed in meta-analysis over different studies. The following is just an idea how this might work. With this basic principle, a generalization should be straight forward.

- Estimates are generated as a linear combination of LSMEANS from the respective effect
- Linear combinations can be obtained using a LSMEANS statement with options / diff cl exp
- However, the linear combinations constructed and evaluated by that statement are less complex than needed in this paper for example only the change in the treatment is considered but not the change of baseline

⇨ The required results could also be obtained by linearly combining results from LSMEANS differences.

Example: For lsmeans determination the Slice-statement was used:

```
ods output SliceDiffs=diffs ;
proc genmod data = ds;
  class ... ;
  model ... ;
  repeated ... ;
  slice trt_no*avisitn / sliceby=avisitn diff oddsratio cl;
run ;
```

We obtain the following output (shortened):

```
The GENMOD Procedure

           Simple Differences of TRT_NO*AVISITN Least Squares Means
                                                Standard
Slice          TRT_NO    _TRT_NO   Estimate      Error    z Value    Pr > |z|
AVISITN 40     6         5           0.1954     0.2947       0.66       0.5073
AVISITN 40     6         4          -0.03586    0.3282      -0.11       0.9130...
AVISITN 40     5         4          -0.2313     0.3094      -0.75       0.4548


       Chi-Square Test for TRT_NO*AVISITN
           Least Squares Means Slice
               Num
Slice           DF    Chi-Square      Pr > ChiSq
AVISITN 20       2          1.73          0.4218


           Simple Differences of TRT_NO*AVISITN Least Squares Means
                                                Standard
Slice          TRT_NO    _TRT_NO   Estimate      Error    z Value    Pr > |z|
AVISITN 20     6         5           0.3461     0.2657       1.30       0.1928
AVISITN 20     6         4           0.2632     0.3039       0.87       0.3866
AVISITN 20     5         4          -0.08289    0.2713      -0.31       0.7600
```

Interpretation:
AVISITN 40 TRT_NO 5 _TRT_NO 4 : ln(odds ratio(Act 200 – PBO at visit 40))
AVISITN 20 TRT_NO 5 _TRT_NO 4 : ln(odds ratio(Act 200 – PBO at baseline))

We need OddsRatio visit 40 to baseline (combine both numbers linearly):
exp(-0.2313 – (-0.08289))=exp(-0.14841)= 0.8621

CONCLUSION

This paper gives a short introduction to the background of the analysis, explains the SAP and shows different ways to program the analysis with one SAS procedure (Proc Genmod). The interesting point is not only how to understand the SAP, but the most surprising part is that one procedure can solve this task with quite different invocations. The teaching part is that if you change one statement in a procedure then you might have to change others as well and that there is not only one correct solution for a task but multiple solutions are available. This can be quite confusing but also re-assuring.

REFERENCES:

[1] www.diseasespictures.com/dactylitis/
[2] The Journal of Rheumatology: Dactylitis of “Sausage-Shaped” Digit from Ignazio Olivieri, Angela Pdula, Enrico Scarano, Raffaele Scarpa
[3] www.WiseGEEK.com/what-is-dactylitis.htm
[4] www.hcgi.nlm.nih.gov/pubmed/16142872 : “Development of an assessment tool for dactylitis in patients with psoriatic arthritis.”
[5] www.hindawi.com/journals/ijr/2012/839425/
[6] Eid Gollwitzer Schmitt: Statistik und Forschungsmethoden, Page 773, Beltz Verlag, Basel 2011, ISBN 978-3-621-27524-8
[7] Fahrmeir Hamerle Tutz 1996: Multivariate statistische Verfahren, Chapter 6.3, Walter de Gruyter, Berlin 2. Auflage, ISBN 3-11-013806-9
[8] http://www.math.wpi.edu/saspdf/stat/chap29.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Irmgard Hollweck
UCB Biosciences GmbH
Alfred-Nobel-Straße 10
40789 Monheim, Germany
Email: irmgard.hollweck@ucb.com