

Analysis of a 3-period 2-treatment pharmacokinetic study to assess scaled average bioequivalence

Jochen Mueller-Cohrs, CSL Behring GmbH, Marburg, Germany

ABSTRACT

In the past years various approaches for assessing bioequivalence of so-called "highly variable drugs" have been debated. These are drugs whose pharmacokinetic profiles vary considerably when given to the same subject. As a consequence of the high variability, average bioequivalence between two such drugs can only be shown with an unfeasibly high number of subjects. Therefore regulatory agencies generally agree that the criterion of average bioequivalence warrants adjustment for highly variable drugs. The mainstream of the most recent discussion favours an approach that is called scaled average bioequivalence. With this approach the usual criterion of average bioequivalence is scaled by the within-subject standard deviation of the pharmacokinetic parameter in question. A parsimonious design for evaluating scaled average bioequivalence is a three-period two-treatment crossover trial where the reference drug is given twice and the investigational drug is given once to each subject. In this paper it is shown how to analyse the data of such a trial with respect to the criterion of scaled average bioequivalence under the assumption of a log-normal distribution of the pharmacokinetic parameter. Explicit SAS® code for the analysis is provided.

INTRODUCTION

"Highly variable drugs" are called drugs whose pharmacokinetic profiles vary considerably when given to the same subject. For highly variable drugs the FDA presented possible alternatives to the average bioequivalence method in a meeting of the Advisory Committee for Pharmaceutical Science (ACPS) in April 2004. The committee favoured a scaled approach where the bioequivalence limits are expanded based on the within-subject variability of the reference product (cf. [1]).

Specifically, let Y denote the log-transformed values (natural logarithm: $\ln$) of the area under the plasma concentration time curve (AUC) or the maximum concentration ($C_{\max}$). Let μ_R and μ_T denote the mean values of Y for the reference and the test product, respectively, and let σ_R denote the within-subject standard deviation of Y for the reference product. The criterion for scaled average bioequivalence requires the following null hypothesis to be rejected at a significance level of 5%:

$$(1) \quad H_0: |\mu_T - \mu_R| > \theta_A (\sigma_R / \sigma_0)$$

where θ_A is $\ln(1.25)$, the limit for average bioequivalence, and σ_0 is a scaling constant. The scaling constant σ_0 needs to be agreed upon. Values between 0.2 and 0.3 are under discussion. A scaling constant of 0.25 would imply that the average bioequivalence criterion θ_A is expanded for drugs with a coefficient of variation of about 25% or higher. In fact, the current definition of highly variable drugs is a within-subject coefficient of variation of 30% or more for the AUC or $C_{\max}$ [2]. For the log-transformed variable Y this corresponds to a within-subject standard deviation of 0.2936 or higher provided Y is normally distributed. Generally, for a log-normally distributed variable X the coefficient of variation γ of X and the standard deviation σ of $\ln(X)$ are related by

$$\gamma = [\exp(\sigma^2) - 1]^{1/2} \quad \text{and} \quad \sigma = [\ln(1 + \gamma^2)]^{1/2}$$

γ and σ are very close for small values and become more discrepant for larger values.

In addition to rejecting the null hypothesis (1) the FDA proposal currently under evaluation requires the point estimate of the geometric mean ratio to fall within the interval [0.80, 1.25] (cf. [2]). This condition is called the point estimate constraint (PEC):

$$|m_T - m_R| \leq \theta_A$$

where m_R and m_T denote the estimated mean values of Y for the reference and the test product, respectively.

At the 2007 PhUSE meeting Tanriverdio and Ring [3] presented a four-period two-treatment crossover study for assessing scaled average bioequivalence. They used a bootstrap method for testing the null hypothesis (1). Hyslop *et al.* [4], within the framework of assessing individual bioequivalence, proposed a more direct approach based on a result of Howe [5]. In the following we apply the method of Hyslop *et al.* [4] for testing the null hypothesis (1) in a three-period two-treatment crossover trial and show how this analysis can be realized with SAS/BASE® and SAS/STAT®. In addition, results of simulation studies are presented that estimate the true type I error probability of the asymptotic test for log-normally distributed data.

STUDY DESIGN AND STATISTICAL MODEL

For testing the null hypothesis (1) the design must allow for estimating the within-subject standard deviation σ_R of Y for the reference product. The most parsimonious design for this purpose advocated by Haider *et al.* [2] is a three-period two-treatment crossover trial where the reference product is given twice to each subject. The design is built on three treatment sequences as follows.

Treatment, observed and expected values for the three treatment sequences

Treatment sequence	Treatment			Observed			Expected		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
1: RRT	<i>R</i>	<i>R</i>	<i>T</i>	Y_{11}	Y_{12}	Y_{13}	μ_{11}	μ_{12}	μ_{13}
2: RTR	<i>R</i>	<i>T</i>	<i>R</i>	Y_{21}	Y_{23}	Y_{22}	μ_{21}	μ_{23}	μ_{22}
3: TRR	<i>T</i>	<i>R</i>	<i>R</i>	Y_{33}	Y_{31}	Y_{32}	μ_{33}	μ_{31}	μ_{32}

Regarding notation, $Y_{k,j}$ and $\mu_{k,j}$ denote the observed and expected values, respectively, for treatment sequence k and treatment number j , where $j=1,2$ refers to the reference product *R* and $j=3$ refers to the test product *T*. The total number N of subjects in the study is randomly assigned to the three treatment sequences so that treatment sequence k is applied to n_k subjects. An observation $Y_{k,j}$ made for subject i is denoted by $Y_{i,k,j}$.

It is assumed that the random triplets $Y_{ik}=[Y_{ik1}, Y_{ik2}, Y_{ik3}]$ are independently normally distributed with the number of location parameters equal to the number of cells in the model matrix, i.e. equal to nine. It is further assumed that there are no carry-over effects so that plausible definitions of the treatment means are

$$(2) \quad \mu_T = (\mu_{13} + \mu_{23} + \mu_{33}) / 3$$

$$(3) \quad \mu_R = (\mu_{11} + \mu_{12} + \mu_{21} + \mu_{22} + \mu_{31} + \mu_{32}) / 6$$

According to reference [6] the observed triplet Y_{ik} for subject i with treatment sequence k is assumed to be the sum of the mean μ_k , a random between-subject effect b_{ik} and a random within-subject effect e_{ik} (residual error):

$$[Y_{ik1}, Y_{ik2}, Y_{ik3}] = Y_{ik} = \mu_k + b_{ik} + e_{ik} = [\mu_{k1}, \mu_{k2}, \mu_{k3}] + [b_{ikR}, b_{ikR}, b_{ikT}] + [e_{ik1}, e_{ik2}, e_{ik3}]$$

where b_{ik} and e_{ik} are independently normally distributed with mean zero and variances as follows.

$$\text{var} (e_{ik}) = \begin{bmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \sigma_R^2 & 0 \\ 0 & 0 & \sigma_T^2 \end{bmatrix}$$

$$\text{var} [b_{ikR}, b_{ikT}] = \begin{bmatrix} \omega_{RR} & \omega_{RT} \\ \omega_{RT} & \omega_{TT} \end{bmatrix}$$

i.e.

$$\text{var} (b_{ik}) = \begin{bmatrix} \omega_{RR} & \omega_{RR} & \omega_{RT} \\ \omega_{RR} & \omega_{RR} & \omega_{RT} \\ \omega_{RT} & \omega_{RT} & \omega_{TT} \end{bmatrix}$$

and therefore

$$\text{var} (Y_{ik}) = \begin{bmatrix} \sigma_R^2 + \omega_{RR} & \omega_{RR} & \omega_{RT} \\ \omega_{RR} & \sigma_R^2 + \omega_{RR} & \omega_{RT} \\ \omega_{RT} & \omega_{RT} & \sigma_T^2 + \omega_{TT} \end{bmatrix}$$

Note that the random subject effects are heteroskedastic and depend on the treatment so that they are actually subject-by-treatment effects. The between-subject effects of the reference product and the test product are identical, i.e. b_{ikR} equals b_{ikT} with probability one if and only if

$$0 = \text{var} (b_{ikT} - b_{ikR}) = \omega_{TT} + \omega_{RR} - 2 \omega_{RT} =: \sigma_D^2$$

σ_D^2 is therefore referred to as the subject-by-treatment interaction (cf. formula (5) in reference [4]).

METHOD OF STATISTICAL ANALYSIS

The null hypothesis (1) is equivalent to

$$H_0: |\mu_T - \mu_R| / \sigma_R > \theta_A / \sigma_0$$

This can be put into the linear form

$$(4) \quad H_0: \delta^2 - \theta^2 \sigma_R^2 > 0$$

where

$$\theta = \theta_A / \sigma_0$$

is a known constant. The within-subject standard deviation σ_R and the treatment effect δ

$$\delta = \mu_T - \mu_R$$

are parameters to be estimated from the data. Thus H_0 is rejected at a significance level of α if an upper $(1-\alpha)$ -confidence limit of $\delta^2 - \theta^2 \sigma_R^2$ is not above zero, i.e. if

$$\text{UCL}(\delta^2 - \theta^2 \sigma_R^2) \leq 0$$

where $\text{UCL}(\eta)$ is a generic notation for an upper $(1-\alpha)$ -confidence limit of a parameter η . We will make the restriction that α is less than 0.5. It will be described below how confidence limits for δ^2 and $\theta^2 \sigma_R^2$ can be calculated by using standard analysis of variance techniques. A general method of Howe [5] can then be applied to derive an upper confidence limit for the difference $\delta^2 - \theta^2 \sigma_R^2$.

The method of Howe (Approximation 1 in [5]) can briefly be described as follows. Let X and Y be independent and (approximately) unbiased estimates of parameters ξ and η , respectively. Further let $X^* > X$ and $Y^* > Y$ be independent upper $(1-\alpha)$ -confidence limits for ξ and η , respectively. Then an approximate upper $(1-\alpha)$ -confidence limit for $\xi + \eta$ is given by

$$\text{UCL}(\xi + \eta) = X + Y + [A^2 + B^2]^{1/2}$$

where

$$A = X^* - X \quad \text{and} \quad B = Y^* - Y$$

To apply Howe's method for testing the null hypothesis (4) one thus needs an upper $(1-\alpha)$ -confidence limit for δ^2 and an upper $(1-\alpha)$ -confidence limit for $-\theta^2 \sigma_R^2$, i.e. a lower $(1-\alpha)$ -confidence limit for $\theta^2 \sigma_R^2$. These confidence limits can be calculated by applying standard analysis of variance techniques to the following variables

$$(5) \quad Z_{ik} = Y_{ik3} - (Y_{ik1} + Y_{ik2}) / 2$$

$$(6) \quad U_{ik} = Y_{ik1} - Y_{ik2}$$

The variables Z_{ik} are normally distributed with mean values

$$E(Z_{ik}) = \xi_k = \mu_{k3} - (\mu_{k1} + \mu_{k2}) / 2$$

and equal variances

$$\sigma_Z^2 = \text{var}(Z_{ik}) = \sigma_R^2 / 2 + \sigma_T^2 + \omega_{RR} + \omega_{TT} - 2\omega_{RT} = \sigma_R^2 / 2 + \sigma_T^2 + \sigma_D^2$$

From (2) and (3) it follows that

$$(\xi_1 + \xi_2 + \xi_3) / 3 = \mu_T - \mu_R = \delta$$

i.e. δ is the grand mean in a one-way analysis of variance model fitted to the variables Z_{ik} . This allows for a straightforward calculation of exact confidence limits for δ and δ^2 . Let $X = (Z_{.1} + Z_{.2} + Z_{.3}) / 3$ be the unbiased point estimate of δ and let $X+L$ be the usual upper $(1-\alpha)$ -confidence limit of δ . Then $|X|+L$ is an upper $(1-\alpha)$ -confidence limit for $|\delta|$ and $(|X|+L)^2$ is one for δ^2 . Note that the point estimate X^2 is only approximately unbiased for δ^2 and that the upper confidence limit $(|X|+L)^2$ becomes conservative as $\sqrt{N}(\delta^2 / \sigma_Z^2)$ approaches zero.

The variables U_{ik} are normally distributed with mean values

$$E(U_{ik}) = \eta_k = \mu_{k1} - \mu_{k2}$$

and equal variances

$$\text{var} (U_{ik}) = 2 \sigma_R^2 .$$

In addition, the variables Z_{ik} and U_{ik} are independent because

$$\begin{bmatrix} 1/2, 1/2, -1 \end{bmatrix} \text{var}(\mathbf{Y}_{ik}) \begin{bmatrix} 1, -1, 0 \end{bmatrix}^T = 0$$

The residual mean square error W obtained from a one-way analysis of variance model fitted to the variables U_{ik} is proportional to a chi-squared distribution with v degrees of freedom, specifically:

$$v W / (2 \sigma_R^2) \sim \chi_v^2$$

where v equals $N-3$. Thus $V=W/2$ is an unbiased point estimate of σ_R^2 and a lower $(1-\alpha)$ -confidence limit (LCL) is given by

$$\text{LCL} (\sigma_R^2) = V (v / c_{1-\alpha, v}) = V - (1 - v / c_{1-\alpha, v}) V$$

where $c_{1-\alpha, v}$ is the $(1-\alpha)$ -quantile of the chi-squared distribution with v degrees of freedom. Let X be the unbiased point estimate of δ as above and let X^2+C be the upper $(1-\alpha)$ -confidence limit of δ^2 , i.e. $C=2|X|L+L^2$; further let $V-E$ be the lower $(1-\alpha)$ -confidence limit of σ_R^2 as above, i.e. $E=(1-v/c_{1-\alpha, v})V$. Then according to Howe's method an approximate upper $(1-\alpha)$ -confidence limit of $\delta^2-\theta^2\sigma_R^2$ is given by

$$\text{UCL} (\delta^2 - \theta^2 \sigma_R^2) = X^2 - \theta^2 V + [C^2 + \theta^4 E^2]^{1/2}$$

which is equal to

$$(7) \quad \text{UCL} (\delta^2 - \theta^2 \sigma_R^2) = X^2 - \theta^2 V + [C^2 + \theta^4 V^2 Q^2]^{1/2}$$

with

$$Q = (1 - v / c_{1-\alpha, v})$$

The null hypothesis (4) is rejected at level α if and only if the upper confidence limit (7) is not above zero. Applying some basic algebra it can be shown that the largest value of θ^2 for which (7) is above zero, i.e. the largest value of θ^2 for which the null hypothesis (4) is not rejected at level α , i.e. an upper $(1-\alpha)$ -confidence limit of δ^2/σ_R^2 is

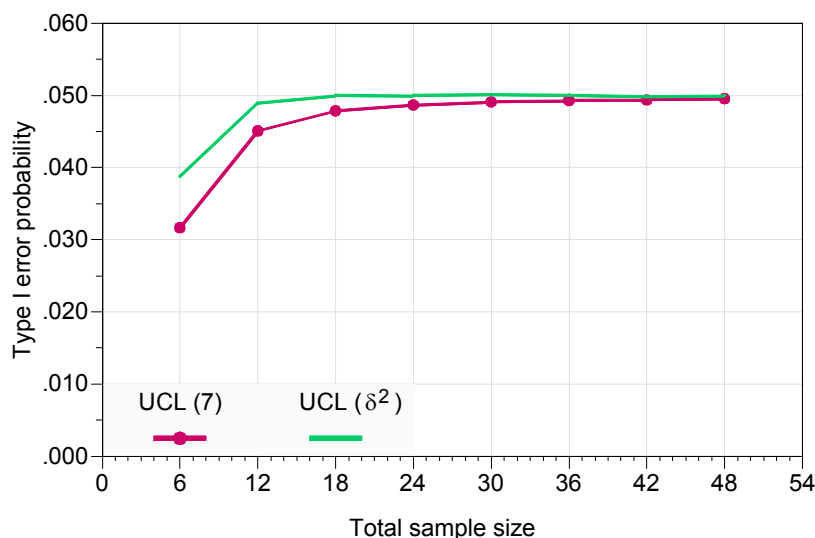
$$(8) \quad \text{UCL} (\delta^2 / \sigma_R^2) = \{ X^2 + [C^2 (1-Q^2) + X^4 Q^2]^{1/2} \} / \{ V (1-Q^2) \}$$

TYPE I ERROR PROBABILITY AND POWER

The above derivation reveals that for a given significance level and a given sample size configuration the operating characteristic of the described test for the null hypothesis (4) depends only on three parameters: δ , σ_R^2 , and $\sigma_T^2 + \sigma_D^2$. The following graph shows the type I error probability simulated for a 5% level test with equal sample sizes per sequence and assuming that $\sigma_R = \sigma_T = 0.3$ and $\sigma_D = 0$. Superimposed is the type I error probability for the upper confidence limit $(|X|+L)^2$ of δ^2 . The confidence limit (7) is conservative for small sample sizes and it appears that this conservatism is due to the conservatism of the confidence limit of δ^2 . We further explored various other scenarios, including some extreme cases. The highest type I error probability we found for the confidence limit (7) was 0.0509. The results are based on 10^7 simulated studies for each scenario.

PhUSE 2009

Type I error probability for the upper confidence limit (7)
and the upper confidence limit of δ^2 (see Text for details)



Haider *et al* [1] investigated the power of the scaled average bioequivalence procedure and contrasted it with the power of the traditional average bioequivalence approach.

SAMPLE SAS PROGRAM

A sample SAS program and an analysis of a sample dataset are provided as Appendices at the end of the paper. The dataset was artificially created only for the purpose of checking the calculations.

REFERENCES

- [1] Haider SH, Makhoul F, Schuirmann DJ, Hyslop T, Davit B, Conner D, Lawrence XY: Evaluation of a scaling approach for the bioequivalence of highly variable drugs. *The AAPS Journal*, 2008, 10(3): 450 - 454.
- [2] Haider SH, Davit B, Cheng ML, Conner D, LaiMing L, Li QH, Lionberger R, Makhoul F, Patel D, Schuirmann DJ, Lawrence XY: Bioequivalence approaches for highly variable drugs and drug products. *Pharmaceutical Research*, 2008, 25(1): 237 - 241.
- [3] Tanriverdio F, Ring A: Implementation of bootstrapping in a replicative bioequivalence study with a highly variable drug. PhUSE 2007, Paper ST03.
- [4] Hyslop T, Hsuan F, Holder DJ: A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine*, 2000, 19: 2885-2897.
- [5] Howe WG: Approximate confidence limits on the mean of $X+Y$ where X and Y are two tabled independent random variables. *Journal of the American Statistical Association*, 1974, 69(347): 789-794.
- [6] Chinchilli VM, Esinhart JD, Barr WH: Analysis of multiple-dose bioequivalence studies. *Journal of Biopharmaceutical Statistics*, 1994, 4(3): 423-435.

CONTACT INFORMATION

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

Jochen Mueller-Cohrs
CSL Behring GmbH
35041 Marburg
Germany
Email: jochen.mueller-cohrs@cslbehring.com

APPENDIX I: SAMPLE SAS PROGRAM

*Three-period two-treatment crossover trial for scaled average bioequivalence (SABE);

*Sample SAS program for the analysis;

options validvarname=v6;

***The sample dataset RAW has the following variables;**

*PARAM name of the pharmacokinetic parameter;

*SUBJECT subject identifier;

*PERIOD period of observation, value of PERIOD from {1,2,3};

*TREATMNT treatment of SUBJECT in PERIOD, value of TREATMNT from {R,R,T};

*VALUE value of the parameter for a given PERIOD and TREATMNT;

data raw;

length param \$8 subject period 8 treatmnt \$1 value 8;

input param--value;

cards;

AUC	1	1	R	20905
AUC	1	2	R	23081
AUC	1	3	T	12011
AUC	2	1	R	21456
AUC	2	2	R	18773
AUC	2	3	T	28972
AUC	3	1	R	29176
AUC	3	2	R	19971
AUC	3	3	T	12542
AUC	4	1	R	17574
AUC	4	2	R	9508
AUC	4	3	T	17561
AUC	5	1	R	38178
AUC	5	2	T	27386
AUC	5	3	R	18666
AUC	6	1	R	18367
AUC	6	2	T	22416
AUC	6	3	R	39906
AUC	7	1	R	46400
AUC	7	2	T	30068
AUC	7	3	R	33210
AUC	8	1	T	21605
AUC	8	2	R	28956
AUC	8	3	R	21948
AUC	9	1	T	34700
AUC	9	2	R	27746
AUC	9	3	R	29383
AUC	10	1	T	37439
AUC	10	2	R	12685
AUC	10	3	R	17711

run;

PhUSE 2009

***From dataset RAW create dataset DIF with variables SEQUENCE, ...;**
***log-transformed values Y1-Y3, and contrasts Zik [formula (5)] and Uik [formula (6)];**

```
proc sort data=raw; by param subject treatmnt period; run;

data dif(keep=param subject sequence y1-y3 zik uik);
  *;
  length sequence $3;
  *;
  array y_(3) y1-y3;
  *;
  i=0;
  *;
  do until(last.subject);
    set raw;
    by param subject treatmnt period;
    i=i+1;
    if length(treatmnt)>1 OR verify(treatmnt,"RRT") then do;
      put / subject= " Treatment not in {R,R,T}";
      abort;
    end;
    if i=3 AND treatmnt="R" then do;
      put / subject= " More than two observations for treatment R";
      abort;
    end;
    if i<3 AND treatmnt="T" then do;
      put / subject= " Fewer than two observations for treatment R";
      abort;
    end;
    if period ^IN(1 2 3) then do;
      put / subject= " Period not in {1,2,3}";
      abort;
    end;
    if i>3 then do;
      put / subject= " Subject has more than three observations";
      abort;
    end;
    if i=3 then do;
      if period=1 then sequence="TRR";
      if period=2 then sequence="RTR";
      if period=3 then sequence="RRT";
    end;
    y_(i)=log(value);
  end;
  *;
  if nmiss(y1,y2,y3) then do;
    put / subject= " Subject has incomplete observations";
    abort;
  end;
  *;
  zik = y3 - (y1+y2)/2;
  uik = y1 - y2;
run;
```

PhUSE 2009

***One-way ANOVA for variables Zik and Uik, estimate of the grand mean for Zik;**

```
proc glm data=dif;
  by param;
  class sequence;
  model uik zik = sequence / ss3;
  estimate "Delta" intercept 1;
  *;
  ods output estimates   =est(where=(dependen="ZIK"));
  ods output overallanova=ano(where=(dependen="UIK" AND source="Error"));
quit;
```

***Putting everything together as described in the paper;**

```
data result(keep=param alpha--sabe);
  length alpha theta df xxx lll ccc vvv qqg ucl 8 nulhypo pec sabe $3;
  *;
  alpha = 0.05;
  theta = log(1.25)/0.25;
  *;
  merge est(keep=param estimate stderr) ano(keep=param df ms);
  by param;
  *;
  tinv= tinv(1-alpha,df);
  cinv= cinv(1-alpha,df);
  *;
  xxx = estimate;          *Variable X in the paper;
  lll = tinv*stderr;       *Variable L in the paper;
  ccc = 2*abs(xxx)*lll + lll**2; *Variable C in the paper;
  vvv = ms/2;             *Variable V in the paper;
  qqg = 1 - df/cinv;      *Variable Q in the paper;
  *;
  *Square root of Formula (8): UCL for |Delta|/Sigma_R;
  *;
  ucl = sqrt( (xxx**2 + sqrt((1-qqg**2)*ccc**2 + (qqg*xxx**2)**2))/((1-qqg**2)*vvv) );
  *;
  *Check of SABE criterion [null hypothesis (4) and point estimate constraint PEC];
  *;
  if n(ucl) then do;
    if ucl > theta then nulhypo="ACC";
    else nulhypo="REJ";
    if abs(xxx) > log(1.25) then pec="NO";
    else pec="YES";
    if nulhypo="REJ" and pec="YES" then sabe="YES";
    else sabe="NO";
  end;
  *;
run;

TITLE "Dataset RESULT";
proc print data=result; id param; format xxx--ucl 12.6; run;
```

PhUSE 2009

APPENDIX II: OUTPUT OF THE SAMPLE SAS PROGRAM

PARAM=AUC

The GLM Procedure

Class Level Information

Class	Levels	Values
SEQUENCE	3	RRT RTR TRR

Number of Observations Read	10
Number of Observations Used	10

PARAM=AUC

The GLM Procedure

Dependent Variable: UIK

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.15239027	0.07619514	0.32	0.7370
Error	7	1.67265177	0.23895025		
Corrected Total	9	1.82504205			

R-Square	Coeff Var	Root MSE	UIK Mean
0.083500	411.4832	0.488825	0.118796

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEQUENCE	2	0.15239027	0.07619514	0.32	0.7370

Parameter	Estimate	Standard Error	t Value	Pr > t
Delta	0.10344231	0.15600489	0.66	0.5285

PhUSE 2009

PARAM=AUC

The GLM Procedure

Dependent Variable: ZIK

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.45146865	0.22573432	1.00	0.4143
Error	7	1.57711585	0.22530226		
Corrected Total	9	2.02858450			

R-Square	Coeff Var	Root MSE	ZIK Mean
0.222554	-8097.025	0.474660	-0.005862

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SEQUENCE	2	0.45146865	0.22573432	1.00	0.4143

Parameter	Estimate	Standard Error	t Value	Pr > t
Delta	0.00973332	0.15148417	0.06	0.9506

Dataset RESULT

PARAM	ALPHA	THETA	DF	XXX	LLL	CCC
AUC	0.05	0.89257	7	0.009733	0.286999	0.087955

PARAM	VVV	QQQ	UCL	NULHYPO	PEC	SABE
AUC	0.119475	0.502386	0.923302	ACC	YES	NO