Evaluation of Replicate Designs for Average Bioequivalence according to EMA's Guideline with Phoenix™ WinNonlin® (2012 Pharsight, A Certara Company, Tripos L.P.)

Helmut Schütz
BEBAC
Consultancy Service for
Bioequivalence and Biovailability Studies
Vienna, March 2013

Version 3.4

Content

1 Disclaimer

Although this white paper contains information of a legal nature, it has been developed for informational purposes only – supporting users in validating software – and does not constitute legal advice as to the current operative laws, regulations, or guidelines of any jurisdiction. In addition, because new standards are issued on a continuing basis, this white paper is not an exhaustive source of all current applicable laws, regulations, and guidelines in the field. While reasonable efforts have been made to assure the accuracy and completeness of the information provided, researchers and other individuals should check with local authorities before starting research activities.

Neither BEBAC nor the author make any representations or warranties regarding the accuracy of its content. Anybody using or relying upon information or opinions set out in this document, does so at their own risk. Validation of the computer system must occur in the user's environment and must account for the user's unique set of user and functional requirements. This document does not replace the user's computer system validation that is expected by the EMA and other regulatory authorities.

2 Introduction

In EMA's current "Guideline on the Investigation of Bioequivalence"[1] some ambiguities in the evaluation of replicate design studies – required if Referenced-Scaled Average Bioequivalence (RSABE) for C_{max} is aimed at – called for clarification, which was provided by the CHMP Pharmacokinetics Working Party in March 2011 (Rev. 3) and is available in the latest version (Rev. 6) published in December 2012 as well.[2] This white paper demonstrates the implementation of the three methods provided (in SAS) in the Q&A document in Phoenix/WinNonlin (PHX/WNL). Two datasets are given in the Q&A document:

- data set I: 4-period 2-sequence (RTRT | TRTR) fully replicated, imbalanced (77 subjects), incomplete (missing periods: one period in six cases, two periods in two cases).
- data set II: 3-period 3-sequence (TRR | RTR | RRT) partial replicated, balanced (24 subjects), complete (no missing periods).

Results of datasets reported by EMA (obtained in SAS 9.1, SAS Institute Inc, NC) were compared to ones obtained in PHX/WNL. The OS was Windows XP Professional, Service Pack 3, all patches (German localization; set during runtime to EN-US) on a DELL Precision 670 Workstation (2 × 2.8 GHz Xeon, 4 GB RAM). Workflows were developed in Phoenix 1.3 beta RC3 and tested in the latest version Phoenix / Win-

Nonlin 6.3 (build 6.3.0.395, released 26 March 2012). Version 2.6 of this document describes analyses in previous versions (6.2.1.51, 6.2.0.495, 6.1.0.173) of Phoenix. A preliminary comparison (not presented) showed similar results to Phoenix ≤6.2.1.51 in the latest release of WinNonlin (5.3).
EMA's data sets are available for download at BEBAC's site in Excel and Phoenix 6.2 format:

- http://bebac.at/downloads/Validation Replicate Design EMA.xls
- http://bebac.at/downloads/EMA Replicate Data Sets.phxproj

3 Procedures

3.1 Data Import

Open the project file, or – alternatively – import the Excel file (if WinNonlin 5.3 is used). All following steps are given for data set I (full replicate), but are applicable to data set II (partial replicate) as well. The only difference in setups – for assessing residuals – is given at the end of Section 3.6.

3.2 Evaluation by EMA's Method C (PHX' default BE Model)

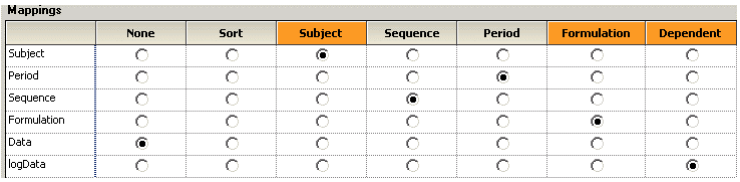
Phoenix' default in evaluating replicate design studies for ABE follows FDA's Guidance "Statistical Approaches to Establishing Bioequivalence".[3] SAS code is given in Appendix E (page 34); the same code is given in the Q&A document (page 21) and referred to as **Method C**.

Note This method is not recommended by the EMA.

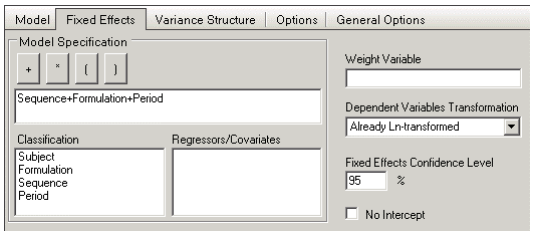
```
proc mixed data=replicate;
classes sequence subject period formulation;
model logDATA= sequence period formulation / ddfm=satterth;
random formulation/type=FA0(2) sub=subject G;
repeated/grp=formulation sub=subject;
estimate 'test-ref' formulation -1 1/ CL alpha=0.10;
run;
```

data set I ➔ Send To ➔ NCA and Toolbox ➔ Bioequivalence

Map logData as Dependent in order to avoid rounding issues (EMA's evaluation was performed on log-transformed data):



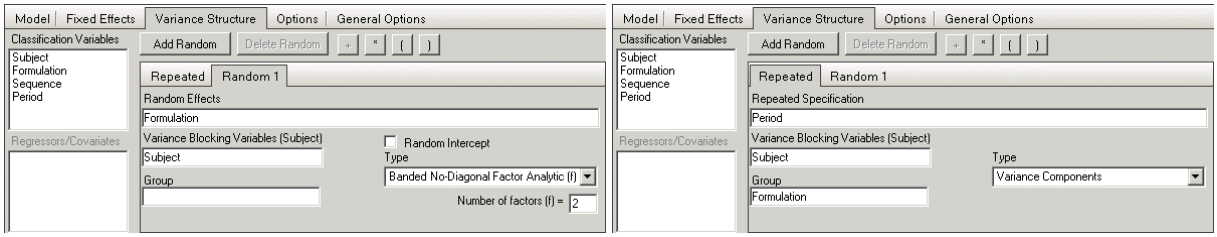
Leave default settings unchanged, except Dependent Variables Transformation ➔ Already Ln-transformed:



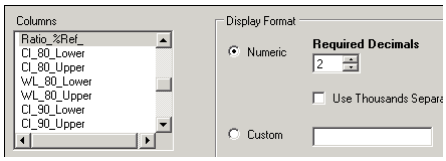
Note Sometimes (*i.e.*, with other datasets) Phoenix fails in 'recognizing' the underlying design. If this happens, specify the model as:

[Fixed Effects]: Sequence+Formulation+Period
Variance Structure [Random 1]: Random Effects = Formulation, Variance Blocking Variables (Subject) = Subject, Type = Banded No-Diagonal Factor Analytic (f), Number of factors (f) = 2

Variance Structure [Repeated]: Repeated Specification = Period, Variance Blocking Variables (Subject) = Subject, Group = Formulation, Type = Variance Components



After executing the Workflow double-click **Output Data** ➔ Average Bioequivalence and change Display Format of fields Ratio_%Ref_, CI_90_Lower, CI_90_Upper to ⦿ Numeric Required Decimals 2.



Close the worksheet.

Ratio_Ref	CI_90_Lower	CI_90_Upper
115.66	107.10	124.89

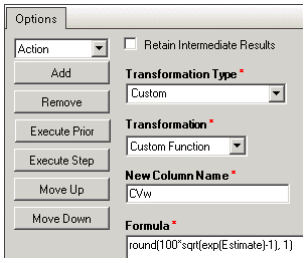
Output Data ➔ Final Variance Parameters

Parameter	Estimate
lambda(2,2)_11	8.3391924E-07
Var(Period*Formulation*Subject)_21	0.20211811
Var(Period*Formulation*Subject)_22	0.11739421

Note Var(Period*Formulation*Subject)_21 and Var(Period*Formulation*Subject)_22 are the estimated within (*intra*-) subject variances of the reference $\left(s^2_{WR} \approx \hat{\sigma}^2_{WR}\right)$ and test formulation $\left(s^2_{WT} \approx \hat{\sigma}^2_{WT}\right)$. lambda(2,2)_11 is the subject × formulation interaction term.*

After filtering results in the Data Wizard for within subject variances (not shown), add a Custom Transformation on the Estimates in order to get the within subject CV% of formulations according to:

$$CV_W\% = 100\sqrt{e^{s^2_W} - 1}$$



New Column Name: CVw

Formula: round(100*sqrt(exp(Estimate)-1), 1)

Note Rounding to one decimal place should only be done in order to facilitate comparison with EMA's results. If the CV is used in sample size estimation for other studies or to judge whether or not scaling for C_{max} is applicable, more decimal places should be employed.

Parameter	Estimate	CVw
Var(Period*Formulation*Subject)_21	0.20211811	47.3
Var(Period*Formulation*Subject)_22	0.11739421	35.3

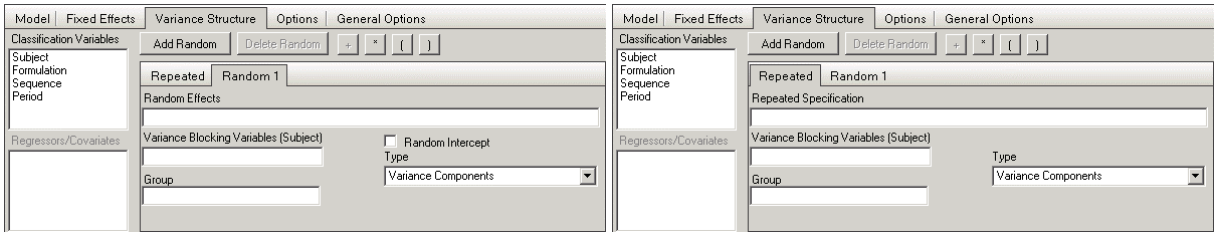
* These parameters are neither mentioned in Phoenix' online-Help nor the User's Guide.

3.3 Evaluation by EMA's Method A (modified BE Model)

EMA's **Method A** treats all effects as fixed (rather than random), assumes equal variances of test and reference, and no subject-by-formulation interaction; only a common within (*intra*-) subject variance $s_W^2 \approx \hat{\sigma}_W^2$ is estimated. SAS-code:*

```
proc glm data=replicate;
class formulation subject period sequence;
model logDATA= sequence subject(sequence) period formulation;
estimate "test-ref" formulation -1+1;
test h=sequence e=subject(sequence);
lsmeans formulation / pdiff=control("R") CL alpha=0.10;
run;
```

Start with PHX' default model. Change [Fixed Effects]: to Sequence+Subject(Sequence)+Period+Formulation and delete under Variance Structure *all* variables – both at [Random 1] and [Repeated].



After executing the Workflow double-click **Output Data** ➔ Average Bioequivalence and change Display Format of fields Ratio_%Ref_, CI_90_Lower, CI_90_Upper to ◉ Numeric Required Decimals 2. Close the worksheet.

Ratio_%Ref_	CI_90_Lower	CI_90_Upper
115.66	107.11	124.89

Output Data ➔ Final Variance Parameters

Parameter	Estimate
Var(Residual)	0.15999519

Note Although the estimated common within subject variance is given in the output and the within subject CV could be calculated according to

$CV_W\% = 100\sqrt{e^{s_W^2} - 1}$,

CV$_W$: 41.65%

EMA requires a different calculation, which is given in Section 3.5.

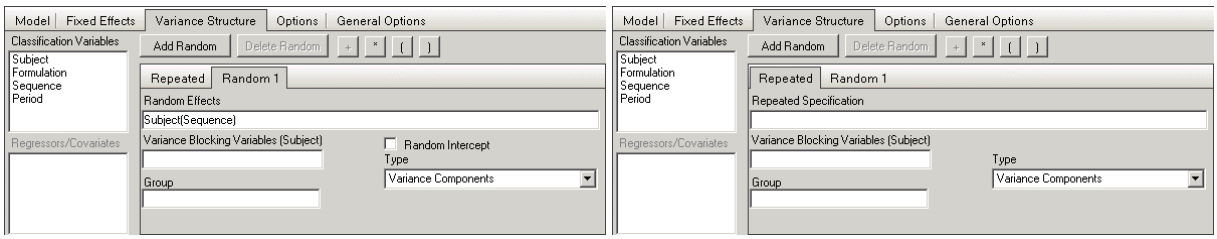
3.4 Evaluation by EMA's Method B (modified BE Model)

Method B is similar to Method A (common variances, no subject-by-formulation interaction), but random effects for subjects. Common within (*intra*-) subject variance $\left(s_W^2 \approx \hat{\sigma}_W^2\right)$ and between (*inter*-) subject variance $\left(s_B^2 \approx \hat{\sigma}_B^2\right)$ are estimated. SAS-code:

```
proc mixed data=replicate;
class formulation subject period sequence;
model logDATA= sequence period formulation;
random subject(sequence);
estimate "test-ref" formulation -1 1 / CL alpha=0.10;
run;
```

* Superfluous `adjust=t` in `lsmeans` statement removed (http://forum.bebac.at/forum_entry.php?id=6794).

Start with PHX' default model. Change [Fixed Effects]: to Sequence+Period+Formulation.
Set at [Variance Structure]: [Random 1] to Subject(Sequence). Delete all variables at [Repeated].



After executing the Workflow double-click **Output Data** ➔ Average Bioequivalence and change Display Format of fields Ratio_%Ref_, CI_90_Lower, CI_90_Upper to ◉ Numeric Required Decimals 2. Close the worksheet.

Ratio_%Ref_	CI_90_Lower	CI_90_Upper
115.73	107.17	124.97

Output Data ➔ Final Variance Parameters

Parameter	Estimate
Var(Sequence*Subject)	0.70693795
Var(Residual)	0.16010032
Intersubject CV	1.0137912
Intrasubject CV	0.41668766

Note Although the common within (*intra*-) subject CV is already given in the output, EMA requires a different evaluation, which is given in the following Section.

CV_w: 41.67%

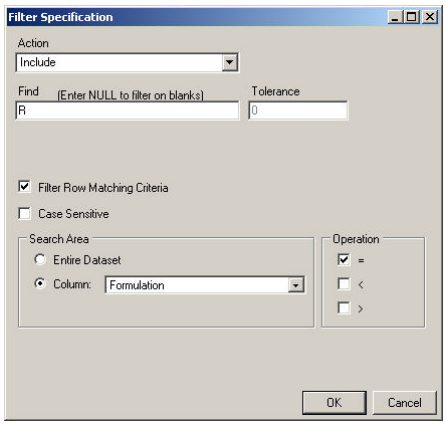
3.5 Calculation of CV_{WR} and Scaled Acceptance Range

Data of the test are removed and repeated reference administrations fitted to a model with fixed effects only. SAS-code (page 23):

```
data var;
set replicate;
if formulation='R';
run;

proc glm data=var;
class subject period sequence;
model logDATA= sequence subject(sequence) period;
run;
```

In the Data Wizard add a Built In Filter including only reference treatments.



Rename Data Wizard to References.

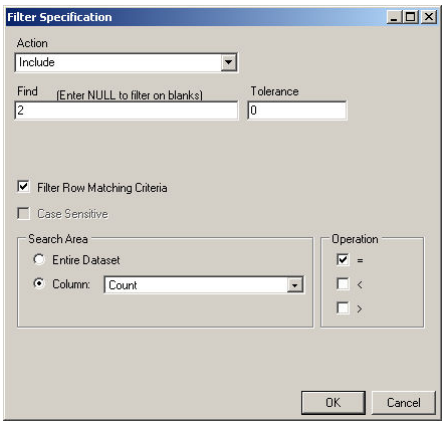
Final Results ➔ Result ➔ Send To ➔ Data ➔ Split Worksheet
Map Subject to Sort; in the Options uncheck ❒ Create unique values worksheets

Get no. of admin's / subject.

Output Data ➔ Unique Values ➔ Send To ➔ Data ➔ Data Wizard

Keep only subjects with two admin's of the reference.

add a Built In Filter including only complete data (*i.e.*, subjects 24, 31, 67, and 71 are excluded).



Rename Data Wizard to Subjects with two administrations.

Final Results ➔ Result ➔ Send To ➔ Data ➔ Join Worksheets

Get the required data of these subjects.

Map Subject to Sort; in the Options check ☑ Inner Join:

Mappings

	None	Sort	Source Column
Subject	○	◉	○
Count	◉	○	○

Map References.Result as the second worksheet. Map Subject to Sort; Period, Sequence, and logData as Source Column:

We don't need the formulation (only R anyway) and untransformed data.

Mappings

	None	Sort	Source Column
Subject	○	◉	○
Period	○	○	◉
Sequence	○	○	◉
Formulation	◉	○	○
Data	◉	○	○
logData	○	○	◉

Rename Data Wizard to Complete data.

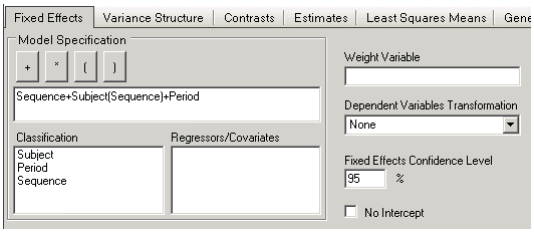
Final Results ➔ Result ➔ Send To ➔ NCA and Toolbox ➔ Linear Mixed Effects

Map Subject, Period, Sequence as Classification and logData as Dependent:

Mappings

	None	Sort	Classification	Regressor	Dependent
Subject	○	○	◉	○	○
Period	○	○	◉	○	○
Sequence	○	○	◉	○	○
logData	○	○	○	○	◉

Specify [Fixed Effects] as Sequence+Subject(Sequence)+Period; keep Dependent Variables Transformation ➔ None, leave all other tabs with their defaults (empty):



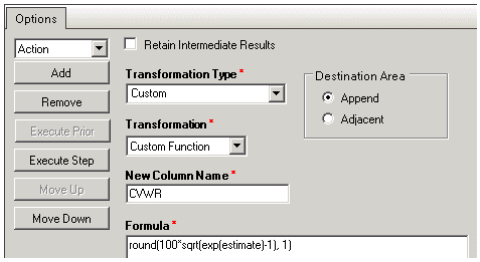
Rename Linear Mixed Effects to LME Ref variability.

After executing the Workflow you should get:

Output Data ➔ Final Variance Parameters

Parameter	Estimate
Var(Residual)	0.1993136

DW In the Data Wizard set up a series of Custom Transformations in order to calculate the within subject CV% of the reference, the scaled Acceptance Range based on the regulatory constant k = 0.760, the percent of the reference to detect, and the Anderson-Hauck limits:



New Column Name: CVWR
Formula: round(100*sqrt(exp/Estimate)-1), 1)
Add ➔ New Column Name: L
Formula: if(CVWR<=30, '80.00 (no scaling)', if(CVWR<=50, round(100*exp(-0.76*sqrt(Estimate)), 2), 69.84))
Add ➔ New Column Name: U
Formula: if(CVWR<=30, '125.00 (no scaling)', if(CVWR<=50, round(100*exp(+0.76*sqrt(Estimate)),2), 143.19))
Add ➔ New Column Name: Pct_to_detect
Formula: if(CVWR<=30, 20.00, 100-L)
Add ➔ New Column Name: AH_LL
Formula: if(CVWR<30, 0.8000, L/100)
Add ➔ New Column Name: AH_UL
Formula: if(CVWR<30, 1.2500, U/100)

Note Rounding CV_{WR} to one decimal place should only be done in order to facilitate the comparison with EMA's results. If CV_{WR} is used in sample size estimation for other studies or to judge whether or not scaling for C_{max} is applicable, more decimal places should be employed.

Note Widening of the Acceptance Range is limited at 30% < CV_{WR} ≤ 50%. If CV_{WR} > 50% is found in the study, the widening is treated *as if* resulting from CV_{WR} = 50% (AR 69.84 – 143.19%).

Parameter	Estimate	CVWR	L	U	Pct_to_detect	AH_LL	AH_UL
Var(Residual)	0.1993136	47.0	71.23	140.40	28.77	0.7123	1.4040

Note Example output for CV 25%, 30%, 50%, 55%:

Parameter	Estimate	CVWR	L	U	Pct_to_detect	AH_LL	AH_UL
Var(Residual)	0.06062462	25.0	80.00 (no scaling)	125.00 (no scaling)	20.00	0.8000	1.2500
Var(Residual)	0.08617770	30.0	80.00 (no scaling)	125.00 (no scaling)	20.00	0.8000	1.2500
Var(Residual)	0.22314355	50.0	69.84	143.19	30.16	0.6984	1.4319
Var(Residual)	0.26428550	55.0	69.84	143.19	30.16	0.6984	1.4319

3.6 Box Plots and QQ-Plots of Model Residuals

According to EMA the applicant has to demonstrate that CV_{WR} > 30% is not caused by outliers. Formal statistical tests are not acceptable; however, box plots were suggested.[4] For details on the analysis of model residuals see Chow and Liu (2009)[5] and Schall *et al.* (2010).[6] The (internally) studentized intra-subject residuals* for two sequences are given by:

$$\tilde{e}_{i1k} = \frac{\hat{e}_{i1k}}{\sqrt{\frac{n_k - 1}{2n_k}\mathrm{MS}_{\mathrm{intra}}}},\ i = 1,2,\ldots,n_k,\ k = 1,2$$

where only the first administration of subject i in sequence k is used.

* Theoretically $\tilde{e}_{i1k}$ has mean 0 and variance 1. Values of data set I are $-5.514 \cdot 10^{-14}$ and 1.014.

Sequences in data set I are RTRT | TRTR. Therefore, data from period 1 (sequence 1: RTRT) and period 2 (sequence 2: TRTR) are required.

Complete data.Result ➔ Send To ➔ NCA and Toolbox ➔ Descriptive Stats

Map Sequence to Sort and logData to Summary. Rename Descriptive Stats to Administrations per sequence.

Admin's / sequence = N.

Output Data ➔ Statistics ➔ Send To ➔ Data Wizard.

Set up an Arithmetic Transformation:

Transformation: x / n

n: 2

map N to x column

New Column Name: nk

We need the number subjects / sequence: n_k = N/2.

Add a filter excluding everything except Variable, Sequence, and n_seq.

Add Properties; Old Column: Variable => New Column Name: Dependent

Change name; needed for later join.

Rename Data Wizard to Subjects per sequence.

Dependent	Sequence	nk
logData	RTRT	36
logData	TRTR	37

Final Results ➔ Result ➔ Send To ➔ Data ➔ Join Worksheets

Get residuals from LME.

Map Dependent and Sequence to Sort and nk to Source Column. In [Options] check ☑ Inner Join.

Map LME Ref variability.Residuals from Section 3.5 as the second worksheet.

Map Dependent and Sequence to Sort; Subject, Period, Predicted, and Residual as Source Column.

Rename Join Worksheets to Join residuals.

Output Data ➔ Result ➔ Send To ➔ Data ➔ Join Worksheets

Get the MSE from LME.

Map Dependent to Sort and all other variables to Source Column. In [Options] uncheck ❒ Inner Join.

Map LME Ref variability.Final Variance Parameters from Section 3.5 as the second worksheet.

Map Dependent to Sort and Estimate as Source Column. Rename Join Worksheets to Join MSE.

Output Data ➔ Result ➔ Send To ➔ Data Wizard.

Only 1st administration in each sequence is needed.

Set up a Custom Inclusion Filter: (Sequence='RTRT' and Period=1) or (Sequence='TRTR' and Period=2), exclude Period.

Set up a Custom Transformation in order to calculate the studentized within subject residuals

Transformation Type: Custom Function

New Column Name: StudentRes

Formula: Residual/sqrt(Estimate*(nk-1)/(2*nk))

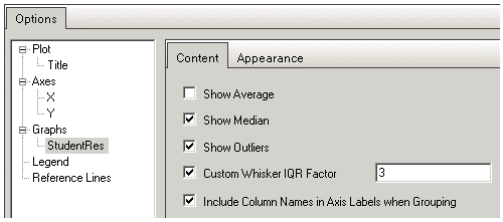
Set up a Built In Filter: Exclude nk and Estimate. Rename Data Wizard to Studentized Residuals.

Ascending by Subject. Only subjects with complete data (without 24, 31, 67, 71).

	Dependent	Residual	Predicted	Subject	Sequence	StudentRes
1	logData	0.2624	7.4722	1	RTRT	0.8429
2	logData	0.1373	7.7218	2	TRTR	0.4410
...	...	...	...	...	...	...
72	logData	-0.0008	7.0179	77	RTRT	-0.0026
73	logData	-0.0076	9.9296	78	RTRT	-0.0246

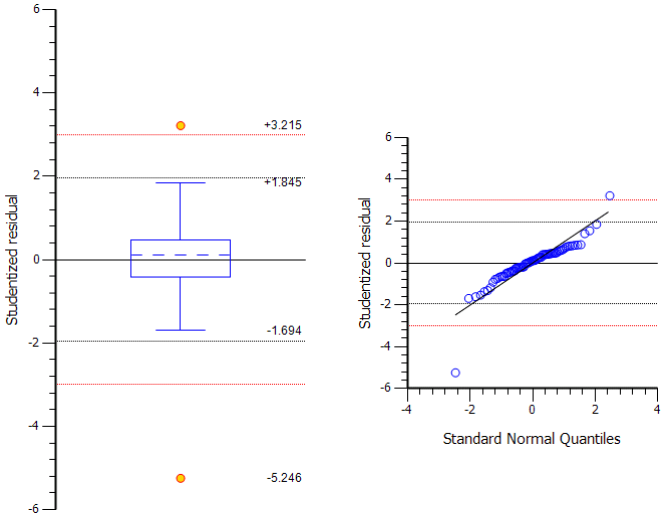
Output Data ➔ Result ➔ Send To ➔ Plotting ➔ Box Plot. Map StudentRes to Y. Click on StudentResid; in [Content] uncheck ❒ Show Average and set ☑ Custom Whisker IQR Factor to 3.

Custom Whisker IQR Factor 3: *'severe'* outliers; (PHX' default is 1.5 ~ *'moderate'* outliers).



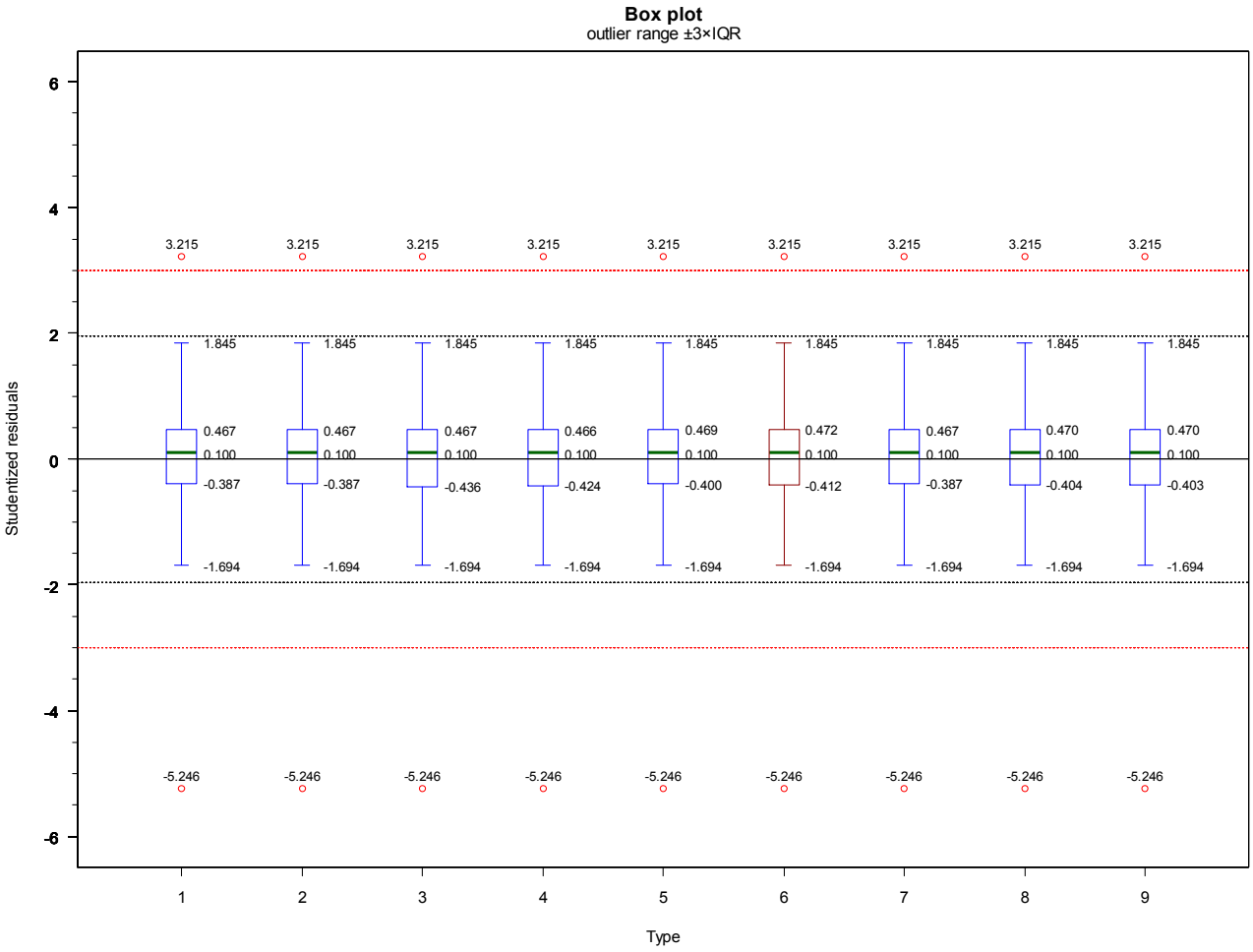
Two severe outliers (outside ±3×IQR) are identified, namely subjects 45 (studentized residual -5.246) and 52 (+3.215).* Add Y Reference Lines to [-3, -1.96, 0, +1.96, +3].†

QQ Optionally **Output Data** ➔ Result ➔ Send To ➔ Plotting ➔ QQ Plot. Map StudentRes to Y. Add Y Reference Lines to [-3, -1.96, 0, +1.96, +3].



Note Different methods to calculate quartiles exists. Phoenix's method agrees with Minitab and SPSS (Type 6), but other algorithms are implemented in, *e.g.*, SAS (default: Type 2), Excel, S+, and R (default: Type 7). If behavior consistent with SAS is desired, export of data to R is suggested, where all of the options mentioned above are available.[7] Below a comparison of different options.

'Type 1–7' acc. to R's terminology.



* Whiskers/hinges of the boxplot are defined as the *smallest* datum ≥Q1–3×IQR and the *largest* datum ≤Q3+3×IQR. Here Q1–3×IQR is –3.063 and Q3+3×IQR is +3.123. The lower whisker is given by subject 46's studentized residual (–1.649); the upper whisker is given by subject 41's (+1.845).

† $\mu \pm 1.96\sigma$ covers 95.00% of normal distributed data; $\mu \pm 3\sigma$ covers 99.73%.

The (internally) studentized intra-subject residuals* for the partial replicate design are given by:

$$\tilde{e}_{i1k} = \frac{\hat{e}_{i1k}}{\sqrt{\frac{n_k - 1}{2n_k}\mathrm{MS}_{\mathrm{intra}}}},\ i = 1,2,\ldots,n_k,\ k = 1,2,3$$

where only the first administration of subject i in sequence k is used.

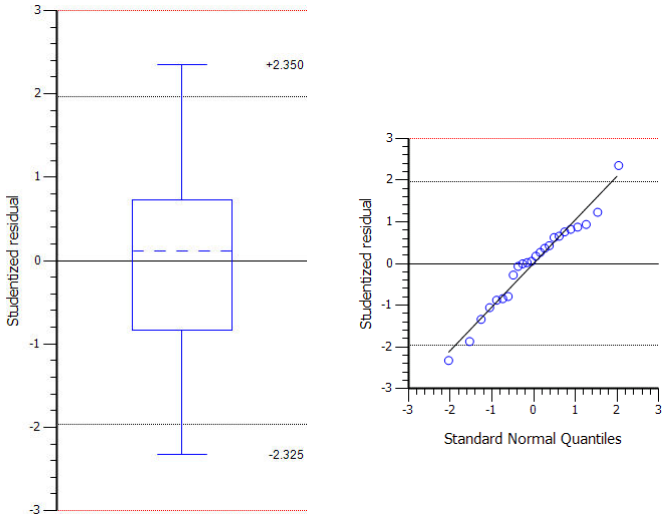
Sequences in data set II are TRR | RTR | RRT. Therefore, data from period 2 (sequence 1: TRR) and period 1 (sequences 2: RTR and 3: RRT) are required. Calculate the studentized residuals as above.

Only 1st administration in each sequence is needed.

Note Which periods of the sequences have to be used depend on the design of the study – and might be different from this data set. Sequences might also be coded differently (in data set II: TRR = 1, RTR = 2, RRT = 3). *Great care must be taken in this step.* Examples:

Design	Filter
RTRT | TRTR†	(Sequence='RTRT' and Period=1) or (Sequence='TRTR' and Period=2)
RRTT | TTRR | RTTR | TRRT	(Sequence='RRTT' and Period=1) or (Sequence='TTRR' and Period=3) or (Sequence='RTTR' and Period=1) or (Sequence='TRRT' and Period=2)
TRR | RTR | RRT†	(Sequence='TRR' and Period=2) or (Sequence='RTR' and Period=1) or (Sequence='RRT' and Period=1)
TRT | RTR‡	(Sequence='TRT' and Period=2) or (Sequence='RTR' and Period=1)
TRT | RTR | TRR | RTT‡	(Sequence='TRT' and Period=2) or (Sequence='RTR' and Period=1) or (Sequence='TRR' and Period=2) or (Sequence='RTT' and Period=1)

No outliers are detected.
However, since CV_{WR} is only 11.2%, the acceptance range may not be widened for this data set.



* Theoretically $\tilde{e}_{i1k}$ has mean 0 and variance 1. Values of data set II are $4.313 \cdot 10^{-3}$ and 1.093.

† Recommended by the FDA.

‡ Allows additionally the estimation of CV_{WT} in a three-period study. Estimates may be less precise since formulations are repeated in only ½ of sequences.

4 Results

4.1 Point Estimates, Confidence Intervals

Method	data set I (full replicate)							
	EMA				PHX BE			
	PE	90% CI		width	PE	90% CI		width
A	115.66	107.11	124.89	17.78	115.66	107.11	124.89	17.78
B	115.73	107.17	124.97	17.80	115.73	107.17	124.97	17.80
C	115.66	107.10	124.89	17.79	115.66	107.10	124.89	17.79

Method	data set II (partial replicate)							
	EMA				PHX BE			
	PE	90% CI		width	PE	90% CI		width
A	102.26	97.32	107.46	10.14	102.26	97.32	107.46	10.14
B	102.26	97.32	107.46	10.14	102.26	97.32	107.46	10.14
C	102.26	97.05	107.76	10.71	102.26	97.05	107.76	10.71

4.2 Within subject CVs, Scaled Acceptance Ranges

Method	data set I (full replicate)				
	EMA		PHX LME		
	CV_{WR}	CV_{WT}	CV_{WR}	CV_{WT}	AR
A/B	47.0	NA	47.0	NA	71.23 – 140.40
C	47.3	35.3	47.3	35.3	71.06 – 140.73

Method	data set II (partial replicate)				
	EMA		PHX LME		
	CV_{WR}	CV_{WT}	CV_{WR}	CV_{WT}	AR
A/B	11.2	NA	11.2	NA	80.00 – 125.00
C	11.5	NA*	11.5	8.65	80.00 – 125.00

5 Discussion and Conclusions

For background on different scaling methods see Haidar *et al.* (2008a, 2008b),[8,9] Endrényi and Tóthfalusi (2009),[10] Tóthfalusi *et al.* (2009),[11] FDA (2010, 2011),[12] Karalis *et al.* (2012).[13]

By modifying Phoenix/WinNonlin's LME BE model (Sections 3.3 and 3.4) results reported by EMA[2] were exactly reproduced – both PEs/CIs and within subject CVs. A bug in previous versions (Phoenix ≤6.2.1.51, WinNonlin ≤5.3) was corrected in the current version. We advise users of previous versions to upgrade or apply the workaround based on LME presented in Version 2.6† of this document.

Evaluation of data set II by Model C (FDA's code) indicates an over-specified model. A warning is issued:

```
Newton's algorithm converged with modified Hessian. Output is suspect.
Model may be over-specified. A simpler model could be tried.
```

A similar statement is obtained in SAS 9.21.

```
Convergence criteria met but final hessian is not positive definite.
```

These problems in convergence (due to non-replicated test treatments) explains the difference in estimated CV_{WT} (PHX 8.65% *vs.* SAS 3.87%) and the wider confidence interval observed for the partial replicate data set if compared to both of EMA's methods. However, the method should provide reliable estimates for a fully replicate 3-period designs (*i.e.*, TRT | RTR | TRR | RTT).

* Although CV_{WT} can be directly estimated in Method C, no value is reported by EMA. SAS 9.21 estimates 3.87% (http://forum.bebac.at/forum_entry.php?id=6770).

† The algorithm for calculating studentized residuals in version 2.6 is only approximate; the setup should be modified according to Section 3.6 of the current version.

EMA's Guideline states (Section 4.1.10, page 17):

> *The applicant should justify that the calculated intra-subject variability is a reliable estimate and that it is not the result of outliers.*

At the 3rd EGA Symposium on Bioequivalence members of the CHMP Pharmacokinetics Working Party expressed their opinion that a formal statistical outlier test is not acceptable. However, box plots seem to be an option.[4]

> *The outlier cannot be removed from evaluation […] but should not be taken into account for calculation of within-subject variability and extension of the acceptance range.*
>
> *On a case by case basis, a study could be acceptable if the bioequivalence requirements are met both including the outlier subject (using the scaled average bioequivalence approach and the within-subject CV with this subject) and after exclusion of the outlier (using the within-subject CV without this subject).*
>
> *An outlier test is not an expectation of the medicines agencies but outliers could be shown by a box plot. This would allow the medicines agencies to compare the data between them.*

In *which* cases such a study *cannot* be accepted?

It remains an open issue how this should be done; we suggest the model's studentized residuals to allow comparisons between studies. Two outliers are evident in data set I (subjects 45 and 52: –5.246, +3.215). If these outliers are excluded, CV_{WR} drops from 47.0% to 32.2% and the Scaled Acceptance Range narrows substantially. Bioequivalence can just be demonstrated:

Artificial data set? The study would even pass conventional (unscaled) ABE…

Method	PE	90% CI		Scaled AR (n=77, full data set CV_{WR} 47.0%)		Scaled AR (n=75, reduced data set, CV_{WR} 32.2%)	
A	115.66	107.11	124.89	71.23	140.40	78.79	126.93
B	115.73	107.17	124.97				

The method to identify outliers must be stated in the statistical protocol.

In the current version of Phoenix it is not possible to add a second graph to a box plot ("overlay"). Annotating the plot (values of whiskers/hinges and outliers) has to be done manually by means of reference lines. An enhancement request is filed at Pharsight. For a procedure to get these values download the project file in PHX 1.3 format:

- http://bebac.at/downloads/EMA ABEL Validation PHX6.3.phxproj

We suggest to calculate in routine use CV_{WR} first (Section 3.5) and subsequently assess the data set for outliers (Section 3.6). The scaled limits (if applicable) can then be copied to the [Options] of the Setup of Models A and B (Percent of Reference to Detect, Anderson-Hauck Lower Limit, Anderson-Hauck Upper Limit).

Phoenix/WinNonlin's standard model for ABE in replicate studies based on FDA's Guidance (2001)[3] is not suitable for RSABE according to FDA's current requirements (2010, 2011).[12] We developed a workflow validated against SAS* which is described in another white paper.

6 Acknowledgments

Jiří Hofmann (Zentiva, Prague, Czech Republic) for in-depth discussions on assessing model residuals and providing the R-code for customized box plots. Jean-Michel Cardot (Université d'Auvergne, Clermont-Ferrand, France) for providing studentized residuals in SAS. Ana Henry (Pharsight, USA) for making a license of PHX 1.3 beta RC3 available and taking care in resolving Bug QC 10856. Simon Davis (Pharsight, UK) for streamlining the workflow in the transition to Phoenix/WinNonlin 1.3.

* http://forum.bebac.at/mix_entry.php?id=7871

7 References

[1] European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP)
Guideline on the Investigation of Bioequivalence
Doc. Ref.: CPMP/EWP/QWP/1401/98 Rev. 1/Corr, London, 20 January 2010
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf

[2] European Medicines Agency, Committee for Human Medicinal Products (CHMP)
Questions & Answers: Positions on specific questions addressed to the pharmacokinetics working party
EMA/618604/2008 Rev. 6, London, 10 December 2012
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf

[3] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER)
Guidance for Industry. Statistical Approaches to Establishing Bioequivalence
Rockville, January 2011
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070244.pdf

[4] European Generics Medicines Association
Revised EMA Bioequivalence Guideline – Questions and Answers
Summary of the Discussions held at the 3rd EGA Symposium on Bioequivalence
London, 1 June 2010
Free online resource (acc. 20 March 2011)

[5] S-C Chow and J-p Liu
Design and Analysis of Bioavailability and Bioequivalence Studies
Chapman & Hall/CRC Biostatistics Series, Boca Raton, pp221–56 (2009)

[6] Schall R, Endrényi L, and A Ring
Residuals and Outliers in Replicate Design Crossover Studies
J Biopharm Stat 20(4), 835–49 (2010) doi: 10.1080/10543401003618876

[7] R Development Core Team, R Foundation for Statistical Computing
R: A Language and Environment for Statistical Computing
ISBN 3-900051-07-0. Vienna, Austria, 2012
http://www.R-project.org
R Documentation: Sample Quantiles
http://stat.ethz.ch/R-manual/R-devel/library/stats/html/quantile.html (acc. 27 July 2012)
R Documentation: Draw Box Plots from Summaries
http://stat.ethz.ch/R-manual/R-devel/library/graphics/html/bxp.html (acc. 27 July 2012)

[8] Haidar SH, Davit B, Chen M-L, Conner D, Lee LM, Li QH, Lionberger R, Makhlouf F, Patel D, Schuirmann DJ, and LX Yu
Bioequivalence Approaches for Highly Variable Drugs and Drug Products
Pharmaceutical Research 25(1), 237–41 (2008) DOI: 10.1007/s11095-007-9434-x
Free online resource (acc. 19 March 2011)

[9] Haidar SH, Makhlouf F, Schuirmann DJ, Hyslop T, Davit B, Conner D and LX Yu
Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs
The AAPS Journal 10(3), 450–4 (2008) DOI: 10.1208/s12248-008-9053-4

[10] Endrényi L and L Tóthfalusi
Regulatory Conditions for the Determination of Bioequivalence of Highly Variable Drugs
J Pharm Pharmaceut Sci 12(1), 138–49 (2009)
Free online resource (acc. 19 March 2011)

[11] Tóthfalusi L, Endrenyi L, and A García-Aríeta
Evaluation of Bioequivalence for Highly Variable Drugs with Scaled Average Bioequivalence
Clin Pharmacokinet 48(11), 725–43 (2009) DOI: 10.2165/11318040-000000000-00000

[12] Food and Drug Administration, Office of Generic Drugs (OGD)
Draft Guidance on Progesterone
Rockville, recommended April 2010; revised February 2011
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf

[13] Karalis V, Symillides M, and P Macheras
Bioequivalence of Highly Variable Drugs: A Comparison of the Newly Proposed Regulatory Approaches by FDA and EMA
Pharm Res 29, 1066–77 (2012) DOI: 10.1007/s11095-011-0651-y