**Equivalence testing using existing reference data: an example with genetically modified and conventional crops in animal feeding studies**

Hilko van der Voet1, Paul W. Goedhart1, Kerstin Schmidt2

1 Wageningen University & Research, Biometris, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands

2 BioMath GmbH, Friedrich-Barnewitz-Str. 8, 18119 Rostock-Warnemünde, Germany

# Supplement 1. Details of analysis

This supplementary document describes the precise steps that were taken in the statistical analysis. All programs were run in the statistical program GenStat Release 18.1, and are available from the authors on request.

**Preparation of a single data file for the 13 weeks data**

The following data files have been retrieved from the CADIMA website at 29-11-2016:

* Study A: “20161121 Grace Data Trial A (Data Definition V4).xlsx”
* Study B: “20161121 Grace Data Trial B (Data Definition V4).xlsx”
* Study C: “20161121 Grace Data Trial C (Data Definition V4).xlsx”
* Study D: “20161121 Grace Data Trial D (Data Definition V4).xlsx”
* Study E: “20161121 Grace Data Trial E (Data Definition V4).xlsx”

GenStat program ***01-Combine\*.gen*** combines these 5 data‑files into a single file named ***Grace.xlsx*** and a GenStat backingstore file ***Grace.bst***. A group factor is defined so that trial/feed combinations can be easily selected:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Study | Control | 11% GMO | 33% GMO | 33% Conv-1 | 33% Conv-2 |
| A | 11 | 12 | 13 | 14 | 15 |
| B | 21 | 22 | 23 | 24 | 25 |
| C | 31 | 32 | 33 | - | 35 |
| D | 41 | 42 | 43 | - | - |
| E | 51 | 52 | 53 | - | - |

Moreover a factor called ‘feedname’ is defined with values according to Schmidt et al (2017):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Study | Control | 11% GMO | 33% GMO | 33% Conv-1 | 33% Conv-2 |
| A | DKC6666 | DKC6667-YG-11 | DKC6667-YG-33 | PR33W82 | SY-NEPAL |
| B | PR32T16 | PR33D48-11 | PR33D48-33 | PR32T83 | DKC6815 |
| C | DKC6666 | DKC6667-YG-11 | DKC6667-YG-33 | - | SY-NEPAL |
| D | DKC6666 | DKC6667-YG-11 | DKC6667-YG-33 | - | - |
| E | PR32T16 | PR33D48-11 | PR33D48-33 | - | - |

Notes:

* The values which are read by GenStat were sorted according to Trial\_ID and Animal\_ID to make sure that the order of the observations is the same.
* Original cage and animal numbers are not unique across trials. They have been made unique such that cage numbers start at 100, 200, 300, 400 and 500 for the respective trials, while animal numbers start at 1000, 2000, 3000, 4000 and 5000.
* Organ weights in Study C are not available after 13 weeks as C is a one‑year study.
* For study C Weights are available for all 160 animals while Haematology and Biochemistry are only available for a limited set of 80 animals.
* Values in sheet “Data exclusions” for all studies were already set to missing in the original Excel files. Note that this is not the case when the comment in that sheet reads “will be considered”. This implies that for trial A 2 of these values are set to missing, for trail B none, for trial C only one as the other values are for observations beyond week 13. For trials D and C all values given in “Data exclusions” were set to missing in the original Excel files.
* Organ Weights are expressed as a percentage of the body weight at week 13. This is done for each individual rat.
* Percentage organ weights for Left and Right were summed before analysis. The Left and Right weights are not statistically analysed.
* There are two HGB variables; one in haematology and one in clinical biochemistry. The latter is called cHGB.

Data at the original scale is stored in Excel file ***Grace.xlsx*** in sheets RawAll (both sexes), RawF (Females) and RawM (Males). The latter two sheets only contain the relevant organ weights. To enable data analysis on the log-scale, log-transformed data are stored in sheets LogRawAll, LogRawF and logRawM. Since cage is the experimental unit in each trial, with 2 animals per cage, cage means were calculated and stored for both the original and the log-transformed scale. This was only done for the “M” and “F” sheets giving sheets MeanRawF, MeanRawM, MeanLogRawF and MeanLogRawM. Each sheet also contains the factors trial, cage, sex, group, feed, feedname and animal (animal is not save for the cage means).

The response variables “Monocyte” and “Eosinophil” contain 3 respectively 28 zeroes. Therefore the value of 0.5 was added to these before taking the logarithm.

All the Excel sheets listed above are also stored in GenStat backingstore file ***Grace.bst*** using the same sheetnames as subfile names. In addition each subfile also contains the following pointers:

* fact the six study design factors trial, cage, sex, group, feed, feedname
* Weights the 14 weights
* Intake the 13 feed intakes
* Haematology the haematology variables
* Clinical the Clinical Biochemistry variables
* Organs the organs weights variables (expressed as percentage) summed over Left and Right when applicable
* OrgansLR the separate Left and Right percentage organ weights when applicable
* all all the response variables listed above

Note that the lists of organs for males and females are not the same. So the relevant structures can be retrieved and analysed response group by response group using the following GenStat program, e.g. for Males using subfile MeanLogRawM:

open 'Grace.bst' ; channel=1 ; filetype=back
retrieve [channel=1 ; subfile=MeanLogRawM] fact, Weights, Intake, Haematology, Clinical, Organs
for responseGroup= Weights, Haematology, Clinical, Organs
 for response= responseGroup[]
 *Do statistical analysis for response*
 endfor
endfor

Or alternatively without grouping of response groups

for response= Weights[], Haematology[], Clinical[], Organs[]
 *Do statistical analysis for response*
endfor

Not all cage means are based on two values due to missing observations. These missing observations are listed in the Table below (Program ***03-MissingDataRaw.gen***). Only for Cage 170 in Study A (Female rats with feed Conv‑1) both values of Ca and Na are missing and these are thus the only missing values for the Cage means (Program ***04-MissingDataMean.gen***). The fact that some means are based on one observation rather than two is ignored in the statistical analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cage | Animal | Study | Sex | Feed | Missing |
| 120 | 1039 | A | M | Conv2 | ALP ALT AST Alb Glu Krea TP Urea CHOL Ca Cl K Na P TAG |
| 123 | 1045 | A | M | Conv2 | K |
| 141 | 1082 | A | F | GMO-33 | ALP ALT AST Alb Glu Krea TP Urea CHOL Ca Cl K Na P TAG |
| 162 | 1123 | A | F | Conv2 | ALP ALT AST Alb Glu Krea TP Urea CHOL Ca Cl K Na P TAG |
| 163 | 1125 | A | F | Conv2 | Ovary\_L Ovary |
| 168 | 1135 | A | F | Conv1 | P |
| 170 | 1139 | A | F | Conv1 | Ca Na |
| 170 | 1140 | A | F | Conv1 | CHOL Ca K Na P TAG |
| 176 | 1152 | A | F | Control | Ca K Na P TAG |
| 216 | 2232 | B | M | GMO-11 | ALP ALT AST Alb Glu Krea TP Urea CHOL Ca Cl K Na P TAG |
| 220 | 2240 | B | M | Conv2 | PLT |
| 240 | 2280 | B | M | Control | Ca K Na P TAG Kidney\_R Kidney |
| 255 | 2310 | B | F | GMO-11 | CHOL Ca Cl K Na P TAG |
| 257 | 2314 | B | F | Conv2 | ALP ALT AST Alb Glu Krea TP Urea CHOL Ca Cl K Na P TAG |
| 264 | 2327 | B | F | Conv2 | Ca K Na P |
| 305 | 3009 | C | M | GMO-33 | ALT |
| 315 | 3029 | C | M | GMO-11 | ALT |
| 325 | 3050 | C | M | Control | WBC RBC HGB HCT MCV MCH MCHC PLT LYMcount Lymphocyte Neutrophil Monocyte Eosinophil |
| 352 | 3103 | C | F | GMO-33 | ALP |
| 370 | 3139 | C | F | GMO-11 | ALT |
| 374 | 3147 | C | F | Control | HGB |
| 381 | 3161 | C | F | Conv2 | CHOL |
| 407 | 4013 | D | M | GMO-11 | Alb TP K |
| 408 | 4016 | D | M | GMO-11 | Alb TP K |
| 410 | 4020 | D | M | GMO-11 | Alb TP K |
| 433 | 4101 | D | F | GMO-33 | Alb TP K |
| 437 | 4109 | D | F | GMO-33 | CHOL |
| 440 | 4116 | D | F | GMO-11 | Alb TP K |
| 443 | 4122 | D | F | Control | CHOL |
| 523 | 5045 | E | M | GMO-11 | Alb TP K |
| 525 | 5050 | E | M | GMO-11 | Alb TP K |

GenStat programs ***Table3-Male.gen*** and **Table6-Male.gen**, and similarly for Females, were used to generate Tables 3 and 6 in Schmidt et al (2017) to check whether the same data have been used. For both tables first cage means were calculated and then the relevant means in Tables 3 and 6. A difference of 1 in the last printed number was ignored in comparing the tables. The following discrepancies were found:

* For Alb, TP and K large discrepancies are found in Table 6. This is due to the fact that Schmidt et al (2017) calculated Table 6 **including** the outliers listed in sheet “Data exclusions” for trials D and E. To put it otherwise, when these “Data exclusions” outliers are copied to the sheet “Clinical biochemistry” in the original Excel files, the values in Table 6 are obtained. The outliers are excluded in the analysis presented here.
* For Na a small discrepancy was found in Table 6. This is due to non-rounding of 5 Na values when table 6 was calculated by Schmidt et al (2017). See Appendix B for details.

**Analysis of weights**

**GenStat program: *01-GrowthRate.gen*.**

For each individual animal the following exponential growth curve was fitted:

 $Weight=α+\left(β-α\right) r^{week}$

The parameters of this curve have the following interpretation:

 $α$ final weight
 $β$ initial weight at week=0
 $r$ growth rate

The fitted curves are displayed in a separate file named ***01-GrowthRate.pdf***. The curve fits very well in almost all cases implying that the 14 weekly weights can be summarized by the three parameters $α$, $β$ and $r$. Note that there are some outlying observations, see e.g. rat 1043 on Page 5, rats 2221 and 2222 on Page 7, rat 2240 on Page 10, rats 5039 and 5040 on page 20, rat 1141 on page 24, rats 2313, 2319 and 2320 on page 30, rats 3124 and 3137 on Page 32. However, such outliers do not seem to affect the fit of the curve too much. Male rat 2240 in cage 220 in trial B with feed Conv-2 seems to be ill at the end of the study (Page 10) which results in a small value of Log(r). In most cases the growth rate $r$ is such that the maximal weight $α$ will be obtained in the not too far future. Notable exceptions, with a Log growth rate larger than -0.1 are rat 3010 (Page 13), and rats 4007 (17), 4008 (17), 5054 (18), 1152(21) and 1122 (25).

**Outlier detection**

**GenStat program: *02-ResidualPlots-Male.gen* and *02-ResidualPlots-Female.gen*.**

The cage means, after a log transformation of the original animal data i.e. in MeanLogRawF and MeanLogRawM, were subjected to analysis of variance for each trial and sex separately with feed as treatment and no blocking effect (the design is completely randomized). Grubbs’ outlier test at the 1% level was applied to the residuals to detect outliers. This was done repeatedly, i.e. outliers were set to missing, the data were re-analysed and Grubbs’ test was performed again, until no further outliers were found. In the second cycle one extra outlier was found for Males and the third cycle did not detect any further outliers. The outliers thus found are given in Appendix C.

The spread of the residuals is graphically displayed in separate files ***02‑ResidualPlots-Male.pdf*** and ***02‑ResidualPlots-Female.pdf***. A single plot consists of the residuals for, from left to right, trial A, B, C, D and E. The colour of the residuals denote the feeding group: black for Control, aqua for GMO-11%, blue for GMO-33%, red for Conv‑1 and orange for Conv‑2. Residuals that are outlying according to Grubb’s test have a larger size; an outlying observation is also depicted by an added X in the title of the graph.

The graphs show that these are clear outliers. Single cages frequently have outliers for multiple responses; see Appendix C. Note that the outlier for the growthRate parameter $r$ was already detected in the analysis of weights (cage 220 in trial B with feed Conv-2). The graphs also show that Grubbs’ test is not well suited for simultaneous detection of two outliers. For example for response K in study C there seem to be two outlying observations which was not detected by Grubbs’ test.

The graphs also reveal that the residual variance is not always homogeneous across studies. This aspect is currently not addressed.

The log-transformed data on the cage level with and without the outliers were saved in GenStat backingstore files ***02‑ResidualPlots-Male.bst*** and ***02‑ResidualPlots-Female.bst***.

**Preparation for equivalence testing**

**GenStat program: *03-PrepareSets.gen***

The GRACE data are used as historical background data for G-TwYST. In order to experiment with the new equivalence testing method using GRACE as historical data **and** current data, five sets of observations are defined as follows were the “current” data always compare the 33% GMO treatment group with the Control group and the “historical” data are the remaining Control and Conventional feeds in studies A, B and C. Note that in the table below the value of the group factor is used.

|  |  |  |  |
| --- | --- | --- | --- |
| Set | Current Study | New data | Historical data |
| 1 | A | 11 13 |  14 15 21 24 25 31 35 |
| 2 | B | 21 23 |  11 14 15 24 25 31 35 |
| 3 | C | 31 33 |  11 14 15 21 24 25 35 |
| 4 | D | 41 43 |  11 14 15 21 24 25 31 35 |
| 5 | E | 51 53 |  11 14 15 21 24 25 31 35 |

For the current data the following summary statistics were saved in current pointer

1. $Name$ the name of the variable
2. $d$ the observed mean difference $Δ$
3. $n\_{T}$ the replication of the treatment group
4. $n\_{C}$ the replication of the control group
5. $SS\_{F}$ the sums of squares for residual
6. $DF\_{F}$ the degrees of freedom for the residual sums of squares

For the historical data the following summary statistics were saved (all corrected for study effects)

1. $Name$ the name of the variable
2. $SS\_{R}$ the between reference groups sums of squares
3. $DF\_{R}$ the degrees of freedom of the between reference groups sums of squares
4. $SS\_{E}$ the sums of squares for residual
5. $DF\_{E}$ the degrees of freedom for the residual sums of squares
6. $N\_{eff}$ the effective replication which is necessary to estimate the between groups variance

This was done separately for Males and Females, and also with and without outliers. Results are stored in backingstore file ***03-PrepareSets.bst*** in subfiles *IncludeOutliersM*, *ExcludeOutliersM*, *IncludeOutliersF* and *ExcludeOutliersF*. Note that each subfile contains a current and historic pointer each containing 5 pointers, one for each set, with the structures given above. For example current[4][] and historic[4][] contains the structures for testing Grace D as current study versus Grace ABC as historic study. Results are also stored in Excel file ***03-PrepareSets.xlsx*** with a separate sheet for each set.

**Genstat program : *03-PrepareSet-ABCDE.gen***

Historical data were also prepared for the full set of Grace trials; these historical data will be used in the analysis of the G-TwYST data. An historic pointer with the six structures as given above is saved in backingstore file ***03-PrepareSet-ABCDE.bst*** and Excel file ***03-PrepareSet-ABCDE.xlsx*** which again contain the four subfiles *IncludeOutliersM*, *ExcludeOutliersM*, *IncludeOutliersF* and *ExcludeOutliersF*.

**Equivalence testing for set 4 (D vs ABC)**

**Genstat program : *04-doDWE0.gen***

This program uses procedure DWE0 to calculate $θ\_{0}$ for set 4 (i.e. D vs ABC) for $n\_{0}$ = 5,8 and for each subfile. The output pointers *results5*, for $n\_{0}$=5, and *results8*, for $n\_{0}$=8, are stored in separate subfiles in backingstore file ***04-doDWE0.bst***. The program uses the following settings $α$=0.05, $power$=0.95, $ndatasets$=10.000 and $ngpq$=10.000. This program takes rather long (4:45 hours on my laptop) due to the large numbers of simulated datasets.

**Genstat program : *05-doDWE.gen***

This program uses the results from the previous program to perform the DWE analysis. The resulting pointers *delta5*, *delta8*, *elsd5* and *elsd8* are stored in backingstore file ***05-doDWE.bst***, again in separate subfiles. Note that *delta5* and *delta8* are back-transformed, using the exp() function, to the ratio scale. These subfiles also contain the *results5* and *results8* pointers. This program only takes 18 seconds to run.

**Genstat program : *06-doInterval.gen***

This program uses results from the previous program to display the results in the usual graphs. File ***06‑doInterval.pdf*** contains 16 graphs: a plot on the RATIO scale and a plot on the ELSD for each combination of 4 subfiles (*IncludeOutliersM*, *ExcludeOutliersM*, *IncludeOutliersF* and *ExcludeOutliersF*) and two values of $n\_{0}$ (5 and 8).

**Genstat program : *07-doInterval.gen***

This program produces an ELSD plot for the two $n\_{0}$ values in the same graph: ***07-doInterval.pdf***.

**Genstat program : *08-doInterval.gen***

This program produces an ELSD plot for Include/Exclude outliers in the same graph: ***08-doInterval.pdf***.

**Appendix C: outliers detected by Grubbs’ test on residuals of a one-way ANOVA per trial**

Outliers found in the second cycle are given in italics (there is only one such outlier)

|  |
| --- |
| **Males: outliers** |
| Study | Variable | Feed | Cage | Value |
| A | HGB | Control | 135 | 2.464 |
| A | MCH | Control | 135 | 2.776 |
| A | MCHC | Control | 135 | 3.353 |
| A | ALT | Conv2 | 123 | 0.9449 |
| A | AST | Conv2 | 123 | 1.321 |
| B | growthR | Conv2 | 220 | -0.2835 |
| B | ALT | Conv2 | 217 | 0.04217 |
| B | Alb | Conv2 | 220 | 3.310 |
| B | P | Conv2 | 217 | 1.523 |
| B | Liver | Control | 240 | 1.075 |
| B | Pancreas | Control | 235 | -2.443 |
| C | MCH | GMO-11 | 317 | 3.090 |
| C | MCHC | GMO-11 | 317 | 3.702 |
| D | PLT | GMO-11 | 407 | 5.005 |
| E | Ca | Control | 528 | 1.094 |
| E | Testis | GMO-11 | 524 | 0.1140 |
| E | Epididymis | GMO-11 | 524 | -0.9447 |
| *B* | *Liver* | *Conv2* | *217* | *1.023* |
| **Females: outliers** |
| Study | Variable | Feed | Cage | Value |
| A | MCHC | Conv1 | 170 | 3.649 |
| A | PLT | Conv1 | 170 | 5.615 |
| A | AST | Conv1 | 168 | 0.9291 |
| A | Glu | Conv1 | 168 | 1.238 |
| A | K | Conv1 | 168 | 2.004 |
| A | Heart | GMO-33 | 148 | -1.512 |
| B | Weight\_4 | Conv2 | 257 | 5.130 |
| B | CHOL | GMO-11 | 255 | 1.115 |
| B | Lung | GMO-33 | 248 | -0.5240 |
| B | Uterus | Conv2 | 262 | -3.028 |
| C | MCH | Control | 374 | 1.634 |
| C | MCHC | Control | 374 | 2.223 |
| D | Weight\_3 | Control | 443 | 5.058 |
| D | PLT | Control | 443 | 5.190 |