G-TWYST
GM PLANTS TWO YEAR SAFETY TESTING

G-TwYST Study A

Combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603

Main statistical report

Paul W. Goedhart & Hilko van der Voet

G-TwYST (EU grant agreement no: 632165) Deliverable 4.3a

Biometris report 32.02.18 February 2018

https://doi.org/10.18174/455205

WAGENINGEN
UNIVERSITY & RESEARCH

G-TwYST Study A

Combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603

Main statistical report

Paul W. Goedhart & Hilko van der Voet

Biometris, one of the largest groups of quantitative scientists in North-Western Europe, develops statistical and mathematical methods for the quantification of biological processes and processes in our living environment. These methods are applied and validated in practice and are often available as software packages. In addition, we provide education at the undergraduate, Master, PhD, and PostDoc levels, as well as training and consultancy for industry. We cover a wide range of application areas, from gene to ecosystem and from product to production chain. Our goal is to bring quantitative methods to life.

Biometris is the full integration of the chair groups Applied Mathematics (Molenaar) and Applied Statistics (Van Eeuwijk) with the Wageningen Plant Research business unit Biometris (Wehrens).

For more information please visit the website www.biometris.nl or contact:

Biometris, Wageningen University & Research
P.O. Box 16
6700 AA Wageningen, The Netherlands

Visiting address:
Buildingnumber 107
Droevendaalsesteeg 1, 6708 PD Wageningen, The Netherlands

Phone: +31 317 480798 or +31 317 486001
E-mail: biometris@wur.nl

February 2018

Contents

Abstract

The purpose of oral toxicity study A in the EU project G-TwYST was to assess the effects of genetically modified (GM) maize NK 603, grown both with and without the use of RoundUp, when fed to rats for a period of up to two years at incorporation rates of 11% and 33% in the feed. The effects were assessed relative to the responses for male and female rats fed the near-isogenic non-GM maize, and the differences were compared, using classical and new statistical methods.

The experimental design was a complete randomized block design with cage as the experimental unit. Five types of feed were administrated to cages with 2 rats in each cage. Experimental units, i.e. cages, were organized in blocks of 5 cages, and the feeding groups were randomized within blocks. Treatment identity was blinded for everyone involved in the experiments and for the statistical analysts until after decisions were made about outliers in the data.

The four feeding groups with GM maize were each compared to the feeding group with non-GM maize for 36 variables at 3 months, 40 variables at 6 months, 44 variables at 12 months and 37 variables at 24 months, and for males and females separately. The variables were body weight, growth rate, feed intake, haematology, clinical chemistry and organ weights. Only data for animals that survived the experiment were statistically analysed. In total, 1256 comparisons on quantitative variables were made. In addition, the times of death were analysed using survival analysis, and an analysis of mortality at 24 months.

Using classical statistical tests at the 5% false positive level ($\alpha = 0.05$), among the 1256 comparisons of GM groups to the non-GM group on quantitative variables, 80 (6.4%) significant differences were found using the parametric t-test, and 72 (5.7%) significant differences using the non-parametric Wilcoxon test. Using the tests at the 1% false positive level ($\alpha = 0.01$), there were 13 (1.0%) significant differences using the t-test (lowest P-value 0.002), and also 13 (1.0%) significant differences using the Wilcoxon test (lowest P=value 0.005). The number of significant test results is therefore equal or slightly higher than the expected false positive rate of the tests, and it was relatively high in the 12 months data (11% for the t-tests, and 9% for the Wilcoxon tests). The patterns of significant results are open for toxicological interpretation.

Statistical tests were also performed based on the factorial structure of the feeding groups using analysis of variance. Using statistical tests at the 5% false positive level ($\alpha = 0.05$), 13 (4%) significant effects were found among 314 comparisons of the average of the four GM groups with the non-GM group, which therefore does not exceed the expected number of false positives given the nominal test level. More significant effects were found within the set of four GM groups. In 32 (10%) of the 314 cases there was a significant interaction effect, i.e. the differences between the GM groups with and without RoundUp were dependent on the GM inclusion rate 11% or 33%. There were 22 (8%) significant RoundUp effects in 282 comparisons, i.e. differences of the two GM groups with RoundUp and the two GM groups without RoundUp. These significant RoundUp effects were relatively more frequent for the data after 3 months (6 out of 67, i.e. 11%) and after 12 months (11 out of 77, i.e. 14%). The number of significant effects found for the GM inclusion rate (11 or 33%) was (6%) 22 in 282 tests, close to the nominal test level. Overall, combined over all factorial tests the rate of significant results was 7% (85 out of 1192). Again, the patterns of significant results are open for toxicological interpretation.

The same type of comparisons (GM groups to non-GM group, and factorial effects) were made for hazard ratios from survival analyses and for mortality rates at 24 months. Among 32 one-sided tests for hazard ratios and 16 one-sided tests for mortality (male or female, feeding groups or factorial structure, animal or cage level), three tests (6%) resulted in a P-value (just) below 0.05, which is close to the theoretically expected rate of false positives (5%). In a factorial analysis on cages with male rats the hazard ratio for the two RoundUp groups was on average higher by a factor 1.73 than for the two GM groups without Roundup (P=0.04). With respect to mortality, a significant difference (P=0.03) was found in males between the Control group (36% mortality) and the GM group with 33% inclusion rate and RoundUp (54% mortality). This also showed up as a significantly higher average mortality in the two RoundUp groups compared to the two GM feeding groups without RoundUp (45% vs. 31%, P=0.03). For females, no significant higher hazards or mortality rates were found, although in fact the GM feeding groups had estimated hazards and mortalities that were equal or significantly lower than for the non-GM feeding group.

Equivalence testing is in principle more adapted to the aim of safety testing than the above tests which are directed to find differences. The G-TwYST project has developed an equivalence testing method that compares the current data to historical reference data. For Study A in G-TwYST, the non-GM data from five 90 days studies in an earlier EU project GRACE were available as reference data. It was found that these data could be used as a reference basis for the 3 and 6 months data of the current study, but this was not the case for the 12 and 24 months data, due to a general (and expected) pattern of increasing between-animal variation over time. Among all 456 equivalence tests for the 3 and 6 months data, equivalence was shown in 433 cases (95%), which corresponds to the set power level (95%) of the test. Nevertheless, in all but one cases equivalence was more likely than non-equivalence. Most failures to show equivalence were due to a lower precision in the current study as compared to the historical data. Restricting the application of the equivalence tests to cases where the residual variance was not more than 150% of the reference residual variance, equivalence was established in 411 of 412 cases (99.8%). This type of equivalence testing would need its own collection of reference data for use with 12 months or 24 months data.

An alternative and easier form of equivalence tests is comparison to externally given target effect sizes. For nine variables such values have been suggested in a recent paper. Against these values equivalence could be established in 211 (94%) of 224 cases, with all failures occurring for the 24 months data, where the variation in the data was much larger. Nevertheless, in all cases equivalence was more likely than non-equivalence.

The results presented in this statistical report are still to be combined with the results from histopathology, and are to be interpreted with respect to their toxicological implications. It is stressed that the statistical analysis has only looked at differences and equivalences with respect to external data or values. It has not used any data regarding benchmark doses that are aimed to distinguish between safe and unsafe values.

1 Introduction

The purpose of oral toxicity study A in the EU project G-TwYST was to assess the effects of genetically modified (GM) maize NK 603, grown both with and without the use of RoundUp, when fed to rats for a period of one year (chronic toxicity study) and two years (carcinogenicity study), at incorporation rates of 11% and 33% in the feed. The effects of the GM maize were assessed relative to the responses for rats fed the near-isogenic non-GM maize. The chronic toxicity and carcinogenicity study ran in parallel. At 12 months part of the rats were euthanized and their organs were obtained, weighted and examined by the G-TwYST histopathological expert. The remaining rats were kept until 24 months, or until premature death, and were then assessed in the same way.

Rat weights and feed intake were measured weekly or bi-weekly. Haematology, clinical biochemistry in blood and urine, differential white blood cell counts, and urine volume and colour were obtained for subsets of animals after 3, 6, 12 and 24 months. Organ weights were observed for the animals in the chronic toxicity 12 month study and for the animals that did not survived the 24 month carcinogenicity study.

This report describes the main results of the statistical analysis of the data from Study A. Detailed results for the data obtained after 3, 6, 12 and 24 months can be found in four separate companion reports (Goedhart % van der Voet, 2018abcd). This main report describes the statistical methods, and summarises the main results. It also provides a discussion of the methodology.

In principle, the statistical analysis was performed according to section 8 of the study plan (Zeljenková and Steinberg 2015). Deviations from the study plan were as follows:

- The study plan specified a statistical analysis of data for males and females together, unless there were prior biological arguments or statistical indications to analyse males and females separately. However, toxicologists preferred separate statistical analysis of males and females for all variables, because it was thought that any specific non-target effect might be sex-specific.
- The study plan anticipated pre-specified limits for use in equivalence testing. However, such limits could not be established in an early phase of the project. Therefore, an alternative method for equivalence testing was developed (van der Voet *et al*. 2017) and applied. This method uses historical non-GM data to obtain reference values for acceptable and normal variation in the observed variables. For the analysis of the data in G-TwYST, the data from non-GM varieties in the preceding GRACE project, which consisted of 3 months studies, were available as historical data. Target effect sizes for a few variables were recently proposed by Hong *et al*. (2017). Although these values have no formal status, equivalence tests were also performed using these effect sizes as originally planned.

Death events were analysed in two ways:

1. Kaplan-Meier estimates of the survival curves were plotted and hazard ratios were analysed using the proportional odds model.
2. Mortality rates after 24 months were analysed using logistic regression.

The remaining results, comparing each GM dose group to the non-GM control group for every variable, are presented according to four schemes of statistical analysis:

1. Equivalence tests, following the method developed in the G-TwYST project (van der Voet *et al*., 2017).
2. For a small number of variables: equivalence tests, based on target effect sizes suggested in Hong *et al*. (2017).
3. Classical tests, in line with OECD Guidance document 116 (2012).
4. Standardised effect sizes, following the methods used in the GRACE project (Schmidt *et al*., 2016, 2017).

In addition, the following results from more integrated analyses were obtained:

5. Factorial analysis, integrating over the five dose group, with main factors GmInclusionRate and RoundUp, and the interaction between these two factors.
6. Correlation analysis, showing co-variation of effects for variables related to the same target organ.

This report is organised as follows. Section 2 describes the data, both the data from the current study and the non-GM data from the GRACE studies as used in the equivalence tests. Section 3 describes data pre-processing procedures, such as summarising the growth and food intakes over time, outlier identification and assumptions checking. Section 4 presents the main results referring to the separate reports for details. Section 5 gives a summary and some evaluation of the methodology.

2 Data

2.1 Data in G-TwYST study A

Study A is a combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603. A full description of the data that have been measured is given in the study plan (document 632165 A/2015/GLP, Zeljenková and Steinberg, 2015). There are five feeding groups which are administrated to cages with 2 rats in each cage. Experimental units, i.e. cages, are organized in blocks of 5 cages, and the feeding groups are randomized over cages within blocks. The design is thus a complete randomized block design with cage as the experimental units. There are ten blocks with male rats, and ten other blocks with female rats, for the one-year chronic toxicity study. In addition, there are 25 blocks with male rats, and 25 other blocks with female rats, for the tow-year carcinogenicity study. For the first year of the study the total of blocks is thus 35. Most of the measurements are on individual animals, only feed intake is measured on the cage level. The number of planned observations per feeding group and sex is given in Table 1. Due to death, or missing data, the actual number of observations were sometimes slightly smaller.

The definition of the five feeding groups is given in Table 2. This includes the coding of the blinded feeds and treatments. In this report the Group labels (Control, NK11-, NK33-, NK11+ and NK33+) will mostly be used.

Male and female rats were analysed separately. Since cage is the experimental unit an analysis of variance employs cage means with degrees of freedom as in Table 3. The main interest is in the difference between each of the four GM maize feeding groups and the non-GM control feeding group.

The observed variables in Study A are given in Table 4. The variables in Table 4 exclude the following measurements in comparison to the study plan:

- Histopathological data. Reason: these were excluded from this statistical analysis in the study plan, and will be separately reported by the histopathological expert in the G-TwYST project.
- Periodic health status observations: morbidity, mortality, clinical signs. Rats were inspected twice daily for evidence of reaction to treatment or ill-health. Health status observations are given in Appendix 1.
- Organ weights: thyroid, parathyroid. These were not determined.

Table 1 Number of planned observations per endpoint category per feeding group and sex at 3, 6, 12 and 24 months in study A. The number of observations at 24 months is subdivided in a group for animals that died prematurely and animals that survived for 24 months.

Month	Number of planned observations per feeding group per sex					
	Weight/Feed	Haematology	ClinChem	diffWBC	Urine	Organ Weights
3	70	40	40	-	20	-
6	70	40	40	40	20	-
12	70	40	40	-	30	20
24 premature	all	partly	partly	-	-	all
24 survived	all	all	all	-	partly	-

Table 2 Feeding groups and coding of feeds in study A.

Factor	Levels / Labels				
Feed coding	Z-1005	Z-1002	Z-1004	Z-1003	Z-1001
Treatment coding	E	B	D	C	A
Group	Control	NK11-	NK33-	NK11+	NK33+
Maize	Control	NK603	NK603	NK603	NK603
AmountNK	0	11	33	11	33
RoundUp	No	No	No	Yes	Yes

Table 3 Skeleton analysis of variance with degrees of freedom for cage means for a single sex.

Source of variation	35 cages/group	20 cages/group	10 cages/group
	d.f.	d.f.	d.f.
Block stratum	34	19	9
Block.Cage stratum			
Group	4	4	4
Residual	137	76	36
Total	174	99	49

Table 4 List of grouped variables with abbreviated names, descriptions and measurement units. Grouping is indicated by the headers in the first column. The Grace column indicates whether the same variable was measured in the GRACE project which consisted of 3 months studies. Variables given in red are not statistically analysed (see companion reports for details). The months 24p for Organ weights denote the animals that died prematurely, i.e. before the end of the trial.

Weights	Description	Unit	Months	Grace
Weight	Body weight at weeks 0, 1 … 13	g/animal	1-3	X
	Body weight at weeks 15,17… 27	g/animal	4-6	-
	Body weight at weeks 29, 31 … 52, 53	g/animal	7-12	-
	Body weight at weeks 55, 57 … 103, 104	g/animal	13-24	-
Feed	Feed intake in weeks 1, 2 … 13	g/cage	1-3	X
	Feed intake in weeks 15,17… 27	g/cage	4-6	-
	Feed intake in weeks 29, 31 … 52, 53	g/cage	7-12	-
	Feed intake in weeks 55, 57 … 103, 104	g/cage	13-24	-
BodyWeight	Body weight at week 13	g/animal	3	X
	Body weight at week 27	g/animal	6	-
	Body weight at week 52	g/animal	12	-
	Body weight at week 104	g/animal	24	-
growthRate	Growth rate fitted to the weights over weeks 0-13	1/week	3	X
	Growth rate fitted to the weights over weeks 14-27	1/week	6	-
	Growth rate fitted to the weights over weeks 27-52	1/week	12	-
FeedMean	Mean of feed intake over period	g/an./day	3,6,12,24	X
Haematology	Description	Unit	Months	Grace
WBC	white blood cells	10^9/L	3,6,12,24	X
RBC	red blood cells	10^{12}/L	3,6,12,24	X
HGB	haemoglobin	g/dL	3,6,12,24	X
HCT	haematocrit	%	3,6,12,24	X
MCV	mean cell volume	fL	3,6,12,24	X
MCH	mean corpuscular haemoglobin	pg	3,6,12,24	X
MCHC	mean corpuscular haemoglobin concentration	g/dL	3,6,12,24	X
PLT	platelets	10^9/L	3,6,12,24	X
LYMR	relative lymphocytes count	%	3,6,12,24	-
LYMA	absolute count of lymphocytes	10^3/uL	3,6,12,24	X
diffWBC	Description	Unit	Months	Grace
Lymphocytes	Percentage of lymphocyte cells in 200 cells	%	6	X
Neutrophils	Percentage of neutrophil cells in 200 cells	%	6	X
Monocytes	Percentage of monocyte cells in 200 cells	%	6	X
Eosinophils	Percentage of eosinophil cells in 200 cells	%	6	X
ClinChem	Description	Unit		Grace
ALP	alkaline phosphatase	µkat/L	3,6,12,24	X
ALT	alanine aminotransferase	µkat/L	3,6,12,24	X
AST	aspartate aminotransferase	µkat/L	3,6,12,24	X
BIL	bilirubin	µmol/L	3,6,12,24	-
ALB	albumin	g/L	3,6,12,24	X
TP	total protein	g/L	3,6,12,24	X
Glu	glucose	mmol/L	3,6,12,24	X
CHOL	cholesterol	mmol/L	3,6,12,24	X

TAG	triglycerides	mmol/L	3,6,12,24	X
Crea	creatinine	mmol/L	3,6,12,24	X
Urea	urea	mmol/L	3,6,12,24	X
cHGB	haemoglobin	mg/dL	3,6,12,24	-
Ca	calcium	mmol/L	3,6,12,24	X
Cl	chloride	mmol/L	3,6,12,24	X
K	potassium	mmol/L	3,6,12,24	X
Na	sodium	mmol/L	3,6,12,24	X
P	phosphorus	mmol/L	3,6,12,24	X
Urine	Description	Unit	Months	Grace
uVol	Urine Volume	ml	3,6,12,24	-
uVolW	Urine Volume / bodyweight	ml /100g	3,6,12,24	-
uColour	Urine Colour (1-light yellow; 2-yellow; 3-dark yellow)	-	3,6,12,24	-
uBil	bilirubin	µmol/L	3,6,12,24	-
uLeu	leukocytes	leu/uL	3,6,12,24	-
uNit	nitrites	neg/pos	3,6,12,24	-
uOsmoll	osmolality	mOsm	3,6,12,24	-
uProtein	total protein	g/L	3,6,12,24	-
uGlu	glucose	mmol/L	3,6,12,24	-
uHemogl	haemoglobin	ery/uL	3,6,12,24	-
uKeton	ketone	mmol/L	3,6,12,24	-
upH	pH	-	3,6,12,24	-
uUrobili	urobilinogen	µmol/L	3,6,12,24	-
Organs	Description; all as percentage of BodyWeight	Unit	Months	Grace
Kidney	Percentage weight of kidney	%	12, 24p	X
Spleen	Percentage weight of spleen	%	12, 24p	X
Liver	Percentage weight of liver	%	12, 24p	X
AdrenGl	Percentage weight of adrenal gland	%	12, 24p	X
Heart	Percentage weight of heart	%	12, 24p	X
Testis	Percentage weight of testis (males)	%	12, 24p	X
Epididymis	Percentage weight of epididymis (males)	%	12, 24p	X
Uterus	Percentage weight of uterus (females)	%	12, 24p	X
Ovary	Percentage weight of ovary (females)	%	12, 24p	X
Brain	Percentage weight of brain	%	12, 24p	X

All variables that are also observed in the GRACE study are statistically analysed. Some of the other variables are not statistically analysed e.g. because all the observed values are (almost) identical; see the companion reports for details. All variables except times of death were transformed to the natural logarithmic scale and then averaged to the cage level. This implies that, rather than looking at differences between feeding group means, ratios between the GM feeds and the Control feed are of interest. Only pH as measured in urine was not log transformed because the pH is already measured on the log scale.

2.2 Reference data in GRACE studies

Data from the GRACE project were used as historic data to set equivalence limits. These data have been analysed before as part of the GRACE project (Schmidt and Schmidtke 2014, Schmidt *et al*. 2015ab, 2016, 2017, Zeljenková *et al*. 2014, 2016). Note that in GRACE a completely randomized

design, i.e. without blocking, was used. The GRACE data were retrieved from the Cadima website (https://www.cadima.info) at 29-11-2016. In GRACE four 90-day studies and one 1-year study, were conducted with several control (or reference) feeds as given in Table 5, see Schmidt et al (2017).

Table 5 Feeds which were used in the five GRACE studies with reference feeds in bold.

GRACE 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	-	-

In studies D and E only a single reference feed was used and, since the equivalence analysis corrects for differences between studies, these studies do not contribute to the between reference variation. In studies A, B and C, the reference feeds DKC6666 and SY-NEPAL were replicated. The degrees of freedom associated with the between reference feeds variance therefore equals 4. The degrees of freedom associated with the residual (between cages) variance varies between 50 and 78 since not all measurements were done on all rats in every study.

We re-analysed the 90-days (3-months) GRACE data from all five studies to enable a comparison with the G-TwYST data. This re-analysis is different from the analysis in the GRACE reports in the following ways:

- For the re-analysis all variables were transformed to the natural logarithmic scale and then averaged to the cage level; the thus obtained cage means were used in the statistical analysis;
- the exponential growth model (see section 3.1) was fitted to the weights observed in GRACE to obtain an estimate of the growth rate γ;
- The sum of the weights of organ pairs was analysed rather than the left and right organs;
- outliers were identified by applying Grubbs' outlier test at the 1% level on residuals of a one-way ANOVA which is conducted separately for each study. These outliers were set to missing.

Details of the re-analysis are given in the report on Study B (Goedhart % van der Voet, 2017). In Table 4 it is indicated which G-TwYST variables have also been measured in the GRACE studies.

3 Data pre-processing

The initial data pre-processing is described in the companion reports (Goedhart & van der Voet 2018abcd).

3.1 Growth curves and feed intake

For each individual rat growth curves were fitted to the observed weights for restricted periods of time. For the data up to 13 weeks (3 months) an exponential growth curve $A + B\ R^{Week}$ was fitted. A re-parameterization of this curve is given by $A + B\ \exp(-\gamma\ Week)$ with the growth rate γ defined by $\gamma = -\log(R)$. For the data between weeks 13 and 27 (6 months), and separately for the data between weeks 27 and 52 (12 months), a simple linear regression, $Weight = \alpha + \beta\ Week$, was

fitted to the observed weights, and the growthRate was defined as $\gamma = \log(\beta)$. In general the growth curves fit very well and it was therefore decided to only analyse the final weight observed after week 13, 27 and 52, further called Weight_13, Weight_27, Weight_52, and the estimated growth rates γ, further called growthRate. No general growth curves could be fitted for the data between weeks 53 and 104 (24 months), and therefore only the final weights for those animals that survived for 24 months, further called Weight_104, were statistically analysed.

The mean weight for each feeding group is given in Figure 1, while the mean weight gain per day per animal in each week is given in Figure 2. The mean feed consumption for each feeding group is given in Figure 3.

3.2 Outliers and checking of ANOVA assumptions

The process for finding potential outliers is described in detail in each of the specific reports. Outlier detection was done before the feeding group codes were unblinded. Essentially, Grubbs' outlier test at the 1% level was sequentially applied to the residuals after an analysis of variance on cage means for the 3/6/12 month data, and on individual animal data for those animals that survived after 24 months. This resulted in a number of outliers which were first presented to the study director and then to the G-TwYST coordinator. Outliers were classified as either (1) typos or physiologically improbable values or (2) values that might be realistic. For the first category the values were set to missing, effectively removing the outlier completely. For the second category a statistical analysis without and with these outlier was performed.

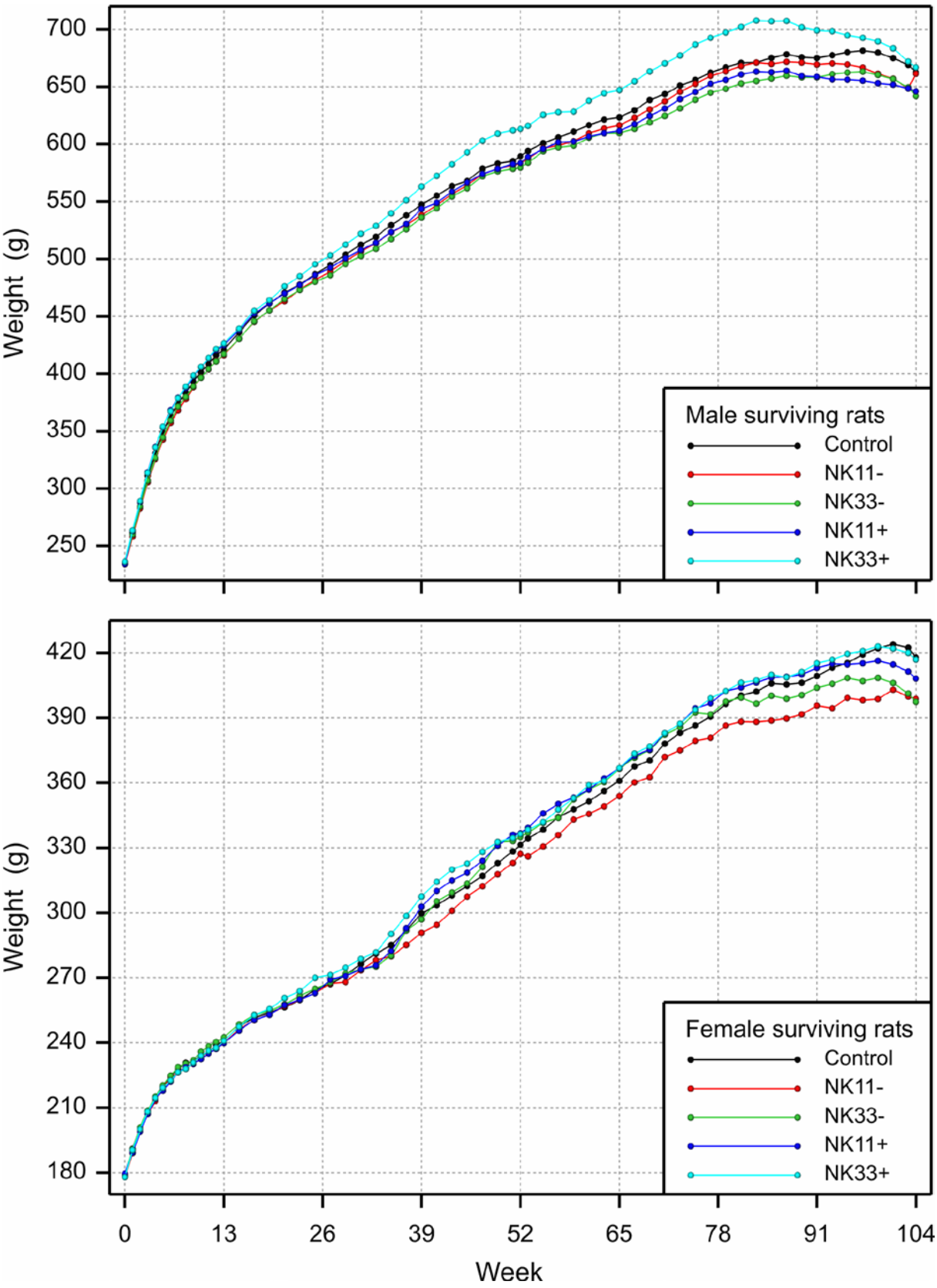

Figure 1 **Mean body weights (g) versus week for each feeding group for male rats (top) and for female rats (bottom).**

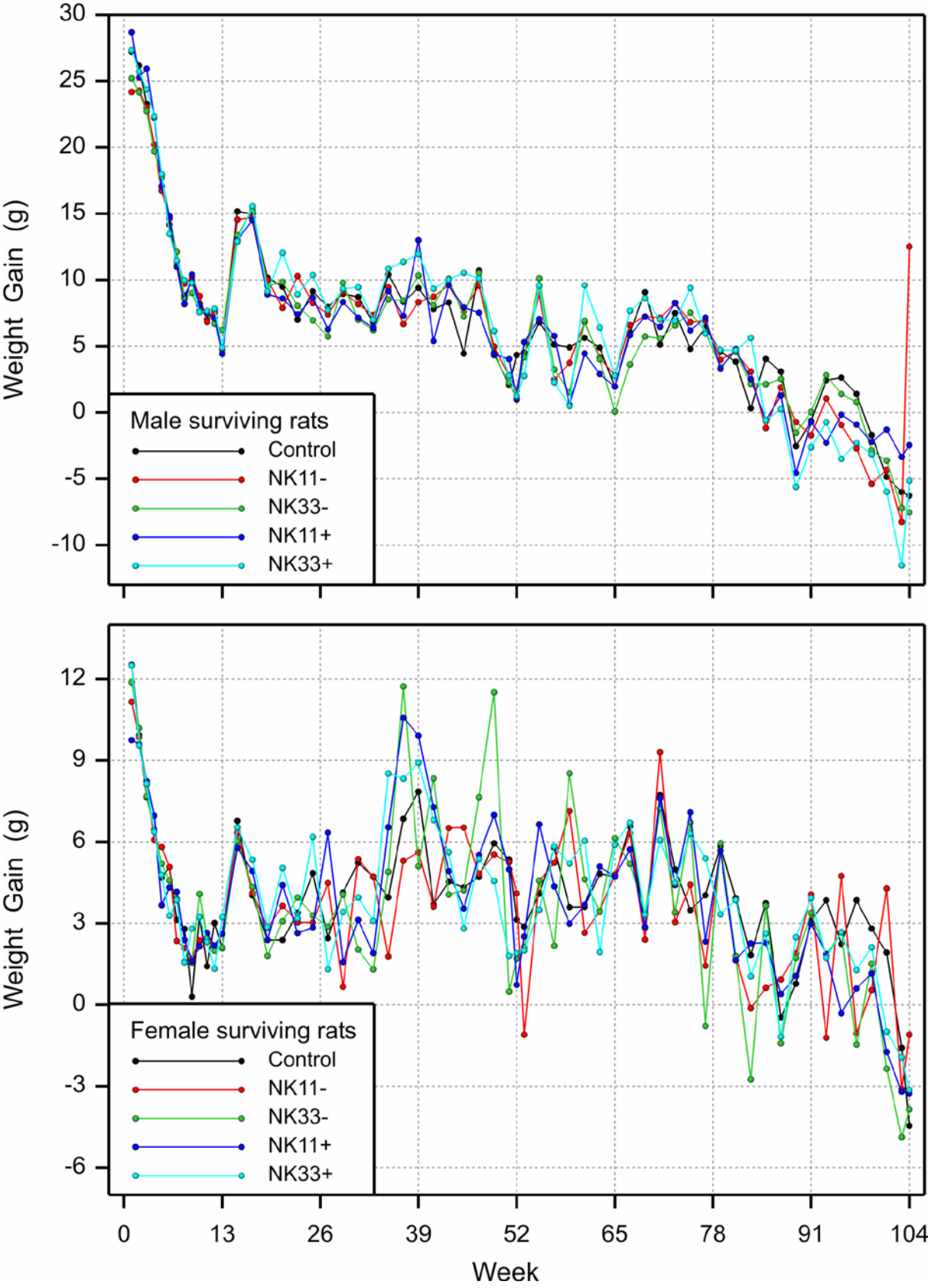

Figure 2 **Mean body weights gain (g/day/animal) versus week for each feeding group for male rats (top) and for female rats (bottom).**

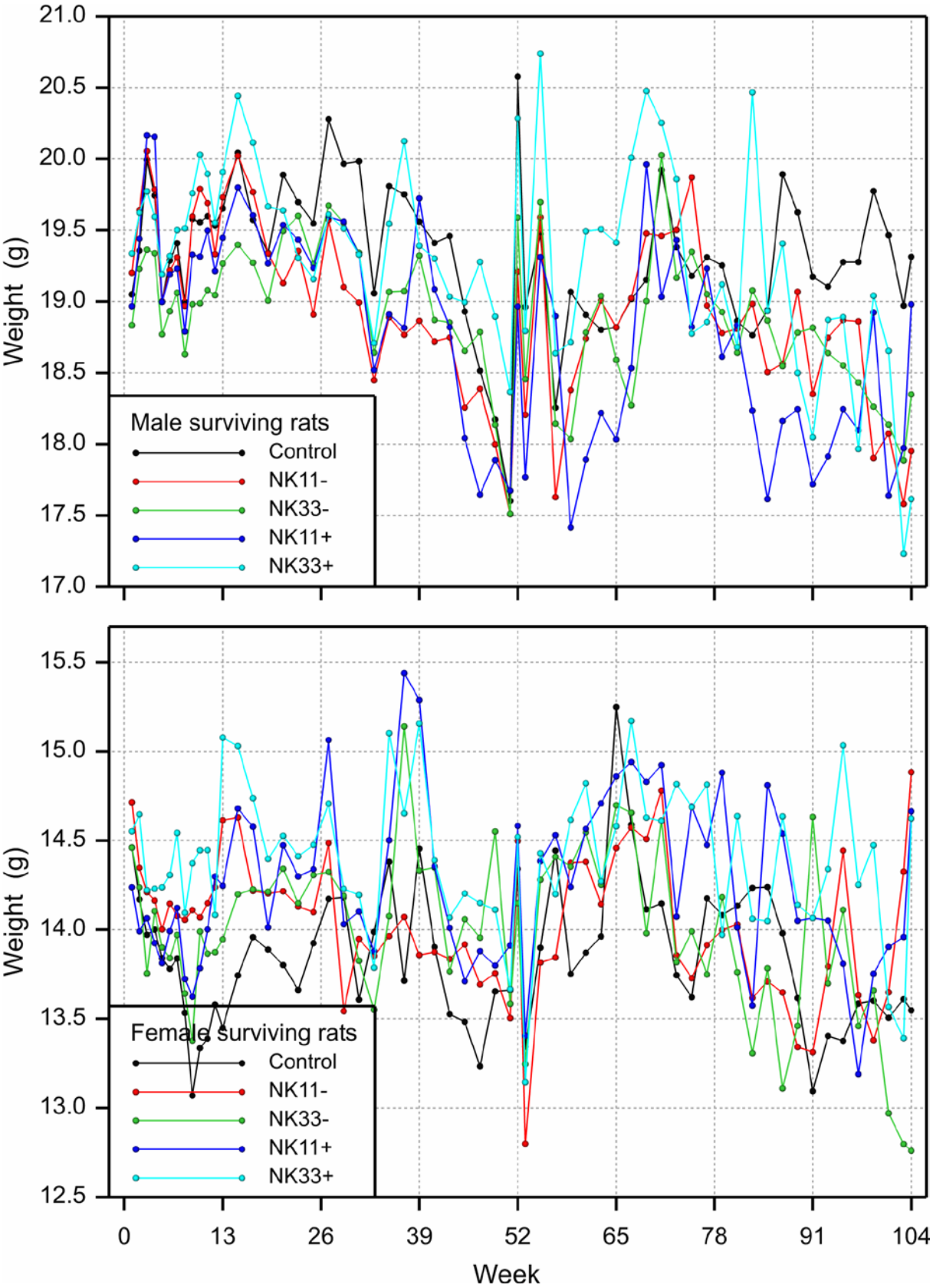

Figure 3 **Mean feed consumption (g/day/animal) versus week for each feeding group for male rats (top) and for female rats (bottom).**

4 Statistical methods

4.1 Analysis of survival curves

4.1.1 Method

Survival analysis was performed at the individual level, which disregards the fact that two animals were housed in the same cage, and at the cage level with 'survival of a cage' defined as survival of both animals in the cage. Kaplan-Meier estimates of the survival curves were plotted for all animals/cages of a single sex for the five dose groups. The differences between the five dose groups were assessed using four overall tests for equality of survival curves (log-rank, Wilcoxon-Breslow, Tarone-Ware and Wilcoxon-Peto-Prentice), as defined in Collett (1994).

Pairwise differences between the control non-GM group and each of the four test dose groups were assessed by means of the proportional odds model (Cox, 1972), with a single survival function and with blocks and dose groups as treatment factors. Based on the latter model pairwise differences of the four test (NK603) groups w.r.t. the control group were assessed using a normal approximation of the logarithms of the hazard ratio. In more detail, if b and s represent the estimate of the hazard logratio and its standard deviation, then $b_L = b - z_{1-\alpha}s$ and $b_U = b + z_{1-\alpha}s$ are one-sided lower and upper $1-\alpha$ confidence limits for the true hazard logratio. Note that equal hazards correspond to a value 0 in the logratio scale. Therefore we find a significant increase in hazard if $b_L > 0$. A one-sided P-value can be calculated based on the normal distribution.

The proportional hazards model was also fitted using a factorial representation of the dose groups. The structure of the G-TwYST study is a 2 by 2 factorial design for the GM feeding groups with factors GM inclusion rate (*IR*, 11% or 33%) and use of Roundup (*RU*, - or +), with an added control for the non-GM control group. This structure allows a more sensitive analysis, integrating over the five dose group, according to the model for the hazard logratio:

$$PH_{ij} = \mu + GM + IR_i + RU_j + int_{ij} \qquad \text{for GM groups, } i = 1,2; j = 1,2.$$

In this model the fixed term GM models the difference between the four GM groups (averaged) and the control. This term is only interpretable if the other three fixed terms can be assumed to be zero. The main effect $IR_2 - IR_1$ models the difference between the groups with GM inclusion rates 33% and 11%, and similarly, the main effect $RU_2 - RU_1$ models the difference between the groups with and without Roundup. These main effects are only useful when there is no interaction inj_{ij} between GM inclusion rate and Roundup.

These survival analyses were performed using procedures KAPLANMEIER, RSTEST, RPROPORTIONAL and RPHFIT, and direct programming in GenStat 18 (VSN International, 2015).

4.1.2 Results

This section summarises the main findings reported in the companion report (Goedhart & van der Voet 2018d).

Kaplan-Meier estimates of the survival curves per sex and per feeding group based on individual rats, and based on cages, are given in Figure 4. The event for a cage was defined as the day at which the first animal died. This definition respects the fact that cage is the experimental unit.

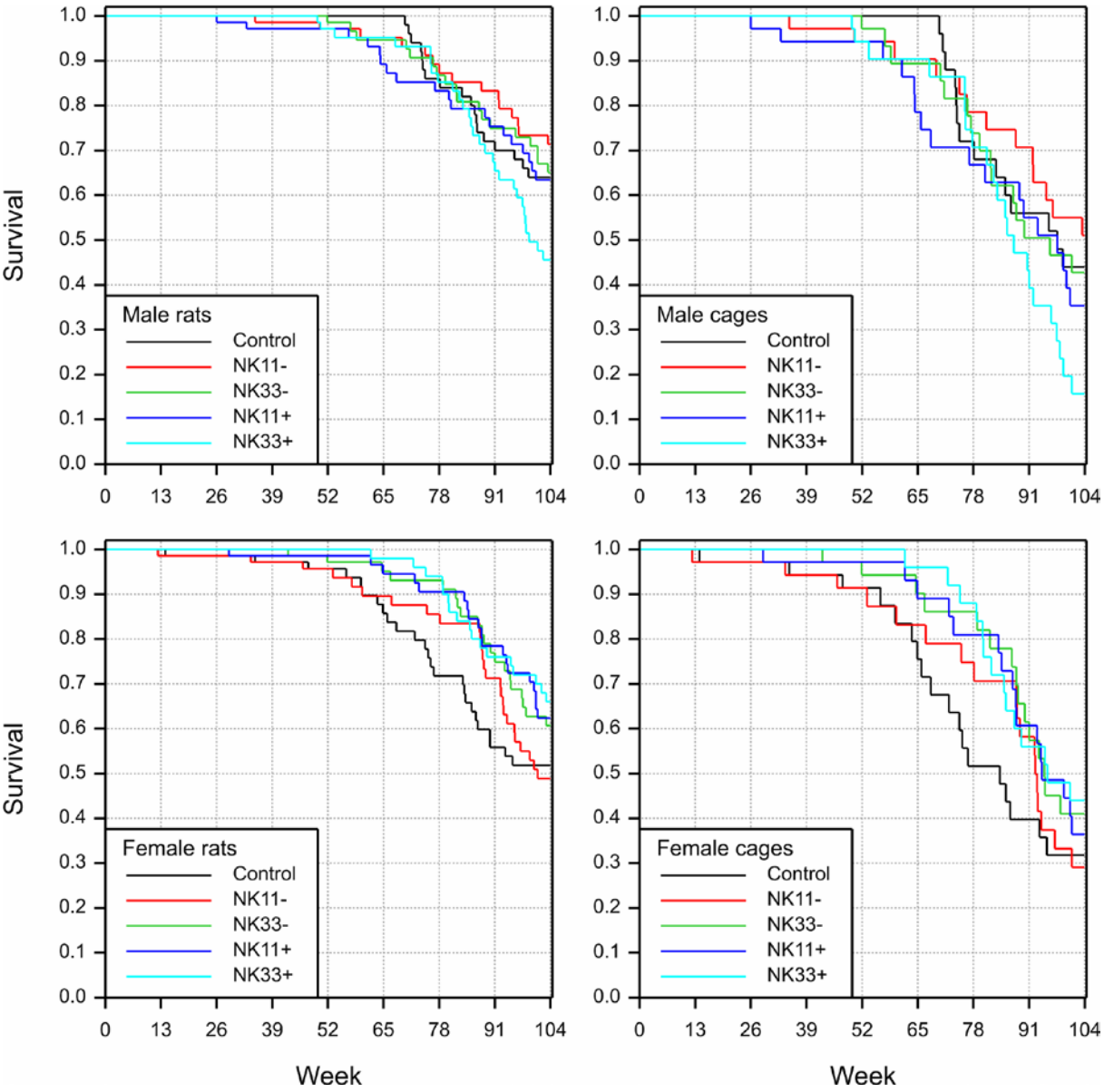

Figure 4 Kaplan-Meier estimates of the survival curves for all 350 rats per sex (left figures), and for cages (right figures) where the event per cage is the day of death of the first animal.

Four overall tests, testing the overall null hypothesis that the survival curves are identical, were performed. According to these tests, which do not account for differences between blocks, there were no overall significant differences between the survival curves at the 5% level.

On a more detailed level, employing the proportional odds model, the one-sided null hypothesis was tested that the hazard ratio of each of the four GM feeding groups relative to the control non-GM group was equal to or smaller than one. The hazard ratios relative to the control non-GM group are graphically depicted in Figure 5. The null hypothesis was never rejected.

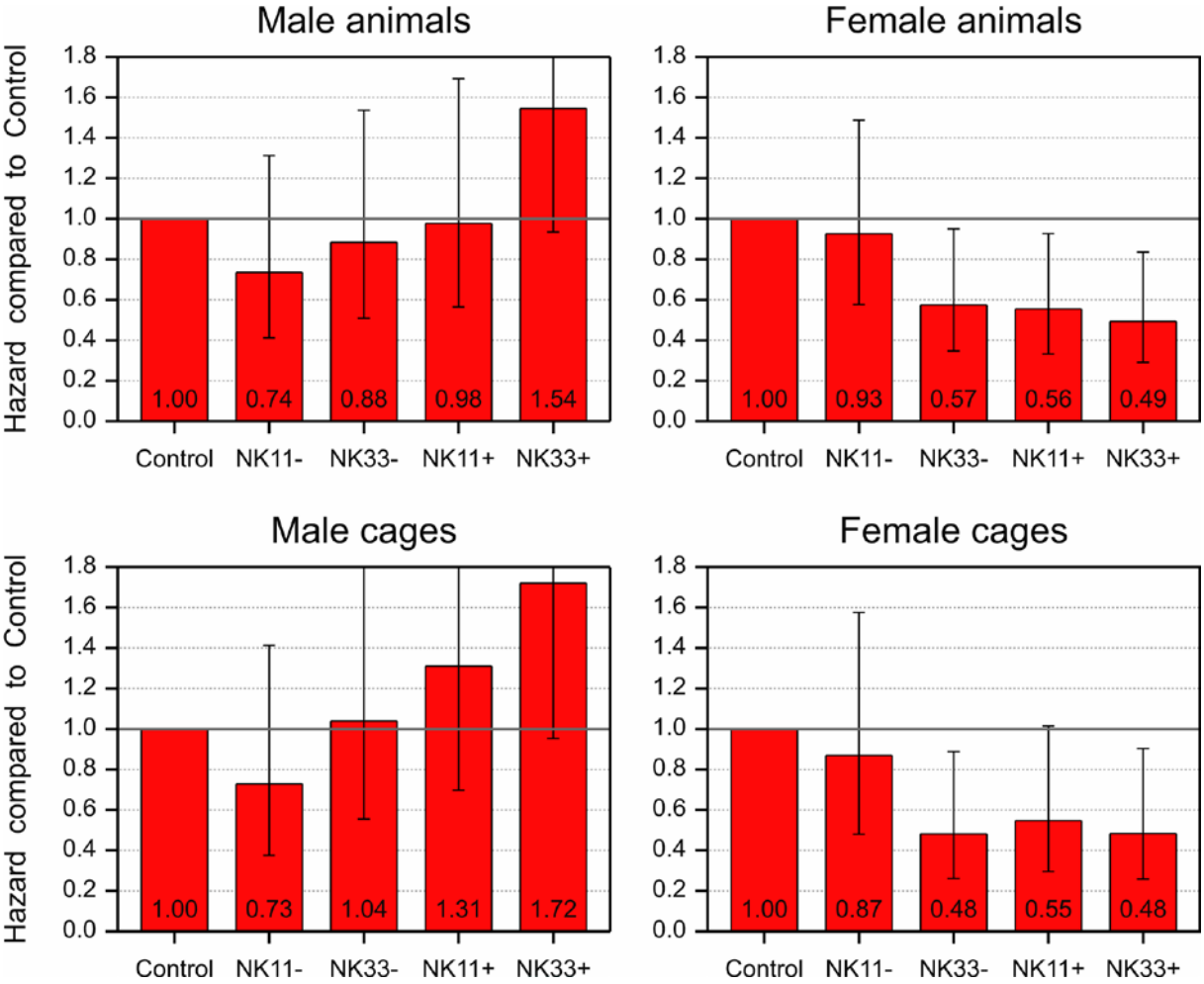

Figure 5 Hazard ratios according to the proportional hazards model relative to the hazard of the control non-GM feeding group. 90% confidence intervals for hazard ratios are indicated by error bars; when the lower limit is larger than 1 the one-sided null hypothesis is rejected. The hazard ratio is also given at the bottom of each bar.

In addition, a factorial analysis has been performed employing chi-squared likelihood ratio testing in the proportional odds model. Factorial effects are graphically depicted in Figure 6. In 16 tests, there was one P-value (just) below 0.05. There is an indication that there might be an effect of Roundup in males (P-values equal to 0.053 and 0.037 in the analysis at the animal and cage level, respectively). The ratio of the hazard for Roundup GM feeds versus non-Roundup GM feeds is estimated by 1.52 for male animals and 1.73 for male cages. This indicates that the non-Roundup GM feeds have a better survival rate. There is also an indication that in females the GM feeds could have a better survival rate than the non-GM control feed (P-values equal to 0.053 and 0.065 in the analysis at the animal and cage level, respectively). The estimated hazard ratio for the GM feeds versus the control feed equals 0.62 for female animals and 0.58 for female cages.

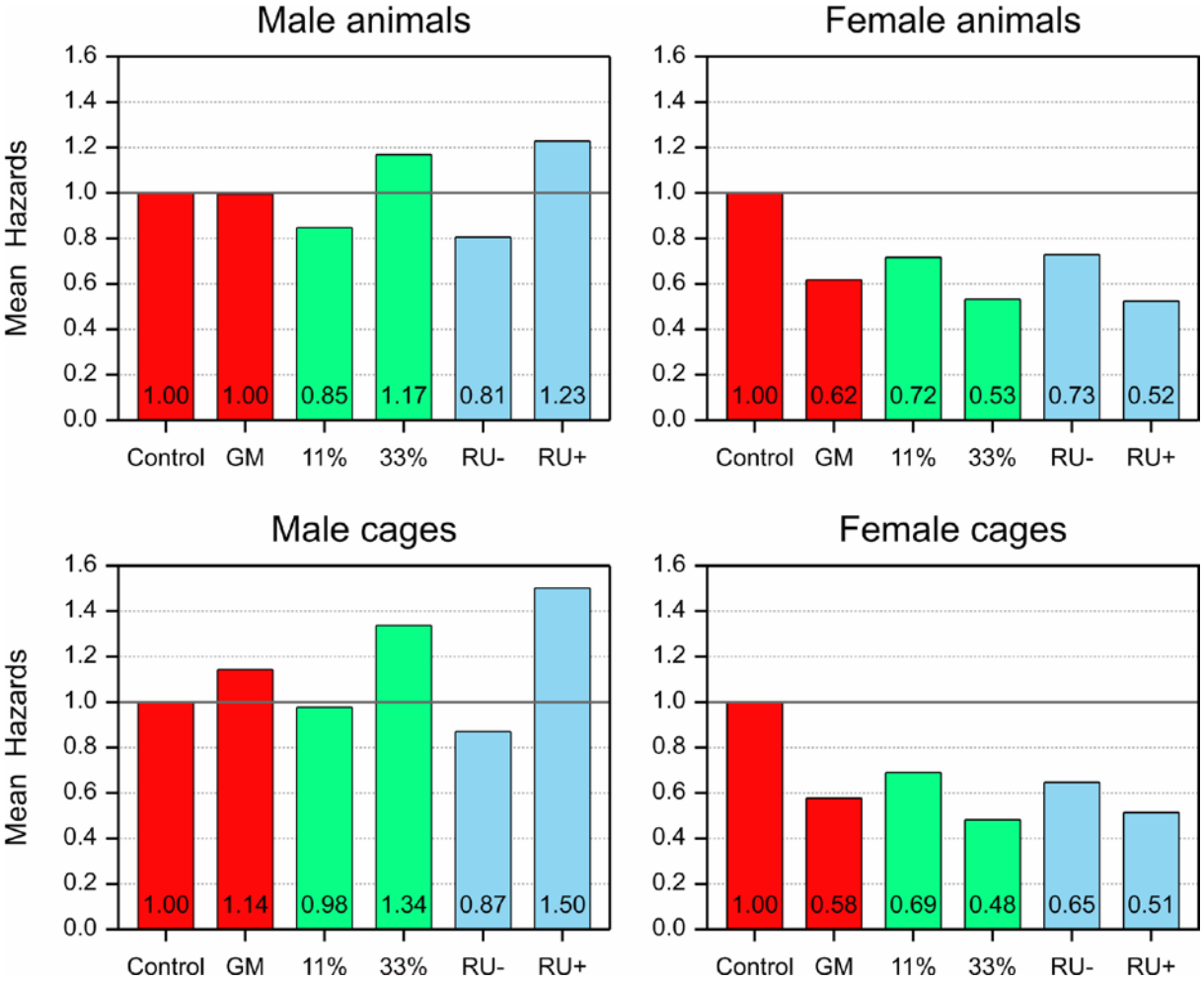

Figure 6 Mean hazard ratios for the overall GM versus non-GM control (red bars), for the 11% and 33% GM inclusion rates (green bars), and for the GM feeds without Roundup (RU-) and with Roundup (RU+, blue bars).

4.2 Analysis of mortality at 24 months

4.2.1 Method

Whereas survival analysis considers the mortality pattern over the time, a more restricted analysis can be done on the mortality rates at specific points in time. In such analyses the exact time of death of dead animals is not considered. Death events of animals were rare up to month 12, with at most 3 dead animals per group of 50 (Table 6,). Therefore only the mortality rates at 24 months were statistically analysed.

In this analysis only the 50 animals per sex and feeding group that were part of the 2-year cohort were statistically analysed. Fitting a beta-binomial regression model (Williams, 1982), by means of maximum likelihood, to the number of dead animals in each cage as response variable revealed that the estimate of the over-dispersion parameter ϕ equals its bounded value of 0.0001 (males and females). This indicates that there was no over-dispersion and therefore the ordinary logistic model (McCullagh and Nelder, 1989) was used to analyse the number of dead animals per cage. After allowing for differences between blocks one-sided pairwise Wald tests were performed with null hypothesis that the mortality probability of a GM feed is equal to or smaller than the mortality

probability of the non-GM control feed. The logistic model was also fitted using a factorial representation of the dose groups.

Table 6 Mortality of males and females after 3, 6, 12 and 24 months. Mortality of a cage is defined as mortality of the fist animal in the cage.

Feed	Male animals (50 total)				Male cages (25 total)			
	3m	6m	12m	24m	3m	6m	12m	24m
Control	-	-	-	18	-	-	-	14
NK11-	-	-	2	14	-	-	2	12
NK33-	-	-	1	17	-	-	1	14
NK11+	-	1	2	18	-	1	2	16
NK33+	-	-	2	27	-	-	2	21
Feed	Female animals (50 total)				Female cages (25 total)			
	3m	6m	12m	24m	3m	6m	12m	24m
Control	-	1	3	24	-	1	3	17
NK11-	1	1	3	26	1	1	3	18
NK33-	-	-	2	20	-	-	2	15
NK11+	-	-	1	10	-	-	1	16
NK33+	-	-	-	17	-	-	-	14

4.2.2 Results

This section summarises the main findings reported in the companion report (Goedhart & van der Voet 2018d). Mortality rates after 24 months are shown in Figure 7. The average mortality at 24 months was 38% for male animals and 39% for female animals. At the level of the experimental unit (cage) the average mortality at 24 months was 62% for males and 64% for females.

Using the logistic model and one-sided testing, there was one significant result ($P<0.05$) in 8 tests (males and females, 4 comparisons with the Control group). There was an indication ($P=0.029$) that the mortality rate for NK33+ in males (54%) is larger than for the control non-GM feed (36%).

The factorial analysis, employing chi-squared likelihood ratio testing in the logistic regression model, found one significant result ($P<0.05$) in 8 tests. There was no indication for an overall higher mortality rate in the GM groups ($P=0.78$ for males, and $P=0.35$ for females). However, there was an indication for a Roundup effect in males ($P=0.030$). Keeping in mind that the mortality in the male Control group was 36%, the mean percentages dead were 31% for NK603 without RoundUp and 45% for NK603 with RoundUp.

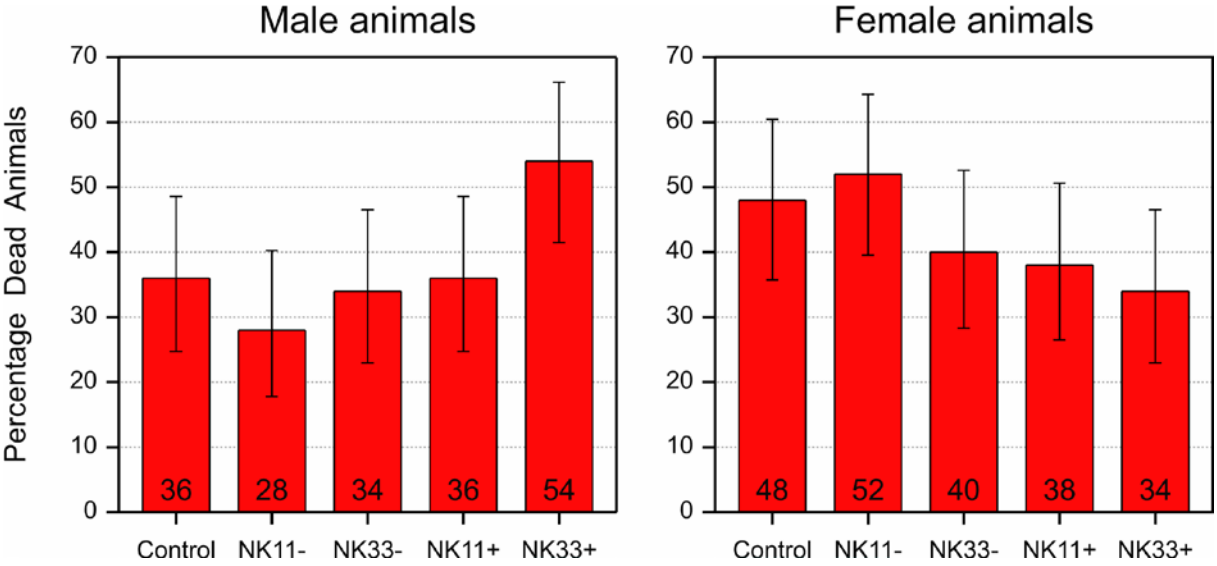

Figure 7 Percentage dead animals after 24 month in the 2-year cohort. 90% confidence intervals for the percentage are indicated by error bars. The percentage is also given at bottom of each bar.

4.3 Equivalence testing using historical data

4.3.1 Method

Equivalence testing was introduced for GM safety assessment for compositional data in the EFSA guidance for risk assessment of food and feed from GM plants (EFSA 2011a). In the context of 90-day studies in rodents, EFSA (2014) recognized the potential advantages of equivalence testing and recommended further investigation. In response to this, an equivalence test was developed in the G-TwYST project. This test compares the difference between a test (T) and a control (C) feed, obtained simultaneously in a current study, to the typical differences between reference (R) varieties obtained in one or more historical studies (van der Voet *et al.*, 2017). The equivalence test is corrected for between-study differences, and the within-study variation between references R is used to set equivalence limits for the difference between T and C in the current study. The so-called Distribution Wise Equivalence (DWE) criterion is used in this test. An equivalence limit for the current study is set using the concept of desired power in a simplified situation, where there is no between-reference variation, where the historical and current studies have the same residual variance, and where the current study is assumed to have a sample size as approved by a regulator. The method is fully described in van der Voet *et al*. (2017) .

The equivalence test of van der Voet *et al*. (2017) requires historic data sets for the endpoints of interest. For the 3 months data in study A the GRACE reference data described in section 2.2 serve this purpose. For the data obtained after 6, 12 and 24 months no corresponding reference data were available. If variance components between and within reference feeds would be stable over time, the 3-months GRACE data could still be used.

Where possible, employing the historic GRACE studies, some statistics are calculated from the historic data: the within-study between reference feeds sums of squares (SS_R), the residual sums of squares (SS_E) and their associated degrees of freedom df_R and df_E. It also involves the effective unit

replication n_{eff} which is necessary to estimate the between reference variance employing the mean squares for feeds and for residuals. The required values based on the five GRACE studies A-E are given in Appendix 2.

The test also requires, for the current G-TwYST study, estimates of differences between the GMO feeds and the control feed, as well as the residual sums of squares and the associated degrees of freedom resulting from an analysis of variance. These are given in the companion reports (Goedhart % van der Voet 2018abcd).

The equivalence limit θ_0 for the DWE criterion is only based on the design values of the historical studies and on three regulatory values: the minimal regulatory sample size n_0, a probability α which defines a $100(1-\alpha)$% two-sided confidence interval for the difference in the current study, and a probability β which defines the desired power $1-\beta$ for the equivalence test. The regulatory sample size n_0 was set to the level of replication in G-TwYST study A, which is different for different variable groups. Values $\alpha = 0.05$ and $\beta = 0.05$ resulting in a power of 0.95. Furthermore the equivalence limit θ_0 is calculated by simulating a large number of datasets in a simplified situation, where for each datasets an upper $100(1-\alpha)$% percentile, θ^0_{upp}, for the DWE criterion is approximated by a large number of so-called GPQ samples. We simulated 40,000 datasets with 15,000 GPQ samples for each dataset. Note that θ_0 is calculated as the upper $100(1-\beta)\%$ percentile of the thus obtained 40.000 values of θ^0_{upp}. The DWE criterion for the current dataset was approximated by means of 100,000 GPQ samples. Note that the equivalence limit θ_0 is calculated assuming a regulatory sample size n_0 which implies $2(n_0-1)$ degrees of freedom for error in the current study. The current G-TwYST study has more degrees of freedom for error, see Table 3.

The DWE equivalence test results in a DWE interval as a so-called equivalence limit scaled difference (ELSD), which can be used both for difference and for equivalence testing. The hypothesis of no difference is rejected in case the interval does not contain zero, while the non-equivalence hypothesis is rejected when the interval fully lies inside the interval (-1,1). In the companion reports (Goedhart % van der Voet 2018abcd) the confidence intervals are also presented at the original ratio scale, with inclusion of the estimated equivalence limits (red bars) and their uncertainty (blue bars). Note that the latter graphs cannot be used directly for performing the equivalence test. However, they show the effects and equivalence limits at a more familiar scale.

4.3.2 Results

This section summarises the main findings reported in the companion reports (Goedhart & van der Voet 2018abcd).

The residual variances for 3, 6, 12 and 24 months data are shown in Figure 8, expressed as a percentage of the corresponding residual variance in the historical GRACE data. From these graphs it is clear, that the variances at month 3 and month 6 are similar, such that the historical 3-months GRACE data can also serve as reference data for equivalence tests with 6 months data. However, at month 12 many variances are much higher, and such use becomes doubtful. At month 24, most variances have become much larger. The conclusion is that the use of the 3-months GRACE data as a historical reference for equivalence testing in G-TwyST should be restricted to the 3 months and 6 months data.

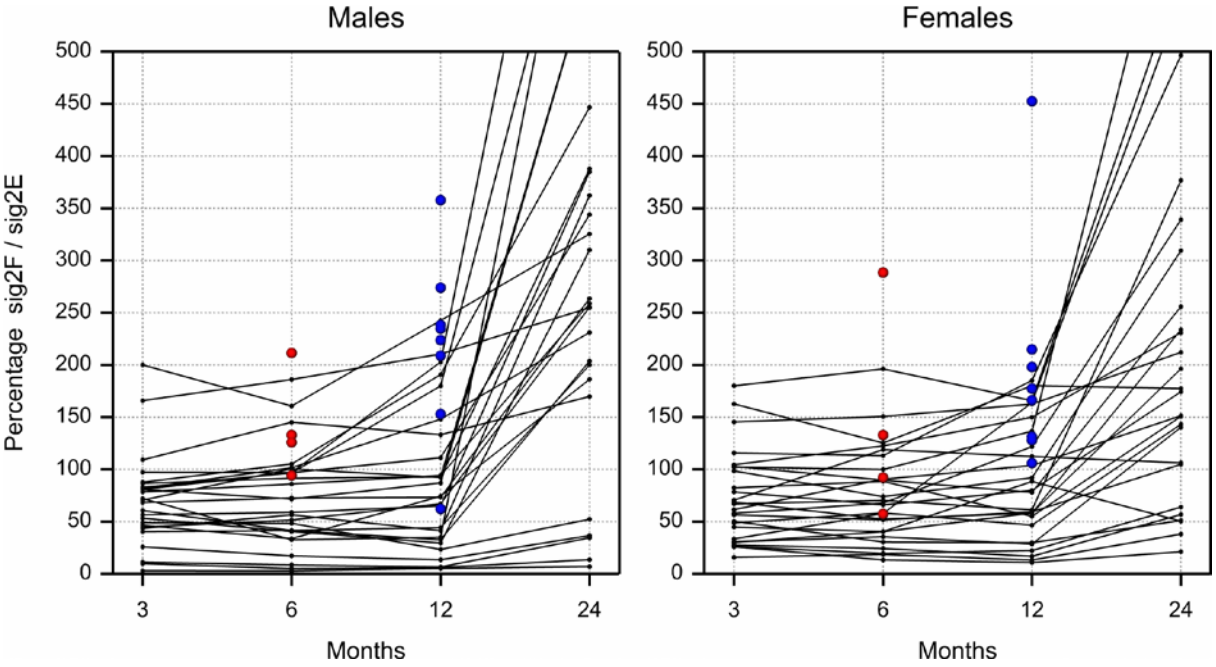

Figure 8 Residual variance for 4 points in time, expressed as a percentage of the historical reference residual variance. The red points are for the diffWBC measurements, which were only observed after 6 months, while the blue points are for the organ weights which were only observed after 12 months.

Equivalence tests were performed given tentative settings for regulatory parameters, $\alpha = 0.05$, $\beta = 0.05$ and n_0 equal to the sample sizes in the current study, i.e. 35 for the weight variables and 20 for haematology, differential white blood cell counts and clinical biochemistry. The results are shown in Figure 9 - Figure 12, and summarised in Table 7.

The performance of the equivalence test was dependent on the ratio of the residual variance in the current study to the residual variance in the historical study. Combined over all 412 equivalence tests with a variance ratio (*VR*) of at most 1.5, equivalence was established in 411 cases (99.8%). However, for the 44 tests with *VR* > 1.5, this was possible in only 12 cases (27%). In Table 7 it is also shown that in 455 of all 456 tests the point estimate was inside the equivalence limits, and therefore equivalence was more likely than not according to the definition of EFSA (2011). Again, the exception was for one of the *VR* > 1.5 cases.

Obviously, large residual variances in the current study relative to the reference data are problematic. Analysing these situations might lead to different conclusions. On the one hand, if the problem is seen as a too low precision obtained in the current study, then the quality of the current study for this variable is sub-standard, and it is just appropriate that this type of equivalence testing fails to prove equivalence. Alternatively, if the high variance ratios are considered to be the result of accidentally very small variation in the reference data, then the lesson would be to improve the reference data collection.

Table 7 Summary of equivalence tests employing historical data. VR denotes the ratio of the residual variance in the current study and the residual variance in the historic GRACE studies.

3 months	**Male rats**		**Female rats**	
	VR ≤ 1.5	*VR* > 1.5	*VR* ≤ 1.5	*VR* > 1.5
#tests	100	8	96	12
#equivalence	100	4 failed MCV 4x	95 failed: MCHC 1x	4 failed: growthRate 4x CHOL 3x P 1x
#equivalence more likely than not	100	8	96	12
6 months	**Male rats**		**Female rats**	
	VR ≤ 1.5	*VR* > 1.5	*VR* ≤ 1.5	*VR* > 1.5
#tests	108	12	108	12
#equivalence	108	2 failed: MCV 2x MCH 4x Lymphocytes 4x	108	2 failed: MCHC 2x Monocytes 4x CHOL 4x
#equivalence more likely than not	108	12	108	11 failed: Monocytes 1x

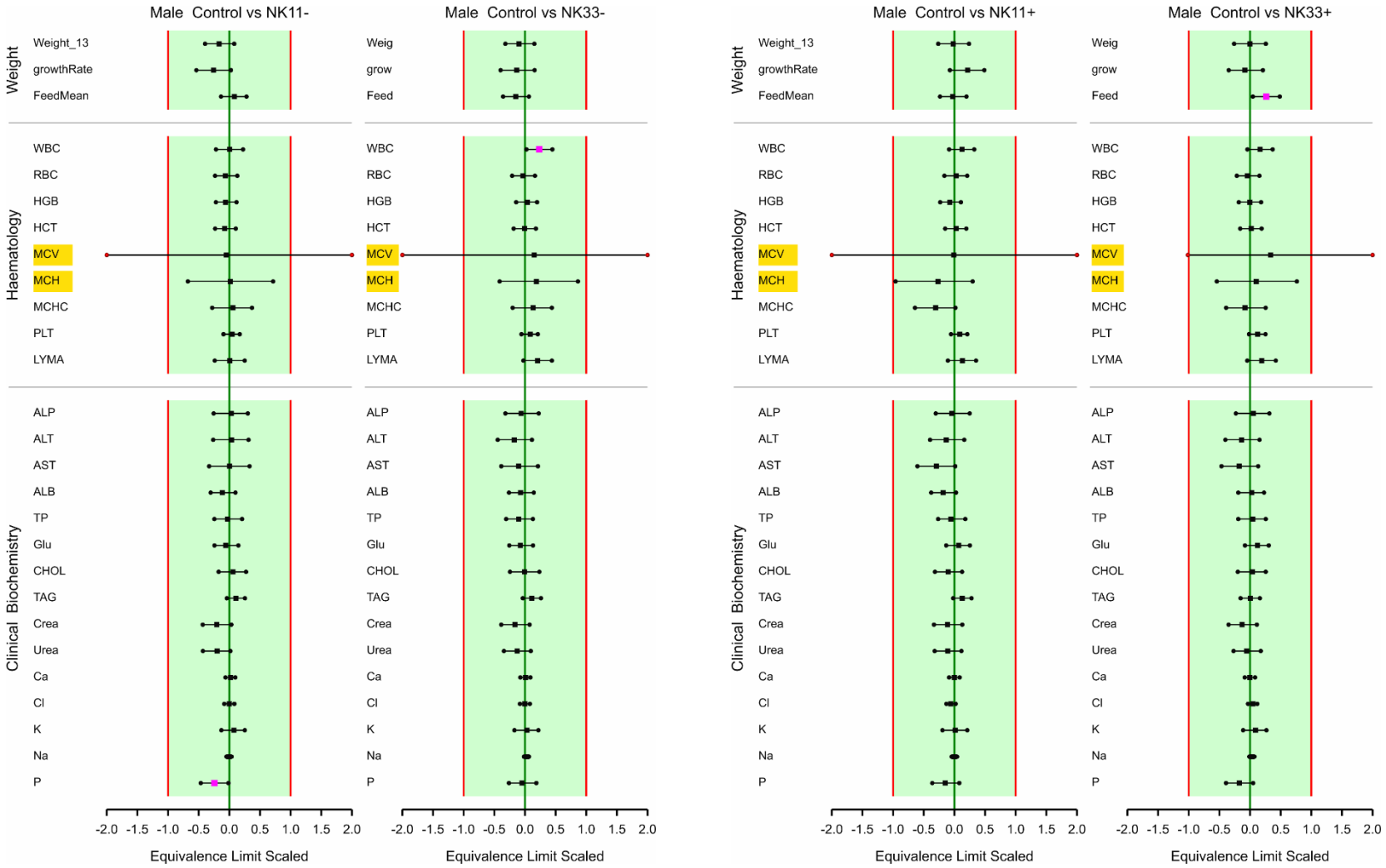

Figure 9 Equivalence testing at 3 months of GM feeding groups versus the non-GM control feed for males. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. Endpoints labelled with a golden background have a large residual variance compared to the historical studies (VR>150%). Fuchsia coloured symbols denote a significant difference.

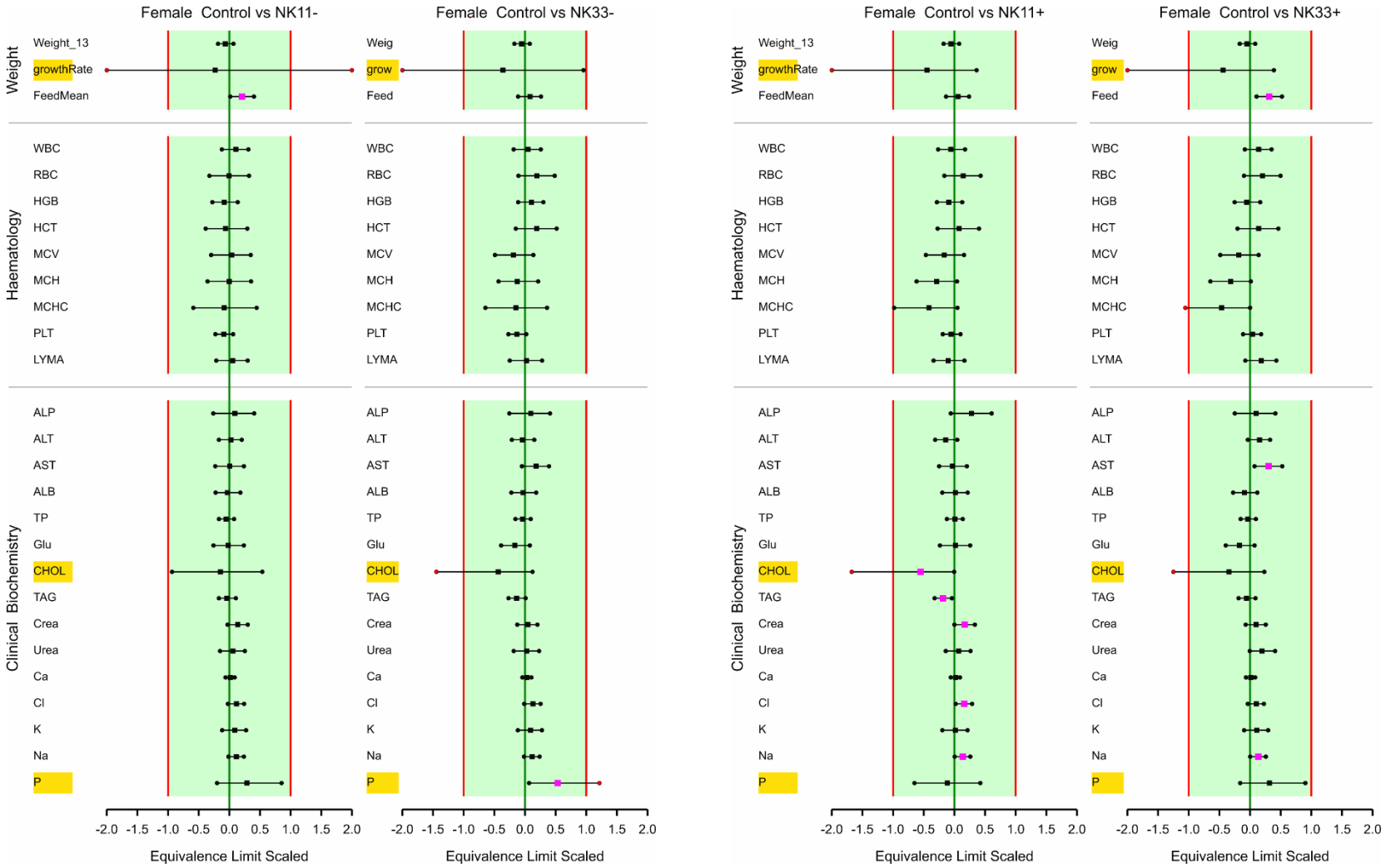

Figure 10 Equivalence testing at 3 months of GM feeding groups versus the non-GM control feed for females. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. Endpoints labelled with a golden background have a large residual variance compared to the historical studies (VR>150%). Fuchsia coloured symbols denote a significant difference.

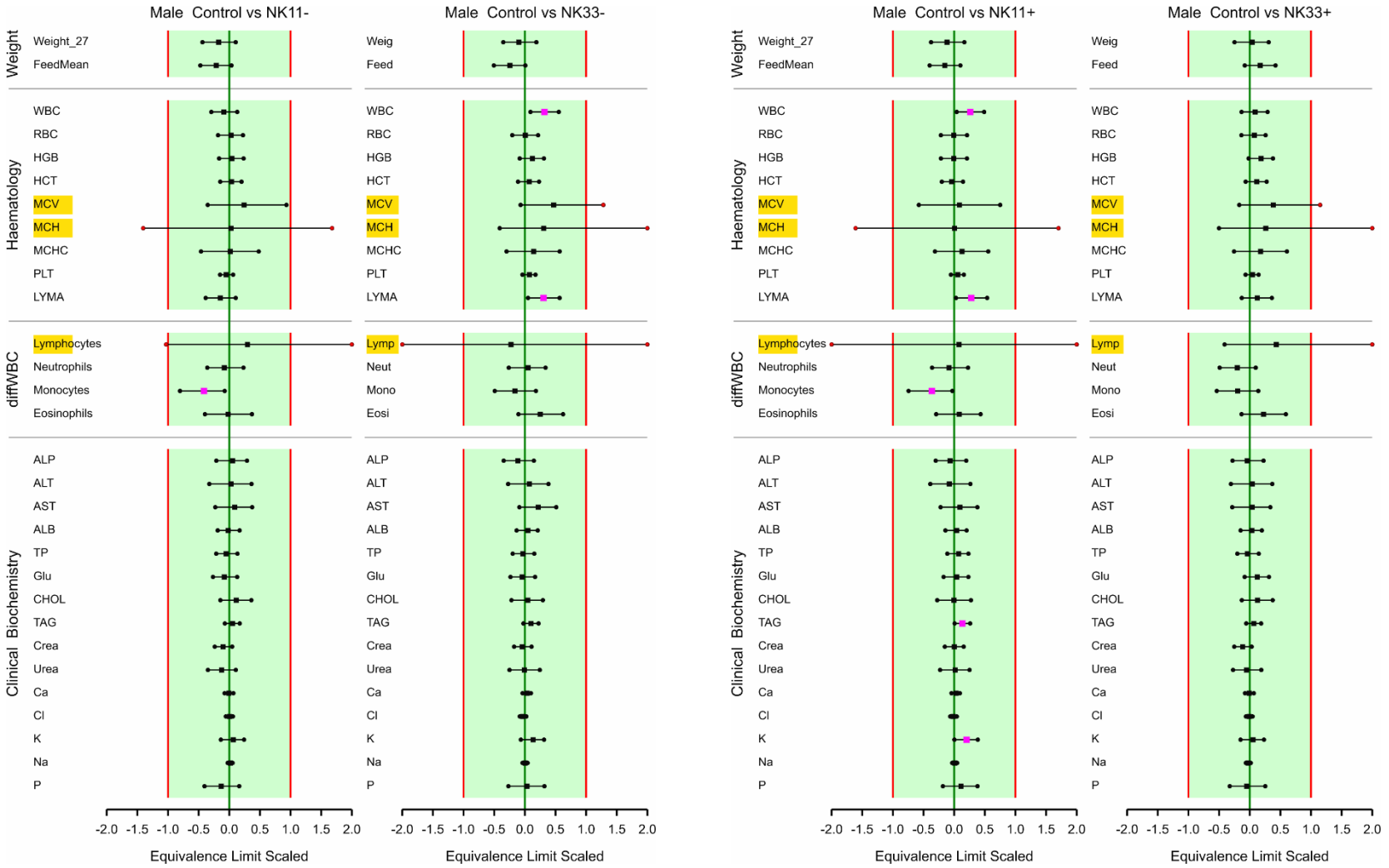

Figure 11 Equivalence testing at 6 months of GM feeding groups versus the non-GM control feed for males. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. Endpoints labelled with a golden background have a large residual variance compared to the historical studies (VR>150%). Fuchsia coloured symbols denote a significant difference.

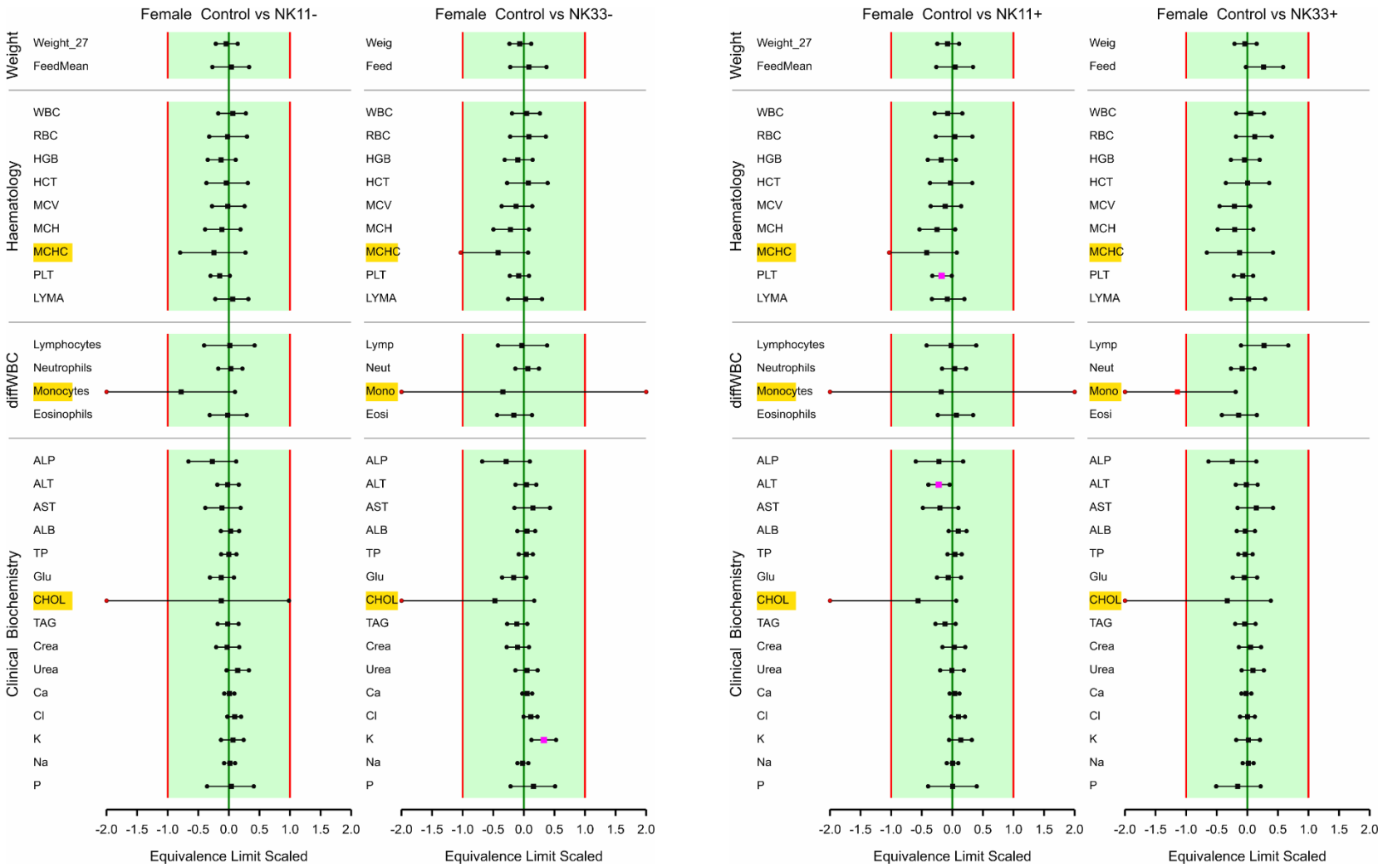

Figure 12 Equivalence testing at 6 months of GM feeding groups versus the non-GM control feed for females. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. Endpoints labelled with a golden background have a large residual variance compared to the historical studies (VR>150%). Fuchsia coloured symbols denote a significant difference.

4.4 Equivalence testing using target effect sizes

4.4.1 Method

For a limited number of variables Hong *et al* (2017) use what they call targeted effect sizes for the purpose of statistical power analysis for a rat sub-chronic feeding study. Although they warn that these effect sizes should not be considered synonymous with biologically or toxicologically relevant effects, these targeted effect sizes were used for equivalence testing. The targeted effect sizes for nine variables that are also relevant in G-TwYST are given in Table 8, along with the implied limits on the log-ratio scale which are used in the equivalence test. Note the asymmetry in these limits: for a targeted effect size of +/- 30%, +30% corresponds to a factor 1.3 which is 0.262 on the log scale, while -30% corresponds to a factor 0.7 which equals -0.357 on the log scale.

Table 8 Targeted effect sizes from Table 1 in Hong *et al* (2017) along with their implied lower and upper limits on the ratio scale and on the log-ratio scale.

Name in Hong *et al* (2017)	G-TwYST Name	Targeted effect size	Ratio scale		Log-ratio scale	
			Lower	Upper	Lower	Upper
Body weight; final non-fasted	BodyWeight	- 10%	0.90	-	-0.105	-
Leukocyte (WBC) count	WBC	+/- 30%	0.70	1.30	-0.357	0.262
Lymphocyte (ALYM) count	LYMA	+/- 30%	0.70	1.30	-0.357	0.262
Alkaline phosphatase (ALKP)	ALP	+ 100%	-	2.00	-	0.693
Creatinine (CREA)	Crea	+ 50%	-	1.50	-	0.405
Blood urea nitrogen (BUN)	Urea	+ 50%	-	1.50	-	0.405
Cholesterol (CHOL)	CHOL	+ 200%	-	3.00	-	1.099
Kidney, % body weight	Kidney	+ 25%	-	1.25	-	0.223
Liver, % body weight	Liver	+ 25%	-	1.25	-	0.223

Denoting the limits on the log-ratio scale as δ_{low} and δ_{upp}, the two-sided non-equivalence null hypothesis reads, with Δ the ratio of the mean of a GMO feed and the mean of the control feed:

$$H_0: \quad \log(\Delta) < \delta_{low} \quad \text{or} \quad \log(\Delta) > \delta_{upp}$$

$$H_1: \quad \delta_{low} \leq \log(\Delta) \leq \delta_{upp}$$

This was tested by means of the TOST approach of Schuirmann (1987) at the 5% level which is equivalent to checking whether the 10% confidence interval for $\log(\Delta)$ lies completely within the interval $(\delta_{low}, \delta_{upp})$ composed of the equivalence limits. For one-sided tests the same confidence interval can be used where only one of the confidence bounds is relevant.

4.4.2 Results

This section summarises the main findings reported in the companion reports (Goedhart & van der Voet 2018abcd).

90% confidence intervals for 7 variables, with equivalence limits according to the targeted effect sizes in Hong *et al.* (2017), are given in Figure 13 - Figure 16 for 3, 6, 12 and 24 months data, each for males and females. Note that these are 90% intervals. The results of the equivalence tests are summarised in Table 9. Equivalence is easily demonstrated for the 3, 6 and 12 months data, but again

the large residual variance of the 24 months data is responsible for 13 cases where equivalence could not be proven.

Table 9 Summary of equivalence tests with fixed target effect sizes.

	Males				Females			
	3m	6m	12m	24m	3m	6m	12m	24m
Number of tests	28	28	36	28	28	28	36	28
Number of tests showing equivalence	28	28	36	22[a]	28	28	36	21[b]
Number of tests showing equivalence more likely than not	28	28	36	28	28	28	36	28

[a] failures for Weight_104 (1x), WBC (1x), LYMA (4x); [b] failures for Weight_104 (3x), WBC (3x), LYMA (1x)

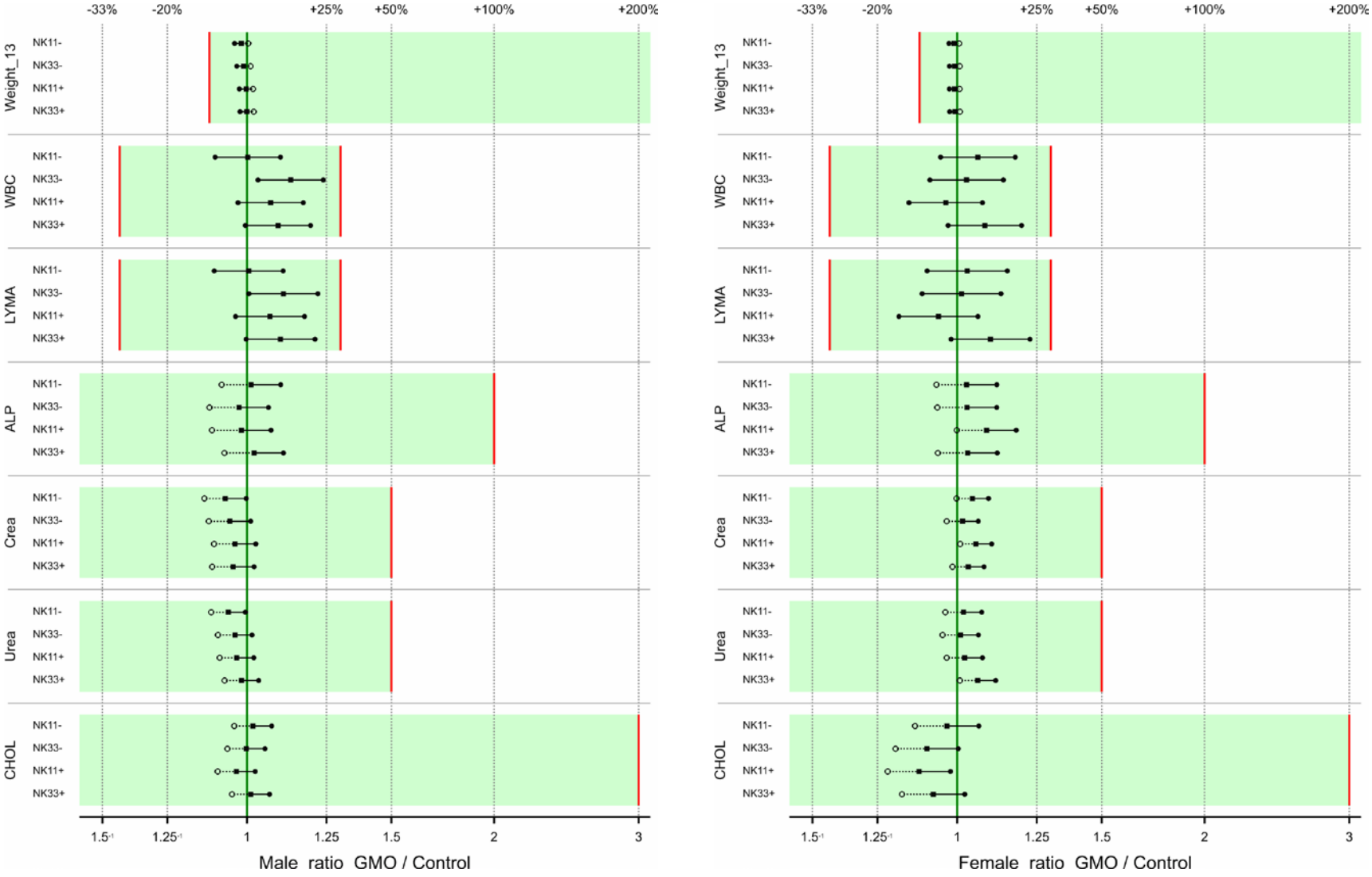

Figure 13 3 months data, 90% confidence intervals for the ratio of the mean of the GMO feed and the control feed for selected variables for males (left) and females (right), along with equivalence intervals defined by targeted effect sizes of Hong *et al.* (2017).

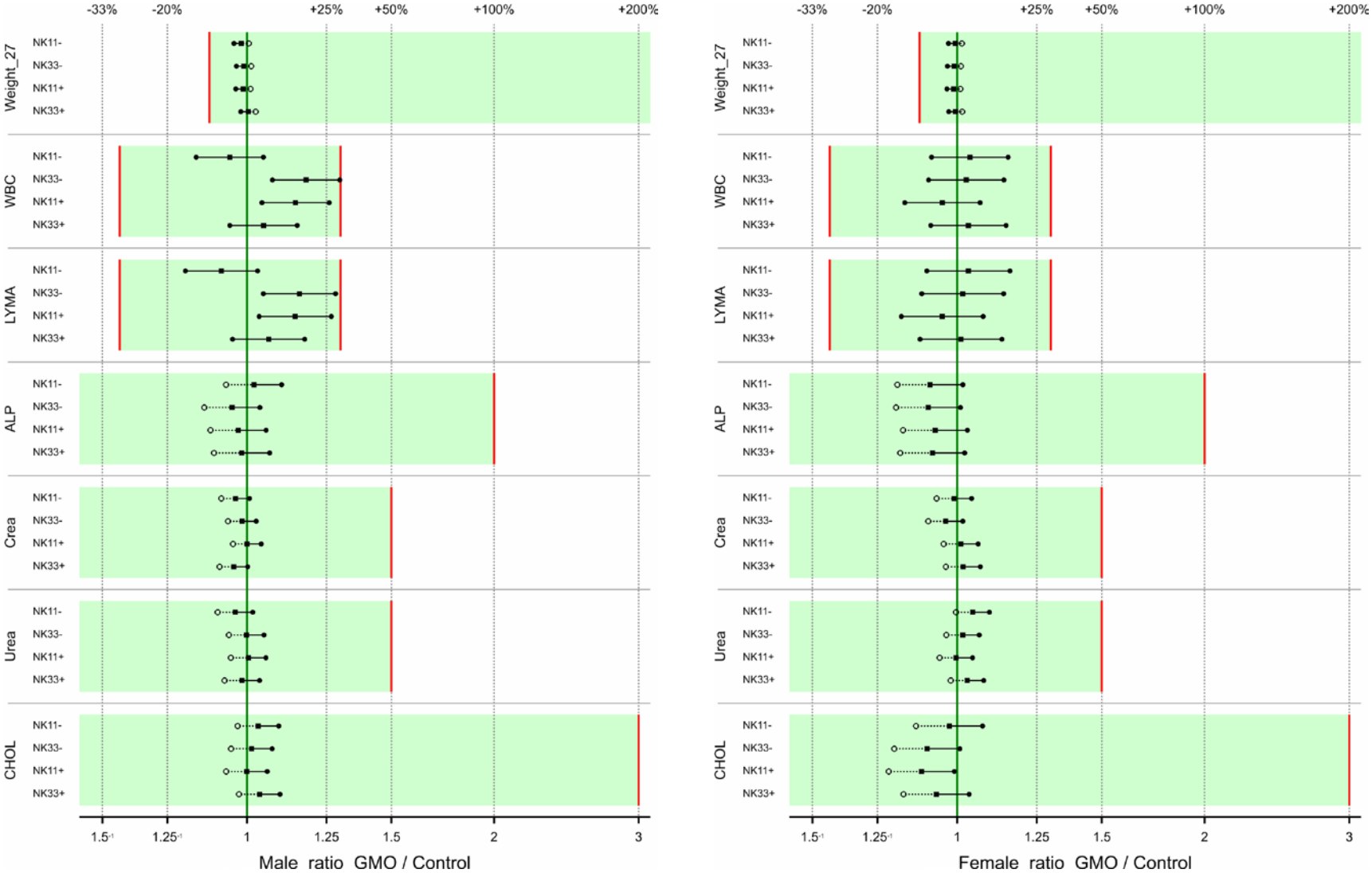

Figure 14 6 months data, 90% confidence intervals for the ratio of the mean of the GMO feed and the control feed for selected variables for males (left) and females (right), along with equivalence intervals defined by targeted effect sizes of Hong *et al.* (2017).

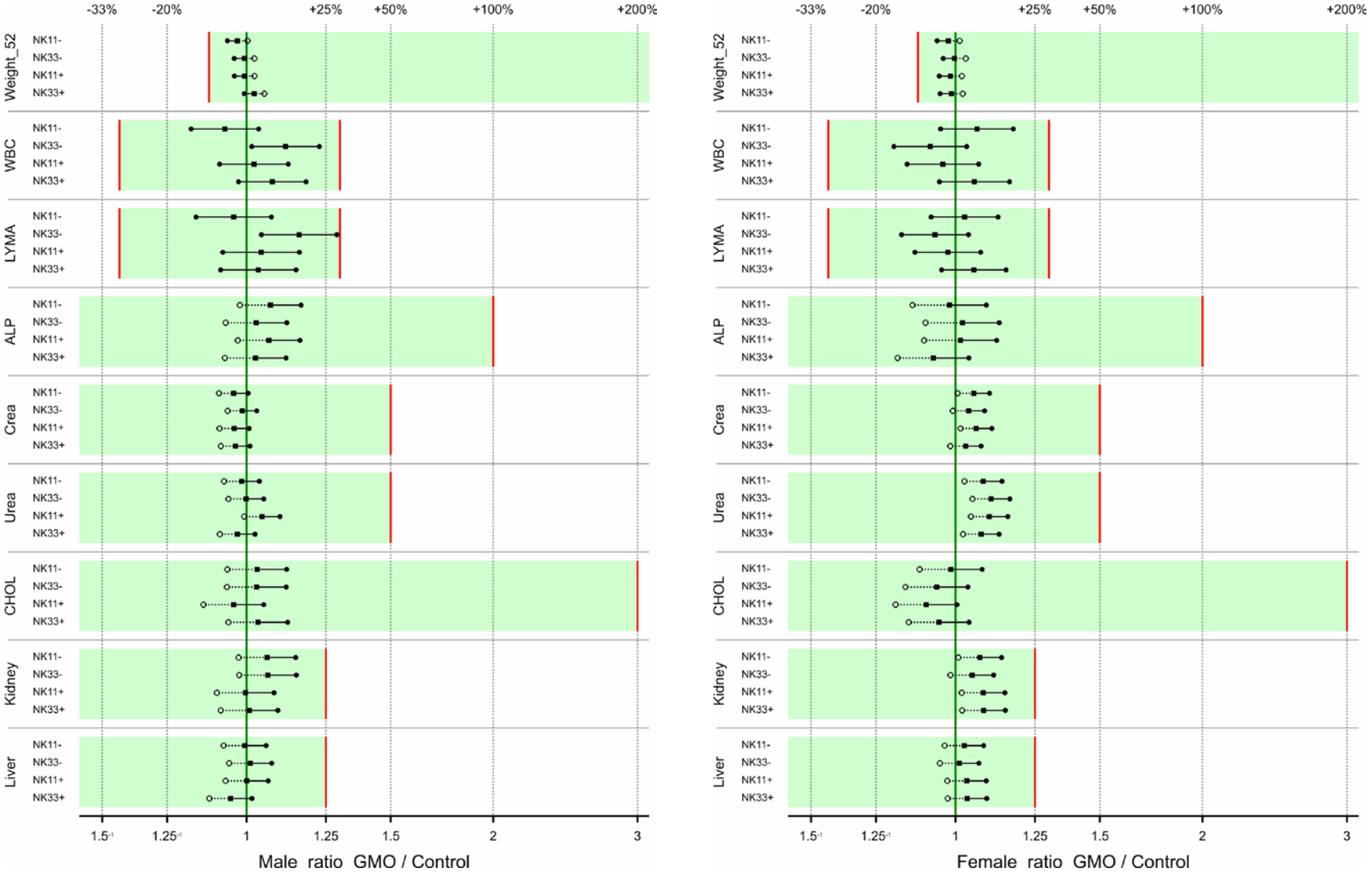

Figure 15 12 months data, 90% confidence intervals for the ratio of the mean of the GMO feed and the control feed for selected variables for males (left) and females (right), along with equivalence intervals defined by targeted effect sizes of Hong *et al.* (2017).

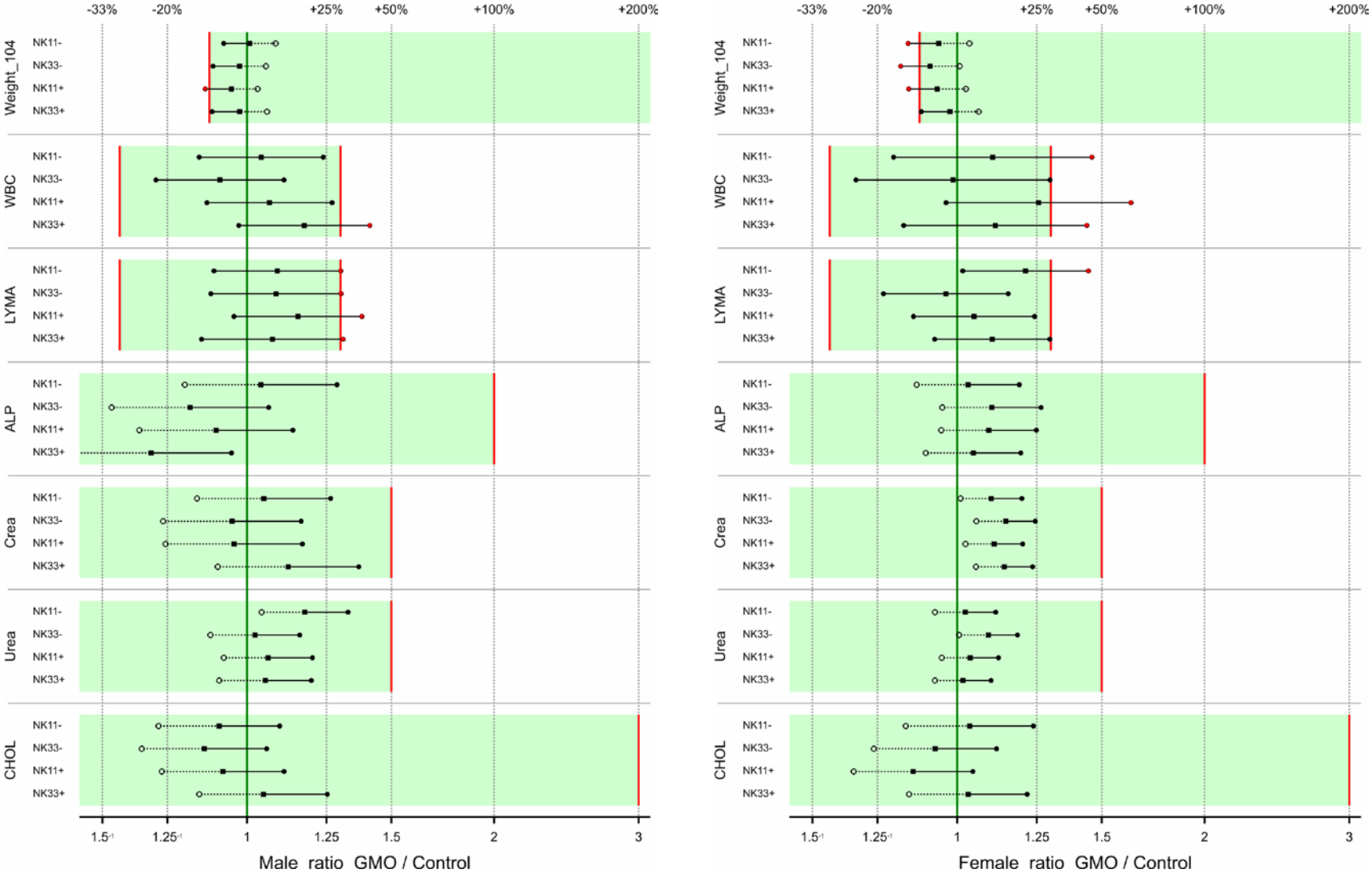

Figure 16 24 months data, 90% confidence intervals for the ratio of the mean of the GMO feed and the control feed for selected variables for males (left) and females (right) that survived for 24 months, along with equivalence intervals defined by targeted effect sizes of Hong *et al.* (2017).

4.5 Classical statistical analysis

4.5.1 Method

G-TwYST study A is based on OECD guidance 453 on combined chronic toxicity and carcinogenicity studies (OECD 2009), EFSA guidance complementing the OECD guidance for whole food/feed studies (EFSA 2011b), and additional EFSA clarifications (EFSA 2013, 2014). OECD guidance 453 (OECD 2009) requires numerical results to be evaluated by an appropriate and acceptable statistical method, but gives no further guidance on statistical analysis. More detailed guidance, although strictly meant for chronic and carcinogenicity studies, is provided in chapter 4 of OECD guidance document 116 (OECD 2012), which describes a flowchart for statistical analysis methods (reproduced in Figure 17).

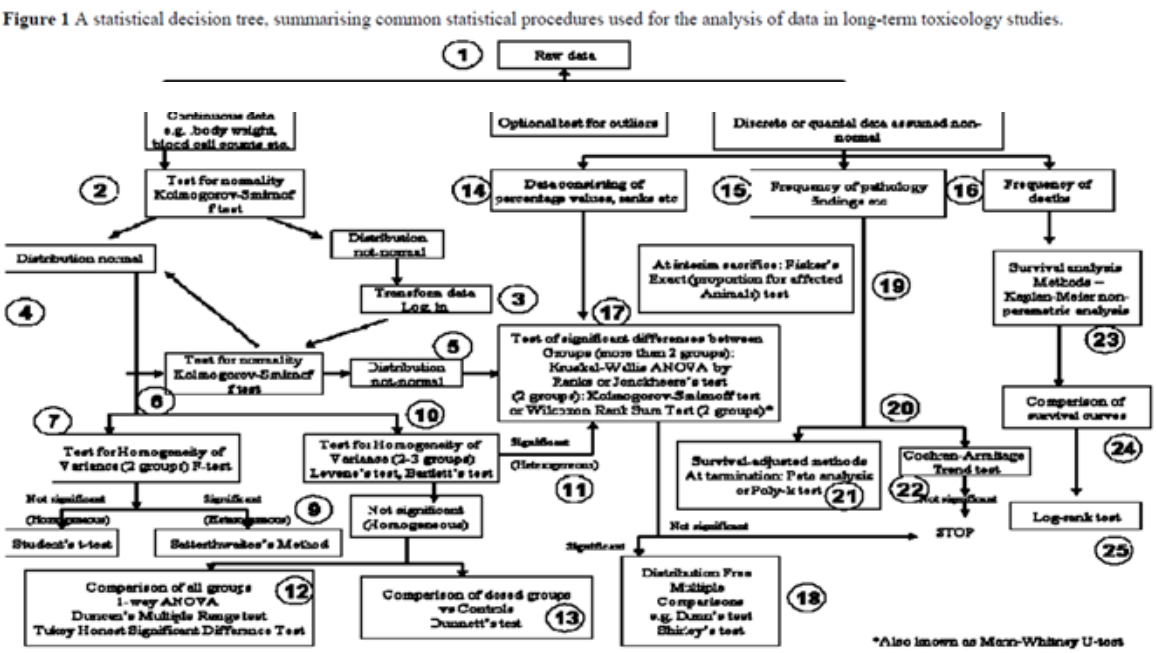

Figure 17 Classical approach to statistical analysis of data in long-term toxicity studies (copied from OECD 2012).

EFSA (2011b) gives further guidance, such as considering cage as the experimental unit, and including block in the model for data from a randomised block design (as is the case for the G-TwYST study). We applied classical statistical methods for continuous data in line with these OECD and EFSA approaches, and very similar to the approaches followed in the GRACE project (Schmidt and Schmidtke 2014, Schmidt *et al* 2015ab).

A classical analysis of variance was performed on the cage means after log transforming the data. This was done in the statistical program R. The R-script which analyses a single response variable is given in Appendix 3. The classical analysis involves:

- Analysis of variance according to the randomized block design employing the model "Block + Treatment" where Treatment defines the five feeding groups. The model was fitted by means of linear regression, using the lm() function in R, because this takes proper account of any missing values. The usual summary statistics are saved as well as estimates for the difference between GMO feeds and the control and corresponding standard errors, t-values and p-values. These are

all calculated using the pooled ANOVA residual standard error which generally has 28 degrees of freedom.

- The ANOVA p-values do not take account of multiple comparisons between the feeds. Therefore Dunnett's test was performed, which also compares the GMO feeds with the control feed, taking account of the number of comparisons made. Dunnett's test is performed by means of the glht() function in the multcomp R-package.
- The residuals of the analysis of variance are checked for normality using the Kolmogorov-Smirnov test as well as the Shapiro-Wilk test for normality. These tests are only approximate since the residuals are not independent. The p-value of the Kolmogorov-Smirnov test is not reported since it is almost always larger that the p-value of the Shapiro-Wilk test, moreover for variables where it is smaller the p-value is far from significant.
- Note that the ANOVA residuals were already assessed by means of a normal probability plot and a plot of residuals versus fitted values (see companion reports).
- The non-parametric Wilcoxon signed rank test is used to test for a difference between each GMO feed and the control feed. Note that this test only uses data of these two feeds and that the test employs the within block difference between the GMO feed and the control feed. The p-value of the test is calculated by means of the wilcox.test() function in R which calculates exact probabilities.
- Homogeneity of variance is assessed by means of Bartlett's test and by means of Levene's test both using the mean and the median. These test do not take blocking into account and basically compare the within feed variances. Note that homogeneity of variance was already assessed by means of a plot of residuals versus fitted values (see companion reports). The p-value of the Levene test with the median is not reported since it is almost always larger that the p-value of the Levene test with the mean, and for variables where it is smaller the p-value is far from significant. Note that both analysis of variance and non-parametric tests require homogeneity of variance.
- Finally, for each feeding group separately, normality was assessed by means of the Kolmogorov-Smirnov test and the Shapiro-Wilk test. The p-value of the Kolmogorov-Smirnov test is always larger than 5% and is thus not reported.

4.5.2 Results

This section summarises the main findings reported in the companion reports (Goedhart & van der Voet 2018abcd). Hypotheses of equality were tested against two-sided alternatives of difference for each of the four GM groups vs. the non-GM Control group. Across the four time points (3, 6, 12, 24 months) and all available data, this amounted to 1256 comparisons, using three statistical methods (Dunnett's test, t-test, Wilcoxon test) for a total of 3768 test results. All tests were performed at the 5% level, so that 0.05 x 3768 = 188 significant results would be in line with the theoretical level of the tests under the joint null hypotheses. Table 10 gives an overview of the numbers and frequencies of significant results, whereas the full results are reported in

Table 10 Absolute numbers and relative frequencies of significant results (P<0.05) in classical tests comparing GM groups to the non-GM group, combined for males and females. For the three individual tests, frequencies significantly higher (P<0.05) than the nominal test level of 5% by an exact binomial test are shown with a red background.

	3 months	6 months	12 months	24 months

Number of comparisons	288		320		352		296	
Dunnett	3	(1%)	3	(1%)	10	(3%)	2	(1%)
t- test	17	(6%)	14	(4%)	38	(11%)	11	(4%)
Wilcoxon	19	(7%)	14	(4%)	32	(9%)	7	(2%)
Any of these	26	(9%)	20	(6%)	46	(13%)	14	(5%)

The highest frequency of significant results is observed for the 12 months data, where the t-test finds 38 significant differences (11%) and Wilcoxon's test 32 (9%), although Dunnett's test, which correct for the multiplicity of four comparisons per endpoint, finds only 10 significant results (3%). In comparison, the frequencies for the 3, 6 and 24 months data are lower, and not significantly different from the nominal test level of 5%.

Table 11 Estimated ratios of GM groups versus the non-GM control group for male rats for months 3, 6, 12 and 24. Significant ratios are marked, with red background colouring, as follows: D: P<0.01 by Dunnett-test, d: P<0.05 by Dunnett-test, T: P<0.01 by t-test but not by Dunnett-test, t: P<0.05 by t-test but not by Dunnett-test, W: P<0.01 by Wilcoxon signed rank test, w: P<0.05 by Wilcoxon signed rank test.

Weights Males	NK11- vs Control				NK33- vs Control				NK11+ vs Control				NK33+ vs Control			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Weight	0.98	0.98	0.97	1.01	0.99	0.99	0.99	0.98	1.00	0.99	0.99	0.96	1.00	1.00	1.02	0.98
growthRate	0.99	0.99	0.96	-	1.00	1.00	1.02	-	1.01	0.97	1.01	-	1.00	1.03	1.13 tw	-
FeedMean	1.01	0.98 w	0.96 dTW	0.98	0.99	0.98 w	0.97 w	0.98	1.00	0.99	0.97 tW	0.95 t	1.03 tw	1.02	1.01	1.03
Haematology Males	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
WBC	1.00	0.95	0.94	1.04	1.13 tW	1.18 dTw	1.12	0.93	1.07	1.15 t	1.02	1.07	1.09	1.05	1.08	1.17
RBC	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.01	1.00	1.00	1.02	1.00	1.00	1.01	1.01	0.99
HGB	0.99	1.00	1.01	0.97	1.00	1.01	1.00	1.02	0.99	1.00	1.01	1.00	1.00	1.02	1.01	1.01
HCT	0.99	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.00	1.00	1.03 tw	0.99	1.00	1.01	1.01	1.00
MCV	1.00	1.01	1.00	1.02	1.00	1.01	1.01	1.02	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.01
MCH	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	0.99	1.00	0.99	1.00	1.00	1.01	1.00	1.01
MCHC	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	0.99	1.00	0.99 tw	1.00	1.00	1.01	1.00	1.01
PLT	1.03	0.97	0.99	1.03	1.05	1.04	1.08 d	1.05	1.05	1.04	1.03	1.05	1.07 w	1.03	1.08 dTW	1.04
LYMR	0.99	0.98	1.00	1.04	0.98	0.99	1.04	1.05	1.00	1.00	1.02	1.04	1.01	1.01	0.98	0.94
LYMA	1.00	0.93	0.96	1.09	1.11	1.16 tw	1.16 t	1.08	1.07	1.15 t	1.04	1.15	1.10	1.06	1.03	1.07
diffWBC Males	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Lymphocytes	-	1.02	-	-	-	0.99	-	-	-	1.00	-	-	-	1.03	-	-
Neutrophils	-	0.97	-	-	-	1.02	-	-	-	0.97	-	-	-	0.93	-	-
Monocytes	-	0.75 tw	-	-	-	0.89	-	-	-	0.78 tw	-	-	-	0.87	-	-
Eosinophils	-	0.98	-	-	-	1.28	-	-	-	1.09	-	-	-	1.24	-	-
ClinChem Males	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
ALP	1.01	1.02	1.07	1.04	0.98	0.96	1.03	0.85	0.99	0.98	1.07	0.92	1.02	0.99	1.03	0.76
ALT	1.01	1.01	1.01	1.03	0.96	1.02	0.98	1.02	0.97	0.98	1.05	1.02	0.97	1.01	1.04	1.05
AST	1.00	1.03	1.02	1.13	0.97	1.07	0.99	1.05	0.91	1.03	0.94	1.05	0.95	1.01	0.99	1.06
BIL	1.05	1.03	0.97	1.03	1.05	1.12 t	0.99	1.04	0.94	1.08	0.94	1.01	1.02	1.07 w	1.02	0.95

ALB	0.99	1.00	0.99	1.00	0.99	1.01	1.01	1.06	0.98 w	1.00	1.00	1.05	1.00	1.00	0.99	1.01
TP	1.00	1.00	0.99	1.01	0.99	1.00	1.01	1.02	1.00	1.01	1.00	1.01	1.00	1.00	1.00	1.01
Glu	0.98	0.97	1.00	1.00	0.98	0.98	0.99	1.02	1.02	1.01	1.05	0.97	1.04	1.04	1.01	0.90
CHOL	1.02	1.03	1.03	0.93	1.00	1.01	1.03	0.89	0.97	1.00	0.96	0.93	1.01	1.04	1.03	1.05
TAG	1.09	1.05	1.03	1.26 w	1.10	1.09	1.12	1.08	1.11	1.12 tw	1.18 tw	1.04	1.01	1.06	1.14	1.22
Crea	0.94	0.97	0.96	1.05	0.95	0.99	0.99	0.96	0.97	1.00	0.97	0.96	0.96	0.96	0.97	1.12
Urea	0.95	0.97	0.99	1.18 t	0.97	1.00	1.00	1.02	0.97	1.00	1.04	1.06	0.99	0.99	0.98	1.05
cHGB	1.12	1.16	0.93	1.45 tw	1.07	1.24	1.06	1.13	0.94	1.24	1.03	1.18	1.06	1.10	1.05	1.34
Ca	1.00	1.00	1.00	1.00	1.00	1.01	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
Cl	1.00	1.00	0.99	1.00	1.00	1.00	0.99 dTW	1.00	0.99	1.00	0.99 tw	1.01	1.01	1.00	0.99 t	1.00
K	1.02	1.02	1.00	1.03	1.01	1.03	1.02	1.01	1.00	1.05 t	1.00	1.05 tw	1.02	1.01	1.00	1.03
Na	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01 w	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00
P	0.95 tw	0.97	0.94 tW	1.09	0.99	1.01	1.02	0.92	0.97	1.02	0.96	1.00	0.96	0.99	0.95 w	1.07
Urine Males	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
uVol	1.02	1.07	1.00	1.10	0.78 t	0.80 t	0.83 t	0.83	0.94	0.91	0.86 t	1.03	0.95	1.04	0.96	1.00
uVolW	1.03	1.08	1.01	1.13	0.78	0.81	0.83 t	0.86	0.94	0.92	0.83 t	1.08	0.93	1.03	0.94	1.04
uLeu	0.95	0.90	0.61	0.98	1.47 t	1.50	1.01	1.03	1.06	1.08	0.60 w	0.73	1.25	0.95	0.58	1.18
uOsmoll	0.94	0.92	1.01	0.96	1.20	1.21	1.18 t	1.07	1.07	1.10	1.14	0.92	1.02	0.96	1.08	1.04
uProtein	-	-	-	1.06	-	-	-	0.95	-	-	-	0.77	-	-	-	1.38
uHemogl	-	-	-	1.64	-	-	-	1.11	-	-	-	1.13	-	-	-	1.08
uKeton	1.14	0.92	1.03	1.20	1.79 t	1.21	1.68 dTw	1.29	1.13	0.92	0.79	1.22	1.19	0.76	1.08	1.76 t
upH	0.84	0.90	0.88	1.07	0.98	0.90	1.05	0.98	0.95	1.03	1.11	1.04	0.86	0.86	0.86	1.18
Organs Males	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Kidney	-	-	1.06 w	-	-	-	1.06	-	-	-	1.00	-	-	-	1.01	-
Spleen	-	-	1.04	-	-	-	1.09	-	-	-	0.99	-	-	-	1.02	-
Liver	-	-	1.00	-	-	-	1.01	-	-	-	1.00	-	-	-	0.96	-
AdrenGl	-	-	1.07 w	-	-	-	1.07	-	-	-	0.94	-	-	-	1.03	-
Heart	-	-	1.04	-	-	-	1.06	-	-	-	1.00	-	-	-	0.99	-
Testis	-	-	1.07	-	-	-	1.02	-	-	-	1.01	-	-	-	1.05	-
Epididymis	-	-	1.12 tw	-	-	-	1.03	-	-	-	0.95	-	-	-	1.03	-
Brain	-	-	1.06	-	-	-	1.02	-	-	-	0.97	-	-	-	1.01	-

Table 12 Estimated ratios of GM groups versus the non-GM control group for female rats for months 3, 6, 12 and 24. Significant ratios are marked, with red background colouring, as follows: D: P<0.01 by Dunnett-test, d: P<0.05 by Dunnett-test, T: P<0.01 by t-test but not by Dunnett-test, t: P<0.05 by t-test but not by Dunnett-test, W: P<0.01 by Wilcoxon signed rank test, w: P<0.05 by Wilcoxon signed rank test.

Weights Females	NK11- vs Control				NK33- vs Control				NK11+ vs Control				NK33+ vs Control			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Weight	0.99	0.99	0.98	0.95	0.99	0.99	1.00	0.93	0.99	0.99	0.99	0.95	0.99	1.00	0.99	0.98
growthRate	0.99	0.99	0.90	-	0.99	1.00	1.05	-	0.99	0.97	0.99	-	0.99	1.03	0.96	-
FeedMean	1.03 tw	1.00	0.98	1.00	1.01	1.01	1.01	0.98	1.01	1.01	0.98	0.98	1.04 DW	1.03	1.02	1.02
Haematology Females	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
WBC	1.06	1.04	1.06	1.11	1.03	1.03	0.93	0.99	0.97	0.96	0.97	1.26	1.08	1.03	1.05	1.11
RBC	1.00	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.00	1.03 w	0.98	1.02	1.01	1.02	1.02
HGB	0.99	0.99	1.00	0.98	1.01	0.99	1.00	0.98	0.99	0.99	1.01	0.99	1.00	1.00	1.01	0.99
HCT	1.00	1.00	1.01	0.98	1.01	1.00	1.02 w	0.97	1.00	1.00	1.02	0.99	1.01	1.00	1.02	0.99
MCV	1.00	1.00	1.01	1.01	0.99	0.99	1.00	0.99	0.99	1.00	0.99	1.00	0.99	0.99	0.99	0.99
MCH	1.00	0.99	1.00	1.01	0.99	0.99	0.99	0.99	0.98 w	0.98	0.98 tW	1.00	0.98	0.99	0.98 t	0.99
MCHC	1.00	0.99	0.99	1.00	1.00	0.99	0.99 tw	0.99	0.99 w	0.99	0.99 tw	0.99	0.99	1.00	0.99	1.00
PLT	0.96	0.94	1.03	0.98	0.95	0.97	1.01	1.08	0.98	0.93 t	1.01	1.10	1.02	0.97	1.03	1.12
LYMR	0.96	0.98	1.01	0.98	0.98	0.99	1.01	0.99	0.98	1.00	1.01	1.01	1.00	0.98	1.02	1.01
LYMA	1.03	1.03	1.03	1.21	1.01	1.02	0.94	0.97	0.95	0.96	0.98	1.05	1.10	1.01	1.05	1.10
diffWBC Females	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Lymphocytes	-	1.00	-	-	-	1.00	-	-	-	1.00	-	-	-	1.03	-	-
Neutrophils	-	1.01	-	-	-	1.03	-	-	-	1.02	-	-	-	0.97	-	-
Monocytes	-	0.76	-	-	-	0.88	-	-	-	0.94	-	-	-	0.67 dW	-	-
Eosinophils	-	0.98	-	-	-	0.85	-	-	-	1.07	-	-	-	0.86	-	-
ClinChem Females	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
ALP	1.03	0.93	0.98	1.03	1.03	0.92	1.02	1.10	1.09	0.94	1.01	1.09	1.03	0.93	0.94	1.05
ALT	1.01	0.99	0.95	0.96	0.98	1.03	0.99	0.98	0.92	0.88 tw	0.88	1.01	1.09	0.99	0.99	1.02
AST	1.00	0.96	0.99	0.95	1.08	1.06	1.04	1.08	0.99	0.92 w	0.97	1.06	1.14 dw	1.06	1.05	1.03
BIL	0.93	1.01	0.92	0.99	1.06	1.03	0.96	1.00	0.96	0.98	0.97	0.95	1.01	1.00	0.95	1.00

ALB	0.99	1.01	1.00	1.00	0.99	1.01	1.01	1.00	1.00	1.02	1.01	0.97	0.98	0.99	1.00	1.01
TP	0.99	1.00	1.00	1.02	0.99	1.01	1.00	1.03	1.00	1.01	1.00	0.99	0.99	0.99	1.00	1.02
Glu	0.99	0.96	1.01	1.01	0.95	0.95	0.97	0.93	1.01	0.98	1.04	0.92	0.95	0.98	1.03	0.93
CHOL	0.97	0.98	0.99	1.04	0.92	0.92	0.95	0.94	0.90 t	0.91	0.92	0.88	0.94	0.94	0.95	1.03
TAG	0.97	0.99	0.97	1.07	0.90 w	0.92	1.00	0.84	0.87 d	0.91 w	0.99	0.92	0.96	0.97	1.02	0.94
Crea	1.04	0.99	1.05	1.10	1.02	0.97	1.04	1.15 dT	1.05 tw	1.01	1.06 t	1.11 t	1.03	1.02	1.03	1.14 dTw
Urea	1.02	1.04	1.08 tW	1.02	1.01	1.02	1.11 Dw	1.09	1.02	1.00	1.10 dTw	1.04	1.06 w	1.03	1.07 tw	1.02
cHGB	0.96	1.05	0.88	0.91	1.14	1.10	0.98	0.93	0.88 w	1.02	0.96	0.89	1.03	0.97	0.97	0.97
Ca	1.00	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.00	1.01	1.01	1.01	1.00	1.00	1.00	1.01
Cl	1.01	1.01	1.01	0.98 tw	1.01 w	1.01	1.00	0.99	1.01 t	1.01	1.00	0.99	1.01	1.00	1.00	0.99
K	1.02	1.02	1.01	0.92 t	1.02	1.08 DW	1.01	0.99	1.00	1.03	1.01	1.02	1.03	1.00	1.01	1.00
Na	1.01	1.00	1.01 t	1.00	1.01 w	1.00	1.01 t	1.00	1.01 tw	1.00	1.00	1.00	1.01 tw	1.00	1.00	1.00
P	1.06	1.01	1.01	1.00	1.12 tW	1.04	1.08	1.11	0.97	1.00	1.05	1.10	1.07	0.96	0.98	0.97
Urine Females	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
uVol	0.94	0.89	0.90	1.02	1.20	1.17	1.00	1.07	0.89	1.02	1.01	0.92	0.85	0.78	0.93	1.16 w
uVolW	0.95	0.89	0.95	1.06	1.18	1.15	1.01	1.09	0.87	1.01	1.03	0.95	0.85	0.78 w	0.97	1.15
uLeu	0.91	1.23	1.04	1.22	1.06	0.95	1.50	0.97	1.06	0.85	0.91	1.30	0.90	1.00	1.12	0.83
uOsmoll	1.08	1.08	1.04	0.93	0.85	0.84	1.00	0.93	1.21	1.00	0.96	0.99	1.25	1.20	1.07	0.90
uProtein	-	-	-	1.99	-	-	-	1.61	-	-	-	1.61	-	-	-	1.12
uHemogl	-	-	-	0.69	-	-	-	0.96	-	-	-	1.12	-	-	-	0.69
uKeton	1.02	1.13	1.17	0.76	1.10	1.06	1.15	0.88	1.43 t	0.91	0.84	1.22	1.16	0.91	1.12	0.76
upH	0.75	0.88	0.89	1.34	1.03	1.16	0.92	1.54 t	0.86	0.95	1.18	1.22	0.98	0.93	0.89	1.18
Organs Females	**NK11- vs Control**				**NK33- vs Control**				**NK11+ vs Control**				**NK33+ vs Control**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Kidney	-	-	1.07	-	-	-	1.05	-	-	-	1.08 tw	-	-	-	1.08 tw	-
Spleen	-	-	1.09	-	-	-	0.98	-	-	-	1.04	-	-	-	1.11	-
Liver	-	-	1.02	-	-	-	1.01	-	-	-	1.03	-	-	-	1.03	-
AdrenGl	-	-	1.03	-	-	-	1.01	-	-	-	1.04	-	-	-	1.06	-
Heart	-	-	1.06 w	-	-	-	1.08 t	-	-	-	1.04	-	-	-	1.07	-
Uterus	-	-	1.05	-	-	-	0.90	-	-	-	1.00	-	-	-	0.99	-
Ovary	-	-	1.11	-	-	-	1.20	-	-	-	1.18	-	-	-	1.23	-
Brain	-	-	1.10 dTW	-	-	-	1.08 t	-	-	-	1.09 dw	-	-	-	1.10 dTW	-

4.6 Standardised effect sizes

4.6.1 Method

EFSA (2011b) defines the standardised effect size (SES) as the effect size measured in SD units, where SD is the standard deviation among experimental units. We will assume that in a randomised block experiment, like the current G-TwYST study, SD refers to the pooled residual variation. The use of SES in EFSA (2011b) was in the context of determination of sample size and power: '*If experience from previous toxicity tests shows an effect size of, say, one SD or less is of little toxicological relevance then this can be used to determine sample size in new situations*' (EFSA 2011b). Zeljenková *et al* (2014) followed this example and, without further toxicological motivation, '*assumed that an SES of 1.0 SD or less is unlikely to be of toxicological importance*'. Consequently, all results of the GRACE studies have been reported as confidence intervals on the SES scale (Zeljenková *et al*, 2014, 2016, Schmidt *et al* 2015, 2017). In this section the same SES graphs are calculated for comparability between GRACE and G-TwYST.

Standardized effect sizes (SES), and their exact 95% confidence intervals, were calculated by means of the conf.limits.nct() function in the MBESS R-package, see section 3 in Kelley (2007). Note that, since the calculated SES confidence interval is exact, the SES interval does not contain zero if and only if the p-value of the corresponding t-test is smaller than 5%.

4.6.2 Results

For the results see the companion reports (Goedhart & van der Voet 2018abcd). The number of intervals that extend outside the -/+ 1 SD limits equals 106 out of 288 (37%) for the 3-months data, 102 out of 320 (32%) for the 6-months data, 141 out of 352 (40%) for the 12-months data and 72 out of 296 (24%) for the 24-months data.

4.7 Factorial analysis

4.7.1 Method

The purpose of oral toxicity study A in the EU project G-TwYST was to assess the effects of genetically modified (GM) maize NK 603, grown both with and without the use of RoundUp, when fed to rats for a period of up to two years at incorporation rates of 11% and 33% in the feed. Table 13 lists the maize type and incorporation rate of the 5 diets.

Table 13 Diets used in the 2-year feeding trial study A with GM inclusion rates 11% and 33%.

Group	Isogenic maize (% of diet)	NK603 only (% of diet)	NK603 + Roundup (% of diet)
Control	33	0	0
NK11-	22	11	0
NK33-	0	33	0
NK11+	22	0	11
NK33+	0	0	33

The structure of diets is a 2 by 2 factorial design for the GM feeding groups with factors GM inclusion rate (*IR*, 11% or 33%) and use of Roundup (*RU*, - or +), with an added control for the non-GM control

group. This structure allows a more sensitive analysis, integrating over the five dose group, according to the model:

$y_{0k} = \mu + \delta_k + \epsilon_{ijk}$ for data in the Control (non-GM) group ($i = 0$)

$y_{ijk} = \mu + GM + IR_i + RU_j + int_{ij} + \delta_k + \epsilon_{ijk}$ for data in GM groups, $i = 1,2; j = 1,2$.

In this model the stochastic terms are δ_k for block effects and ϵ_{ijk} for residual effects. The fixed term GM models the difference between the four GM groups (averaged) and the control. This term can only be interpreted if the other three fixed terms can be assumed to be zero. The main effect $IR_2 - IR_1$ models the difference between the groups with inclusion rates 33% and 11%, and similarly, the main effect $RU_2 - RU_1$ models the difference between the groups with and without Roundup. These main effects are only useful when there is no interaction, modelled by int_{ij}, between GM inclusion rate and Roundup.

4.7.2 Results

This section summarises the main findings reported in the companion reports (Goedhart & van der Voet 2018abcd). Estimated ratios of factorial effects are shown in Table 15 and Table 16, for male and female rats, respectively. The frequencies of the significant results among all 1192 tests are shown in Table 14. RoundUp effects at 3 and 12 months, as well as interactions at 6 and 12 months were seen in more than 5% of tests.

Table 14 Absolute and relative frequencies of significant results (P<0.05) in testing factorial effects. Males and females combined. Frequencies significantly higher (P<0.05) than the nominal test level of 5% by an exact binomial test are shown with a red background.

	3 months	6 months	12 months	24 months	all
GM vs. non-GM	3/72 (4%)	1/80 (1%)	8/88 (9%)	1/74 (1%)	13/314 (4%)
GM incl. rate	7/67 (10%)	2/70 (3%)	6/77 (8%)	3/68 (4%)	18/282 (6%)
RoundUp	8/67 (11%)	3/70 (4%)	11/77 (14%)	0/68 (0%)	22/282 (8%)
interaction	5/72 (5%)	10/80 (12%)	11/88 (12%)	6/74 (8%)	32/314 (10%)
all	23/278 (8%)	16/300 (5%)	36/330 (11%)	10/284 (4%)	85/1192 (7%)

Table 15 Estimated ratios of factorial effects for male rats. Only ratios for significant effects (P<0.05) are shown with a red background. Cells containing 'x' refer to non-significant effects, while cells containing '-' indicate no data or no analysis.

Males Weights	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Weight	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.99 / 0.96 / 1.04
growthRate	x	x	x	x	x	x	1.10	x	x	x	1.08	x	1 / 1.01 / 1.01 / 1.01	x	x	x
FeedMean	x	x	x	x	x	x	1.03	x	x	1.02	1.03	x	1 / 0.98 / 0.99 / 1.02	x	x	x
Males Haematology	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
WBC	x	x	x	x	x	x	1.12	x	x	x	x	x	x	1 / 1.24 / 1.20 / 1.10	x	x
RBC	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
HGB	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
HCT	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MCV	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MCH	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MCHC	x	x	x	x	x	x	x	x	0.99	x	x	x	x	x	x	x
PLT	x	x	x	x	x	x	1.07	x	x	x	x	x	x	x	x	x
LYMR	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
LYMA	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 1.24 / 1.23 / 1.14	1 / 1.20 / 1.08 / 1.07	x
Males diffWBC	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Lymphocytes	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Neutrophils	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Monocytes	-	0.82	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Eosinophils	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Males ClinChem	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
ALP	x	x	x	x	x	x	x	0.83	x	x	x	x	x	x	x	x
ALT	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
AST	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
BIL	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ALB	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 1.06 / 1.05 / 1.01

TP	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Glu	x	x	x	x	x	x	x	x	1.06	1.05	x	x	x	x	x	x
CHOL	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
TAG	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Crea	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Urea	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
cHGB	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Ca	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.98 / 0.98 / 1.00
Cl	x	x	0.99	x	x	x	x	x	x	x	x	x	x	x	x	x
K	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Na	x	x	x	x	1	x	x	x	x	x	x	x	x	x	x	x
P	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 1.08 / 1.02 / 1.01	1 / 0.85 / 0.92 / 0.99
Males Urine	**GM vs. non GM**				**33% vs. 11%**				**RU+ vs. RU-**				**Interaction (NK11- / NK33- / NK11+ / NK33+)**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
uVol	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.75 / 0.85 / 0.97	1 / 0.83 / 0.85 / 0.96	x
uVolW	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.75 / 0.85 / 0.95	1 / 0.83 / 0.83 / 0.93	x
uLeu	x	x	x	x	1.35	x	x	1.30	x	x	x	x	x	x	x	x
uOsmoll	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 1.32 / 1.19 / 1.05	1 / 1.17 / 1.13 / 1.07	x
uProtein	-	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x
uHemogl	-	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x
uKeton	x	x	x	x	x	x	1.49	x	x	x	0.70	x	x	x	x	x
upH	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 1.20 / 1.26 / 0.98	x
Males Organs	**GM vs. non GM**				**33% vs. 11%**				**RU+ vs. RU-**				**Interaction (NK11- / NK33- / NK11+ / NK33+)**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Kidney	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Spleen	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Liver	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
AdrenGl	-	-	x	-	-	-	x	-	-	-	0.92	-	-	-	x	-
Heart	-	-	x	-	-	-	x	-	-	-	0.95	-	-	-	x	-
Testis	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Epididymis	-	-	x	-	-	-	x	-	-	-	x	-	-	-	1 / 0.92 / 0.85 / 0.92	-
Brain	-	-	x	-	-	-	x	-	-	-	0.95	-	-	-	x	-

Table 16 Estimated ratios of factorial effects for female rats. Only ratios for significant effects (P<0.05) are shown with a red background. Cells containing 'x' refer to non-significant effects, while cells containing '-' indicate no data or no analysis.

Females Weights	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Weight	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
growthRate	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
FeedMean	1.02	x	x	x	x	x	1.03	x	x	x	x	x	1 / 0.98 / 0.98 / 1.01	x	x	x
Females Haematology	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
WBC	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.88 / 0.91 / 0.99	x
RBC	x	x	x	x	x	x	x	x	x	x	1.02	x	x	x	x	x
HGB	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
HCT	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
MCV	x	x	x	x	x	x	x	0.98	x	x	0.99	x	x	x	x	x
MCH	x	x	x	x	x	x	x	x	0.99	x	0.99	x	x	x	x	x
MCHC	x	x	0.99	x	x	x	x	x	0.99	x	x	x	x	x	x	x
PLT	x	x	x	x	x	x	x	x	1.04	x	x	x	x	x	x	x
LYMR	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
LYMA	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.88 / 0.91 / 0.99	x
Females diffWBC	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Lymphocytes	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Neutrophils	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Monocytes	-	x	-	-	-	x	-	-	-	x	-	-	-	1 / 1.16 / 1.22 / 0.88	-	-
Eosinophils	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-	-
Females ClinChem	GM vs. non GM				33% vs. 11%				RU+ vs. RU-				Interaction (NK11- / NK33- / NK11+ / NK33+)			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
ALP	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ALT	x	x	x	x	x	1.08	x	x	x	0.93	x	x	1 / 0.96 / 0.91 / 1.08	x	x	x
AST	x	x	x	x	1.12	1.13	x	x	x	x	x	x	x	x	x	x
BIL	x	x	x	x	1.09	x	x	x	x	x	x	x	x	x	x	x
ALB	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

TP	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Glu	x	x	x	x	x	x	x	x	x	x	1.05	x	x	x	x	x
CHOL	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
TAG	x	x	x	x	x	x	x	x	x	x	x	x	1 / 0.93 / 0.89 / 0.99	x	x	x
Crea	x	x	1.04	1.12	x	x	x	x	x	x	x	x	x	x	x	x
Urea	x	x	1.09	x	x	x	x	x	x	x	x	x	x	x	x	x
cHGB	x	x	x	x	1.19	x	x	x	x	x	x	x	x	x	x	x
Ca	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Cl	1.01	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
K	x	x	x	x	x	x	x	x	x	x	x	x	x	1 / 1.06 / 1.02 / 0.99	x	1 / 1.08 / 1.12 / 1.09
Na	1.01	x	1.01	x	x	x	x	x	x	x	x	x	x	x	x	x
P	x	x	x	x	1.08	x	x	x	x	x	x	x	x	x	1 / 1.03 / 1.32 / 0.99	1 / 1.11 / 1.10 / 0.97
Females Urine	**GM vs. non GM**				**33% vs. 11%**				**RU+ vs. RU-**				**Interaction (NK11- / NK33- / NK11+ / NK33+)**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
uVol	x	x	x	x	x	x	x	x	0.82	x	x	x	x	1 / 1.32 / 1.15 / 0.88	x	x
uVolW	x	x	x	x	x	x	x	x	0.81	x	x	x	x	1 / 1.30 / 1.14 / 0.88	x	x
uLeu	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
uOsmoll	x	x	x	x	x	x	x	x	1.29	x	x	x	x	1 / 0.78 / 0.93 / 1.11	x	x
uProtein	-	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x
uHemogl	-	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x
uKeton	x	x	x	x	x	x	x	x	x	x	0.84	x	x	x	x	x
upH	x	x	x	x	1.24	x	x	x	x	x	x	x	x	x	1 / 1.03 / 1.32 / 0.99	x
Females Organs	**GM vs. non GM**				**33% vs. 11%**				**RU+ vs. RU-**				**Interaction (NK11- / NK33- / NK11+ / NK33+)**			
	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
Kidney	-	-	1.07	-	-	-	x	-	-	-	x	-	-	-	x	-
Spleen	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Liver	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
AdrenGl	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Heart	-	-	1.06	-	-	-	x	-	-	-	x	-	-	-	x	-
Uterus	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Ovary	-	-	x	-	-	-	x	-	-	-	x	-	-	-	x	-
Brain	-	-	1.09	-	-	-	x	-	-	-	x	-	-	-	x	-

4.8 Correlation analysis

4.8.1 Method

For single variables the difference between a GM feed group and the control group is quantified by the ratio of the responses. These can then be compared to given limits (as in Figure 13 - Figure 16) or to limits calculated from historical data (as in Figure 9 **-** Figure 12).

For a toxicological interpretation it may be helpful to see results for variables that simultaneously relate to the same pathological endpoint. Bivariate plots were prepared showing the patterns for each pair for three variables related to liver disorder (relative liver weight, ALP, CHOL) and three variables related to kidney disorder (relative kidney weight, Urea, Crea). The ten points in each graph are based on the cage means in the ten blocks of the study. This analysis was only performed for the 12 months data, which was the only time point where organ weights have been measured.

For comparison, the proposed target effect sizes of Hong *et al* (2017) are included in the plots as horizontal and vertical lines (together with lines at ratio 1 for reference). It can be noted that similar plots could have been made using the equivalence limit scaled differences (ELSDs) as presented in Figure 9 **-** Figure 12**.**

4.8.2 Results

This section summarises the main findings reported in the companion report (Goedhart & van der Voet 2018c).

The correlation plots for three liver-related and three kidney-related variables in males and females are shown in Figure 18 - Figure 21. For males there were significant positive correlations between Urea and Crea for all comparisons (Figure 19), without accompanying correlations with the relative Kidney weights. Further significant (positive) correlations were found between ALP and CHOL in males, and between Kidney and Urea in females. Simultaneous exceedance of the threshold for two variables was only observed for Kidney and Urea for a single cage in females; the accompanying Crea value was close to the threshold for this cage.

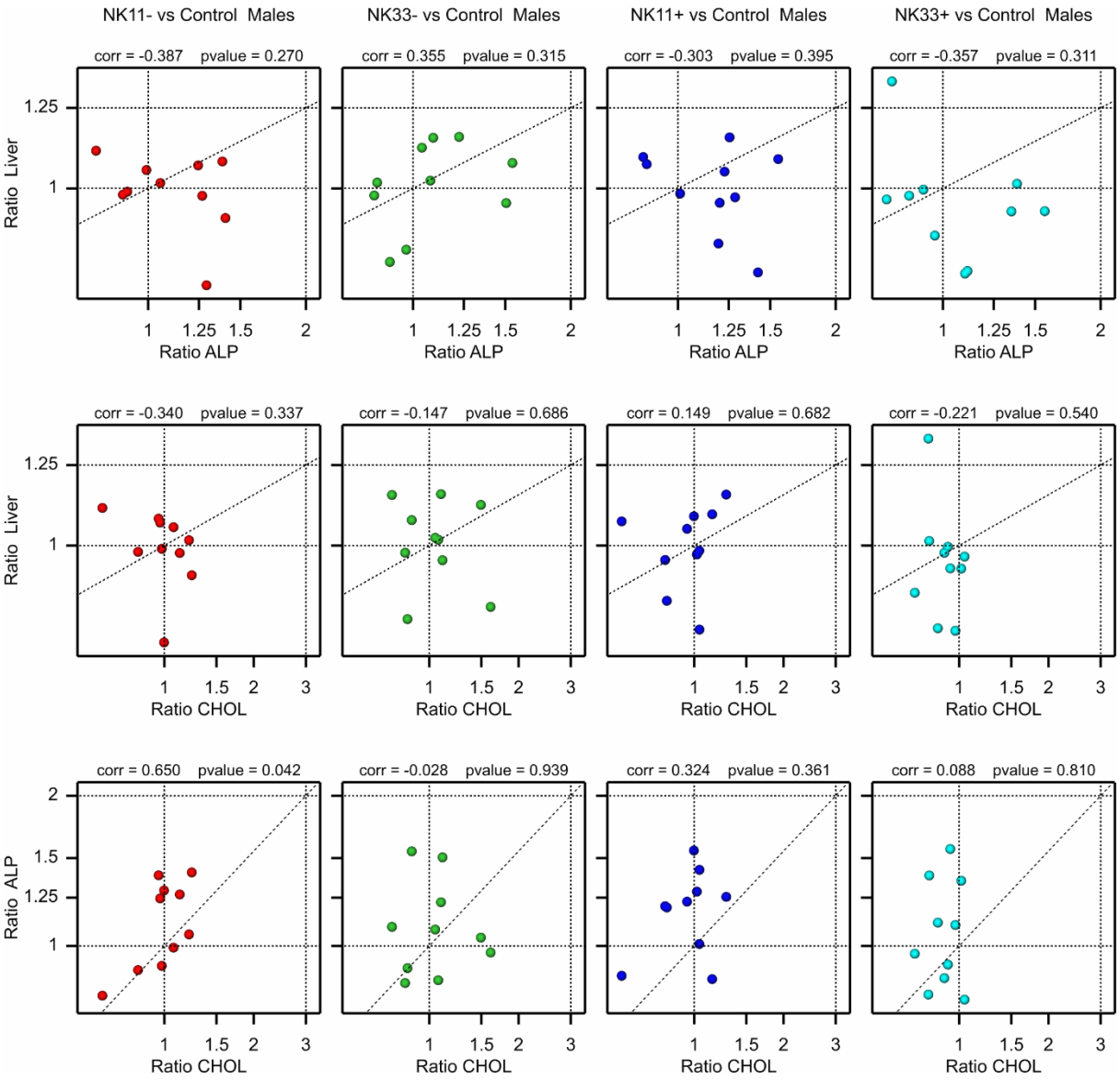

Figure 18 Pairwise results for variables with set target effect sizes related to liver damage in Males after 12 months. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in ten blocks. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong et al. (2017).

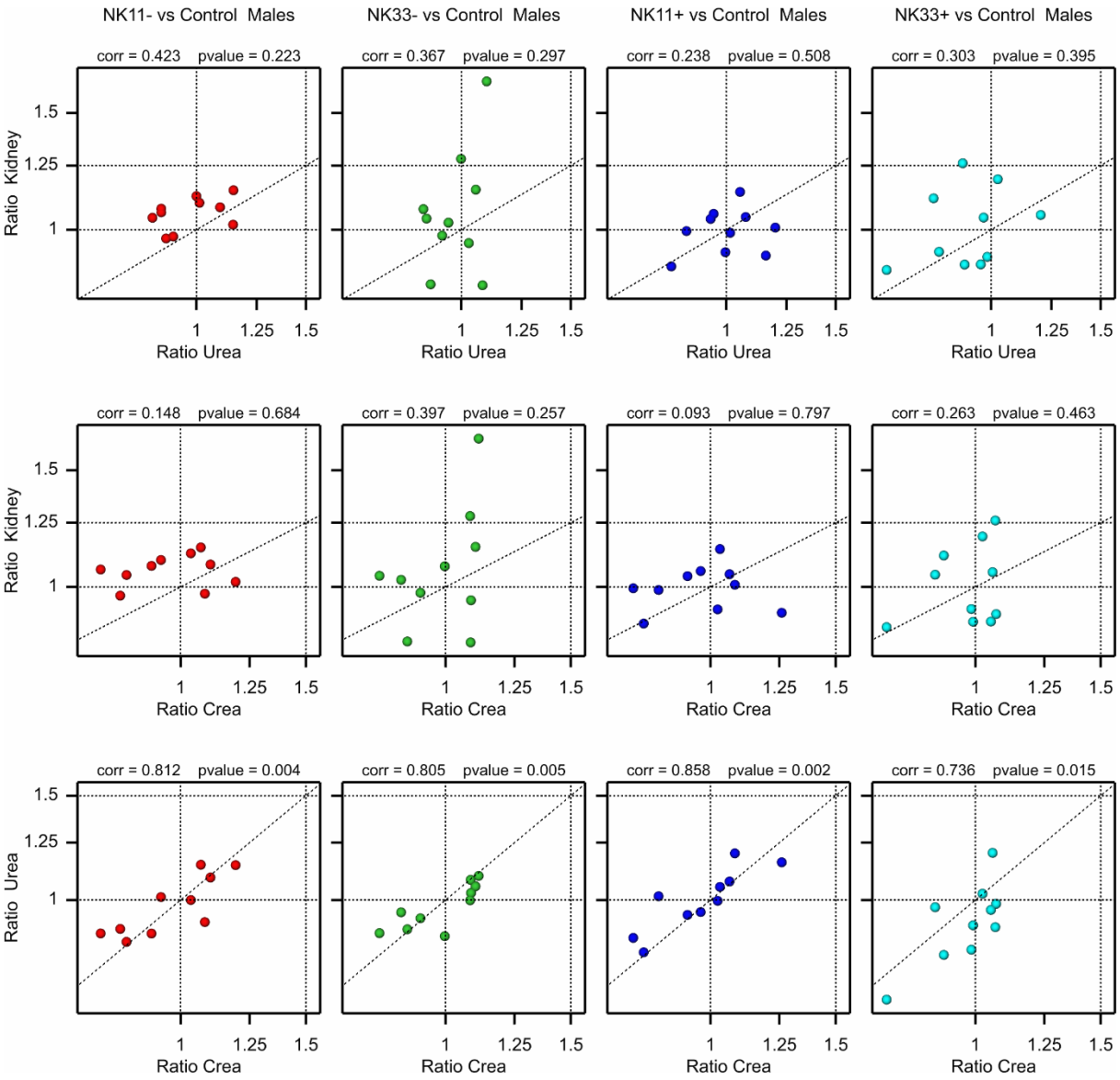

Figure 19 Pairwise results for variables with set target effect sizes related to kidney damage in Males after 12 months. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in ten blocks. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong et al. (2017).

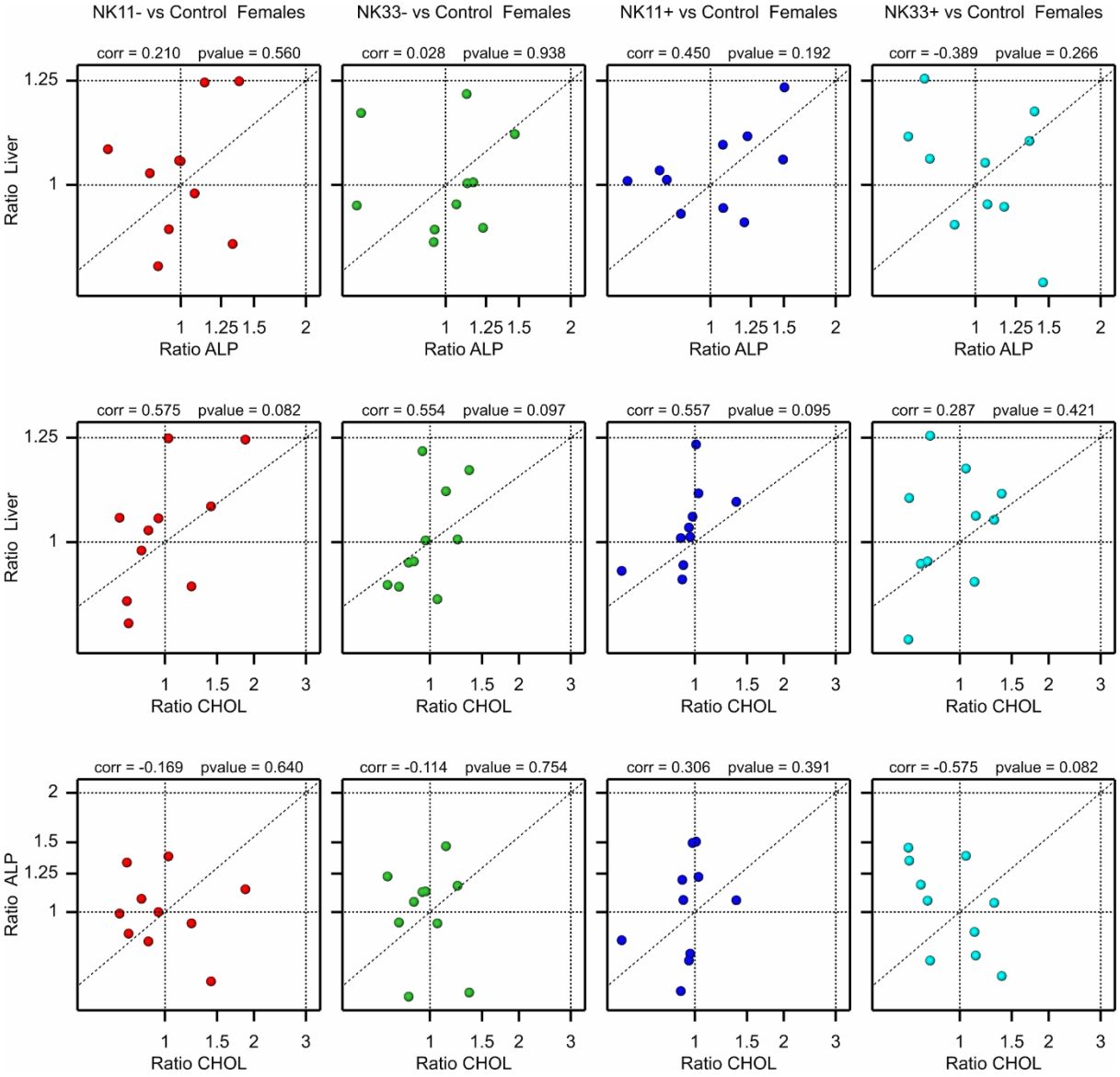

Figure 20 Pairwise results for variables with set target effect sizes related to liver damage in Females after 12 months. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in ten blocks. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong et al. (2017).

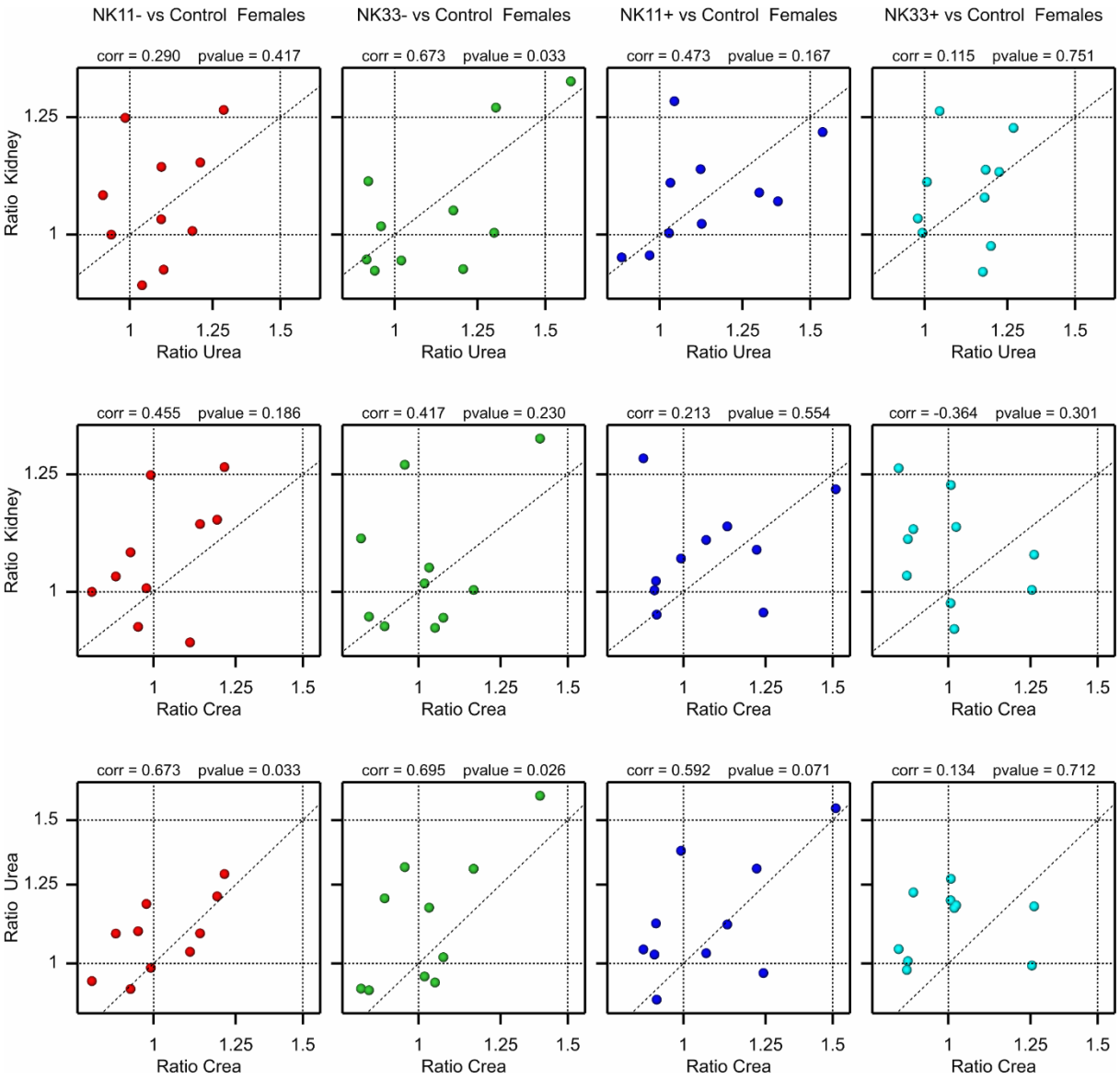

Figure 21 Pairwise results for variables with set target effect sizes related to kidney damage in Females after 12 months. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in ten blocks. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong et al. (2017).

5 Summary and discussion

In this report the data from G-TwYST study A have been analysed following eight approaches. For comparisons between a GM feeding group and the control feed for a single variable, these approaches were two forms of analysing survival/mortality (4.1, 4.2), two forms of equivalence analysis (4.3 4.4), the traditional approach focusing on significant differences (4.5), and the standardised effect size (SES) approach (4.6). In addition, a factorial analysis (4.7, also 4.1, 4.2) and a correlation analysis (4.8) allowed some limited forms of integration over dose groups or variables, respectively.

Survival analysis (4.1) showed few differences between the five feeding groups. The mortality rate at 24 months (4.2) was increased for male animals in the NK33+ group relative to the Control group. In a factorial analysis, there was an indication that RoundUp could have increased the hazard and the mortality rate at 24 months for the males. It should be noted that 3 significant results (with P-values 0.03, 0.03 and 0.04, respectively) found in 48 comparisons amount to 6%, which is close to the nominal level of the tests (5%).

Among the two forms of equivalence analysis, the approach with given external equivalence limits (4.4) is the simpler one. It could be preferred if toxicologists were able to set external equivalence limits for all relevant variables based on their expert knowledge. In the current report it was applied to nine variables, for which Hong *et al* (2017) recently proposed targeted effect sizes. Obviously, the uncertainty in setting these targeted effect sizes is not accounted for in the equivalence analysis using these fixed limits.

However, external equivalence limits are often not available, and toxicologists notice many uncertainties about the impact of toxicological effects. Moreover, they find it often difficult to come to a conclusion on such equivalence limits. For such cases, the equivalence analysis which bases equivalence on normal variation in historical non-GM data (4.1) may be an attractive alternative. This approach assumes that test facilities perform whole-food studies with rodents on a routine basis, such that variations between non-GM foods and between experimental units which are seen in historical studies have a relevance for the current study. In the current report, the approach could be applied to 27 variables measured at 3 months and 30 variables measured at 6 months. These variables were also observed in five preceding studies at 3 months in the same test facility in the GRACE project. Van der Voet *et al* (2017) discuss this new method which was developed in the G-TwYST project. In the results from study A it appeared that the residual variance within the feeding groups was generally increasing after 6 months. For this reason the equivalence testing method against historical data could not be applied with success for the data obtained at 12 months and 24 months.

For the data at 3 months and 6 months, and using the approach based on the historical GRACE data given tentative settings for regulatory parameters, equivalence was established in 99.8% of cases (411/412) as long as the residual variance in the current study was not larger than 150% of the residual variance in the historical studies. The regulatory parameters are the testing level (set at 5%) for all approaches, and two parameters for the new G-TwYST equivalence approach: the desired power (set at 95%) and the minimum sample size per group (set equal to the sample sizes in the current study, i.e. 35 for the weight variables and 20 for haematology, differential white blood cell

counts and clinical biochemistry). Note that test results could be different if these parameters were chosen differently.

For the 44 cases where the variance in the current study was more than 150% of that in the historical study, the conclusion must be that these discrepancies must be solved first before this approach to equivalence testing would be feasible. Either, the precision was insufficient in the current study or the precision was under-estimated in the historical study, and better reference data should be gathered. This is obviously true at 12 and 24 months, for which currently no historical reference data from the same test facility were available. If on the other hand it would be thought that the lower precision in the current study has to be accepted, then the study should be repeated with higher numbers of experimental units.

Using the alternative equivalence testing method using fixed target effect sizes, equivalence was established in 94% of cases (211/224), with the large residual variation in body weight, WBC and LYMA at 24 months being responsible for the negative results.

In all cases where equivalence was not established by means of the statistical test and the variance ratio was below 1.5, the median estimate was still in the equivalence region. Therefore, in the terminology of EFSA (2011a), these cases are still classified as 'equivalence more likely than lack of equivalence'.

As noted by EFSA (2011b), separate analysis of many endpoints, most of which are not expected to differ between treatment groups, results in a large number of statistical tests. This will lead to the issue of multiple testing (multiplicity). As long as the variance in the current study was not larger than 150% of the variance in the historical studies, there was only one non-significant equivalence test and therefore the rate of non-equivalence was lower than the nominal level of the tests (which was set at 5%). In this report, we have not tried to adapt equivalence tests for multiplicity. However, it should be pointed out that a recently proposed approach to adjusting for multiplicity based on the False Discovery Rate (FDR) is not appropriate. Hong *et al* (2017) used adjusted p values using the FDR method for multiplicity adjustment. This means that effectively most p values are much larger (indicating less significant differences) than in a standard unadjusted analysis. This may explain why they report that 'no treatment-related differences were observed', although there were some 150 continuous endpoints in total. This complete absence of statistically significant differences is very much at odds with what is commonly found (e.g. in the GRACE, G-TwYST, and GMO90+ studies). Indeed, the absence of significant differences in Hong *et al* (2017) could be a direct consequence of using the FDR adjustment. It is doubtful whether the use of the FDR-correction makes sense in food safety testing (EFSA 2010, van der Voet 2018). It controls false discoveries, and is therefore connected to difference testing, where false positives are considered as error of the first kind. i.e. one wants to have a small probability of erroneously reporting a difference. In the context of equivalence testing the statistical hypotheses are reversed, and false negatives are the error of the first kind, i.e. one wants to have a small probability of erroneously reporting equivalence. Consequently, the FDR concept is addressing the wrong type of error.

Classical analysis following OECD guidance is only focusing on finding differences, not equivalences. Only for the data at 12 months the rate of significant results was higher than the nominal 5% significance level, i.e. 11% when the t-test was applied. The scheme advocated by OECD contains several adaptations. First, a multiplicity correction by using Dunnett's test rather than the straight-

forward t-test is proposed for the fact that four groups are compared at the same time to the control group. Thus, applying Dunnett's test the number of significant cases at the 5% significance level was reduced to 3% of all comparisons for 12 months data, i.e. similar to the nominal error level). However, a multiplicity correction may be wrong for the same reason why the FDR method was wrong: if we are primarily interested in safety and equivalence, then the roles of the statistical hypotheses are reversed, and corrections as used in Dunnett's test address the wrong type of error.

In this work confidence intervals were also expressed and plotted as Standardised Effect Size (SES), see EFSA (2011b), in order to allow a comparison with SES results for the preceding GRACE project (Schmidt and Schmidtke 2014, Schmidt *et al* 2015ab, 2016, 2017, Zeljenková *et al* 2014, 2016). SES, also known as Cohen's d, is often used in meta-analyses to show the results of different variables in the same plot. Reporting and graphically displaying effect sizes was described in Schmidt *et al* (2016) as a way 'to avoid the yes/no decision trap of statistical tests and to illustrate the size of effects in the context of biological relevance'. However, in the absence of clear limit values for biological relevance, these authors had to build on the arbitrary EFSA example, where effects of ± 1 SD were assumed to be unlikely to be of toxicological importance. Schmidt *et al* (2016) already concluded that the pooled standard deviation SD of individual measurements 'is a priori not expected to be directly related to biological relevance', and Schmidt *et al* (2017) warned that 'it should therefore be kept in mind that future decisions on relevant equivalence limits may influence the equivalence results'. The results of the current G-TwYST study, where 97% of all intervals extended outside the ± 1 SD limits, confirms the pattern observed in GRACE. Whereas, displaying the confidence intervals indeed gives a richer view on the results than just reporting yes/no decisions, the scale of the SES plots does not seem the best choice for equivalence assessments. As Hong *et al* (2017) remark, the value of SES to support data interpretation is limited. Alternatively scaled effect sizes, such as those presented in section 4.3 can be preferred, because the scaling factor (the equivalence limit) is based on data analysis of in this case historical data, rather than being an arbitrary value. It can be noted, however, that this approach was not available for the GRACE project, because of lack of historical data in the same test facility.

Factorial analyses for single variables allowed to consider effects pooled over more than two groups, thus providing more powerful tests for main effects in the absence of interaction. However, this approach was in the current work restricted to the testing of differences. In principle, it could be further developed for the equivalence tests.

Most statistical analyses in this report have considered variables one by one, collecting the results only in a joint table or plot for ease of interpretation. However, toxicologists often stress that effects should be judged together. Wherever a prior hypothesis exists that links multiple variables, these may sometimes be translated in a function of those variables. For example, there is a biological connection between the pancreas and the regulation of glucose, which leads to a prior expectation of a negative correlation between pancreas weight and serum glucose. It may then be sensible to perform difference and equivalence testing for an additional variable such as the ratio or log-ratio of these variables. Such ratios have not been defined in the current study.

Another tool to study variables together is pairwise plotting of results per experimental unit. In section 4.8 this was done for three variables related to liver damage and three variables related to kidney damage. The rationale was that correlations between variables would show up in these plots,

but this was hardly observed in these cases. To assist in the interpretation, the effects were plotted together with proposed target effect sizes. Most effects were below these limits also at the cage level (as was already observed for the means in section 4.4). We may conclude that correlations between these variables related to the same organ are nevertheless not prominent as long as the effects are within the targeted range. Of course, correlations could be (and are expected to be) more evident for effect sizes that would exceed the limits by large amounts.

A more detailed approach to testing than reported here would also be possible based on a more detailed consultation with toxicologists. For example, nephrotoxic effects can lead to decreased or increased kidney weights. However, in both of these cases, the toxicologists would expect to see increased urea (Urea) and/or creatinine (Crea) levels. In addition, there might be a decreased level of glucose in the urine (Glu) or an increased level of amino acids, but these effects are less predictable. Increases in Urea or Crea may indicate nephrotoxic effects that are not yet discernible as deviating kidney weights. It is concluded that increased Urea and/or Crea levels are the primary indicators of kidney damage, and only increased levels represent a toxicological concern. Considering observed normal ranges, an increase by 50% in at least one of the two key variables could be seen as potentially concerning, and provide a level to be used as equivalence limit. Specific hypotheses to be tested for the differences Δ (on the log scale) between the treatment groups (GM vs. comparator) would then be as follows.

Difference tests:

$$H_0\colon\ E(\Delta_{Urea}) = 0 \qquad \text{vs.} \qquad H_1\colon\ E(\Delta_{Urea}) > 0$$

$$H_0\colon\ E(\Delta_{Crea}) = 0 \qquad \text{vs.} \qquad H_1\colon\ E(\Delta_{Crea}) > 0$$

Equivalence tests:

$$H_0\colon\ E(\Delta_{Urea}) = EL_{Urea} \quad \text{vs} \quad H_1\colon\ E(\Delta_{Urea}) < EL_{Urea}$$

$$H_0\colon\ E(\Delta_{Crea}) = EL_{Crea} \quad \text{vs.} \quad H_1\colon\ E(\Delta_{Crea}) < EL_{Crea}$$

where $EL_{Urea} = EL_{Crea} = \log(1.5)$.

In cases when a difference is found or an equivalence cannot be shown, the other variables (kidney weight, Glu, amino acids) may provide further interpretation to the toxicologist. These variables are therefore considered as secondary: the results can be summarised in terms of absolute values and confidence intervals for Δ (also shown graphically), but they would not be part of the testing framework based on primary variables. However, fine-tuning of statistical analyses as suggested here demands a large investment of time from both toxicologists and statisticians, and it will be very difficult to perform such exercises across the whole spectrum of endpoints.

References

EFSA (2010). Statistical considerations for the safety evaluation of GMOs. EFSA Journal 2010, 8, 1250. http://dx.doi.org/10.2903/j.efsa.2010.1250.

EFSA (2011a). Scientific Opinion on Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal 2011;9(5): 2150. [37 pp.] http://dx.doi.org/10.2903/j.efsa.2011.2150.

EFSA (2011b). EFSA guidance on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed. EFSA Journal 2011;9(12):2438 [21 pp.] doi:10.2903/j.efsa.2011.2438.

EFSA (2013). Considerations on the applicability of OECD TG 453 to whole food/feed testing. EFSA Journal 2013;11(7):3347, 18 pp. doi:10.2903/j.efsa.2013.3347

EFSA (2014). Explanatory statement for the applicability of the Guidance of the EFSA Scientific Committee on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed for GMO risk assessment. EFSA Journal 2014;12(10):3871, 25 pp., doi:10.2903/j.efsa.2014.3871

Goedhart, P.W. & van der Voet, H. (2017). G TwYST Study B. a 90-day toxicity study in rats fed GM maize NK603. Statistical report. Report 31.10.17, Biometris, Wageningen, The Netherlands.

Goedhart, P.W. & van der Voet, H. (2018). G TwYST Study A. Combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603. Statistical report, 3 months. Report 33.02.18, Biometris, Wageningen, The Netherlands.

Goedhart, P.W. & van der Voet, H. (2018). G TwYST Study A. Combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603. Statistical report, 6 months. Report 34.02.18, Biometris, Wageningen, The Netherlands.

Goedhart, P.W. & van der Voet, H. (2018). G TwYST Study A. Combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603. Statistical report, 12 months. Report 35.02.18, Biometris, Wageningen, The Netherlands.

Goedhart, P.W. & van der Voet, H. (2018). G TwYST Study A. Combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603. Statistical report, 24 months. Report 36.02.18, Biometris, Wageningen, The Netherlands.

Hong B, Du Y, Mukerji P, Roper JM, Appenzeller LM (2017). Safety assessment of food and feed from GM crops in Europe: Evaluating EFSA's alternative framework for the rat 90-day feeding study. Journal of Agricultural and Food Chemistry, 65(27): 5545-5560. http://dx.doi.org/10.1021/acs.jafc.7b01492

Kelly K (2007). Confidence intervals for Standardized Effect Sizes: Theory, Application and Implementation. Journal of Statistical Software, 20(8).

McCullagh, P. & Nelder, J.A. (1989). Generalized linear models. Chapman and Hall. London.

OECD (1998). Test Guideline 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264070707-en

OECD (2014). Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453: Second edition, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264221475-en

Schmidt K, Schmidtke J (2014). Statistical analysis report project GRACE, Statistical Analysis of A) 90-day feeding study in rats with Monsanto MON810 maize B) 90-day feeding study in rats with Pioneer MON810 maize. Version 2.01. https://www.cadima.info/index.php/area/selectStudy/3

Schmidt K, Schmidtke J, Schmidt P (2015a) Statistical analysis report project GRACE, Statistical analysis of chronic toxicity (1-year) study of rats with MON810 maize. Version 2.02. https://www.cadima.info/index.php/area/selectStudy/4

Schmidt K, Schmidtke J, Schmidt P (2015b) Statistical analysis report project GRACE, Statistical analysis of repeated dose 90-day oral toxicity/longitudinal study in rodents with MON810 maize. Version 2.00. https://www.cadima.info/index.php/area/selectStudy/5

Schmidt K, Schmidtke J, Kohl C, Wilhelm R, Schiemann J, van der Voet H, Steinberg P (2016). Enhancing the interpretation of statistical P values in toxicology studies – implementation of linear mixed models (LMMs) and standardized effect sizes (SESs). Archives of Toxicology, 90: 731–751. http://dx.doi.org/10.1007/s00204-015-1487-8

Schmidt K, Schmidtke J, Schmidt P, Kohl C, Wilhelm R, Schiemann J, van der Voet H, Steinberg P (2017). Variability of control data and relevance of observed group differences in five oral toxicity studies with genetically modified maize MON810 in rats. Archives of Toxicology, 91(4): 1977-2006. https://dx.doi.org/10.1007/s00204-016-1857-x

van der Voet H (2018). Safety assessments and multiplicity adjustment: comments on a recent paper. J. Agric. Food Chem. http://dx.doi.org/10.1021/acs.jafc.7b03686

van der Voet H, Goedhart PW, Schmidt K (2017). Equivalence testing using existing reference data: an example with genetically modified and conventional crops in animal feeding studies. Food and Chemical Toxicology, 109, 472-485. https://doi.org/10.1016/j.fct.2017.09.044

VSN International (2015). Genstat for Windows 18th Edition. VSN International, Hemel Hempstead, UK. www.Genstat.co.uk

Williams, D.A. (1982). Extra-binomial variation in logistic linear model. Applied Statistics, 31, 144-148.

Zeljenková D, Ambrušová K, Bartušová M, Kebis A, Kovrižnych J, Krivošíková Z, Kuricová M, Líšková A, Rollerová E, Spustová V, Szabová E, Tulinská J, Wimmerová S, Levkut M, Révajová V, Ševčíková Z, Schmidt K, Schmidtke J, La Paz JL, Corujo M, Pla M, Kleter GA, Kok EJ, Sharbati J, Hanisch C, Einspanier R, Adel-Patient K, Wal J-M, Spök A, Pöting A, Kohl C, Wilhelm R, Schiemann J, Steinberg P (2014). 90-day oral toxicity studies on two genetically modified maize MON810 varieties in Wistar Han RCC rats (EU 7th Framework Programme project GRACE). Archives of Toxicology, 88: 2289-2314 https://dx.doi.org/10.1007/s00204-016-1798-4

Zeljenková D, Aláčová R, Ondrejková J, Ambrušová K, Bartušová M, Kebis A, Kovrižnych J, Rollerová E, Szabová E, Wimmerová S, Černák M, Krivošíková Z, Kuricová M, Líšková A, Spustová V, Tulinská J, Levkut M, Révajová V, Ševčíková Z, Schmidt K, Schmidtke J, Schmidt P, La Paz JL, Corujo M, Pla M, Kleter GA, Kok EJ, Sharbati J, Bohmer M, Bohmer N, Einspanier R, Adel-Patient K, Spök A, Pöting A, Kohl C, Wilhelm R, Schiemann J, Steinberg P (2016). One-year oral toxicity study on a genetically modified maize MON810 variety in Wistar Han RCC rats (EU 7th Framework Programme project GRACE). Archives of Toxicology, 90: 2531-2562. https://dx.doi.org/10.1007/s00204-016-1798-4

Zeljenková D, Steinberg P (2015). Final study plan combined chronic toxicity and carcinogenicity study in rats fed GM maize NK603. G-TwYST GMP Two Year Safety Testing, Study 632165 B/2016/GLP. https://www.g-twyst.eu/files/Final%20study%20plans/Combined_chronic_toxicity-carcinogenicity_%20feeding_trial_study_plan_25.05.2015_final.pdf.

List of Tables

List of Figures

Appendix 1. Health status observations

Animal	Cage	Sex	Group	Death	Cens.	Organ, sys.	Finding	Start	End
2	1	M	NK33+	347	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition	04-07-16	22-07-16
21	11	M	NK33+	707	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition	23-01-17	17-07-17
21						skin	mass- 1cm, lumbal area	23-01-17	17-07-17
22	11	M	NK33+	728	1	skin	mass- multiple, 1-5cm, axillary region	01-10-16	07-08-17
23	12	M	NK33+	533	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea	06-12-16	24-01-17
25	13	M	NK33+	728	1	skin	1. mass- lumbal area, 1,5cm 2.mass- axillary area, 2cm	08-07-17	07-08-17
26	13	M	NK33+	682	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, torticollis, abnormal breathing	24-04-17	23-06-17
29	15	M	NK33+	693	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition	04-07-17	-
32	16	M	NK33+	532	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition	03-01-17	24-01-17
33	17	M	NK33+	352	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition	29-07-16	-
34	17	M	NK33+	668	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, forelimbs cramps	07-06-17	10-06-17
36	18	M	NK33+	728	1	skin	defect- 2,5cm in diamether, red coloured	27-07-17	09-08-17
37	19	M	NK33+	532	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea	04-01-17	25-01-17
39	20	M	NK33+	544	0	maxilla, nervous sys.	weight loss, mass- left maxilla, 3cm in diameter	26-01-17	07-02-17
40	20	M	NK33+	595	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, blepharospasmus	11-03-17	30-03-17
41	21	M	NK33+	728	1	skin	mass- 1cm, femoral area chromodacryorrhea	27-07-17	10-08-17
42	21	M	NK33+	728	1	skin	mass- 1cm, thoracic area	26-01-17	00-08-17
43	22	M	NK33+	474	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	16-11-16	29-11-16
43						skin	mass- right thoracic area, 1cm in diamether	26-11-16	29-11-17
44	22	M	NK33+	621	0	skin	alopecia- symetric., abdominal area bilat., 4cm in diamether, round shaped	25-04-17	-
46	23	M	NK33+	579	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	06-01-17	15-03-17
47	24	M	NK33+	634	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture, cramps	03-02-17	09-05-17
48	24	M	NK33+	728	1	skin	mass- 2cm, abdominal area	18-11-16	10-12-16
49	25	M	NK33+	598	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture, cramps	06-01-17	03-04-17
53	27	M	NK33+	637	0	eye	Right eye- enlargement, white coloured weight loss	03-04-17	15-05-17
55	28	M	NK33+	612	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture, eating bedding material, epistaxis	11-03-17	21-04-17
55						skin	mass- right mandible, 4-4,5cm	07-03-17	21-04-17
57	29	M	NK33+	685	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, torticollis, paraplegia of hind limbs	13-06-17	03-07-17

58	29	M	NK33+	601	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, torticollis	10-04-17	-
61	31	M	NK33+	728	1	skin	mass- multiple, 1-2cm	26-10-16	16-08-17
63	32	M	NK33+	728	1	skin	mass- mandible, 0,5cm	01-12-16	17-08-17
66	33	M	NK33+	568	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, dyspnoe, abnormal vocalisation	09-03-17	10-03-17
66						skin	scrub- 3cm in diamether, lumbal area	08-03-16	30-07-16
68	34	M	NK33+	644	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, torticollis	26-05-17	-
70	35	M	NK33+	687	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, torticollis	30-06-17	08-07-17
85	43	M	NK11-	351	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	02-08-16	04-08-16
92						skin	mass- 0,5cm, axillary area;	07-08-17	-
97	49	M	NK11-	546	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	04-02-17	07-02-17
99	50	M	NK11-	728	1	skin	mass- inguinal area, 2cm	07-02-17	08-08-17
101	51	M	NK11-	245	0	spleen, liver	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, anemia, mass- abdominal cavity	12-04-16	-
107	54	M	NK11-	728	1	skin/fur	weight loss, alopecia in neck area bilateral symmetric	15-06-16	09-08-17
113	57	M	NK11-	417	0	nervous sys.	weight loss, piloerection, lethargy	01-09-16	03-10-16
117	59	M	NK11-	728	1	skin	cicatrix	28-08-16	11-08-17
119	60	M	NK11-	728	1	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	10-08-17	11-08-17
120	60	M	NK11-	728	1	skin	mass- 7cm, abdominal area	27-07-17	11-08-17
120						nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	10-08-17	11-08-17
129	65	M	NK11-	728	1	head/ skin	mass- 3 cm, temporal area	29-07-17	16-08-17
130	65	M	NK11-	728	1	head/ maxilla	maloclusion- dentes incisivi	29-07-17	16-08-17
130						nervous sys.	weight loss, piloerection, lethargy, poor overall condition	15-08-17	16-08-17
133	67	M	NK11-	615	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	11-03-17	26-04-17
134	67	M	NK11-	728	1	skin/fur	alopecia- antebrachium, bilateral	12-02-17	17-08-17
135	68	M	NK11-	728	1	skin	mass- shoulder blade area, 0,5 cm	26-11-16	17-08-17
139	70	M	NK11-	728	1	skin	defect, scubs- head, pruritus	31-12-16	18-08-17
145	73	M	NK11+	364	1	skin	alopecia- axillary region, bilat, symetric, round shape, max 10-4cm, red coloured skin	09-02-16	10-08-16
148	74	M	NK11+	231	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, blepharospasmus, hunched posture, dyspnoe, paraplegia of hind limbs	30-03-16	31-03-16
162	81	M	NK11+	728	1	skin	mass- abdominal area, 2,5 cm	27-07-17	07-08-17

164	82	M	NK11+	621	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition	10-04-17	22-04-17
167	84	M	NK11+	728	1	skin	mass- scapular area, 0,5 cm	05-07-16	26-11-16
167						skin	mass- lumbal area, 1,5cm	27-07-17	08-08-17
169	85	M	NK11+	450	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	11-10-16	03-11-16
172	86	M	NK11+	476	0	skin	tail- defect, scub, bleeding	25-10-16	29-11-16
178	89	M	NK11+	449	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	26-10-16	03-11-16
179	90	M	NK11+	683	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, hunched posture	20-06-17	26-06-17
181	91	M	NK11+	704	0	head/skin	mass- 6cm, frontal area	07-07-17	17-07-17
187	94	M	NK11+	565	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	03-02-17	01-03-17
188	94	M	NK11+	728	1	abdominal cavity	mass- in abdominal cavity, 12cm, abdominalgy	23-06-17	11-08-17
190	95	M	NK11+	728	1	skin	mass- abdominal area, 1cm, green coloured	23-12-16	11-03-17
190						skin	red coloured skin, scubs - tail, scrotum	04-02-17	11-08-17
190						eye	white coloured, bilateral	27-07-17	11-08-17
195	98	M	NK11+	651	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	16-05-17	30-05-17
199	100	M	NK11+	182	0	nervous sys.	paraplegia- hind limbs	15-02-16	17-02-16
205	103	M	NK11+	562	0	skin	scub- lumbal area, 2cm	08-03-16	30-07-16
209	105	M	NK11+	460	0	nervous sys.	piloerection, lethargy, poor overall condition	23-11-16	-
212	106	M	NK33-	364	1	back area	alopecia- 0,5x0,4 cm	12-03-16	29-04-16
214	107	M	NK33-	364	1	back area	bite wound	20-10-15	-10-2015
216	108	M	NK33-	364	1	ear lobe	scub	17-02-16	12-03-16
227	114	M	NK33-	364	1	skin	scub- 2cm, round shape, lumbal area	08-03-16	30-07-16
233	117	M	NK33-	536	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, hunched posture	23-01-17	27-01-17
235	118	M	NK33-	728	1	tail	biteing off tail	22-05-17	07-08-17
236	118	M	NK33-	575	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, dyspnoe, hunched posture	07-03-17	-
238	119	M	NK33-	492	0	head- bones	mass- bleeding, scub,putrefactive odour, 3cm; anemia, piloerection, hunched posture	14-08-16	15-12-16
239	120	M	NK33-	542	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, dyspnoe, hunched posture	11-10-16	03-02-17
240	120	M	NK33-	728	1	skin/mammary gl.	mass- abdominal area, 3cm	27-05-17	08-08-17
245	123	M	NK33-	574	0	skin/fur	alopecia-bilatereral, symetric, neck area	04-11-15	30-07-16
246	123	M	NK33-	728	1	skin	mass/ scub- abdominal area	19-07-17	09-08-17
247	124	M	NK33-	707	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, dyspnoe, hunched posture	21-06-17	19-07-17
250	125	M	NK33-	728	1	praeputial lymphno.	left- enlargement	09-08-17	10-08-17
251	126	M	NK33-	616	0	skin	mass- abdominal area, 7cm, bleeding	09-03-17	20-04-17

253	127	M	NK33-	728	1	skin	mass- femoral area, 2cm	11-03-17	10-08-17
254	127	M	NK33-	728	1	skin	mass- shoulder blade area, 1 cm	01-04-17	10-08-17
255	128	M	NK33-	728	1	skin	mass- multiple, 0,5cm, lumbal and inguinal area	21-01-17	11-08-17
256	128	M	NK33-	728	1	skin	mass- lumbal area, 0,5cm, green	23-11-16	04-02-17
258	129	M	NK33-	724	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture, aggressive behavior	01-10-16	07-08-17
259	130	M	NK33-	728	1	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture, blind, tremor	10-08-17	-
263	132	M	NK33-	498	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, dyspnoe, hunched posture	18-12-16	27-12-16
269	135	M	NK33-	401	0	eye	enargement, bleeding,	12-09-16	23-09-16
273	137	M	NK33-	695	0	skin	mass- 1cm, lumbal area	01-04-17	15-07-17
284	142	M	Control	364	1	skin/mammary gl.	mass- 1,5-2cm, inguinal area	02-06-16	09-06-16
297	149	M	Control	364	1	skin	scub- left masseter area	-10-2015	-11-2015
304	152	M	Control	604	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	13-01-00	05-04-17
309	155	M	Control	583	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	16-03-17	-
315	158	M	Control	728	1	skin	mass- left shoulder blade area, red; bleeding, musous membranes anemia	31-12-16	09-08-17
316	158	M	Control	598	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	01-03-17	01-04-17
319	160	M	Control	490	0	skin	mass- inguinal area, 2,5cm, scub,	29-09-16	15-12-16
319						skin	mass- thoracic area, 4cm, algecis; polydipsia, hunched posture	10-12-16	15-12-16
320	160	M	Control	728	1	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture, blind, tremor	27-07-17	10-08-17
324	162	M	Control	516	0	skin	scubs, multiple	24-06-16	10-01-17
324						skin	mass- 10cm, black coloured, abdominal area	06-11-16	10-01-17
326	163	M	Control	669	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus	07-10-16	13-06-17
327	164	M	Control	728	1	head- masseter muscle	mass- 3cm, bleeding, scub; weight loss	18-12-16	11-08-17
329	165	M	Control	728	1	skin	scub- frontal area	22-07-16	11-08-17
332	166	M	Control	496	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea	08-12-16	21-12-16
340	170	M	Control	518	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea	14-12-16	18-01-17
343	172	M	Control	637	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, anemia, hunched posture, blepharospasmus	23-02-17	18-05-17
345	173	M	Control	728	1	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, blepharospasmus, hunched posture	16-08-17	17-08-17
347	174	M	Control	500	0	maxilla	mass- 2cm, white- yellow, putrefactive odour, chromodacryorrhea, weight loss	18-11-16	2-01-17

Animal	Cage	Sex	Group	Death	Cens.	Organ, sys.	Finding	Start	End
506	503	F	NK33+	364	1	tail	skin- red, scubs, necrosis, biteing of tail, anemia, poor overall condition, hunched posture, piloerection	29-06-16	24-08-16
521	511	F	NK33+	728	1	fur	whole body- alopecia	29-01-16	21-08-17
521						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	21-08-17	-
522	511	F	NK33+	667	0	skin/mammary gl.	mass- multiple, 2cm	10-04-17	21-06-17
523	512	F	NK33+	613	0	skin/mammary gl.	mass- 5cm, algesic, dark red; lethargy	25-03-17	22-04-17
528	514	F	NK33+	728	1	fur	alopecia- bilateral symmetric, forelimbs	23-02-16	22-08-17
531	516	F	NK33+	728	1	skin/mammary gl.	mass- 4cm	21-08-17	22-08-17
532	516	F	NK33+	663	0	skin/mammary gl.	mass- caudal part of mammary gland, 9ccm, bleeding, dark red	12-12-16	18-06-17
535	518	F	NK33+	728	1	skin/mammary gl.	mass- 2-3cm, inguina	08-02-17	23-08-17
535						skin	mandible- 0,4cm	30-07-17	23-08-17
536	518	F	NK33+	728	1	skin/mammary gl.	mass- 7-9cm, dark red skin	01-04-17	23-08-17
537	519	F	NK33+	704	0	fur	alopecia- sternum	28-07-16	04-09-16
537						nervous sys.	weight loss, lethargy	30-07-17	-
540	520	F	NK33+	728	1	reproductive sys.	mass- vulvar area, 2,5cm	27-07-17	24-08-17
541						skin/mammary gl.	mass- inguina, 2cm	30-07-17	24-08-17
543	522	F	NK33+	551	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, hunched posture	20-01-17	28-02-17
544	522	F	NK33+	728	1	skin	wounds, scubs- neck, lumbal area	30-06-16	24-08-17
545	523	F	NK33+	728	1	skin/mammary gl.	mass- 1cm	24-08-17	25-08-17
546	523	F	NK33+	599	0	skin/mammary gl.	mass- 5cm, caudal part of mammary gland, dark red- black, anemic, hunched posture, poor overall condition	26-11-16	18-04-17
547	524	F	NK33+	728	1	eye	enlargement, torticollis	30-07-17	25-08-17
548	524	F	NK33+	728	1	skin/mammary gl.	mass- caudal part of mammary gland, 4cm, bleeding	30-07-17	25-08-17
549	525	F	NK33+	728	1	skin/mammary gl.	mass- axilla, 4cm	24-08-17	25-08-17
554	527	F	NK33+	562	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, hunched posture	22-02-17	15-03-17
555	528	F	NK33+	552	0	fur	alopecia- femoral area, 3cm	31-12-16	06-03-17
555						nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, hunched posture	21-02-17	06-03-17
556	528	F	NK33+	524	0	fur	alopecia- 2cm, neck	02-10-13	06-02-17
556						skin/mammary gl.	mass- multiple, 3-7cm, dark red-black, bleeding	22-11-16	06-02-17
557	529	F	NK33+	596	0	nervous sys.	weight loss, chromodacryorrhea	10-04-17	19-04-17
558	529	F	NK33+	728	1	skin/mammary gl.	mass- muiltiple, 4cm	30-07-17	05-09-17
559	530	F	NK33+	728	1	skin/mammary gl.	mass- multiple, axilla, inguina, 3-4 cm	28-02-17	06-09-17
560	530	F	NK33+	728	1	skin	mass- neck, 4 cm, dark red	27-05-17	06-09-17
561	531	F	NK33+	713	0	fur	alopecia- forelimbs, hindlimbs- bilateral symetric	23-09-15	12-03-16

561						eye	left- exophtalmus, enlargement, defect, bleeding	30-07-17	07-08-17
562	531	F	NK33+	561	0	fur	alopecia- neck- bilateral, symetric	23-09-15	16-03-17
562						skin/mammary gl.	mass- inguina, 4 cm, bleeding, open wound	06-11-16	16-03-17
562						nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, hunched posture, anemia	04-02-17	16-03-17
563	532	F	NK33+	434	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, hunched posture	10-11-16	-
568	534	F	NK33+	728	1	skin/mammary gl.	mass- 1,5cm	07-09-17	08-09-17
569	535	F	NK33+	728	1	skin	mass- axilla, 1,5cm	07-09-17	08-09-17
570	535	F	NK33+	504	0	nervous sys.	weight loss, piloerection, lethargy, poor overall condition, chromodacryorrhea, hunched posture	25-11-16	20-01-17
579	540	F	NK11-	364	1	skin	wound- 1cm, lumbal area	06-07-16	28-07-16
583	542	F	NK11-	364	1	maxilla, incisors	maloclusion- overgrown upper incisors	06-11-15	06-09-16
585	543	F	NK11-	364	1	skin	mass- inguina, 3cm	29-06-16	07-09-16
591	546	F	NK11-	728	1	mandibular lymphnodes	enlargement	30-07-17	21-08-17
592	546	F	NK11-	728	1	skin/mammary gl.	mass- multiple, 3-8cm, dark red skin	10-04-17	21-08-17
593	547	F	NK11-	613	0	eye	exophtalmus, torticollis	04-12-16	28-04-17
593						skin/mammary gl.	mass- 4cm, bleeding, open wound, anemia	11-03-17	28-04-17
595	548	F	NK11-	701	0	skin/mammary gl.	mass- multiple, 1-4cm	20-06-16	25-07-17
595						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	25-07-17	-
596	548	F	NK11-	679	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	03-07-17	-
597	549	F	NK11-	526	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, torticollis	01-02-17	-
598	549	F	NK11-	619	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, torticollis	03-05-17	5-05-17
599	550	F	NK11-	616	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition	28-03-17	02-05-17
603	552	F	NK11-	728	1	skin/mammary gl.	mass-caudal part, 7cm, black coloured skin	21-06-17	23-08-17
605	553	F	NK11-	728	1	skin/mammary gl.	mass- 5cm, caudal part of mammary gland	21-06-17	23-08-17
606	553	F	NK11-	650	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	17-05-17	06-06-17
606						skin/mammary gl.	mass- 6cm, dark red skin	28-01-00	23-08-17
608	554	F	NK11-	657	0	skin/mammary gl.	mass- axilla, 5cm, black, open wound	09,11,16	13-06-17
609	555	F	NK11-	403	0	skin/mammary gl.	mass- axilla, 12cm, poor body condition	14-07-16	03-10-16
611	556	F	NK11-	651	0	skin	scubs- multiple, shoulder blade area and head	04-02-17	01-04-17
611						skin/mammary gl.	mass- neck, 2cm	18-05-17	08-06-17

612	556	F	NK11-	670	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, torticollis	27-05-17	27-06-17
613	557	F	NK11-	323	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	14-07-16	15-07-16
614	557	F	NK11-	728	1	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	23-08-17	24-08-17
615	558	F	NK11-	669	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	20-06-17	27-06-17
616	558	F	NK11-	372	0	skin/fur	whole body- alopecia, red coloured skin, scubs, conjuctivitis	01-01-16	03-09-16
617	559	F	NK11-	728	1	skin/mammary gl.	mass- 2cm	24-08-17	25-08-17
618	559	F	NK11-	728	1	skin/mammary gl.	mass- multiple, 1-2cm	04-02-17	25-08-17
619	560	F	NK11-	728	1	skin/mammary gl.	mass- inguina, 4cm	30-07-17	25-08-17
620	560	F	NK11-	707	0	skin/mammary gl.	mass- 5cm	30-07-17	04-08-17
620						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	30-07-17	04-08-17
621	561	F	NK11-	646	0	skin/mammary gl.	mass- 12x10cm, dark red, poor body condition	26-11-16	07-06-17
622	561	F	NK11-	728	1	skin/mammary gl.	mass- axilla, 1,5cm	04-09-17	-
624	562	F	NK11-	728	1	nervous sys.	torticollis	04-09-17	-
624						skin/mammary gl.	mass- 5cm, dark red skin	04-09-17	-
627	564	F	NK11-	694	0	skin/mammary gl.	mass- 10 cm, open wound, bleeding, putrefactive odour	31-12-16	26-07-17
628	564	F	NK11-	647	0	skin/mammary gl.	mass- 2,5cm, caudal part	04-02-17	09-06-17
628						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	01-06-17	09-06-17
629	565	F	NK11-	728	1	skin	mass- lumbal area, 1cm; epistaxis	30-07-17	06-09-17
632	566	F	NK11-	238	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	20-04-16	27-04-16
633	567	F	NK11-	420	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	10-10-16	27-10-16
636	568	F	NK11-	617	0	skin/mammary gl.	mass- multiple, 4-7cm, poor body condition, hunched posture	07-04-16	12-05-17
638	569	F	NK11-	728	1	skin	mass- shoulder blade area, 3 cm, bleeding, defect, open wound	09-20016	08-09-17
640	570	F	NK11-	728	1	skin/mammary gl.	mass- 4cm, abdominal area	07-09-17	08-09-17
661	581	F	NK11+	728	1	skin/mammary gl.	mass- inguina, 3cm	21-08-17	-
662	581	F	NK11+	728	1	skin/mammary gl.	mass-2,5cm, axilla	21-08-17	-
663	582	F	NK11+	728	1	skin/mammary gl.	mass- 6-7cm, multiple scubs	10-12-16	21-08-17
664	582	F	NK11+	728	1	skin/mammary gl.	mass- 3cm, black skin, bleeding, anemia, poor body condition, hunched posture	17-04-17	21-08-17
665	583	F	NK11+	650	0	skin/mammary gl.	mass- 8x6cm, dark red- black, poor body condition	10-01-17	04-06-17
669	585	F	NK11+	707	0	skin	scubs- lumbal area, 4cm	15-09-16	06-11-16
669						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	01-08-17	-

670	585	F	NK11+	728	1	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, dyspnoe, epistaxis,anemia	11-03-17	22-08-17
671	586	F	NK11+	728	1	skin, fur, eye	whole body- alopecia, red coloured skin, scubs, conjuctivitis	28-01-16	22-08-17
672	586	F	NK11+	587	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	28-03-17	03-04-17
673	587	F	NK11+	513	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	21-12-16	20-01-17
674	587	F	NK11+	704	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	29-07-17	30-07-17
675	588	F	NK11+	728	1	skin/mammary gl.	mass- multiple, 0,5-1cm	22-08-17	23-08-17
676	588	F	NK11+	592	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	29-03-17	09-04-17
677	589	F	NK11+	506	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	18-12-16	13-01-17
678						skin/mammary gl.	mass- inguina,3cm	30-07-17	23-08-17
679	590	F	NK11+	594	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, epistaxis	19-01-17	12-04-17
680	590	F	NK11+	454	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, epistaxis	29-09-16	23-11-16
681	591	F	NK11+	656	0	skin/mammary gl.	mass- 5-6cm, dark red, open wound, anemia	26-01-17	13-06-17
683	592	F	NK11+	434	0	mammary gl.	mass- 6-8cm, green scubs on skin, algesic	29-09-16	03-11-16
685	593	F	NK11+	616	0	skin/mammary gl.	mass- 1cm, dark red, scub	02-04-17	05-05-17
685						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, tremor	02-04-17	05-05-17
687	594	F	NK11+	728	1	skin	wounds, scubs- multiple, lumbal area	30-12-15	25-08-17
690	595	F	NK11+	728	1	skin/mammary gl.	mass- inguina, 10cm, dark red skin	15-07-17	25-08-17
691	596	F	NK11+	701	0	skin/mammary gl.	mass- axilla, 3 cm	20-02-17	01-08-17
691						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, epistaxis	30-07-17	01-08-17
692	596	F	NK11+	658	0	skin/mammary gl.	mass- 2,5cm, caudal part	20-06-16	19-06-17
692						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	28-05-17	19-06-17
693	597	F	NK11+	610	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	02-04-17	02-05-17
694	597	F	NK11+	728	1	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, vaginal secretion- green	04-09-17	-
695	598	F	NK11+	728	1	skin/mammary gl.	mass- inguina, 5cm	20-08-17	05-09-17
696	598	F	NK11+	694	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	05-05-17	26-07-17

697	599	F	NK11+	615	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	05-05-17	08-05-17
698						skin/mammary gl.	mass- 1cm	04-09-17	05-09-17
699	600	F	NK11+	728	1	skin/mammary gl.	mass- multiple, 3-6cm	28-05-17	06-09-17
703	602	F	NK11+	728	1	skin	scubs- multiple, lumbal area and head	15-03-16	07-09-17
705	603	F	NK11+	728	1	skin/mammary gl.	mass- axilla, 1,5cm	06-09-17	07-09-17
707	604	F	NK11+	704	0	skin/mammary gl.	mass- 2cm, caudal part, open wound	30-07-17	08-08-17
707						nervous sys.	weight loss , piloerection, lethargy, poor overall condition	08-08-17	-
709	605	F	NK11+	728	1	vulva	mass- dark red, bleeding, anemia	30-07-17	08-09-17
714	607	F	NK33-	364	1	auricle	scubs, red coloured	25-02-16	12-03-16
727	614	F	NK33-	364	1	skin	wounds, scubs- thoracic area	01-10-15	08-09-16
730	615	F	NK33-	364	1	skin	wound- 0,5cm	09-03-16	29-04-16
731	616	F	NK33-	576	0	skin/mammary gl.	mass- 12cm, inguina, dark coloured skin, open wound	23-08-16	22-03-17
732	616	F	NK33-	466	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, abdominalgy	31-10-16	02-12-16
733	617	F	NK33-	688	0	fur	alopecia-hind limbs, lumbal area	-12-2015	12-07-17
733						skin/mammary gl.	mass- multiple, 1-3cm	06-11-16	12-07-17
733						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	07-07-17	12-07-17
734	617	F	NK33-	728	1	fur	alopecia- frontal area	14-12-15	21-08-17
735	618	F	NK33-	728	1	fur	alopecia- neck area, shoulder blade area- bilateral symetric	19-10-15	21-08-17
735						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	21-08-17	-
737	619	F	NK33-	728	1	skin	scubs, wounds	23-09-15	22-08-17
737						skin/mammary gl.	mass- multiple, 1-2,5cm, bleeding	02-04-17	22-08-17
738	619	F	NK33-	637	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	23-05-17	-
740	620	F	NK33-	728	1	skin/mammary gl.	mass- 4-5cm, red skin	28-05-17	22-08-17
741	621	F	NK33-	721	0	skin/mammary gl.	mass- 2,5cm, axilla	25-03-17	15-08-17
741						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	30-07-17	15-08-17
742	621	F	NK33-	653	0	skin/mammary gl.	mass- 2,5cm, cranial part of mammary gland	31-01-17	08-06-17
742						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	28-05-17	08-06-17
744	622	F	NK33-	728	1	skin/mammary gl.	mass- multiple, 2-3 cm	04-02-17	23-08-17
744						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea, torticollis	22-08-17	23-08-17
745	623	F	NK33-	299	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	15-06-16	20-06-16

749	625	F	NK33-	662	0	skin/mammary gl.	mass- 4-5cm, axilla, dark red skin	18-05-17	19-06-17
751	626	F	NK33-	728	1	skin/mammary gl.	mass- multiple, axilla, 2-4 cm	023-08-	24-01-00
752	626	F	NK33-	663	0	skin/mammary gl.	mass- 4-5cm, caudal part	01-06-17	10-06-17
752						nervous sys.	weight loss , piloerection, lethargy, hunched posture, chromodacryorrhea, torticollis	10-06-17	20-06-17
755	628	F	NK33-	630	0	skin/mammary gl.	mass- multiple, 2-3cm	21-01-17	19-05-17
755						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	19-05-17	-
757	629	F	NK33-	573	0	head	mass- 4cm, dental maloclusion, weight loss	13-01-17	23-03-17
758	629	F	NK33-	728	1	skin/mammary gl.	mass- axilla, 1,5cm	25-08-17	-
759	630	F	NK33-	681	0	skin/mammary gl.	mass- 3cm, shoulder blade area	09-06-17	09-07-17
759						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	07-07-17	09-07-17
760	630	F	NK33-	609	0	skin/mammary gl.	mass- caudal part, 2cm,	15-07-16	28-04-17
760						nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	28-04-17	-
761	631	F	NK33-	728	1	eye	chromodacryorrhea	14-15-15	04-09-17
761						skin/mammary gl.	mass- 3cm, axilla; tremor	10-04-17	04-09-17
762	631	F	NK33-	728	1	eye	chromodacryorrhea	21-11-15	04-09-17
763	632	F	NK33-	728	1	reproductive sys.	vaginal bleeding, anemia	28-05-17	04-09-17
764	632	F	NK33-	616	0	skin/mammary gl.	mass- axilla, 6 cm, bleeding	12-03-17	08-05-17
765	633	F	NK33-	619	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	18-04-17	12-05-17
766	633	F	NK33-	728	1	auricle	scub- bilateral	28-09-16	10-12-16
767	634	F	NK33-	728	1	skin/mammary gl.	mass- cranial part of mammary gl., 0,5cm	04-09-17	05-09-17
773	637	F	NK33-	728	1	skin/mammary gl.	mass- multiple, 3-6cm	26-01-17	07-09-17
775	638	F	NK33-	552	0	fur	alopecia- sternal area	28-07-16	08-03-17
776	638	F	NK33-	581	0	fur	alopecia- head, unilat.	02-04-16	-
776						nervous sys.	piloerection, lethargy, hunched posture	06-04-17	-
778	639	F	NK33-	683	0	skin/mammary gl.	mass- multiple, 2-5cm, open wound, dark coloured	26-11-16	18-07-17
779	640	F	NK33-	728	1	skin/mammary gl.	mass- axilla, 2,5 cm	07-09-17	08-09-17
780	640	F	NK33-	728	1	skin/mammary gl.	mass- inguina, 4cm	04-09-17	08-09-17
784	642	F	Control	364	1	auricle	scubs, red coloured	28-02-16	12-03-16
790	645	F	Control	364	1	skin	wounds, scubs- lumbal area	15-07-16	26-08-16
792	646	F	Control	364	1	skin/mammary gl.	mass- 6cm, caudal part of mammary gland	01-08-16	05-09-16
798	649	F	Control	364	1	skin	tail- bite wounds	28-07-16	21-08-16
801	651	F	Control	728	1	skin/mammary gl.	mass- axilla, 1,5cm	21-08-17	-
802	651	F	Control	666	0	skin/mammary gl.	mass- multiple, 2-4cm	06-11-16	20-06-17
802						nervous sys.	piloerection, hunched posture	20-06-17	-
802						fur	alopecia- 5-6cm, unilat.	05-02-17	20-06-17

804	652	F	Control	332	0	skin/fur	scubs/ wound- lumbal area, 5cm, weight loss, poor body condition	17-10-15	21-07-16
805	653	F	Control	418	0	skin/mammary gl.	mass- 1cm, caudal part	30-07-16	15-10-16
806	653	F	Control	420	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	10-10-16	17-10-16
808	654	F	Control	476	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	13-12-16	-
809	655	F	Control	589	0	skin/mammary gl.	mass- 5cm, axilla, dark red, poor overall condition	21-01-17	05-04-17
813	657	F	Control	394	0	nervous sys.	piloerection, lethargy, poor overall condition, hunched posture	23-09-16	-
814	657	F	Control	609	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	26-04-17	-
814						skin/mammary gl.	mass- 5cm, dark coloured skin	28-02-17	26-04-17
817	659	F	Control	506	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea,	17-12-16	13-01-17
820	660	F	Control	728	1	skin/mammary gl.	mass- axilla, 3cm	30-07-17	24-08-17
823	662	F	Control	728	1	skin/mammary gl.	mass- multiple, axilla, inguina, 2-2,5cm	30-07-17	24-08-17
824	662	F	Control	728	1	mammary gl.	mass- 5cm	30-07-17	24-08-17
826	663	F	Control	537	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	03-02-17	15-02-17
826						skin/mammary gl.	mass- 5cm, axilla, dark red coloured	21-08-16	15-02-17
828	664	F	Control	728	1	skin/mammary gl.	mass- inguina, 2cm	30-07-17	25-08-17
830	665	F	Control	629	0	nervous sys.	weight loss , piloerection, lethargy, poor overall condition, hunched posture, chromodacryorrhea	18-05-17	-
831						skin/mammary gl.	mass- multiple, 4-7cm, porr body condition, hunched posture	30-07-17	04-09-17
832	666	F	Control	654	0	skin/mammary gl.	mass- 5-7cm	20-02-17	15-06-17
833	667	F	Control	728	1	skin/mammary gl.	mass- axilla, 4-5cm	20-03-17	04-09-17
834	667	F	Control	728	1	skin/mammary gl.	mass- inguina, 1cm, poor body condotion, hunched posture	30-07-17	04-09-17
835	668	F	Control	528	0	skin/mammary gl.	mass- 10cm, dark coloured skin, axilla,	02-08-16	10-02-17
836	668	F	Control	728	1	skin/mammary gl.	mass- dark coloured, 2cm	30-07-17	05-09-17
837	669	F	Control	728	1	skin/mammary gl.	mass- dark red, 1cm	04-09-17	05-09-17
838	669	F	Control	728	1	skin/mammary gl.	mass- multiple, 6cm, bleeding, open wound; poor overall condition	05-02-17	05-09-17
840	670	F	Control	454	0	skin/mammary gl.	mass- 9cm, dark coloured, scub	30-07-16	29-11-16
841	671	F	Control	523	0	skin-tail	mass- 4-5cm, bleeding	11-05-16	06-02-17
841						lymph.nodes	inguinal lymphn.- enlargement bilat.	31-12-16	06-02-17
842	671	F	Control	532	0	skin/mammary gl.	mass- 10-12cm, dark coloured skin, poor body condition	25-09-16	15-02-17
843	672	F	Control	445	0	skin	scubs/wounds- lumbal area, 8x8cm area, poor body condition, weight loss	04-01-16	21-11-16
845	673	F	Control	728	1	fur	alopecia- bilateral symmetric, auricle area	18-05-17	07-09-17
847	674	F	Control	728	1	skin/mammary gl.	mass- axilla, 4cm	07-09-17	08-09-17

Appendix 2. Sums of squares, degrees of freedom and effective replication for the GRACE data

The values below are based on a simultaneous statistical analysis of the five GRACE studies A-E on cage means after a log-transform and after removal of outliers. The columns have the following interpretation:

- SS_R the between reference feeds sums of squares
- SS_E the residual sums of squares
- df_E the degrees of freedom for the residual sums of squares
- n_{eff} the effective replication
- %R/S the between reference feeds estimated standard error (σ_R) as a percentage of the estimated residual standard error (σ_E), where empty cells denote zero values.

Variable	Male rats					Female rats				
	SS_R	SS_E	df_E	n_{eff}	%R/S	SS_R	SS_E	df_E	n_{eff}	%R/S
BodyWeight	0.008387	0.19554	69	10.50		0.014574	0.26913	69	10.50	
growthRate	0.000206	0.02434	68	10.33		0.000751	0.06011	69	10.50	
FeedMean	0.022190	0.16515	69	10.50	35	0.032193	0.24197	69	10.50	35
WBC	0.489441	2.61275	59	9.25	44	0.356101	3.36561	59	9.25	25
RBC	0.004304	0.14638	59	9.25		0.007652	0.07286	59	9.25	24
HGB	0.008819	0.09808	58	9.08	18	0.003452	0.05478	59	9.25	
HCT	0.005987	0.13627	59	9.25		0.004400	0.04452	59	9.25	22
MCV	0.000242	0.01864	59	9.25		0.000925	0.02226	59	9.25	
MCH	0.001610	0.03980	58	9.08		0.002266	0.05291	58	9.11	
MCHC	0.001347	0.01505	58	9.08	18	0.000187	0.01436	57	8.94	
PLT	0.056905	3.74573	59	9.25		0.057845	1.66444	57	9.08	
LYMA	0.402469	2.53641	59	9.25	38	0.320270	3.15442	59	9.25	23
Lymphocytes	0.008561	0.11920	59	9.25	8	0.012897	0.12741	59	9.25	23
Neutrophils	0.101555	1.54710	59	9.25		0.214056	1.72788	59	9.25	30
Monocytes	0.932225	6.25457	59	9.25	36	0.948050	4.73942	59	9.25	46
Eosinophils	1.831756	12.92053	59	9.25	34	1.006901	14.28882	59	9.25	7
ALP	0.058075	1.80285	59	9.25		0.050742	1.41079	59	9.25	
ALT	0.040433	0.76309	57	8.90		0.143187	3.83379	59	9.25	
AST	0.082408	1.33797	58	9.08		0.108613	2.28004	58	9.08	
ALB	0.003719	0.16461	58	9.08		0.002401	0.49386	59	9.25	
TP	0.003316	0.09254	59	9.25		0.044110	0.30963	59	9.25	35
Glu	0.137923	1.20832	59	9.25	27	0.068248	1.26028	58	9.08	
CHOL	0.157821	0.72566	59	9.25	49	0.066835	0.91290	59	9.25	9
TAG	1.512480	4.78566	59	9.25	63	0.455467	6.38891	59	9.25	7
Crea	0.100802	1.01337	59	9.25	22	0.169402	0.84663	59	9.25	46
Urea	0.162169	0.56082	59	9.25	59	0.212305	0.68252	59	9.25	62
Ca	0.021413	0.21007	58	9.25	23	0.069969	0.13967	58	9.08	83
Cl	0.010708	0.13568	59	9.25	13	0.001771	0.04791	59	9.25	
K	0.025453	0.68615	59	9.25		0.063943	0.62498	58	9.08	23
Na	0.011611	0.16216	59	9.25	8	0.003438	0.04081	58	9.08	16
P	0.071064	0.54837	58	9.08	31	0.084107	0.93345	59	9.25	19

Variable	Male rats					Female rats				
	SS_R	SS_E	df_E	n_{eff}	%R/S	SS_R	SS_E	df_E	n_{eff}	%R/S
Kidney	0.008755	0.15764	50	8.00		0.002113	0.14971	50	8.00	
Spleen	0.028870	0.32638	50	8.00	11	0.007828	0.48420	50	8.00	
Liver	0.011485	0.10964	48	7.66	18	0.021902	0.24719	50	8.00	12
AdrenGl	0.054960	0.63981	50	8.00	10	0.021734	0.71168	50	8.00	
Heart	0.015268	0.12327	50	8.00	26	0.005002	0.16094	50	8.00	
Thymus	0.135724	1.59675	50	8.00	9	0.168167	0.75330	50	8.00	47
Testis	0.012233	0.26694	50	8.00		-	-	-	-	
Epididymis	0.012314	0.28836	50	8.00		-	-	-	-	
Uterus	-	-	-	-		0.063741	1.37482	49	7.83	
Ovary	-	-	-	-		0.187232	0.81620	50	8.00	48
Brain	0.017742	0.11613	50	8.00	34	0.012478	0.19660	50	8.00	

Appendix 3. R-script for the classical statistical analysis.

```
# Classical Statistical analysis of a single response for G-TwYST studies
# Define some settings
alpha <- 0.95            # Confidence level for intervals
friedman.mc <- 20000     # Number of MC samples for p.value of Friedman test
friedman.limit <- 0.25   # Do MC when Asymptotic p.value is smaller than limit
set.seed(492193917)      # Initialize random generator
alpha2 <- (1+alpha)/2

# Define a dataframe to save results of test-statistics
testStats <- as.data.frame(matrix(nrow=9,ncol=2))
colnames(testStats) <- c("statistic", "value")
rownames(testStats) <- c("fAnova", "pAnova", "pKS", "pSW",
    "pFriedmanAs", "pFriedmanMc","pBartlett", "pLeveneMean", "pLeveneMedian")
testStats[,1] <- c("fAnova", "pAnova", "pKS", "pSW",
    "pFriedmanAs", "pFriedmanMc","pBartlett", "pLeveneMean", "pLeveneMedian")

# Get data, define factors and sort (necessary for wilcox.test())
data <- read.csv("RscriptInput.csv")
data$block <- as.factor(data$block)
data$treat <- as.factor(data$treat)
newlevels <- levels(data$treat)[c(1,2,4,3,5)]
data$treat <- factor(data$treat, newlevels)
data <- data[order(data$block, data$treat),]

# Load libraries
suppressMessages(library(multcomp))  # For Dunnett multiple comparison after ANOVA
suppressMessages(library(MBESS))     # For SES intervals
suppressMessages(library(NSM3))      # For p.value of non-parametric Friedman test
suppressMessages(library(car))       # For Levene test for homogeneity of variance
library(broom)                       # For the tidy() function

# ============================================================================
# Randomized block ANOVA; note that lm() takes proper account of any unbalance
# Estimates for treat are differences between GMO feeds and the Control feed
# ============================================================================
lm0 <- lm(response ~ block, data)
lm  <- lm(response ~ block + treat, data)
aov <- anova(lm, lm0)
testStats["fAnova","value"] <- aov$F[2]
testStats["pAnova","value"] <- aov$'Pr(>F)'[2]
estimates <- tidy(lm)
estimates

# Extract results for treatment differences, rename columns, add Residual Df and Se
tAnova <- estimates[startsWith(estimates[,1],"treat"),
    c("term", "estimate", "std.error", "statistic", "p.value")]
colnames(tAnova)[c(3,4,5)] <- c("se", "tvalue", "pvalue")
tAnova$term <- gsub("treat", "", tAnova$term)
tAnova$term <- gsub("'", "", tAnova$term)
tAnova$dfRes <- df.residual(lm)
tAnova$seRes <- summary(lm)$sigma

# Add confidence interval for differences
edt <- qt(alpha2, tAnova$dfRes)
tAnova$CIlower <- tAnova$estimate - edt*tAnova$se
tAnova$CIupper <- tAnova$estimate + edt*tAnova$se
```

```
# Add Dunnett p.values
dunnett <- glht(lm, linfct=mcp(treat="Dunnett"))
tAnova$pdunnett <- summary(dunnett)$test$pvalues

# Add SES and its confidence interval
tAnova$SES <- tAnova$estimate/tAnova$seRes
tAnova$SESlower <- NA * tAnova$SES
tAnova$SESupper <- NA * tAnova$SES
tnobs <- aggregate(response ~ treat, data=data, FUN=function(x) sum( !is.na(x) ))
nobs  <- as.vector(tnobs[,2])
mult  <- sqrt( (nobs[1] + nobs[c(2,3,4,5)]) / (nobs[1]*nobs[c(2,3,4,5)]) )
for (ii in 1:4) {
  CInct <- conf.limits.nct(tAnova$tvalue[ii], tAnova$dfRes[ii], conf.level=alpha)
  tAnova$SESlower[ii] <- mult[ii] * CInct$Lower.Limit
  tAnova$SESupper[ii] <- mult[ii] * CInct$Upper.Limit
}

# Do normality checks for lm() residuals
# This is only approximate because residuals are not independent
res <- residuals(lm)
ks <- ks.test(res, "pnorm", mean(res), sd(res))
testStats["pKS","value"] <- ks$p.value
sw <- shapiro.test(res)
testStats["pSW","value"] <- sw$p.value

# ============================================================================
# Non-parametric Friedman test and pairwise Wilcoxon signed rank tests
# ============================================================================
fried <- friedman.test(data$response, groups=data$treat, blocks=data$block)
testStats["pFriedmanAs","value"] <- fried$p.value

if (fried$p.value < friedman.limit) {
  blk <- as.numeric(data$block)
  trt <- as.numeric(data$treat)
  pFrd <- pFrd(x=data$response, b=blk, trt=trt, method="Monte Carlo",
n.mc=friedman.mc)
  testStats["pFriedmanMc","value"] <- pFrd$p.val
} else {
  testStats["pFriedmanMc","value"] <- NA
}

# Pairwise Wilcoxon signed rank test. Note that ordering of data is (block,treat)
# Also note that wilcox.test() can handle NA
sublevels <- levels(data$treat)[1]
controlData <- data$response[data$treat %in% sublevels]
for (ii in 1:4) {
  sublevels <- levels(data$treat)[1+ii]
  treatData <- data$response[data$treat %in% sublevels]
  wt <- wilcox.test(controlData, treatData, paired=TRUE, exact=TRUE)
  tPairwise$pwilcoxon[ii] <- wt$p.value
}
```

```
# ============================================================================
# Bartlett test for homogeneity of variance
# ============================================================================
bt <- bartlett.test(response ~ treat, data=data)
testStats["pBartlett","value"] <- bt$p.value
lv <- levene.test(response ~ treat, data=data, center=mean)
testStats["pLeveneMean","value"] <- lv$'Pr(>F)'[1]
lv <- leveneTest(response ~ treat, data=data, center=median)
testStats["pLeveneMedian","value"] <- lv$'Pr(>F)'[1]

# ============================================================================
# Kolomogorov-Smirnov & Shapiro-Wilks tests for normality
# ============================================================================
normality <- as.data.frame(matrix(nrow=5,ncol=3))
colnames(normality) <- c("term", "pKS", "pSW")
normality$term <- levels(data$treat)
normality[, seq(2,ncol(normality))] <- NA
for (ii in 1:5) {
  sublevels <- levels(data$treat)[ii]
  treatData <- data$response[data$treat %in% sublevels]
  treatData <- treatData[!is.na(treatData)]
  if (var(treatData) > 0) {
    ks <- ks.test(treatData, "pnorm", mean(treatData), sd(treatData))
    normality$pKS[ii] <- ks$p.value
    sw <- shapiro.test(treatData)
    normality$pSW[ii] <- sw$p.value
  }
}

# ============================================================================
# Output results for further processing in GenStat
# ============================================================================
for (ii in 1:1) {
  cat("\n")
  print(testStats);  cat("\n\n")
  print(tAnova);     cat("\n\n")
  print(tPairwise);  cat("\n\n")
  print(normality);  cat("\n\n")
}
write.csv(testStats, file="RscriptTest.csv",      row.names=FALSE, quote=FALSE)
write.csv(tAnova,    file="RscriptAnova.csv",     row.names=FALSE, quote=FALSE)
write.csv(tPairwise, file="RscriptPairwise.csv",  row.names=FALSE, quote=FALSE)
write.csv(normality, file="RscriptNormality.csv", row.names=FALSE, quote=FALSE)
```