G-TWYST
GM PLANTS TWO YEAR SAFETY TESTING

G-TwYST Study B

a 90-day toxicity study in rats fed GM maize NK603

Statistical report

Paul W. Goedhart & Hilko van der Voet

G-TwYST (EU grant agreement no: 632165) Deliverable 4.1

Biometris report 31.10.17

October 2017
updated 22-02-2018

https://doi.org/10.18174/455203 with appendices https://doi.org/10.18174/455204

WAGENINGEN
UNIVERSITY & RESEARCH

G-TwYST Study B

a 90-day toxicity study in rats fed GM maize NK603

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

October 2017

Contents

Abstract

The purpose of oral toxicity study B 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 90 days at incorporation rates of 11% and 33% in the feed. The effects were assessed relative to the responses for rats fed the near-isogenic non-GM maize, and the differences were compared, using a recently developed method for equivalence testing, to differences between non-GM feeds obtained in previous studies performed in the EU project GRACE.

Given tentative settings for regulatory parameters, for a set of 320 comparisons involving body and organ weights, haematology and clinical chemistry, equivalence was established in 94% of cases, close to the nominal confidence level of the test which was 95%. Equivalence was found to be more likely than lack of equivalence in 100% of cases.

In addition to this primary analysis, the report also contains results for equivalence testing using external target effect sizes for a limited set of variables, results for classical statistical analysis of differences, graphs of standardised effect sizes such as performed in the GRACE study, results of factorial analysis and graphs showing correlations between variables associated with liver or kidney damage.

1 Introduction

The purpose of oral toxicity study B 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 90 days at incorporation rates of 11% and 33% in the feed. The effects were assessed relative to the responses for rats fed the near-isogenic non-GM maize.

This report describes the results of the statistical analysis of the data from Study B. 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 makes use of 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 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.

The results comparing for each variable each GM dose group to the non-GM control group 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, results from more integrated analyses were obtained:

5. Factorial analysis, integrating over the five dose group, with main factors GM inclusion rate 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 A-E 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 according to the six schemes of statistical analysis. Section 5 gives a summary and some evaluation of the methodology. Appendices to this report are provided as a separate document.

2 Data

2.1 Data in G-TwYST study B

Study B is a 90-day (sub-chronic) toxicity 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 B/2016/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 within blocks. The design is thus a complete randomized block design with cage as the experimental units. There are eight blocks with male rats and eight other blocks with female rats. Most of the measurements are on individual animals, only feed intake is measured on the cage level. Some specific measurements, notably immunology and cytokines, are only done on a limited number of rats.

The definition of the five feeding groups is given in Table 1. This includes the coding of the blinded treatments. In this report the Group labels (Control, NK11-, NK33-, NK11+ and NK33+) will mostly be used.

Table 1 Feeding groups used in study B.

Factor	Levels / Labels				
Treat (blinded)	XE	XA	XC	XD	XB
Group	Control	NK11-	NK33-	NK11+	NK33+
Maize	Control	NK603	NK603	NK603	NK603
AmountNK	0	11	33	11	33
RoundUp	No	No	No	Yes	Yes

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 2. The main interest is in the difference between each of the four GM maize feeding groups and the control feeding group.

Table 2 Skeleton analysis of variance for cage means for a single sex.

Source of variation	d.f.
Block stratum	7
Block.Cage stratum	
Group	4
Residual	28
Total	39

The observed variables in Study B are given in Table 3.

Table 3 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 study. Variables given in red are not statistically analysed (see section 4.3.1 for details).

Weights	Description	Unit	Grace
Weight 1-14	Body weight at weeks 0, 1 … 13	g/animal	X
Feed 1-13	Feed intake in week 1, 2 … 13	g/cage	X
BodyWeight	Body weight at the end of the trial, i.e. at week 13	g/animal	X
growthRate	Growth rate fitted to the weight over all weeks	1/week	X
FeedMean	Mean of feed intake over 13 weeks	g/animal/day	X
Haematology	Description	Unit	Grace
WBC	white blood cells	10^9/L	X
RBC	red blood cells	10^{12}/L	X
HGB	haemoglobin	g/dL	X
HCT	haematocrit	%	X
MCV	mean cell volume	fL	X
MCH	mean corpuscular haemoglobin	pg	X
MCHC	mean corpuscular haemoglobin concentration	g/dL	X
PLT	platelets	10^9/L	X
LYMR	relative lymphocytes count	%	-
LYMA	absolute count of lymphocytes	10^3/uL	X
diffWBC	Description	Unit	Grace
Lymphocytes	Percentage of lymphocyte cells in 200 cells	%	X
Neutrophils	Percentage of neutrophil cells in 200 cells	%	X
Monocytes	Percentage of monocyte cells in 200 cells	%	X
Eosinophils	Percentage of eosinophil cells in 200 cells	%	X
Basophils	Percentage of basophil cells in 200 cells	%	X
ClinChem	Description	Unit	Grace
ALP	alkaline phosphatase	µkat/L	X
ALT	alanine aminotransferase	µkat/L	X
AST	aspartate aminotransferase	µkat/L	X
BIL	bilirubin	µmol/L	-
ALB	albumin	g/L	X
TP	total protein	g/L	X
Glu	glucose	mmol/L	X
CHOL	cholesterol	mmol/L	X
TAG	triglycerides	mmol/L	X
Crea	creatinine	mmol/L	X
Urea	urea	mmol/L	X
cHGB	haemoglobin	mg/dL	-
Ca	calcium	mmol/L	X
Cl	chloride	mmol/L	X
K	potassium	mmol/L	X
Na	sodium	mmol/L	X
P	phosphorus	mmol/L	X
Urine	Description	Unit	Grace
uVol	Urine Volume	ml	
uVolW	Urine Volume / bodyweight	ml /100g	
uColour	Urine Colour (1 – light yellow; 2 – yellow; 3 – dark	-	
uBil	bilirubin	µmol/L	
uLeu	leukocytes	leu/uL	
uNit	nitrites	neg/pos	

uOsmoll	osmolality	mOsm	
uProtein	total protein	g/L	
uGlu	glucose	mmol/L	
uHemogl	haemoglobin	ery/uL	
uKeton	ketone	mmol/L	
upH	pH	-	
uUrobili	urobilinogen	µmol/L	
Organs	Description; all as percentage of BodyWeight	Unit	Grace
Kidney	Percentage weight of kidney	%	X
Spleen	Percentage weight of spleen	%	X
Liver	Percentage weight of liver	%	X
AdrenGl	Percentage weight of adrenal gland	%	X
Heart	Percentage weight of heart	%	X
Thymus	Percentage weight of thymus	%	X
Testis	Percentage weight of testis (males)	%	X
Epididymis	Percentage weight of epididymis (males)	%	X
Uterus	Percentage weight of uterus (females)	%	X
Ovary	Percentage weight of ovary (females)	%	X
Brain	Percentage weight of brain	%	X
Immunology	Description	Unit	Grace
Granulocytes	Phagocytic activity of granulocytes	%	
RespirBurst	Respiratory burst of phagocytes	%	
Con	Proliferative activity of lymphocytes stimulated with mitogen Concanavalin A	cpm (counts per minute)	
PHA	Proliferative activity of lymphocytes stimulated with mitogen phytohaemmagglutinin	cpm	
PWM	Proliferative activity of lymphocytes stimulated with pokeweed mitogen	cpm	
Medium	Proliferative activity of non-stimulated lymphocytes	cpm	
IprConA	Ratio of proliferative activity of lymphocytes stimulated with Concanavalin A versus proliferative activity of non-stimulated lymphocytes	-	
IprPHA	Ratio of proliferative activity of lymphocytes stimulated with phytohaemmagglutinin versus proliferative activity of non-stimulated lymphocytes	-	
IprPWM	Ratio of proliferative activity of lymphocytes stimulated with pokeweed mitogen versus proliferative activity of non-stimulated lymphocytes	-	
Cytokines	Description	Unit	Grace
IL1a	Interleukin 1 alpha	pg/mL	
IL1b	Interleukin 1 beta	pg/mL	
IL2	Interleukin 2	pg/mL	
IL4	Interleukin 4	pg/mL	
IL5	Interleukin 5	pg/mL	
IL6	Interleukin 6	pg/mL	
IL10	Interleukin 10	pg/mL	
IL12p70	Interleukin 12p70	pg/mL	
IL13	Interleukin 13	pg/mL	
IL17A	Interleukin 17A	pg/mL	
GCSF	Granulocyte colony-stimulating factor	pg/mL	
GMCSF	Granulocyte-macrophage colony-stimulating factor	pg/mL	

TNFa	Tumour necrosis factor alpha	pg/mL	
IFNg	Interferon gamma	pg/mL	
CellPhenotype	Description	Unit	Grace
sp3	Spleen: percentage of cells, not labelled with monoclonal antibody anti-rat CD3	%	
sp3_4	Spleen: percentage of T-cytotoxic lymphocytes, cells double labelled with monoclonal antibodies anti-rat CD3 and anti-rat CD4	%	
sp3_8	Spleen: percentage of T-helper lymphocytes, cells double labelled with monoclonal antibodies anti-rat CD3 and anti-rat CD8	%	
sp3_45	Spleen: percentage of B- lymphocytes, cells anti-rat CD3 antigen negative and anti-rat CD45R antigen positive	%	
sp3_161	Spleen: percentage of NK-cells, cells anti-rat CD3 antigen negative and anti-rat CD161 antigen positive	%	
ln3	Lymph node: percentage of cells, not labelled with monoclonal antibody anti-rat CD3	%	
ln3_4	Lymph node: percentage of T-cytotoxic lymphocytes, cells double labelled with monoclonal antibodies anti-rat CD3 and anti-rat CD4	%	
ln3_8	Lymph node: percentage of T-helper lymphocytes, cells double labelled with monoclonal antibodies anti-rat CD3 and anti-rat CD8	%	
ln3_45	Lymph node: percentage of B- lymphocytes, cells anti-rat CD3 antigen negative and anti-rat CD45R antigen positive	%	
ty3	Thymus: percentage of cells, not labelled with monoclonal antibody anti-rat CD3	%	
ty3_4	Thymus: percentage of T-cytotoxic lymphocytes, cells double labelled with monoclonal antibodies anti-rat CD3 and anti-rat CD4	%	
ty3_8	Thymus: percentage of T-helper lymphocytes, cells double labelled with monoclonal antibodies anti-rat CD3 and anti-rat CD8	%	
Hormone	Description	Unit	Grace
Testosteron	Testosteron (males)	ng/ml	
betaEstr	17β – Estradiol (females)	pg/ml	
T3	Triiodothyronine hormone	nmol/L	
T4	Thyroxine hormone	nmol/L	
eCycle	Description (females)	Unit	Grace
Cycle1	Duration of 1st oestrous cycle (missing = unfinished)	days	
Cycle2	Duration of 2nd oestrous cycle (missing = unfinished)	days	
Regular	Number of regular cycles out of 2 cycles	count	
Irregular	Number of irregular cycles out of 2 cycles	count	

The variables in Table 3 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. No deviations from normal findings were observed in experimental groups and control group of rats, except for 2 animals: (1) male animal number 41 (feeding group NK33-) was euthanized in week 7 following paresis of hind limbs, and (2) female animal number (control feeding group) had oedema on the front left limb caused mechanically by strangulation on the lid; it is assumed that this is not connected to the experiment.
- Total and anti-maize/CP4 EPSPS specific antibody levels (IgG, IgM, IgE). These were not determined because there is no commercially available ELISA kit to do so. To be able to develop a new ELISA test system, it would have been necessary to have the purified CP4 EPSPS protein, and there was no trustful source that was independent of any plant biotechnology company.
- Organ weights: sternum with bone marrow, thyroid, parathyroid. These were not determined and this is in line with the OECD Test Guideline 408, which does not foresee such measurements. These measurement were erroneously included in the original study plan.

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 identical; see section 4.3.1 for details. All variables 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 are 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 the GRACE studies 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 five studies were conducted with several control (or reference) feeds as given in Table 4, see Schmidt et al (2017).

Table 4 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 GRACE data 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 Appendix 8. In Table 3 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 Appendix 1, which resulted in two Excel data files with combined data: one for male rats and one for female rats.

3.1 Growth curves and feed intake

For each individual rat an exponential growth curve $A + B\ R^{Week}$ was fitted to the observed weights. A re-parameterization of this curve is given by $A + B\ \exp(-\gamma\ Week)$ with the growth rate γ defined by $\gamma = -\log(R)$. In Appendix 2 the observed weights are graphically displayed along with the fitted curve and, in the bottom right corner, the resulting estimate of the growth rate γ. In case an absolute standardized residual is larger than 3, the accompanying weight is set to missing and the curve is fitted again. Such weight values, the re-estimated growth curve and the re-estimated growth rate γ are given in red in Appendix 2. This results in two sets of parameters which are both given in the combined Excel data files. Note that only when there are standardized residuals larger than 3 the two sets of parameters are different. SZU remarked on the weights with large residuals that : "*Weights were checked again, no wrong values (no typing errors) were indicated. Explanation for "outliers" might be: loss of appetite of particular rat, fighting between animals, irregular stool etc*". It was therefore decided to keep the weights with the large residuals. Note that in general the exponential curve fits very well and it is therefore decided to only analyse the final weight observed after week 13, further called BodyWeight, and the estimated growth rate γ, further called growthRate.

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. Feed consumption for each cage in units g/animal/day is graphically depicted in Appendix 3. The mean feed consumption for each feeding group is given in Figure 3.

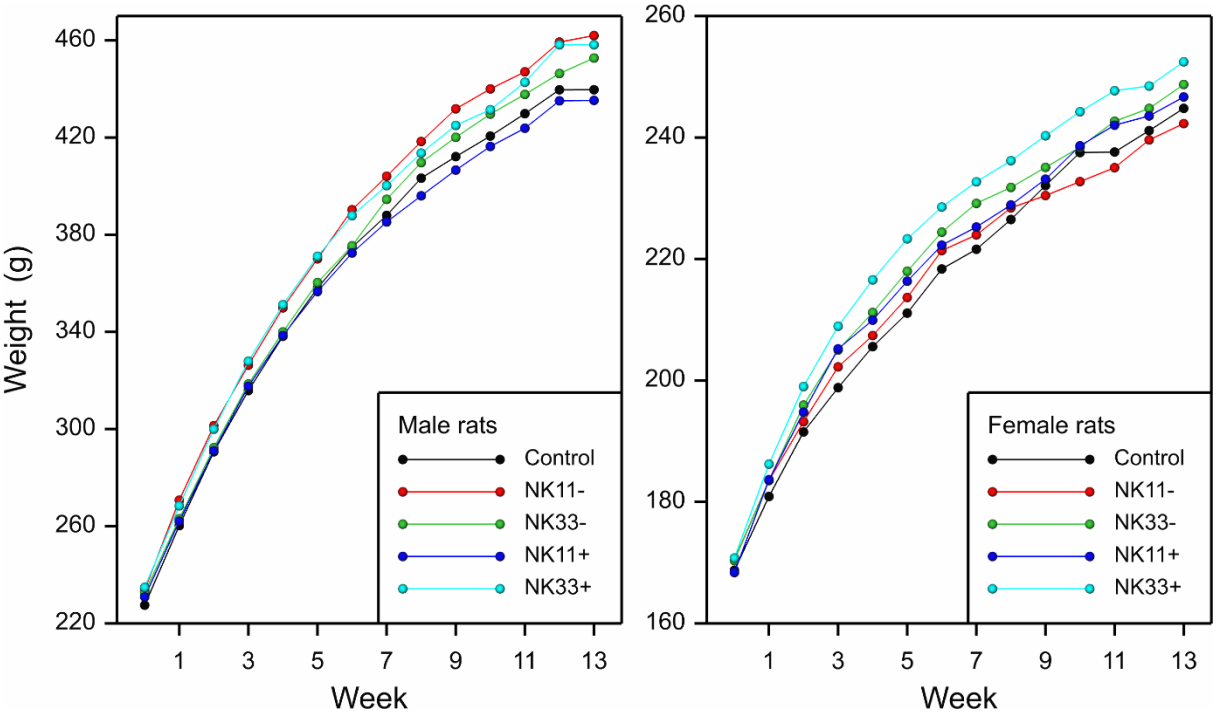

Figure 1 **Mean body weights versus week for each feeding group for male rats (left) and for female rats (right).**

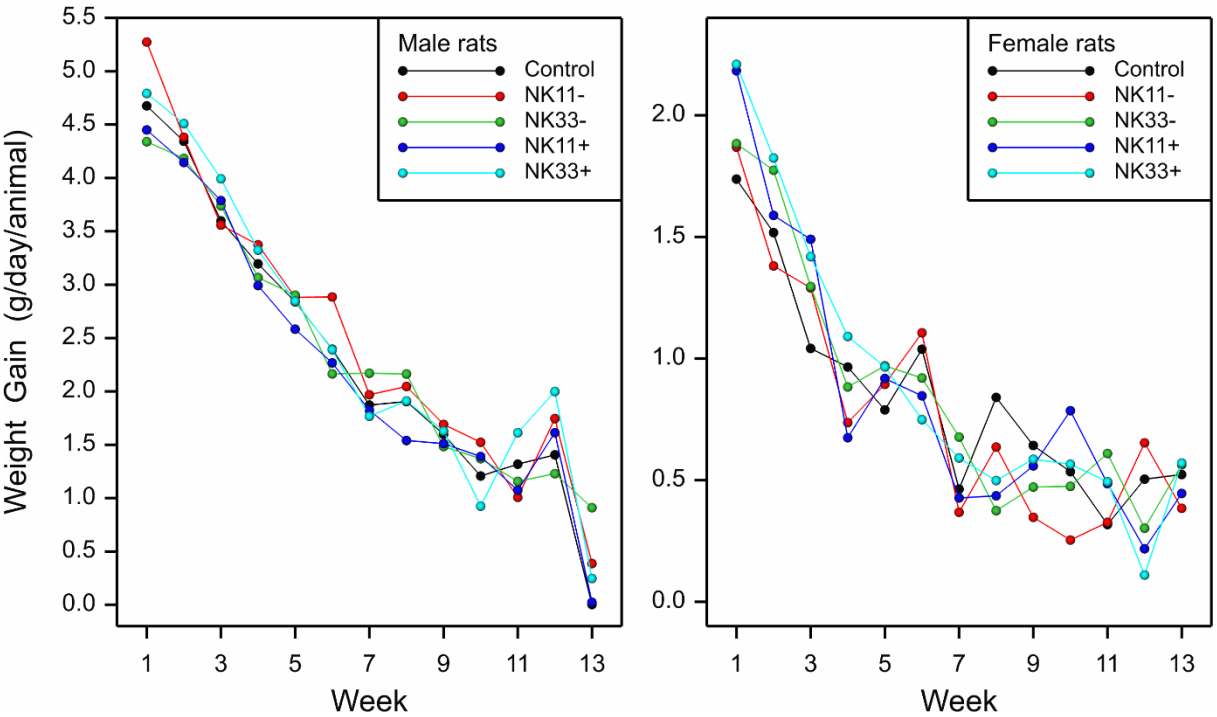

Figure 2 **Mean body weights gain (g/day/animal) versus week for each feeding group for male rats (left) and for female rats (right).**

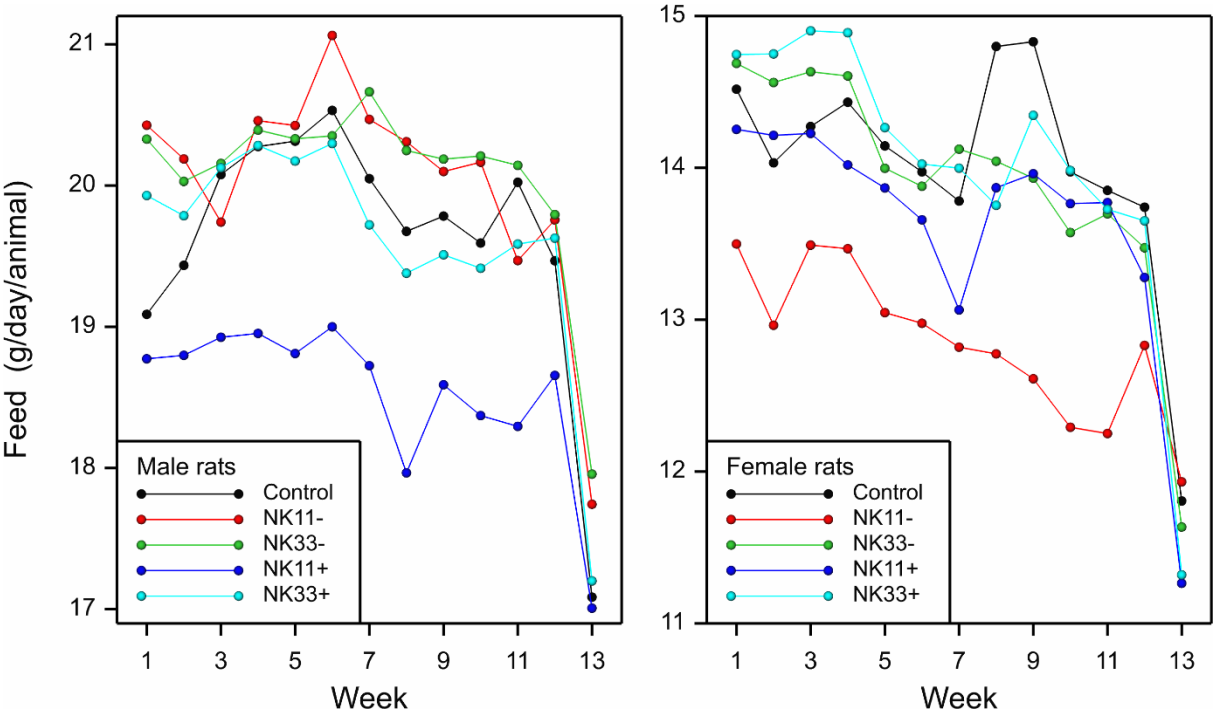

Figure 3 Mean feed consumption (g/day/animal) versus week for each feeding group for male rats (left) and for female rats (right).

3.2 Outliers and checking of ANOVA assumptions

The cage means, after log transformation, for each observed variable are statistically analysed by means of an analysis of variance using the model "Block + Group" according to the randomized block design. Grubbs' outlier test at the 1% level was applied to the residuals to detect outliers. This resulted in six outliers which are depicted by the solid symbols in Figure 4. Table 5 lists the outliers along with the two values that make up the outlying cage mean on the log scale. Only the bold values are considered to be outliers because they represent rather unusual values. The individual values for Lymphocytes and for Neutrophils are not very unusual; they are only picked up by Grubbs' outlier test because the mean is somewhat unusual. Note that Lymphocytes and Neutrophils are correlated because the sum of the WBC values is 200. This implies that e.g. an unusual large Lymphocytes count for an individual must be accompanied with a smaller Neutrophils count, such that the sum of the WBC values remains 200 for that individual.

The outlying values in Table 5 were presented to SZU who remarked that "*Male HGB – repeated measurements, no technical problems, Differential white blood cell count – % Lymphocytes, neutrophils animals No- 53,54 – method – light microscopy – values checked again, no technical problems detected*". SZU did not comment on the bold values for Liver (animal 37), Brain (animals 31 and 73). So given that there were no technical problems with the outliers in Table 5, the potential outliers were presented to the toxicological expert and G-TwYST coordinator (Pablo Steinberg, Max Rubner Institut) who confirmed the choice of outliers.

Outlier summary: in addition to the values that were set to missing because there were technical problems, see Appendix 2, the bold values in Table 5 were also set to missing.

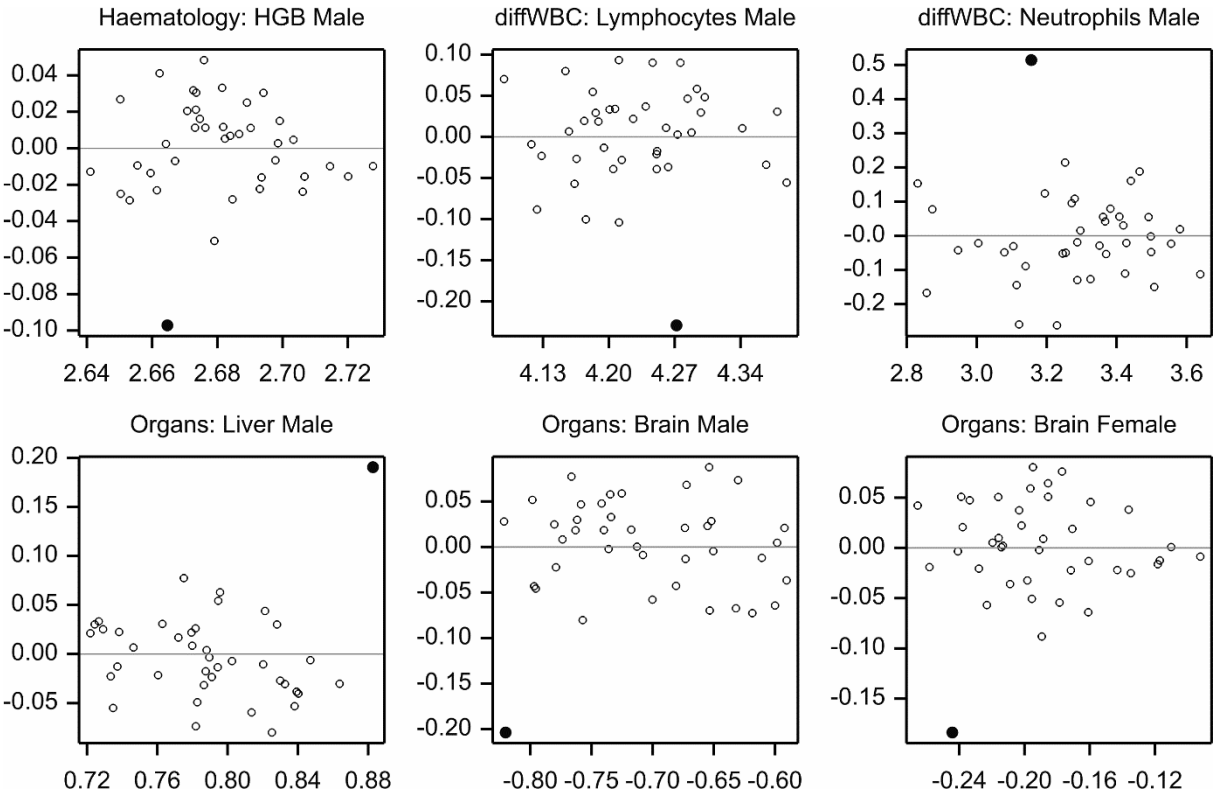

Figure 4 **Residuals along the y-axis versus fitted values along the x-axis resulting from an analysis of variance on log transformed cage means. Solid symbols denote outliers found by Grubbs' outlier test at the 1% level applied to the residuals.**

Table 5 **Cage and animal numbers for which Grubbs' outlier test is significant at the 1% level along with the two values that make up the cage mean. Only the bold values are finally considered to be outliers.**

Male/Female	Variable	Cage	Animal	Log-Value	Value
Male	HGB	7	13	2.70	14.9
Male	***HGB***	***7***	***14***	***2.43***	***11.4***
Male	***Liver***	***19***	***37***	***1.24***	***3.462***
Male	Liver	19	38	0.90	2.471
Male	***Brain***	***16***	***31***	***-1.20***	***0.301***
Male	Brain	16	32	-0.85	0.429
Male	Lymphocytes	27	53	4.71	111
Male	Lymphocytes	27	54	4.76	117
Male	Neutrophils	27	53	4.38	80
Male	Neutrophils	27	54	4.34	77
Female	Brain	73	159	-0.12	0.885
Female	***Brain***	***73***	***160***	***-0.73***	***0.480***

Without these outliers cage means on the original scale are given in Appendix 4, while cage means after a log transformation of the individual data are given in Appendix 5. Normal probability plots of

the ANOVA residuals, of an analysis on cage means after log transformation, are given in Appendix 6. To aid interpretation a 99% envelope is added to the probability plots, such that only values outside the envelop might be suspicious. Appendix 7 gives plots of residuals versus fitted values after the same analysis of variance. These residual plots are generally satisfactory implying that the ANOVA assumptions, homogeneity of variance and less importantly normality, are generally fulfilled.

3.3 Summary tables

Summary tables, on the original non-transformed scale, of number of observations, means, standard deviations and coefficients of variation (%), classified by the feeding groups, are given in Table 6 for males and in Table 7 for females. These tables were obtained by first calculating cage means and then calculating the summary statistics. The number of cages per feeding group is generally 8, except for the Immunology variables with 5 cages, the Cytokine variables with 5 cages for males and 3 cages for females, the CellPhenotype variables with 5 cages and the Hormone data with 6 cages for males and 8 cages for females.

Table 6 Summary statistics for male rats classified by the feeding groups: number of cages (N), means (Mean), standard deviations (Sd) and coefficients of variation (CV). The summary statistics are obtained from cage means.

Weights	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
BodyWeight	8	440	29.8	6.8	8	462	38.6	8.4	8	449	29.0	6.5	8	435	24.8	5.7	8	458	49.2	10.7
growthRate	8	0.14	0.025	17.0	8	0.14	0.016	12.0	8	0.13	0.027	20.4	8	0.15	0.034	23.4	8	0.14	0.017	11.9
FeedMean	8	19.6	1.34	6.8	8	20.0	1.86	9.3	8	20.0	1.04	5.2	8	18.5	0.61	3.3	8	19.6	1.88	9.6
Haematology	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
WBC	8	9.10	1.92	21.1	8	8.85	1.58	17.8	8	9.42	1.63	17.3	8	9.75	1.22	12.5	8	9.57	1.04	10.9
RBC	8	7.79	0.30	3.8	8	7.77	0.44	5.6	8	7.77	0.27	3.4	8	7.67	0.35	4.6	8	7.83	0.20	2.6
HGB	8	14.6	0.41	2.8	8	14.7	0.29	2.0	8	14.7	0.44	3.0	8	14.4	0.55	3.8	8	14.8	0.18	1.2
HCT	8	42.4	1.21	2.8	8	42.8	1.96	4.6	8	43.0	1.40	3.3	8	42.4	1.60	3.8	8	43.2	0.84	2.0
MCV	8	54.4	1.07	2.0	8	55.2	1.44	2.6	8	55.4	1.16	2.1	8	55.4	1.38	2.5	8	55.1	0.63	1.1
MCH	8	18.8	0.46	2.5	8	18.7	0.52	2.8	8	19.0	0.69	3.6	8	18.8	0.78	4.2	8	19.0	0.45	2.4
MCHC	8	34.6	0.28	0.8	8	33.9	0.94	2.8	8	34.2	0.78	2.3	8	33.9	0.59	1.7	8	34.4	0.76	2.2
PLT	8	842	75	8.9	8	792	120	15.1	8	862	125	14.5	8	851	83	9.7	8	847	115	13.6
LYMR	8	73.0	3.74	5.1	8	72.5	4.97	6.9	8	70.9	5.89	8.3	8	72.0	3.41	4.7	8	73.4	5.06	6.9
LYMA	8	6.64	1.55	23.3	8	6.42	1.24	19.3	8	6.68	1.34	20.0	8	7.00	0.70	10.0	8	7.04	1.09	15.4
diffWBC	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Lymphocytes	8	71.4	6.45	9.0	8	68.1	6.48	9.5	8	65.8	5.86	8.9	8	69.8	6.53	9.4	8	70.6	6.28	8.9
Neutrophils	8	25.7	6.58	25.6	8	28.8	7.01	24.3	8	30.5	5.17	16.9	8	26.8	6.83	25.5	8	26.4	6.60	25.0
Monocytes	8	1.50	0.46	30.9	8	1.66	0.44	26.7	8	1.78	0.92	51.7	8	1.53	0.41	26.8	8	1.31	0.42	31.8
Eosinophils	8	1.44	0.58	40.3	8	1.41	0.72	51.1	8	1.84	1.32	71.7	8	1.91	0.90	47.0	8	1.72	0.78	45.6
ClinChem	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
ALP	8	1.25	0.14	11.3	8	1.28	0.18	13.9	8	1.27	0.14	10.7	8	1.25	0.13	10.2	8	1.44	0.19	13.2
ALT	8	0.56	0.040	7.1	8	0.57	0.061	10.6	8	0.55	0.077	13.9	8	0.50	0.058	11.5	8	0.56	0.080	14.3
AST	8	2.40	0.60	24.9	8	2.30	0.47	20.6	8	2.23	0.36	16.2	8	2.12	0.38	17.8	8	2.22	0.45	20.3
BIL	8	7.30	1.97	27.0	8	7.08	0.98	13.8	8	6.92	0.42	6.1	8	7.16	0.82	11.4	8	6.48	0.34	5.3
ALB	8	36.4	0.81	2.2	8	36.6	1.77	4.8	8	36.9	0.79	2.1	8	37.8	0.99	2.6	8	36.4	1.30	3.6
TP	8	64.4	1.13	1.7	8	65.1	2.32	3.6	8	65.0	1.08	1.7	8	66.2	1.66	2.5	8	64.6	1.36	2.1

Glu	8	5.38	0.85	15.8	8	5.34	0.82	15.3	8	5.02	0.69	13.7	8	5.15	0.36	6.9	8	5.49	0.53	9.6
CHOL	8	2.05	0.23	11.2	8	2.01	0.27	13.5	8	2.02	0.22	10.8	8	1.94	0.19	10.0	8	2.04	0.17	8.5
TAG	8	1.02	0.25	24.6	8	1.10	0.14	12.8	8	1.05	0.33	31.3	8	1.03	0.31	30.5	8	1.09	0.39	35.9
Crea	8	40.1	2.09	5.2	8	41.5	4.78	11.5	8	40.8	3.61	8.8	8	43.8	4.66	10.6	8	41.9	4.15	9.9
Urea	8	5.13	0.32	6.3	8	4.90	0.55	11.2	8	4.99	0.28	5.6	8	4.88	0.43	8.9	8	4.96	0.50	10.0
cHGB	8	69.7	46.1	66.1	8	61.1	20.6	33.7	8	58.4	12.8	21.8	8	52.6	15.5	29.5	8	49.4	13.5	27.4
Ca	8	2.40	0.040	1.7	8	2.40	0.030	1.2	8	2.41	0.056	2.3	8	2.41	0.045	1.9	8	2.39	0.036	1.5
Cl	8	102	1.33	1.3	8	102	1.19	1.2	8	102	1.03	1.0	8	102	0.70	0.7	8	102	1.22	1.2
K	8	4.99	0.23	4.5	8	5.09	0.37	7.2	8	5.25	0.20	3.9	8	5.22	0.31	5.9	8	5.11	0.24	4.8
Na	8	145	1.65	1.1	8	145	1.46	1.0	8	145	1.16	0.8	8	146	1.12	0.8	8	145	1.39	1.0
P	8	2.37	0.31	13.1	8	2.44	0.19	7.9	8	2.50	0.15	6.1	8	2.41	0.16	6.5	8	2.31	0.18	7.7
Urine	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
uVol	8	19.2	3.58	18.6	8	19.1	5.53	28.9	8	22.3	7.15	32.0	8	21.1	8.59	40.7	8	19.2	5.09	26.5
uVolW	8	4.59	0.95	20.7	8	4.31	1.16	26.9	8	5.17	1.69	32.6	8	5.05	2.01	39.7	8	4.39	1.23	28.1
uLeu	8	28.1	35.8	127.4	8	20.3	13.3	65.3	8	31.2	34.1	109.0	8	26.6	35.6	134.1	8	18.8	9.4	50.4
uOsmoll	8	489	148	30.2	8	474	94	19.9	8	448	108	24.2	8	455	139	30.5	8	480	90	18.8
uKeton	8	0.42	0.31	73.9	8	0.78	0.41	52.5	8	0.75	0.93	123.4	8	0.53	0.65	121.8	8	0.31	0.55	175.0
upH	8	7.12	0.30	4.2	8	7.00	0.40	5.7	8	6.88	0.13	1.9	8	6.94	0.32	4.6	8	6.75	0.19	2.8
Organs	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Kidney	8	0.52	0.038	7.2	8	0.52	0.028	5.4	8	0.54	0.029	5.4	8	0.55	0.038	6.9	8	0.52	0.041	7.9
Spleen	8	0.18	0.015	8.2	8	0.19	0.010	5.6	8	0.18	0.016	9.3	8	0.19	0.015	7.8	8	0.18	0.015	8.1
Liver	8	2.16	0.10	4.7	8	2.19	0.13	6.0	8	2.24	0.13	5.6	8	2.24	0.10	4.5	8	2.17	0.04	1.8
AdrenGl	8	0.014	0.0013	9.1	8	0.013	0.0020	15.4	8	0.013	0.0015	11.3	8	0.013	0.0015	11.6	8	0.013	0.0014	11.0
Heart	8	0.23	0.007	2.8	8	0.24	0.009	3.6	8	0.24	0.014	5.9	8	0.24	0.015	6.2	8	0.23	0.013	5.7
Thymus	8	0.09	0.012	13.7	8	0.10	0.017	17.4	8	0.10	0.017	17.5	8	0.10	0.018	17.7	8	0.10	0.017	17.3
Testis	8	0.81	0.084	10.3	8	0.80	0.054	6.8	8	0.82	0.038	4.6	8	0.83	0.054	6.5	8	0.81	0.063	7.8
Epididymis	8	0.27	0.024	8.9	8	0.26	0.011	4.3	8	0.28	0.020	7.1	8	0.27	0.016	5.9	8	0.27	0.030	11.0
Brain	8	0.49	0.041	8.4	8	0.48	0.034	7.0	8	0.51	0.021	4.1	8	0.51	0.031	6.1	8	0.49	0.054	11.0
Immunology	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Granulocytes	5	73.8	13.0	17.7	5	77.5	9.6	12.4	5	73.6	12.3	16.8	5	73.7	14.3	19.3	5	73.5	11.8	16.1
RespirBurst	5	64.4	12.6	19.6	5	66.3	9.9	14.9	5	60.1	11.8	19.6	5	60.6	15.7	26.0	5	63.2	9.4	14.8

Con	5	77166	21348	27.7	5	78118	16358	20.9	5	70681	15091	21.4	5	73844	16697	22.6	5	83659	17987	21.5
PHA	5	50436	17825	35.3	5	51083	6848	13.4	5	50736	10519	20.7	5	47416	9575	20.2	5	53347	7981	15.0
PWM	5	25656	15280	59.6	5	36631	11884	32.4	5	25415	14163	55.7	5	25404	10910	42.9	5	46487	29478	63.4
Medium	5	2370	902	38.1	5	2324	554	23.9	5	1944	601	30.9	5	1812	242	13.3	5	3003	1407	46.9
IprConA	5	34.8	8.7	25.0	5	37.4	7.4	19.7	5	40.1	6.1	15.2	5	41.5	14.2	34.1	5	34.7	15.0	43.3
IprPHA	5	21.8	2.30	10.5	5	26.6	4.79	18.0	5	28.2	2.75	9.7	5	27.2	9.58	35.2	5	21.8	9.36	42.9
IprPWM	5	10.5	3.04	29.0	5	15.3	3.30	21.6	5	12.7	3.36	26.4	5	13.9	5.69	40.8	5	15.1	4.55	30.0
Cytokines	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
IL2	5	4082	936	22.9	5	5378	496	9.2	5	4499	658	14.6	5	4477	960	21.4	5	5501	853	15.5
IL4	5	28.7	24.2	84.4	5	14.6	5.1	34.9	5	10.9	0.7	6.9	5	11.3	1.7	15.3	5	14.1	2.0	14.2
IL10	5	10265	8214	80.0	5	9151	7329	80.1	5	3256	1579	48.5	5	3677	3005	81.7	5	13753	13026	94.7
IL17A	5	184	121	66.0	5	146	47	32.6	5	119	48	39.9	5	98	46	46.7	5	198	118	59.4
TNFa	5	17.9	7.25	40.5	5	25.1	1.74	6.9	5	17.3	2.92	16.9	5	19.2	5.85	30.4	5	26.0	7.08	27.2
IFNg	5	14401	2655	18.4	5	19549	4549	23.3	5	12162	1740	14.3	5	13601	3607	26.5	5	16342	3902	23.9
CellPhenotype	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
sp3	5	36.0	5.20	14.4	5	41.7	6.15	14.8	5	41.0	8.66	21.1	5	43.8	4.37	10.0	5	43.0	6.35	14.8
sp3_4	5	21.4	3.84	17.9	5	24.2	5.14	21.2	5	26.2	6.70	25.6	5	25.6	4.03	15.7	5	26.4	6.25	23.7
sp3_8	5	14.7	3.05	20.7	5	17.1	2.30	13.4	5	14.9	4.08	27.3	5	18.1	1.42	7.9	5	16.5	1.96	11.9
sp3_45	5	24.9	1.07	4.3	5	25.6	0.48	1.9	4	24.7	1.35	5.5	5	24.6	1.34	5.4	5	25.2	0.72	2.9
sp3_161	5	13.6	1.15	8.4	5	13.1	0.76	5.8	4	13.8	1.00	7.3	5	13.5	0.86	6.4	5	12.5	0.71	5.7
ln3	5	57.8	4.0	7.0	5	59.2	3.9	6.6	5	53.9	11.3	20.9	5	52.7	12.6	23.9	5	54.6	10.5	19.2
ln3_4	5	42.7	3.1	7.2	5	41.1	2.1	5.0	5	39.5	9.4	23.7	5	37.4	10.7	28.6	5	39.5	9.6	24.2
ln3_8	5	15.5	1.31	8.5	5	17.9	2.06	11.5	5	15.0	1.95	13.0	5	16.2	2.68	16.5	5	15.6	1.21	7.8
ln3_45	5	34.5	8.1	23.4	5	29.3	9.3	31.9	5	29.1	7.8	26.8	5	34.0	11.8	34.5	5	30.8	8.2	26.5
ty3	5	18.4	4.66	25.3	5	18.7	3.72	20.0	5	19.3	2.40	12.5	5	20.0	2.27	11.4	5	20.2	2.94	14.5
ty3_4	5	14.2	3.17	22.4	5	14.4	2.60	18.1	5	15.1	2.38	15.7	5	15.5	1.40	9.1	5	15.9	2.27	14.3
ty3_8	5	7.93	1.71	21.6	5	8.39	1.98	23.6	5	8.24	0.68	8.3	5	8.35	1.06	12.7	5	9.00	1.62	18.0
Hormone	Control				NK11-				NK33-				NK11+				NK33+			
Male	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Testosteron	6	2.58	0.88	34.3	6	2.78	1.64	59.1	6	1.97	1.12	56.6	6	3.03	1.93	63.7	6	2.17	0.56	25.6
T3	6	0.75	0.06	8.1	6	0.82	0.14	16.7	6	0.74	0.09	12.1	6	0.75	0.09	12.6	6	0.81	0.08	10.0
T4	6	55.2	4.74	8.6	6	54.0	5.40	10.0	6	48.7	3.65	7.5	6	55.1	3.77	6.8	6	53.4	5.29	9.9

Table 7 Summary statistics for female rats classified by the feeding groups: number of cages (N), means (Mean), standard deviations (Sd) and coefficients of variation (CV). The summary statistics are obtained from cage means.

Weights	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
BodyWeight	8	245	17.2	7.0	8	242	18.7	7.7	8	249	14.8	5.9	8	247	13.6	5.5	8	252	12.4	4.9
growthRate	8	0.12	0.032	26.5	8	0.16	0.056	35.5	8	0.15	0.038	24.8	8	0.15	0.047	30.6	8	0.18	0.023	12.8
FeedMean	8	14.0	1.01	7.2	8	12.8	1.51	11.7	8	13.9	1.17	8.4	8	13.6	0.58	4.2	8	14.0	0.92	6.5
Haematology	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
WBC	8	5.52	1.48	26.7	8	5.37	0.97	18.0	8	6.06	1.16	19.2	8	4.86	0.70	14.4	8	5.43	1.26	23.3
RBC	8	6.73	0.22	3.3	8	6.72	0.55	8.2	8	6.87	0.17	2.5	8	6.81	0.50	7.3	8	6.66	0.30	4.5
HGB	8	13.7	0.45	3.3	8	13.6	1.03	7.6	8	14.1	0.18	1.3	8	13.8	0.81	5.9	8	13.6	0.65	4.8
HCT	8	39.4	1.09	2.8	8	39.1	3.29	8.4	8	40.5	0.95	2.3	8	39.9	2.57	6.4	8	39.6	1.81	4.6
MCV	8	58.7	1.40	2.4	8	58.2	1.79	3.1	8	59.1	1.03	1.8	8	58.7	1.63	2.8	8	59.6	1.32	2.2
MCH	8	20.4	0.69	3.4	8	20.2	0.54	2.7	8	20.6	0.44	2.1	8	20.4	0.84	4.1	8	20.4	0.62	3.0
MCHC	8	34.8	0.83	2.4	8	34.7	0.75	2.2	8	34.8	0.64	1.8	8	34.7	0.93	2.7	8	34.3	0.47	1.4
PLT	8	902	113	12.6	8	825	144	17.5	8	901	91	10.1	8	813	100	12.3	8	904	94	10.4
LYMR	8	75.2	5.81	7.7	8	75.1	3.52	4.7	8	77.5	3.52	4.5	8	76.6	4.92	6.4	8	77.1	4.96	6.4
LYMA	8	4.22	1.37	32.4	8	4.04	0.69	17.2	8	4.69	0.91	19.4	8	3.71	0.69	18.5	8	4.15	0.97	23.4
diffWBC	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Lymphocytes	8	69.2	5.98	8.6	8	69.5	6.24	9.0	8	70.4	3.79	5.4	8	71.9	5.15	7.2	8	71.1	4.08	5.7
Neutrophils	8	28.1	4.73	16.8	8	28.5	5.76	20.2	8	26.8	4.46	16.6	8	26.1	5.02	19.2	8	26.1	4.18	16.1
Monocytes	8	1.53	0.76	49.7	8	1.12	0.65	58.2	8	1.44	0.83	57.9	8	1.12	0.38	33.6	8	1.25	0.57	45.4
Eosinophils	8	1.16	1.20	104.0	8	0.91	0.71	77.9	8	1.31	0.94	71.8	8	0.88	1.03	117.3	8	1.56	0.98	62.7
ClinChem	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
ALP	8	0.58	0.14	23.3	8	0.58	0.06	9.6	8	0.65	0.07	10.0	8	0.58	0.07	12.1	8	0.58	0.10	18.1
ALT	8	0.47	0.06	12.9	8	0.42	0.04	10.3	8	0.53	0.12	23.0	8	0.45	0.13	29.6	8	0.42	0.08	19.1
AST	8	2.30	0.67	29.0	8	2.22	0.34	15.5	8	2.33	0.25	11.0	8	2.21	0.53	24.1	8	2.37	0.44	18.7
BIL	8	7.57	0.97	12.8	8	7.11	0.50	7.1	8	8.13	1.12	13.8	8	7.84	0.97	12.4	8	8.51	2.06	24.2
ALB	8	42.5	3.53	8.3	8	41.5	2.48	6.0	8	44.6	2.71	6.1	8	44.1	2.64	6.0	8	43.9	2.84	6.5
TP	8	68.7	3.38	4.9	8	67.2	2.90	4.3	8	71.0	2.55	3.6	8	70.9	2.41	3.4	8	70.5	2.97	4.2

Glu	8	5.19	0.35	6.8	8	5.12	1.18	23.1	8	4.40	0.68	15.4	8	5.05	1.12	22.2	8	5.62	1.03	18.3
CHOL	8	1.91	0.21	11.0	8	2.05	0.28	13.9	8	2.01	0.28	14.0	8	2.00	0.25	12.3	8	1.99	0.34	16.9
TAG	8	0.52	0.13	24.7	8	0.62	0.20	31.9	8	0.65	0.09	13.4	8	0.60	0.12	19.6	8	0.57	0.09	15.1
Crea	8	40.9	2.43	5.9	8	42.0	2.84	6.8	8	42.9	3.98	9.3	8	40.8	5.06	12.4	8	41.9	5.67	13.5
Urea	8	5.86	0.35	6.0	8	5.94	0.85	14.4	8	5.32	0.67	12.6	8	5.29	0.64	12.1	8	5.59	0.57	10.2
cHGB	8	41.8	22.2	53.2	8	39.8	9.7	24.4	8	49.5	19.0	38.4	8	49.3	24.1	48.9	8	56.4	31.9	56.6
Ca	8	2.47	0.070	2.8	8	2.45	0.035	1.4	8	2.49	0.040	1.6	8	2.50	0.053	2.1	8	2.49	0.049	2.0
Cl	8	102	1.03	1.0	8	102	1.54	1.5	8	101	1.06	1.1	8	102	1.15	1.1	8	102	1.88	1.8
K	8	4.45	0.22	4.8	8	4.46	0.20	4.6	8	4.52	0.29	6.4	8	4.54	0.21	4.7	8	4.70	0.33	7.0
Na	8	145	1.94	1.3	8	144	1.22	0.8	8	145	1.16	0.8	8	145	1.79	1.2	8	145	2.68	1.9
P	8	2.04	0.23	11.3	8	1.96	0.23	11.7	8	2.15	0.36	16.8	8	1.93	0.21	11.1	8	1.96	0.22	11.4
Urine	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
uVol	8	17.1	4.23	24.7	8	17.1	4.52	26.5	8	16.8	7.06	42.2	8	16.7	6.55	39.2	8	13.6	2.99	22.1
uVolW	8	7.42	1.84	24.8	8	7.49	2.05	27.3	8	7.09	3.00	42.3	8	7.13	2.93	41.1	8	5.70	1.45	25.5
uLeu	8	10.9	10.4	95.4	8	7.8	9.3	119.0	8	6.2	9.4	151.2	8	20.3	24.0	118.3	8	10.9	17.0	155.0
uOsmoll	8	413	112	27.2	8	417	104	25.0	8	435	123	28.2	8	414	128	31.0	8	427	63	14.7
uKeton	8	0.16	0.19	119.0	8	0.03	0.09	282.8	8	0.16	0.27	169.7	8	0.03	0.09	282.8	8	0.00	0.00	-
upH	8	6.62	0.27	4.0	8	6.44	0.29	4.5	8	6.50	0.23	3.6	8	6.44	0.26	4.0	8	6.47	0.16	2.5
Organs	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Kidney	8	0.60	0.051	8.5	8	0.57	0.050	8.7	8	0.57	0.046	8.0	8	0.58	0.034	5.8	8	0.58	0.025	4.3
Spleen	8	0.24	0.018	7.2	8	0.23	0.028	12.4	8	0.24	0.022	9.0	8	0.22	0.022	10.0	8	0.25	0.018	7.4
Liver	8	2.44	0.14	5.8	8	2.55	0.19	7.6	8	2.50	0.32	12.8	8	2.53	0.17	6.6	8	2.39	0.08	3.5
AdrenGl	8	0.031	0.0033	10.5	8	0.028	0.0024	8.7	8	0.028	0.0018	6.2	8	0.028	0.0028	10.0	8	0.029	0.0017	5.8
Heart	8	0.32	0.017	5.4	8	0.31	0.041	13.0	8	0.31	0.019	6.2	8	0.31	0.021	6.7	8	0.32	0.032	9.9
Thymus	8	0.12	0.020	17.2	8	0.12	0.017	14.2	8	0.12	0.025	20.8	8	0.12	0.021	18.5	8	0.12	0.016	13.3
Uterus	8	0.25	0.049	19.5	8	0.26	0.083	32.4	8	0.24	0.048	19.6	8	0.25	0.058	22.8	8	0.31	0.053	16.8
Ovary	8	0.038	0.0026	6.9	8	0.034	0.0028	8.4	8	0.035	0.0057	16.1	8	0.036	0.0028	7.8	8	0.032	0.0012	3.7
Brain	8	0.84	0.057	6.8	8	0.85	0.040	4.7	8	0.83	0.045	5.4	8	0.83	0.052	6.3	8	0.82	0.024	2.9
Immunology	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
Granulocytes	5	76.8	3.18	4.1	5	71.0	5.87	8.3	5	77.7	3.24	4.2	5	73.5	3.14	4.3	5	73.1	3.65	5.0
RespirBurst	5	56.3	13.2	23.5	5	52.4	1.8	3.4	5	54.5	6.1	11.2	5	52.3	11.7	22.3	5	57.9	7.8	13.5

Con	5	51683	19272	37.3	5	61830	17753	28.7	5	54694	25288	46.2	5	65130	25868	39.7	5	63123	20824	33.0
PHA	5	29498	13084	44.4	5	36438	17117	47.0	5	27484	15738	57.3	5	34743	10510	30.3	5	36871	8194	22.2
PWM	5	18769	15085	80.4	5	23672	6233	26.3	5	18101	9659	53.4	5	22491	10344	46.0	5	25289	15832	62.6
Medium	5	1429	545	38.1	5	1685	219	13.0	5	1422	646	45.4	5	1788	533	29.8	5	1870	695	37.2
IprConA	5	38.0	10.0	26.3	5	37.5	10.9	29.0	5	37.8	7.9	20.8	5	36.6	5.8	15.8	5	36.0	4.9	13.7
IprPHA	5	21.4	5.57	26.0	5	21.8	9.32	42.8	5	19.9	5.56	28.0	5	21.7	7.81	36.0	5	21.7	6.15	28.4
IprPWM	5	11.4	5.42	47.7	5	13.6	3.75	27.5	5	12.7	2.56	20.2	5	12.2	2.60	21.3	5	13.2	5.55	42.1
Cytokines	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
IL2	3	3821	583	15.2	3	4491	285	6.3	4	3240	1267	39.1	3	3936	243	6.2	4	3434	1320	38.4
IL4	3	14.2	2.0	14.3	3	12.1	2.1	17.6	4	9.5	4.8	50.0	3	25.9	14.3	55.2	4	11.1	7.0	62.9
IL10	3	10919	11461	105.0	3	9553	9833	102.9	4	4401	2669	60.6	3	11809	8515	72.1	4	8175	11715	143.3
IL17A	3	201	43.4	21.5	3	115	28.1	24.4	4	103	55.3	53.6	3	185	49.9	27.0	4	105	69.1	66.1
TNFa	3	19.9	5.19	26.2	3	15.6	1.08	6.9	4	12.2	5.86	48.0	3	21.7	1.02	4.7	4	13.7	6.41	46.8
IFNg	3	12364	4658	37.7	3	9788	2098	21.4	4	8656	5507	63.6	3	13476	1975	14.7	4	9123	6293	69.0
CellPhenotype	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
sp3	5	37.1	2.17	5.8	5	41.8	3.75	9.0	5	38.6	4.84	12.5	5	39.7	5.88	14.8	5	41.1	3.56	8.7
sp3_4	5	20.6	1.42	6.9	5	24.9	3.97	15.9	5	21.7	5.59	25.8	5	24.0	5.11	21.3	5	23.2	2.81	12.1
sp3_8	5	16.1	0.91	5.6	5	15.7	1.57	10.0	5	15.5	0.74	4.8	5	15.2	1.68	11.1	5	17.2	1.99	11.6
sp3_45	5	25.6	0.36	1.4	5	25.7	0.63	2.5	5	25.5	0.73	2.9	5	25.1	0.26	1.0	5	25.7	0.87	3.4
sp3_161	5	11.1	0.61	5.5	5	10.9	0.31	2.8	5	11.0	0.57	5.2	5	10.9	0.11	1.0	5	11.3	1.08	9.6
ln3	5	61.4	2.56	4.2	5	61.1	4.69	7.7	5	61.9	2.81	4.5	5	61.9	3.97	6.4	5	64.8	2.82	4.4
ln3_4	5	44.8	2.48	5.5	5	45.5	4.26	9.4	5	45.1	3.99	8.9	5	46.1	3.95	8.6	5	47.0	2.77	5.9
ln3_8	5	17.0	0.61	3.6	5	16.0	1.27	7.9	5	16.9	1.74	10.3	5	16.3	0.58	3.6	5	18.1	1.91	10.5
ln3_45	5	20.4	6.90	33.8	5	25.6	9.26	36.1	5	21.7	7.71	35.4	5	22.0	8.27	37.5	5	18.4	5.49	29.9
ty3	5	20.0	4.11	20.5	5	19.7	3.50	17.8	5	20.0	2.08	10.4	5	20.6	4.79	23.2	5	19.9	3.56	17.9
ty3_4	5	15.2	3.75	24.6	5	15.3	3.03	19.8	5	15.8	1.87	11.9	5	15.4	3.79	24.6	5	14.8	2.47	16.7
ty3_8	5	8.71	2.49	28.6	5	7.94	1.71	21.5	5	8.33	2.61	31.3	5	8.64	2.40	27.8	5	8.11	2.32	28.7
Hormone	Control				NK11-				NK33-				NK11+				NK33+			
Female	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV	N	Mean	Sd	CV
betaEstr	8	7.74	4.51	58.3	8	4.98	3.45	69.4	8	4.63	1.35	29.2	8	5.04	2.08	41.1	8	4.64	1.29	27.8
T3	8	0.76	0.16	21.4	8	0.77	0.12	15.2	8	0.69	0.08	11.9	8	0.73	0.13	18.1	8	0.63	0.12	18.8
T4	8	34.2	7.9	23.2	8	31.0	8.5	27.4	8	34.0	11.4	33.4	8	37.8	9.0	24.0	8	25.4	8.5	33.4

4 Statistical analysis

4.1 Equivalence testing using historical data

4.1.1 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, employing the historic GRACE studies, calculation of 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 9.

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 Appendix 10.

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 $100(1-\alpha)$% confidence interval for the difference in the current study, and a probability β which defines the desired power $1-\beta$ for the equivalence test. We used values $n_0 = 8$, $\alpha = 0.05$ and $\beta = 0.05$ resulting in a power of 0.95. Note that the regulatory sample size $n_0 = 8$ equals the replication, i.e. the number of cages, for most variables in both the GRACE and the G-TwYST studies. 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 $n_0 = 8$ which implies 14 degrees of freedom for residual in the

current study. The current G-TwYST study indeed has replication 8 but has 28 degrees of freedom for residual.

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). For further interpretation, 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.

4.1.2 Results

Each GMO feed was tested for equivalence with respect to the control feed. The DWE intervals showing the main results of the equivalence tests for 42 variables (with two variables only for males, and two others only for females) are given in Figure 5 to Figure 8. For further interpretation the 95% confidence intervals for the ratios are given in Table 8 and Table 9, and these intervals and limits at the ratio scale are given in Figure 9 to Figure 16.

The DWE equivalence test depends, among other things, on the ratio of the residual variance of the current study and the residual variance of the historical studies. In case this ratio is small the corresponding DWE interval will generally be short. The ratio of the residual variances is given in Figure 17. Small ratios are observed for e.g. Ca, Cl, and Na for both sexes. Large ratios are observed for Lymphocytes in both sexes, and HCT, RBC, Eosinophils and some organ weights in females.

Among 320 equivalence tests, there were 19 failures (6%) to prove equivalence (i.e. reject the hypothesis of non-equivalence), which is close to the 5% level of the test. In all these 19 cases the median estimate was within the equivalence limits, therefore equivalence is still more likely than lack of equivalence according to the terminology of EFSA (2011a). These 19 cases are observed for Lymphocytes in males (4x), while the remaining 15 cases are observed in females for Lymphocytes (4x), HCT (4x), Eosinophils (2x), kidney weight (2x), RBC (1x), uterus weight (1x) and growthRate (1x). From Figure 17 it can be seen that these are all cases where the G-TwYST study B was less precise than the historical studies on average (residual variance 1.5 to 3 times higher than in the historical datasets).

Although not the primary result of the equivalence analysis, it can also be observed from the graphs and tables that, for those variables for which the equivalence test is performed, the number of significant differences, employing t-tests, equals 25 (8% of 320 difference tests), which is again close to the 5% level of the test. Only in three of these cases (Lymphocytes males NK33- and growthRate females NK11- and NK33+) there was both a significant difference and a failure to show equivalence.

For all 632 difference tests, i.e. including those for which the equivalence test was not performed, 55 t-tests were significant which is 9% of the tests.

.

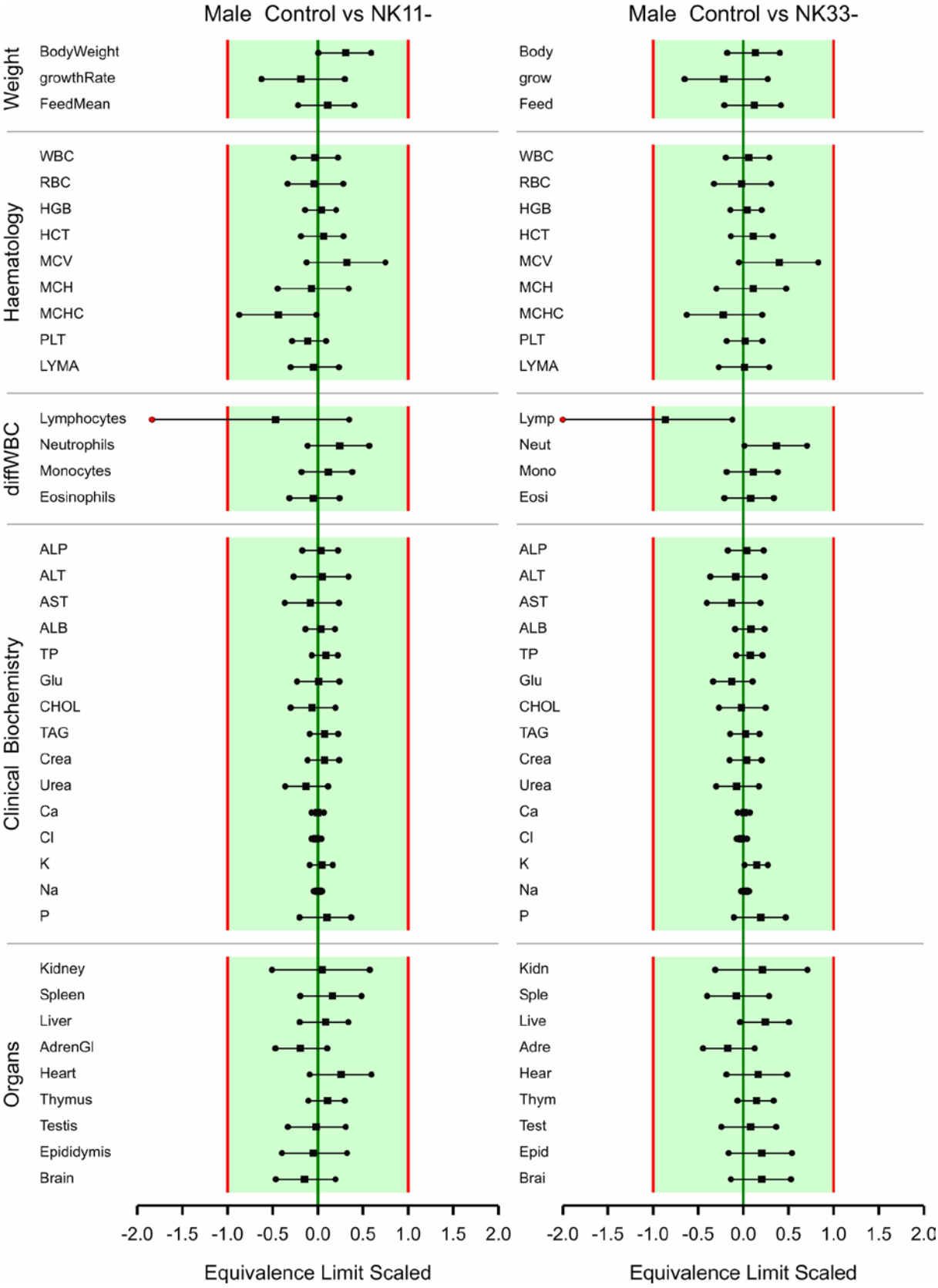

Figure 5 Equivalence testing of NK11- and NK33- versus the control feed for males. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. See Table 8, Figure 9 and Figure 10 for further interpretation.

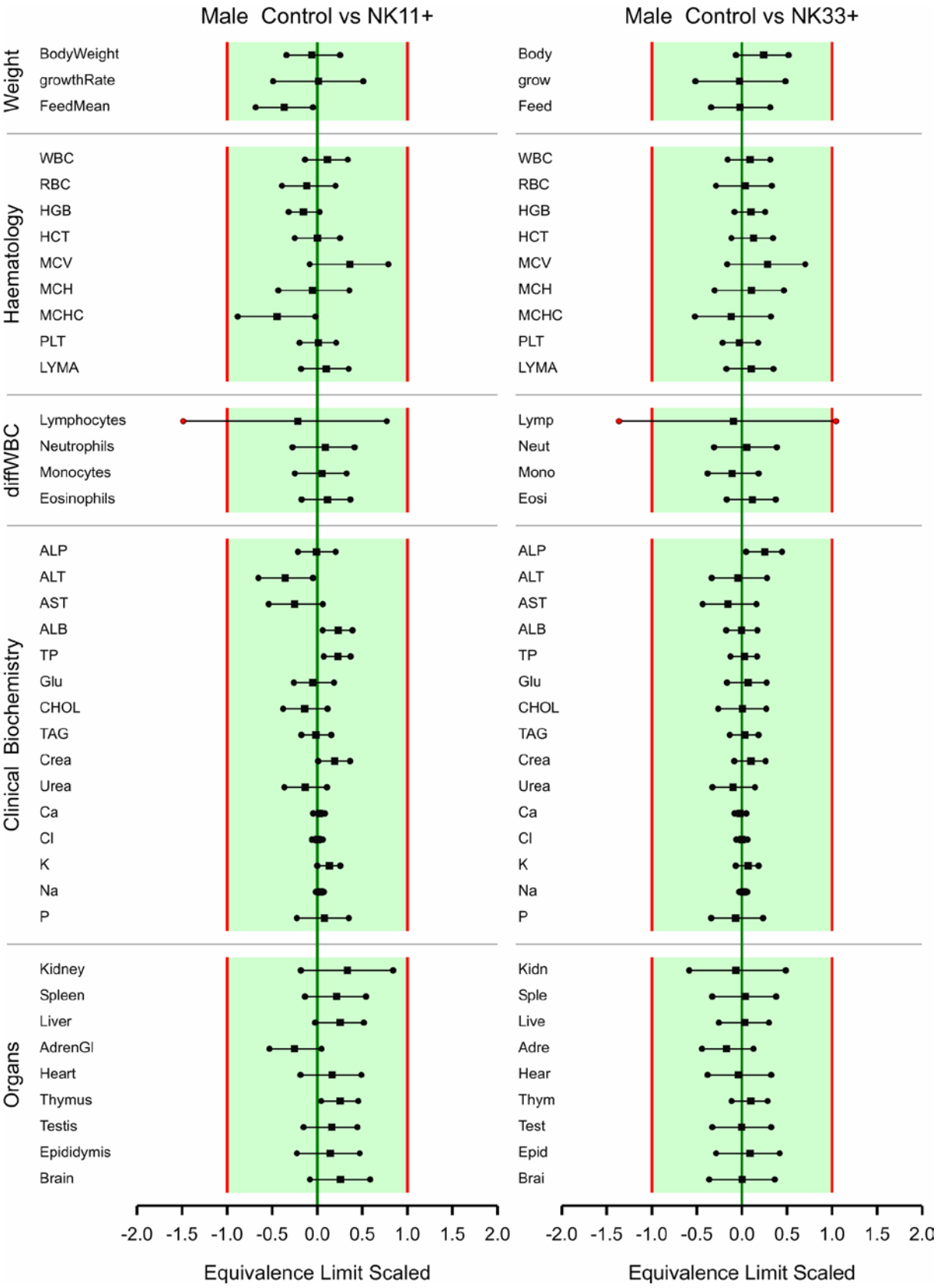

Figure 6 Equivalence testing of NK11+ and NK33+ versus the control feed for males. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. See Table 8, Figure 11 and Figure 12 for further interpretation.

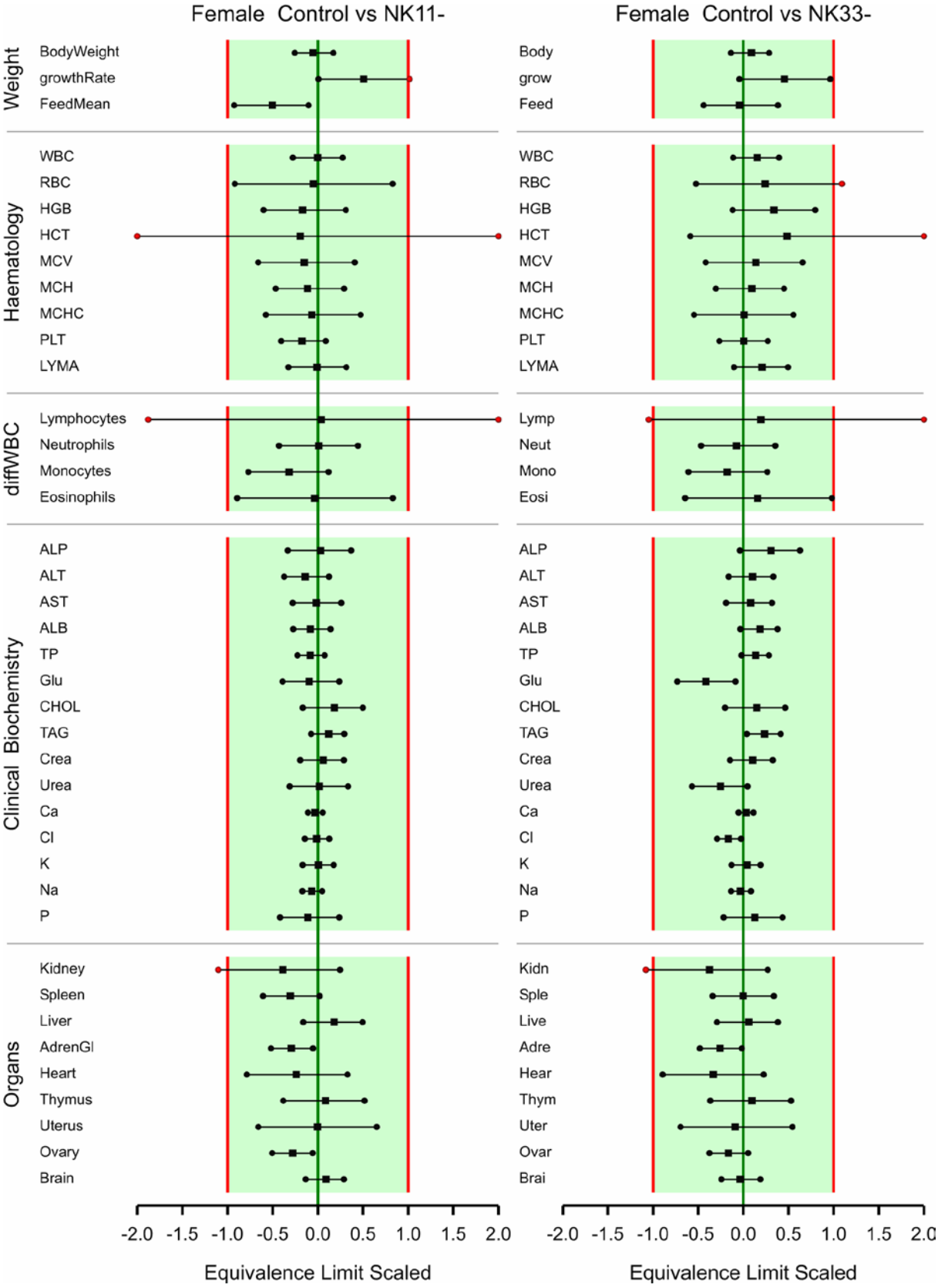

Figure 7 Equivalence testing of NK11- and NK33- versus the control feed for females. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. See Table 9, Figure 13 and Figure 14 for further interpretation.

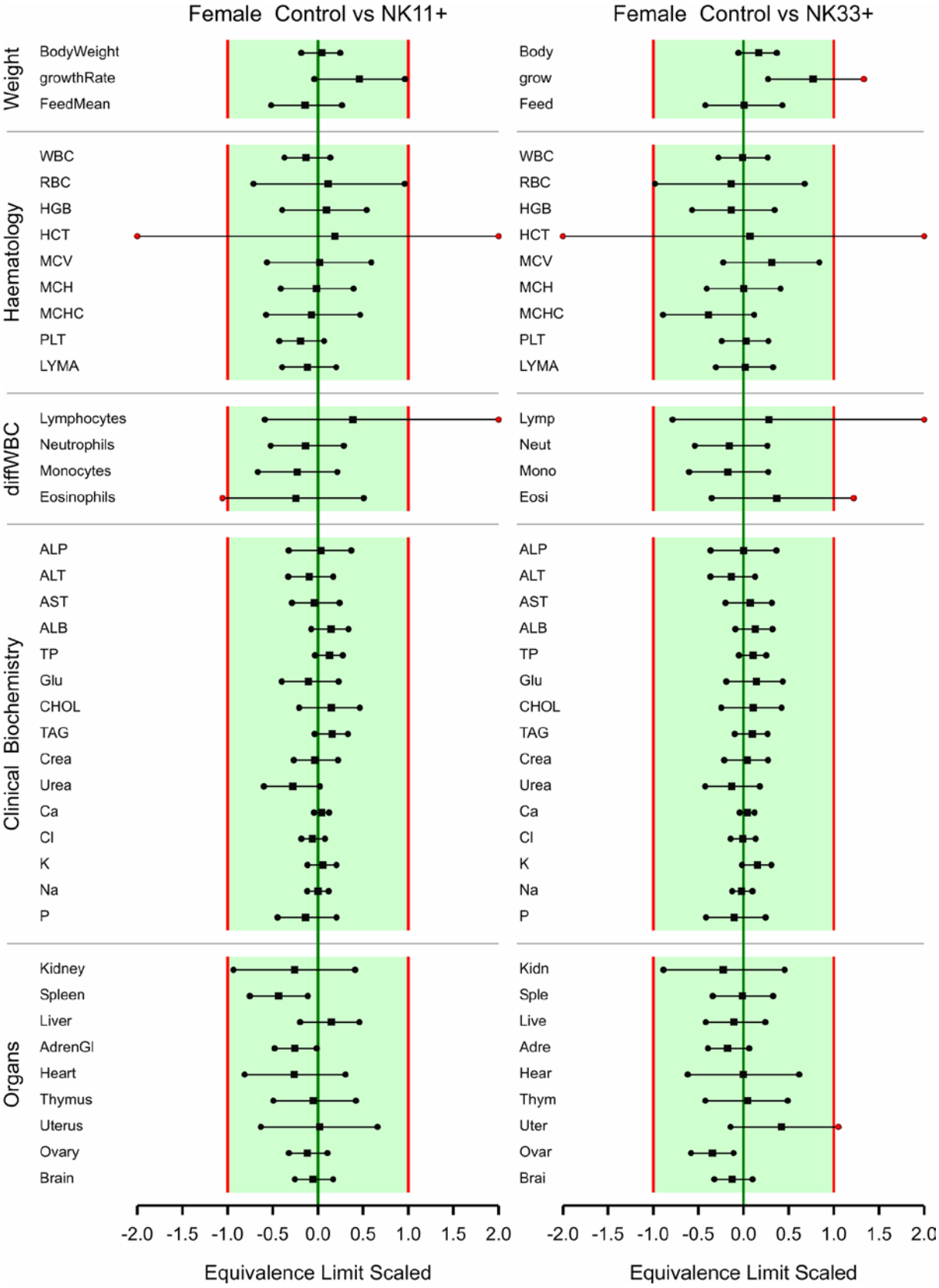

Figure 8 **Equivalence testing of NK11+ and NK33+ versus the control feed for females. For estimates (square symbols) on the left of zero the GM feed has a smaller mean than the control feed. See Table 9, Figure 15 and Figure 16 for further interpretation.**

Table 8 95% Confidence interval plus estimate for the ratio Δ of the GMO feeds versus the Control feed for males. Intervals are based on an ANOVA with 5 feeding groups. Ratios with corresponding Intervals that do not encompass the value 1 are coloured red; this is equivalent to a significant difference according to a t-test with significance level 5%.

Weights	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
BodyWeight	1.001	1.049	1.101	0.974	1.021	1.071	0.944	0.991	1.039	0.990	1.038	1.089
growthRate	0.968	0.991	1.014	0.967	0.990	1.013	0.978	1.001	1.024	0.976	0.999	1.022
FeedMean	0.968	1.017	1.069	0.970	1.019	1.071	0.899	0.945	0.993	0.948	0.996	1.047
Haematology	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
WBC	0.817	0.976	1.167	0.875	1.046	1.250	0.909	1.086	1.299	0.894	1.069	1.278
RBC	0.951	0.994	1.040	0.954	0.997	1.043	0.940	0.983	1.029	0.961	1.005	1.052
HGB	0.982	1.006	1.030	0.982	1.006	1.030	0.957	0.980	1.003	0.989	1.013	1.038
HCT	0.974	1.010	1.046	0.981	1.016	1.053	0.965	1.000	1.037	0.983	1.019	1.056
MCV	0.994	1.015	1.037	0.998	1.019	1.041	0.996	1.017	1.039	0.993	1.014	1.035
MCH	0.967	0.995	1.024	0.980	1.008	1.038	0.968	0.996	1.025	0.979	1.008	1.037
MCHC	0.961	0.980	0.999	0.971	0.990	1.009	0.961	0.980	0.999	0.975	0.994	1.014
PLT	0.787	0.919	1.072	0.870	1.016	1.185	0.866	1.010	1.179	0.840	0.980	1.143
LYMR	0.926	0.994	1.067	0.899	0.965	1.036	0.920	0.988	1.060	0.938	1.006	1.080
LYMA	0.804	0.969	1.168	0.837	1.009	1.217	0.889	1.072	1.292	0.892	1.075	1.296
diffWBC	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Lymphocytes	0.885	0.954	1.029	0.850	0.917	0.988	0.908	0.979	1.055	0.918	0.990	1.068
Neutrophils	0.950	1.122	1.327	1.006	1.189	1.406	0.882	1.043	1.233	0.868	1.026	1.213
Monocytes	0.832	1.133	1.544	0.827	1.127	1.536	0.777	1.059	1.443	0.651	0.888	1.209
Eosinophils	0.601	0.925	1.421	0.737	1.132	1.741	0.774	1.190	1.830	0.780	1.198	1.842
ClinChem	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
ALP	0.914	1.020	1.139	0.914	1.020	1.140	0.891	0.995	1.112	1.026	1.146	1.279
ALT	0.916	1.018	1.131	0.874	0.972	1.080	0.797	0.886	0.984	0.886	0.985	1.095
AST	0.837	0.963	1.108	0.820	0.943	1.085	0.775	0.892	1.027	0.810	0.932	1.073
BIL	0.880	0.991	1.117	0.866	0.975	1.099	0.892	1.005	1.132	0.813	0.916	1.032
ALB	0.978	1.006	1.035	0.986	1.014	1.043	1.010	1.039	1.068	0.972	0.999	1.028
TP	0.992	1.011	1.030	0.991	1.010	1.029	1.009	1.029	1.048	0.985	1.004	1.023

Glu	0.902	1.004	1.118	0.845	0.941	1.048	0.878	0.977	1.088	0.928	1.033	1.150
CHOL	0.883	0.974	1.076	0.899	0.992	1.095	0.857	0.946	1.044	0.908	1.003	1.107
TAG	0.908	1.089	1.305	0.858	1.029	1.234	0.819	0.982	1.178	0.866	1.038	1.245
Crea	0.954	1.032	1.117	0.939	1.016	1.099	1.004	1.087	1.176	0.964	1.044	1.129
Urea	0.872	0.953	1.041	0.891	0.973	1.062	0.871	0.952	1.039	0.882	0.964	1.052
cHGB	0.706	0.954	1.289	0.702	0.949	1.282	0.628	0.849	1.148	0.588	0.795	1.075
Ca	0.986	0.999	1.013	0.988	1.002	1.016	0.991	1.005	1.019	0.982	0.996	1.010
Cl	0.988	0.997	1.006	0.988	0.997	1.006	0.991	1.001	1.010	0.991	1.000	1.010
K	0.970	1.016	1.064	1.004	1.052	1.102	1.000	1.047	1.096	0.977	1.023	1.072
Na	0.993	1.001	1.009	0.996	1.004	1.012	0.997	1.006	1.014	0.995	1.003	1.011
P	0.942	1.032	1.131	0.970	1.063	1.165	0.935	1.025	1.124	0.893	0.979	1.073
Urine	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
uVol	0.704	0.954	1.292	0.780	1.056	1.430	0.765	1.035	1.402	0.725	0.982	1.330
uVolW	0.669	0.908	1.232	0.764	1.036	1.406	0.769	1.043	1.414	0.697	0.945	1.282
uLeu	0.522	0.934	1.669	0.559	1.000	1.788	0.522	0.934	1.669	0.559	1.000	1.788
uOsmoll	0.763	0.975	1.246	0.709	0.906	1.158	0.726	0.928	1.186	0.788	1.008	1.288
uKeton	0.646	1.867	5.394	0.437	1.261	3.643	0.365	1.055	3.049	0.231	0.667	1.928
upH	0.688	0.882	1.132	0.607	0.779	0.999	0.647	0.829	1.063	0.536	0.687	0.881
Organs	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Kidney	0.935	1.007	1.085	0.958	1.032	1.112	0.975	1.051	1.132	0.919	0.990	1.067
Spleen	0.956	1.040	1.132	0.902	0.981	1.068	0.968	1.054	1.146	0.928	1.009	1.098
Liver	0.972	1.013	1.056	0.995	1.037	1.081	0.997	1.039	1.084	0.964	1.005	1.048
AdrenGl	0.842	0.934	1.037	0.848	0.941	1.044	0.826	0.916	1.016	0.849	0.942	1.045
Heart	0.987	1.041	1.098	0.973	1.026	1.082	0.973	1.026	1.082	0.942	0.994	1.048
Thymus	0.944	1.065	1.201	0.964	1.088	1.227	1.026	1.157	1.305	0.938	1.058	1.193
Testis	0.931	0.996	1.065	0.952	1.018	1.089	0.969	1.036	1.108	0.935	1.000	1.069
Epididymis	0.914	0.989	1.069	0.967	1.046	1.131	0.954	1.032	1.116	0.943	1.020	1.103
Brain	0.928	0.977	1.029	0.980	1.032	1.088	0.988	1.041	1.096	0.950	1.000	1.054
Immunology	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Granulocytes	1.001	1.062	1.127	0.943	1.000	1.061	0.939	0.996	1.057	0.944	1.002	1.063
RespirBurst	0.900	1.037	1.194	0.809	0.932	1.074	0.805	0.927	1.068	0.858	0.988	1.138
Con	0.645	1.038	1.671	0.634	1.020	1.643	0.564	0.908	1.462	0.758	1.221	1.966

PHA	0.762	1.131	1.678	0.758	1.124	1.668	0.624	0.927	1.375	0.799	1.186	1.759
PWM	0.829	1.445	2.518	0.589	1.026	1.788	0.584	1.017	1.773	1.053	1.835	3.197
Medium	0.612	0.972	1.542	0.550	0.872	1.384	0.505	0.801	1.272	0.795	1.261	2.001
IprConA	0.667	1.069	1.712	0.750	1.201	1.925	0.707	1.132	1.813	0.602	0.965	1.546
IprPHA	0.792	1.173	1.738	0.879	1.302	1.929	0.792	1.174	1.739	0.643	0.953	1.412
IprPWM	1.098	1.487	2.014	0.878	1.189	1.611	0.945	1.280	1.734	1.080	1.463	1.981
Cytokines	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
IL2	1.124	1.374	1.678	0.937	1.145	1.399	0.912	1.115	1.362	1.146	1.400	1.711
IL4	0.538	0.794	1.171	0.418	0.617	0.910	0.417	0.616	0.908	0.528	0.778	1.148
IL10	0.457	1.216	3.238	0.236	0.629	1.675	0.201	0.535	1.424	0.728	1.939	5.161
IL17A	0.512	1.122	2.460	0.418	0.916	2.009	0.295	0.646	1.415	0.646	1.416	3.104
TNFa	1.140	1.602	2.251	0.789	1.109	1.559	0.815	1.146	1.611	1.174	1.650	2.319
IFNg	1.216	1.702	2.383	0.765	1.071	1.500	0.774	1.084	1.518	1.022	1.432	2.005
CellPhenotype	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
sp3	0.977	1.150	1.353	0.904	1.064	1.252	1.035	1.218	1.434	1.010	1.188	1.399
sp3_4	0.942	1.120	1.332	0.955	1.136	1.351	1.010	1.202	1.429	1.024	1.217	1.448
sp3_8	0.972	1.167	1.400	0.794	0.952	1.143	1.034	1.241	1.489	0.942	1.131	1.357
sp3_45	0.975	1.027	1.082	0.941	0.996	1.053	0.936	0.986	1.039	0.962	1.014	1.068
sp3_161	0.887	0.968	1.056	0.925	1.016	1.116	0.912	0.995	1.086	0.844	0.921	1.006
ln3	0.848	1.020	1.227	0.765	0.920	1.106	0.742	0.893	1.074	0.775	0.933	1.122
ln3_4	0.773	0.959	1.189	0.730	0.906	1.123	0.677	0.840	1.042	0.727	0.902	1.119
ln3_8	0.984	1.157	1.361	0.832	0.979	1.151	0.894	1.051	1.236	0.869	1.022	1.202
ln3_45	0.673	0.819	0.997	0.688	0.837	1.018	0.779	0.948	1.154	0.723	0.879	1.070
ty3	0.861	1.024	1.217	0.903	1.073	1.276	0.941	1.119	1.329	0.947	1.126	1.338
ty3_4	0.865	1.018	1.199	0.919	1.082	1.274	0.955	1.125	1.325	0.968	1.140	1.343
ty3_8	0.875	1.050	1.261	0.878	1.054	1.265	0.881	1.058	1.270	0.953	1.144	1.373
Hormone	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Males	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Testosteron	0.561	0.900	1.443	0.420	0.673	1.080	0.608	0.975	1.564	0.499	0.800	1.283
T3	0.971	1.081	1.203	0.890	0.990	1.103	0.889	0.989	1.101	0.965	1.074	1.196
T4	0.891	0.977	1.072	0.806	0.884	0.969	0.912	1.000	1.097	0.883	0.969	1.063

Table 9 95% Confidence interval plus estimate for the ratio Δ of the GMO feeds versus the Control feed for females. Intervals are based on an ANOVA with 5 feeding groups. Ratios with corresponding Intervals that do not encompass the value 1 are coloured red; this is equivalent to a significant difference according to a t-test with significance level 5%.

Weights	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
BodyWeight	0.948	0.989	1.033	0.974	1.017	1.062	0.966	1.009	1.053	0.989	1.033	1.078
growthRate	1.001	1.039	1.079	0.997	1.035	1.074	0.997	1.035	1.075	1.021	1.060	1.100
FeedMean	0.851	0.913	0.981	0.924	0.992	1.065	0.907	0.974	1.046	0.933	1.001	1.075
Haematology	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
WBC	0.816	0.998	1.219	0.920	1.124	1.373	0.740	0.904	1.105	0.813	0.993	1.214
RBC	0.941	0.996	1.054	0.965	1.021	1.081	0.954	1.010	1.069	0.934	0.988	1.046
HGB	0.950	0.986	1.024	0.990	1.029	1.068	0.971	1.008	1.047	0.952	0.989	1.027
HCT	0.941	0.989	1.039	0.978	1.028	1.081	0.962	1.011	1.063	0.955	1.004	1.055
MCV	0.967	0.992	1.019	0.981	1.007	1.033	0.975	1.001	1.027	0.990	1.016	1.042
MCH	0.958	0.990	1.024	0.975	1.008	1.042	0.966	0.999	1.032	0.968	1.000	1.034
MCHC	0.977	0.997	1.018	0.980	1.000	1.021	0.977	0.997	1.018	0.965	0.984	1.004
PLT	0.800	0.915	1.045	0.877	1.002	1.145	0.793	0.906	1.035	0.888	1.014	1.159
LYMR	0.938	0.998	1.063	0.971	1.033	1.100	0.958	1.019	1.085	0.965	1.027	1.093
LYMA	0.797	0.993	1.238	0.931	1.160	1.446	0.738	0.920	1.147	0.814	1.014	1.264
diffWBC	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Lymphocytes	0.925	1.004	1.089	0.941	1.021	1.107	0.959	1.041	1.129	0.949	1.030	1.117
Neutrophils	0.817	1.005	1.236	0.782	0.962	1.183	0.758	0.932	1.146	0.750	0.923	1.135
Monocytes	0.514	0.754	1.105	0.580	0.851	1.247	0.556	0.815	1.194	0.583	0.855	1.253
Eosinophils	0.445	0.962	2.081	0.555	1.202	2.599	0.348	0.753	1.629	0.711	1.538	3.327
ClinChem	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
ALP	0.872	1.015	1.182	0.983	1.145	1.333	0.873	1.016	1.183	0.859	1.001	1.165
ALT	0.736	0.897	1.095	0.886	1.081	1.318	0.762	0.930	1.134	0.740	0.903	1.101
AST	0.844	0.990	1.161	0.894	1.049	1.230	0.834	0.978	1.147	0.891	1.045	1.226
BIL	0.826	0.938	1.066	0.942	1.071	1.216	0.910	1.033	1.174	0.961	1.092	1.241
ALB	0.921	0.978	1.038	0.991	1.052	1.116	0.980	1.040	1.104	0.976	1.036	1.100
TP	0.942	0.979	1.018	0.995	1.035	1.076	0.993	1.032	1.074	0.988	1.027	1.068

Glu	0.833	0.959	1.104	0.726	0.836	0.963	0.830	0.955	1.100	0.924	1.064	1.225
CHOL	0.942	1.070	1.214	0.931	1.056	1.199	0.931	1.056	1.198	0.917	1.040	1.181
TAG	0.927	1.133	1.385	1.044	1.276	1.560	0.965	1.179	1.441	0.905	1.106	1.352
Crea	0.927	1.026	1.136	0.943	1.044	1.156	0.891	0.986	1.092	0.919	1.017	1.126
Urea	0.893	1.007	1.135	0.802	0.904	1.019	0.794	0.896	1.010	0.842	0.949	1.070
cHGB	0.691	0.977	1.382	0.817	1.155	1.634	0.806	1.140	1.612	0.859	1.215	1.719
Ca	0.975	0.993	1.011	0.989	1.007	1.026	0.991	1.010	1.028	0.991	1.009	1.028
Cl	0.987	0.999	1.011	0.973	0.985	0.998	0.982	0.994	1.007	0.987	0.999	1.012
K	0.946	1.002	1.061	0.958	1.014	1.074	0.962	1.019	1.080	0.995	1.054	1.116
Na	0.984	0.994	1.004	0.987	0.997	1.007	0.990	1.000	1.010	0.988	0.998	1.008
P	0.844	0.959	1.090	0.924	1.050	1.193	0.835	0.949	1.078	0.845	0.960	1.091
Urine	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
uVol	0.709	0.958	1.295	0.685	0.926	1.252	0.681	0.920	1.244	0.593	0.801	1.082
uVolW	0.710	0.970	1.324	0.667	0.911	1.244	0.667	0.911	1.244	0.568	0.775	1.058
uLeu	0.550	0.872	1.382	0.513	0.814	1.290	0.723	1.147	1.819	0.550	0.872	1.382
uOsmoll	0.754	0.993	1.309	0.782	1.030	1.357	0.758	0.999	1.316	0.803	1.059	1.395
uKeton	0.572	0.760	1.010	0.692	0.919	1.221	0.572	0.760	1.010	0.534	0.709	0.943
upH	0.650	0.829	1.057	0.692	0.882	1.126	0.650	0.829	1.057	0.671	0.855	1.091
Organs	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Kidney	0.878	0.951	1.030	0.879	0.952	1.031	0.893	0.967	1.047	0.897	0.971	1.052
Spleen	0.837	0.917	1.005	0.912	0.999	1.095	0.806	0.884	0.969	0.909	0.997	1.092
Liver	0.967	1.039	1.117	0.943	1.013	1.089	0.960	1.032	1.109	0.910	0.977	1.050
AdrenGl	0.825	0.899	0.981	0.835	0.911	0.993	0.836	0.912	0.995	0.860	0.938	1.023
Heart	0.894	0.966	1.044	0.882	0.953	1.030	0.891	0.963	1.040	0.925	1.000	1.080
Thymus	0.874	1.034	1.222	0.879	1.039	1.229	0.828	0.979	1.158	0.861	1.018	1.204
Uterus	0.788	0.998	1.264	0.760	0.963	1.219	0.797	1.009	1.278	0.943	1.194	1.512
Ovary	0.795	0.879	0.973	0.838	0.927	1.025	0.856	0.947	1.048	0.772	0.854	0.945
Brain	0.974	1.018	1.063	0.951	0.993	1.037	0.947	0.989	1.033	0.934	0.976	1.019
Immunology	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
Granulocytes	0.848	0.914	0.985	0.939	1.012	1.091	0.888	0.957	1.031	0.879	0.947	1.021
RespirBurst	0.777	0.950	1.160	0.804	0.982	1.200	0.769	0.940	1.148	0.857	1.047	1.279
Con	0.690	1.289	2.407	0.496	0.926	1.729	0.702	1.312	2.450	0.718	1.340	2.503

PHA	0.657	1.220	2.265	0.457	0.848	1.574	0.695	1.290	2.394	0.753	1.399	2.597
PWM	0.880	1.644	3.075	0.618	1.156	2.161	0.829	1.550	2.899	0.892	1.668	3.118
Medium	0.836	1.287	1.983	0.612	0.943	1.453	0.861	1.326	2.042	0.905	1.394	2.147
lprConA	0.719	1.019	1.444	0.685	0.970	1.375	0.718	1.018	1.442	0.699	0.990	1.403
lprPHA	0.561	0.938	1.569	0.543	0.908	1.519	0.580	0.970	1.622	0.594	0.993	1.662
lprPWM	0.839	1.252	1.870	0.820	1.224	1.828	0.756	1.128	1.684	0.790	1.180	1.762
Cytokines	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
IL2	0.899	1.166	1.513	0.746	0.960	1.235	0.799	1.036	1.344	0.791	1.018	1.310
IL4	0.379	0.872	2.009	0.414	0.928	2.081	0.631	1.452	3.344	0.367	0.824	1.848
IL10	0.173	0.819	3.885	0.178	0.802	3.620	0.245	1.162	5.507	0.231	1.042	4.700
IL17A	0.279	0.466	0.778	0.376	0.619	1.017	0.546	0.912	1.524	0.357	0.587	0.966
TNFa	0.530	0.773	1.127	0.505	0.727	1.048	0.778	1.134	1.653	0.584	0.841	1.212
IFNg	0.442	0.896	1.818	0.459	0.911	1.807	0.601	1.220	2.476	0.442	0.878	1.741
CellPhenotype	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
sp3	0.977	1.123	1.291	0.900	1.034	1.189	0.921	1.058	1.216	0.956	1.099	1.263
sp3_4	0.968	1.195	1.476	0.835	1.031	1.273	0.919	1.135	1.402	0.902	1.113	1.375
sp3_8	0.867	0.979	1.105	0.853	0.964	1.088	0.836	0.944	1.066	0.947	1.070	1.208
sp3_45	0.972	1.005	1.039	0.966	0.999	1.032	0.949	0.981	1.014	0.972	1.004	1.039
sp3_161	0.909	0.979	1.055	0.916	0.987	1.063	0.906	0.977	1.052	0.939	1.012	1.091
ln3	0.917	0.991	1.071	0.933	1.009	1.090	0.934	1.009	1.090	0.976	1.055	1.140
ln3_4	0.912	1.011	1.121	0.904	1.002	1.112	0.925	1.026	1.138	0.943	1.046	1.160
ln3_8	0.862	0.944	1.034	0.909	0.996	1.090	0.879	0.962	1.054	0.971	1.063	1.164
ln3_45	0.915	1.245	1.694	0.783	1.066	1.451	0.794	1.080	1.469	0.668	0.909	1.237
ty3	0.863	0.984	1.122	0.881	1.005	1.146	0.897	1.023	1.166	0.873	0.996	1.135
ty3_4	0.872	1.010	1.171	0.906	1.051	1.218	0.869	1.007	1.168	0.848	0.983	1.140
ty3_8	0.822	0.928	1.049	0.840	0.949	1.072	0.882	0.997	1.126	0.824	0.930	1.051
Hormone	NK11- vs Control			NK33- vs Control			NK11+ vs Control			NK33+ vs Control		
Females	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper	Lower	Ratio	Upper
betaEstr	0.416	0.617	0.914	0.446	0.662	0.981	0.443	0.657	0.974	0.426	0.632	0.937
T3	0.863	1.033	1.237	0.778	0.932	1.116	0.820	0.982	1.175	0.704	0.843	1.009
T4	0.594	0.856	1.235	0.642	0.926	1.335	0.793	1.144	1.649	0.467	0.673	0.971

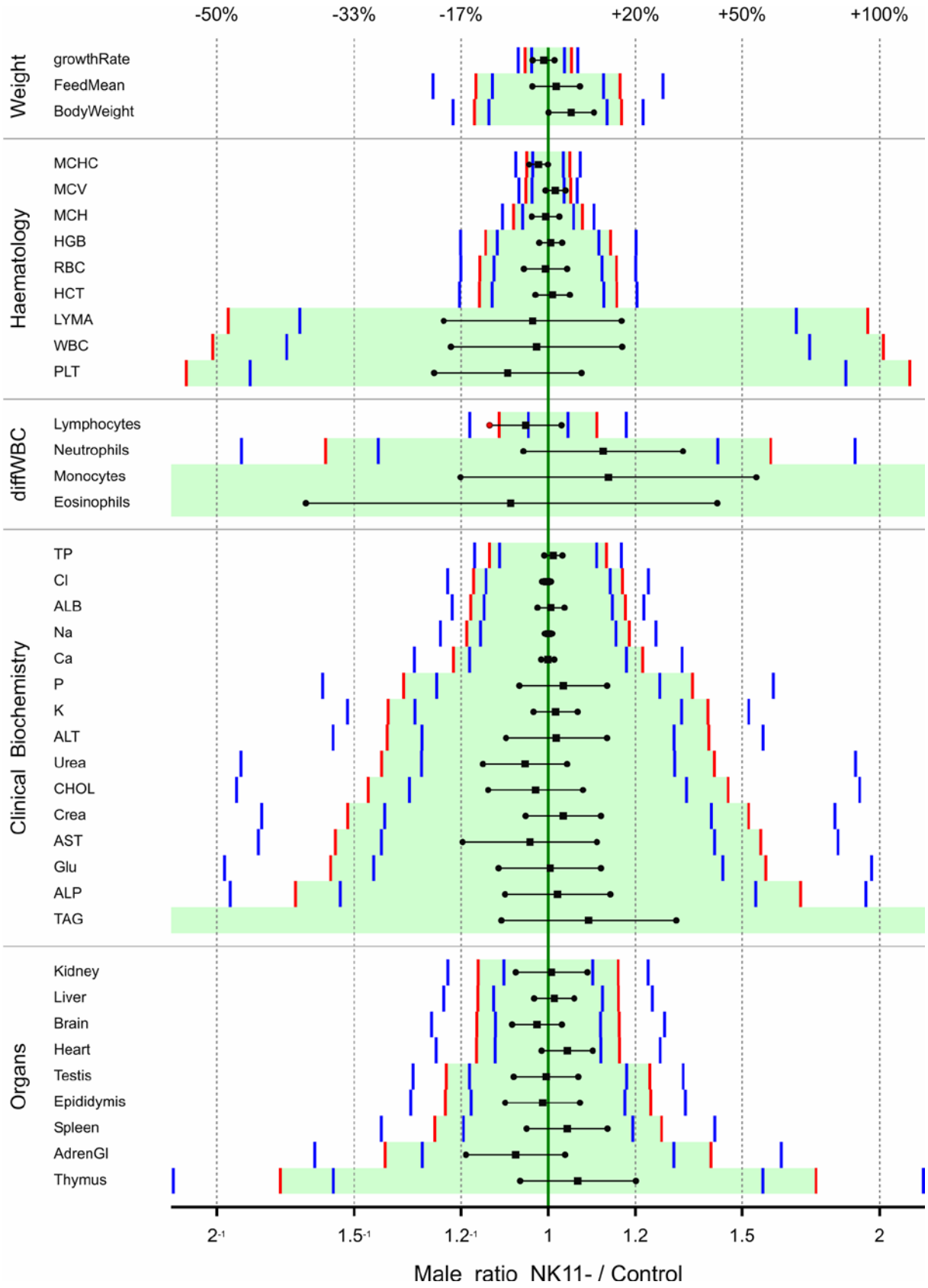

Figure 9 **Confidence intervals for the ratio of NK11- and the Control feed for males with added intervals for the equivalence limits (blue and red bars, see text).**

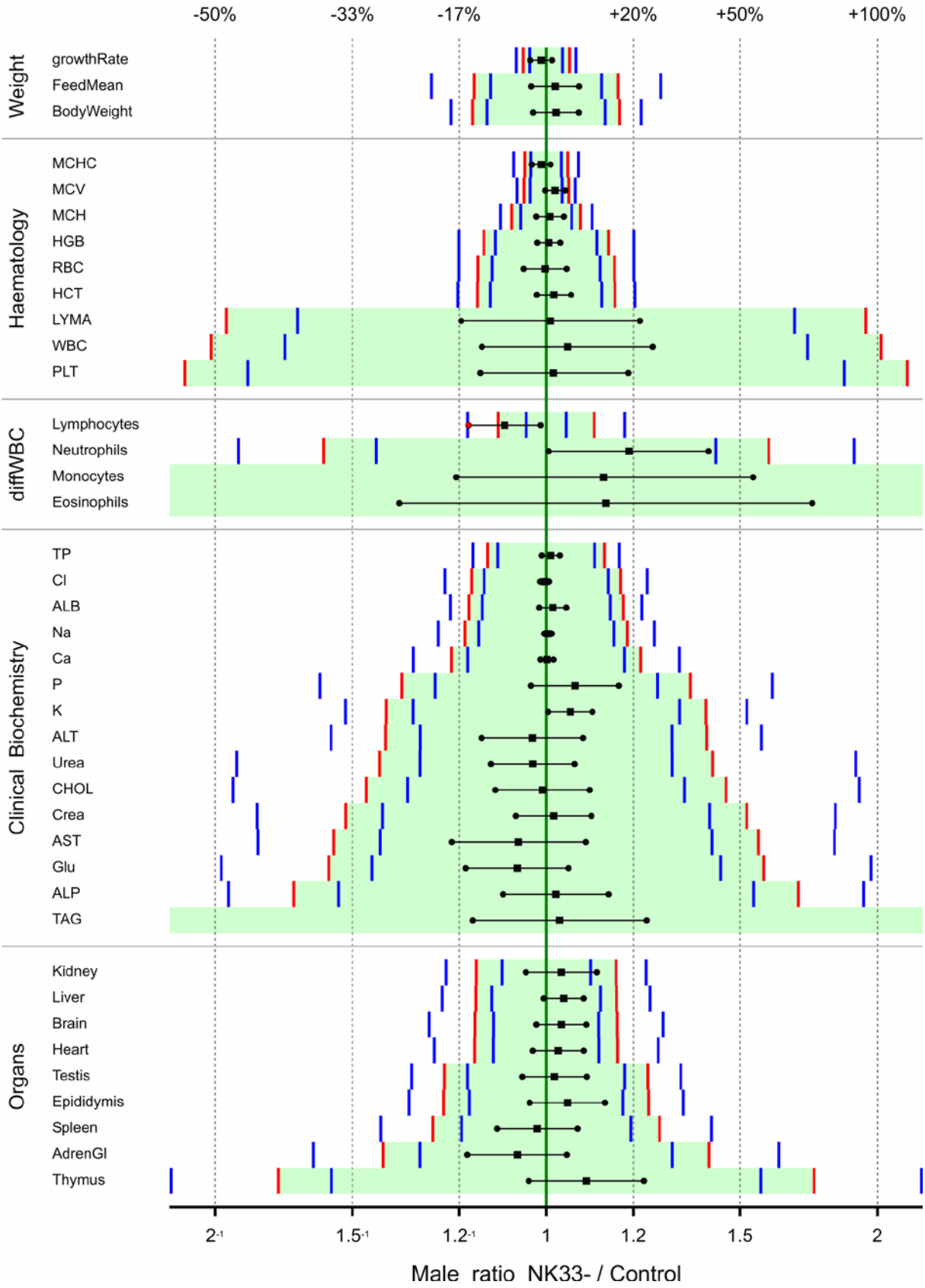

Figure 10 Confidence intervals for the ratio of NK33- and the Control feed for males with added intervals for the equivalence limits (blue and red bars, see text).

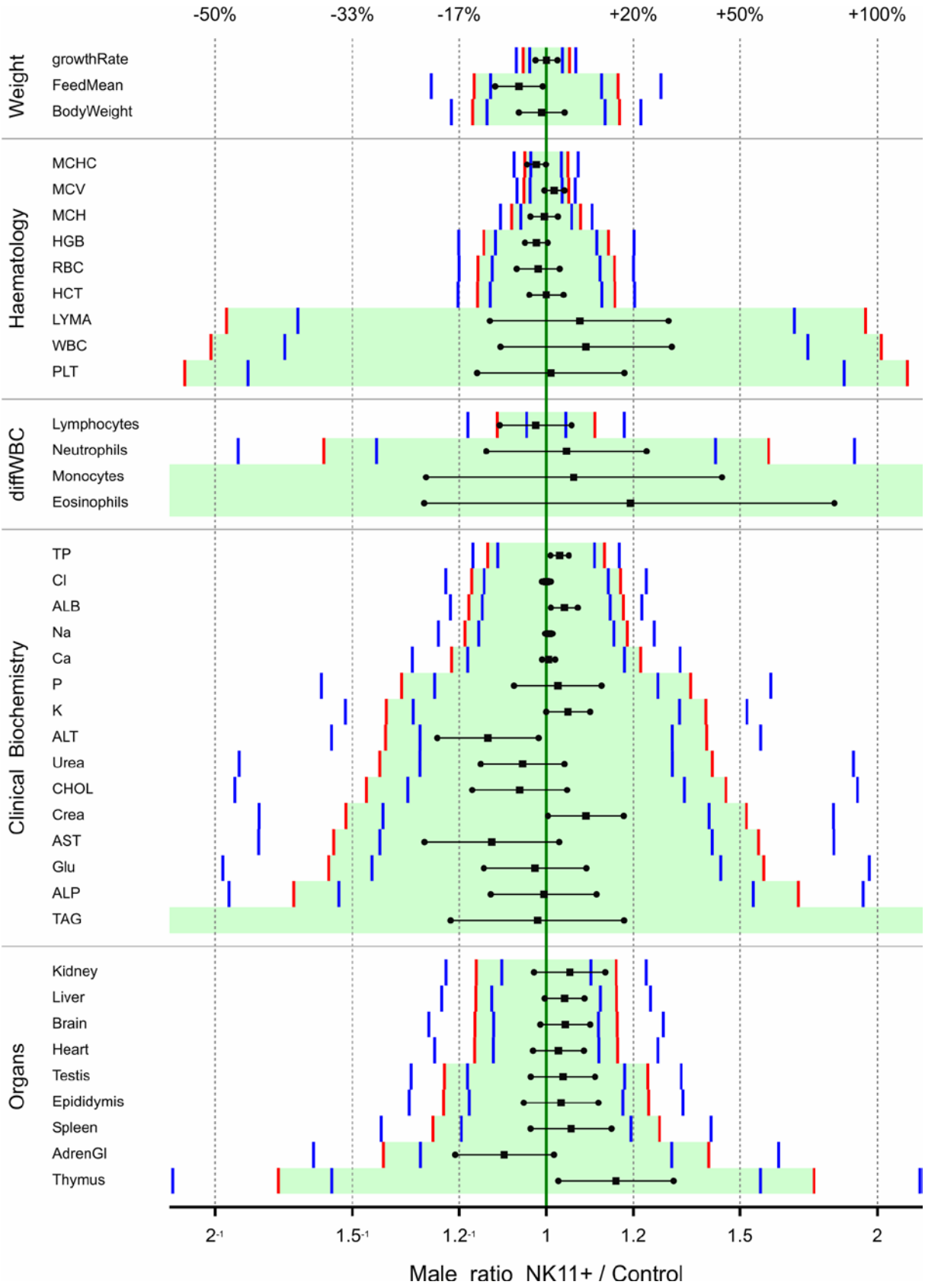

Figure 11 Confidence intervals for the ratio of NK11+ and the Control feed for males with added intervals for the equivalence limits (blue and red bars, see text).

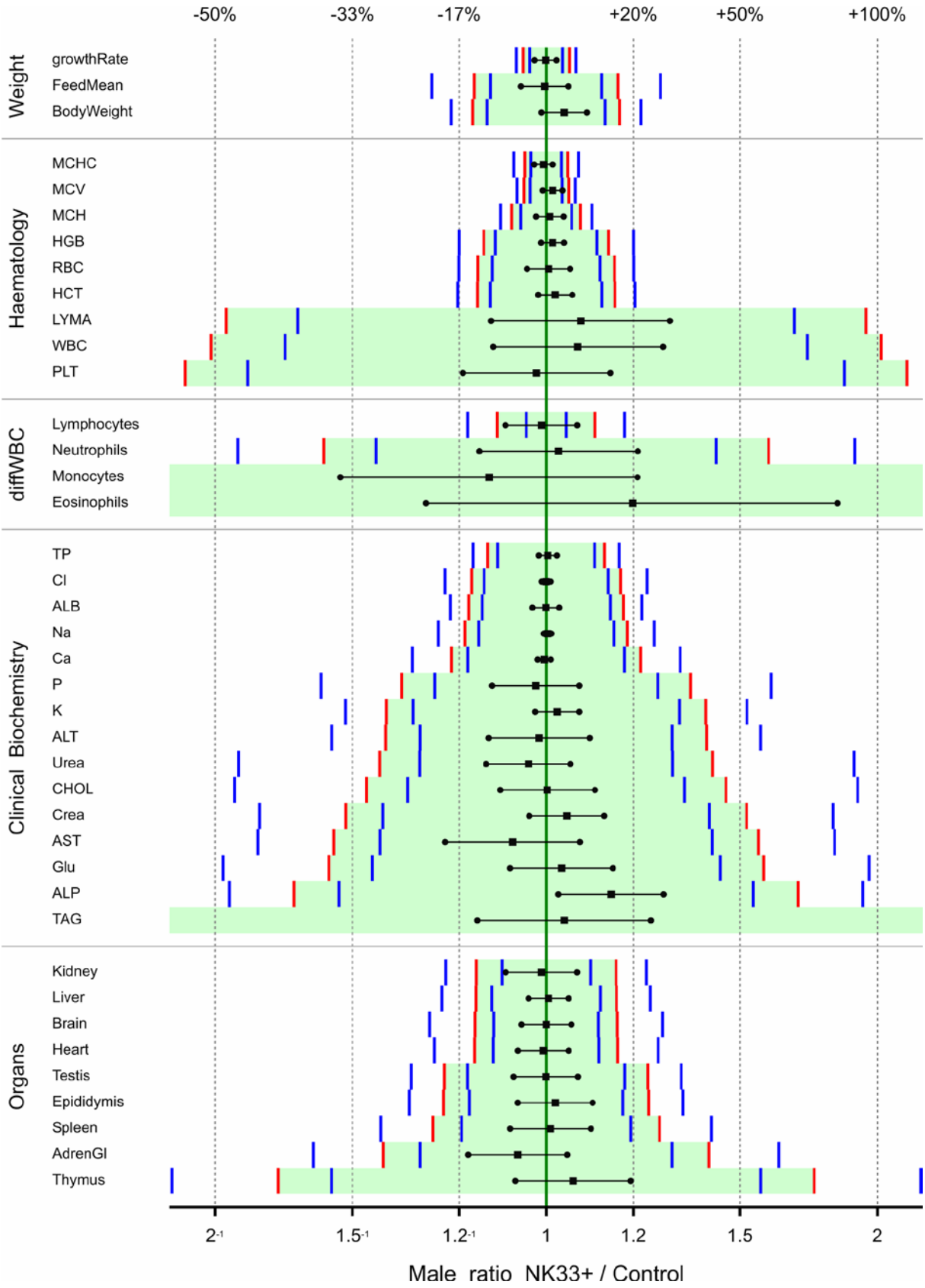

Figure 12 Confidence intervals for the ratio of NK33+ and the Control feed for males with added intervals for the equivalence limits (blue and red bars, see text).

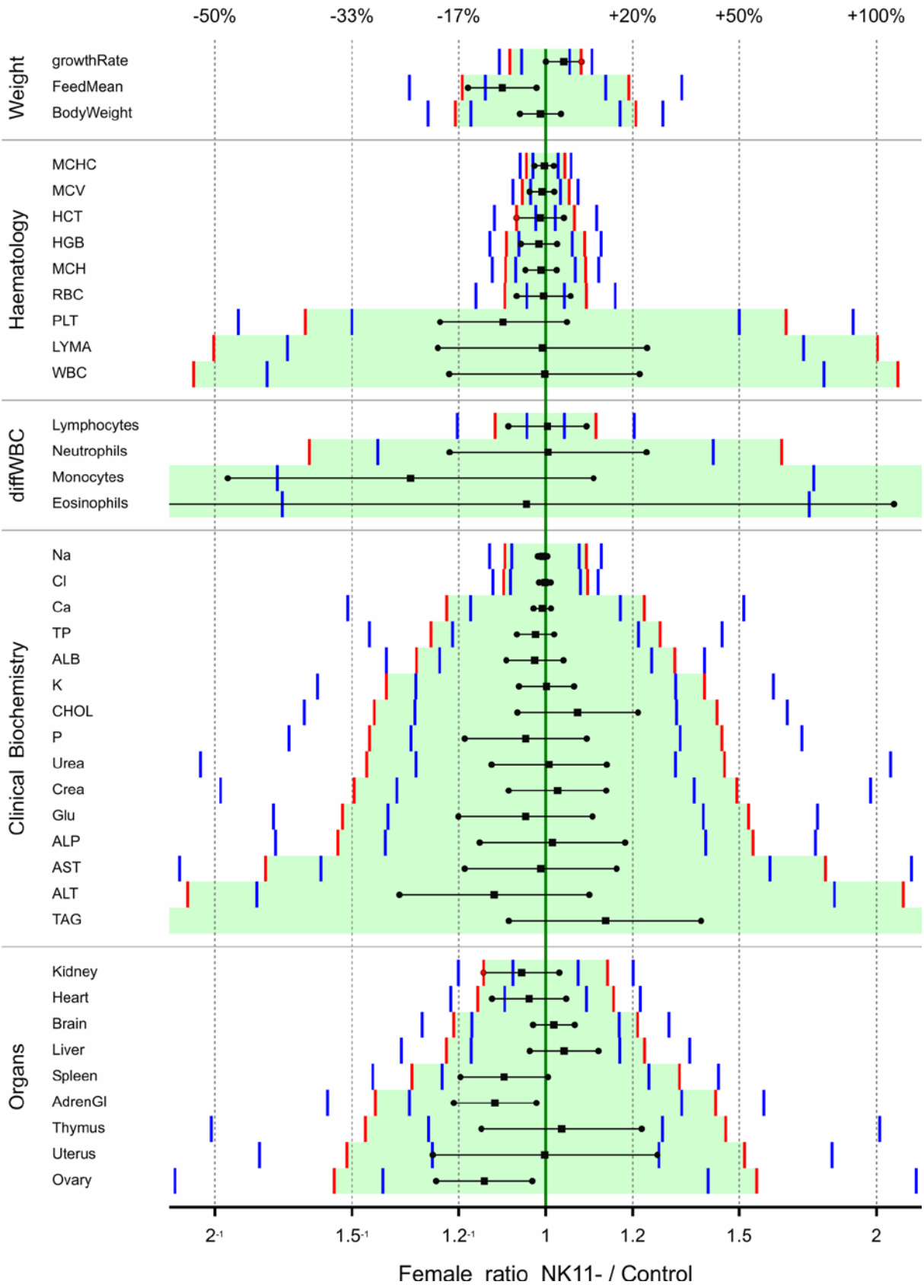

Figure 13 Confidence intervals for the ratio of NK11- and the Control feed for females with added intervals for the equivalence limits (blue and red bars, see text).

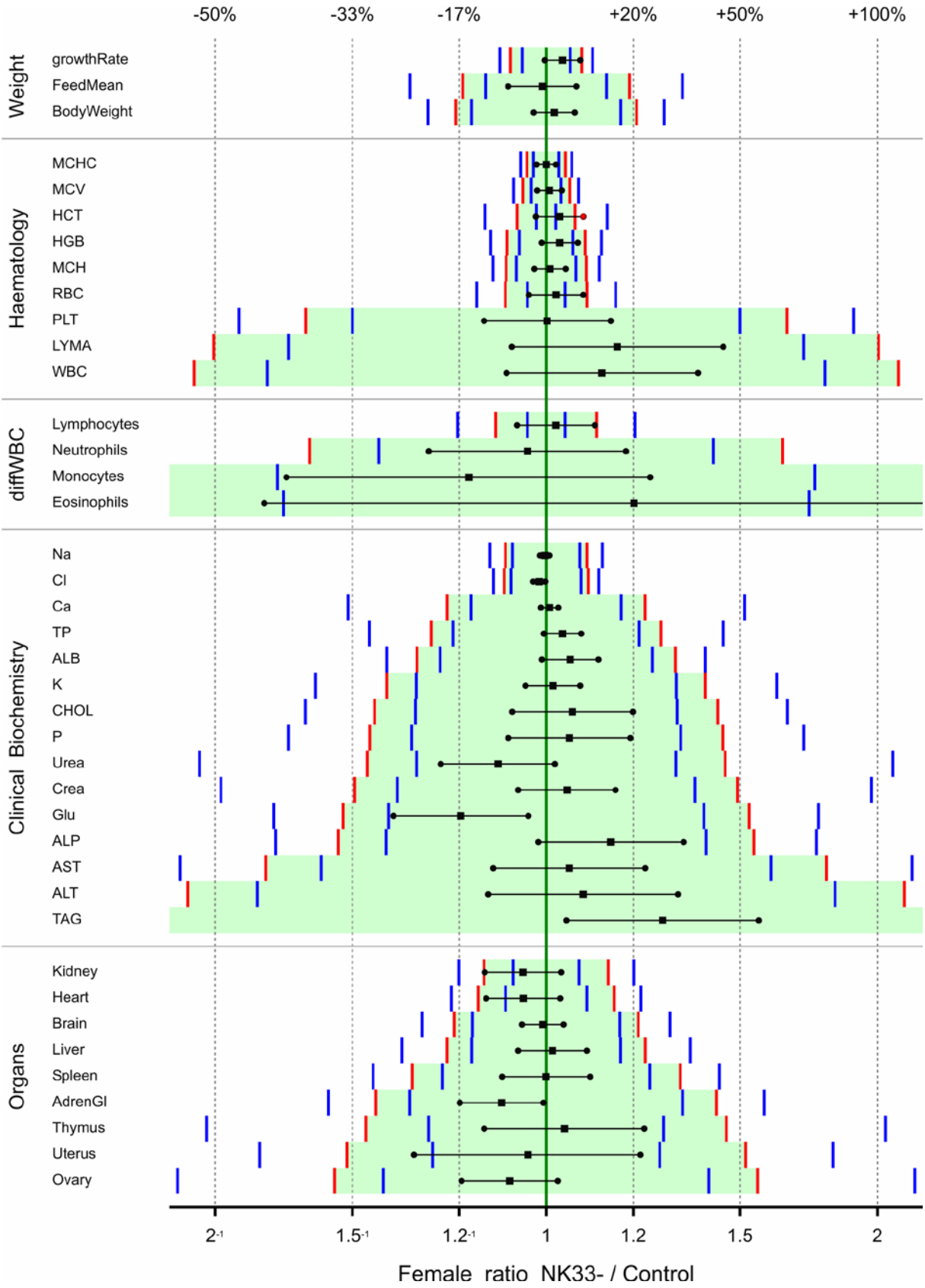

Figure 14 Confidence intervals for the ratio of NK33- and the Control feed for females with added intervals for the equivalence limits (blue and red bars, see text).

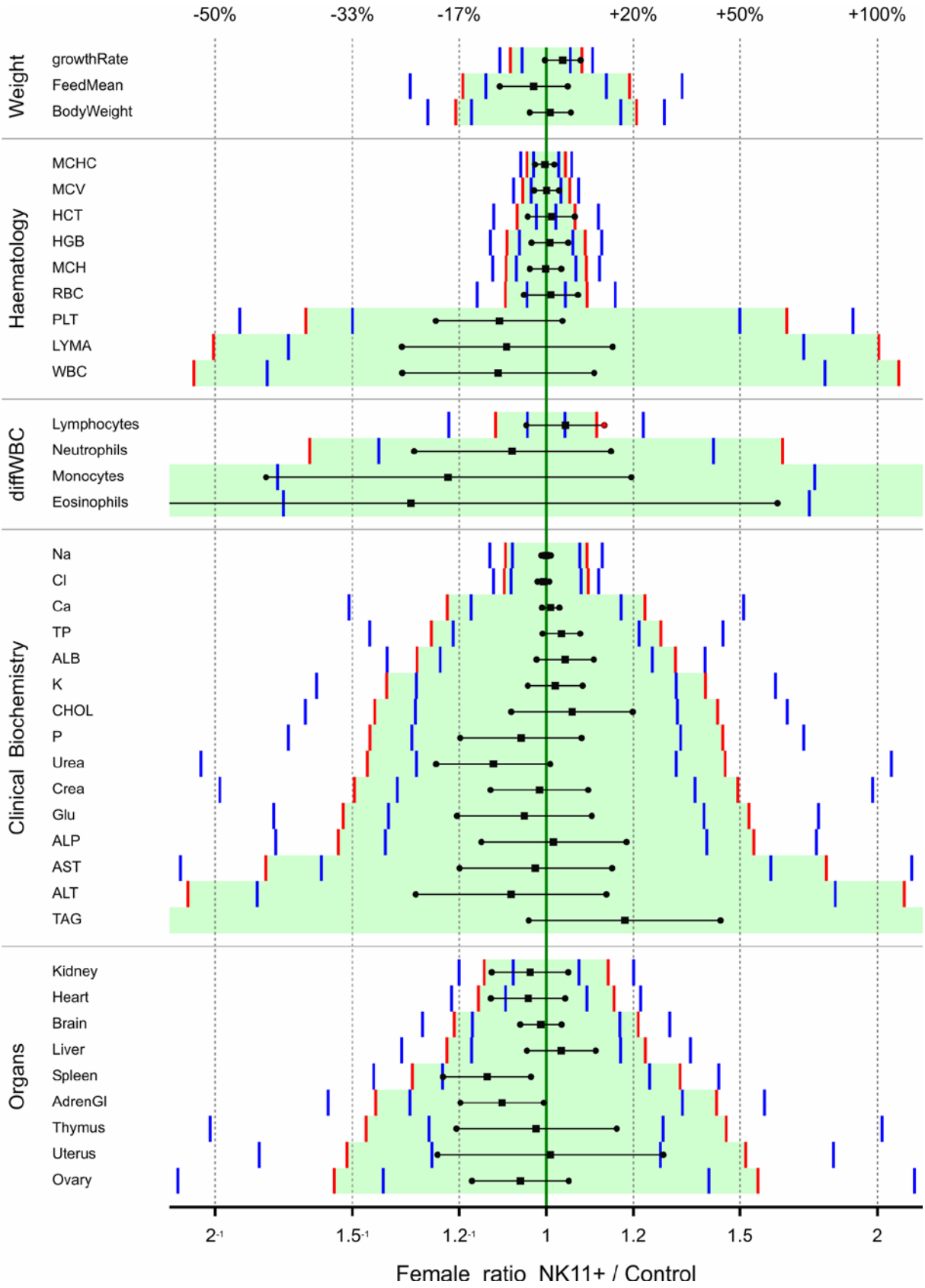

Figure 15 Confidence intervals for the ratio of NK11+ and the Control feed for females with added intervals for the equivalence limits (blue and red bars, see text).

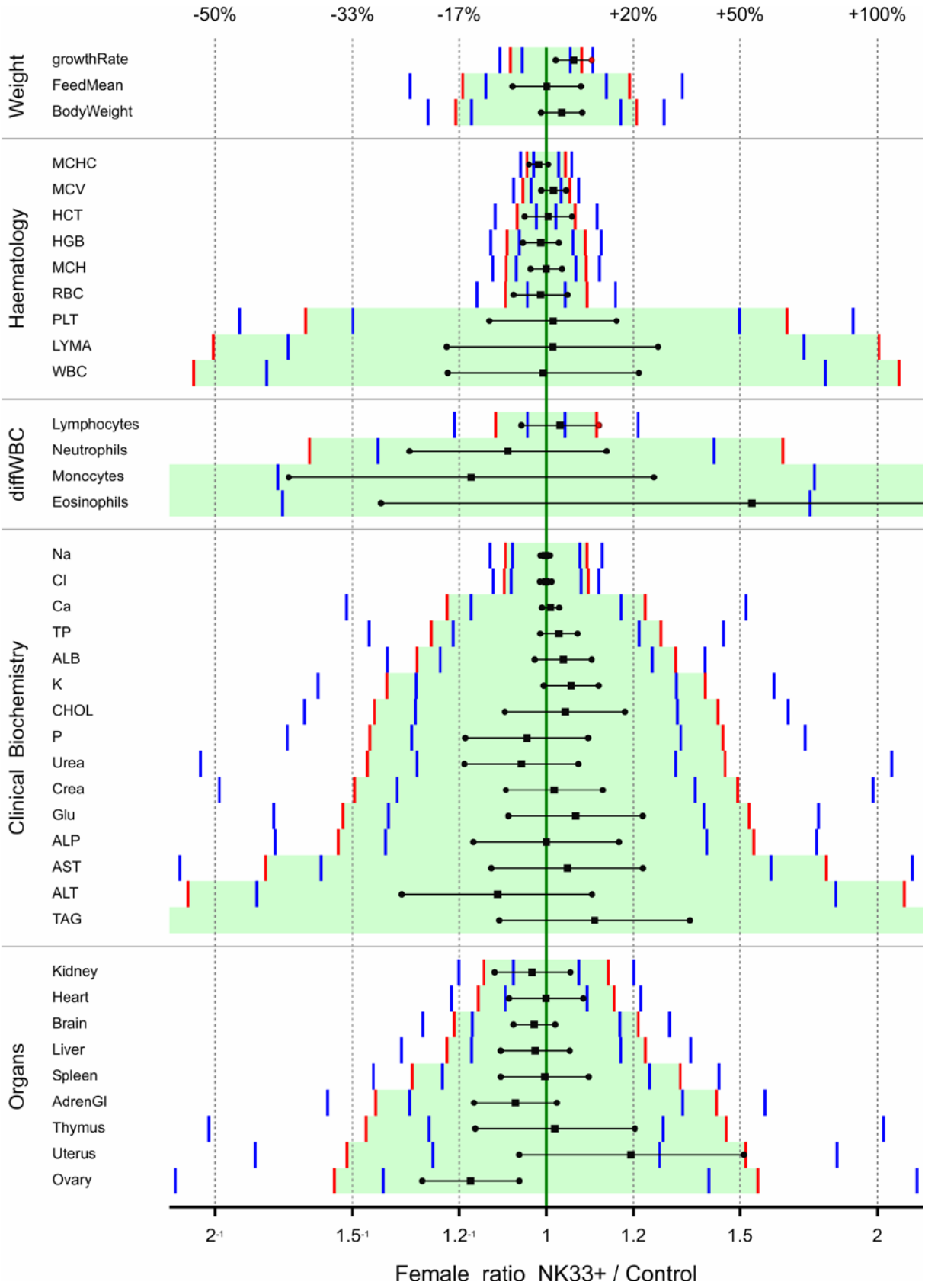

Figure 16 Confidence intervals for the ratio of NK33+ and the Control feed for females with added intervals for the equivalence limits (blue and red bars, see text).

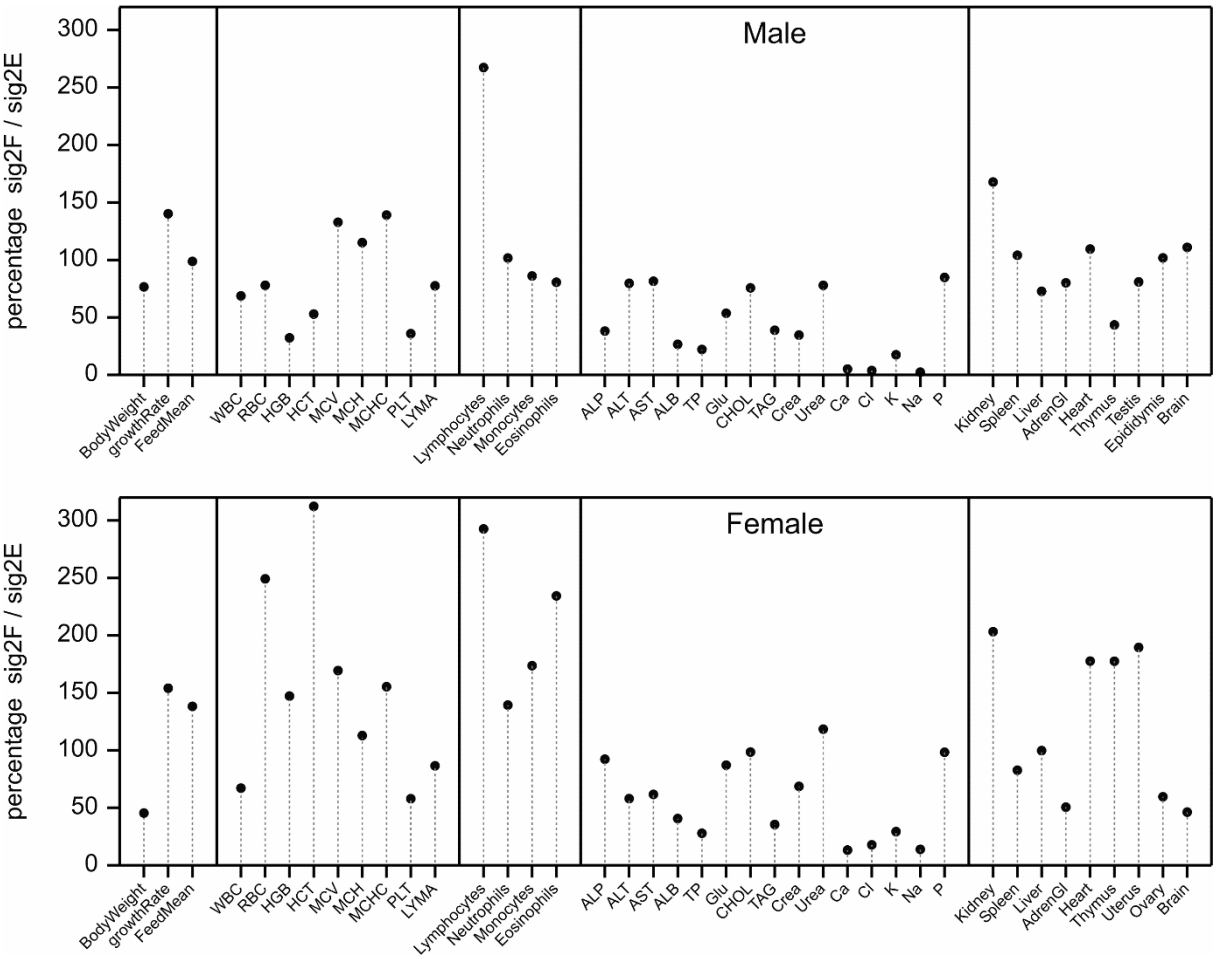

Figure 17 Residual variance (sig2F or σ_F^2) in the current G-TwYST B study as a percentage of the residual variance (sig2E or σ_E^2) in the historical GRACE studies for males (top panel) and females (bottom panel).

4.2 Equivalence testing using target effect sizes

4.2.1 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 10 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 10 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)	Krea	+ 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.

4.2.2 Results

The confidence intervals for the 9 variables in Table 10, as well as the equivalence limits are given in Figure 18 for males and in Figure 19 for females. Only for WBC and LYMA for GMO feed NK33- administered to females the null hypothesis of non-equivalence is not rejected. In all other cases (70 out of 72, or 97%) non-equivalence was rejected, and thus equivalence accepted, with generally small p-values (Table 11). For WBC and LYMA in Females NK33- equivalence was still more likely than not according to the terminology of EFSA (2011a).

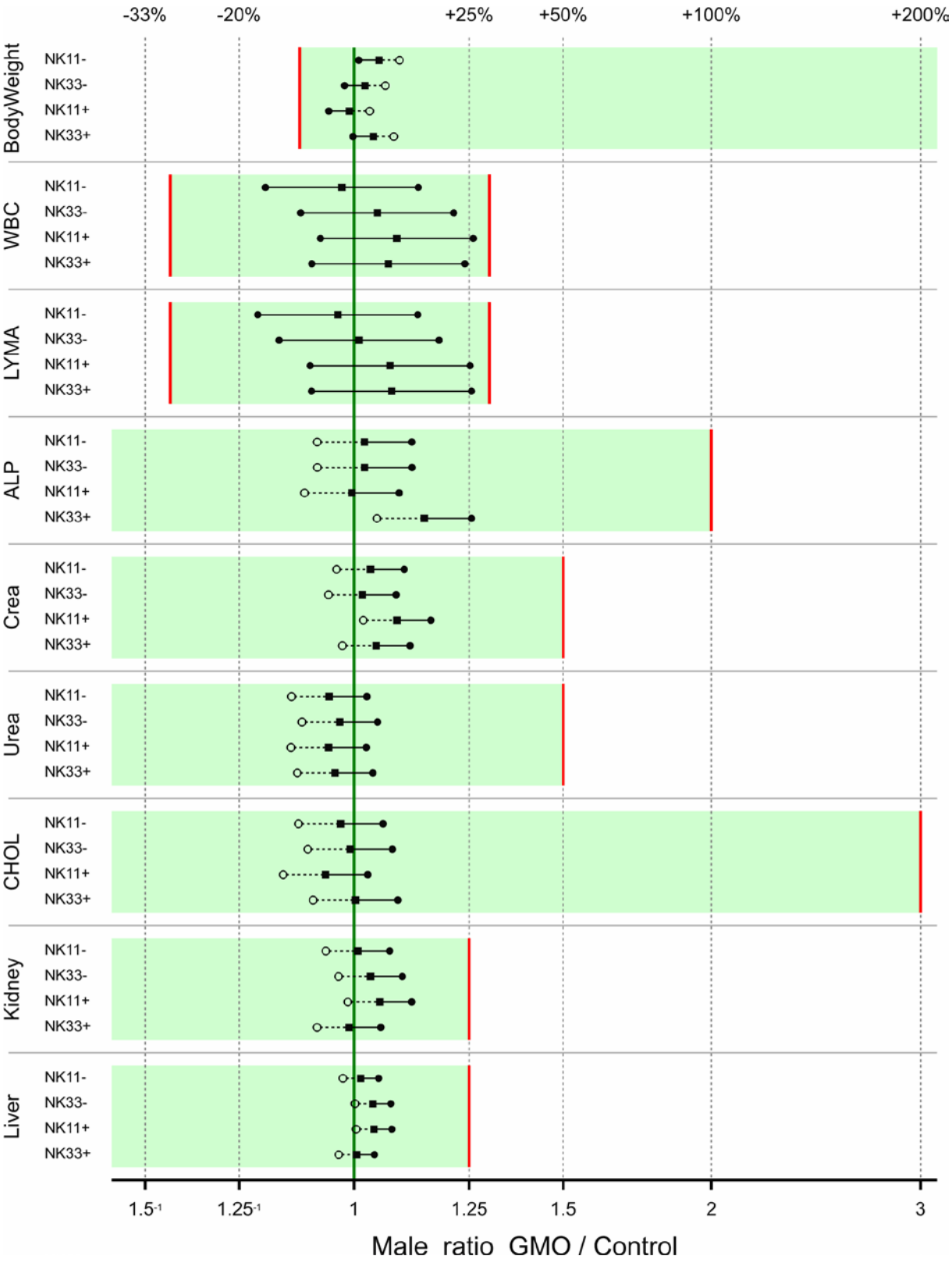

Figure 18 90% confidence intervals for the ratio of the mean of the GMO feed and the control feed for selected variables for males along with equivalence intervals defined by targeted effect sizes of Hong et al (2017).

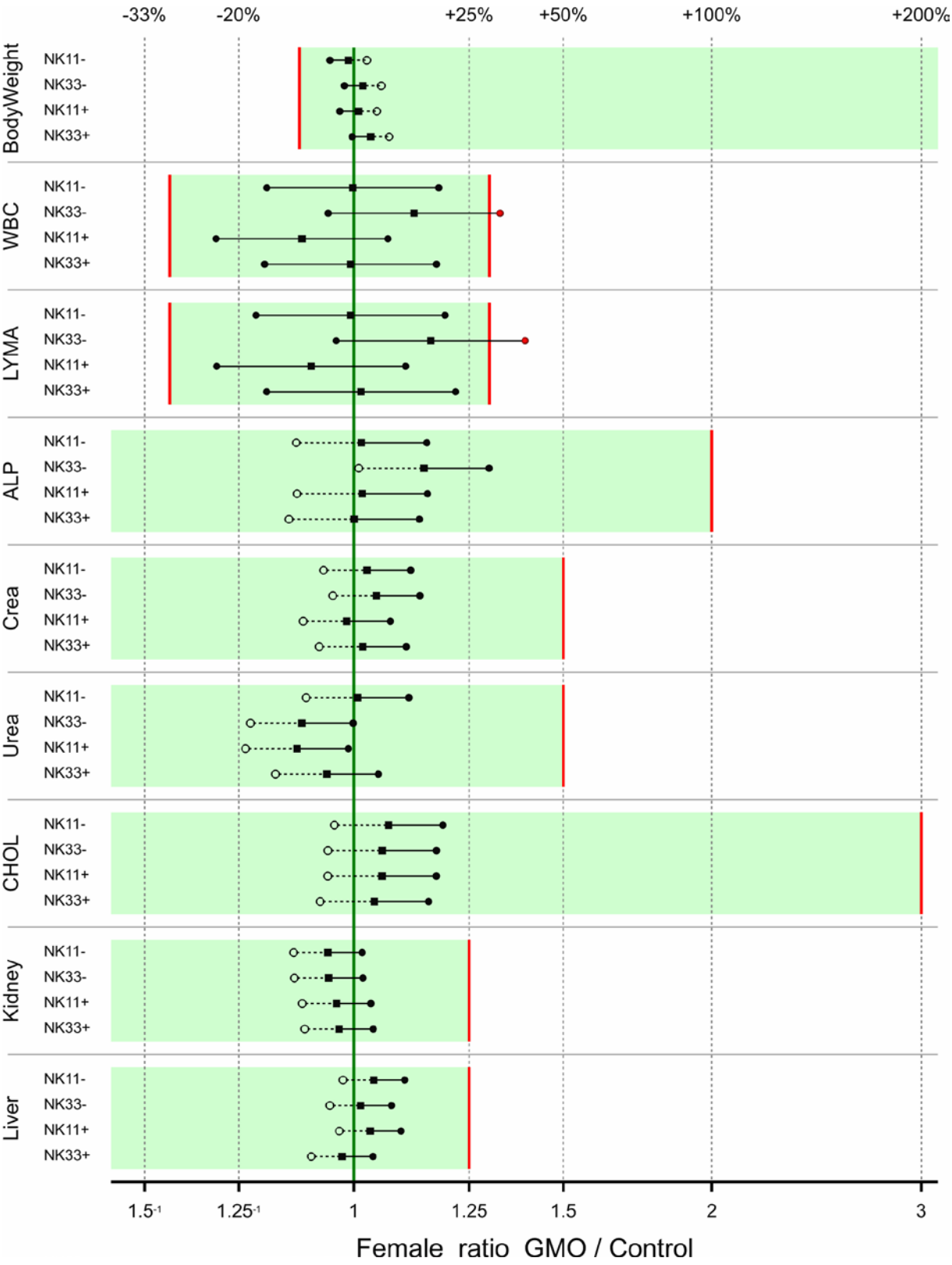

Figure 19 90% confidence intervals for the ratio of the mean of the GMO feed and the control feed for selected variables for females along with equivalence intervals defined by targeted effect sizes of Hong et al (2017).

Table 11 P-values of equivalence tests for the ratio of the mean of the GMO feed versus the mean of the control feed using targeted effect sizes of Hong et al (2017), see Table 10, as equivalence limits. P-values smaller than 0.01/0.05 have a gold/yellow background.

Variable	Males				Females			
	NK11-	NK33-	NK11+	NK33+	NK11-	NK33-	NK11+	NK33+
BodyWeight	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WBC	0.001	0.009	0.024	0.016	0.006	0.074	0.007	0.005
LYMA	0.002	0.005	0.022	0.023	0.009	0.149	0.008	0.014
ALP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Krea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Urea	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CHOL	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Kidney	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Liver	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

4.3 Classical statistical analysis

4.3.1 Method

G-TwYST study B is based on OECD guidance 408 on repeated dose 90-day oral toxicity studies in rodents (OECD 1998), EFSA guidance complementing the OECD guidance for whole food/feed studies (EFSA 2011b), and additional EFSA clarifications (EFSA 2013, 2014). OECD guidance 408 (OECD 1998) 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 20).

Figure 1 A statistical decision tree, summarising common statistical procedures used for the analysis of data in long-term toxicology studies.

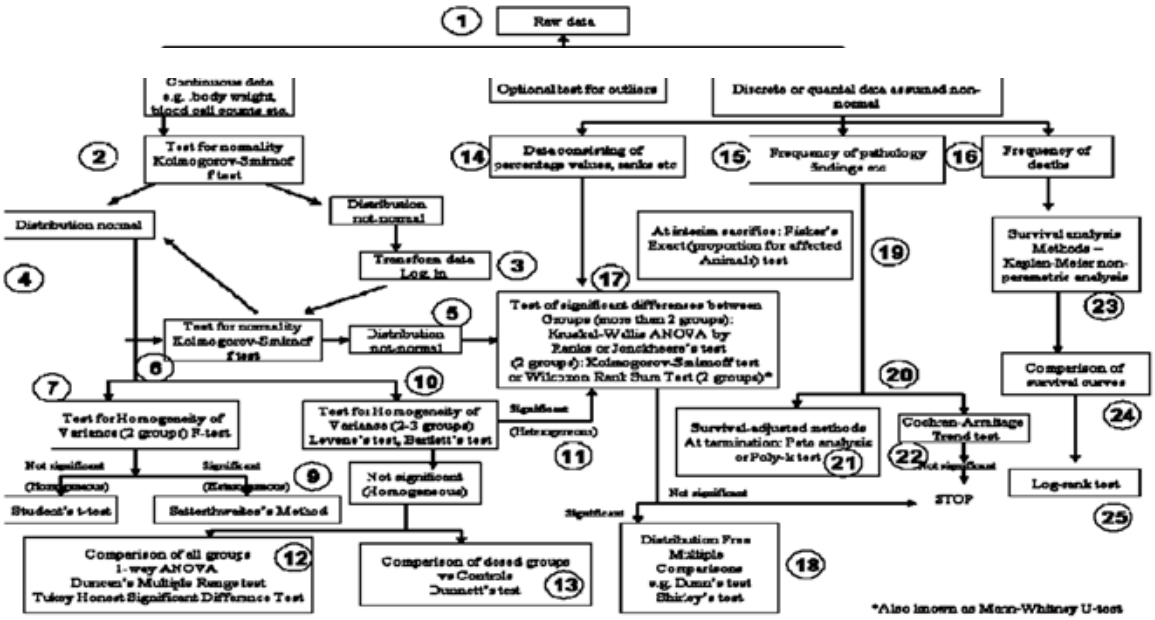

Figure 20 Classical approach to statistical analysis of data in long-term toxicity studies (copied from OECD 2012).

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).

In the current section we apply 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).

The following variables were not statistically analysed:

- The observed weights in each week. The weights are summarized by means of the growth curve resulting in the growthRate variable and the final BodyWeight variable (section 3.1);
- The observed feed intake in each week. These are summarized in the FeedMean variable which is the mean feed intake over the thirteen weeks;
- Basophils: the observed number of basophils white blood cells is zero for all rats;
- uColour: the colour of the urine was yellow (score 2) for all rats except for female rats 179 and 180 which had light yellow urine (score 1). These two rats are both housed in cage 83 and received the Control feed;
- uBil: all measurements of bilirubin in urine are equal to zero;
- uNit: all measurements of nitrites in urine equal zero except for the five male rats 16 (NK11-), 48 (NK33-), 63 (NK11+), 64 (NK11+) and 79 (Control) which have the value one;
- uProtein: all measurements of total protein in urine are equal to zero;
- uGlu: all measurements of glucose in urine are equal to one;
- uHemogl: all measurements of haemoglobin in urine are equal to zero, except for male rat 47 (NK33-) which has a value 25 and for female rat 147 (NK33-) which has a value 1;
- uUrobili: all measurements of urobilinogen in urine are equal to one;
- IL1a: cytokine Interleukin 1 alpha equals 11.2 for most rats except male rats 73 (26.43, Control) and 76 (13.94, Control), and female rat 154 (14.1, NK11+);
- IL1b: cytokine Interleukin 1 beta equals 15.9 except for male rat 37 (159.3, Control) which is most probably a typo;
- IL5: cytokine Interleukin 5 equals 2.4 except for male rats 73 (5.2, Control) and 74 (3.3, Control), and female rats 105 (3.03, NK11-), 119 (2.55, NK33+) and 165 (3.42, Control);
- IL6: cytokine Interleukin 6 equals 4.9 except for male rat 73 (7.34, Control);
- IL12p70: cytokine Interleukin 12p70 equals 9.8 for 55 rats except for the 11 male rats 5 (16.49, NK11-), 8 (14.27, NK11-), 10 (19.04, NK11-), 12 (12.89, NK11-), 13 (12.35, NK11-), 28 (10.20, NK33+), 70 (39.21, Control), 71 (39.21, Control), 73 (44.60, Control), 74 (20.71, Control), 75 (16.83, Control) and 78 (18.60, Control), and the 14 female rats 101 (31.35, NK11-), 105 (17.89, NK11-), 118 (14.72, NK33+), 119 (16.83, NK33+), 149 (12.95, NK11+), 150 (40.35, NK11+), 151 (61.66, NK11+), 152 (53.68, NK11+), 153 (42.68, NK11+), 154 (25.51, NK11+), 165 (50.13, Control), 166 (22.17, Control), 167 (20.70, Control) and 168 (30.39, Control);
- IL13: cytokine Interleukin 13 equals 4.9 for 66 rats except for the 11 male rats 7 (5.58, NK11-), 8 (7.41, NK11-), 12 (6.81, NK11-), 13 (6.26, NK11-), 22 (14.62, NK33+), 27 (10.37, NK33+), 53 (7.07, NK11+), 69 (7.74, Control), 70 (6.37, Control), 71 (9.82, Control) and 73 (12.43, Control), and the 4 female rats 119 (5.05, NK33+), 150 (5.85, NK11+), 153 (7.07, NK11+) and 165 (13.21, Control);
- GCSF: cytokine granulocyte colony-stimulating factor equals 3.7 for all rats;

- GMCSF: cytokine granulocyte-macrophage colony-stimulating factor equals 9.8 for most rats except for the 7 male rats 7 (14.48, NK11-), 8 (18.05, NK11-), 14 (13.07, NK11-), 22 (14.08, NK33+), 70 (12.87, Control), 71 (12.55, Control) and 73 (19.18, Control), and the 3 female rats 151 (13.30, NK11+), 154 (29.40, NK11+) and 165 (11.30, Control);

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 12; Appendix 13 contains an example dataset for the R-script. 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 variables in the eCycle group are analysed separately.
- 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 (Appendix 6) and a plot of residuals versus fitted values (Appendix 7).
- 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.
- The non-parametric Friedman test, which is applicable to a randomized block design, is used to test for overall differences between the five feeds.
- 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 (Appendix 7). 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.

For the oestrous cycle data (in group eCycle) the following variables were statistically analysed:

- testing whether the probability of having an unfinished cycle is different between the feeds; Fisher's exact test was used for this purpose;
- testing whether the probability of having a cycle of 5 days is different between the feeds; Fisher's exact test was used for this purpose;
- testing whether the probability of having an irregular cycle is different between the feeds; this was done by means of logistic regression.

4.3.2 Results

Table 12 (males) and Table 13 (females) present the results of the t-tests, of Dunnett's tests and of Wilcoxon tests for the 79 variables divided in ten groups. For ease of interpretation results are expressed as means and coefficients of variation on the original scale, rather than as means and standard deviations on the log scale. Note however that 95% confidence intervals on the ratio scale are given in Table 8 (males) and Table 9 (females). It can be seen that the relative precision of variables ranges from high precision, e.g. CV 1-3% for e.g. HGB, HCT, MCV, Ca, Cl, Na, to low precision, e.g. CV 40-175% for e.g. uLeu, uKeton, IL10, Eosinophils.

Results of the t-test and Dunnett's test for the difference tests, with 28 degrees of freedom for residual, are summarized by letters which indicate significance at 5% and at 1%. Results for the non-parametric Wilcoxon test, where each test only uses data for the specific GMO feed and the Control feed, are summarized in the same way. Exact p-values for these tests are given in Appendix 14. In 62 cases (9.8% of the 632 comparisons) a difference was significant by at least one of the tests at the 5% level. On their own Dunnett's test resulted in 17 significant differences (2.7%), the t-test resulted in 55 significant differences (8.7%), Wilcoxon's test resulted in 26 significant differences (4.1%). Note that cells coloured red in Table 8 and Table 9, with confidence intervals for ratios, correspond to cells coloured red in Table 12 and Table 13 with letters t, T, d or D.

Results of the Shapiro-Wilks test for normality and Bartlett's and Levene's test for homogeneity of variance are given in Appendix 15. The non-normality as indicated by the Shapiro-Wilks test for separate feeding groups is frequently significant. The graphs of cage means on the log scale in Appendix 5 indicate that significance of non-normality is mostly due to one outlying observation in a feeding group.

Table 12 Means and coefficient of variation (CV) for male rats. Means of GMO feeds which are significantly different from the Control feed 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. Note that Dunnett- and t-tests are based on a ANOVA with 5 treatment groups.

Weights	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
BodyWeight	439.7	6.8	462.0	8.4	t	448.9	6.5		435.3	5.7		458.2	10.7	
growthRate	0.145	17.0	0.136	12.0		0.134	20.4		0.146	23.4		0.144	11.9	
FeedMean	19.99	5.8	20.20	8.2		20.04	4.9		18.53	3.3	DW	19.79	8.9	
Haematology	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
WBC	9.100	21.1	8.850	17.8		9.419	17.3		9.750	12.5		9.575	10.9	
RBC	7.790	3.8	7.768	5.6		7.768	3.4		7.672	4.6		7.829	2.6	
HGB	14.64	2.8	14.72	2.0		14.72	3.0		14.36	3.8		14.84	1.2	
HCT	42.36	2.8	42.84	4.6		43.05	3.3		42.40	3.8		43.16	2.0	
MCV	54.41	2.0	55.25	2.6		55.44	2.1		55.36	2.5		55.14	1.1	
MCH	18.82	2.5	18.74	2.8		18.97	3.6		18.76	4.2		18.96	2.4	
MCHC	34.57	0.8	33.90	2.8	t	34.22	2.3		33.87	1.7	tw	34.38	2.2	
PLT	842.5	8.9	792.1	15.1		861.6	14.5		851.0	9.7		847.1	13.6	
LYMR	73.02	5.1	72.49	6.9		70.93	8.3		71.97	4.7		73.42	6.9	
LYMA	6.644	23.3	6.419	19.3		6.681	20.0		7.000	10.0		7.037	15.4	
diffWBC	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Lymphocytes	71.38	9.0	68.09	9.5		65.84	8.9	t	69.81	9.4		70.59	8.9	
Neutrophils	25.69	25.6	28.84	24.3		30.53	16.9	t	26.75	25.5		26.38	25.0	
Monocytes	1.500	30.9	1.656	26.7		1.781	51.7		1.531	26.8		1.312	31.8	
Eosinophils	1.438	40.3	1.406	51.1		1.844	71.7		1.906	47.0		1.719	45.6	
ClinChem	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
ALP	1.252	11.3	1.280	13.9		1.273	10.7		1.246	10.2		1.439	13.2	t
ALT	0.563	7.1	0.572	10.6		0.550	13.9		0.501	11.5	tw	0.557	14.3	
AST	2.400	24.9	2.300	20.6		2.226	16.2		2.121	17.8		2.218	20.3	
BIL	7.300	27.0	7.075	13.8		6.919	6.1		7.156	11.4		6.481	5.3	

ALB	36.41	2.2	36.65	4.8		36.89	2.1		37.80	2.6	dTw	36.38	3.6	
TP	64.39	1.7	65.11	3.6		64.99	1.7		66.22	2.5	dT	64.60	2.1	
Glu	5.381	15.8	5.337	15.3		5.018	13.7		5.153	6.9		5.491	9.6	
CHOL	2.047	11.2	2.006	13.5		2.022	10.8		1.936	10.0		2.041	8.5	
TAG	1.021	24.6	1.104	12.8		1.051	31.3		1.028	30.5		1.087	35.9	
Crea	40.13	5.2	41.55	11.5		40.82	8.8		43.79	10.6	t	41.94	9.9	
Urea	5.135	6.3	4.899	11.2		4.987	5.6		4.882	8.9		4.965	10.0	
cHGB	69.73	66.1	61.09	33.7		58.42	21.8		52.57	29.5		49.35	27.4	
Ca	2.402	1.7	2.400	1.2		2.406	2.3		2.414	1.9		2.392	1.5	
Cl	102.1	1.3	101.8	1.2		101.8	1.0		102.2	0.7		102.1	1.2	
K	4.987	4.5	5.087	7.2		5.250	3.9	t	5.225	5.9		5.106	4.8	
Na	144.8	1.1	144.9	1.0		145.3	0.8		145.6	0.8		145.2	1.0	
P	2.374	13.1	2.443	7.9		2.499	6.1		2.408	6.5		2.313	7.7	
Urine	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
uVol	19.25	18.6	19.12	28.9		22.31	32.0		21.12	40.7		19.19	26.5	
uVolW	4.595	20.7	4.306	26.9		5.171	32.6		5.051	39.7		4.389	28.1	
uLeu	28.12	127.4	20.31	65.3		31.25	109.0		26.56	134.1		18.75	50.4	
uOsmoll	489.0	30.2	473.7	19.9		447.5	24.2		455.4	30.5		480.3	18.8	
uKeton	0.419	73.9	0.781	52.5	w	0.750	123.4		0.531	121.8		0.312	175.0	
upH	7.125	4.2	7.000	5.7		6.875	1.9	t	6.938	4.6		6.750	2.8	dTw
Organs	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Kidney	0.521	7.2	0.524	5.4		0.537	5.4		0.548	6.9		0.516	7.9	
Spleen	0.179	8.2	0.186	5.6		0.177	9.3		0.188	7.8		0.181	8.1	
Liver	2.157	4.7	2.186	6.0		2.238	5.6		2.242	4.5		2.167	1.8	
AdrenGl	0.0140	9.1	0.0131	15.4		0.0131	11.3		0.0128	11.6		0.0131	11.0	
Heart	0.232	2.8	0.242	3.6	w	0.239	5.9		0.239	6.2		0.231	5.7	
Thymus	0.091	13.7	0.096	17.4		0.096	17.5		0.104	17.7	tw	0.096	17.3	
Testis	0.808	10.3	0.802	6.8		0.819	4.6		0.834	6.5		0.806	7.8	
Epididymis	0.267	8.9	0.262	4.3		0.279	7.1		0.274	5.9		0.272	11.0	
Brain	0.493	8.4	0.481	7.0		0.507	4.1		0.512	6.1		0.494	11.0	
Immunology	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Granulocytes	73.80	17.7	77.50	12.4	t	73.57	16.8		73.69	19.3		73.55	16.1	

RespirBurst	64.42	19.6	66.30	14.9		60.14	19.6		60.58	26.0		63.17	14.8	
Con	77166	27.7	78118	20.9		70681	21.4		73844	22.6		83659	21.5	
PHA	50436	35.3	51083	13.4		50736	20.7		47416	20.2		53347	15.0	
PWM	25656	59.6	36631	32.4		25415	55.7		25404	42.9		46487	63.4	t
Medium	2370	38.1	2324	23.9		1944	30.9		1812	13.3		3003	46.9	
IprConA	34.82	25.0	37.43	19.7		40.15	15.2		41.48	34.1		34.65	43.3	
IprPHA	21.84	10.5	26.64	18.0		28.23	9.7		27.21	35.2		21.85	42.9	
IprPWM	10.48	29.0	15.28	21.6	d	12.71	26.4		13.95	40.8		15.14	30.0	t
Cytokines	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
IL2	4082	22.9	5378	9.2	dT	4499	14.6		4477	21.4		5501	15.5	D
IL4	28.71	84.4	14.62	34.9		10.87	6.9	t	11.32	15.3	t	14.06	14.2	
IL10	10265	80.0	9151	80.1		3256	48.5		3677	81.7		13753	94.7	
IL17A	183.6	66.0	145.5	32.6		119.3	39.9		98.0	46.7		197.9	59.4	
TNFa	17.91	40.5	25.07	6.9	dT	17.26	16.9		19.24	30.4		26.02	27.2	dT
IFNg	14401	18.4	19549	23.3	dT	12162	14.3		13601	26.5		16342	23.9	t
CellPhenotype	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
sp3	35.98	14.4	41.68	14.8		41.03	21.1		43.76	10.0	t	43.04	14.8	t
sp3_4	21.44	17.9	24.18	21.2		26.18	25.6		25.63	15.7	t	26.40	23.7	t
sp3_8	14.72	20.7	17.11	13.4		14.94	27.3		18.07	7.9	t	16.54	11.9	
sp3_45	24.90	4.3	25.57	1.9		24.70	5.5		24.58	5.4		25.23	2.9	
sp3_161	13.60	8.4	13.14	5.8		13.81	7.3		13.52	6.4		12.51	5.7	
ln3	57.78	7.0	59.17	6.6		53.89	20.9		52.69	23.9		54.56	19.2	
ln3_4	42.66	7.2	41.15	5.0		39.50	23.7		37.43	28.6		39.54	24.2	
ln3_8	15.47	8.5	17.85	11.5		15.02	13.0		16.21	16.5		15.58	7.8	
ln3_45	34.54	23.4	29.27	31.9	t	29.10	26.8		34.04	34.5		30.80	26.5	
ty3	18.41	25.3	18.65	20.0		19.29	12.5		19.98	11.4		20.23	14.5	
ty3_4	14.16	22.4	14.36	18.1		15.13	15.7		15.51	9.1		15.88	14.3	
ty3_8	7.925	21.6	8.390	23.6		8.240	8.3		8.350	12.7		8.995	18.0	
Hormone	Control		NK11-			NK33-			NK11+			NK33+		
Males	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Testosteron	2.579	34.3	2.784	59.1		1.974	56.6		3.026	63.7		2.170	25.6	
T3	0.750	8.1	0.816	16.7		0.745	12.1		0.748	12.6		0.805	10.0	w
T4	55.16	8.6	54.03	10.0		48.65	7.5	dw	55.08	6.8		53.36	9.9	

Table 13 Means and coefficient of variation (CV) for female rats. Means of GMO feeds which are significantly different from the Control feed 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. Note that Dunnett- and t-tests are based on a ANOVA with 5 treatment groups.

Weights	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
BodyWeight	244.8	7.0	242.3	7.7		248.7	5.9		246.7	5.5		252.5	4.9	
growthRate	0.120	26.5	0.158	35.5	t	0.154	24.8	w	0.154	30.6		0.177	12.8	dTW
FeedMean	14.31	9.6	12.96	11.8	d	13.91	8.4		13.77	5.1		14.03	6.5	
Haematology	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
WBC	5.519	26.7	5.369	18.0		6.063	19.2		4.862	14.4		5.425	23.3	
RBC	6.726	3.3	6.718	8.2		6.866	2.5		6.808	7.3		6.657	4.5	
HGB	13.72	3.3	13.56	7.6		14.11	1.3	w	13.84	5.9		13.58	4.8	
HCT	39.41	2.8	39.10	8.4		40.53	2.3		39.92	6.4		39.63	4.6	
MCV	58.66	2.4	58.23	3.1		59.06	1.8		58.72	2.8		59.59	2.2	
MCH	20.41	3.4	20.21	2.7		20.57	2.1		20.38	4.1		20.41	3.0	
MCHC	34.81	2.4	34.71	2.2		34.81	1.8		34.71	2.7		34.26	1.4	
PLT	901.9	12.6	824.6	17.5		901.4	10.1		813.0	12.3		903.6	10.4	
LYMR	75.17	7.7	75.06	4.7		77.48	4.5		76.59	6.4		77.11	6.4	
LYMA	4.225	32.4	4.037	17.2		4.688	19.4		3.713	18.5		4.150	23.4	
diffWBC	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Lymphocytes	69.22	8.6	69.50	9.0		70.41	5.4		71.88	7.2		71.12	5.7	
Neutrophils	28.09	16.8	28.47	20.2		26.84	16.6		26.12	19.2		26.06	16.1	
Monocytes	1.531	49.7	1.125	58.2		1.438	57.9		1.125	33.6		1.250	45.4	
Eosinophils	1.156	104.0	0.906	77.9		1.312	71.8		0.875	117.3		1.562	62.7	
ClinChem	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
ALP	0.581	23.3	0.578	9.6		0.654	10.0		0.584	12.1		0.579	18.1	
ALT	0.470	12.9	0.417	10.3		0.525	23.0		0.453	29.6		0.424	19.1	
AST	2.299	29.0	2.225	15.5		2.326	11.0		2.205	24.1		2.366	18.7	
BIL	7.575	12.8	7.113	7.1		8.131	13.8		7.844	12.4		8.506	24.2	

ALB	42.49	8.3	41.50	6.0		44.59	6.1		44.12	6.0		43.94	6.5	
TP	68.67	4.9	67.24	4.3		71.01	3.6		70.86	3.4		70.49	4.2	
Glu	5.188	6.8	5.117	23.1		4.404	15.4	dw	5.047	22.2		5.625	18.3	
CHOL	1.906	11.0	2.054	13.9		2.009	14.0		1.997	12.3		1.992	16.9	
TAG	0.524	24.7	0.621	31.9		0.647	13.4	tW	0.602	19.6		0.566	15.1	
Crea	40.91	5.9	42.01	6.8		42.86	9.3		40.75	12.4		41.90	13.5	
Urea	5.864	6.0	5.943	14.4		5.321	12.6		5.285	12.1		5.590	10.2	
cHGB	41.76	53.2	39.85	24.4		49.50	38.4		49.29	48.9		56.36	56.6	
Ca	2.472	2.8	2.454	1.4		2.489	1.6		2.495	2.1		2.494	2.0	
Cl	102.1	1.0	102.0	1.5		100.6	1.1	tw	101.6	1.1		102.1	1.8	
K	4.450	4.8	4.463	4.6		4.519	6.4		4.537	4.7		4.700	7.0	
Na	145.1	1.3	144.2	0.8		144.7	0.8		145.1	1.2		144.9	1.9	
P	2.040	11.3	1.964	11.7		2.153	16.8		1.931	11.1		1.956	11.4	
Urine	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
uVol	17.12	24.7	17.06	26.5		16.75	42.2		16.69	39.2		13.56	22.1	
uVolW	7.418	24.8	7.490	27.3		7.089	42.3		7.128	41.1		5.698	25.5	
uLeu	10.94	95.4	7.81	119.0		6.25	151.2		20.31	118.3		10.94	155.0	
uOsmoll	412.8	27.2	417.2	25.0		434.6	28.2		413.6	31.0		427.1	14.7	
uKeton	0.156	119.0	0.031	282.8		0.156	169.7		0.031	282.8		0.000	*	t
upH	6.625	4.0	6.438	4.5	w	6.500	3.6		6.438	4.0		6.469	2.5	
Organs	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Kidney	0.598	8.5	0.570	8.7		0.570	8.0		0.579	5.8		0.581	4.3	
Spleen	0.245	7.2	0.225	12.4		0.245	9.0		0.217	10.0	dTw	0.245	7.4	
Liver	2.444	5.8	2.554	7.6		2.498	12.8		2.531	6.6		2.391	3.5	
AdrenGl	0.0312	10.5	0.0280	8.7	t	0.0283	6.2	t	0.0284	10.0	t	0.0292	5.8	
Heart	0.320	5.4	0.311	13.0		0.306	6.2	w	0.308	6.7		0.320	9.9	
Thymus	0.117	17.2	0.120	14.2		0.122	20.8		0.115	18.5		0.117	13.3	
Uterus	0.251	19.5	0.256	32.4		0.244	19.6		0.254	22.8		0.313	16.8	
Ovary	0.0381	6.9	0.0338	8.4	dw	0.0355	16.1		0.0362	7.8		0.0324	3.7	dTW
Brain	0.840	6.8	0.854	4.7		0.833	5.4		0.830	6.3		0.817	2.9	
Immunology	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Granulocytes	76.77	4.1	70.96	8.3	t	77.69	4.2		73.50	4.3		73.06	5.0	

RespirBurst	56.32	23.5	52.40	3.4		54.54	11.2		52.28	22.3		57.86	13.5	
Con	51683	37.3	61830	28.7		54694	46.2		65130	39.7		63123	33.0	
PHA	29498	44.4	36438	47.0		27484	57.3		34743	30.3		36871	22.2	
PWM	18769	80.4	23672	26.3		18101	53.4		22491	46.0		25289	62.6	
Medium	1429	38.1	1685	13.0		1422	45.4		1788	29.8		1870	37.2	
IprConA	38.00	26.3	37.47	29.0		37.84	20.8		36.62	15.8		35.99	13.7	
IprPHA	21.40	26.0	21.77	42.8		19.88	28.0		21.70	36.0		21.66	28.4	
IprPWM	11.38	47.7	13.62	27.5		12.69	20.2		12.18	21.3		13.20	42.1	
Cytokines	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
IL2	3821	15.2	4491	6.3		3240	39.1		3936	6.2		3434	38.4	
IL4	14.16	14.3	12.13	17.6		9.53	50.0		25.94	55.2		11.11	62.9	
IL10	10919	105.0	9553	102.9		4401	60.6		11809	72.1		8175	143.3	
IL17A	201.4	21.5	115.3	24.4	dT	103.2	53.6		184.6	27.0		104.6	66.1	t
TNFa	19.85	26.2	15.62	6.9		12.20	48.0		21.70	4.7		13.71	46.8	
IFNg	12364	37.7	9788	21.4		8656	63.6		13476	14.7		9123	69.0	
CellPhenotype	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
sp3	37.10	5.8	41.81	9.0		38.60	12.5		39.74	14.8		41.06	8.7	
sp3_4	20.61	6.9	24.91	15.9		21.71	25.8		24.00	21.3		23.20	12.1	
sp3_8	16.12	5.6	15.67	10.0		15.53	4.8		15.19	11.1		17.16	11.6	
sp3_45	25.56	1.4	25.68	2.5		25.53	2.9		25.06	1.0		25.68	3.4	
sp3_161	11.15	5.5	10.88	2.8		10.97	5.2		10.86	1.0		11.28	9.6	
ln3	61.37	4.2	61.10	7.7		61.87	4.5		61.86	6.4		64.77	4.4	
ln3_4	44.76	5.5	45.52	9.4		45.08	8.9		46.07	8.6		46.96	5.9	
ln3_8	17.00	3.6	16.04	7.9		16.91	10.3		16.25	3.6		18.07	10.5	
ln3_45	20.39	33.8	25.62	36.1		21.75	35.4		22.03	37.5		18.39	29.9	
ty3	20.05	20.5	19.65	17.8		20.00	10.4		20.63	23.2		19.91	17.9	
ty3_4	15.24	24.6	15.27	19.8		15.77	11.9		15.40	24.6		14.80	16.7	
ty3_8	8.715	28.6	7.940	21.5		8.325	31.3		8.645	27.8		8.110	28.7	
Hormone	Control		NK11-			NK33-			NK11+			NK33+		
Females	Mean	CV	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig	Mean	CV	Sig
Estradiol	7.740	58.3	4.975	69.4	t	4.634	29.2	tw	5.044	41.1	tw	4.644	27.8	tW
T3	0.757	21.4	0.766	15.2		0.690	11.9		0.731	18.1		0.635	18.8	
T4	34.20	23.2	31.02	27.4		33.97	33.4		37.79	24.0		25.41	33.4	tw

For the oestrous cycle data (in group eCycle) the following was found.

The duration of the first oestrous cycle is mostly 5 days with the exception of the following four female rats: 114 (6 days, NK11-), 116 (6 days, NK11-), 130 (3 days, NK33+) and 162 (unfinished, NK11+). According to Fishers exact test 2 out of 16 with a different cycle duration (NK11-) is not significantly different from 0 out of 16 (Control), with p-value equal to 0.484. So there is no indication that the duration of the first cycle is different between the GMO feeding groups and the control group. Note that Fishers exact test assumes that animals (even within a cage) have independent cycle duration.

The duration of the second oestrous cycle is given in Table 14; there are many rats with an unfinished cycle. The two NK11- rats with a 6 days second cycle are different from the ones with a first cycle of 6 days. Pairwise testing whether the probability of an unfinished second cycle is equal, again with Fishers Exact test and assuming that animals (even within a cage) have independent cycle duration, reveals that there are significant differences between Control and NK11+ (p=0.009), between NK11- and NK33- (p=0.032), and between NK11- and NK11+ (p=0.003). There is also an indication of a difference between NK11+ and NK33+ (p=0.054). Testing, in the same way, whether the probability of a second cycle of 5 Days is equal results in less significant results: there is only a significant difference between Control and NK11+ (p=0.023) and an indication for a difference between NK11- and NK11+ (p=0.054).

Table 14 Number of rats in each feeding group with specified length of the second oestrous cycle.

Group	unfinished	3 Days	4 Days	5 Days	6 Days
Control	6	-	1	9	-
NK11-	5	1	-	8	2
NK33-	12	-	-	4	-
NK11+	14	-	-	2	-
NK33+	8	-	1	7	-

The number of irregular cycles out of two cycles follows a binomial distribution with binomial total 2 and probability π, and the question is whether these probabilities are different between groups. This was tested with logistic regression, again assuming that rats are independent with respect to irregularity of cycles, by fitting the Block + Group model. Pairwise differences between groups were tested by a likelihood ratio test, giving mean probabilities and p-values for differences with the Control in Table 15. There is some indication (p=0.061) that the Control group has a larger probability of an irregular cycle than the NK33- and NK33+ groups. There are no significant differences between the GMO feed groups.

Table 15 Estimated probabilities of having an irregular oestrous cycle and p-values for differences between the control feed and the GMO feeds resulting from a logistic regression.

Group	Mean Probability	p-value
Control	0.2188	-
NK11-	0.1250	0.307
NK33-	0.0625	0.061
NK11+	0.1250	0.307
NK33+	0.0625	0.061

4.4 Standardised effect sizes

4.4.1 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), again using the residual standard error with 28 degrees of freedom, and their exact 95% confidence intervals were calculated. This employed 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 t-test is smaller than 5%.

4.4.2 Results

SES intervals were calculated for all 79 variables. Results for the four treatment groups, separately for males and females, are given in Figure 21 to Figure 24. It can be noted that almost all intervals (616 out of 632, i.e. 97.5%) extend outside the +/-1 SD limits.

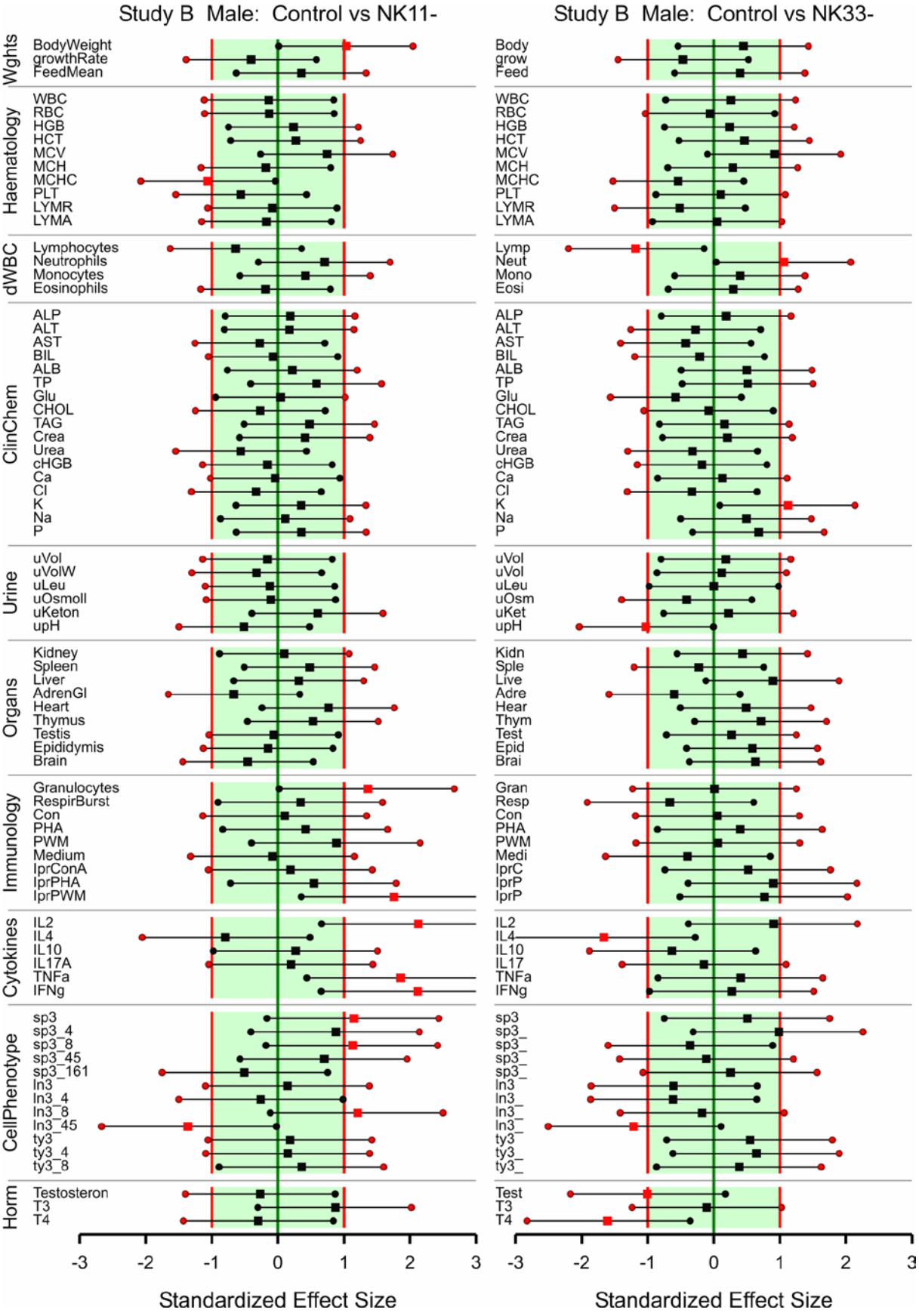

Figure 21 Confidence intervals for Standardized Effect Sized (SES) for male rats for GMO feeds NK11- and NK33- versus the control feed.

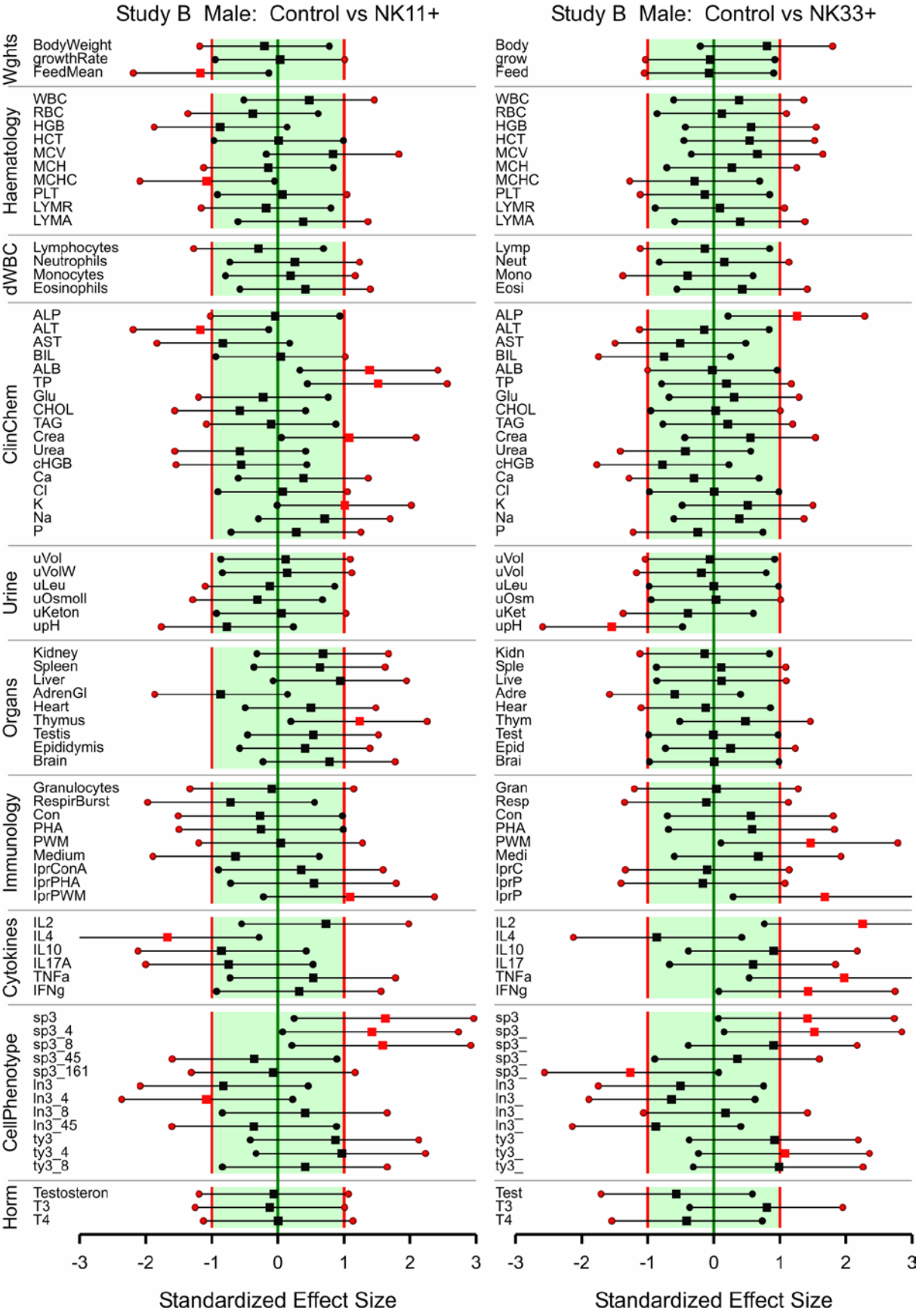

Figure 22 Confidence intervals for Standardized Effect Sized (SES) for male rats for GMO feeds NK11+ and NK33+ versus the control feed.

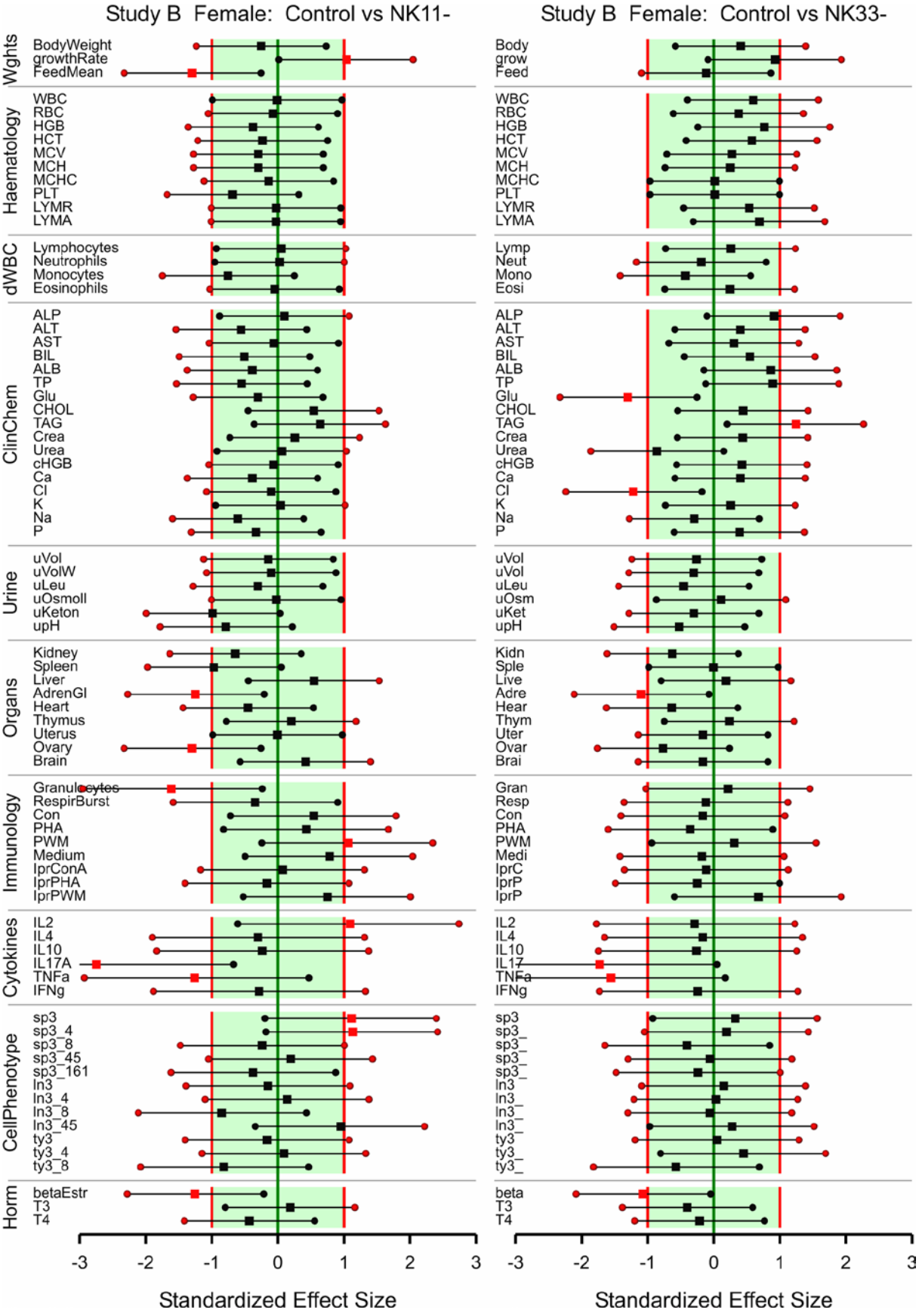

Figure 23 Confidence intervals for Standardized Effect Sized (SES) for female rats for GMO feeds NK11- and NK33- versus the control feed.

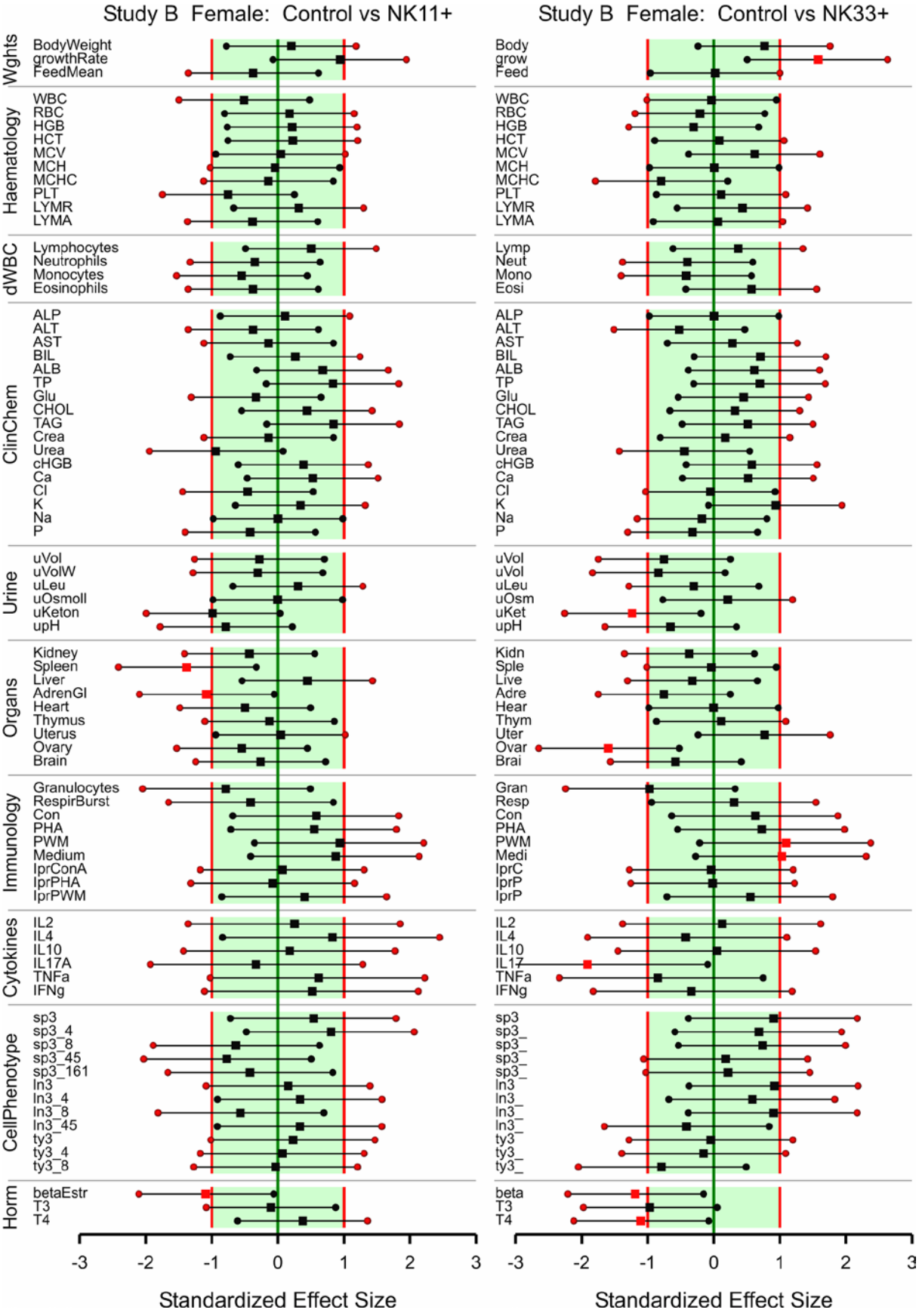

Figure 24 Confidence intervals for Standardized Effect Sized (SES) for female rats for GMO feeds NK11+ and NK33+ versus the control feed.

4.5 Factorial analysis

4.5.1 Method

The purpose of this oral toxicity study is to assess the effects of GM maize NK 603 when fed to rats for a period of 90 days at an incorporation rate of 11% and 33% in the feed. Table 16 lists the maize type and incorporation rate of the 5 diets as well as the number of animals used for every diet.

Table 16 Diets used in the 90-day feeding trial study 1 with GM inclusion rates 11% and 33%.

Group	Isogenic maize (% of diet)	NK603 only (% of diet)	NK603 + Roundup (% of diet)	No. of Males	No. of Females
Control	33	0	0	16	16
NK11-	22	11	0	16	16
NK33-	0	33	0	16	16
NK11+	22	0	11	16	16
NK33+	0	0	33	16	16

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:

$y_{0k} = \mu + \delta_k + \epsilon_{ijk}$ 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 estimates the difference between the four GM groups (averaged) and the control. This is only a useful estimate if the other three fixed terms can be assumed to be zero. The main effect $IR_2 - IR_1$ estimates the difference between the groups with GM inclusion rates 33% and 11%, and similarly, the main effect $RU_2 - RU_1$ estimates the difference between the groups with and without Roundup. These main effects are only useful when there is no interaction between GM inclusion rate and Roundup.

4.5.2 Results

The significance results from fitting the factorial model to the 79 variables are given in Table 18. The ratios for the significant cases are shown in Table 19 and Table 20.

As an example, consider the results for growthRate, HGB and PLT for females; the means, after the usual log transformation, for these variables are given in Table 17. For growthRate the means suggest that there is an interaction between GM inclusion rate and Roundup, because there is hardly an effect of GM inclusion rate for RoundUp-, while there appears to be an effect of GM inclusion rate for RoundUp+. However this interaction is, according to Table 18, not significant with p-value 0.301. The main GMO effect, i.e. comparing the mean of the four GMO feeds, which equals 0.161, to the control feed, is however very significant with p-value 0.008. The corresponding ratio equals exp(0.161 – 0.120) = 1.04, and this ratio is given in Table 20.

For HGB there is a significant interaction (p=0.026, Table 18) between GM inclusion rate and RoundUp with, for RoundUp- increasing means when GM inclusion rate is increased, and for RoundUp+ decreasing values when GM inclusion rate is increased (Table 17). This implies that the effect of GM inclusion rate depends on whether or not RoundUp is applied. In Table 20 the ratios for HGB are expressed relative to the NK11- feeding group.

For PLT there is not a significant interaction because the effect of GM inclusion rate, see Table 17, is very similar for RoundUp- (difference equals 0.09) and RoundUp+ (difference equals 0.11,). The mean effect across the two roundup treatment equals 0.10, and this is significantly different from zero (p=0.035, Table 18). The corresponding ratio in Table 20 equals exp(0.10) = 1.10.

Table 17 Means for three example variables in females which exhibit significant main effects or a significant interaction between GM inclusion rate of the GMO feed (IR-11 and IR-33) and application of RoundUp (-/+) when applying the factorial model.

	growthRate			HGB			PLT		
	Control	IR-11	IR-33	Control	IR-11	IR-33	Control	IR-11	IR-33
Control	0.120	-	-	2.62	-	-	6.79	-	-
RoundUp-	-	0.158	0.154	-	2.60	2.65	-	6.70	7.79
RoundUp+	-	0.154	0.177	-	2.63	2.61	-	6.69	6.80

The p-values in Table 18 can be summarized as follows. In 17 cases (11% of all 158 cases) there was a significant difference at the 5% level between GM and non-GM. In 26 cases (16% of all 158 cases) there was a significant interaction term at the 5% level. In these latter cases the main effects should not be used for assessing effects of GM inclusion rate or Roundup. Among the remaining 132 cases, i.e. where the interaction is not significant, there were 10 cases (8%) with a significant differences between GM inclusion rates, and 3 cases (2%) with a significant difference between the GMO feeds with and without roundup.

Table 18 P values for significance of effects obtained with the factorial model. P-values smaller than 0.01/0.05 have a gold/yellow background.

Response	Males				Females			
Weights	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
BodyWeight	0.198	0.558	0.220	0.032	0.483	0.093	0.257	0.891
growthRate	0.574	0.832	0.240	0.968	0.008	0.463	0.440	0.301
FeedMean	0.756	0.118	0.009	0.145	0.272	0.034	0.144	0.273
Haematology	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
WBC	0.542	0.676	0.307	0.495	0.982	0.135	0.120	0.857
RBC	0.779	0.425	0.923	0.549	0.869	0.927	0.639	0.246
HGB	0.918	0.051	0.276	0.052	0.852	0.390	0.503	0.026
HCT	0.422	0.316	0.800	0.636	0.682	0.353	0.964	0.190
MCV	0.056	0.999	0.811	0.625	0.691	0.115	0.343	0.997
MCH	0.882	0.219	0.974	0.940	0.956	0.405	0.987	0.491
MCHC	0.070	0.075	0.742	0.712	0.506	0.485	0.257	0.263
PLT	0.742	0.518	0.587	0.231	0.412	0.035	0.970	0.818
LYMR	0.666	0.823	0.471	0.327	0.433	0.342	0.734	0.532
LYMA	0.681	0.735	0.212	0.769	0.832	0.109	0.177	0.704
diffWBC	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
Lymphocytes	0.165	0.593	0.059	0.328	0.462	0.921	0.430	0.633
Neutrophils	0.178	0.720	0.065	0.525	0.569	0.708	0.418	0.816
Monocytes	0.704	0.401	0.161	0.429	0.183	0.525	0.757	0.785
Eosinophils	0.550	0.486	0.306	0.514	0.808	0.090	0.998	0.364
ClinChem	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
ALP	0.323	0.076	0.243	0.077	0.484	0.325	0.214	0.205
ALT	0.375	0.418	0.096	0.045	0.508	0.264	0.302	0.128
AST	0.206	0.810	0.372	0.509	0.810	0.272	0.889	0.938
BIL	0.531	0.196	0.554	0.363	0.530	0.043	0.198	0.392
ALB	0.198	0.122	0.369	0.024	0.274	0.105	0.256	0.073
TP	0.087	0.059	0.396	0.087	0.246	0.075	0.104	0.034
Glu	0.776	0.904	0.383	0.115	0.355	0.763	0.021	0.018
CHOL	0.574	0.272	0.780	0.562	0.278	0.756	0.753	0.978
TAG	0.640	0.994	0.460	0.378	0.050	0.696	0.463	0.199
Crea	0.166	0.312	0.160	0.652	0.650	0.491	0.357	0.855
Urea	0.242	0.593	0.866	0.895	0.177	0.552	0.415	0.055
cHGB	0.299	0.736	0.170	0.776	0.411	0.342	0.400	0.669
Ca	0.914	0.472	0.996	0.234	0.507	0.279	0.153	0.268
Cl	0.716	0.923	0.304	0.926	0.256	0.321	0.262	0.040
K	0.068	0.703	0.939	0.083	0.329	0.262	0.173	0.596
Na	0.293	0.937	0.496	0.328	0.499	0.858	0.313	0.494
P	0.504	0.785	0.168	0.243	0.667	0.257	0.266	0.386
Urine	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
uVol	0.960	0.816	0.964	0.467	0.369	0.411	0.379	0.617
uVolW	0.874	0.873	0.829	0.283	0.335	0.306	0.305	0.647
uLeu	0.879	0.735	1.000	1.000	0.633	0.290	0.290	0.523
uOsmoll	0.617	0.959	0.735	0.367	0.854	0.624	0.863	0.912
uKeton	0.760	0.255	0.111	0.929	0.034	0.540	0.198	0.198
upH	0.021	0.079	0.284	0.719	0.092	0.581	0.854	0.854
Organs	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
Kidney	0.502	0.498	0.986	0.113	0.199	0.911	0.511	0.955
Spleen	0.532	0.092	0.493	0.794	0.141	0.003	0.538	0.594
Liver	0.161	0.725	0.834	0.057	0.593	0.119	0.394	0.565

AdrenGl	0.095	0.629	0.791	0.772	0.013	0.507	0.477	0.801
Heart	0.311	0.213	0.224	0.634	0.322	0.662	0.415	0.345
Thymus	0.071	0.419	0.512	0.193	0.791	0.704	0.526	0.773
Testis	0.646	0.768	0.648	0.225	-	-	-	-
Epididymis	0.493	0.422	0.747	0.219	-	-	-	-
Uterus	-	-	-	-	0.689	0.426	0.176	0.221
Ovary	-	-	-	-	0.013	0.468	0.918	0.033
Brain	0.548	0.664	0.391	0.014	0.712	0.214	0.132	0.710
Immunology	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
Granulocytes	0.417	0.103	0.060	0.052	0.063	0.034	0.611	0.012
RespirBurst	0.476	0.576	0.481	0.036	0.722	0.200	0.620	0.492
Con	0.779	0.283	0.859	0.231	0.330	0.362	0.257	0.300
PHA	0.483	0.264	0.493	0.243	0.406	0.399	0.108	0.192
PWM	0.139	0.410	0.444	0.006	0.047	0.409	0.365	0.216
Medium	0.780	0.175	0.480	0.035	0.132	0.270	0.084	0.133
IprConA	0.552	0.867	0.526	0.282	0.992	0.684	0.921	0.909
IprPHA	0.264	0.621	0.152	0.151	0.748	0.977	0.657	0.839
IprPWM	0.004	0.581	0.726	0.040	0.151	0.917	0.515	0.752
Cytokines	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
IL2	0.002	0.669	0.941	0.001	0.477	0.061	0.566	0.109
IL4	0.006	0.931	0.917	0.030	0.963	0.147	0.252	0.081
IL10	0.844	0.241	0.564	0.002	0.864	0.825	0.333	0.886
IL17A	0.948	0.178	0.780	0.030	0.002	0.445	0.012	0.005
TNFa	0.008	0.985	0.733	0.001	0.085	0.034	0.005	0.134
IFNg	0.019	0.312	0.379	0.001	0.827	0.275	0.341	0.234
CellPhenotype	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
sp3	0.009	0.249	0.068	0.549	0.086	0.552	0.991	0.122
sp3_4	0.008	0.772	0.147	0.991	0.094	0.152	0.823	0.264
sp3_8	0.055	0.007	0.027	0.269	0.736	0.106	0.307	0.044
sp3_45	0.715	0.890	0.417	0.047	0.774	0.335	0.320	0.107
sp3_161	0.329	0.553	0.149	0.015	0.609	0.290	0.573	0.487
ln3	0.274	0.547	0.238	0.148	0.508	0.152	0.149	0.523
ln3_4	0.119	0.900	0.247	0.277	0.500	0.848	0.305	0.621
ln3_8	0.321	0.036	0.540	0.122	0.720	0.006	0.096	0.349
ln3_45	0.028	0.610	0.077	0.362	0.477	0.061	0.082	0.917
ty3	0.130	0.563	0.154	0.660	0.965	0.934	0.677	0.497
ty3_4	0.091	0.401	0.097	0.591	0.777	0.850	0.385	0.426
ty3_8	0.194	0.409	0.367	0.446	0.179	0.479	0.437	0.177
Hormone	GMO	InclRate	Roundup	Interact	GMO	InclRate	Roundup	Interact
Testosteron	0.241	0.094	0.372	0.743	-	-	-	-
betaEstr	-	-	-	-	0.007	0.906	0.950	0.692
T3	0.373	0.943	0.909	0.014	0.421	0.049	0.233	0.696
T4	0.158	0.024	0.046	0.221	0.390	0.084	0.909	0.023

Table 19 Ratios for significant variables at the 5% level in the factorial analysis for males. Main effects are "GM vs Contrl" (ratio of the mean of the four GM feeds vs the control feed), "33 vs 11" (ratio of the two GM feeds with 33% GM inclusion rate vs the two feeds with 11% GM inclusion rate) and "+RU vs -RU" (ratio of the two GM feeds with roundup vs the two feeds without roundup). The interaction ratios are scaled such that NK11- equals 1. The InclRate and RndUp main effects are only given when the interaction is not significant.

Males Group	Variable	GM vs Contrl	33 vs 11	+RU vs -RU	Interaction			
					NK11-	NK33-	NK11+	NK33+
Weights	BodyWeight	-	-	-	1.00	0.97	0.94	0.99
Weights	FeedMean	-	-	0.95	-	-	-	-
ClinChem	ALT	-	-	-	1.00	0.95	0.87	0.97
ClinChem	ALB	-	-	-	1.00	1.01	1.03	0.99
Urine	upH	0.79	-	-	-	-	-	-
Organs	Brain	-	-	-	1.00	1.06	1.06	1.02
Immunology	RespirBurst	-	-	-	1.00	0.90	0.89	0.95
Immunology	PWM	-	-	-	1.00	0.71	0.70	1.27
Immunology	Medium	-	-	-	1.00	0.90	0.82	1.30
Immunology	IprPWM	1.35	-	-	1.00	0.80	0.86	0.98
Cytokines	IL2	1.25	-	-	1.00	0.83	0.81	1.02
Cytokines	IL4	0.70	-	-	1.00	0.78	0.78	0.98
Cytokines	IL10	-	-	-	1.00	0.52	0.44	1.59
Cytokines	IL17A	-	-	-	1.00	0.82	0.58	1.26
Cytokines	TNFa	1.35	-	-	1.00	0.69	0.72	1.03
Cytokines	IFNg	1.30	-	-	1.00	0.63	0.64	0.84
CellPhenotype	sp3	1.15	-	-	-	-	-	-
CellPhenotype	sp3_4	1.17	-	-	-	-	-	-
CellPhenotype	sp3_8	-	0.86	1.12	-	-	-	-
CellPhenotype	sp3_45	-	-	-	1.00	0.97	0.96	0.99
CellPhenotype	sp3_161	-	-	-	1.00	1.05	1.03	0.95
CellPhenotype	ln3_8	-	0.91	-	-	-	-	-
CellPhenotype	ln3_45	0.87	-	-	-	-	-	-
Hormone	T3	-	-	-	1.00	0.92	0.92	0.99
Hormone	T4	-	0.94	1.06	-	-	-	-

Table 20 Ratios for significant variables at the 5% level in the factorial analysis for females. Main effects are "GM vs Contrl" (ratio of the mean of the four GM feeds vs the control feed), "33 vs 11" (ratio of the two GM feeds with 33% GM inclusion rate vs the two feeds with 11% GM inclusion rate) and "+RU vs -RU" (ratio of the two GM feeds with roundup vs the two feeds without roundup). The interaction ratios are scaled such that NK11- equals 1. The InclRate and RndUp main effects are only given when the interaction is not significant.

Females Group	Variable	GM vs Contrl	33 vs 11	+RU vs -RU	Interaction			
					NK11-	NK33-	NK11+	NK33+
Weights	growthRate	1.04	-	-	-	-	-	-
Weights	FeedMean	-	1.06	-	-	-	-	-
Haematology	HGB	-	-	-	1.00	1.04	1.02	1.00
Haematology	PLT	-	1.11	-	-	-	-	-
ClinChem	BIL	-	1.10	-	-	-	-	-
ClinChem	TP	-	-	-	1.00	1.06	1.05	1.05
ClinChem	Glu	-	-	-	1.00	0.87	1.00	1.11
ClinChem	TAG	1.17	-	-	-	-	-	-
ClinChem	Cl	-	-	-	1.00	0.99	1.00	1.00
Urine	uKeton	0.78	-	-	-	-	-	-
Organs	Spleen	-	1.11	-	-	-	-	-
Organs	AdrenGl	0.91	-	-	-	-	-	-
Organs	Ovary	0.90	-	-	1.00	1.05	1.08	0.97
Immunology	Granulocytes	-	-	-	1.00	1.11	1.05	1.04
Immunology	PWM	1.49	-	-	-	-	-	-
Cytokines	IL17A	0.63	-	-	1.00	1.33	1.96	1.26
Cytokines	TNFa	-	0.84	1.30	-	-	-	-
CellPhenotype	sp3_8	-	-	-	1.00	0.98	0.96	1.09
CellPhenotype	ln3_8	-	1.08	-	-	-	-	-
Hormone	betaEstr	0.64	-	-	-	-	-	-
Hormone	T3	-	0.88	-	-	-	-	-
Hormone	T4	-	-	-	1.00	1.08	1.34	0.79

4.6 Correlation analysis

4.6.1 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 18 and Figure 19) or to limits calculated from historical data (as in Figure 9 - Figure 16, or rescaled, as in Figure 5 - Figure 8).

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 eight points in each graph are based on the cage means in the eight blocks of the study.

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 5 - Figure 8.

4.6.2 Results

The correlation plots for three liver-related and three kidney-related variables in males and females are shown in Figure 25 - Figure 28. In most cases there appears to be no clear correlation. Exceptions are a significant positive correlation for liver weight versus CHOL for NK33+ in females (corr = 0.925, p-value = 0.001), and for NK33- in females significant negative correlations for kidney weight versus Urea (corr = -0.754, p-value=0.031) and Crea (corr = -0.719, p-value=0.045). Note that in the latter case there is also a positive correlation for Urea versus Crea (corr = 0.881, p-value = 0.004). Exceedance of a threshold level (kidney weight in NK33- Males, liver weight in NK33- Females) is not accompanied by exceedance of the threshold in the other plotted variable.

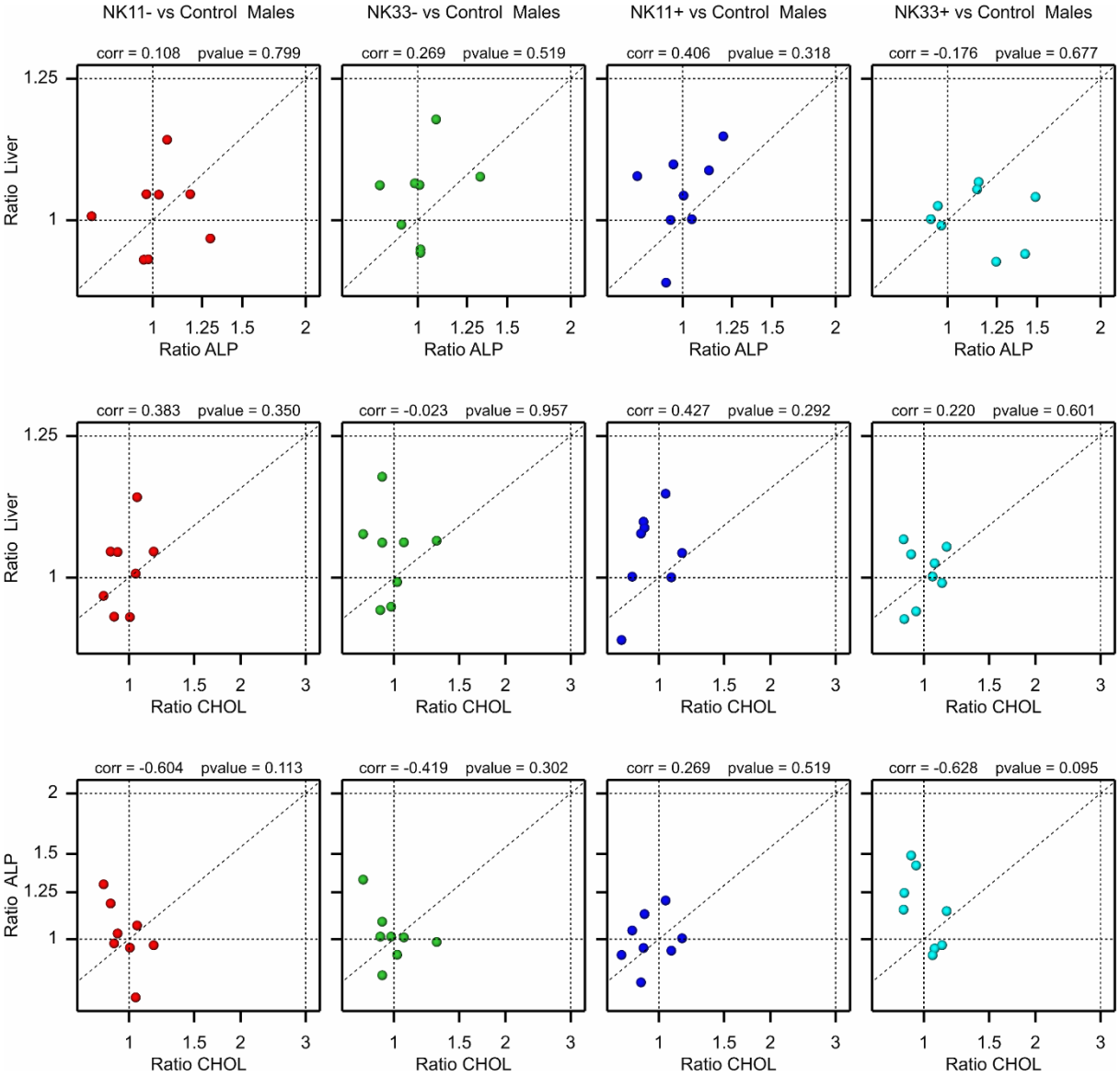

Figure 25 Pairwise results for variables with set target effect sizes related to liver damage in Males. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in the eight blocks in the study. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong *et al* (2017).

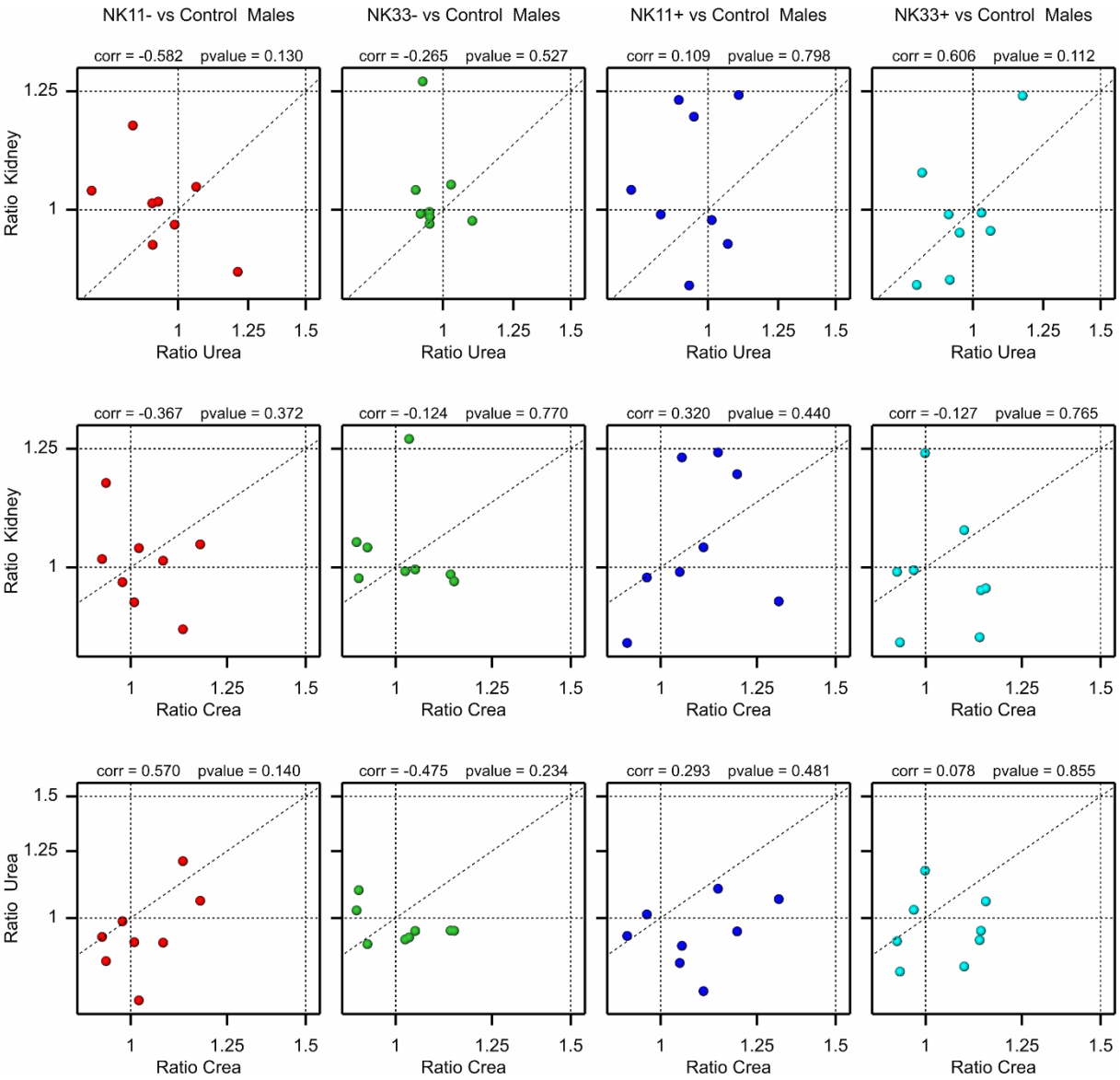

Figure 26 Pairwise results for variables with set target effect sizes related to kidney damage in Males. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in the eight blocks in the study. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong *et al* (2017).

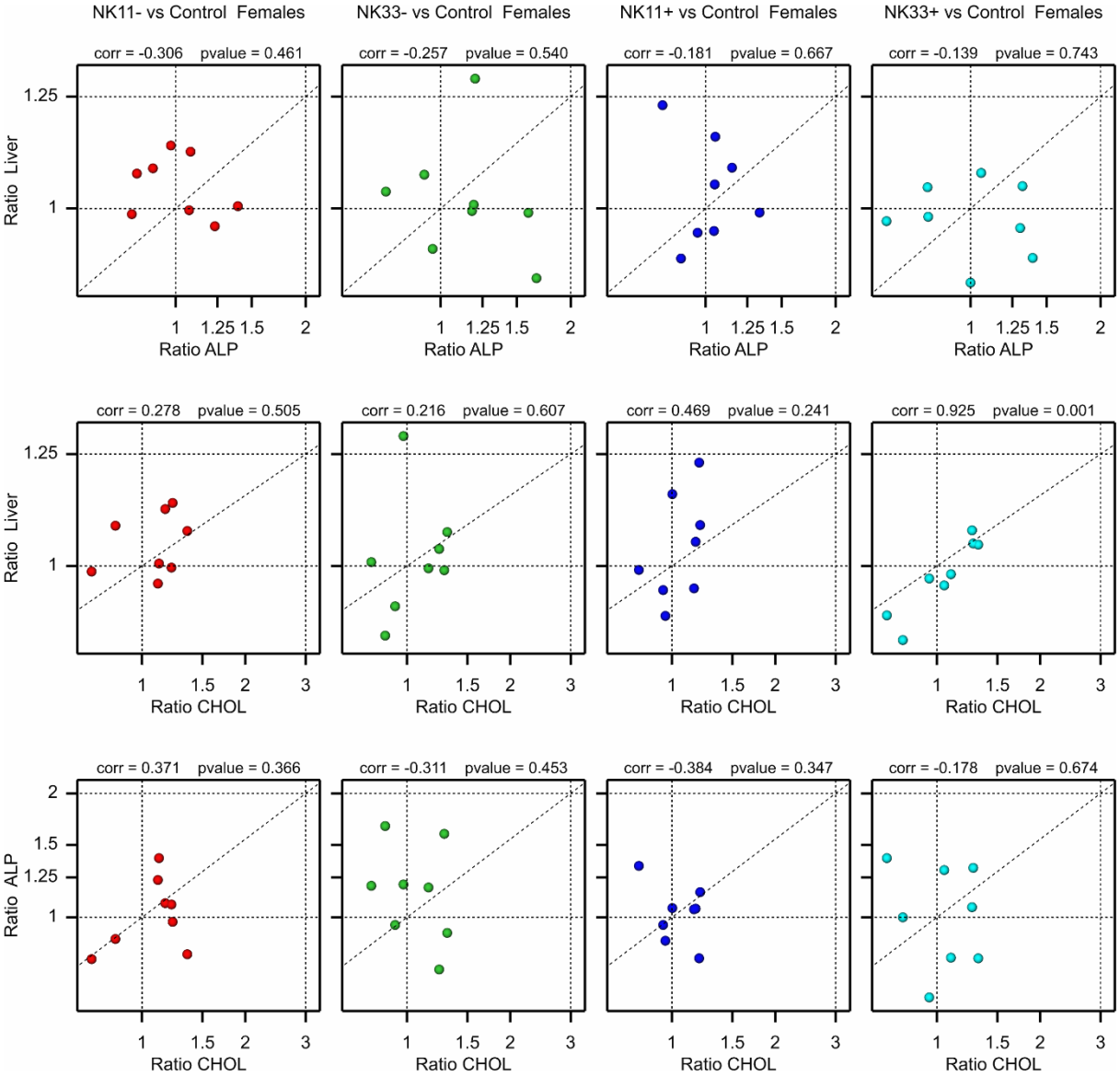

Figure 27 Pairwise results for variables with set target effect sizes related to liver damage in Females. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in the eight blocks in the study. Horizontal and vertical lines represent a ratio of 1 and the target effect sizes from Hong *et al* (2017).

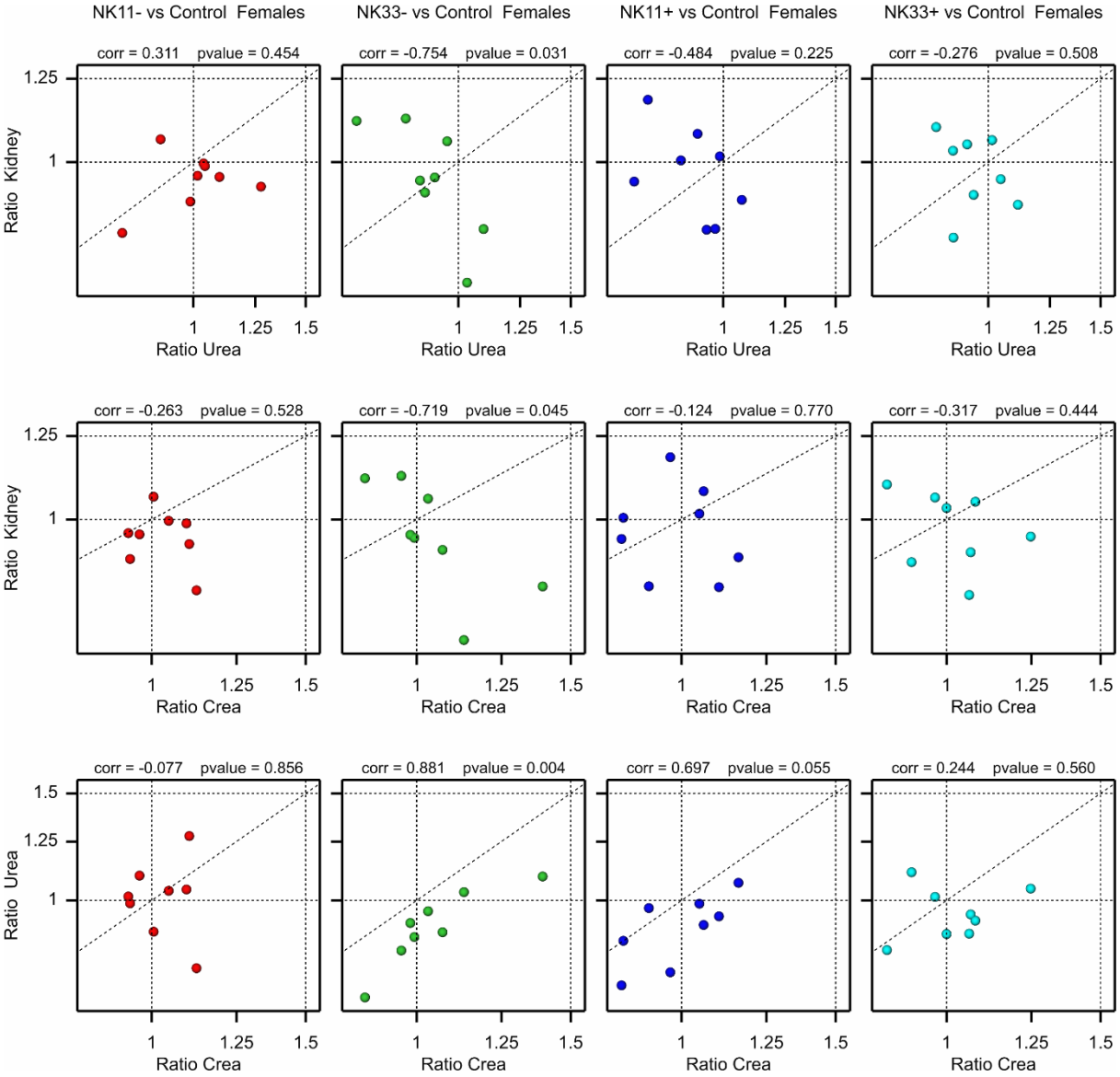

Figure 28 Pairwise results for variables with set target effect sizes related to kidney damage in Females. In each row there are four graphs corresponding to the four GM feed groups. Points represent the ratio of the cage mean for the GM group vs. the cage mean for the control feed in the eight blocks in the study. 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 B have been analysed following six approaches. For comparisons between a GM feeding group and the control feed for a single variable, these approaches were two forms of equivalence analysis (4.1, 4.2), the traditional approach focusing on significant differences (4.3), and the standardised effect size (SES) approach (4.4). In addition, a factorial analysis (4.5) and a correlation analysis (4.6) allowed some limited forms of integration over dose groups or variables, respectively.

Among the two forms of equivalence analysis, the approach with given external equivalence limits (4.2) is the most simple 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 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 42 variables, which were also observed in five preceding studies in the same test facility in the GRACE project. Van der Voet *et al* (2017) discusses this new method which was developed in the G-TwYST project.

Given tentative settings for regulatory parameters, equivalence was established in 97% of cases (70/72) for the approach with external equivalence limits and for 94% of cases (301/320) for the approach based on the historical GRACE data. 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 at 8 experimental units). Note that test results could be different if these parameters were chosen differently.

In all cases where equivalence was not established by means of the statistical test, 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). The proportion of non-significant equivalence tests (3% or 6%) was close to 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). 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. If performed using t-tests after applying an ANOVA model, there were 55 significant differences at the 5% significance level. This is 9% of the total number of comparisons (632), and higher than the nominal 5% level that could be expected. 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 17 (3% of all comparisons, 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.

OECD (2012) also proposes non-parametric tests in case of non-normality or heterogeneity of variance. For the current set of variables (158, i.e. 79 for males and 79 for females) 13 variables (8%) showed non-normality of ANOVA residuals in a Shapiro-Wilks test at the 5% level, while at the 1% level only 3 variables were significant in a Shapiro-Wilks test (Appendix 15). In 48 cases the Shapiro-Wilks test was significant in one or more single dose groups, and 31 variables (15%) showed variance heterogeneity in a Bartlett's or Levene's test. Among all 632 comparisons the non-parametric Wilcoxon's test resulted in 26 significant differences (4%); note that this tests also requires homogeneity of variance. Despite these findings, i.e. non-normality or variance heterogeneity, the normal probability plots (Appendix 6) and the plots of residuals versus fitted values (Appendix 7) were generally satisfactory. This suggests that parametric t-tests and Dunnett tests can safely be applied.

In this report 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 4.1 (Figure 5 - Figure 8) 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 4.6 (Figure 25 - Figure 28) 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 4.2). 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: $E(\Delta_{Urea}) = 0$ vs. H_1: $E(\Delta_{Urea}) > 0$

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

Equivalence tests:

H_0: $E(\Delta_{Urea}) = EL_{Urea}$ vs H_1: $E(\Delta_{Urea}) < EL_{Urea}$

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

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

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).

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, 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

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 90-day subchronic toxicity study 1 in rats fed GM maize NK 603. GMP Two Year Safety Testing, Study 632165 B/2016/GLP. https://www.g-twyst.eu/files/Final%20study%20plans/90-day_NK603_feed_trial_1_study_plan_final.pdf.

List of Appendices

The following appendices are given in a separate document

List of Tables

List of Figures