EFSA review of statistical analyses conducted for the assessment of the MON 863 90-day rat feeding study

June 2007

Table of Contents

Executive Summary

General

In 2004 the EFSA GMO Panel gave its opinion (EFSA, 2004a,b) on the safety of MON 863 maize for import and processing, and released a statement on the safety of MON 863 shortly after (EFSA, 2004c). The EFSA GMO Panel based its opinions and statement on a wide range of evidence, which included data from a 90-day rat feeding study. This study was performed by Covance Laboratories, in compliance with internationally agreed GLP principles (OECD, 1998), for the applicant Monsanto. Subsequently Hammond *et al*. (2006) published a scientific paper based on the Monsanto report (2002), which provided fewer details than the original Monsanto report.

In a re-analysis of the MON 863 90-day rat study Séralini *et al*. (2007) claimed to have revealed significant variations in growth for both genders, as well as signs of hepatorenal toxicity in rats, and alleged that it cannot be concluded that MON 863 maize is a safe product.

The European Commission (DG SANCO, 2007) asked EFSA what impact the analysis performed by Séralini *et al.* (2007) study might have on its earlier opinion on MON 863 maize.

This EFSA report presents an assessment of the statistical methodology as applied by Monsanto (2002) and Séralini *et al.* (2007). It takes account of contributions from Member States, in particular the reports provided by the Commission du Génie Biomoléculaire of France (CGB) and information provided by the Agence Française de la Sécurité Sanitaire des Aliments of France (AFSSA).

The experimental design in the Monsanto study, adapted from the OECD Guideline 408 to make it more fit for the purpose of assessing GMOs, involved three factors, each at two levels: genotype (MON 863 and non-transgenic control LH82 x A634 maize); gender (Male and Female); and dose (11% and 33% level of MON 863 in the diet for the GM-fed rats). In addition to the near-isogenic control diets, rats were fed diets from six non-GMO reference lines, i.e. commercial hybrid maize genotypes.

Previous studies of these data have involved two separate statistical analyses: (1) body weight, body weight gain and food consumption; and (2) hematology, clinical-chemistry and urinalysis parameters, histopathology and organ weights.

Body weight development

The data in the Monsanto (2002) report were in accordance with the general requirements for this type of study and sufficient to allow the GMO Panel to draw adequate conclusions for the safety assessment. The analysis was clear and simple. However, the week-by-week analysis had relatively low statistical power, because it ignored the temporal structure of the data and was therefore potentially wasteful of information. Also, it engendered subsequent problems of multiple testing with associated problems of an increased possibility of false positives.

Séralini *et al*. (2007) did not account for temporal autocorrelation between measurements made at successive weekly intervals on the same individual rats. In addition, Séralini *et al.* (2007) analysed mean rat weights per group and hence did not account for any variability between individual rats within groups. Furthermore, no mention was made of heterogeneity of variances between rats over time. In such analyses the likelihood of spurious significant results that represent false positives is increased, and statements concerning probabilities must be viewed with caution.

The statistical test chosen by Séralini *et al*. (2007) yielded large and statistically significant differences between the Gompertz growth curves of GMO-fed and control-fed rats, both for males fed the 11% diet and for females fed the 33% diet. This was a portmanteau test of all three Gompertz growth curve

parameters simultaneously and there was no exploration of the source of any differences in the growth curves.

Two recent and independent statistical analyses both allowed for temporal autocorrelation and for heterogeneity of variances and examined the robustness of the Séralini *et al*. (2007) analyses with reference to the above weaknesses. One study was carried out by Monod (2007) at the request of the CGB. The other was conducted by EFSA and is reported here. Both studies employed modern sophisticated statistical methodology, capable of detecting even small differences.

From the Monod (2007) analysis the following points emerged: (1) there were no significant differences between GMO-fed and control rats for three of the four genotype-dose combinations; (2) for the fourth genotype-dose combination, namely female rats at a dose of 33%, the confidence limits showed a weak but marginally statistically significant difference in the shape of the growth curve. This was caused by a difference in the rate at which weight was gained, exhibited in the parameter *b* of the fitted Gompertz curve. It does not imply or demonstrate any difference in the final weight or total weight gain. In fact, the confidence interval for the estimate of the difference of the final weight, exhibited in the parameter a, showed no statistically significant difference.

In the EFSA analysis the following points emerged: (1) there was no significant difference in body weight over the 14-week period between GMO-fed and control rats when data were averaged over the genders and doses; (2) there was no significant effect of dose (11% versus 33%) or any interaction between dose and any other factor; (3) in all weeks except for males at 12 weeks, differences in body weights between GMO and control were consistent with random variation; (4) body weights of the animals fed GM maize and control maize were within the variations found in the additional study groups fed a range of non-GM maize commercial varieties.

In conclusion, the assumptions underlying the statistical tests performed by Séralini *et al*. (2007) did not hold, and so their tests tended to detect more significant results than analyses based on more robust techniques. The weekly growth data showed transient differences within the study period, but no overall differences in weight could be demonstrated. Séralini *et al*. (2007) reported that food consumption between GMO-fed and control rats were not noticeably different, but as shown by Monsanto (2002) and further illustrated by AFSSA (2007), transient changes in food intake were the most likely source of the transient differences in body weights.

Hematology, clinical-chemistry and urinalysis parameters, histopathology and organ weights

Séralini *et al*. (2007) compared their results with those of Hammond *et al*. (2006) which reported a less comprehensive range of data and analyses. Séralini *et al*. (2007) did not compare their results to those of the study reported in the Monsanto application (2002) that the GMO Panel assessed. Séralini *et al*. (2007) imposed a threshold to exclude potentially incidental differences, which accords with the distinction stressed by the GMO Panel (EFSA 2004 a,b) between statistical and biological significance. 479 endpoints were comparable between the studies of Séralini *et al*. (2007) and Monsanto (2002). Séralini *et al*. (2007) and Monsanto (2002) both reported significant differences between GMO-fed and control-fed rats in the same 25 endpoints and these represent the statistically most robust results.

In addition, Monsanto (2002) reported a further 10 significant differences not reported by Séralini *et al*. (2007); whilst Séralini *et al*. (2007) reported a further 13 significant differences not reported by Monsanto (2002). Differences between the two analyses were caused by the use of different variance estimates. Neither method is considered greatly superior to the other.

Séralini *et al*. (2007) found 40 significant differences out of 494 tested endpoints and claimed that only 25 would be expected by chance alone. However, this statement is correct only if: (1) the endpoints are independent and uncorrelated and (2) there are absolutely no systematic differences

between GMO-fed and control-fed rats.

In both the Séralini and Monsanto analyses tests were performed assuming that no correlation existed between endpoints. However, the EFSA analysis reposted here shows that sufficient correlation was present to demonstrate that the probability of 40 significant differences by chance alone was not negligible. Furthermore, this analysis shows that, given the fact that variable means from different genotypes may sometimes actually be slightly different under a substantial equivalence argument, the probability that 40 positive results in a set of 494 tests were obtained by chance alone is substantial.

EFSA investigated the variability in the data in detail. This took into account variation observed from both within and between maize variety studies and animal test populations. Toxicologists use the variability from such feeding studies to aid the interpretation of statistical significance tests. This approach provides a biological context for any statistical differences found between genotypes. For the majority of endpoints the variability between GMO-fed rats and those fed the near-isogenic control was considerably smaller than that shown by the six reference maize lines.
Furthermore, this EFSA study emphasises that statistically significant effects must be evaluated with respect to their biological significance (EFSA 2004 a,b). Finally those statistically significant differences that were found did not show consistency of patterning over dose and gender.

In conclusion, the results reported by Séralini *et al*. on the biochemical parameters, clinical pathology and organ weights were largely consistent with the findings previously assessed by the GMO Panel and reported in the Monsanto application.

Introduction and Objectives

In 2004 the EFSA GMO Panel published its opinion on the safety of MON 863 maize for import, processing, food and feed uses (EFSA, 2004a,b), and released a statement on the same topic shortly thereafter (EFSA, 2004c). For the opinions and statement the EFSA GMO Panel took into consideration data from a 90-day rat feeding study performed by Covance Laboratories for Monsanto (2002). Subsequently, Hammond *et al*. (2006) published a scientific paper based on the same Monsanto 90-day rat feeding study, but provided less detailed statistical analysis results than as described in the original Monsanto (2002) report.

More recently, Séralini *et al*. (2007) published a statistical re-analysis of the data from the Monsanto 90-day feeding study. Their main conclusions were the presence of significant differences in the growth curves for both male and female rats in one of the two different levels of MON863 in the diet in both male and female rats and signs of hepato and renal toxicity in animals fed diet containing MON863 maize.

Shortly after the publication by Séralini *et al.* (2007), the European Commission asked EFSA what consequences this paper and its conclusions may have on the EFSA GMO Panel opinions (EFSA 2004a,b). EFSA followed up with two initiatives:

1. Member States were asked to provide any analyses and comments that may contribute to the consideration of the statistical analysis of MON 863 data. Reports and comments provided by Member States are available in Appendix 1 of this document.
2. EFSA created an *ad hoc* task force with internal and external statisticians to assess the statistical methodology applied by Séralini *et al.* (2007) and its possible impact on the conclusions of the EFSA GMO Panel opinions (EFSA 2004a,b). During this process, also the statistical methods and results reported in Monsanto (2002) were considered as was the work conducted on these data by organisations in Member States.

This document, together with its appendices, addresses the statistical aspects of the question posed by the European Commission and it represents the outcome of the work carried out by the EFSA *ad hoc* task force.

The analysis of the 90-day rat feeding study data published by Séralini *et al*. (2007) is divided into two main parts:

1. Evaluation of weekly body weight data; and
2. Evaluation of biochemical parameters and organ weights, referred to as 'other variables' in this report.

This document follows the same structure, and is divided into two main parts accordingly.

Study design

The Monsanto (2002) study is publicly available[1]. The study design was adapted from the OECD Guideline No. 408 for a repeated dose 90-day oral toxicity study in rodents. It was conducted in compliance with Good Laboratory Practice (GLP) principles.

Briefly, the experimental design involved three factors, each at two levels:

- **Gender** (male or female);

[1] http://www.monsanto.com/monsanto/content/products/technicalandsafety/fullratstudy.pdf

- **Genotype** (MON 863 or non-transgenic control LH82 X A634). The MON 863 Genotype is referred to in this report as **GMO** and the non-transgenic control LH82 X A634 as **Control;** and
- **Dose** (11% or 33% level of maize in diet). The 11% diets were supplemented with 22% non-transgenic maize.

In addition to the near-isogenic control, other groups of rats were fed diets from six non-genetically modified commercial hybrid maize genotypes. These are referred to in this report as **reference genotypes**.

A group of rats with the same Gender, Genotype, and Dose are referred to as a **treatment group** in this report. For each Gender, there were 10 treatment groups each consisting of 20 Crl:CD®(SD)IGS BR rats. Only ten rats per group were measured for the haematological and chemistry parameters.

Test and control diets were formulated to compositionally match the standard rodent diet PMI #5002 that is comprised of 33% commercial maize. Analyses showed that the test diets were compositionally equivalent to the near-isogenic control and reference diets (Monsanto 2002).

The feeding study covered a period of fourteen weeks. At week 5 and at study termination blood samples were taken from 10 animals per treatment group for clinical pathology investigations.

Evaluation of body weight

Monsanto (2002) assessed the statistical significance of differences of body weight for each of the fourteen weeks separately.

To identify significant differences in weight gain patterns over time Séralini *et al*. (2007) modelled the entire fourteen week observation period by fitting one three-parameter Gompertz curve to the weight data of each treatment group i.e. each combination of Genotype, Gender, and Dose. This model was fitted to fourteen data points representing the *mean* body weight for each week (Séralini *et al*., personal communication, Appendix 2).

The fitting of a Gompertz curve to the weight data thus takes into consideration the measurements conducted each week into a single analysis rather than analysing each week separately. Whereas the fitting of a Gompertz curve to the weight data may therefore represent an interesting approach, it was deemed necessary to examine whether the assumptions required to conduct a valid statistical analysis were met. Key assumptions to be able to carry out valid statistical tests are that the residuals are independent and characterized by the same distribution.

Specifically, it was investigated whether the approach chosen by Séralini *et al*. (2007) for the analysis of the growth curves took into account the autocorrelation present in the dataset. Autocorrelation is the correlation between data when the same variable is repeatedly measured on the same subject (e.g. same subject measured over time). When such data correlation is not addressed then the model residuals may also be correlated and the assumption of their independence is thus violated. Ignoring this increases the likelihood of finding significant differences between treatment groups when, in reality, they might not be present.

It was also verified whether the increase in the variance of the weight data, associated to the increasing weights over time, was properly considered. In addition, use of the weekly mean weight data as done by Séralini *et al*. (2007) removed all variability between rats from the dataset and, as a result, makes valid statistical testing of differences between treatment groups questionable.

Because of these concerns, the weight gain data have been further investigated using alternative approaches.

Evaluation of other variables

For the other variables, the Monsanto (2002) report includes a one-way analysis of variance (ANOVA), one for each Week and Gender, in which all study groups (GMO at 2 Dose levels, Control at 2 Dose levels, and 6 references) were included. Contrasts included a test for differences between the GMO and the Control for each Dose level, and a comparison of the GMO at the 33% Dose level to the average of the 6 reference lines.

Séralini *et al.* (2007) compared, for each Dose level, the GMO group with its respective Control group using univariate tests. The authors point to a large number of significant results not previously reported by Monsanto, but referred for this to the Hammond *et al*. (2006) paper rather than the Monsanto (2002) study. However, as discussed above, the Hammond *et al*. (2006) provided only a subset of the results provided in the Monsanto (2002) report. Hence, it was determined which significant results in Séralini *et al.* (2007) were previously not already reported in Monsanto (2002) and *vice versa*. In addition, any discrepancy in significant results between Monsanto (2002) and Séralini *et al.* (2007) was further examined in order to investigate if these could be attributed to errors in the statistical analysis.

Séralini *et al*. (2007) also pointed out that of the 494 comparisons carried out on these 'other variables' 40 differences (8%) were statistically significant at the 5% level while only 25 would be expected to be significant under the global null hypothesis of no differences between GMO and Control groups. This statement required further examination. Firstly, it must be established how likely it is to obtain a number of significances as extreme as 40 from a sample of 494 tests at the 5% level. Secondly possible correlation within the dataset must be taken into account when examining the likelihood of finding more significant results than expected. For example, if the globulin concentration is statistically significantly increased whereas the albumin concentration is unchanged, then it would not be surprising that the albumin/globulin ratio is statistically significantly decreased as a result. Hence, it must also be assessed how likely the finding of 40 would be from the set of tests that were conducted, given the correlation structure of the data. Thirdly, as also highlighted by Séralini *et al*. (2007), it is reasonable to define acceptable margins of measured differences between the tested GMO and its comparator. Hence, it is relevant to assess how likely the finding of 40 would be from a set of tests given that the null hypothesis of no effect was replaced by an assumption that a small difference between the GMO and Control groups can be expected and is therefore acceptable. Consequently, the EFSA evaluation examines the likelihood of obtaining 40 significant results by answering these three questions above.

Data and Methods

Data

The data of the 90-day rat feeding study (Monsanto, 2002) were provided by Monsanto to EFSA in electronic format.

Approach

To answer the question from the European Commission, EFSA considered all available information. This included, in addition to the Séralini *et al*. (2007) paper, the Monsanto (2002) report, as well as reports on additional statistical analyses carried out by Member States' organizations.

EFSA contacted the authors of the Séralini *et al*. (2007) paper for further clarifications on the methods used for their analyses.

The description of the approaches and methodologies adopted in this report, which follows below, does not provide an exhaustive description of the statistical details. It is meant to summarize the main issues in a compact format. Supporting statistical information is available in larger detail in the appendices supporting this document.

Part1: Evaluation of body weight and food consumption data

Summary of Monsanto and Séralini approaches

As indicated, Monsanto (2002) assessed the statistical significance of differences of rats' weight for each of the fourteen weeks separately. In addition, Monsanto (2002) also assessed weight gain since study onset, and food intake for each of the fourteen weeks.

As indicated, to identify significant differences in weight gain patterns over time Séralini *et al*. (2007) modelled the entire fourteen week observation period by fitting a statistical three-parameter Gompertz curve to each treatment group. This model was fitted to fourteen data points representing the mean for each week (Séralini *et al*., personal communication, Appendix 2).

Séralini *et al*. (2007) also report to have modelled the food consumption data using a multivariate analysis.

Analyses performed in Member States: non-linear mixed model

The Commission du Génie Biomoléculaire (CGB) in France commissioned a statistical analysis of the weight data from the 90-day safety study. For this purpose Monod (2007) provided an extensive statistical analysis of the MON 863 data (Appendix 1).

In his report Monod (2007) emphasized that to be able to carry out valid statistical tests the underlying statistical assumptions must be met. Monod (2007) examined whether these assumptions were met when fitting a single Gompertz curve to the data from all the rats in a treatment group. He showed:

- That there is a large individual variability in the data between rats and this needs to be considered in the model;
- That the assumption of independence of residuals was violated. This violation of residual assumptions was present because the measurements were correlated, and such correlation invalidated the assumptions of independence; and
- That there was evidence of heteroskedasticity in the rat weight data, in the sense that, as expected, the variability of weight between rats increased with time i.e. with body weight itself.

Subsequently, Monod (2007) used the same Gompertz curve as Séralini *et al*. (2007), but taking into consideration the variability between rats. This was done by fitting a non-linear mixed effect model to the data, with rat as a random factor such that a separate Gompertz curve was fit to the dataset from each rat. The comparison between groups could then also be performed by comparing parameter estimates from different groups of rats. With this approach he was able to address both the heteroskedasticity and the autocorrelation of these repeated measures data.

Besides the Monod (2007) study, also the results on the weight and consumption data presented in the AFSSA (2007) report were considered.

Analyses performed by EFSA: one-way ANOVA and linear mixed model approach

A preliminary investigation of the weight data done with a factorial ANOVA performed week by week where Gender (male or female), Dose (11% or 33% maize in diet), Genotype (GMO or Control) were the considered factors. Furthermore, a non parametric analysis, free from the assumption of

normality and more robust against heteroskedasticity, was carried out according to the proposal of Scheirer el al (1976).

In addition to confirming the presence of autocorrelation in these repeated measurements data, the factorial ANOVA revealed some departures from normality and a few cases of heteroskedasticity.

It was decided to use a linear mixed model for longitudinal data in order to:

- Take into account all these factors and thus ensure a reliable assessment of the data;
- Consider all the weight data; and
- Not impose a prefixed growth curve, but consider a more generic form of autocorrelation allowing accommodating subjects' variation in growth behaviour.

In this mixed model analysis, Rat was considered as a random factor whereas Gender (male or female), Dose (11% or 33% maize in diet), Genotype (GMO or Control) and Week (time in week from the start of experiment, 14 levels) were considered fixed factors. To estimate the model parameters the restricted maximum likelihood (REML) procedure was used.

Six different covariance models were considered in this report. A brief description of these models is provided in Appendix 3. For selecting the best covariance structure with respect to the dataset under consideration four model selection criteria were used (see Appendix 3).

These analyses were done using SPSS version 15.01.

Part 2: Evaluation of other variables

Description of the data

The variability present in the data was visualised through standard box-plots for every variable in each treatment group. This was done with the Splus 7.0 software package.

Furthermore, since the largest number of significant cases was found for chemical endpoints, a compact visualization was proposed to summarize the overall information corresponding to the 33 variables into a few combinations of them. This was made using a principal component analysis (using Matlab 7.0.2), and allowed apprehension of any obvious cluster or difference between groups, as well as the background variability.

Finally, correlation coefficients were calculated between all the endpoints measured in the same week. For each endpoint, the correlation coefficient between measurements in weeks 5 and 14, were also calculated.

Further details on the approach can be found in Appendix 4.

Comparison of Monsanto and Séralini et al. results

As mentioned above, the 2004 EFSA opinions were based on evidence from the Monsanto (2002) report which was issued prior to the summary analysis of Hammond *et al*. (2006) which contained only a subset of the results.

To assess the agreement and discrepancies between the reported significant variables in the Monsanto (2002) report and the Séralini *et al*. (2007) paper, the following steps were taken:

- Clarifications were sought on the steps taken by Séralini *et al*. (2007) for their analysis. The authors clarified that the choice of univariate test was based on the outcome of the Shapiro test for normality on the GMO group, the Shapiro test on the Control group, and the F tests for equal variances between the GMO and the Control group. When one of the Shapiro tests was

significant ($p \leq 0.05$) the Mann-Whitney test was used. If not, the Student t test was used except for the cases where the F test was significant ($p \leq 0.05$) in which case the Student-Welch test was used. In addition, on the request of EFSA Séralini *et al*. provided a table indicating for each variable the p value and the statistical test that had been used (Appendix 2).

- Séralini *et al.* (2007) reported that they found 40 significant differences, but only identified 33 of them, as they considered 7 of those not relevant because the differences between treatment groups were considered too small (<5 %).
- EFSA verified whether any analyses had been reported only by Monsanto (2002) or only by Séralini *et al*. (2007). For those variables the statistical analyses were redone. For this purpose, the ests reported for the variable by Séralini *et al*. (Appendix 2) were redone (Splus 7.0). For the one-way ANOVA reported by Monsanto (2002) PROC GLM in SAS version 9.1 was used.
- All the statistically significant results from Monsanto (2002) and Séralini *et al.* (2007) were then tabulated along with their p-values (Table 1,2 and 3).
- As shown in Table 4 the tested endpoints were then cross-tabulated into:
  1. those which in both reports were reported as statistically not significant,
  2. those which in both reports were reported as statistically significant,
  3. those which indicated significance in Séralini *et al*. (2007) but not in Monsanto (2002), and
  4. those which did not indicate significance in Séralini *et al*. (2007) but did so in Monsanto (2002).

Simulation studies

To assess whether the 40 or so significant observations could be due to random variation alone it was considered worthwhile to estimate by simulation how likely it was to observe 40 significant results in the following three situations:

- Under the assumption that the endpoints are independent and that GMO and Control means are exactly the same (as assumed by Séralini *et al.*);
- Under the same null hypothesis but given the correlation structure as estimated from the data (derived from 119 x 4 = 476 endpoints at weeks 5 and 14); and
- Under the assumption that GMO and Control means might in fact be slightly different, given a distribution of acceptable differences. In these simulations the degree of acceptable difference was characterised by the ratio of between-group to within-group standard deviation. Simulations were performed firstly using various pre-defined degrees of acceptable difference, and secondly using a distribution of acceptable differences as estimated by a random effect model from the data of the 6 reference groups in the MON863 study (Genstat release 9.2).

The simulation studies and the models used are described in detail in Appendix 5.

Results

Part 1: Evaluation of body weight and food consumption data

Summary of Monsanto and Séralini results

Séralini *et al*. (2007) found no statistically significant differences between treatment groups for food consumption. Séralini *et al*. (2007) reported differences in growth curves that were statistically significantly different for two of the four GMO groups when compared to their respective Controls. Specifically, they reported differences indicative of:

- Lower weight curves in the Male treatment group fed the 11% Dose GMO diet, when compared to the 11% Dose Control Males;

- Higher weight curves in the Female treatment group fed the 33% Dose GMO diet, when compared to 33% Dose Control Females.

The Monsanto (2002) analysis showed that:

- For the 11% Dose Male GMO treatment group, neither the final body weight nor the overall body weight gain was statistically significantly different from those of their corresponding Control group. Body weights were statistically significantly lower for the 11% Dose Male GMO group in weeks 3 and 6 and their weight gain was also significantly lower than in their Control group during weeks 3 to 6. The Monsanto (2002) report also shows significantly lower food consumption for this 11% Dose Male GMO treatment group in weeks 3 and 4 as well as in week 10, when compared to its corresponding Control group.
- For the 33% Dose Female treatment group, neither the final body weight nor the overall body weight gain was statistically significantly different from their corresponding Control group. It is noted that in the pre-treatment period and in week 8 of the study food intake was significantly higher in this treatment group than in its Control group.

Analyses performed in Member States: non-linear mixed model

A first analysis with the Gompertz model fitted to all the data showed a number of outliers. These values, as well the subsequent measurements, were eliminated from the analysis dataset.

The main findings reported by Monod (2007)(Appendix 1) are as follows:

- Neither with males nor with females there was a significant difference in the value of the 3 model parameters when considering the four GMO and Control treatment groups in either Males or Females;
- There was no significant difference in the growth curve parameters when comparing the treatment groups the Males fed 11% GMO diet to their Controls;
- The growth curve parameters were significantly different between the Female treatment group fed the 33% GMO diet and its Control. The level of significance was only ($p=0.045$) though. The difference was attributable to the slope parameter b for the fitted Gompertz curve and not to the parameter a indicating the final weight.

Those findings were also valid when the outliers were included in the analysis.

Analyses performed by EFSA: one-way ANOVA

A detailed description of the results can be found in Appendix 3. The main findings are reported below.

Three rats, were excluded from the analysis because either they were identified as outlier (B38656) or, were only present in the trial until week 5 (B38923 and B38967) (Appendix 3). Another four rats (B38612, B38642, B38690, and B38789) were characterized by unusual growth patterns (Appendix 3). In order to assess the potential influence of data from latter four animals on the statistical results, all the analyses reported in this study were carried out twice: (i) including these four rats (referred to as analyses 'with 4 potential outliers'), and (ii) excluding them (referred to as analyses 'without 4 potential outliers').

The one-way ANOVA, in which weight data were analyzed separately week by week, showed, as expected, a significant main effect of Gender. There also was a significant interaction ($0.01 \leq p \leq 0.05$) of Gender and Genotype (GMO vs. Control) in (up to) 4 of the 14 study weeks. Females fed the GMO diet showed a slightly greater body weight compared to their Controls at those weeks. The opposite was the case for Males. However, for both Genders differences were small *i.e.* below 3% and 4%, respectively. Given that the data were affected by some heteroskedasticity problems, results of the ANOVA may show an increased type 1 error rate. With the non-parametric approach, none of the Gender by Genotype interactions were significant.

Results on linear mixed model approach

A detailed description of the results can be found in Appendix 3. The main findings are reported below.

Based on the different information criteria used, the better covariance structure was the heterogeneous Toeplitz. There was a significant interaction between Gender and Week showing that weights differed between the two genders over the course of the fourteen weeks. A significant interaction between Week and Genotype was also present, reflecting a difference in rats' response to diet at different weeks. Finally, there was a significant three-way interaction between Gender, Week, and Genotype, indicating lack of consistency between Males and Females to Genotype at different Weeks.

The mixed model analyses conducted on the data from reference groups and the Control groups (that is, without the GMO groups) also showed the two same significant 2-way interactions of Genotype by Week and Gender by Week to be present.

Part 2: Evaluation of other variables

Description of the data

Descriptive statistics are reported in Appendix 4. Box plots of all significant results by Séralini *et al.* (2007) and/or Monsanto (2002) visualize the background variation in the GMO groups, the Control groups, and the reference lines (RefA33 to RefF33). Variability between GMO groups and their corresponding Control groups appeared to be considerably lower than the overall variability depicted by the six reference lines.

The first four principal components on the chemical data (Appendix 4) allowed to visually discriminate between the two Genders and between the different Weeks, but no visual difference between Genotype groups was present.

The correlation coefficients calculated varied across endpoints and about half of them were significantly different from zero. The correlations between organ endpoints (median=0.44) and the time-correlations between the data measured at week 5 and week 14 for a same variable (median=0.35) were generally higher in absolute values than correlations among other endpoints (medians at week 14=0.12 and 0.14 respectively for chemical and haematological endpoints).

Comparison of results reported by Monsanto and Séralini *et al.*

Whereas the Séralini *et al.* (2007) paper stated that there were 40 significant results, the authors only provided to EFSA details on 39 statistically significant results. In addition, Séralini *et al.* (2007) conducted comparisons on both the percentage and the absolute number of neutrophils, lymphocytes, monocytes, and eosinophils, whereas the Monsanto (2002) report provided the results for the absolute number but not for the closely related % data for these four variables. Hence, Séralini *et al.* (2007) conducted 16 tests not reported by Monsanto (2002). One of these was statistically significant (monocytes females 11% week 5). This left 479 tests that were reported to EFSA by both Monsanto (2002) and Séralini *et al*. (2007). For these 479 tests, Monsanto (2002) reported 35 significant results while Séralini *et al*. (2007) reported 38 significant results.

Tables 1 to 3 give an overview of the significant p values (bold) reported in the Monsanto (2002) report and the ones provided to EFSA by the authors of Séralini *et al*. (2007) (Appendix 2). A comparison of the results from Monsanto (2002) and Séralini *et al*. (2007) shows (Table 4) that the latter paper:

- Confirms 25 significant results from Monsanto (2007);
- Identifies 13 new ones; and

- Does not report as significant 10 results that were previously identified as significant by Monsanto (2000).

After receiving the full set of the statistical results from the authors of Séralini *et al*. (2007), EFSA reproduced the analysis of Séralini *et al.* (2007) and found only discrepancies in the male 33% groups. Consultation with the authors clarified that this was due to the exclusion of the data of one rat which died during the study whereas, in the original Monsanto analysis, the rat was replaced with another rat of the same group. After taking this into account, EFSA could reproduce the complete analysis perfectly. Next, EFSA was also able to reproduce the Monsanto statistical analyses which were significant in Séralini *et al*. (2007) but not in Monsanto (2002), except for Urea Nitrogen for females at dose 11% in week 14 where the p-value turned out to be 0.01.

For the 13 new significant results found by Séralini *et al.* (2007) (Table 4) Monsanto (2002) reported three with p-values below 0.1, eight with p-values between 0.1 and 0.2 while the remaining two had p-values of 0.428 and 0.334. Similarly, for variables found significant in Monsanto (2002) and not reported as significant by Séralini *et al*. (2007) the same pattern of p-values in the range between 0.05 and 0.2 was reported (Appendix 4) by Séralini *et al*. (2007).

Regarding haematological values, Séralini *et al*. (2007) showed a significant increase in neutrophil count in males in week 14 at the 33% dose. Monsanto (2002) did not report this as significant, but showed that the white blood cell count was significantly higher in that group at that time point.

With respect to blood chemistry, Séralini *et al*. (2007) reported the following significant results not previously reported by Monsanto (2002):

- An increase in globulin concentration in the 11% female group at week 5 and higher blood urea nitrogen concentrations in week 14; and
- An increase in the albumin/globulin ratio and the total protein in the 11% male group at weeks 5 and 14, respectively.

For urine chemistry the findings by Séralini *et al*. (2007) not reported by Monsanto (2002) are:

- Higher urine potassium concentrations in the 11% male group in week 5;
- Lower urine sodium concentration and excretion in the 33% male group in week 14, and lower urine phosphorus concentration in weeks 5 and 14.

For organ weights Séralini *et al*. (2007) newly identified a higher liver weight and liver/brain ratio in the 11% female group.

Simulation Studies

Details of the simulation studies are reported in Appendix 5. The simulation studies were performed assuming multivariate normality and homoskedasticity. Whereas these assumptions may sometimes fail for the actual data, the simulation study was intended to illustrate two general points, and the results have relevance for any specific structure of data.

Details on how the correlation was computed are reported in Appendix 4. When accounting for the correlation between endpoints and time-points (weeks 5 and 14), the probability to observe 38 or more (false positive) significant results was found to be 5.5%, i.e. about 20 times greater than without accounting for this correlation. Such results were derived under the assumption that the means of simulated endpoints were exactly the same for both groups, which is conservative and is not expected to be true for all variables in equivalent groups.

Simulations were also performed accounting for acceptable difference between groups. First, when using various predefined degrees of between group variability, it turned out that the probability to observe at least 40 significant cases ranged from 3.8% to 86% when the ratio of between-group to within-group standard deviation ranged from 0.1 to 0.2. Second, the distribution of differences estimated from the data of the 6 reference groups was estimated. Among these groups the between-

group standard deviation was estimated as negligible in 100 of the investigated chemical and haematological endpoints, but it had a positive value up to 50 % of the within-group standard deviation in the remaining 84 endpoints. Using this empirical distribution of 184 ratios to characterise acceptable variability between-groups, at least 40 significant cases were observed out of 494 comparisons in 54% of the simulated datasets.

In reality, both correlation and background variability are present. This leads to the conclusion that it should be expected to observe 40 significant cases.

Discussion

First, it is noted that Séralini *et al*. (2007) found the descriptive statistics to be consistent between their data and the results reported by Hammond *et al.* (2006).

Part 1: Evaluation of body weight and food consumption

Model choice

Séralini *et al*. (2007) stated that a multivariate analysis had been conducted on the food consumption and the body weight data. Séralini *et al.* (personal communication) confirmed that they performed a multivariate analysis on the food consumption data, which was however not described in the paper. Their statistical approach for the analysis of the body weight data consisting of the fitting of a non-linear growth curve (Gompertz curve) to a series of weekly treatment group does not represent what is known as a multivariate analysis. Nevertheless, the fitting of a Gompertz curve to the weight data represents an interesting approach. However, the fact that this curve was fitted on the mean weights of each group essentially removed all variability between rats from the data. This greatly inflated the probability to find statistically significant results because it ignored the variability of the rat weights within each group. In addition, this approach did not allow for testing of independency of residuals with acceptable power, as the number of residuals was very low (n=14).

Indeed, other statistical methods such as linear or non linear mixed-effects modelling proposed in this report are preferable because these approaches take into account the correlation structure of the data.

Mixed models were considered very suitable to analyze the weight data of the MON863 90-day rat feeding study for the following reasons:

- The between-rat variability was accounted for;
- The models allowed to handle for the presence of correlated data and non-constant variability as well as some missing observations. Mixed models offer also the possibility to use different covariance structures within subjects over time; and
- They are robust to deviations from the normality assumption, unless the datasets are strongly unbalanced (which was not the case for the weight data)

The possibility of testing different covariance structure models in the linear mixed model represented an advantage in this specific case because it allowed overcoming the problem of lack of fit of non linear fitting in the weeks following the blood collection that took place during the 5th week. Interestingly, although heterogeneous Toeplitz could be clearly identified as the most appropriate covariance structure model, there was consistency of results among all the six covariance structure models tested in this report, suggesting the obtained results are robust.

Main factors affecting rats body weight

The results of the one-way ANOVA and mixed model analyses, as performed by EFSA, showed large

significant effects of Week and Gender on body weight. The latter were expected, given the well known difference in growth rate between male and female rats.

The presence of a significant interaction between Gender and Week in the linear mixed model analysis demonstrates that growth rate patterns differed between the two genders over the course of the fourteen weeks. Again this is not surprising.

Potential effects of Genotype and Dose

Several questions can be raised with regard to the identification of potential differences in body weight that might be attributable to differences between the GMO groups and their corresponding Control groups. These include:

- Were there differences in the final weights?
- Were there differences in the average weight or the weight gain during the 14 week study period?
- Were there differences in the weight gain patterns during the study period?

The Monod (2007) and the EFSA results confirm the Monsanto (2002) report that there were no statistically significant differences between the GMO treatment groups and their respective Controls in the average body weight during the study nor in the final weight at the end of the study, respectively.

Monod (2007) showed that, when the Gompertz model was used with appropriate model assumptions, the parameters of the growth curve models were not significantly different between the GMO and the Control treatment groups, neither for males nor for females, with the exception of one treatment group (33% Females) which showed a slight ($p=0.045$) statistically significant difference. This is in contrast with Séralini *et al*. (2007) who reported that the growth curves for two of the four GMO treatment groups showed a strong statistical difference ($p<0.001$), when compared to their respective Controls.

The significant interaction of Gender and Genotype in (up to) five of the fourteen weeks in the one-way ANOVA as performed by EFSA and the significant two-way interaction between Genotype and Week and the three-way interaction between Genotype, Gender, and Week in the linear mixed model analysis reflect:

- Differences between treatment groups in different weeks; and
- Lack of consistency between Males and Females for those differences.

This could indicate an effect of GMO on the growth rate of rats. However, when the analysis was carried out excluding the GMO from the dataset (i.e. analysis only on the commercial lines and the Controls) the two-way interactions of Week with Gender and Week with Genotype were also present. Therefore, it can not be concluded that the latter interaction id due to the presence of MON863 in the diet.

The 3-way interaction might suggest a pattern of higher increase in body size of the Male GMO treatment groups when compared to their Control groups and a pattern of slightly higher body weight gain of Female GMO treatment groups, when compared to their Control groups. However, for both Male and Female genders the weight differences between GMO treatment groups and Controls very small i.e. below 3% and 4%, respectively.

The results of the linear and the non-linear mixed model analyses therefore confirm the results of Monsanto (2002) in that there are differences for the 33% Female and 11% Male treatment groups in some weeks.

The weekly data on food consumption indicate that some of these transient differences in weight and weight gain are consistent with the observed patterns in the significant differences in food intake between these GMO treatment groups and their Control counterparts (Monsanto, 2002):

- There was a significantly lower food consumption in weeks 3-6 with 11% Males; and
- A significantly higher food intake in week 8 in the 33% Male.

The AFSSA report (2007) illustrates that the food consumption in the 11% Dose GMO Male group was almost always lower than in the equivalent Control group throughout the study. Thus it is not surprising that the weight gain in that group tended to be numerically lower. The AFSSA (2007) report also illustrates that from week 7 the food intake in the Male 33% GMO group was higher than in its Control group. This is consistent with a pattern of higher weight gain in that period.

Part 2: Other variables

For the other variables, Séralini *et al*. (2007) highlight a number of differences to be significantly different from zero. Many (n=25) of the significant results on variables tested in both studies were already previously reported in the Monsanto (2002) report examined by the EFSA GMO panel. Thus, in effect, the Séralini *et al*. (2007) study provides independent confirmation for many of the previous findings.

Nevertheless, among the 40 reported variables some were found to be significant in the Séralini *et al*. (2007) paper and not in the Monsanto (2002) report and vice versa. There are two main reasons for these differences:

- Differences in statistical significance appeared when a t-test was used in both studies. In the Monsanto (2002) study the comparison between the GMO and Controls was based on an ANOVA approach in which all ten study groups were included. In contrast, Séralini *et al.* (2007) compared the mean of the GMO11% versus Control 11% group and the GMO 33% versus Control 33% using simple univariate tests. Hence, where they occur these differences are due to the use of different variance estimates in both approaches. Determining which of these variance estimates is the more appropriate one to use for the t test is a matter of debate and depends on how much the variance of the reference lines is considered to provide information that is of relevance for the hypothesis being tested.
- In addition, instead of using at t-test, Séralini *et al*. (2007) performed a Student-Welch test when the equal variance assumption between the two groups was not met and a non parametric test (Mann-Whitney) in case of non-normality. While the use of these tests may be quite appropriate it is not the case that they led to the identification of more statistically significant results. In other words, sometimes results were not identified as significant by Séralini *et al*. (2007) using these other tests whereas the t-test used in the Monsanto (2002) indicated the result to be significant.

More relevant is the concordance of both approaches in the identification of 25 statistically significant variables. These represent, from a statistical perspective, the more robust results.

There is no evidence that the test methods used by Séralini *et al*. (2007) resulted in the identification of more significant differences than in the approach used by Monsanto (2002). The total number of significant results was in fact very similar for both (Table 4). Séralini *et al.* (2007) wrote that on average 25 'false positive' significant results were expected, while 40 were found. However in both the Séralini and Monsanto analysis the statistical tests were performed assuming that no differences exist between the treatment groups in the averages of the baseline values of their test parameters. Due to the correlation between the variables (which was clearly present in this study) and the fact that variable means from different treatment groups may sometimes actually be slightly different, the probability that 40 or more positive results in a set of 494 tests are obtained by chance alone is substantial.

As the aim of safety studies is to identify any potential adverse effect, they should be organised and focussed on finding differences with primary attention for restricting the false negative error rate (finding no difference where a biologically relevant difference does in fact exist). This may often come at the cost of finding statistically significant results which are not biologically relevant and thus need careful review and interpretation.

When a multiplicity of statistical tests is performed, one on each of a large number of variables, it is to

be expected that some will attain statistical significance by chance alone. There are several ways one can make allowance for this:

- Classically this is done by adjusting the level at which significance is assessed, e.g. Bonferroni adjustments (Bonferroni, 1935), or by using protected pairwise tests which are conducted only after a significant omnibus F-test. Recently proposed approaches to address massively multiple comparisons are based on controlling false discovery rates (FDR) as introduced by Benjamini and Hochberg (1995).
- Monsanto (2002), although they reported every significant contrast, also used a protected approach. This was done by only considering for further evaluation those variables for which the overall test results of the one-way ANOVA on 10 diets was significant. If this was the case, then they tested whether the 33% GMO diet was significantly different from the means of the reference lines. Finally, if these were significant, then the means of the 33% GMO diets were compared to the range of the commercial diets to assess whether they were biologically meaningful. They reported that there were 19 findings that satisfied all the above criteria.
- Séralini *et al*. (2007) used unprotected tests and reported the significant results of each of the 494 variables tested except for those with a difference of less than 5%. They then assessed whether differences could be due to differences in the diet by comparing the control groups to the mean of the reference groups and finally compared the remaining significant parameters to the mean of the six reference groups. They reported that seven results satisfied all the above criteria.
- Of the seven variables proposed as key findings by Séralini *et al*. (2007) (highlighted in Tables 1) three were also significant in the Monsanto (2002) analysis. The remaining four concern urine parameters in the 33% males (Table 1) (Séralini *et al*., 2007): urine phosphorus (weeks 5 and 14), and urine sodium (week 14), and urine sodium excretion (week 14).

Statistical analysis is a useful tool to point out possible safety problems. However, it does not in itself constitute a sufficient condition to demonstrate presence of an ill effect. For that purpose it is necessary to interpret these results and consider whether the differences are of a biologically relevant magnitude. The allowed discrimination between genders and between weeks, but no visual difference between treatment groups could be found.

This report contains the following information that provides a biological context for any differences found between treatment groups and can thus be used to aid the interpretation of statistical significance tests (Appendix 4):

- The results on variability of the data and the presence of outliers within each treatment group (Box plots);
- The correlation matrices between variables; and
- The principal components analysis which helped identified the main sources of variation in the dataset (Gender and Week) and thus helps to put into perspective any findings attributed to Genotype.

Conclusions

Overall, the various studies that were conducted on body weight show a consistent picture in that:

- Neither final weight nor average weight were different between the GMO and their Control groups;
- During the study the differences in the weight data between the GMO and their Control groups sometimes varied by Week and Gender, but differences were always below 4 %. These differences are probably attributable to fluctuations in food intake.

The results of the statistical analyses on haematology, serum and urine chemistry, and organ weight data reported by Séralini *et al*. (2007) are largely consistent with the findings previously reported by Monsanto (Monsanto, 2002). The differences resided in the interpretation of these results i.e. which

statistically significant results were considered biologically meaningful. The number of statistically significant test results found (in both studies) is not higher than might be expected by chance alone.

Acknowledgements

EFSA acknowledges the members of the task force created to assess the statistical analysis of MON 863 data: Marco Acutis (external expert; member of the GMO statistics working group), Billy Amzal (EFSA staff member), Hubert Deluyker (EFSA staff member), Claudia Paoletti (EFSA staff member), Joe Perry (member of the GMO panel), Hilko van der Voet (external expert; member of the GMO statistics working group), and Didier Verloo (EFSA staff member).

EFSA is grateful for information received from Member States. EFSA wishes to thank Hervé Monod for his assistance with interpretation of statistical results and Gilles-Eric Séralini, Dominique Cellier and Joël Spiroux de Vendomois for their willingness to discuss their findings and to provide further clarifications.

References

AFSSA, 2007. Agence Française de la Sécurité Sanitaire des Aliments. Dossier n 2007-SA-0109 http://www.afssa.fr/Ftp/Afssa/40728-40729.pdf

EFSA, 2004a. Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the Notification (Reference C/DE/02/9) for the placing on the market of insect-protected genetically modified maize MON 863 and MON 863 x MON 810, for import and processing, under Part C of Directive 2001/18/EC from Monsanto. The EFSA Journal, 49, 1-25. http://www.efsa.europa.eu/etc/medialib/efsa/science/gmo/gmo_opinions/381.Par.0001.File.dat/opinion_gmo_06_en1.pdf

EFSA, 2004b. Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the safety of foods and food ingredients derived from insect-protected genetically modified maize MON 863 and MON 863 x MON 810, for which a request for placing on the market was submitted under Article 4 of the Novel Food Regulation (EC) No 258/97 by Monsanto. The EFSA Journal, 50, 1-25. http://www.efsa.europa.eu/etc/medialib/efsa/science/gmo/gmo_opinions/383.Par.0001.File.dat/opinion_gmo_07_en1.pdf

EFSA, 2004c. Statement of the Scientific Panel on Genetically Modified Organisms on the evaluation of the 13-week rat feeding study on MON 863 maize, submitted by the German authorities to the European Commission. http://www.efsa.europa.eu/etc/medialib/efsa/science/gmo/statements/666.Par.0001.File.dat/sr_gmo01_statement_study_MON 863_en1.pdf

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. Roy. Statist. Soc. B 57, 289–300.

Bonferroni, C. E., 1935. Il calcolo delle assicurazioni su gruppi di teste. In: Studi in Onore del Professore Salvatore Ortu Carboni. Rome: Italy, 13-60.

Hammond, B.G., Dudek, R., Lemen, J.K., Nemeth, M.A., 2006. Results of a 90-day safety assurance study with rats fed grain from corn borer-protected corn. Food Chem. Toxicol., 44, 1092-1099.

Monod H., 2007. Expérience sur rats menée par Monsanto en 2001-2002 avec certains régimes comportant du maïs génétiquement modifié: analyse statistique des courbes d'évolution du poids. Study conducted upon request of the CGB (Commission du Génie Moléculaire, France). http://www.ogm.gouv.fr/experimentations/evaluation_scientifique/cgb/autres_avis/Avis_CGB_MON 863_15juin2007.pdf

Monsanto, 2002. 13-Week Dietary Subchronic Comparison Study with MON 863 Corn in Rats Preceded by a 1-Week Baseline Food Consumption Determination with PMI Certified Rodent Diet #5002 http://www.monsanto.com/monsanto/content/products/technicalandsafety/fullratstudy.pdf

Scheirer J.C., Ray W.S., Hare N., 1976. The Analysis of Ranked Data Derived from Completely Randomized Factorial Designs Biometrics, 32, 429-434.

Séralini, G.E., Cellier, D., de Vendomois, J.,S., 2007. New analysis of a rat feeding study with a genetically modified maize reveals signs of hepatorenal toxicity. Arch. Environ. Contam. Toxicol., 52, 596-602.

List of Tables

Table 1. Chemistry data

Endpoint	MON Male 11%	SER Male 11%	MON Male 33%	SER Male 33%	MON Female 11%	SER Female 11%	MON Female 33%	SER Female 33%
Week 5								
Albumin	0.251	0.4888	0.403	0.4911	0.388	0.363	**0.018**	**0.0102**
Albumin/Globulin Ratio	0.116	**0.0117**	0.61	0.4359	0.142	0.0736	0.527	0.1294
Globulin	**0.028**	**0.0178**	0.711	0.5148	0.109	**0.049**	0.865	0.876
Total Protein	**0.023**	0.0515	0.866	0.8569	0.504	0.3884	0.073	0.0602
Triglycerides	0.081	0.192	0.895	0.887	0.705	0.4154	**0.001**	**0.0023**
Urine Calcium	**0.027**	0.0633	0.972	0.7621	0.632	0.6045	0.133	0.1789
Urine Chloride Excretion	0.157	0.1928	0.906	0.8909	**0.032**	**0.0363**	0.051	**0.0378**
Urine Phosphorus	0.928	0.893	0.11	**0.0455**	0.525	0.8534	0.384	0.5212
Urine Potassium	0.129	**0.036**	0.456	0.3422	0.906	0.869	0.583	0.4078
Number Significant	*3*	*3*	*0*	*1*	*1*	*2*	*2*	*3*
Week 14								
Alanine Aminotransferase	**0.023**	**0.0275**	0.584	0.5021	0.577	0.5447	0.771	0.8197
Albumin	0.237	0.161	0.163	0.2202	**0.047**	**0.0437**	0.097	0.1641
Albumin/Globulin Ratio	0.224	0.3638	0.403	0.6444	**<0.001**	**0.0024**	0.22	0.1617
Calcium	0.383	0.3425	**0.044**	0.105	0.74	0.9696	0.986	0.8914
Chloride	0.676	0.6061	**0.026**	**0.0128**	0.44	0.4118	0.149	0.1274
Creatinine	0.174	0.0827	**0.019**	**0.0438**	**0.01**	**0.0168**	0.708	0.7545
Globulin	0.058	0.1572	0.07	0.1675	**0.004**	**0.0144**	0.667	0.6329
Glucose	0.378	0.2166	0.084	0.5387	**0.041**	**0.0441**	**0.021**	**0.0073**
Sodium	**<0.001**	**0.001**	0.624	0.5704	0.055	0.0819	0.148	0.0792
Specific Gravity	**0.018**	**0.025**	0.298	0.195	0.55	0.9318	0.779	0.7805
Total Protein	**0.018**	**0.0255**	**0.015**	**0.0387**	0.818	0.8367	0.251	0.3713
Triglycerides	0.346	0.4051	0.937	0.7802	**0.01**	**0.0438**	0.693	0.613
Urea Nitrogen	0.125	0.0586	0.638	0.6653	0.058	**0.0483**	0.874	0.8135
Urine Creatinine	0.141	0.0753	0.857	0.211	0.84	0.3445	0.702	0.5288
Urine Phosphorus	0.051	0.0968	0.21	**0.0435**	0.782	0.6842	0.805	0.7913
Urine Sodium	0.337	0.2458	0.428	**0.0498**	0.788	0.6579	0.331	0.2291
Urine Sodium Excretion	0.878	0.8914	0.129	**0.011**	0.215	0.2215	0.304	0.2583
Number Significant	*4*	*4*	*4*	*6*	*6*	*7*	*1*	*1*

Summary table of p values obtained by Monsanto (2002) and Séralini *et al*. (2007). The significant (≤ 0.05) p values are bold faced. The seven significant results identified by Séralini *et al*. 2007 as key significant differences between MON863 and their respective control groups are underlined.

Table 2. Hematology data

Trait	MON Male 11%	SER Male 11%	MON Male 33%	SER Male 33%	MON Female 11%	SER Female 11%	MON Female 33%	SER Female 33%
Week 5								
Hematocrit	0.411	0.1744	0.219	**0.0392**	0.397	0.762	0.062	0.1499
Hemoglobin	0.355	0.3574	0.092	0.0814	0.398	0.3226	**0.024**	0.1607
Monocytes*		0.8982		0.0561		**0.01429**		0.7263
Red Blood Cell Count	0.402	0.2457	0.313	0.1422	0.175	0.0756	**0.012**	0.0694
Reticulocyte Count	0.969	0.9722	0.640	0.4742	0.364	0.1292	**0.022**	0.1491
Reticulocyte Count Abs	0.811	0.8255	0.567	0.4081	0.290	0.0658	**0.045**	0.2205
Seg Neutrophils Abs	0.799	0.2855	0.334	**0.0385**	0.306	0.2337	0.852	0.8633
Number Significant	*0*	*0*	*0*	*2*	*0*	*1*	*4*	*0*
Week 14								
Basophils Absolute			**0.004**				**0.032**	
Eosinophils Absolute	0.065	0.2047	**0.032**	**0.0305**	0.329	0.4549	1.000	1
Lymphocytes Absolute	0.283	0.2833	**0.042**	0.0933	0.072	0.0585	0.160	0.2036
Prothrombin time 14	0.495	0.3977	0.886	0.8124	**0.015**	**0.009**	0.092	0.0952
Reticulocyte Count	0.550	0.5158	0.676	0.304	0.105	0.0837	**0.011**	**0.0232**
Reticulocyte Count Abs	0.543	0.5401	0.589	0.4673	0.104	0.0782	**0.016**	**0.0162**
White Blood Cell Count	0.173	0.1909	**0.033**	0.0693	0.090	0.0862	0.151	0.1945
Number Significant	*0*	*0*	*4*	*1*	*1*	*1*	*3*	*2*

Summary table of p values obtained by Monsanto (2002) and Séralini *et al*. (2007). The significant (≤ 0.05) p values are bold faced. *The variable Monocytes was not reported in the Monsanto (2002) report.

Table 3. Organ data

Trait	MON Male 11%	SER Male 11%	MON Male 33%	SER Male 33%	MON Female 11%	SER Female 11%	MON Female 33%	SER Female 33%
Heart Wt	0.697	0.638	0.899	0.8925	0.591	0.8283	**0.020**	0.0598
Kidney / Brain Ratio	0.179	0.2883	**0.017**	**0.0147**	0.659	0.6361	0.886	0.8767
Kidney Wt	0.324	0.2928	**0.034**	**0.0246**	0.365	0.3515	0.446	0.4399
Kidney, % Body Wt	0.743	0.6889	**0.046**	**0.034**	0.590	0.5588	0.636	0.6889
Liver / Brain Ratio	0.650	0.6872	0.549	0.5992	0.142	**0.036**	0.297	0.3437
Liver Wt	0.892	0.8815	0.572	0.6441	0.078	**0.0104**	0.126	0.1555
Number Significant	*0*	*0*	*3*	*3*	*0*	*2*	*1*	*0*

Summary table of p values obtained by Monsanto (2002) and Séralini *et al*. (2007). The significant (≤ 0.05) p values are bold faced.

Table 4. Cross tabulation of the Séralini and Monsanto results

Study		**Séralini *et al.***	
		Significant	Non-Significant
Monsanto	Significant	25	**10**
	Non-Significant	**13**	431

List of Appendices