EUROPEAN COMMISSION

Community Research

SEVENTH FRAMEWORK PROGRAMME

GRACE Stakeholder Consultation on animal feeding studies in GMO risk assessment: Chronic toxicity (1-year) study and subchronic toxicity and longitudinal metabolomics study

FP7 Collaborative Project
GRACE 311957

Responses of GRACE team members to questions and comments raised by stakeholders

July, 2014

GRACE

GMO Risk Assessment and Communication of Evidence

This report “GRACE Stakeholder Consultation on animal feeding studies in GMO risk assessment: Chronic toxicity (1-year) study and subchronic toxicity and longitudinal metabolomics study” dated July 2014 holds the responses of GRACE team members to questions and comments raised by stakeholders participating in the written consultation in December 2013. Please, note that this compilation is not yet complete. Questions and comments still lacking a response are provided in the second section of this report. A complete version of responses will replace this document.

“GMO Risk Assessment and Communication of Evidence”, GRACE is a Collaborative Project of the Seventh Framework Programme of the European Community for Research, Technological Development and Demonstration Activities.
Grant agreement no: 311957
Project duration: 1st June 2012 – 30th November 2015
GRACE website: www.grace-fp7.eu

Table of contents

Introduction

On 6 December 2013 the FP7 project GRACE started a written stakeholder consultation procedure on the draft study plans of two animal feeding studies conducted with GM maize MON810:

- Chronic toxicity (1-year) study
- Subchronic toxicity and longitudinal metabolomics study[1]

The design, conduct and interpretation of animal feeding studies with whole food/feed in the context of GMO risk assessment has been highly debated over recent years. The aim of this consultation in the context of the GRACE project was to consider the inputs from stakeholders and external experts on the draft study plans describing the design, conduct and analysis of the above mentioned animal feeding studies. Drafted study plans were put at the GRACE website (http://www.grace-fp7.eu/content/start-stakeholder-consultation-draft-study-plans-chronic-and-sub-chronic-toxicity-studies-gm) and around 700 stakeholder organisations Including competent authorities, industry, farming organisations, professional organisations, civil society organisations, and academia were invited to provide comments by e-mail. 243 comments and questions were subsequently received from eight stakeholder organisations (list provided in the Annex). Stakeholder comments were thoroughly evaluated, discussed and responses were provided by GRACE team members. The draft study plans were revised by taking into account the provided comments and the final versions of the study plans were subsequently made available via the GRACE website (http://www.grace-fp7.eu/content/reports-study-plans-consultation-documents). Based on the revised study plans, the chronic toxicity (1-year) study and the subchronic toxicity and longitudinal metabolomics study were started in January and February 2014 respectively. Once the results of these studies will be available another stakeholder consultation will be organised in 2015 in order to discuss results and draft interpretations. More information on GRACE and GRACE stakeholder consultations can be found at http://www.grace-fp7.eu/.

On the following pages stakeholder comments and questions as well as responses of the GRACE team are provided with an indication whether they were leading to an adaptation of the draft study plans. A few comments/questions (list provided in Part II, p.57) still remain unanswered. An update to this report will be provided as soon as all responses arrived.

For the sake of clarity, the comments received were categorised and re-arranged according to topic areas. References to page numbers in the comments/questions refer to the draft study plans, references to page numbers in the responses refer to the final versions available at the GRACE website.

[1] A specific study (based on the 90-day feeding study design) to consider whether (additional) metabolic endpoints improve the evaluation of toxic effects through transgenic crops. Since the collection of samples of blood and urine throughout the (longitudinal) study means additional disturbance for the rats a separate study was launched using the same plant material as for previous 90-day feeding studies. This study takes into account OECD TG408 and the EFSA standard protocol on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed.

Part I: Responses to questions and comments

Chronic toxicity (1 year) study

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	General		
1	It is not clear why the length of the study has been reduced to 1 year from the original 2 years. We suggested previously to reduce the length of the study period to 18 months if Wistar rats are used as study objects.	Following from the tasks defined in the Call GRACE is going to evaluate and compare the approach "90-day feeding study" with other potential tools or sources of information for the risk assessment of whole food/feed derived from GM crops. Outcomes from 90-day feeding studies will be compared with outcomes of the compositional analysis, ~omics of plant material and animal tissues, *in vitro* studies with animal tissues, longitudinal metabolomics and extended feeding studies, and literature data. The exact duration of the extended study was not defined beforehand. The decision to perform a 1 year study is based on the general rationale of subsequent feeding studies and the given resources of the project.	N
2	It is appreciated that the testing strategies have been extended to a chronic toxicity study. However, with the given study batteries and designs, no final evidence is possible with reference to long-term (especially appropriate for foodstuffs), including possible carcinogenic, to reproductive or developmental effects of the whole food and/or feed.		
3	Recommends to state and thoroughly detail the *rationale and the objectives* of the proposed study. This lack does not allow to understand if and how this study could provide a continuum in the assessment of the safety profile of MON810 within the GRACE project. A *lack of consistency* between the proposed 1-year study and the former 90-day are noted (e.g. possible different test site for histopathology). A clear description for the rationale is needed in order to justify the use of animals in the context of 3Rs.	The overarching aim of GRACE is not to assess the safety profile of MON810 but to reconsider the design, execution and interpretation of 90-day or extended feeding studies as well as *in vitro* and *in silico* approaches with whole food/feed in order to provide recommendations on the appropriateness of these tools for being considered in the risk assessment of GM plants. The rationale of this study has to be seen in the light of this overall context. Therefore, the outcomes of the chronic toxicity study will be important firstly to draw conclusion on the adequacy of extending the testing period with diets containing whole food/feed. Secondly the results of the chronic toxicity study will help to determine the biological relevance of possible alteration observed 1) by the different omics	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		approaches being part of the previous 90-day feeding trial and 2) by the upcoming subchronic toxicity and longitudinal metabolomics study. The reason for a possible deviation regarding the histopathological test site is given under comment 15.	
4	There are several places in the document where the *90-day study results* should be reviewed to determine if a planned endpoint or interval is necessary. The *rationale for those endpoints* should be based on the previously conducted 90-day study of which the results are apparently not known yet, or at the least not considered in this protocol. Are the histopath slides not evaluated, yet?	The objective of GRACE is not to reassess the safety profile of MON810 but to reconsider the value of 90-day or extended feeding studies (and possible alternatives) for the risk assessment of whole GM food/feed. The different approaches are applied in the course of GRACE to enable their comparative evaluation. This is not precluding a strategy of safety testing, whether an approach is obsolete, valuable on its own or only complementary useful. For the purpose of the project triggers are not considered *a priori*. Due to financial constraints the omics analyses of animal tissues and fluids will only be performed if results from the previous 90 day feeding study (Study No: 311957 - A / 13/ GLP) carried out with the same maize varieties indicate relevant effects to direct further testing.	N
5	These are really important points because it indicates decisions are being made based on information available to the scientists rather than a check-the-box approach. Moreover, it is in-line with information in OECD TG 452, which indicates previous data should inform study design.		
6	The proposed protocol *lacks clarity* and *relevant information* on the study conduction, diet stability assessment, methodological details, data collection and analysis, and reporting, with difficulties therefore in judging on the robustness of the proposed study.	The study plan was revised by taking the provided comments into account.	Y
7	A *lack of full compliance to GLPs* is noted even for standard evaluations, as well as and lack of adherence to the principles of *multisite study* conduction with possible impact on the overall quality of the proposed study.	The study plan was revised by taking the provided comments into account as far as possible. Nevertheless, the project has the nature of EU-funded research project with given limitations and constraints.	Y
8	In general in compliance with the OECD test guideline 452.	No response needed.	
9	In general, there is *room for improvement* of style, language and structure of the protocol (sometimes not clear, sometimes statements that are not usually made in toxicology test protocols).	No response needed.	
10	It is impossible to take this study seriously when the protocol includes a *disclaimer* (p.5) that states: The authors reserve the right not to be	Disclaimer was only included for the public consultation phase and will be excluded from the finalized Study Plan. It is neither an official	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	responsible for the correctness and completeness of the information provided. If there is no reasonable assurance of the correctness or completeness of the study, it is difficult to justify the use of animals to generate these data.	document from the EU nor a guideline to be followed for other purposes beside GRACE.	
11	This study will not end the debate, but could act as a precedent, risking to become mandatory. What is our responsibility as scientists in this matter?	As scientists, we have to ensure for the adequacy of a chosen experimental approach to generate a sound and solely evidence-based answer to a scientific issue/question under investigation. In the case of rodent feeding trials with whole food/feed the testing approach has never been validated explicitly. GRACE provides steps in this direction to underpin scientific arguments for choosing an adequate methodology.	N
12	Given that the OECD TG is designed for chemicals not complex mixtures, how to make sure that any differences that may be measured stem from the intended variation in the diet?	Animal feeding trials with whole food/feed are analysed for statistical significance and biological (pathological) relevance. The diets administered ideally differ (only) in their targeted test plant material and the comparator used, while the diets are nutritionally balanced. Moreover, the composition of diets and plant material is analysed in detail. Nevertheless, due to the complexity of the test material feeding trials may not provide full evidence in terms of distinct causes and effects. Further on targeted investigations may be necessary to substantiate the GMP effects. GRACE is supposed to highlight such critical issues in the comparative analysis of approaches.	N
13	*Title*: lack of consistency with the title of the previous 90-day rat study.	No response needed.	
	Specific		
	Staff, GLP, Data recording		
14	*Clinical Pathology* analysis is performed in a laboratory that it is not GLP compliant and no ISO compliance is mentioned. This is neither in line with EFSA Guidance for Risk assessment of food and feed from genetically modified plants (2011) nor with the Commission Implementing Regulation (EU) 503/2013. It is highlighted that accreditation by a National Accreditation System is not equivalent to GLP compliance.	ISO compliance was specified on p.17.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
15	Why is it not clear who will perform the histopathological evaluation? Usually all labs which take part in studies of this complexity should be determined beforehand.	Due to the principle "best value for money" applied to all EU funded projects and the local allocation guidelines at SZU a tender had to be launched.	N
16	*Histology evaluation (p.11)*: Please indicate whether or not the *pathologists* evaluating the slides will be *board-certified*. Board-certification is an excellent way to demonstrate professional depth and breadth, and might even provide some information regarding the individuals familiarity with the animal model under investigation. All critical factors in determining the qualifications of the individual chosen to conduct this analysis.	As the call for tenders is pending, details about the histopathologists will be provided in an amendment.	N
17	The *Histopathology test site* has not been identified yet. Histopathology must be conducted taking into account results from previous studies to guarantee consistency in the terminology used and in the findings interpretation, allowing a continuum in the safety assessment of the test item. If TOPALAB will not be chosen as test site for histopathology, a peer review of this study should be conducted by the same peer reviewer of the previous 90-day study. The *peer review* and the peer reviewer are to be indicated in the protocol. In addition a *Pathology Working Group* is also suggested (Mann and Hardisty, Toxicologic Pathology 2013).	Peer review will be conducted by the he same peer reviewer of the previous 90-day study as specified on p.20 (Prof. DVM. F. Jelinek, PhD, Dipl. ECVP, Veterinary Histopathological Laboratory Prague, member of European Society of Veterinary Pathologists). Furthermore, all raw data will be made available and will be open for scrutiny. Beyond this the project is limited in its financial resources and additional boards cannot be installed.	Y
18	*Histology evaluation (p.11)*: Also, a *peer review* is an excellent way to determine if one or more pathologists agree with the opinions of the first pathologist to read the study. It allows the group of pathologists to arrive at a consensus opinion and limits the influence of any one individual in the diagnostic process. Please advise if it will be included.	See comment 17.	Y
19	*Histopathology*: *Peer review* is a best practice highly recommended. The peer reviewer pathologist should be indicated in the protocol.	See comment 17.	Y
20	A *best practice* in toxicological settings is to have a *toxicological pathologist supervising* the necropsy session/s, to guarantee consistency among macroscopic observations. This is not indicated.	Specified on p. 36.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
21	No indication of the computerised systems used for *data evaluation and reporting* is provided (p.31).	Biochemical and haematological endpoints are measured by lab analysers connected to a PC and print outs are provided. All other endpoints (e.g. food consumption, animal weight etc.) are recorded manually and transferred to excel sheets.	N
	Objective		
22	The *rationale* and the *objectives* of this study are not clear but should be clearly defined and the experimental protocol fully aligned with these. Without a clear rationale and clear objectives, it is also difficult to justify the use of animals for a repetition of an already completed study in the context of 3Rs.	See comment 3.	N
23	*Objective (p.10)*: EFSA guidance indicates, "Depending on the outcome of the 90-day feeding study, further toxicity studies may be needed". Yet, 90-day studies have been conducted by GRACE, but not yet reported. The conduct of a chronic study before the results of the 90-day study are known appears inconsistent with EFSA guidance. Please consider reporting the results of the first study and determining whether additional testing is necessary based on those results before proceeding with a longer term test.	See comments 4 and 5.	N
24	*Objective* (p.10): The study plan claims to be in accordance with OECD Test Guideline 452, but there are main problems with/differences between the long-term testing strategy of, generally speaking, chemical substances (OECD) and that of whole GM food/feed:	We completely agree that internationally accepted test guidelines as provided by OECD where designed to test the toxicity of chemicals and not that of whole food/feed. Following the recent discussions the OECD guidelines were chosen as backing methodological guidelines for testing whole GM food/feed, though the overall concept to determine MTD cannot be realized. Consequently their direct application for the testing of whole food/feed is accompanied by some essential limitations compared to the OECD strategy. One example would be the sensitivity issue as the maximal incorporation level of the test matrix in the diets is not determined by the concept of MTD but by ensuring for their nutritional balance. In order to determine the appropriateness of different test designs/approaches for the safety evaluation of GM crops,	N
25	• Top dose and dose selection/dose level spacing *OECD Guidance Document 116* (Conduct and design of chronic toxicity and carcinogenicity studies) supports the concept of the *Maximum Tolerated Dose* (MTD), conventionally defined as the highest dose to produce toxic effects without causing death and to decrease body weight gain by no more than 10 % relative to controls, as top dose. The problem *with GMOs is that such a dose normally cannot be reached*, because even medium-term studies which are typically used for		

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	establishing the MTD show no or only slight (and unclear or statistically disputable) adverse effects with the highest tolerable doses in those tests. In contrast to pure chemical substances, with GMOs the tolerability normally doesn't depend on the toxicity of the active principle but on avoidance of nutritional imbalance of the whole food/feed, thus allowing administration of the active principle in the mg-range, at most. Therefore, the top dose is limited more by nutritional balance and metabolic non-disturbance than by classic toxicity (of the chemical test substance).	GRACE will compare and evaluate the gain in quality resulting from difference experimental designs including subchronic and chronic toxicity studies as well as "omic" and alternative *in vitro* approaches using whole food/feed as the test substance. The rationale for the inclusion of a control group getting only ordinary rodent diet for the safety assessment of GM crops using whole food/feed as the test matrix is rather questionable due to the following reasons. Following the guidance for risk assessment of food and feed from genetically modified plants provided by EFSA "the risk assessment strategy … seeks to deploy appropriate methods and approaches to compare GM plants and derived food and feed with their appropriate comparators". This directly implies that the only differences in the diets fed to the test animals have to stem from the incorporated plant matrix under investigation (in our case maize), whereas the inclusion level has to be tightly controlled in order to guarantee for the comparability of the gathered results. As GRACE is using maize as the test material, it has to be further insured that no other maize derived ingredients besides to the plant material under investigation were used for diet preparation as this may affect the nutritional balance of the final product. When considering the above mentioned issues and the fact that the overall maize content in "ordinary rodent diets" provided by Harlan range from 35-40% (considering both maize and maize gluten meal), an inclusion of such a test group would neither be appropriate to serve as a "zero test substance group" nor would it help to facilitate the interpretation of the generated results as deviations in the composition when compared to the other test diets having a 33% inclusion of maize *in toto* and no inclusion of maize gluten meal would jeopardize the comparability of the outcomes. Concerning the adequacy of implementing an additional conventional	
26	But not to be misunderstood, when starting with the highest nutritionally tolerable dose (for instance 30 % whole GM food/feed), possible long-term effects may become obvious even with low doses. In the "classic" OECD chronic toxicity test design also doses below the MTD which cause no effects in shorter-term studies are provided (and can produce long-term effects).		
27	None of the specific OECD guidelines was developed for the testing of whole foods but rather for testing the toxicity of isolated chemicals when feeding rats with standardized, nutritionally balanced feed. In most cases, the test substance is administered separately from the normal feed or mixed with the feed in a minor volume. As long as the addition of the test substance to the feed does not alter significantly the relative concentration of the feed components, a single control with zero test substance is sufficient. But such *an essential zero test substance group (control group getting only reference diet, i.e. ordinary rodent diet)* is missing. On the other hand, a conventional maize getting group is added. The actual effect of introducing the use of conventional controls lies more in providing an instrument to declare technology-induced significant differences as irrelevant against the background of unspecified 'noise' (stemming from broad, unrelated natural variation) than in setting up a		

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	sound methodology for risk assessment. In practice, adding genetically unrelated controls will increase the unspecific background variability in the data and make the statistical detection of specific differences due to the insertion of the transgenic trait at least more difficult. Moreover, the scientific opinion by EFSA regarding "Guidance on conducting repeated-dose-90-day oral toxicity study in rodents on whole food/feed" – for this subject also applicable - states (although partly because of other reasons) that "in the case of GM food, the inclusion in the experimental design of reference groups, fed with a diet containing commercially available food/feed similar to the test food/feed, in order to estimate the natural variability of endpoints is in general not recommended…." To summarize, instead of a conventional maize getting group an ordinary rodent diet group (reference group) should be employed.	variety in the test design, the EFSA scientific opinion on the "Guidance on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed" further clarifies that an inclusion of further conventional varieties "…should be considered if no acceptable historical background data is available". As no background data are available at SZU its incorporation is in line with EFSA recommendations.	
28	Regarding the sample size, EFSA prefers in its guidance document on conducting repeated-dose 90-day oral toxicity study in rodents an approach which is known as the "standardized effect size" (SES). Using this method, an appropriate sample size can be estimated from the SES, the significance level (usually set at 5 %), the power (often set at 80-90 %), and the alternative hypothesis (one or two-sided). This approach should be followed also with the long-term test, or an equivalent strategy for setting the number of animals.	For the GRACE project the chosen setting is a starting point for further critical considerations. More specifically: the lab has not been conducting such feeding studies with such diets before and historical data allowing to choose pathological relevant effect size(s) representative for the lab were not available *a priori*. Therefore the planning of the statistical power is based on recommendations from standard guidelines. A power analysis involves a mathematical relationship between six variables. Specify any five of them in order to calculate the sixth one. To determine sample size: 1. the power (say 80%), 2. the significance level (say 5%), 3. the alternative hypothesis (say two-sided), 4. the standard deviation and 5. the effect size of clinical importance need to be specified. Put these together and 6. the sample size can be calculated. However, in a toxicity test it is impossible to pre-specify the magnitude of difference between treated and control likely to be of clinical importance for every one of the 50 or so biomarkers which are measured because it is the pattern of response among all of them which is of importance. The standardised effect size (SES) is the	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		effect size (5 above) divided by the standard deviation (4 above). What EFSA did was to assume (somewhat subjectively) that an SES of 1.0 is unlikely to be of toxicological importance. If this is true, then the power analysis shows that for the above specification N=16.7. The suggested number for the chronic study is n=20 , which is not materially different but follows the OECD guidelines since an EFSA guideline for chronic toxicity studies with whole food/feed is not available. But uncertainties remain. Is it true that an SES of one or less is unlikely to be of clinical importance? What is the consequence of housing two rats per cage? The fundamental research on such questions has not yet been done. While the SES method is helpful, it is not a magic way of determining sample size. We hope that the GRACE project might begin to answer some of these questions.	
	Diet preparation and analysis		
29	*Pesticide treatment* maize production, if different, may generate *additional variatio*n in the diets, which may not become apparent from the compositional analysis.	Besides to the compositional analysis, a comprehensive list of e.g. chemical contaminants including different pesticides will be analyzed as mentioned in the study plan. Furthermore, all maize varieties were grown under the same conditions according to standard cultural practices and all treatments were recorded.	**N**
30	Will the *herbicide* used and possible metabolites be *dosed* in the food/feed?	No herbicides or possible metabolites will be dosed in the diets. Diets will be tested for presence of pesticides. Nevertheless, crop management conditions applied for the varieties are equal and typical for the region.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
31	*Diet formulation* (p.17): The maximal maize content of the diet is suggested to be just *33%*. In our view, it is *too low to be able to detect changes* which could be obvious with a higher dietary maize inclusion. The nutrient composition of maize makes possible a 50-60% inclusion comfortably. It would be advisable to increase the proportion of GM and/or conventional maize content of the diets to at least 50%, and have the 33% as the lower level of inclusion with, or without the 11% maize content.	The provider of the diet (Harlan) suggested to limit maize to maximum 40%. Commonly used are 33%. The actual factor between 33% and 40% is toxicologically negligible. We choose the standard amounts. Nevertheless, the maximum amounts of plant material used should be reconsidered for each crop with regards to a well-balanced diet and maximum "dosage".	N
32	*Diets (p.17)*: The rodent diets industry prepares for 90-day studies are considered stable for 6 months according to the manufacturer's specifications. Likewise we anticipate that the diets prepared by Harlan have a similar shelf-life, but we anticipate that it may not be for the entire length of this study. To *ensure the "freshness" of the diets* throughout the study, it may be desirable to stagger diet production throughout the course of the study.	The mode of diet provision by the diet manufacturer was specified on p.27. Two follow-up batches per dosage group will be provided and the diets can be considered stable for up to 9 months following the manufacturer's specification.	Y
33	Please advise if the text on p.17 'Formulation is carries out […]' is correct. Per the table on Page 21, the diets will consist of 67% reference diet, and 33% maize in total. This could be interpreted to mean that regular Harlan diet (or reference diet), which presumably already contains some maize, is being supplemented with an additional 33% maize. That would make the total maize content higher, and may impact nutritional balance. That could confound study results. Please revise the section p.17 and the p.21 table.	See specification on p. 26 and table on p. 31.	Y
34	No information on the *method for assessing the stability of the feed* material is provided, only information on the storage conditions are provided. Considering the duration of the study, it is recommended to elaborate a strategy for assessing stability of the feed material.	As the feed will be provided in 2 follow-up batches per dosage group, no specific stability assessment will be performed as the diets can be considered stable for up to 9 months following the manufacturer's specification	N
35	It is not clear from the study protocol whether the *compositional analysis* (p.18) of the diets will be, or was, completed prior to or following *gamma irradiation*. Should be clarified.	Specified on p. 27.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
36	Not indicated how the *potential nutritional imbalance* as a result of dietary incorporation of the test item is assessed.	Detailed analysis of diets will be performed in order to show their nutritional balance.	
37	Please create separate bullet points for *GMO DNA and proteins* as key parameters (p.18) as these are dietary components that are *not anticipated to lead to adverse effects* like mycotoxins, heavy metals, or other contaminants.	Adopted on p.27.	Y
38	A similar comment to the one above on '[…]as for the presence of genetically modified organisms […]' (p.18), indicating that *GMOs are not anticipated to be toxic* given the existing safety data. Consequently, their inclusion with contaminants known to cause adverse effects in the diet seems unusual and perhaps prejudicial.	Adopted on p.27.	Y
	Test and control groups		
39	*Control crops (p.15)*: The use of genetically unrelated reference varieties in GM feeding trials in the context of the EU GMO legislation and risk assessment system is inappropriate. The actual effect of introducing the use of unrelated reference controls lies more in providing an instrument to declare technology-induced significant differences as irrelevant against the background of unspecified 'noise' (stemming from broad, unrelated natural variation) than in setting up a sound methodology for risk assessment under the applicable legal framework. (more in ENSSER email and Meyer and Hilbeck).	As stated by EFSA for the conduct of 90-day feeding trials with rodents, the "Inclusion of reference groups should be considered if no acceptable historical background data is available". As no historical data are available at SZU, an additional conventional variety was included. Furthermore, seed companies have introduced the MON810 event into a large number of commercial varieties by conventional breeding methods while tests may be performed with a selection of pre-market lines. It is a precondition for the risk assessment that such an approach demanded for authorisation of a GM crop is robust towards the genetic background.	N
40	Control crops (p.15): one reference variety is not contained in the current EU seed variety catalogue, according to ENSSER's information. The draft plan does not describe for what purpose the data gained with this reference variety will be used.	The three varieties used in the chronic assay are registered (for more information please visit the plant variety database of the EC).	N
	Test System		
41	The *RCC Han Wistar strain* is the test system vs. the test facility has experience in the use of Wistar rat. Please clarify if the test facility has experience on RCC Han Wistar strain.	Mentioned on p. 26.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
42	Test system (p.16): If the *initial difference in the weights* of the animals within the same group, and/or within treatments is in the range of ±20% as it is suggested, an additional 5% difference is needed to be able to detect any significant differences between the end points applied for the different treatments. This means that within a year, instead of the ±5% weight difference a change of ± 20% plus 5% should be occurring between treatments for the differences to become significant. Therefore, in order to be able to detect a difference of $p<0.05$, the initial weight of the rats should not be more than ±5% within a treatment group and also between the different treatment groups.	It is standard practice in toxicity tests to specify that the animals within a group should vary no more than ±20%. Whether there is a statistically significant difference in mean body weight at the end of the test depends on the observed magnitude of the difference, the within group standard deviation, the number per group, the significance level and the alternative hypothesis. We can't predict the outcome of the experiment. If we could do so, we would not need an experiment.	N
43	We noted that *group size* is 20 rats of each sex by group.	Correct.	N
44	*Historical data* from the same rat strain and the same kind of study (1 year) performed by the same test site available?	No historical data for chronic toxicity trials with maize-rich diets are available at SZU.	N
45	It is indicated that the number of animals per group was determined by a power analysis for quantitative biomarkers, with 20 rats/group offering a 90% of chance of detecting a SES of 1.10; note that: • it is not clear what are the *quantitative biomarkers* used for the analysis; • power analysis is not taking into account the *gender*. This approach is *inconsistent* with the one proposed in the subchronic longitudinal study; furthermore provide clarification on the validity of the above power analysis in the case of gender related effect on parameters.	Specified on p.25. Also in the case of the subchronic toxicity and longitudinal metabolomics study, males and females will be combined for power analysis.	Y
46	1) *Haematology, clinical chemistry and urinalysis*: According to the Anses and EFSA recommendations, a group size of *16 rats* of each sex was suggested to have a 80% chance of detecting a standardized effect size of 1 SD assuming a 5% significance level. Consequently, why is the number of animals *10* by group for the blood analysis ? 2) Moreover, from an ethical point of view, how do you explain the sample size of 20 rats of each sex in the study design?	1) The study plan was adapted insofar, that at study termination all animals/group will be subjected to blood analysis. 2) The recommended number of 16 animals per group does apply for subchronic (90 day) toxicity studies only. The use of 20 animals/sex for the chronic toxicity study was determined by a power analysis considering both, quantitative and qualitative biomarkers. Furthermore, a group size of 20 animal/sex is recommended by TG 452.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
47	Housing *males and females in different rooms* IS NOT a standard practice in general toxicology studies. It does introduce an *additional source of variation*. Justification for this choice should be provided.	There seems to be a difference of opinion here. The sexes often are maintained in different rooms in toxicity tests, there is no general rule. Any sex difference will be confounded with differences between rooms. Sex differences as such are not of particular interest, but sex x treatment interactions may be of interest. These can still be calculated. The rooms are not sufficiently large to house all animals in the same room, so room effects will need to be considered in any case.	N
48	Statistical calculation of the *group size*: if calculated assuming only 1 variable (GMO), how would additional variation (pesticides, soil, mycotoxins …) affect the statistics?	The power calculations assume two groups are being compared (GMO vs control). If the experiment had more treatment groups then a decision would be needed on the way that a power analysis should be applied. However, the variables mentioned appear more like ones which increase the variation rather than being treatments. If variation is increased, power will go down.	N
49	There is no justification included for use of *multiple animal rooms*, or for housing animals in separate rooms, by gender (p.20).	See comment 47.	N
50	The *sample size* was determined by power analysis, however no information on the *mathematical model* and on the *data* used are provided. In particular it is not clear which are the *quantitative biomarkers* used for the power analysis.	Specified on p. 25.	Y
	Experimental design		
51	Regarding the number of test sites, the Good Laboratory Practice must take into account the *multi-site procedure*.	See comment 52.	Y
52	This is a *multisite study*, it should therefore be conducted according *OECD 13* (Consensus Document of the Working Group on Good Laboratory Practice - The Application of the OECD Principles of GLP to the Organisation and Management of Multi-Site Studies, 2002). In particular, the following points are to be addressed: • a *single final report* under the responsibility of the Study Director is to be produced; clarification on how data will be provided to the Study Director for inclusion in the study is to be provided;	All analysis performed at test sites located outside Slovakia are not in compliance with GLP principles and thus seeking the advice of the respective national GLP compliance monitoring authority where the site is located does not need to be considered. If the to be identified histopathological test site will be located outside Slovakia, the proposed specifications will be amended to the study plan. The Lead Quality Assurance Manager was specified on p. 19. A single final report will be compiled.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	• as this multi-site study is conducted in more than one country, the opportunity to *seek the advice of the national GLP compliance monitoring authority* where the site is located is to be considered; • a *Lead Quality Assurance* is to be identified, please specify this in the protocol.		
53	Is one *control group* sufficient to build up a *control data base* that can be used to position the results obtained from the GM treated groups?	The decision for the inclusion of only one conventional variety is based on the spatial capacities available at SZU. The inclusion of 2 control groups (the near-isogenic line and the conventional variety) will not be sufficient to build up a comprehensive control data base at this stage but it will certainly give an indication about the degree of possible variations in specific endpoint between 2 maize varieties with a history of safe use.	N
54	No *rationale for dose selection* beyond OECD TG 452 (e.g. omics and immunological testing). Basis should be the results of the 90-day study.	GRACE will reconsider the overall discussion on the usefulness of rodent feeding trials with whole food/feed and not preclude on "senseful" approaches. There are statements that 90-day feedings studies are not sensitive and extended are generally needed (e.g. discussions during the first stakeholder workshop). Anyway, because of the limited resources and as mentioned in the study plan "Omics" analyses of animal tissues and feed materials will only be performed on selected candidate metabolites, transcripts and proteins if results from a previous 90 day feeding study (Study No: 311957 - A / 13/ GLP) carried out with the same feed material and dose levels indicate significant effects."	N
57	For the *remaining 10 animals/sex* being used as sentinels (p.20): There is insufficient information in the protocol about how these animals will be treated, what diet they will be fed, or which data will be collected for them in support of health evaluations.	Specified on p. 25.	Y
58	"To minimize the chance of mistakes being made, cages of the same treatment groups will be clustered in vertically arranged groups"(p.20). This sounds like there will be a *lack of proper randomisation* which could seriously undermine the principles of good experimental design.	This is a reasonable compromise between the need to minimize possible mistakes and the need to ensure that each cage has, on average, the same environment within the animal room.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
59	Method and objectives do not describe the experimental design (Mixing diet GM1, GM2+non-GM, non-GM and conventional)(p.20).	We are unclear about the reference of the comment.	N
60	Food *ad libitum* (p.21): No nutritional experiment can be published unless all animals consume the same amount of food, or at least all animals placed in the same cage should consume the same amount of food per cage. To do this food intake should be monitored daily. The minimum food intake consumed by all animals in a cage on the first day should determine what amount of food is put into all the cages the following day. If on the following day all the food offered was consumed in all the cages, than additional food should be offered. This is usually 2-4g per cage. If this has also been eaten, more food should be given using the same principles. Again, the lowest consumption per cage should determine the amount of food offered the following day. Throughout the experiment the amount of food put into all the cages each day should be the minimum amount of food consumed per a cage the previous day. Therefore, the lowest amount consumed per cage should determine the amount for all the animals in a cage the following day.	Following from the OECD standard testing procedure TG 452, measurements of food consumption and food efficiency should be made at least weekly for the first 13 weeks and at least monthly thereafter, but not daily. Other approaches cannot be comparatively tested in the course of GRACE.	N
61	Rodents are sometimes known to dig in and spill their food. This can lead to *erroneous food consumption measurements* if it is not observed and noted. To minimize the potential for confounding the study data, please advise if steps will be taken to determine if food spillage will be documented, and to what extent it will be estimated to avoid the possibility that spillage will be recorded as excess consumption (p.21).	Specified on p. 30. Furthermore, only little food spillage was observed in feeding trials with the RCC Han Wistar strain conducted at SZU.	Y
62	No clear description how *feed consumption* will be reported and how *food conversion* is calculated.	Specified on p. 30.	Y
63	The measures to be taken to *avoid feed spillage* should be described.	See comment 61.	Y
64	*Water consumption* is not measured.	Following TG452 this is only necessary if the test substance is administered in drinking water.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
65	The *"double blind" code* could be easily identified, according to the RIKILT and INRA analysis.	When setting up the experiment, it was indeed anticipated that the outcomes of the analysis for transgenic DNA and proteins might actually help identify the GM maize and diets containing it if the sample contained the same code as the diet fed to the laboratory animals. This scenario was avoided by assigning different codes to the samples from the diets used for analysis as compared to the codes used on the diets provided to the animal testing facility. So even though the presence of MON810 could be linked to an analytical sample, the code used for that sample should be different from the code used on the same diet fed to the animals.	N
	Periodic Observations		
70	*Detailed physical examination and functional assessment (p.22)*: OECD TG 452 states, "For chemicals where previous repeated dose 28-day and/or 90-day toxicity tests indicated the potential to cause neurotoxic effects..." This is another reason why it is critical to know the *results of the 90-day study before starting the chronic study*. If no neurotoxic effects are noted on the 90-day study discovering the substance is neurotoxic on a longer study is unlikely.	See comments 4 and 5.	N
71	*Ophtalmoscopic examination*: no procedural details are presented.	Specified on p. 31.	Y
72	*Ophthalmologic assessments* are not needed in a 1 year chronic toxicology studies when this has already been done in a previous sub-chronic toxicity study.	Required when following TG 452; see comment 54.	N
73	*Body weight (p.22)*: Food consumption (FC) values are being recorded weekly for the first 13 weeks and then at least every two weeks thereafter. But body weight (BW) are being recorded weekly for the first 13 weeks and monthly thereafter. Suggest making the *FC and BW intervals the same* to enable the calculation of food efficiency at each interval.	Intervals for the recording of food consumption and body weight were harmonized and will be recorded weekly for the first 13 weeks and every second week thereafter.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
74	*Body weight (p.22)*: it is advised that body weight should be recorded for all animals throughout the experiment weekly. Fewer data points after 3 months will not provide all information in case of effects appearing only after a later date than 13 weeks.	Not needed when following TG 452; see comment 60.	N
75	There are no procedures described in the protocol to account for *incompatible animal pairs* to address fighting or other types of cage aggression (food hoarding, etc.)(p.20).	SOP number was specified on p.32 (SOP No. TOX/V 004) describing the procedure for the separation of incompatible animal pairs.	Y
76	No mention of *order of observations* – needs to be in random order, in a manner that takes proper account of the fact that 40 animals per day will be necropsied (p.23).	It was specified that the selection of animals will be accomplished by a random selection of cage pairs.	Y
	Procedures for sample collection		
77	The *selection process for animals assigned to clinical pathology sample collection* is not specified in the protocol (p.23). First 10 animals/group? Last 10? Random selection of individual animals? Random selection of cage pairs? What if a previously selected animal does not survive for all of the scheduled sample time points? Will it be replaced with a different animal? Fewer samples for a group and gender that experiences an early termination? How would a replacement animal's values be treated in the context of the overall data interpretation? If animals are not replaced, would the lack of some data at certain study intervals impact the statistical power for certain variables to an extent that it would be difficult to make conclusions from the data?	SOP number is given on p.34. The selection is randomized for cage pairs. Animals dying in the course of the study will not be replaced and missing data points will consequently reduce the power of the study.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
78	The *objectives of sample collection* should be better described, and should be in accordance with the study objectives.	As already mentioned before (see comment 4), this study has to be seen as a stand-alone experiment and thus all sampling times as described in TG 452 (also after 3 months) have to be included. As one objective of GRACE is to evaluate the gain in quality resulting from the implementation of a bunch of "omics" technologies, additional sampling is needed at study termination. Due to financial limitations, the respective analysis will be restricted to candidate genes, proteins and metabolites, possibly showing significant alterations in a previous 90-day feeding study, using the same feed material but planted in 2012.	N
80	General comment on *sample collection (p.23)*: OECD TG 452, paragraph 44, page 10 indicates the following: "It is generally considered that baseline haematological and clinical biochemistry variable are needed before treatment for dog studies, but need not be determined in rodent studies (30). However, if historical baseline data (see paragraph 50) are inadequate, consideration should be given to generating such data." Since it is not apparent how robust the *historical baseline data* is for 1 year whole food feeding chronic studies at the SMU, serious consideration should be given to adding a pre-treatment sample collection interval. This will not affect the overall number of sampling intervals if the 3 month intervals are dropped in accordance with the comments in the haematology, clinical biochemistry, and urinalysis sections. Briefly put, OECD TG 452 indicates that interval is not necessary if the results of a previous 90-day study indicate no effects at that time point.	As already mentioned before (see comment 4), this study has to be seen as a stand-alone experiment and thus all sampling times as described in TG 452 (also after 3 months) have to be included. Thus, the addition of a pre-treatment sample collection interval is unfortunately not feasible.	N
81	*Haematology (p.23)*: Per OECD TG 452, if no effect was seen on haematology in a previous 90-day study there is no need to conduct the haematology evaluation at 3 months. Please consider the results from the first 90-day study with MON 810 before including this evaluation interval.	As already mentioned before (see comment 4), this study has to be seen as a stand-alone experiment. Moreover haematology analysis after 3 months may help to assess the robustness of the system by comparing it with the results from the previous 90 day study.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
82	*Haematology (p.23)*: Although only small amounts of blood are taken each time, it stresses the animals. Our suggestion is that initial blood sample should be taken from all animals and then alternating the 10 animals for sampling at a later time point. It is also advisable to collect the final blood samples from all animals; it will increase the power of statistical analysis greatly.	The suggestion to collect final blood samples from all animals was incorporated in the study plan. The proposed alternating sampling mode was not adopted as TG 452 recommends to use the same 10 animals throughout the whole study (see comment 60).	Y
83	*Clinical chemistry (p.24)*: Per OECD TG 452, if no effect was seen on clinical chemistry in a previous 90-day study there is no need to conduct the clinical chemistry evaluation at 3 months. Please consider the results from the first 90-day study with MON 810 before including this evaluation interval.	See comment 81.	N
84	*Urinalysis (p.25)*: Per OECD TG 452, if no effect was seen on clinical chemistry in a previous 90-day study there is no need to conduct the clinical chemistry evaluation at 3 months. Please consider the results from the first 90-day study with MON 810 before including this evaluation interval.	No urinalysis was performed in the course of the previous 90 day study; see comment 81.	N
85	"At months 3 and 6 and 7 days before sacrifice, *blood samples* from the tail vein will be taken from 10 male and 10 females per group after 16 hours fasting using the same animals throughout for haematological (as well as clinical chemistry) examination"(p.23). Which 10 - one from each of 10 cages or two from each of 5 cages? Same issue applies to Clinical Chemistry and Urinalysis.	The 10 animals from the same 5 cages will be sampled.	Y
86	*Clinical chemistry (p.24)*: sampling from a different anatomical site as indicated here can have an effect on the parameters under investigation. Therefore, it is recommended to use the same collection site for all intervals to minimize the potential for variability in the endpoints.	The rationale for the use different sites for sample collection (tail vein vs abdominal vessel) is determined by the amount of blood needed to be collected for the respective analysis at study termination.	N
87	*Clinical pathology* is performed on a subset of animals (10 out of 20 rats/sex/group). It is highlighted that bleeding procedures on a subset of rats introduce a variable, possibly impacting on consistency in	The number of animal used for clinical chemistry is given by TG 452, recommending the use of at least 10 animals/sex/group. To determine a possible impact on the health status of the test animals, haematology	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	histopathology analysis between animals in the group.	and clinical biochemistry will be performed at the end of the study on all test animals. Furthermore, possible inconsistencies in the outcomes of the histopathological analysis within groups will be checked against the history of each animal.	
88	*Urinalysis (p.25)*: sampling from a different anatomical site as indicated here can have an effect on the parameters under investigation. Therefore, it is recommended to use the same collection site for all intervals to minimize the potential for variability in the endpoints.		N
89	Detail on the work organization of the *sampling* of fluids and tissues should be in the appropriate *SOP*, not the protocol.	The respective details were included due to stakeholder requests on the previous 90-day study plan.	N
90	EDTA will be used as an *anticoagulant* for the measurement of PT and APTT. *Citrate* may be more appropriate?	Adopted on p. 35	Y
91	*Clinical biochemistry measurements* in serum or in plasma? If plasma, Li heparin should be used as an *anticoagulant*.	Will be performed on serum and was specified on p. 35.	Y
92	No specific *gravity and microscopic analysis* of the sediment are planned for *urinalysis*.	The chosen endpoints are in accordance with TG 452.	N
	Omics		
94	*Purpose* of the study: to determine the added scientific value of chronic toxicity compared to sub-chronic toxicity. Therefore we consider that omics must be performed whatever the results of the 90-day toxicity study are. If significant omics signatures are observed only after a 1-year study, it is a relevant information for the limit of the 90-day as sentinel study.	Due to financial restrictions, the full set of omics analysis cannot be performed in the course of the chronic toxicity study. A scientifically sound comprise will be applied by focussing on selected candidate metabolites, transcripts and proteins if results from a previous 90 day feeding study (Study No: 311957 - A / 13/ GLP) carried out with the same feed material and dose levels indicate significant effects.	N
95	As already proposed for the 90-day study, the supplementary analyses at INRA will not be performed according to GLP requirements.	This is correct and that`s why test site 5 is not listed when conferring to test sites being in compliance with GLP principles. Analyses will be performed using the Quality reference system developed and used for research and experimentations at INRA in order to meet the objectives of INRA's Quality policy: traceability of research activities and reliability of measurable results.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
96	*Metabolomics, Transcriptomics and Proteomics (p.29)*: No mention is made of whether or not these studies will be *GLP compliant* or what *procedures* will be followed in the Protocol. Given their tentative nature this is probably OK for the time being. However, if included in the study the absence of analytical specifics for these samples and the endpoints under investigation would be highly unusual. Such details are provided in protocols or study plans to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs. If included in the study it will be necessary to amend the protocol to add those details.	If these analyses will be included in the study, the mentioned details will be amended to the study plan. Omics analyses are not conducted with GLP compliance.	N
97	*"Omics"*: *historical data* for this rat strain?	No	N
98	*"Omics"*: What kind of "*standards*" are proposed for the *experimental quality and data handling* of "omic data" (e.g. such as those defined under the acronyms of MIAME).	As only the sampling is considered being part of the study covered by the GLP-constrained study plan and due to the tentative (explorative) nature of the analysis, supplementary details will be provided as an addendum. At this level of performing this study, no need for "standards" is necessary. Indeed, we need to be aware of the high quality of the experimental design, particularly to discard any confounding factor that could be detrimental in the interpretation of results. In addition, some "analytical" controls in the data production for building of omic (metabolomic) raw databases will be used to get unequivocal quality parameters to check some deviations of data with time (rank of analysis) and the precision of the mass (m/z) of analytes that characterises metabolomic variables.	N
99	Immunological studies of plasma for INRA (p.29): This is not a standard endpoint in studies compliant with OECD TG 452. If included, please clarify why this parameter was chosen for inclusion in this study.	Although this is not a standard endpoint, this is an important issue considering the high prevalence of food allergies, and it corresponds to a societal demand to guarantee that the genetic modification will not increase/create allergenicity risk - please see answer to comment 79 - subchronic/metabolomic.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
100	Additional detail must be provided on analytical specifics for these samples for INRA and the endpoints under investigation in order to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs (p.29).	As only the sampling is considered being part of the study covered by the study plan and due to the tentative nature of the analysis, additional details will be provided as an addendum.	N
101	Metabolomic studies of tissue samples for INRA (p.29): This is not a standard endpoint in studies compliant with OECD TG 452. If included, please clarify why this parameter was chosen for inclusion in this study.	This is typically an explorative part of the research devoted to the characterisation of any metabolic (physiological) deviation or disruption induced by a subchronic (90 days) feeding exposure of animals to a GM-based feed. Indeed, for some of scientists involved in the GRACE partnership, toxicological studies used to characterize possible physiological deviations related to such GM-based feeding that are described in the OECD TG452 recommendations would be better performed considering omic techniques like metabolomics which a sound phenotyping tool usable to detect some subtle metabolic deviations. For such a subchronic study, only the late stage (90 days) is considered. In the longitudinal study, a longitudinal follow-up is organized to reinforce modelling of the variance with a random part attached to every individuals submitted to this follow-up (mixed-effect models). In addition, with a fine interpretation of factorial maps coming from multivariate statistical analyses, it will be easier to address part of the variance explained by a gender effect, a GM effect, a dose effect in the GM one if present, or last by a cultivar effect, in three different matrices (plasma, liver, kidney). If deviations are observable in these matrices, a canonical factorial analysis would help to detect what variables are involved in the explanation of deviations that are shared by two matrices.	N
102	Additional detail must be provided on analytical specifics for these samples for INRA and the endpoints under investigation in order to determine if the researchers followed the path set by the Study Director.	See comments 96 and 100.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs (p.29).		
103	"*omics" samples* to be collected from the multiple animal *organs/tissues (p.29)*: The protocol specifies the anatomic location of the sites for collection of liver and kidney tissue. However, the protocol also indicates that all gross lesions will be evaluated microscopically. It is not uncommon in long-term (subchronic and/or chronic) toxicity studies for animals to have idiopathic, or age-related gross lesions of the liver and/or kidneys. This protocol does not include specifications to address the likely potential that a gross lesion may be present at a location within a tissue that is scheduled to be sampled for "omics". Is it more important to understand the morphology of a lesion or the molecular profile of tissues containing a lesion? How would molecular analysis of lesion tissue potentially impact the overall results and interpretation of the "omics" evaluations, if conducted?	We have no doubt that such "abnormal" observations, even arising by chance, will be duly documented. If the molecular content of such a tissue is evidently modified in connection with the gross observation obtained previously, we would obtain an atypical metabolic fingerprint of this sample. Therefore, in multivariate analyses such as in principal component analysis, this individual should be projected in the factorial map as an outlier. Indeed, when we will obtain such factorial projections, even in absence of coding key to get assignment of individuals to any treatment group, we will question colleagues for any living animal observation or macroscopic tissue observation that would be indicative of a higher probabilistic outlier status of such animals. Clearly, this modus operandi will be in line with GLP recommendations that govern ongoing of such studies.	N
104	*Transcriptomics of tissue samples* for FUB (p.30): This is not a standard endpoint in studies compliant with OECD TG 452. Please clarify why this parameter was chosen for inclusion in this study.	Transcriptomics analysis will only be performed in selected cases of specific questions focusing on distinct pathways. This is not following a routine OECD procedure, but we are furthermore interested in research and scientific investigations. Therefore, new findings will be of interest for a chronic study to discuss suitable analytical necessities.	N
105	If included, additional detail must be provided on analytical specifics for these samples for FUB and the endpoints under investigation in order to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs (p.30).	See comments 96 and 100.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Pathology		
106	It is very important that the chosen pathology lab provides a *sound historical control data basis*.	The contract will only be awarded to a GLP certified laboratory with a long lasting experience in the histopathological evaluation of rat tissues.	
107	*Gross necropsy* (p.26): Growth of the body and internal organ growth and development depend on the amount of food consumed. It is important that all animals should consume the same, or nearly the same amount of food. Rather, animals in all the cages should consume the same amount of food. Since food interacts with the gut, it is of crucial importance and therefore suggested that the wet weight of the stomach, duodenum, jejunum, ileum, caecum and colon is also weighed and recorded. For this it is necessary to wash out the parts of the gut with ice cold distilled water. The detailed procedure is as follows: cut out the gut, remove the stomach, the small intestine (in one piece), the caecum and colon. Measure the full length of the small intestine (s.i., cm). Fill a siring with ice cold distilled water and wash out the small intestinal content with it. Lay down the empty small bowel on a glass plate, and cut out the sections needed for measuring tissue weight and for histological examination. Histological examination should be performed on all parts of the bowel. The recommended sections to be taken for histology are the followings: Starting from the pylorus, the 3-5 cm part is the tissue section for the duodenum, and the part from 23-25 cm is the tissue section for the jejunum. The part of 3-5 cm from the caeco-ileal junction is the section for the terminal ileum, and the 23-25cm part from the caeco-ileal junction is the section for the ileum. For determining tissue weight weigh, the weighs of the 0-3 cm piece (as duodenum) and that of the 5-23 cm piece from the pylorus (as the jejunum), as well as the weighs of the 0-3 cm piece from the caeco-ileal	One major aim of the GRACE project is to provide the results of well conducted feeding studies with whole food/feed derived from GM plants, carried out according to relevant recommendations, i.e. OECD and EFSA guidance. Therefore the focus is on the analysis of reliable, internationally accepted parameters (as recommended by OECD TG 452), namely the histopathological analysis of gastrointestinal tissues. It is noted that the general lack of information on the normal variation of the stomach and intestinal weights would make it difficult if not impossible to evaluate any differences with regard to their toxicological relevance.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	junction (as the terminal ileum) and the 5-23 cm piece from the caeco-ileal junction (as the ileum), plus the weight of the remaining part between the jejunum and the ileum should be added together to get the weight of the small bowel (preferable corrected for the missing parts used for the histological examination). Since the water content of the bowel tissue might differ as a result of washing, it would be nice to freeze dry the intestinal samples and weigh them again to get the dry weight. In our view, this would give better indication of existing differences, if any, between treatments. The small intestine should be cut to the following sections: 0-5cm duodenum /5-25cm jejunum /remaining part / 23-5cm ileum /5-0 cm terminal ileum /0-3 cm duodenal tissue/ 5-23 cm jejunal tissue /remnant of s.i. tissue /23-5cm ileal tissue /3-0cm terminal ileal tissue 3-5 cm histology for the duodenum/23-25 cm histology for the jejunum /25-23cm histology for the ileum /5-3 cm histology for ileal tissue.		
108	*Lungs* (p.26) are not typically weighed as part of studies conducted in accordance with *OECD TG 408*. Furthermore, lungs are typically inflated with fixative promptly after harvest to ensure optimal tissue preservation for the histopathological analysis. This could add significant variability to the organ weight collection and provides another reason for not collecting this weight.	Lungs are routinely weighed at SZU before inflating them with the fixative. No negative effect on tissue preservation was observed in the past.	N
109	In the *Histopathology* section, it should be stated what samples will be analysed if they find significant differences between high dose test and control group. Would the next step be histopathology of the low dose test? In such a circumstance, would the reference group also be evaluated? If not, under what circumstances would it?	Specified on p. 39.	Y
110	The *procedure* followed for *gross pathology recording* (macroscopic observation) is not indicated.	Specified on p.37.	N
111	No information is provided on the *approach and software system* (if any) used to record organ weights and for macroscopic observations.	Specified on p.37.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
112	No information is provided if organs are weighed separately or paired (p.26).	Specified on p.38.	Y
113	*Histopathology*: *Trimming procedures* vary from laboratory to laboratory, differences should be evaluated carefully by a professional pathologist.	A supervising toxicological pathologist will be present as specified on p. 37.	Y
114	"Additional tissues may need to be investigated based on clinical or any other findings"(p.27): Any additional tissues to be investigated should be added by Protocol Amendment to ensure transparency.	Will be done.	
115	*Histopathology* (p.28): The biggest alterations occur in the stomach, jejunum and ileum, they also should be examined. Villi and crypts length/cell numbers should be measured/counted.	In the study internationally agreed parameters (according to OECD TG452) will be analysed, which includes a histopathological analysis of gastrointestinal organs and tissues.	N
116	*Formalin* is not an appropriate fixative for eyes and male reproductive tissues (p.28). *Bouin's or Davidson's solutions* should be considered to preserve these tissues for microscopic examination.	Adopted on p. 37.	Y
117	*Histopathology*: it is not clear from reading, but it sounds as though the *study pathologist* may be *blind* to treatment. This is typically not the case and would not follow best practices developed and agreed by toxicology pathologists. Pathologists typically start by evaluating control slides first to understand background incidence and severity of histological lesions (normal variation in the broad sense and within a specific study) before evaluating the treatment groups.	In order to reduce bias and increase confidence it was decided to unblind the study after the histopathological evaluation. Blinding will certainly reduce bias and thus increase the validity of the results.	N
	Data evaluation and statistics		
118	No indication whether statistical analysis will be conducted on individual organ weights or paired organ weights (p.31).	Data will be collected on individual organs. These will be analysed separately.	Y
119	"*A one-way analysis with planned or post hoc comparisons will be used*"(p.31). It is not clear why a choice of planned or post hoc comparisons is proposed. If this choice is to be retained then clear guidance needs to be provided on when the comparison should be planned and when conducted post hoc.	Specified on p.42.	Y

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
120	*Statistical analysis procedures* to be used influence study design constraints, and should be *specified prior to conduct the study*. This is especially critical for full transparency of the intent of the study. "Planned or post-hoc comparisons" and determining which analyses are appropriate "decided on a case-by-case basis" imply that study data will be collected and examined prior to selective application of unspecified statistical tests. The protocol should provide sufficient detail regarding the planned analyses such that the study statistician receives clear guidance; otherwise, *post-hoc* bias (intentional or unintentional) may be introduced at this stage (p.31).	Specified on p. 42.	Y
121	If the compositional analysis would show relevant differences in the presence of toxicants or anti-nutritional factors, how can this be taken into account such that any health *effects* or other differences can be attributed to the correct *cause*?	Various considerations come into play: Which effects are observed and is it plausible that these are linked to the differences in composition based on the magnitude of the difference and the known properties of the particular compound being different (for example, at which dietary level can toxicity be expected to occur or not). Maize kernels do not contain intrinsic toxicants, whereas various anti-nutrients, particularly phytic acid, can occur. Other sources of toxicants and anti-nutrients could be other ingredients in the diet (mainly wheat and soybean), which are kept to the same inclusion levels across diets. Besides intrinsic plant compounds, contaminations, such as mycotoxins formed by moulds that might infect the plants could theoretically play up. In another European research project, it was indeed found that contamination of GM maize MON810 with the DON mycotoxin accounted for the effects in salmon fed diets containing this maize, so the effects indeed could be related to presence of DON and not particularly linked to the presence of GM. Any observed differences will be taken into account when analysing the results of the trials.	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
122	Final report: The study report should include all of the raw data, not “relevant” raw data. Who is to decide, in interpretation of study data, what is relevant and what is not? Providing the entire raw data set will ensure adequate transparency and allow the public to evaluate the results in an unbiased manner (p.33).	"Relevant" was deleted.	Y

Subchronic toxicity and longitudinal metabolomics study

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	General		
1	Recommendation to state and thoroughly detail the *rationale and the objectives* of the proposed study. This lack does not allow to understand if and how this study could provide a continuum in the assessment of the safety profile of MON810 within the GRACE project. A lack of consistency between the proposed 90-day study and the former 90-day are noted. A clear description for the rationale is needed in order to justify the use of animals in the context of 3Rs.	This study is a pilot study planned to determine the added scientific value of implementing additional metabolic endpoints in the experimental design of 90-day studies used for the risk assessment of GM food/feed using Monsanto and Pioneer maize MON810. Further, it aims to provide guidance for the use and improvement of existing and new assessment tools for GM food and feed safety evaluation. For that purpose, a longitudinal follow-up of every animal will be performed concerning their metabolic and immunological status and their respective variations. An untargeted metabolomic exploration will be performed in plasma, urine and tissues and will privilege extensive multivariate data mining to get the hidden statistical structuring of datasets summarized in statistical factorial maps. Structural identification of putative biomarkers involved in such a structuring of the metabolomic information will be tentatively performed. The more frequent sampling in the course of the 90-day feeding trial in comparison to the standard protocol might disturb / stress the animals and alter the study results accordingly. Therefore a separate pilot study was conducted.	**Y**
2	The proposed protocol *lacks clarity* and relevant information on the study conduction, diet stability assessment, methodological details, data collection and analysis, and reporting, with difficulties therefore in judging on the robustness of the proposed study.	The study plan was revised by taking the provided comments into account	**Y**
3	There are several *big concerns*, concerns that have been already sent to the GRACE team in our initial comments on the 90-day protocol last year and which have not been taken into account.	Following the first consultation round on the design of 90-day feeding trials in December 2012, a total of 147 comments were provided by the different stakeholder groups. For practical reasons, a pre-screening was performed according to their anticipated need for being considered/incorporated before study initiation. Out of the 147	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		comments, 127 were discussed internally and the study plan was revised accordingly if it seemed appropriate in the frame of the project.	
4	There is lack of detail in the document. If the protocol is supposed to guide all study conduct, the level of detail provided for many of the metabolomic and immunostimulatory endpoints probably will not be sufficient.	The level of detail was increased, taking the comments on the study plan into account (see "specific" comments).	**Y**
5	A *lack of full compliance to GLPs* is noted even for standard evaluations, as well as and lack of adherence to the principles of *multisite study* conduction with possible impact on the overall quality of the proposed study.	See comments 13, 41 and 42.	**Y**
6	In general in compliance with OECD test guideline 408.	No response needed.	
7	The protocol is referring to the EFSA Guidance (2011) which gives new indications for the realization of 90-days studies, especially on statistics.	No response needed.	
8	In general, there is *room for improvement* of style, language and structure of the protocol (sometimes not clear, sometimes statements that are not usually made in toxicology test protocols).	No response needed.	
9	It is impossible to take this study seriously when the protocol includes a *disclaimer* that states (p.5): The authors reserve the right not to be responsible for the correctness and completeness of the information provided. If there is no reasonable assurance of the correctness or completeness of the study, it is difficult to justify the use of animals to generate these data.	Disclaimer was only included for the public consultation phase and will be excluded from the finalized Study Plan. It is neither an official document from the EU nor a guideline to be followed for other purposes beside GRACE.	**N**
10	"The study will be carried out in accordance with OECD Test Guideline 408" (p.7): the present study shows difficulties and differences as with the chronic study.	See comment 24-27 on the chronic toxicity (1 year) study.	**N**
11	*Title*: lack of consistency with the titles of the other rat studies within GRACE.	No response needed.	

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Specific		
	Staff, GLP, Data recording		
12	The *Histopathology test site* has not been identified yet. Histopathology must be conducted taking into account results from previous studies to guarantee consistency in the terminology used and in the findings interpretation, allowing a continuum in the safety assessment of the test item. If TOPALAB will not be chosen as test site for histopathology, a peer review of this study should be conducted by the same peer reviewer of the previous 90-day study. The *peer review* and the peer reviewer are to be indicated in the protocol. In addition a *Pathology Working Group* is also suggested (Mann and Hardisty, Toxicologic Pathology 2013).	Peer review will be conducted by the he same peer reviewer of the previous 90-day study as specified on p.20 (Prof. DVM. F. Jelinek, PhD, Dipl. ECVP, Veterinary Histopathological Laboratory Prague, member of European Society of Veterinary Pathologists). Furthermore, all raw data will be made available and will be open for scrutiny. Beyond this the project is limited in its financial resources and additional boards cannot be installed.	**Y**
13	*Clinical Pathology* analysis is performed in a laboratory that it is not GLP compliant and no ISO compliance is mentioned. This is neither in line with EFSA Guidance for Risk assessment of food and feed from genetically modified plants (2011) nor with the Commission Implementing Regulation (EU) 503/2013. It is highlighted that accreditation by a National Accreditation System is not equivalent to GLP compliance.	ISO compliance was included on p. 17.	**Y**
14	*Histopathology* (p.26): please indicate whether or not the pathologists evaluating the slides will be *board-certified*. Board-certification is an excellent way to demonstrate professional depth and breadth, and might even provide some information regarding the individual's familiarity with the animal model under investigation.	As the call for tenders is pending, details about the histopathologist will be provided in an amendment.	**N**
15	No indication of the computerised systems used for *data evaluation and reporting* is provided.	Biochemical and haematological endpoints are measured by lab analysers connected to a PC and print-outs are provided. All other endpoints (e.g. food consumption, animal weight etc.) are recorded manually and transferred to excel sheets.	

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Objective		
16	The *rationale* and the *objectives* of this study are not clear but should be clearly defined and the experimental protocol fully aligned with these. Without a clear rationale and clear objectives, it is also difficult to justify the use of animals for a repetition of an already completed study in the context of 3Rs.	See comment 1.	**Y**
17	*Objective* (p.9): Closely examining the *scientific value of metabolic endpoints* for the safety assessment of GM crops is a worthwhile goal. Metabolic or metabolomic studies, like many –omic analyses, are a relatively new technology, for which biologically relevant differences and diagnostic utility have not been proven. Thus, it is noteworthy that this is being done on the second feeding study with these test substances. Furthermore, the first study, "90-day feeding study of rats with Monsanto MON 810 maize", included metabolic endpoints over the objections of EuropaBio. Examining the scientific value of an analysis after already including it as a potential diagnostic tool in a previous study is highly unusual. Nonetheless, this more in-depth review is consistent with feedback provided by EuropaBio prior to the conduct of the first feeding study, where caution was raised that -omics technologies have not been validated – or more simply put, proven robust and rugged enough to produce reliable, reproducible results – and are therefore not ready for use as a diagnostic tool. This new study appears to be an attempt to address that concern after-the-fact with another feeding study. Another feeding study is not needed; the financial resources, effort, and time of the GRACE Project would be better spent on a pilot study to validate metabolomic changes associated with a well-characterized toxic mechanisms (e.g., drug induced liver injury via acetaminophen) or an animal disease model. Published standards for such undertakings are available from the International Organization for Standardization (ISO). Once metabolic	This comment translates with no doubt the gap there is when we try to access to a comprehensive and unequivocal register (database) of unitary disruptions that would be anticipated at the metabolic level when considering now metabolomics and the classical approach used in toxicity assessment. In our opinion, it is a deep error to try to compare what is done in this explorative study with "classical" methods and what would be anticipated given some increase in the protocol as it is suggested. This is easily explainable by the fact that conclusions are only accessible thanks to a sophisticated datamining based on multivariate statistical analyses. Clearly, with a convenient mining procedure, only main factorial axes need to be interpreted in terms of physiological signatures given the fact that we have access to projection of group of individuals and hence to the main homeostatic metabolic adjustments. If the experimental design is well built, no confounding factor would prevent to get a hierarchy in the respective effects of the different controlled factors or their main interactions. Only, repeated approaches would be helpful in building a share and comprehensive repertoire of metabolic signatures when we are comparing similar GM traits or events. This could be a very sound objective of regulatory agencies such as EFSA in EU to compile all these data produced in a similar experimental frame (subchronic + longitudinal) to get then a compendium of complex metabolic signatures (coming from understanding of factorial axes) that are built by the coordinate combination of main metabolic biomarkers.	

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	differences are reliably associated with toxicities those endpoints can be monitored of safety studies to determine if similar effects are seen in response to the application of the test substance to the test species. Similar methods were used to validate the current battery of clinical chemistry, haematology, and urinalysis diagnostic assays; and they have proven reliable in detecting toxicities over the past several decades.		
18	Objective (p.9): it is not clear why there is any need to compare the safety profile of MON810 maize to its non-GM counterpart or to commercial varieties, as MON810 has already undergone a thorough safety evaluation. The rationale for conducting the proposed studies and the main objective of the project has not been considered in sufficient detail. This important point was raised during the stakeholder consultation by many participants, including several Member State Competent Authorities, EFSA, NGOs and industry members, notably in terms of significant concerns with repeating existing work and animal welfare issues. Numerous 90-day feeding studies have already been conducted with the MON810 maize event as well as many other products containing this event, notably in independent contracting laboratories and with no reported adverse health effects. Thus, the scientific rationale for continued testing is not apparent to EuropaBio and appears to be inconsistent with Article 38 of EU Directive 2010/63/EU.	The purpose of this pilot study is not to determine the safety profile of MON810 but to assess the adequacy and validity of complementary metabolomics approaches to inform the safety evaluation GM food and feed (for more detail see comment 1).	**Y**
19	Objective (p.9): Current -*omics profiling studies* are highly heterogeneous (Ricroch, 2012), and should be *standardized and independently validated* to reach sound conclusions regarding their ability to detect relevant effects (Blankenburg et al., 2009). In general, -omics technologies generate thousands of data points, most of which are un-interpretable because of a lack of familiarity and understanding of the biological significance because of the limited data available in the test species. A fundamental flaw with using –omics technologies for	It is false to contest such an approach by saying that there is no hypothesis to test. We should be aware of the necessity to perfectly design the study before performing it, particularly to foresee what should be attributable to a confounding factor effect. As for classical toxicological studies, we need to get controlled factors to perform the study. The only difference is the fact we work directly at a multidimensional level, not at a univariate one. But even we are dealing with hundred to thousand variables, calculation of the degree of	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	hazard identification is that there is no hypothesis to test, which ultimately leads to a bias toward false positive results (Chassy, 2010). At this point in time, all relevant health effects are evaluated in the safety assessments for biotech crops (Codex 2009). It is not known if any of these additional analytes, generated by the -omics technologies have any relevance to safety assessment.	freedom of the study which is necessary to weigh the residual variance is always given by the total number of individuals and secondary by the number of controlled factors and their levels. This rule is highly evident since pioneering works of Ronald Fisher in the twenties of the 20th century and is still true! What is important is to interpret what are the "metabolic" movements (adjustments) linked to a change of feeding when it is based on GM cultivars. Probably it will be of minor importance to get access to some potent biomarkers which would mainly sign a metabolic (homeostatic) adjustment that would be not necessary a toxicological or subtoxicological one. The wish to try to access to these biomarkers without remembering that they are enclosed in a set of biomarkers that are co-ordinately adjusted is highly risky because this will open the possibility to over-interpret these biomarkers.	
20	The *objectives of sample collection* should be better described, and should be in accordance with the study objectives.	In our opinion, it is well described in the introductive part of the study (also see comments 17 and 19).	**N**
	Diet preparation and analysis		
21	All *crops*: *production* during the 2012 season, all in a small area in the Empordà […]'(p.13): Please indicate whether or not EFSA field trial specifications were used to determine if the grains were grown under similar conditions. Indeed, a robust field trial design must be used (including several replicated plots for each corn entry) in order to be able to link any possible difference observed in the metabolomics analysis with the type of corn entry, and not with any other variable, such as the field location.	All varieties were cultured in a very small area in the Mas Badia agricultural experimental station, that has been used for many years by Mas Badia researchers and it is very well known as being very uniform in terms of soil profile and agricultural conditions. Being such a small area and having all varieties been cultured under the very same agricultural conditions (same sowing date, density and type of sowing, irrigation type and patterns, herbicide and fertilization treatments, harvesting machinery and date, etc.), climatic conditions were the same and possible differences due to environmental conditions were minimized. In addition, a number of parameters were monitored in every variety, including production yields, health, morphological parameters, etc. and all values were as expected, in comparison with the many seasons' experience of Mas Badia in comparing many different maize varieties in the very same area (using split-plot designs). Thus, the main differences	

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		between the grains of different varieties used to prepare feed should certainly be the variety itself.	
22	*Pesticide treatment* maize production, if different, may generate *additional variatio*n in the diets, which may not become apparent from the compositional analysis.	Besides to the compositional analysis, a comprehensive list of e.g. chemical contaminants including different pesticides will be analyzed as mentioned in the study plan. Furthermore, all maize varieties were grown under the same conditions according to standard cultural practices and all treatments were recorded.	N
23	Formulation was carried out according to the *diet composition* (p.15): Please advise if this text is correct. Per the table on Page 19, the diets consist of 67% reference diet, and 33% maize *in toto*. This could be interpreted to mean that regular Harlan diet, which presumably already contains some maize, is being supplemented with an additional 33% maize. That would make the total maize content higher, and may impact nutritional balance. That could confound study results. Please revise this section or the table on page 19.	Besides to the milled maize harvested within the GRACE project, no further maize ingredients where used for diet preparation. For clarification, p. 25 and the table on p. 29 were revised.	**Y**
24	Key parameters, isoflavones and lectins (p.16): Is there any lectin, and is it an antinutrient in maize?	Lectin was measured in diets because they contained both wheat and soy ingredients, which are known to be sources of lectins (e.g. as mentioned in OECD consensus documents).	N
25	"GMOs (DNA), Cry1Ab protein, pesticide […]"(p.16): Please create separate bullet points for *GMO DNA and proteins* as key parameters as these are dietary components that are *not anticipated to lead to adverse effects* like mycotoxins, heavy metals, or other contaminants.	Adopted in the study plan.	**Y**
26	"As well as for the presence of genetically modified organisms"(p.16): A similar comment to the one above, indicating that *GMOs are not anticipated to be toxic* given the existing safety data. Consequently, their inclusion with contaminants known to cause adverse effects in the diet seems unusual and perhaps prejudicial.	Adopted in the study plan.	**Y**
27	*Homogeneity analysis* of the prepared pelleted diets was not specified, nor *stability of the diets* under relevant conditions of use and storage during the subchronic study. Both analyses are recommended in the	As regards the issue of stability, the type of rodent diets used in the study is generally known to be stable over the experimental period. For the diets that have been conserved, after the first series of experiments,	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	EFSA Scientific Opinion (p.16).	for the follow-up experiments (for omics analysis of the animals fed maize-based diets), stability will be verified based on analysis of gross composition, fatty acids, amino acids, and vitamins (as indicated in the study plan) in samples taken from these diets after storage and before the follow-up experiments. As regards homogeneity, the issue was considered at a very early stage of the project. Unlike the testing of pure chemicals, for which the concentration of the chemical in different samples taken from different parts of the produced batch could provide an indication of homogeneity, there is no such marker for homogeneity of a crop component (maize). Moreover, all materials used for diet preparation (that consisted of milled maize and the basal mix containing the other dietary ingredients) and which were to become combined in the diet, were thoroughly mixed before processed into pellets.	
	Test and control groups		
28	*Control crops (p.15)*: The use of genetically unrelated reference varieties in GM feeding trials in the context of the EU GMO legislation and risk assessment system is inappropriate. The actual effect of introducing the use of unrelated reference controls lies more in providing an instrument to declare technology-induced significant differences as irrelevant against the background of unspecified 'noise' (stemming from broad, unrelated natural variation) than in setting up a sound methodology for risk assessment under the applicable legal framework. (more in ENSSER email and Meyer and Hilbeck)	Not relevant for this study plan.	**N**
29	Is one *control group* sufficient to build up a *control data base* that can be used to position the results obtained from the GM treated groups?	Not relevant for this study plan.	**N**
30	Housing *males and females in different rooms* (p.18) IS NOT a standard practice in general toxicology studies. It does introduce an *additional source of variation*. Justification for this choice should be provided.	There are divergent opinions. The sexes often are maintained in different rooms in toxicity tests, there is no general rule. Any sex difference will be confounded with differences between rooms. Sex differences as such are not of particular interest, but sex x treatment	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		interactions may be of interest. These can still be calculated. The rooms are not sufficiently large to house all animals in the same room, so room effects will need to be considered in any case. (Taken from chronic study).	
31	There is no *justification* included for use of *multiple animal rooms*, or for housing animals in separate rooms, by gender (p.18).	See comment 30.	
	Test System		
32	The *RCC Han Wistar strain* is the test system vs. the test facility has experience in the use of Wistar rats. Please clarify if the test facility has experience on RCC Han Wistar strain.	Mentioned on p.24.	**N**
33	Statistical calculation of the *group size*: if calculated assuming only 1 variable (GMO), how would additional variation (pesticides, soil, mycotoxins …) affect the statistics?	The power calculations assume two groups are being compared (GMO vs control). If the experiment had more treatment groups then a decision would be needed on the way that a power analysis should be applied. However, the variables mentioned appear more like ones which increase the variation rather than being treatments. If variation is increased, power will go down.	
34	The *sample size* was determined by power analysis, however no information on the *mathematical model* and on the *data* used are provided. In particular it is not clear which are the *quantitative biomarkers* used for the power analysis.	Specified on p. 24.	**Y**
35	It is indicated that the number of animals per group was determined by a power analysis for quantitative biomarkers, with 10 rats/group offering a 90% of chance of detecting a SES of 1.5 standard deviation and 80% power of detecting a SES of 1.5 standard deviation; note that: • it is not clear what are the quantitative biomarkers used for the analysis.	See comment 34.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
36	The proposal to use 16 animals of each sex was based on an effect size of 1 SD. The choice to use 1.3 or 1.5 SD could be justified when compared to the natural variation of the parameters, but it differs from the principle to use 1 SD, that was chosen in the Anses (2011) and EFSA (2011) recommendations to explain the number of 16 animals of each sex by group (p.14) .	This is a pilot study. We have no background data on which to base a power analysis. One of the aims of the study is to provide such information. As metabolomics results a large number of biomarkers it is unlikely that a separate statistical analysis of each biomarker for each sex will be appropriate. Ten animals in each treatment group seem to be a reasonable compromise which should provide adequate data without using too many animals.	**N**
37	Please clarify the results of the *power analysis to determine group sizes*: although the cage is specified as the experimental unit, power analysis indicated that "10 animals" was adequate to detect the stated standardized effect differences. If the power analysis indicated an N of 10, this would be 10 experimental units, not 10 animals; therefore, 20 animals/sex/group would need to be placed on study to satisfy the required group size, since statistical analysis will be conducted on a cage basis (2 animals/cage)(p.14).	See comment 36.	
38	Even if the housing in TECNIPLAST cages (p.14) is usual, there are questions about the neutrality of this material in case of gnawing by the rats (due to the risk of depolymerization after multi-sterilization in autoclave). The same kind of question arises with the plastics in the toys, even if toys are mandatory in the context of animal welfare (CEE 86/609 regulation).	TECHNIPLAST cages are routinely used for the conduct of feeding trials with rodents. Referring to the recommendations provided by the producer "the plastic is highly resistant and can be autoclaved at 134°C". There is and will always be a compromise between the pros and cons regarding the properties of the materials used within science.	**N**
39	Not indicated how the *potential nutritional imbalance* as a result of dietary incorporation of the test item is assessed.	Compositional analysis will be performed in order to guarantee for the nutritional balance of the diets.	**N**
40	The rationale behind the power analysis for qualitative data (i.e. pathology) is debatable: a background of 5% is absolutely arbitrary, with chronic nephropathy, for example, rising well above the 5 % in rats of this age.	These sample sizes are not really adequate for the analysis of qualitative characters unless there are substantial differences among treatment groups, which is unlikely. However, this is a pilot study of the application of metabolomics which provides a quantitative readout of many biomarkers.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Experimental design		
41	This is a *multisite study*, it should therefore be conducted according *OECD 13* (Consensus Document of the Working Group on Good Laboratory Practice - The Application of the OECD Principles of GLP to the Organisation and Management of Multi-Site Studies, 2002). In particular, the following points are be addressed: - a *single final report* under the responsibility of the Study Director is to be produced; clarification on how data will be provided to the Study Director for inclusion in the study is to be provided; - as this multi-site study is conducted in more than one country, the opportunity to *seek the advice of the national GLP compliance monitoring authority* where the site is located is to be considered; - a *Lead Quality Assurance* is to be identified, please specify this in the protocol.	All analysis performed at test sites located outside Slovakia are not in compliance with GLP principles and thus seeking the advice of the respective national GLP compliance monitoring authority where the site is located does not need to be considered. If the to be identified histopathological test site will be located outside Slovakia, the proposed specifications will be amended to the study plan. The Lead Quality Assurance Manager was specified on p. 19. A single final report will be compiled.	**Y**
42	This is a multisite study (5 test sites besides the SMU test facility), therefore (p.7): - The study director must endorse the global conformity of the study to GLP; - The multi-site procedure should be applied, or at least, the sites that don't have a GLP recognition should be clearly identified as such.	The conformity of the study plan to GLP principles was approved by the Head of the QAU of the Test Facility. All test sites with GLP accreditation or ISO compliance are made explicit in the study plan.	**N**
43	Experimental design (p.18): It is important to stress that the *initial weight difference between animals within the same group and within treatments should not be more than ±5%.* If is the ±5% weight difference is extended, in order to be able to detect significant differences between treatments one would need 5% or more, +5% difference between treatments.	Following OECD TG 408 the difference in weight should not exceed ± 20%. Based on the long lasting experiences at SZU it is expected to be less than 20%.	**N**
45	*Randomisation* (p.18): when the first animal in a cage has got its random number, another animal of the same weight (or as near to it as possible) should be selected out and placed in the same cage. The least difference it is between animals in a cage the tighter the results will be for all	This is incorrect if the cage with two rats in it is the experimental unit, as is the case with these experiments. In this case differences between cages (not rats) should be minimised. Random assignment of the two rats should help to achieve this.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	parameters measured.		
47	"To minimize the chance of mistakes being made, cages of the same treatment groups will be clustered in vertically arranged groups"(p.18). This sounds like there will be a lack of proper randomisation which could seriously undermine the principles of good experimental design.	This is a reasonable compromise between the need to minimize possible mistakes and the need to ensure that each cage has, on average, the same environment within the animal room.	
49	The referenced OECD 408 and EFSA guidelines do not include specifications for *sentinel animals* for monitoring animal health during the study.(p.18)	Sentinels are included to serve as a veterinary control and will be fed with the standard diet as provided by the diet manufacturer.	**Y**
50	There is insufficient information in the protocol about how sentinel animals will be treated, what diet they will be fed, or which data will be collected for them in support of health evaluations.(p.18)	See comment 49.	**Y**
51	Method and objectives do not describe the *experimental design* (Mixing diet GM1, GM2+non-GM, non-GM and conventional).(p.19)	We're unclear about the reference of the comment.	**N**
52	*Food ad libitum* (p.19): Growth of the body and, internal organ growth and development depend on the amount of food consumed. It is of crucial importance that all animals should consume the same, or nearly the same amount of food. Rather, animals in all the cages should consume the same amount of food. To achieve this, food intake should be monitored daily. The minimum amount of food consumed by animals in a cage on the first day should determine the amount of food offered for all cages the following day. If the following day all food which was offered were consumed in each cage, then an additional amount of food (usually 2 g per cage) should be offered to all. If this was also consumed by all, some more food should be given, using the same principle. The amount of food offered each day should be the least amount of food consumed by the animals in a cage. Therefore the lowest food intake per cage should determine the amount for the following day.	Following from the OECD standard testing procedure TG 452, measurements of food consumption and food efficiency should be made at least weekly for the first 13 weeks and at least monthly thereafter but not daily. Other approaches cannot be comparatively tested in the course of GRACE.	**N**
53	No clear description how *feed consumption* will be reported and how *food conversion* is calculated.	Specified on p. 29.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
54	The measures to be taken to *avoid feed spillage* should be described.	Specified on p. 29.	**Y**
55	Rodents are sometimes known to dig in, and spill their food. This can lead to erroneous food consumption measurements if it is not observed and noted. To minimize the potential for confounding the study data, please advise if steps will be taken to determine if *food spillage* will be documented, and to what extent it will be estimated to avoid the possibility that spillage will be recorded as consumption.(p.19)	Specified on p. 29.	**Y**
56	*Water consumption* is not measured.	As the test substance is not administered via the drinking water, water consumption does not need to be measured following TG 408.	**N**
	Periodic Observations		
57	There are no procedures described in the protocol to account for *incompatible animal pairs* to address fighting or other types of cage aggression (food hoarding, etc.)(p.14 and 18).	SOP number was specified on p.30 (SOP No. TOX/V 004) describing the procedure for the separation of incompatible animal pairs.	
61	Please consider adding *body weight* collection at randomization and using this as part of the randomization criteria to minimize the possibility of widely different body weight means and ranges in the treatment groups (p.20).	These young animals are quite uniform in body weight. It would be counter-productive to match the animals within a cage for body weight when the cage is the experimental unit as this would increase the inter-cage variance and reduce the power of the experiment. It is also unlikely that the expression of the metabolomic markers will depend on body weight, but this can be tested later using covariance analysis.	**N**
62	*Ophtalmoscopic examination*: no procedural details are presented9-1-2014.	Specified on p.30.	**Y**
	Procedures for sample collection		
63	"Sample collection is detailed in attachment 1"(p.21): these details cannot be found there. But from the time schedule depicted in attachment 1 it can be deduced that the *administration will start on a staggered basis* not beginning the treatment of the animals in the different groups at the same time. This might be reasonable from a practical point of view but differs considerably from state-of-the-art toxicity study designs, bringing in *possible time-dependent variations*, and lowers the statistical significance and power of the results. In this	Timing of sample collection for metabolomics/immunological analysis was specified in attachment 1B and possible day-to-day variations will be taken into account.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	case, someone has to be very careful to stick to a very standardized and scrupulously observed study protocol not to misinterpret day-to-day variations.		
64	It is not clearly stated that *blood sample collection* will be performed on all the animals of the study (10 rats/sex/group, group=6) this should be clarified.	Specified on p.32.	**Y**
66	It is highly unusual to *repeatedly sample blood* from the main study group on a 90-day toxicity study as appears to be indicated in this protocol (p.21). If repeated sampling is necessary, satellite groups are often used to collect those samples with the main study groups not being sampled until the end of the study for clinical pathology evaluations. Satellite groups often consist of fewer animals and they are typically used for a single endpoint (e.g. blood sample collection for toxicokinetic analyses at one or more study intervals). Repeated sampling from the same individual animals can deplete blood volume, and these effects would be detectable in the haematology analysis at the end of the study and thereby confound its interpretation. It is recommended to use a satellite group of animals for this sample collection and analysis.	As already mentioned earlier, this study should be seen as a pilot study in order to determine the value of implementing additional metabolic endpoints in the experimental design of 90-day studies used for the risk assessment of GM food/feed. Consequently, the sampling procedures have to be adapted accordingly and deviations from TG 408 and EFSA "Guidance on conducting repeated-dose 90-day oral toxicity study in rodents on whole food/feed" are indispensable. Furthermore, due to spatial limitations at the SZU (the longitudinal metabolomics study and the chronic toxicity study are running in parallel) an inclusion of additional satellite groups is not feasible. But as all animals from all dosage groups experience the same treatment throughout the study, possible treatment effects on specific endpoint are expected to affect the whole test population equally. In order to determine a possible consequence on the measured endpoints, the order of magnitude of the different responses can be compared to the outcomes of the 2 previous 90 day trials as the same feed material (GM plus near-isogenic variety) was fed to the test animals.	**N**
67	Storage of samples is not described in the document.	Specified on p. 39	**Y**
68	*Haematology* (p.22): OECD TG 408 states that "a measure of blood clotting time/potential" should be included in the haematology analysis. Please add this endpoint for compliance with the guidance.	These parameters cannot be measured at SZU and are thus not included in the study plan.	**N**
69	No measurement of *blood coagulation parameters.*	See comment 68.	**N**
70	*Urine collection*: It is not indicated if animals in metabolic cages are feed deprived.	Specified on p. 32.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Omics/Immunology		
71	Because of the relative friability of samples intended for molecular analysis, it seems most prudent to have the "omics" samples collected first and frozen as quickly as possible after collection of organ weights to preserve molecular integrity. These functions are included in the study protocol as being conducted by "person 11", who seems to be a bit far down the processing line to ensure adequate preservation of tissues.(p.23)	Adopted on p. 33.	**Y**
72	It is noted that *blood sampling* is performed after the animals have been in the metabolic cages for urine collection. It is well recognised that the permanence in metabolic cages is a *stressful condition* for the animals (individual housing, different environment, starved condition). Furthermore, blood sampling from the tail implies warming of the animal or of the tail, providing further stress to the animals. It is therefore questioned if these samples are *suitable* for immunological and metabolomics evaluations, besides considerations on the *welfare of the animals*.	As all test animals are treated the same way throughout the pilot study, possible variations caused by the sampling procedure will consequently affect all test animals equally.	**N**
73	*Time points* for metabolomics blood samples versus dosing time are unclear.	Specified on p. 32.	**Y**
74	Has the toxicology laboratory sufficient *historical control data* available to position, e.g. the immunological parameters?	No historical data are available for diets with whole food/feed.	**N**
75	*Metabolite identification* and *analytical methods* are not described in sufficient detail.	Some details concerning 1) the analytical processing of samples (extraction, dilution, etc.), 2) the randomized selection of animals, 3) the metabolomic fingerprinting procedures used to get data, 4) the statistical procedures used to mine raw data and 5) the identification steps used to characterize structure of some putative biomarkers have been added in the study plan.	**Y**
76	Sampling, preservation and transportation methods for *leucocyte phenotyping* are unclear.	Additional detail was provided on p. 38.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
77	*Phagocytic activity and respiratory burst of leukocytes* (p.24): This appears to be the first time this analysis is mentioned and it is not a standard endpoint in studies compliant with OECD TG 408. Please clarify why this parameter was chosen for inclusion in this study. Also please clarify if SPP/IMU/M008 refers to an SOP or some other internal document.	SPP/IMU/M008 refers to an SOP. Specified in the protocol as SOP: SPP/IMU/M008.	**Y**
78	*Phagocytic activity and respiratory burst of leukocytes*: *procedural details* are not provided, it is not possible to comment on this.	Explained in more detail p. 35.	**Y**
79	*Humoral immunological studies of plasma* (p.24): Studies claiming immunoreactivity to GMO crops have been published, but have been widely debunked. Consequently, the reason for including such analyses in these studies is not apparent.	These studies are of special importance when considering the high prevalence of food allergies, and they correspond to a societal demand to guarantee that the genetic modification will not increase/create allergenicity risk. This has been highlighted by different EFSA reports (see Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed; EFSA Panel on Genetically Modified Organisms (GMO Panel); EFSA Journal 2010; 8(7):1700). The chosen parameter will allow assessing the Cry1Ab allergenicity/immunogenicity, and the effect of the genetic modification on the intrinsic allergenicity/immunogenicity of maize. It is all the more relevant that the way of administration, i.e. the whole food via the oral route, will reproduce the potential route of exposures in human. Moreover, within GRACE, these studies will be combined with IgE binding studies using sera from maize-allergic patients and digestibility tests, both performed with GM vs conventional maize. This combined approach will then give relevant and useful information to strengthen the allergenic risk assessment, following EFSA guidance.	**N**
80	Test Site 5 should provide a *validation report* for the immunospecific assays (p.24). Such assays are known to be highly variable, and their inclusion in studies intended for regulatory submission has been limited by this variability. Presuming that Test Site 5 can develop a valid assay (a remarkable achievement given the previously mentioned difficulties), it	Immunospecific assays have been validated using commercial standard isotypes and lab-made naive vs Cry1Ab and maize - rat antisera. This will be provided with the first sub-chronic studies.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	might be adoptable in the wider scientific community if the development and validation processes are completely transparent.		
81	*Metabolomic studies of plasma and urine samples for INRA* (p.24): No mention is made of whether or not these studies will be GLP compliant or what procedures will be followed in the Protocol. Also, the absence of analytical specifics for these samples and the endpoints under investigation is highly unusual. Such details are provided in protocols or study plans to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs.	Some details concerning 1) the analytical processing of samples (extraction, dilution, etc.), 2) the randomized selection of animals, 3) the analytical methods to fingerprint samples, 4) the statistical methods used to model the metabolomic information and select some putative biomarkers, and 5) to tentatively identify the chemical structure of these biomarkers have been added to the study plan. These procedures are routinely used in labs of the Test Site 5. Endpoints of the study have better described in the objectives part of the Study Plan. However, only the animal part of the study leading to production of samples is submitted to full GLP regulation, not complementary studies performed outside the Test Facility.	**Y**
82	*Metabolomic studies of plasma and urine samples* for INRA (p.24) and *Metabolomic studies of tissue samples* (p.26): There is a need for method harmonization. Additionally, the study includes only a metabolomic approach, without transcriptomic or proteomic studies. Do we have metabolomic data obtained on kidney and liver of Wistar rat with known hepatotoxic or nephrotoxic substances that could be used as references? Indeed, the main limitations of the metabolomic approaches are the lack of reproducibility of the methods, insufficient informatics tools for the data treatment and a lack of knowledge on the metabolomic signatures. The need of standardization of the methods is clearly explained in a paper of Robertson *et al.* (Metabolomics in Toxicology: Preclinical and Clinical Applications Toxicol Sci., 2011): "*CURRENT NEEDS/FUTURE DIRECTIONS* *Clearly, metabolomics has grown and changed over the past several years evolving into a technology with wide application across the spectrum of biological sciences. Although toxicology is no longer at the*	The main problem concerning omic studies and particularly metabolomics is mostly the weakness of statistical modelling and mining of intrinsic information contained in raw data, even classical tools are existing and are easily accessible. We can use external samples coming from the same rat line (Wistar) to get additional confidence parameters only for data production and data processing but this will be of minor interest. What is more important is really the building of the experimental design to address specific effect of any controlled factor (gender, cultivar, GM, dose of GM), and also possibly some main interactions, with the necessity to highlight possibility to get a confounding factor situation. It is the main difficulty concerning regulatory assessment of feed or food with explorative methods coming strictly from the drug world. Concerning the need of method harmonization, it would be urgent to get alternative means to phenotype metabolome of animals to compare them and to get lastly a consensual harmonization. At present time, it is not the case. Then, contrary to the main feeling that there is a lack of knowledge signatures,	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	center of the metabolomics universe, the technology has demonstrated tremendous potential for toxicological applications. Current needs for the technology include some sort of standardization in MS approaches so that different laboratories running different instrumentation can still generate comparable data on the same set of samples. Currently, this is not the case as different MS configurations and separation strategies can give very different biochemical pictures of a biofluid or tissue extract. The current atmosphere has a bit of a "Wild West" feel to it with most investigators using their own approaches that tend to vary from laboratory to laboratory. Although this is probably a necessary stage of development of the technology, as no single approach or strategy can yet claim superiority, it is inhibiting to those wishing to pursue regulatory applications of the technology. Although some attempts have been made to standardize MS approaches and their reporting (Jones et al., 2007; Kanani et al., 2008; Sumner et al., 2007; Want et al., 2010b), we still have a long way to go. Reviews by Gomase (Gomase et al., 2008) and Go (2010) provide nice summaries of current database resources for metabolomic analysis. Although these databases are increasingly valuable, they are dependent on the approaches used to generate them so that unless identical methods are employed, they can become difficult to utilize."	we know that with an in-depth data mining, it is possible to access to valuable physiological (metabolic) signatures given the fact that the experimental design is well built (well-balanced, sufficient number of individuals in unitary cells, possibly a longitudinal follow-up, high degree of symmetry in the building of the experimental design itself, etc.). To conclude, we are aware of severe limitations coming from a too much abrupt transposition of guidelines that are well-adapted to xenobiotics assessment to the feed or food assessment domain for which it is easy to notice important methodological difficulties (dose effect for example which is the main one).	
83	*Cellular immunological studies* (p.26): This is not a standard endpoint in studies compliant with OECD TG 408. Please clarify why this parameter was chosen for inclusion in this study.	Cellular immunological studies were included to assess the possible immunomodulatory potential of the tested diet. Originally, the methods were used in human clinical immunology to examine the changes in cellular immunity for diagnosis of immune diseases. Later on, the methods were adapted for the assessment of immunomodulation and immunosafety in animal studies. They are applied in laboratory for several years to monitor the immune system in Wistar rats.	
84	Cellular immunological studies (p.26): Additional detail must be provided on analytical specifics for these samples and the endpoints under	Explained in more detail on p. 38.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	investigation in order to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs.		
85	*Cellular immunological studies*: *In vitro production of cytokine*: actually it is an ex vivo assay, not an in vivo test. It is not specified which is the *ELISA method* (e.g. ELISPOT) followed in the determination of cytokines from splenocytes, it is not possible to fully comment on this. It is not indicated if the methods have been *validated*. It is not clear why *no specific challenge/s* have been included in the spleen test.	Further detail is given on p. 38.	**Y**
86	*Phenotypic analysis of leucocytes* (p.26): This is not a standard endpoint in studies compliant with OECD TG 408. Please clarify why this parameter was chosen for inclusion in this study.	Methods are included for the assessment of immunotoxicity as recommended by Burleson et al. (1995): Methods in Immunotoxicology. Panel of CD markers were selected according to the recommendation of the Guidance for Industry: Immunotoxicology Evaluation (2002). http://www.fda.gov/cder/guidance/index.htm	**N**
87	*Phenotypic analysis of leucocytes* (p.26): Additional detail must be provided on analytical specifics for these samples and the endpoints under investigation in order to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs.	Further detail is given on p. 38.	**Y**
88	Mesenteric lymph-node cytometry: no information on the flow cytometry method is provided (e.g. tissue preparation), it is not possible to comment on this. Furthermore it is not indicated if the method has been validated. No rationale on the choice of cytokines is provided of (e.g. CD5R and CD161).	Mesenteric lymph node cytometry will follow SOP: ŠPP/IMU/M007. Shortly, lymph nodes will be placed into the RPMI medium with heparine. Lymph nodes will be homogenized in medium using Eppendorf tubes, centrifuged, washed with medium, standardized, labelled with monoclonal antibodies and analysed using flow cytometry. CD45R and CD161 are leukocyte antigens, not cytokines. CD45R and CD161 antigens were selected for analysis of rat B-cells and	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		NK-cells since those cells play a very important role in the immune defence. The battery of CD markers were selected according to the recommendation of the Guidance for Industry: Immunotoxicology Evaluation (2002) (http://www.fda.gov/cder/guidance/index.htm). Production of cytokines (TNF-alpha, IL-1beta) will follow SOP: ŠPP/IMU/M009. Rationale for selection of cytokines: TNF-alpha and IL-1beta are important pro-inflammatory cytokines. TNF-alpha is involved in systemic inflammation and stimulates the acute phase reaction. It is produced by activated macrophages, and many other cell types such as CD4+ lymphocytes, NK cells. IL-1beta is an important mediator of the inflammatory response involved in cellular activities, including cell proliferation, differentiation, and apoptosis.	
89	Metabolomics studies of tissue samples: no information on the endpoints selected for evaluation is provided.	Objectives have been detailed concerning aim of the metabolomic study. As untargeted study, it is not possible to reach some quantitative characterization of some prior selected biomarkers when they are unknown before such a study is performed. What is more important in this study is the way to try to get a global metabolic information structuring based on a longitudinal follow-up of every animal. This will be a way to get sound comparisons between groups of treated animals from their ability to adjust their metabolism to feed prepared from GMO-maize or not and considering a possible dose effect in the metabolic deviation whatever the cultivar studied.	**N**
90	*Metabolomic studies of tissue samples* (p.26): This is not a standard endpoint in studies compliant with OECD TG 408. Please clarify why this parameter was chosen for inclusion in this study.	It is an additional study for which the explorative side is high and would bring thanks to an adapted data mining procedure some interesting results in the respective weight of controlled factors in the explanation of the total variance.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
91	*Metabolomic studies of tissue samples* (p.26): Additional detail must be provided on analytical specifics for these samples and the endpoints under investigation in order to determine if the researchers followed the path set by the Study Director. Without them it is impossible to determine if the Protocol was followed and if the Study Director truly was the "the single point of study control" as required by OECD GLPs.	Further detail is given on p. 38/39.	**Y**
92	The *Cellular immunological studies* section of the protocol (p.26) indicates that a portion of the spleen, the mesenteric lymph nodes, thymus and bone marrow will be collected. However, these organs/tissues are also intended for microscopic examination. There are no specifications for the amount or anatomic locations of tissue to be collected and allocated for each of the intended analyses (histopathology and molecular). The site of bone marrow collection is not specified, and there could be potential differences depending on whether central or peripheral marrow is sampled and evaluated. Storage conditions prior to analysis of samples placed into PBS buffer are not specified in the protocol.	Specified on p.38.	**Y**
93	*"omics" samples* to be collected from the multiple animal organs/tissues (p.26-27): The protocol specifies the anatomic location of the sites for collection of liver and kidney tissue. However, the protocol also indicates that all gross lesions will be evaluated microscopically. It is not uncommon in long-term (subchronic and/or chronic) toxicity studies for animals to have gross lesions of the liver and/or kidneys. This protocol does not include specifications to address the likely potential that a gross lesion may be present at a location within a tissue that is scheduled to be sampled for "omics". Is it more important to understand the morphology of a lesion or the molecular profile of tissues containing a lesion? How would molecular analysis of lesion tissue potentially impact the overall results and interpretation of the "omics" evaluations?	We have no doubt that such "abnormal" observations, even arising by chance, will be duly documented. In the molecular content of such a tissue is evidently modified in connection with the gross observation obtained previously, we would obtain an atypical metabolic fingerprint. Therefore, in multivariate analyses such as principal component analysis, this individual should be projected in the factorial map as an outlier. Indeed, when we will obtain such factorial projections, even in absence of coding key to get assignment of individuals to any treatment group, we will question colleagues for any living animal observation or macroscopic tissue observation that would be indicative of a higher probabilistic outlier status of such animals. Clearly, this modus operandi will be in line with GLP recommendations that govern ongoing of such	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	Similarly, lesions of the spleen and/or thymus could occur, and it needs to be determined a priori whether these lesions should be evaluated microscopically or collected for molecular analyses.	studies. Concerning spleen and thymus, no metabolomic observation has been foreseen. Only a gross examination and an histological one will be performed as recommended in OECD guidelines.	
	Pathology		
94	*Gross necropsy* (p.25): Since food first interacts with the gut, it is of crucial importance and therefore suggested that the wet weight of the stomach, the small and large intestines should also be measured. Preferable the duodenum, jejunum, ileum, caecum and colon weights should be recorded separately. Before weighing it is necessary to wash out the different parts of the gut with ice cold distilled water.	One major aim of the GRACE project is to provide the results of well conducted feeding studies with whole food/feed derived from GM plants, carried out according to relevant recommendations, i.e. OECD and EFSA guidance. Therefore the focus is on the analysis of reliable, internationally accepted parameters (as recommended by OECD TG 408), namely the histopathological analysis of gastrointestinal tissues. It is noted that the general lack of information on the normal variation of the stomach and intestinal weights would make it difficult if not impossible to evaluate any differences with regard to their toxicological relevance.	**N**
95	*Tissue specimens* (p.25): In several cases differences were found in gut weight and morphology between GM and isogenic line containing food fed animals. Therefore, histological examination should be performed on the gut tissue from different parts of the rat gut. The recommended sections to be taken for histology are the following: Starting from the pylorus, the 3-5 cm part is the tissue section for the duodenum, and the part from 23-25cm is the tissue section for the jejunum. The part of 3-5 cm from the caeco-ileal junction is the section for the terminal ileum, and the 23-25cm part from the caeco-ileal junction is the section for the ileum. For determining tissue weight, the weights of the 0-3 cm piece (as duodenum) and that of the 5-23 cm piece from the pylorus (as the jejunum), as well as the weights of the 0-3 cm piece from the caeco-ileal junction (as the terminal ileum) and the 5-23 cm piece from the caeco-ileal junction (as the ileum), plus the weight of the remaining part	The methodological details are made explicit in the respective SOP and will not be explained in detail in the study plan.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	between the jejunum and the ileum should be added together to get the weight of the small bowel (preferable corrected for the missing parts used for the histological examination). Since the water content of the bowel tissue might differ as a result of washing, it would be nice to freeze dry the intestinal samples and weigh them again to get the dry weight. This would give a better indication of existing differences, if any, between treatments. The small intestine should be cut into the following sections: 0-5cm duodenum /5-25cm jejunum /remaining part / 23-5cm ileum/5-0 cm terminal ileum 0-3 cm duodenal tissue/ 5-23 cm jejunal tissue /remnant/23-5cm ileal tissue of s.i.tissue /3-0cm terminal ileal tissue 3-5 cm histology for the duodenum /23-25 cm histology for the jejunum/25-23cm histology for the ileum /5-3 cm histology for t.i.		
96	*Lungs* (p.25) are not typically weighed as part of studies conducted in accordance with OECD TG 408. Furthermore, lungs are typically inflated with fixative promptly after harvest to ensure optimal tissue preservation for the histopathological analysis. This could add significant variability to the organ weight collection and provides another reason for not collecting this weight.	Lungs are routinely weighed at SZU before inflating them with the fixative. No negative effect on tissue preservation was observed in the past.	**N**
97	*Tissue specimens* include (p.25): Please indicate the three levels to be collected per OECD TG 408.	Comment is not clear.	**N**
98	*Tissue specimens* include (p.25): accessory sex organs are listed in OECD TG 408, but not included in this list. For the sake of compliance they should also be collected and evaluated.	Included.	**Y**
99	If the *mesenteric lymph nodes* were to be collected, it would be useful to look for the transgenic DNA and proteins in it, since the mesenteric lymph nodes are the most likely places to find the rDNA and/or transgenic protein if it is taken up by the body.(p.25)	This will not be feasible due to a lack of material.	**N**
100	The *procedure* followed for *gross pathology recording* (macroscopic observation) is not indicated.	Specified on p.36.	**N**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
101	A *best practice* in toxicological settings is to have a *toxicological pathologist supervising* the necropsy session/s, to guarantee consistency among macroscopic observations. This is not indicated.	Specified on p. 36.	**Y**
102	No information is provided on the *approach and software system* (if any) used to record organ weights and for macroscopic observations.	Specified on p. 36.	**Y**
103	No information is provided if organs are weighed separately or paired.(p.25)	Specified on p. 37.	**Y**
104	*Formalin* is not an appropriate fixative for eyes and male reproductive tissues. *Bouin's or Davidson's solutions* should be considered to preserve these tissues for microscopic examination (p.25).	Adopted on p. 37.	**Y**
105	*Histopathology* (p.26): Peer reviewed journal articles indicate that *histopathologists should not be blinded* during the histopathological evaluation, and these references have been provided to GRACE. Please advise why this best practice is being ignored.	In order to reduce bias and increase confidence it was decided to unblind the study after the histopathological evaluation. Blinding will certainly reduce bias and thus increase the validity of the results.	**N**
106	*Histopathology* (p.26): a *peer review* is an excellent way to determine if one or more pathologists agree with the opinions of the first pathologist to read the study. It allows the group of pathologists to arrive at a consensus opinion and limits the influence of any one individual in the diagnostic process. Please advise if it will be included.	Peer review pathologist was specified on p.20.	**Y**
107	*Histopathology*: *Peer review* is a best practice highly recommended. The peer reviewer pathologist should be indicated in the protocol.	Peer review pathologist was specified on p.20.	**Y**
108	*Histopathology*: *Trimming procedures* vary from laboratory to laboratory, differences should be evaluated carefully by a professional pathologist.	Specified on p. 36.	**Y**
	Data evaluation and statistics		
109	The study design does not support a *combined-gender statistical analysis*. Males and females are not of the same age at receipt (staggered start by gender); females will acclimate to the testing facility for one week longer than the males; males and females do not undergo evaluations or sample collections at the same time. Further, there is no	Agreed. The analysis will have to be done separately for each sex. However, with the large number of biomarkers detected using metabolomics a multivariate statistical analysis such as principle components, discriminant function or a cluster analysis will need to be used. Determining the best way to analyse the results is part of the	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	accounting for potential animal room bias introduced by housing the genders in separate facility rooms. The use of multiple animal rooms could introduce variability to the study that does not appear to be accounted for in the statistical analysis section of the protocol (p.18 and 28).	project, along with developing suitable logistical methods.	
110	A *one-way analysis with planned or post-hoc comparisons* will be used (p.28). It is not clear why a choice of planned or post hoc comparisons is proposed. If this choice is to be retained then clear guidance needs to be provided on when the comparison should be planned and when conducted post hoc.	A multivariate analysis will be needed as noted in comment 109.	**Y**
112	*Statistical analysis procedures* to be used influence study design constraints, and should be *specified prior to conduct the study* (p.28). This is especially critical for full transparency of the intent of the study, and is consistent with current EFSA guidance. "Planned or post-hoc comparisons" and determining which analyses are appropriate "decided on a case-by-case basis" imply that study data will be collected and examined prior to selective application of unspecified statistical tests. The protocol must provide sufficient detail regarding the planned analyses such that the study statistician receives clear guidance; otherwise, bias (intentional or unintentional) may be introduced at this stage. This deviation from EFSA guidance should also be described in the Regulatory Test Guidance section of the protocol.	This is a pilot study. These methods have not previously been used for toxicity testing of GM crops. One of the objectives are to determine the best way of analysing the results.	**N**
113	No indication whether statistical analysis will be conducted on individual organ weights or paired organ weights.	Data will be collected on individual organs. These will be analysed separately.	**Y**
114	If the compositional analysis would show relevant differences in the presence of toxicants or antinutritional factors, how can this be taken into account such that any health *effects* or other differences can be attributed to the correct *cause*?	Various considerations come into play: Which effects are observed and is it plausible that these are linked to the differences in composition based on the magnitude of the difference and the known properties of the particular compound being different (for example, at which dietary level can toxicity be expected to occur or not). Maize kernels do not contain intrinsic toxicants, whereas various antinutrients, particularly	N

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
		phytic acid, can occur. Other sources of toxicants and anitnutrients could be other ingredients in the diet (mainly wheat and soybean), which are kept to the same inclusion levels across diets. Besides intrinsic plant compounds, contaminations, such as mycotoxins formed by moulds that might infect the plants could theoretically play up. In another European research project, it was indeed found that contamination of GM maize MON810 with the DON mycotoxin accounted for the effects in salmon fed diets containing this maize, so the effects indeed could be related to presence of DON and not particularly linked to the presence of GM. Any observed differences will be taken into account when analysing the results of the trials.	
115	Given that the OECD TG is designed for chemicals, not complex mixtures, how to make sure that any differences that may be measured stem from the intended variation in the diet?	Animal feeding trials with whole food/feed are analysed for statistical significance and biological (pathological) relevance. The diets administered ideally differ (only) in their targeted test plant material and the comparator used, while the diets are nutritionally balanced. Moreover, the composition of diets and plant material is analysed in detail. Nevertheless, due to the complexity of the test material feeding trials may not provide full evidence in terms of distinct causes and effects. Furtheron targeted investigations may be necessary to substantiate the GMP effects. GRACE is supposed to highlight such critical issues in the comparative analysis of approaches.	**N**
116	There is no calibration of the metabolomic techniques described in this type of experimentation. Differences will be found. It is not possible to link these differences with health parameters. What if no statistical significant differences in the general health parameters are found (see COGEM report)?	In an untargeted metabolomic study, there is no need to use any calibration of metabolomic techniques. However, we are aware of the fact that production of metabolomic data is done in a quantitative way.	**N**
117	The study report should include all of the raw data, not “relevant” raw data. Who is to decide, in interpretation of study data, what is relevant and what is not? Providing the entire raw data set will ensure adequate	"Relevant" was deleted.	**Y**

Q/A	Comment/suggestion	Brief explanation/argumentation	Adopted in study plan (Y/N)
	transparency and allow the public to evaluate the results in an unbiased manner (p.30).		
118	Attachment 1 does not include the extraneous blood and urine sample collection dates for the metabolomics analysis (p.31).	Included on p. 44-46.	**Y**

Part II: List of open questions

Chronic toxicity (1 year) study

Q/A	Comment/suggestion
	Experimental design
56	Specify the *selection criteria for placing animals on study* (or for excluding animals from placement on study)(p.20). Will *animal substitutions* be allowed after assignment to group but prior to the first day of feeding? Under what conditions would substitutions be required? If animals are not assigned to study because they have exclusionary criteria, would they still be useful as sentinels for health monitoring, or would they be otherwise disposed?
55	Cage numbers and the *random diet assignment* (p.20): When the first animal in a cage has got its random number, another animal of the same weight (or as near to it as possible) should be selected out without randomization and placed in the same cage. The reasoning is that the lesser the difference between animals in a cage the tighter the results will be for all parameters measured.
66	A *completely randomised design (p.20)* is specified, which is fine if the initial *body weights* are all very similar. However, if this is not the case then it would make more sense to adopt a randomised complete block design by stratifying according to initial body weight.
67	A *completely random design* (p.20) makes it possible that treatment groups will have vastly different body weight means if all the heavier animals wind up one dose group. This can be avoided by using *body weight stratification* to assign animals to the treatment groups. In this case, weigh all of the animals on the day of randomization, and then randomly assign one of the 4 lightest animals of each gender to each dose group. This would be repeated with cohorts of 4 animals groups by body weight, until the groups were completely populated. If a completely random design is necessary then please indicate whether body weight will be considered a blocking factor in the statistical analysis.
	Periodic Observations
68	*Periodical Health Status Observation (p.22)*: There is no mention of the *order of observations* – observations need to be made in random order so as to avoid confounding treatment effects with temporal effects and/or other potential sources of bias.
69	Is the functional assessment derived from FOB or Irwin tests?
	Procedures for sample collection
79	*Storage of samples* is not described in the document.
93	In contrast to the specifications for the liver and kidney, this protocol (p.30) also indicates that sections of gut, lymph nodes and spleen, and that the sections should be of a pre-specified size and mass; however the *mass is not indicated* in the protocol. Additionally, although animals are fasted prior to necropsy, the gut typically has contents. This protocol does not specify whether the *intestinal contents* should be removed prior to collection, weighing and preservation of gut samples. Gut contents are typically removed prior to fixation of samples for histologic processing.

Subchronic toxicology and longitudinal metabolomics study

Q/A	Comment/suggestion
	Experimental design
44	Specify the *selection criteria for placing animals on study* (or for excluding animals from placement on study)(p.18). Will animal substitutions be allowed after assignment to group but prior to the first day of feeding? Under what conditions would substitutions be required? If animals are not assigned to study because they have exclusionary criteria, would they still be useful as sentinels for health monitoring, or would they be otherwise disposed?
46	A *completely randomised design* is specified, which is fine if the initial *body weights* are all very similar. However, if this is not the case then it would make more sense to adopt a randomised complete block design by stratifying according to initial body weight (p.18).
48	A *completely random design* makes it possible that treatment groups will have vastly different body weight means if all the heavier animals wind up one dose group. This can be avoided by using *body weight stratification* to assign animals to the treatment groups. In this case, weigh all of the animals on the day of randomization, and then randomly assign one of the 6 lightest animals of each gender to each dose group. This would be repeated with cohorts of 6 animals groups by body weight, until the groups were completely populated. If a completely random design is necessary then please indicate whether body weight will be considered a blocking factor in the statistical analysis (p.18).
	Periodic Observations
58	There is no mention of the *order of observation* – observations need to be made in random order so as to avoid confounding treatment effects with temporal effects and/or other potential sources of bias(p.20).
59	While not required by OECD TG 408, it is advisable to include a *baseline functional assessment* before administration of the test substance to the test group. Having baseline data on all animals allows the investigator to see if differences existed between groups before treatment and thus determine if differences between groups are truly test substance related(p.20).
60	Will the *functional assessment* be derived from FOB or Irwin tests?
	Procedures for sample collection
65	*Sample collection* (p.21): A blocking structure is described on p.23 but there is still a need to implement randomisation of samples within any given day.
	Data evaluation and statistics
111	There is no mention of taking account of the blocking structure that is described on p.23.

Annex: List of comment providers

The following organizations provided written comments:

AGES (Österreichische Agentur für Gesundheit und Ernährungssicherheit), Austria.
ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail, France.
BelTox (Belgian Society of Toxicology and Ecotoxicology), Belgium.
EFSA (European Food Safety Authority), Italy.
ENSSER (European Network of Scientists for Social and Environmental Responsibility), Germany.
EuropaBio, Belgium.
Hungarian Ministry of Rural Development, Hungary.
VIB (Vlaams Instituut voor Biotechnologie), Belgium.