Comments by the Federal Office of Consumer Protection und Food Safety

The development and establishment of agreed quality criteria are very welcome and we think that GRACE and G-TwYST are good projects to do so. In the main, we support the quality criteria presented in the publication under consideration (Schmidt *et al.*, 2016). However, there are some aspects we would like to comment on (see comments below).

At large, it is not always quite clear what reference point the publication is exactly focused on. The headline refers to *"mandatory rat and mouse feeding trials"*. In the text it is written that *"the present letter focuses on feeding trials falling under the EFSA scenario 2"* (no relevant changes and/or specific hazards identified). Elsewhere it is said that *"the described set of proposed key quality criteria applies to 90-day feeding trials as well as to long term feeding trials"*. Finally, in the concluding remarks it is stated that *"only in case a trigger is available from the initial molecular, compositional, phenotypic and/or agronomic analyses and therefore the rationale of the study prior to testing is formulated in form of hypotheses regarding specific endpoints, feeding trials with whole food/feed may provide an added scientific value for the risk assessment of GM crops".* We would like to point out that only 90-day feeding trials are currently mandatory according to the Implementing Regulation No. 503/2013/EC. The terms of EFSA scenario 1 and 2 are also merely used with regard to 90-day feeding trials. In contrast, long-term animal feeding trials are currently not a mandatory requirement for risk assessment of GM food/feed. Concerning this matter, the present publication fails to give examples on what might trigger a scientifically justified necessity for a long-term animal feeding trial.

General comments on animal feeding trials with whole food/feed (as a tool for the risk assessment (RA) of GMOs)

- To begin with, it should be noted that the gain of knowledge from animal feeding trials with whole food/feed as a tool for the RA of GMOs is of very special but limited informative value due to inherent system restrictions.

  The OECD guidelines 408/451/452/453 were primarily developed to test single substances with low molecular weight and are very well suited to identify potential toxic and/or carcinogenic effects of defined single substances. In contrast, food and feed consist of a plurality of individual substances and therefore have a complex composition. Single substances of low molecular weight can be administered in very high doses achieving a very large margin of safety. In contrast, due to its nature (large volume resulting in a limitation of administration by reaching saturation), whole food/feed can - compared to single sub-

stances - only be administered in a lower multiple of the expected human exposure. Moreover, the amount of the administrable dose of whole plant material is limited by the fact that an unbalanced composition of the feed must be avoided in order to prevent secondary effects.

- Ethical aspects of animal trials and the 3R principle (refine, reduce, replace) should be kept in mind (see Directive 2010/63/EU on the protection of animals used for scientific purposes). (Socio-)political arguments are insufficient to justify the performance of animal experiments (particularly with regard to long-term animal feeding trials) if an appropriate scientific rationale is missing.

General comments on the experimental design of animal feeding trials with whole food/feed

- Especially with regard to long-term animal feeding trials, an appropriate scientific rationale is an indispensable prerequisite and, therefore, should be a quality criterion (at the beginning of each scientific study a clear question/hypothesis is needed to conceive the design of the study accordingly).
- One should be aware that there is no such thing like “the one and only perfect design” to carry out an animal feeding trial. There are trade-offs between different variants. Particularly with regard to a two-year animal feeding trial, it is very difficult to perform an experiment without any methodological problems.
- A certain level of standardization is very useful and, therefore, agreed quality criteria are welcomed from the perspective of a risk assessor. However, it should be noted that especially long-term animal feeding trials will be (and should only be) conducted on a case-by-case basis. The specific problem formulation may be different depending on the individual case. It is currently not advisable to set all details of the design in stone. Some kind of a flexibility of the case-specific design should be possible, but a justification of the selected criteria should be given in any case. Schmidt *et al.* (2016) state that *“the present letter focuses on feeding trials falling under EFSA scenario 2”* (no specific hypothesis to be tested). In this regard, we would like to point out that EFSA addresses a scenario 2 only with respect to 90-day feeding trials, but not in terms of long-term animal feeding trials.
- For each animal trial the following applies: as many animals as necessary and as few animals as possible.

Comments on the proposed criteria

1. The design of the feeding trial is based on internationally recognized test guidelines, but adapted for specific needs of whole food/feed studies and non-targeted testing

- We support that internationally recognized test guidelines are taken as a basis to design the experiments, but adapted for specific needs of whole food/feed studies.
- As the available EFSA recommendations are already adapted for specific needs of whole food/feed studies, any deviation between the EFSA recommendations and the recommendations in the publication under consideration should be scientifically justified.

2. An analysis of the plant materials and diets including, among others, macro- and micronutrients, biological and chemical contaminants as well as the identification and quantification of the event, is performed.

- Generally, we support a comprehensive analysis of the plant materials and diets.
- In any case, test and control material should be tested with regard to the event under consideration. However, is it really necessary to test BOTH the starting plant material AND the diets for the presence of the event at any rate? Moreover, we do not agree that the presence of (all existing) other events as background contamination in the whole food/feed must necessarily be tested (please keep the large number of existing events in mind). In our opinion, small admixtures of seeds with events other than the introduced event are equated with accidental admixtures of seeds from other non-genetically modified varieties and, therefore, does not affect the risk assessment due to their low proportion.
- Reference to and compliance with the MRL (Maximum Residue Levels) Regulation shall be considered.
- For the sake of completeness, it should be noted that the application of complementary herbicides (e.g. glyphosate) to genetically modified herbicide tolerant plants and the effects possibly arising from pesticide residues (and their metabolites) is an independent question that - in case of need - must be considered separately in the study design.
- What about choice and preparation of the diet?
  The product to be consumed may vary between human food and target animal feed, particularly with regard to the degree of processing. Therefore, it should be discussed, which indicators have an effect on the decision whether processed or unprocessed whole product (or even both) should be utilised?

3. The highest level of the plant material that can be incorporated in the animal diets without leading to a nutritional imbalance is tested.

- The recommendations of Schmidt *et al.* (2016) differ in some points from the recommendations provided by EFSA. No distinction of cases is made between 90-day feeding studies on the one hand and long-term animal feeding trials on the other hand. The scientific necessity of the alterations/modifications is not quite clear and should be explained by the authors. E.g., Schmidt *et al.* (2016) state that *"in principle, it is not necessary to use two dose levels"*. With regard to 90-day feeding trials, EFSA (2011) recommends two dose levels (high dose and low dose) when testing whole food/feed. In this context, the high dose level should correspond to the highest level of the whole food/feed that can be incorporated in the animal diets whilst avoiding nutritional imbalances. The low dose level could be half to a quarter of the high dose and should always be above the anticipated human intake. However, according to an explanatory statement of EFSA (2014), in the absence of a test hypothesis (scenario 2) a scientifically justified option is to use only one dose level of the GM test material at the maximum incorporation rate. With regard to long-term animal feeding trials (chronic toxicity and carcinogenicity studies), EFSA (2013) does not recommend any specific dose group and/or dose selection approach, which should be defined on a case-by-case basis. Nevertheless, EFSA recommends two dose levels (high as well as low dose group) as a minimum requirement for long-term animal feeding trials. This point is still worth to be discussed.

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

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

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

- The incorporation rates for the high dose group require further discussion. With regard to 90-day studies, EFSA suggests incorporation rates for several crops. In contrast, Schmidt *et al.* (2016) recommend to perform a preliminary inclusion rate study in either case. In this regard, the authors should explain (differentiated according to the crop under consideration) why they don't accept the incorporation rates given by EFSA (2014) for a) rice, b) maize, c) soybean, d) rapeseed, and d) potatoes. Nothing is said about the possibility of a literature search to get adequate incorporation rate data for a certain crop. Moreover, the authors should give guidance/quality criteria on how to perform an acceptable inclusion rate finding study (e.g., do we really need a 2-year inclusion rate finding study ahead of a

2-year carcinogenicity study in order to test whether a certain inclusion rate of a given crop has long-term effects on the animals?).

- Testing (only) two dose levels can lead to a situation where the effect of the low dose group is already within the range of the stagnation phase. Therefore, selection of a reasonable interval between the high and the low dose level is of importance. Moreover, the criterion for the lowest dose level (= comparable to or higher than the anticipated intake) may vary between human food and target animal feed.

4. A non-GM line with a comparable genetic background is used as a control.

- We agree that a non-GM line with a comparable genetic background should be the first choice of comparator.

  With regard to stacked events we agree on the following exception given by EFSA: *"[…] However, where applicants can demonstrate that a conventional counterpart for the GM plant cannot be made available, then applicants could use as comparators either a negative segregant(s), but only where such segregant is derived from crosses between GM plants containing events which have been risk assessed and which are all stacked in the GM plant under assessment; or any set of GM plants that have all been risk assessed on the basis of experimental data collected according to the principles of EFSA MC and FF risk assessment. This set of GM plants must include between them all of the events stacked in the GM plant under assessment, and no others. Additional comparators may be included if deemed useful to support the risk assessment."*

- The authors should specify what is meant by *"a well-established history of safe use"*? Possibly, the near-isogenic control line isn't a commercially used line.

- Nothing is said about non-GM reference lines/diets (→ natural variability).

5. Specific aspects regarding the choice and housing of the laboratory animals used in the feeding trials are considered.

- The choice of the animal model and the strain (or even substrain) is of great importance (quality criterion) and should be specifically justified.

  Nothing is said about the use of strain-specific data from the scientific literature to get information on the natural variability of the endpoints. In this respect, it should be defined what is accepted as proper historical background data and where to get this from (e.g., is it acceptable to cite the tumour incidence of certain strains from the scientific literature?).

- It should be noted that only certain numbers of animals are manageable within one feeding trial. Too many animals within one experiment increase the probability of error and/or bias due to external factors. Particularly inbred strains are extremely sensitive to environmental effects.
- Animals should be certified as SPF (Specific Pathogen Free) according to FELASA guidelines.
- Based on the individual problem formulation it can be reasonable to test only one sex in particular cases (e.g. when the endpoint to be tested is sex-related and only found in one sex, like cancer of male and female sex organs, respectively). Such approach should be possible (justification needed).

6. Appropriate randomization techniques are applied.

- We agree that appropriate randomization techniques are very important. In this regard, recommendations/criteria on how to perform an adequate randomization would be welcomed.

7. A reliable and appropriate sample collection and processing strategy is implemented.

- We agree that a reliable and appropriate sample collection and processing strategy are very important.
- The probability of error and/or bias due to methodical factors in terms of sample collection and processing strategy should be as low as possible (e.g., same fasting period for all animals, considering the circadian rhythm → do not take samples from the first animal in the morning and from the last animal in the evening).

8. The staff performing the feeding trial and the analysis of the plant materials, diets and animal samples is "blind" with respect to the identity of the diets.

- We agree that the study should be blinded. For the sake of clarity, the authors should explain why the dose groups are unblinded for the histopathological evaluation.

9. Appropriate statistical methods are applied to evaluate the power of the study and to analyze the obtained results.

- We agree that this criterion is very important.

- Based on the total number of animals within one feeding trial using whole food/feed (which is limited by technical reasons) it makes more sense to include fewer groups with a higher number of animals per group instead of more groups with a lower number of animals per group.
- It should be noted that on the one hand higher numbers of animals increase the statistical power. On the other hand only certain numbers of animals are manageable within one feeding trial. Too many animals within one experiment increase the probability of error and/or bias due to external factors.
- In our opinion, a scenario 2 (no relevant changes and/or specific hazards identified) does not justify the performance of a long-term animal feeding trial.
- Schmidt *et al.* (2016) state that *"raw data are made available to third parties to allow for an independent analysis"*. In this regard, the authors should clarify what/who is meant by *"third parties"*. We do not consider that it is always necessary to make all raw data available to the whole public. There are good arguments to do so in projects like GRACE and G-TwYST, but this should not be a standard for any performed feeding trial. Please note that the raw data of animal feeding trials in the frame of a regulatory decision process are available to both EFSA and the risk assessors of the EU Member States.

Concluding remarks

- Schmidt *et al.* (2016) conclude that *"it is important to note that a feeding trial does not automatically provide useful information simply because it meets the nine proposed criteria".* We would like to point out that this conclusion does not fit to cases with an appropriate scientific rationale (scenario 1 given). If there is a clear-cut question/hypothesis and an accordingly well designed study, then the results should provide useful information to answer this question (→ verification or falsification of the underlying hypothesis). On the other hand, it is also possible to perform an adequate study with a useful study outcome without following all nine criteria down to the last detail.
- With regard to 90-day rodent feeding trials, we fully support the authors' conclusion that *"Only in case a trigger is available from the initial molecular, compositional, phenotypic and/or agronomic analyses and therefore the rationale of the study prior to testing is formulated in form of hypotheses regarding specific endpoints, feeding trials with whole food/feed may provide an added scientific value for the risk assessment of GM crops".* However, this is a conclusion from the GRACE project [appropriate citation is missing] which can't be deduced from the present publication that focuses on feeding trials falling under the EFSA scenario 2. Furthermore, Schmidt *et al.* (2016) fail to give examples on

what might trigger a scientifically justified necessity for a long-term animal feeding trial (e.g., indications of adverse effects from the 90-day feeding study).