SCIENTIFIC REPORT OF EFSA

Considerations on the applicability of OECD TG 453 to whole food/feed testing[1]

European Food Safety Authority[2,3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

Upon request from the European Commission, the European Food Safety Authority prepared a scientific report that would aid the future establishment of protocols for chronic toxicity and/or carcinogenicity studies in rodents with whole food/feed. This scientific report provides a commentary on OECD TG 453 with considerations on its applicability to support the safety assessment of long term consumption of a given food with respect to its chronic toxicity or carcinogenicity potential. The decision to conduct chronic toxicity and/or carcinogenicity studies with whole food/feed should be taken on a case-by case basis. It should be based on the evaluation of all the available information on the whole food/feed resulting from compositional analyses and any other available nutritional and toxicological studies. The conduct of the study and its reporting should be in line with good laboratory practice standards. Preparation of appropriate test diets is a key element of the experiment with respect to characterisation of the starting material and of the diet, level of inclusion of whole food/feed, nutritional balance, processing and storage. Statistical considerations are discussed to assist in estimating the number of animals necessary to obtain a suitable sample size capable of detecting biologically relevant effects with a pre-specified power and significance level. A comprehensive set of endpoints as set out in the OECD TG 453 should be measured during and at the end of the study, as appropriate. The collection of data and reporting should ensure a thorough biological and statistical evaluation. Recommendations on the relevant issues to be considered when designing chronic toxicity and/or carcinogenicity studies in rodents with whole food/feed are provided throughout the report and summarised in the conclusions.

KEY WORDS

2-year feeding study, whole food/feed, carcinogenicity, chronic toxicity, experimental design, statistical analysis

[1] On request from EC, Question No EFSA-Q-2013-00316, approved on 21 March 2013 and endorsed by the EFSA Scientific Committee on 16 July 2013.

[2] Correspondence: scientific.committee@efsa.europa.eu

[3] Acknowledgement: EFSA wishes to thank: the EFSA staff Saghir Bashir, José Cortiñas Abrahantes, Daniele Court Marques, Wolfgang Gelbmann, Andrea Germini, Yi Liu, Claudia Paoletti and Didier Verloo for the preparatory work on this scientific output, and the external reviewers Alicja Mortensen, Hanspeter Naegeli, Joe Perry and Josef Schlatter and the Member State representatives of the EFSA GMO Network for the comments provided to this scientific output.

Suggested citation: European Food Safety Authority, 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

Available online: www.efsa.europa.eu/efsajournal

SUMMARY

Upon request from the European Commission, the European Food Safety Authority (EFSA) was asked to develop guiding principles that would aid the future establishment of protocols for chronic toxicity and/or carcinogenicity studies in rodents with whole food/feed. It was agreed that this EFSA scientific report would be a commentary on OECD TG 453, with specific considerations related to whole food/feed. The mandate was assigned to an internal EFSA task force composed of staff members with different key scientific expertise and was endorsed by the EFSA Scientific Committee. Experts in the role of external reviewers and Member State representatives of the EFSA GMO Network provided valuable comments during the development of the document.

This scientific report comments on the applicability of OECD Test Guideline 453 for the carcinogenicity and chronic toxicity testing of chemicals (OECD, 2009a) for the purpose of testing whole food/feed. The report includes specific recommendations for performing and reporting experiments carried out with whole food/feed. These recommendations are in line with the guiding principles previously developed in the EFSA scientific opinion on conducting repeated-dose 90-day oral toxicity studies in rodents with whole food/feed (EFSA Scientific Committee, 2011a).

OECD Test Guideline 453 was established for the chronic toxicity and carcinogenicity testing of chemicals including the provision of dose-response data on chemicals. Its adaptation to whole food/feed presents limitations that should be considered when developing the test protocol such as: limitations in the maximum whole food/feed dose that can be incorporated into the experimental diet; the fact that toxic effects can be detected only if these are present at levels compatible with a balanced diet formulation; and the availability of proper historical control data.

The evaluation of potential hazards identified by compositional analyses and any other available nutritional and toxicological studies on the whole food/feed would provide indications on the decision to conduct chronic toxicity and/or carcinogenicity testing of a whole food/feed. The study should be carried out according to Good Laboratory Practice (GLP) standards and should take into account animal welfare issues.

Rodents are the preferred test animals and should be housed in cages in pairs of the same sex throughout the duration of the study. The starting material and the diet should be thoroughly characterised and the storage conditions and stability of the feeds should be documented over the course of the whole test.

The decision on the most appropriate study design should be made on a case-by-case basis according to the objectives of the study and the particular hypotheses to be tested. The number of animals per group depends upon the specific objectives of the study and should be adequate to conduct a thorough biological and statistical evaluation. Statistical considerations are discussed that aid the derivation of appropriate sample sizes for continuous and categorical variables.

The parameters to be considered for both the chronic toxicity phase and the carcinogenicity phase should be in line with those detailed in OECD TG 453. Additional markers of potentially adverse nutritional and/or metabolic effects should be considered on a case-by-case basis, according to the available body of evidence and the type of whole food/feed under investigation.

All details of the design, the conduct and the analysis should be clearly justified, documented and reported. A specific chapter on assumptions and uncertainty analysis should be included in the study report.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY THE EUROPEAN COMMISSION

On 25 February 2013, a qualified majority of the Member States voted in favour of the European Commission (EC) proposal on a Regulation concerning applications for the authorisation of GM food and feed which request applicants to carry out an obligatory 90-day oral toxicity study for each submitted GMO. Depending on the outcome of previous studies, a 2-year study in rats may also be requested, on a case-by-case basis.

As of today, no standardised protocol or guidelines exist for this type of study and applicants have to adapt protocols (such as the OECD Test Guideline (TG) 451 protocol) designed for pure chemical substances. Against this background, there is a need for guidance on the design and conduct of 2-year carcinogenicity studies.

In order to provide guidance to applicants on this matter, it is appropriate that the European Food Safety Authority (EFSA) develops a protocol for 2-year carcinogenicity studies. The protocol to be developed is also important to assist a planned DG RTD research project focusing on a 2-year carcinogenicity feeding study in rodents with genetically modified (GM) feed for which a call is expected to be launched soon.

TERMS OF REFERENCE AS PROVIDED BY THE EUROPEAN COMMISSION

EFSA is requested, according to Article 31 of Regulation (EC) No 178/2002, to develop a protocol for a 2-year carcinogenicity study with whole food/feed.

CLARIFICATIONS TO THE TERMS OF REFERENCE

During clarifications between EFSA and the EC it was highlighted that to be scientifically sound a study design for carcinogenicity trial in rats with whole food/feed should take into account previous results of any *in vitro* or *in vivo* toxicity test. Further assessment should only be carried out after initial information on toxicity or potential carcinogenicity has been obtained from 90-day feeding studies or other available nutritional and toxicological studies. Hence, a carcinogenicity trial in rats with whole food/feed would be a confirmatory study with a well-defined objective to be tested. However, EFSA understands that the purpose of the long term feeding trial in rats with whole food/feed as proposed in this mandate is, in contrast, of an exploratory nature and lacks a specific test hypothesis. Regardless of the nature of the study, EFSA is of the opinion that the general objectives of a long-term feeding trial with whole food/feed are consistent with those laid down in the applicable OECD Guideline for the Testing of Chemicals.

In light of the background provided by the EC it was also determined that, if longer term feeding studies are deemed necessary to investigate potential hazards identified, then OECD TG 453, which examines the combined carcinogenicity and chronic toxicity of chemicals, would be more appropriate than OECD TG 451, on the carcinogenicity of chemicals, as a reference guideline for the scope of this report. This is because it would provide a more comprehensive long term assessment of the safety of any given whole food/feed. Nevertheless the final design of the study would depend on the specific objectives and hypotheses, and a combined carcinogenicity and chronic toxicity study might not always result appropriate.

Following the above clarifications it was agreed that this EFSA Scientific Report would be a commentary on OECD TG 453, with specific considerations related to whole food/feed.

CONTEXT OF THE SCIENTIFIC OUTPUT

Risk assessment of whole food for human consumption or of whole feed for animal use is currently based on integrated approaches where information generated through various type of tests on a number of characteristics and properties is required. These include the assessment of source material, production and processing methods, chemical composition, contaminants as well as nutritional and toxicological profile.

In the context of this report, "whole food" or "whole feed" refers to a product to be consumed in its entirety by humans or by animals. Whole food/feed are complex matrices ranging from plant-based products, such as maize or potatoes, through more refined products such as fruit juices or flour, to food and feed consisting of microorganisms, as well as animal-derived food products such as meat and milk (EFSA Scientific Committee, 2011a). The interpretation of the term whole food/feed as used in this report aims to differentiate a whole food/feed from food/feed ingredients or fractions that, in the context of animal testing, could be administered at much higher dietary levels. As such, individual ingredients are outside the scope of this report.

It is considered that where it is necessary to assess the safety of long term consumption of whole food for humans or of whole feed for animals, a 90-day feeding study is required (FAO/WHO, 2000). Additional studies with whole food/feed may be required on a case-by-case basis if compositional analyses, 90-day feeding studies or any other available nutritional and toxicological study have identified potential hazards. Depending on the potential hazard identified, the additional testing of the whole food/feed might need to focus on effects on reproductive or endocrine tissues, effects on development, chronic toxicity, carcinogenicity, etc. Therefore, protocols adapted from standardised guidelines for toxicity testing such as those described in OECD test guidelines, in line with relevant EC directives, are considered the most appropriate for whole food/feed.

Additionally, it has been reported that conventional toxicological feeding trials are of limited value in assessing whole food/feed because of limitations in the upper level of maximum doses of the whole food/feed that can be incorporated into an experimental diet without causing adverse effects due to nutritional imbalance (FAO/WHO, 2000). The main difference between testing chemicals and whole food/feed is that chemicals can be administered to the test animals at dose levels which are much higher than the likely human exposure levels, whereas such a testing approach is not always possible with whole food/feed due to its bulk. In fact, administering high dose levels of whole food/feed is likely to result in satiation and/or unbalanced diets and potential adverse effects observed during administration may then be the result of dietary imbalance rather than of inherent toxicity *per se*. Therefore, careful consideration should be given to designing the control and the experimental diet to prevent nutritional imbalance.

EFSA's scientific activities on animal feeding trials with whole food/feed

EFSA has already published a scientific opinion describing the guiding principles for 90-day whole food/feed studies in rodents (EFSA Scientific Committee, 2011a). There EFSA complements the OECD TG 408 (OECD, 1998) developed for chemicals and provides specific advice for performing and reporting experiments carried out with whole food/feed. The scientific opinion gives guidance on several aspects of 90-day feeding studies in rodents including: the design of the study, the preparation of the test diets, and the sample size to be analysed to detect a pre-specified biologically relevant effect size with a pre-specified power.

In the present scientific report, EFSA comments on the applicability of OECD TG 453 (OECD, 2009a) to support the safety assessment of long term consumption of a given food identifying those aspects of the OECD guideline that are not applicable to whole food/feed. The report summarises and directly quotes both the OECD TG 453 on the combined carcinogenicity and chronic toxicity of chemicals and the EFSA scientific opinion on conducting repeated-dose 90-day oral toxicity study in rodents with whole food/feed. In each section of the present report a summary of the specific OECD

recommendations is followed by EFSA's considerations on its applicability to whole food/feed and by EFSA's complementary recommendations, when needed.

CONSIDERATIONS

1. Introduction

OECD TG 453 provides guidance on how to perform combined chronic toxicity/carcinogenicity studies with chemicals. The present report is intended to comment on the applicability of OECD TG 453 to support the safety assessment of long term human consumption of a given food with respect to potential chronic toxicity or carcinogenicity. In this respect the safety assessment of whole food/feed, as discussed in the present report is intended to address consumer safety, and does not cover the safety assessment of feed for animal use; nevertheless information derived from the studies described in the present report may be informative also for the assessment of feed.

OECD TG 453 applies primarily to studies carried out in rodent species, since the majority of chronic toxicity and carcinogenicity studies are carried out in these species.

The design of the trial as outlined in OECD TG 453 is composed of two parallel phases. The first is a chronic toxicity phase, where the test substance is administered daily in graduated doses to several groups of test animals, one dose level per group, normally for a period of 12 months. This allows any effect of cumulative toxicity to become manifest, without the confounding effects of geriatric changes. The second is a carcinogenicity phase, where the test substance is administered daily in the same doses as in the chronic toxicity phase to several groups of test animals for the majority of their normal life span.

A study carried out according to OECD TG 453 provides information on the major toxic effects of the substance including potential carcinogenicity, and indicates target organs, progressive toxic effects and the possibility of delayed toxicity. The use of a combined test as described in OECD TG 453 provides greater efficiency in terms of time and cost, and some reduction in animal use, compared to conducting two separate studies, without compromising the quality of the data in either the chronic phase or the carcinogenicity phase (OECD, 2012). In the case of whole food/feed the final design of the study would depend on the specific objectives and a combined chronic toxicity/carcinogenicity study might not always be appropriate. If chronic toxicity or carcinogenicity of a whole food/feed needs to be addressed, depending on the specific objectives, it might be more appropriate to perform a standalone chronic toxicity study according to OECD TG 452 (OECD, 2009b) or carcinogenicity study according to OECD TG 451 (OECD, 2009c). The considerations on OECD TG 453 provided in this report with regard to whole food/feed testing would also be applicable in the case of studies performed according to OECD TG 451 and 452.

It should be emphasised that OECD TG 453 was established for the toxicity testing of chemicals to provide dose-response data on the chemical and its adaptation to whole food/feed (consisting of a complex mixture of thousands of components potentially consumed as bulk) requires careful considerations and a case-by-case approach, depending on the whole food/feed to be tested. In fact, in the case of whole food/feed, due to the inherent limitations in dosing the test animals, a proper characterisation of a dose-response relationship is usually difficult or impossible to establish. As a consequence, except for specific whole food/feed (e.g. some type of berries, juices, etc.), it may prove difficult or impossible to identify no-observed-adverse-effect levels (NOAELs) or other toxicologically relevant reference points as further discussed later in the report.

The most relevant exposure route to be considered in the case of whole food/feed is the oral route. Dermal and inhalation exposure are two alternative routes of occupational exposure which are not covered by this report.

In the assessment and evaluation of the potential chronic toxicity and carcinogenicity of whole food/feed, all available information on the test product should be considered by the testing laboratory prior to conducting the study. This would allow focusing the design of the study on specific characteristics to efficiently test the toxicological properties of the food/feed and to minimise animal

usage. It should be emphasized that chronic toxicity and/or carcinogenicity studies should only be carried out after initial information on toxicity has been obtained from repeated dose 28-day toxicity studies, 90-day feeding studies in rodents or other available nutritional or toxicological studies, and when deemed necessary. Additional studies with whole food/feed may be required if potential hazards are identified which warrant further in-depth investigation e.g. as part of a chronic toxicity and/or carcinogenicity study.

A weight of evidence approach, based on all available compositional, nutritional and toxicological information, must be used to inform the decision on the most appropriate toxicological test strategies to be chosen to investigate the potential hazard. The specific objectives of the study and the rationale for choosing a chronic toxicity and/or carcinogenicity study with whole food/feed over alternative *in vitro* or *in vivo* test strategies should be clearly defined before conducting the experiment.

To ensure reliability and comparability of the results, the study should be performed and documented according to the principles of Good Laboratory Practice (GLP) described in Directive 2004/10/EC.[4]

Animal welfare issues should also be taken into account. In particular the testing strategy should adhere to the principles described in Directive 2010/63/EU on the protection of animals used for scientific purposes.[5]

2. Description of the method

a. Selection of animal species

Rodents are the preferred animals to be used in long term studies, due to their relatively short life span, their widespread use in such studies, their susceptibility to tumour induction, their relative ease of handling and the availability of well characterised strains. According to OECD TG 453 rats are the preferred species; however the use of mice should not be precluded, even if it may have limitations. The use of non-rodent species is limited to specific cases and should be justified based on the principles of OECD TG 453 together with those of OECD TG 409, repeated dose 90-day oral toxicity study in non-rodents (OECD, 2009a).

OECD TG 453 recommends that young healthy adult animals (shortly after weaning) not subjected to previous experimentation should be used; the females should be nulliparous and non-pregnant. Consideration should be given to using a strain of animal that has an acceptable survival rate for the long-term study. The study should be carried out in animals from the same strain and source as those used in the preceding toxicity study(ies) of shorter duration, unless scientifically justified (OECD, 2009a).

EFSA agrees with the OECD recommendations with regard to the selection of animal strain.

b. Housing and feeding

According to OECD TG 453, animals should be caged in small groups of the same sex, although individual housing may be considered if scientifically justified (e.g. in the case of male mice). Temperature and relative humidity have to be monitored throughout the study; sequential lighting of 12 hours light, 12 hours dark is maintained through artificial light. Feeding with conventional laboratory diets may be used with an unlimited supply of drinking water. Diet and water have to be analysed periodically regarding contaminant levels such as pesticide residues, persistent organic

[4] Directive 2004/10/EC of the European Parliament and of the Council on the harmonisation of laws, regulations and administrative provisions relating to the application of the principles of good laboratory practice and the verification of their applications for tests on chemical substances. OJ L 50, 20.2.2004, p. 44-59.

[5] Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes. OJ L 276, 20.10.2010, p. 33-79.

pollutants, phytoestrogens, heavy metals and mycotoxins. The diet has to meet all nutritional requirements of the species tested and must be periodically analysed for its nutrients levels; the choice of the diet may still be influenced by specific needs regarding a suitable admixture of the test substance (OECD, 2009a).

Rodents are social animals; therefore EFSA proposes that animals of the same sex are housed in cages in pairs, unless scientifically justified. A group of animals within a cage is referred to statistically as the experimental unit (ExpU). The ExpUs are assigned to the treatments at random, and it must be possible for any two ExpU to receive different treatments. Animals in the same cage cannot receive different treatments when these are supplied in the diet. The experiment should be designed to be unbiased, with no systematic differences among groups apart from the treatment. This is mainly controlled by appropriate randomisation and by using coding (where practical) to blind the staff to the treatment group to which an ExpU belongs. Further details on the principles for experimental design can be found in (EFSA Scientific Committee, 2011a).

EFSA agrees with the OECD recommendations regarding feeding and suggests that in the case of testing with whole food/feed a thorough characterisation of the starting material and of the diet should be performed (EFSA Scientific Committee, 2011a). Particular attention should be paid to the analysis of naturally occurring contaminants and the residues resulting from specific agronomic treatments (together with their relevant metabolites) the amounts of which should be documented in the analysis report. The starting test material used in the preparation of the animal diet should be representative of the final product to which humans or animals would be exposed in terms of agricultural practices and/or industrial processing.

c. Preparation of animals

According to OECD TG 453, animals have to undergo acclimatisation to laboratory conditions for at least 7 days. The animals should be characterised by species, strain, source, sex, weight and age. Dosing should begin as soon as possible after weaning and acclimatisation. For rodents, preferably before the animals are 8 weeks old. At the beginning of the study, body weight variation should not exceed 20 % of the mean weight of all animals by sex, and after randomisation and assignment to the control and treatment groups, there should be no differences of biological concern in mean body weights between groups within each sex. Each animal should be assigned with a unique identification number (OECD, 2009a).

EFSA finds the recommendations of OECD TG 453 for the preparation of animals directly applicable in the case of whole food/feed testing.

3. Procedure

a. Dose groups and dosage

OECD TG 453 refers to OECD Guidance Document (GD) No. 116 (OECD, 2012) for all aspects of dose selection and dose level spacing. According to OECD GD 116 at least three dose levels and a concurrent control should be used when testing chemicals, for both the chronic toxicity and carcinogenicity phases. The selection of dose levels will generally be based on the results of shorter-term repeated dose or range finding studies, and should take into account any existing toxicological and toxicokinetic data available for the test substance or related materials. Dose levels and dose level spacing may be selected to establish a dose-response, a NOAEL or other intended outcome of the study. The selection of dose level spacing will depend on the objectives of the study and the characteristics of the test item. However OECD guidelines suggest that two to four-fold intervals frequently provide good test performance and that the use of factors greater than 10 should be avoided or justified if used. The control group may be an untreated group or a vehicle-control group if a

vehicle is used to administer the test item. Animals in the control group should be handled in an identical manner as those in the test groups (OECD, 2009a).

EFSA does not recommend any specific dose group and/or dose selection approach, which should be defined on a case-by-case basis according to the objectives of the study and the specific hypotheses to be tested.

Given the inherent limitations on dosing test animals with whole food/feed, a dose-response relationship is often difficult or impossible to establish. Furthermore, in the case of whole food/feed toxic effects can be detected only if these are present at levels compatible with the diet formulation.

The considerations provided by OECD GD 116 on dose level spacing and on how to derive an appropriate mid dose level may be applicable to whole food/feed in some specific cases (e.g. leaves, juice concentrates, seeds, etc.). However, for the great majority of whole food/feed products these considerations are not applicable due to the bulky nature of the food and due to the high human intake estimates. These factors may narrow the space between the high and the low dose group to such an extent that the OECD GD 116 considerations for the mid dose group (i.e. half or quarter of the high dose, or the geometric mean of the low and high dose) become either of limited additional value for the experiment, or not applicable. In such cases EFSA recommends the use of two dose levels, a high dose and a low dose, together with the concurrent control group(s).

According to OECD TG 453, for the chronic toxicity phase of the study, three dose levels may not be necessary if it can be anticipated that testing at a dose level equivalent to at least 1000 mg/kg body weight/day is unlikely to produce adverse effects. In the case of chemicals this can be established considering information from preliminary studies and lack of toxicity documented by data from structurally related substances. For whole food/feed however, human exposure may indicate the need for a higher dose level, as the limit of 1000 mg/kg applies only in rare cases.

In the case of testing of chemicals for chronic toxicity, according to OCED (2000) the highest dose level should be chosen so that toxic effects, including the principal target organs, can be identified while avoiding severe toxicity, morbidity, or death. For chemicals, this is usually possible by increasing the dose up to the maximum tolerated dose. This approach is not applicable to whole food/feed, if no toxic effects are anticipated and the only possibility to increase the power of a study to detect effects would be to increase the number of experimental units in the study (see section 3c).

In the case of whole food/feed testing, the highest 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 which might preclude the correct evaluation of results. The lowest dose level should always be at least at a level comparable to the anticipated (high percentile) human intake. High incorporation levels of whole food/feed in the animal diet can be achieved, depending on the specificities of the test whole food/feed, provided that the diet is properly nutritionally balanced. When the diet is balanced, potential variability arising from the difference in composition of one or several nutrients is minimised/removed, which is a prerequisite for the detection of unintended effects (EFSA GMO Panel, 2008).

Considerations should also be given as to whether or not the whole food/feed contains inherent anti-nutritional components in a relatively high concentration (e.g. trypsin inhibitor in unprocessed soybean meal; glyco-alkaloids in potatoes; gossypol in cotton; etc.). This can be predicted from compositional analysis data, a review of the literature or preliminary studies and should be taken into account in the diet formulation. The presence of anti-nutritional components, or other substances, in the whole food/feed to be tested may be the limiting factor that determines its maximum inclusion level into the test diet (EFSA Scientific Committee, 2011a).

For whole food/feed where no adequate information exists, a small preliminary tolerance test with a limited number of animals should be performed. Pilot shorter-term studies may need to be conducted

on a case-by-case basis to investigate the appropriateness of the dose levels and control group(s) to be used in the study.

When testing GM crops, in the case that an appropriate conventional counterpart is available (EFSA GMO Panel, 2011), it is recommended to include this conventional counterpart in the formulation of the diet of the control group(s) to minimize possible confounding effects due to the use of different ingredients in the diets. When an appropriate conventional counterpart for the GM plant is not available (e.g. GM plants expressing quality traits possibly leading to extensive compositional alterations; GM plants expressing new traits which facilitate adaptation to environmental stress conditions; etc.) the choice of suitable control groups should take into account the characteristics of the GM plant and its derived products (EFSA GMO Panel, 2011). The inclusion of additional control groups, fed diets based on non-GM commercial varieties in order to estimate the variability of the endpoints, is in general not recommended for two reasons. Firstly, this would substantially increase the number of test animals. Secondly, the identification of differences in the group fed the GM diet(s) is based on the comparison with the concurrent control group(s). The relevance of these differences is a biological issue, which should be addressed by toxicologists according to their likelihood of raising safety concerns (EFSA Scientific Committee, 2011b).

Similar considerations apply when testing novel foods. In this case control diets should be selected on a case-by-case basis, depending on the nature of the novel food to be tested and on considerations provided in the respective Commission Recommendation 97/618/EC.[6]

Information on the variability of the endpoints should primarily be obtained from properly collected historical control data, from the same testing facility, and using data collected no more than five years prior to the study. Nonetheless, the use of historical control data should be considered with caution. The historical controls might not be useful because the incidences of neoplastic (or non-neoplastic) lesions would possibly be from control animals kept on different diets than the diet applied in whole food/feed study, and because the diet itself (high/low fat, type of fat, % of carbohydrate, type of carbohydrate, etc.) can influence the formation of neoplastic or non-neoplastic lesions. Where the diet formulation used in the experiment for the control groups cannot be demonstrated to be equivalent to that used for the generation of historical control data, the inclusion may be considered of an additional control group (as similar as possible to the historical controls), in addition to the concurrent control group(s).

b. Preparation and administration of doses

According to OECD TG 453, the preferred route of exposure is the oral administration via the diet, drinking water or by gavage. If the test substance needs to be dissolved or suspended in a suitable vehicle, the use of an aqueous dilution/suspension is preferred and considerations should be given to the characteristics of the vehicle, its potential influence on toxicokinetics, toxicity of the test compound, or on food or water consumption, or on the nutritional status of the animals. For vehicles other than water, the toxicological properties of the vehicle should be known. Information should be given on the stability and homogeneity of the test substance in the dosing solution or diet (OECD, 2009a).

EFSA is of the view that these recommendations of OECD TG 453 are applicable also in the case of whole food/feed. In addition in the case of whole food/feed it is recommended that animals should be fed *ad libitum*. Where the whole food/feed is given alongside the diet, or by gavage, the same kind of consideration about balancing the diet should be taken into account as in the case where the whole food is incorporated into the diet. Administration by gavage is not common for whole food/feed, but

[6] 97/618/EC: Commission Recommendation of 29 July 1997 concerning the scientific aspects and the presentation of information necessary to support applications for the placing on the market of novel foods and novel food ingredients and the preparation of initial assessment reports under Regulation (EC) No 258/97 of the European Parliament and of the Council. OJ L 253, 16.9.97, p. 1-36.

could be considered in certain instances such as poor palatability or poor stability of the feed or in cases where an exact dosing is needed.

Detailed information on the storage conditions and stability of the feed over the course of the whole test should be provided.

c. Number and sex of animals

OECD TG 453 recommends that the number of animals should be adequate to conduct a thorough biological and statistical evaluation of the results and that both male and female animals should be tested.

According to OECD TG 453, for rodents, the following recommendations apply when testing chemicals:

- for the carcinogenicity phase of the study, a minimum of 50 animals per sex per dose group and concurrent control group should be used;
- for the chronic toxicity phase, a minimum of 10 animals per sex per dose group should be used, which corresponds to the number of animals needed for satellite and/or interim kills groups.

The observations carried out on all animals on both phases of the study are then combined (OECD, 2009a).

EFSA emphasizes that part of the planning phase of the study should include defining clear objective(s) and a prospective power analysis in order to establish the number of animals needed for the experiment. This requires selecting a specific effect size of interest and estimating the standard deviation (in case of continuous variables). However, it is often difficult to establish an effect size of interest for every endpoint because of the multiplicity of endpoints usually assessed in toxicity testing. Further the chosen effect size should be biologically relevant (EFSA 2011b). EFSA recommends that a statistician is consulted from the start of the study planning.

One possible approach was proposed by EFSA (EFSA Scientific Committee, 2011a) for the analysis of continuous variables (e.g. biochemical parameters). It is to change the scale on which effects are measured to define an effect in terms of standard deviations. The resulting "standardised effect size" (SES) is the ratio of the variation between treatment groups divided by the standard deviation (SD) among experimental units (ExpUs). The SES may be regarded as a signal/noise ratio. Such an approach is suitable for the analysis of continuous variables, but is not applicable to categorical variables, such as tumour prevalence, or to survival (time to event) analysis. In the case of categorical variables alternative methods are needed to allow the estimation of appropriate sample sizes, depending on the specific objectives of the study.

Below an example is provided of a method to estimate appropriate sample sizes in order to find a difference in tumour prevalence assuming that one sex only is to be tested and that the design includes two dose levels and two concurrent control groups, when rats are housed in pairs within a cage. In this case, it is the cage which is termed the 'experimental unit', because randomisation of treatments (i.e. the diets being compared) to experimental units can only occur at the cage level and not at the individual rat level. If either of the two or both rats have a tumour then the cage (ExpU) prevalence is unity; otherwise it is zero. The calculation is based on assuming a binomial distribution of tumour occurrence (Armitage, 2002 – Section 4.6). For mathematical reasons, it is suggested to use ExpUs consisting of two animals per cage but, for ease of interpretation, the cage prevalence rates (π_1) are converted to animal prevalence rates (π_{rat}), assuming that animals within cage are independent, using the following formula:

$$\pi_{rat} = 1 - \sqrt{(1 - \pi_1)}$$

The example is presented for a design with four groups: two treatment groups at different doses and two control groups. It is emphasized that this example is for illustration only; other designs are possible and EFSA does not recommend any specific experimental design in this commentary. Table 1 presents in bold type the number of ExpUs needed based on control group tumour prevalence (i.e. prevalence among individual rats) in order to ensure that a two-sided test of whether the treatment group prevalence is higher by a defined percentage amount (Detectable difference) than the control group prevalence, has a statistical power of 80 % at a significance level of 5 %. This is done for a range of expected prevalences of affected rats in the control group (Control group tumour prevalence). For example, detecting a difference of 10 % with a control group prevalence of 15 % (i.e. a treatment group prevalence of less than 5 % or greater than 25 %) would require 140 ExpUs (assuming 2 rats per cage). In the example provided, 140 ExpUs corresponds to a total number of 1120 rats for an experiment in one sex, with two treatments and two control groups.

Table 1: Examples of the number of experimental units and the total number of rats (in brackets) needed to perform an experiment as function of the detectable difference between one treatment group and its concurrent control group. The design includes two groups treated at different dose levels and two concurrent control groups for one sex. The statistical power is 80 % and the significance level is 5 %.

	Number of experimental units per group (Total number of animals in the experiment)			
Detectable difference	**Control group tumour prevalence**			
	5 %	15 %	30 %	45 %
1 %	**4 197** (33 576)	**11 222** (89 776)	**20 288** (162 304)	**27 588** (220 704)
5 %	**225** (1 800)	**500** (4 000)	**854** (6 832)	**1 138** (9 104)
10 %	**73** (584)	**140** (1 120)	**226** (1 808)	**296** (2 368)
20 %	**26** (208)	**42** (336)	**63** (504)	**80** (640)
30 %	**14** (112)	**22** (176)	**31** (248)	**40** (320)

The example provided in this section is for a design with a clear pre-specified hypothesis (i.e. testing for a difference in the prevalence rates, at a particular time point between a treatment group and its concurrent control group). Adding groups or other factors that change underlying assumptions (e.g. multiple testing) could inflate the sample size. The issue of sex, especially in relation to sex specific tumours, should also be addressed and justified.

Because of the limitations in dosing animals with whole food/feed discussed in Section 3.a, the magnitude of the detectable differences that can be observed in these studies is generally expected to be smaller than those typically detectable when testing chemicals. Therefore in order to provide a meaningful study design to reliably detect small differences with enough statistical power, a larger number of animals would usually be necessary when testing whole food/feed than when testing chemicals.

The decisions made during the planning phase of the study design and the sample size needed for the experiment should be justified and documented in the report.

In the case of comparison of tumour incidences among groups OECD GD 116 strongly advocates for the use of survival adjusted methods in order to take into consideration survivability of the animals (OECD, 2012); details on the methods are provided in OECD GD 116 and consideration on their implementation in the experiment should be given at the design stage.

Further considerations on the power of the experiment and the use of blocking to help control variability and increase power, which are also applicable in the case of long term feeding trials with whole food/feed, are provided by the EFSA Scientific Committee (2011a).

d. Provision for interim kills, satellite groups and sentinel animals

According to OECD TG 453 interim kills may be used, e.g. at 6 months for the chronic toxicity phase, to provide information on the progression of non-neoplastic changes and mechanistic information, if scientifically justified. The animals used in the chronic toxicity phase of the study, normally lasting 12 months, provide interim kill data for the carcinogenicity phase of the study. Satellite groups, normally limited to the highest dose level tested and the control group, may also be included to monitor the reversibility of any toxicological change induced by the chemical under investigation. An additional group of sentinel animals (typically 5 animals per sex) may be included to monitor the disease status during the study. Interim kill and satellite animals should normally undergo the same observations, as the animals in the chronic toxicity phase, although provision may also be made (in the interim kill groups) to restrict the measurements to specific key measures such as neurotoxicity or immunotoxicity (OECD, 2009a).

EFSA is of the view that the principles recommended by OECD TG 453 are directly applicable in the case of whole food/feed testing. The number of animals for satellite and interim kills groups, if included, should be justified.

e. Duration of study

According to OECD TG 453 the period of dosing and duration is normally 12 months for the chronic toxicity phase, and 24 months for the carcinogenicity phase in the case of rodents, although for specific strains of mice duration of 18 months for the carcinogenicity phase may be more appropriate. Deviations should be justified, particularly in case of shorter durations. All dose groups allocated to the chronic phase should be terminated at the designated time for evaluation of chronic toxicity and non-neoplastic pathology. Satellite groups, included to monitor the reversibility of effects, should be maintained without dosing for a period of between 4 weeks and one third of the total study duration after cessation of exposure. The duration of the carcinogenicity phase should represent the majority of the normal life span of the animals used. Termination of the study may be considered under certain circumstances, when the number of survivors in the lower dose groups or the control group falls below 25 % or when the data available are no longer sufficient to enable a statistically valid evaluation. Survival should be considered separately for each sex (OECD, 2009a).

EFSA is of the view that the recommendations of OECD TG 453 are directly applicable also in the case of whole food/feed.

4. Observations

Details of the parameters to be considered for both the chronic toxicity phase and the carcinogenicity phase in terms of

i. clinical signs, ophthalmological examination, body weight, food/water consumption and food efficiency

ii. haematology and clinical biochemistry

iii. pathology, including gross necropsy and histopathology

are listed in the relevant section of OECD TG 453 (OECD, 2009a).

EFSA finds the recommendations of OECD TG 453 directly applicable also in the case of whole food/feed. Additional endpoints described in the OECD TG 407 on repeated-dose 28-day oral toxicity study in rodents (OECD, 2008) may be considered for assessment, depending on the nature of the food/feed being tested and the available information. Additional markers of potentially adverse nutritional and/or metabolic effects should be considered on a case-by-case basis, according to objectives of the experiment, the available body of evidence and the type of whole food/feed under investigation.

5. Data and reporting

OECD GD 116 provides guidance on the statistical issues associated with the design of animal feeding studies for chronic toxicity and carcinogenicity, the analysis of the resulting data and the interpretation of the outcomes. EFSA finds the recommendations given in OECD GD 116 applicable also in the case of whole food/feed.

a. Data

OECD TG 453 recommends the use of appropriate statistical methods for data analysis that should be defined at the design stage. Individual animal data and comprehensive summary statistics should be provided for all parameters. EFSA finds the recommendations of OECD TG 453 directly applicable also in the case of whole food/feed. Furthermore EFSA Scientific Committee (2011a) recommends that all the details of the design, the conduct and the analysis should be clearly documented (i.e. in a protocol, statistical analysis plan and study report). Descriptive (summary) statistics should also be presented, both on the original and standardised effect size scales. Summary statistics should be provided by treatment group and by sex.

According to OECD guidelines, if information on the variation in endpoint values is required, it should primarily be obtained from historical control data collected from the same testing facility, generated in similar experiments during the five years preceding the study in question and relate to animals of the same age and strain.

Further guidance on the use of appropriate statistical methods for data analysis is given in Section 6 of EFSA Scientific Committee guidance (2011a) and in OECD GD 116 (OECD, 2012). It is recommended that the issues of animal flow through the experiment, missing data and multiplicity are also addressed.

b. Analysis report

OECD GD 116 recognises that interpreting the results of a long-term bioassay is complex. A critical issue is the practical problem of the low power of the design when the tumour incidence is rare together with the multiple comparisons issues arising from the investigation of several tissues possibly from both sexes. As a result, there is a risk of both Type I (false positives) and Type II (false negative) errors. OECD GD 116 also recognises the necessity to integrate the results of the full battery of statistical tests, significant or otherwise, and the importance of a series of biological issues in the assessment of the result (OECD, 2011).

The recommendations provided in OECD GD 116 and TG 453 with regard to data interpretation and reporting are also relevant in the case of whole food/feed. Further guidance on the interpretation of results of animal studies is given in Section 7 of EFSA Scientific Committee guidance (2011a).

A specific chapter on assumptions and uncertainty analysis should be included in the study report. Any uncertainties in the design of the experimental model which might influence the power of the experiment should be highlighted and quantified as far as possible. The assumptions underlying the statistical analysis should be reported and tested for robustness (EFSA Scientific Committee, 2011a).

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

In the present scientific report EFSA provides a commentary on OECD TG 453 with considerations on its applicability to support the safety assessment of long term consumption of a given food with respect to its chronic toxicity or carcinogenicity potential.

EFSA recognises that the general principles applicable for the design of chronic toxicity and/or carcinogenicity studies with whole food/feed are laid down in the OECD Test Guideline 453 for the chronic toxicity and carcinogenicity testing of chemicals (OECD, 2009a) and in the OECD Guidance Document 116 on the conduct and design of chronic toxicity and carcinogenicity studies (OECD, 2012). Furthermore guiding principles on the performance of animal feeding studies with whole food/feed were previously provided by EFSA in its guidance on conducting repeated-dose 90-day oral toxicity studies in rodents with whole food/feed (EFSA Scientific Committee, 2011a); some of these are also applicable in the case of chronic toxicity and/or carcinogenicity studies.

EFSA emphasizes that the implementation of the guiding principles described in the present scientific report into a test protocol for performing chronic toxicity and/or carcinogenicity studies in rodents is strictly dependent on the objectives of the study. These objectives should be defined based on the potential hazards identified by the compositional analyses of the whole food/feed, by 28-day and/or 90-day feeding studies or by any other available nutritional and toxicological study. It should be emphasized that the final design of the study would depend on the specific objectives and a combined chronic toxicity/carcinogenicity study might not always be appropriate.

OECD TG 453 was established for the toxicity testing of chemicals to provide dose-response data on the chemical. In its adaptation to whole food/feed, several limitations have to be taken into consideration when developing the test protocol.

The highest whole food/feed dose that can be incorporated into the experimental diet is one of the critical factors posing limitation when designing protocols for whole food/feed testing. To account for the limitation in dosing whole food/feed, a larger number of animals would usually be necessary when testing whole food/feed than when testing chemicals in order to provide a meaningful study design with enough statistical power.

Furthermore in the case of whole food/feed, toxic effects can be detected only if these are present at levels compatible with the diet formulation.

It should also be considered that suitable historical control data informing on the natural variability of the endpoints measured in the study may not be available due to the specificities of whole food/feed testing. In this respect EFSA emphasizes that the interpretation of identified differences between test and control groups is primarily a biological issue, which should be addressed by toxicologists according to the magnitude of those differences and their likelihood of raising safety concerns.

RECOMMENDATIONS

EFSA specifically recommends taking into account the following principles when planning chronic toxicity and/or carcinogenicity studies in rodents with whole food/feed:

- The decision to conduct chronic toxicity and/or carcinogenicity studies with whole food/feed should be taken on a case-by case basis. It should be based on the evaluation of all the available information on the whole food/feed resulting from compositional analyses and any other available nutritional and toxicological studies. All aspects of the study should be fully justified and clearly documented.

- The specific objectives and the rationale for choosing a chronic toxicity and/or carcinogenicity study over alternative *in vitro* or *in vivo* experimental models should be clearly defined and documented before conducting the experiment.

- The study should be performed and documented according to GLP principles and should take into account animal welfare issues.

- Rodents, and in particular rats, are the preferred test animals for the study and should be housed in cages in pairs of the same sex throughout the duration of the study. Alternative options may be acceptable if scientifically justified. The reason for choosing a particular animal strain or stock should be justified.

- A thorough characterisation of the starting material and of the diet should be performed. Storage conditions should be recorded and documented. The stability of the feeds over the course of the whole test should be ensured. The starting material should be representative of the final product to which humans or animals would be exposed.

- The strategy for dosing animals should take into account the considerations provided in OECD GD 116, though for most cases two dose groups fed with the whole food/feed, together with the concurrent control group(s), will be appropriate. No specific dose group and/or dose selection approach is recommended; these should be defined on a case-by-case basis according to the objectives of the study and the particular hypotheses to be tested.

- Experiments and statistical analysis should always be pre-planned. All the objectives of the study, the hypotheses to be tested, the details of the design, the conduct and the analysis should be fully justified and clearly documented (i.e. in a protocol, statistical analysis plan and study report).

- The study should be appropriately randomised and (where practical) blinded to ensure that the experiment is unbiased.

- A prospective power analysis should be performed in order to establish the number of animals needed for the experiment. The number of animals per group, which will depend upon the specific objectives of the study, must be adequate to conduct a thorough biological and statistical evaluation.

- The parameters to be considered for both the chronic toxicity phase and the carcinogenicity phase should be in line with those detailed in OECD TG 453. Additional markers of potentially adverse nutritional and/or metabolic effects should be considered on a case-by-case basis, according to the available body of evidence and the type of whole food/feed under investigation.

- Historical control data, where appropriate, should be used to inform on the variability of endpoint values.

- A specific chapter on assumptions and uncertainty analysis should be included in the study report.

REFERENCES

Armitage P, Berry G and Matthews JNS, 2002. Statistical Methods in Medical Research. 4th ed. Blackwell, Oxford, UK, 817pp.

EFSA GMO Panel, 2008. Safety and nutritional assessment of GM plants and derived food and feed: The role of animal feeding trials. Food and Chemical Toxicology, 46, S2–S70.

EFSA GMO Panel, 2011. Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal 2011;9(5):2150, 37 pp. doi: 10.2903/j.efsa.2011.2150.

EFSA Scientific Committee, 2011a. 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 Scientific Committee, 2011b. Statistical Significance and Biological Relevance. EFSA Journal 2011;9(9):2372, 17 pp. doi:10.2903/j.efsa.2011.2372.

FAO/WHO (Food and Agriculture Organization of the United Nations / World Health Organisation), 2000. Safety aspects of genetically modified foods of plant origin. Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, 29 May - 2 June 2000 Geneva, Switzerland.

OECD (Organisation for Economic Co-operation and Development), 1998. Repeated Dose 90-Day Oral Toxicity Study in Rodents. Test Guideline No. 408, Guidelines for the Testing of Chemicals, OECD, Paris.

OECD (Organisation for Economic Co-operation and Development), 2000. Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Human Endpoints for Experimental Animals Used in Safety Evaluation. Series on Testing and Assessment No. 19, OECD, Paris.

OECD (Organisation for Economic Co-operation and Development), 2008. Repeated Dose 28-day Oral Toxicity Study in Rodents. Test Guideline No. 407, Guidelines for the Testing of Chemicals, OECD, Paris.

OECD (Organisation for Economic Co-operation and Development), 2009a. Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing.

OECD (Organisation for Economic Co-operation and Development), 2009b. Test No. 452: Chronic Toxicity Studies, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing.

OECD (Organisation for Economic Co-operation and Development), 2009c. Test No. 451: Carcinogenicity Studies, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing.

OECD (Organisation for Economic Co-operation and Development), 2012. Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting test guidelines 451, 452 and 453, 2nd edition, OECD, Paris.